

| | |
|--------------|---|
| Title | 分子動力学を用いたユビキチン分子の熱力学的挙動に関する考察 |
| Author(s) | 春田, 篤志 |
| Citation | |
| Issue Date | 2001-03 |
| Type | Thesis or Dissertation |
| Text version | author |
| URL | http://hdl.handle.net/10119/708 |
| Rights | |
| Description | Supervisor:小長谷 明彦, 知識科学研究科, 修士 |

修 士 論 文

分子動力学を用いたユビキチン分子の
熱力学的挙動に関する考察

指導教官 小長谷 明彦 教授

北陸先端科学技術大学院大学
知識科学研究科知識システム基礎学専攻

950072 春田 篤志

審査委員： 小長谷 明彦 教授（主査）
佐藤 賢二 助教授
本多 卓也 教授

2001 年 2 月

目次

| | |
|--------------------------------|----|
| 第1章 はじめに..... | 1 |
| 1.1 研究の背景と目的 | 1 |
| 1.2 本論文の構成..... | 2 |
| 第2章 タンパク質の立体構造・機能とダイナミクス | 3 |
| 2.1 はじめに..... | 3 |
| 2.2 タンパク質の立体構造と機能..... | 3 |
| 2.3 タンパク質のフォールディング | 5 |
| 2.4 シミュレーションでのモデル化..... | 7 |
| 2.5 分子動力学について..... | 9 |
| 2.5.1 分子動力学概説..... | 9 |
| 2.5.2 運動方程式..... | 11 |
| 2.5.3 数値積分 | 11 |
| 2.5.4 境界条件 | 12 |
| 2.5.5 原子間ポテンシャル | 13 |
| 2.5.6 熱力学的物理量..... | 15 |
| 第3章 ユビキチンの MD シミュレーション | 16 |
| 3.1 はじめに..... | 16 |
| 3.2 ユビキチンプロテアソーム系..... | 16 |
| 3.2.1 タンパク質分解系 | 16 |
| 3.2.2 ユビキチン様タンパク質修飾システム | 19 |
| 3.3 数値実験モデル | 20 |
| 3.3.1 数値実験方法 | 20 |

| | | |
|--------------|------------------------------|-----------|
| 3.3.2 | 数値実験で用いた式 | 22 |
| 3.3.3 | トラジェクトリの解析 | 23 |
| 3.4 | シミュレーションパラメータ | 24 |
| 3.4.1 | 初期構造 | 24 |
| 3.4.2 | 数値実験でのシミュレーション時間 | 24 |
| 3.4.3 | 炭素原子の材料定数 | 24 |
| 3.5 | 数値実験結果 | 25 |
| 3.5.1 | Ubi の NVE アンサンブル | 25 |
| 3.5.2 | 各タンパク質の主鎖安定構造 | 26 |
| 3.6 | 考察 | 29 |
| 第 4 章 | 分子力学の並列化 | 30 |
| 4.1 | 並列化の必要性 | 30 |
| 4.2 | 並列処理について | 31 |
| 4.2.1 | 並列コンピューティングモデル | 31 |
| 4.2.2 | 並列化効率 | 32 |
| 4.3 | 分子力学の並列化 | 32 |
| 4.3.1 | 粒子分割法 | 33 |
| 4.4 | ベンチマーク | 34 |
| 4.4.1 | 用いたクラスタ構成 | 34 |
| 4.4.2 | 原子数に関してのベンチマーク | 35 |
| 4.4.3 | Time Step に関してベンチマーク | 36 |
| 4.4.4 | NAS Parallel Benchmark | 37 |
| 4.5 | 考察 | 39 |
| 第 5 章 | まとめ | 40 |
| | 参考文献 | 43 |
| | 付録 | 45 |

目 次

| | | |
|--------|--|----|
| 図 2.1 | CATH によるタンパク質の分類 | 5 |
| 図 2.2 | フォールディング・ファネル..... | 6 |
| 図 2.3 | 理論と実験, シミュレーション | 8 |
| 図 2.4 | タンパク質モデル吟味 | 8 |
| 図 2.5 | 分子力学の手順 | 10 |
| 図 2.6 | 原子に働く力..... | 11 |
| 図 3.1 | ユビキチン-プロテアソーム系 | 17 |
| 図 3.2 | ユビキチン化のしくみ | 18 |
| 図 3.3 | ユビキチンと Ubl (SUMO-1, NEDD8) | 19 |
| 図 3.4 | Ubiquitin (PDB Entry : 1D3Z) | 21 |
| 図 3.5 | SUMO-1 (PDB Entry : 1A5R)..... | 21 |
| 図 3.6 | NEDD8 (PDB Entry : 1NDD)..... | 21 |
| 図 3.7 | NVE アンサンブルでの Energy (Ubiquitin)..... | 25 |
| 図 3.8 | ユビキチン主鎖安定構造..... | 26 |
| 図 3.9 | ユビキチン 構造揺らぎ | 26 |
| 図 3.10 | SUMO-1 主鎖安定構造 | 27 |
| 図 3.11 | SUMO-1 構造揺らぎ | 27 |
| 図 3.12 | NEDD8 主鎖安定構造..... | 28 |
| 図 3.13 | NEDD8 構造揺らぎ | 28 |
| 図 4.1 | 並列計算の方法 | 31 |
| 図 4.2 | 粒子分割法 | 33 |
| 図 4.3 | 原子数に関するの応答時間 | 35 |
| 図 4.4 | TimeStep に関するの応答時間 | 36 |
| 図 4.5 | 様々な計算機の NPB | 37 |
| 図 4.6 | NPB(EP) on Alta Cluster..... | 38 |

| | | |
|--------|----------------------|----|
| 付録図 1 | Ubiquitin at 400 [K] | 45 |
| 付録図 2 | Ubiquitin at 300 [K] | 46 |
| 付録図 3 | Ubiquitin at 200 [K] | 47 |
| 付録図 4 | Ubiquitin at 100 [K] | 48 |
| 付録図 5 | Ubiquitin at 0 [K] | 49 |
| 付録図 6 | SUMO-1 at 400 [K] | 50 |
| 付録図 7 | SUMO-1 at 300[K] | 51 |
| 付録図 8 | SUMO-1 at 200[K] | 52 |
| 付録図 9 | SUMO-1 at 100[K] | 53 |
| 付録図 10 | SUMO-1 at 0[K] | 54 |
| 付録図 11 | NEDD8 at 400 [K] | 55 |
| 付録図 12 | NEDD8 at 300 [K] | 56 |
| 付録図 13 | NEDD8 at 200 [K] | 57 |
| 付録図 14 | NEDD8 at 100 [K] | 58 |
| 付録図 15 | NEDD8 at 0 [K] | 59 |

表 目 次

| | | |
|-------|------------------------|----|
| 表 3-1 | NVT アンサンブルでの計算条件 | 24 |
| 表 3-2 | 材料定数..... | 24 |
| 表 3-3 | 定数..... | 24 |
| 表 4-1 | クラスタ構成 | 34 |

第 1 章

はじめに

1.1 研究の背景と目的

今日のヒトゲノム解析に代表されるように、さまざまな生物種の塩基配列が明らかにされ、膨大な量の DNA 塩基配列、タンパク質のアミノ酸配列がゲノムデータベースに蓄積されるようになってきた。しかし、配列を読み取ったからといって生命の営みのしくみが明らかになったわけではない。現在では、これらのゲノムに書き込まれている機能を解明する研究がはじまっている。生命活動は、タンパク質の相互作用で成り立っており、タンパク質は細胞内においてダイナミックな構造変化をしている。生命の営みを明らかにするにはこのタンパク質の働きを知ることが非常に重要である。そのためには、タンパク質の機能にかかわる立体構造、そしてフォールディング(折りたたみ)過程を明らかにすることが必要である。最近の研究では、がん、脳疾患といった病気はユビキチン-プロテアソーム系と呼ばれるタンパク分解システムの異常で起こることが知られている。この分解系においてユビキチン(Ub)はユビキチン化と呼ばれる一連の複合酵素反応で標的タンパク質のリシン(Lys)残基にイソペプチド結合し、またユビキチン相互でこの反応を繰り返し結合する。このユビキチン化修飾の意味は、標的タンパク質に分解のシグナルを付与することである。また、ユビキチン様タンパク質(Ubl)と呼ばれるユビキチンと相同性のあるタンパク質も様々なタンパク質と結合し、タンパク質の機能制御に関与することが明らかになってきた[1]。しかし、ユビキチンシステムがいかにして標的分子を的確に選別するかという分子機構は未だ明らかにされていない。それを解明する有力なアプローチの一つとして、タンパク質の一つ一つの原子についてまで熱力学的

挙動を解析することができる分子動力学 (MD) シミュレーションがある。このようなことから、我々は (1) タンパク質分解系におけるシグナル付加において重要な働きをするユビキチンの標的タンパクの分子構造選別機構、(2) 分子動力学を用いたタンパク質のフォールディング理解に注目している。そこで本研究では、MD シミュレーションを用いユビキチンとユビキチン様タンパク質の構造変化に関して考察した。具体的には、ユビキチンとユビキチン様たんぱく質 (SUMO-1, NEDD8) の各立体構造の自由エネルギー最小化に伴う安定構造を温度制御 (エネルギー制御) MD でその系の熱力学的エネルギーにおける構造変化を明らかにし、各分子のトポロジーとその標的タンパク結合部位の位置について調べ、ユビキチン化と標的タンパク質認識に関する構造の特徴を探ることを目的とした。実際には、次のことを行った。(1) ユビキチンと Ubl の主鎖安定構造を MD シミュレーションにより予測した。また、MD シミュレーションは、計算量が膨大で時間がかかる。ところが、近年コストパフォーマンスに優れた PC クラスタが注目され始めてきた。そこで、本研究では、PC クラスタ上で MD シミュレーションを行うため MPI ライブラリを使用してこれまでに開発した MD プログラムについて (2) 粒子分割法による分子動力学並列化を行い、その並列化効率の評価についても考察を行った。

1.2 本論文の構成

本論文では、まず第 2 章ではタンパク質立体構造とその機能・ダイナミクスの概要について述べ、本研究で用いた分子動力学法の概要について述べる。第 3 章では、ユビキチン-プロテアソーム系と呼ばれるタンパク分解系の概要を述べ、ユビキチン化のしくみについて述べる。そして、ユビキチン(Ub)とユビキチン様タンパク質 (Ubl) の分子動力学シミュレーションについて、数値実験方法と実験結果について述べる。4 章では、分子動力学の並列化手法とその実装方法について述べ、数値実験を行いベンチマークをとり性能の評価をした。最後に 5 章で本研究のまとめを行い、問題点と今後の課題を述べる。

第 2 章

タンパク質の立体構造・機能とダイナミクス

2.1 はじめに

生命活動はタンパク質の相互作用で成り立っている。アミノ酸配列は、自由エネルギーが最小となる構造（ネイティブ構造）に折りたたまれる。ネイティブ構造を安定化する力は、配列上で近くに位置するアミノ酸間の相互作用（局所的相互作用）と離れたアミノ酸間の相互作用（非局所的相互作用）であり、ネイティブ構造はこの相互作用の総和によってきまる自由エネルギーの最小状態のことである。立体構造こそがタンパク質の機能において重要な役割を果たす。タンパク質はアミノ酸配列によって定まる固有の構造に折りたたまれることによって、タンパク質を構成する各原子はある特定の位置に配置する。この結果として、タンパク質はある特定の機能をもつ。本章では、タンパク質の立体構造とそのダイナミクスに関して述べるとともに、本研究で用いた分子動力学法の概要についても述べる。

2.2 タンパク質の立体構造と機能

これまでたくさんの研究者によってタンパク質の立体構造、機能およびアミノ酸が決定されてきた。その結果、タンパク質立体構造データベースである PDB (Protein Data Bank) [2]には、膨大な数の構造や機能解明のすすんだ情報が蓄積された。現在では 12000 個以上のタンパク質の立体構造が登録されている。これらと比較したところ、よく似た立体構造をもつタンパク質が多数存在することが知られている。タンパク質の複雑な立体構造は、一次構造、二次構造、

三次構造，四次構造の4つのレベルに分けて考えられている。一次構造は，タンパク質を構成する個々のアミノ酸の配列である。二次構造は， α ヘリックスや β シートと呼ばれるポリペプチド鎖（アミノ酸がペプチド結合で鎖状につながった高分子）の部分部分に見られる局所的な構造である（図 2.1）。三次構造は，折りたたまれたポリペプチド鎖の全体的なトポロジーのことである。四次構造は複数のサブユニットが会合してできたタンパク質の立体構造を表す。タンパク質の類似性は α ヘリックス， β ストランドといった二次構造のアミノ酸配列上の出現順序とそれらの空間配置（トポロジー）の類似度によって議論される。

二次構造の出現順序と空間配置のことをフォールドと呼ぶ。すなわちフォールドの類似性から立体構造の類似性を推測するし，また立体構造の類似は細胞内での機能も似ている確率が高いと考えられる。現在，タンパク質の構造を分類した SCOP[3]，CATH[4]，Dali[5]，Entrez[6]と呼ばれるデータベースが存在する。これらは，データベース開発者が考案した方法に従ってタンパク質のフォールドを分類しており，それぞれのフォールドには名前がつけられている。タンパク質の機能を知るにはまずその立体構造を知ることから始まるのである。代表的なデータベースの1つの CATH では，タンパク質を4つの階層で分類している。最初の階層 Class で全タンパク質を α ， β ， $\alpha\beta$ 混合，例外の4つに分類し，次の階層 Architecture で35個に細分化される。その次の階層が Topology で，現時点で672個のトポロジーが見つかっている。この段階まではアミノ酸配列の情報を使わないで分類できるが，4つ目の階層は配列相同性（Homology）を基にしたものであり，タンパク質のファミリーに相当する。（図 2.1）

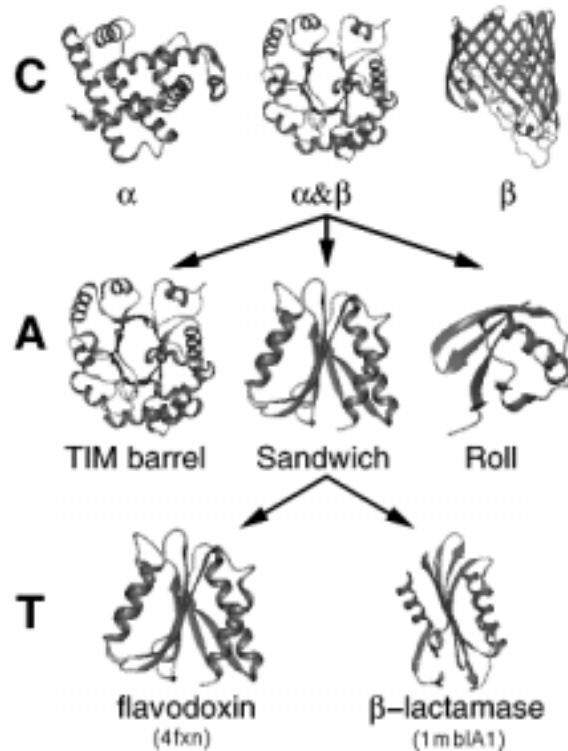


図 2.1 CATHによるタンパク質の分類

2.3 タンパク質のフォールディング

タンパク質は、細胞内において様々な構造をとることができ、細胞内でダイナミックな構造変化をしている。例えば、実験環境下ではタンパク質水溶液に酸やアルカリを加えると、固有の立体構造（ネイティブ構造）が壊れ、変性状態と呼ばれるランダムな主鎖構造となる。しかし、この状態から生理的環境に戻してやると、自発的に天然状態の構造を形成する。これをタンパク質のフォールディングという。このように細胞内でポリペプチド鎖が折りたたまれることにより、タンパク質としての活性を持つようになる。最近では、タンパク質は必ずしも自発的に折りたたまれるものばかりでなく、細胞内で折りたたみを助けるシャペロニンというタンパク質が介助していることもわかってきている [7]。いずれにしても、タンパク質のネイティブ構造は、アンフィンセンの実験 [8]によりアミノ酸配列と溶媒環境によって決まる自由エネルギー最小の状態

あることが示されている。

フォールディングの理論的な研究

タンパク質のフォールディング研究の流れとして、格子モデルを用いたモンテカルロシミュレーションや、ヘテロ高分子鎖の統計力学に基づく理論的な研究の成果として、「フォールディング・ファネル」という考えが生まれてきた[9][10] (図 2.2)。タンパク質のフォールディングは非常に複雑な系であるので、折れたたまるには、特定の経路上の限られた数の中間体を経て折れたたまるのではなく、そこには無数の経路が存在し、折れたたまりが進行するにつれて経路の数が減少して天然構造へと折れたたまれていく。タンパク質の個々のミクロな構造状態を多次元のコンフォメーション空間内の点とするならば、タンパク質の取り得るエネルギー曲面は、個々の構造状態のエネルギーを縦軸にとり、図 2.2(a)のように表すことができる。このモデルではタンパク質のフォールディングはファネルで表されるようなエネルギー曲面にしたがってエネルギーが落ちて行くように進行していく。このような理論背景に従うことで分子動力学といった熱統計力学的な分子シミュレーション手法を用いることができ、フォールディングの過程を計算機の中で定量的に予測することができる。

フォールディングの分子動力学シミュレーション

上で述べたようにタンパク質の細胞中での天然構造は、1つの立体構造に固定された「硬い」ものではなく、熱揺らぎによって生じる様々な立体構造をとる「柔らかい」ものである。天然構造には、エネルギー的に準安定な多数の構

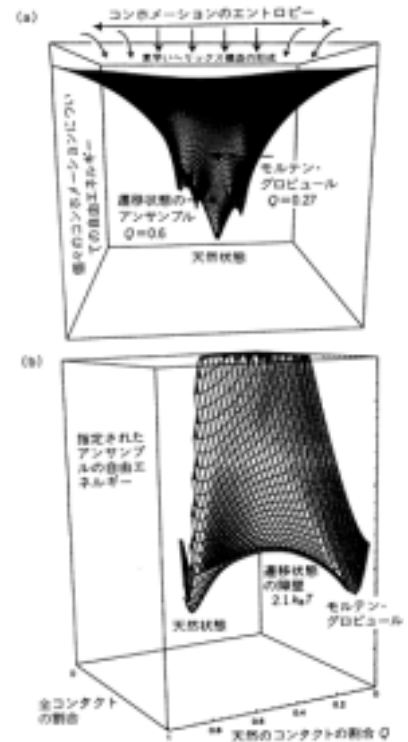


図 2.2 フォールディング・ファネル
出所: 参考文献[11]

造状態が存在し、さらにそれらが複雑な階層構造を形成している。タンパク質のダイナミクスは、1つの構造状態近傍でのゆらぎだけでなく、構造状態間の遷移によっても生じてくる。この複雑さは、タンパク質分子のみに起因するのではなく、溶媒環境によっても大きく影響される。タンパク質のダイナミクスは、ピコ(10^{-12})秒 (ps) からフェムト(10^{-15})秒(fs)のオーダーで起こるため、実験環境下で観察するのはむづかしい。そこで、このダイナミクスを知る有効な手段として分子動力学法を用いた分子シミュレーション法が有効であることが知られている。タンパク質の性質をシミュレーションによって定量的に正しく再現するためには、モデルをどのように構築するかが重要となる。分子シミュレーションでは、系を構成する各原子を取り扱う最小単位とし、原子間の相互作用エネルギーを表すポテンシャル関数をあらかじめ定義しておく。このように、タンパク質分子・周辺の溶媒分子の原子座標を与えると、系のエネルギーを求めることができる。このような古典的モデルの分子動力学シミュレーションモデルでは、電子状態の変化などは取り扱えないが、ゆらぎ・構造安定性などを考察するには十分であることが知られている。分子動力学シミュレーション法は、計算量が膨大で、これまでは限られた系のシミュレーションしか行うことができなかったが、近年の計算機能力の向上により、創薬といった実践領域でも、その必要性・有用性が認められ分子動力学シミュレーションを適用したの実用的なシステムを開発することは今後ますます重要である。

2.4 シミュレーションでのモデル化

物理学・化学では伝統的な研究手段として、実験と理論が互いに相補的な役割を果たし発展を遂げてきた。近年では、計算機シミュレーションがより現実的な事象を模擬することができるようになり、実験では行うことが困難な環境下の計算機実験や、ミクロな原子などの動的挙動を考察ができるようになってきた。しかし、実際に計算機を用いて現実的な系を再現するにはその妥当性を検証するためにいくつかのことを考えなければならない。タンパク質の物性を評価する上でも図 2.3 に示すように、実験、理論、そしてシミュレーションと相

互の比較検討が重要である.



図 2.3 理論と実験, シミュレーション

特にタンパク質のシミュレーションを行う上でも, 分子シミュレーションによって得られたタンパク質のモデルは, その信頼性が評価されなければならない. そのためにタンパク質モデルの特性を吟味する上で欠かせない要素は立体化学的な精度, 充填特性, 折りたたみ構造の信頼性といったものが挙げられる (図 2.4). 構築されたモデルの立体化学的な性質は, 結合長, 結合角, ねじれ角といったパラメータの精度に基づいて評価される. また, タンパク質における疎水性残基と親水性残基の分布はタンパク質モデルの信頼性を判断する上で重要な要素となる[11].

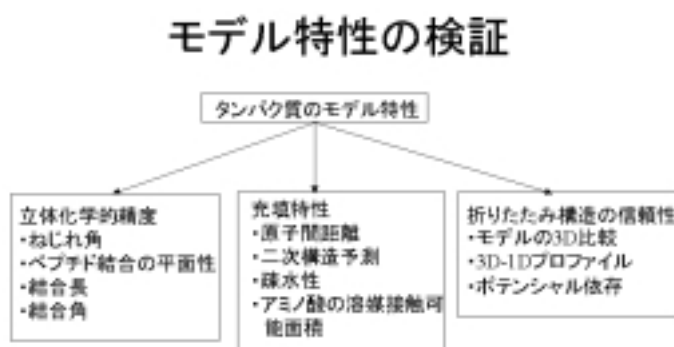


図 2.4 タンパク質モデル吟味

2.5 分子動力学について

タンパク質分子は種々の立体構造をとるが、その熱力学的な安定性を考えるときには各立体構造が作られるときの自由エネルギーの変化を考えればよい。すなわち、ある立体構造での自由エネルギーは、その近傍の立体構造の自由エネルギーより低い、つまりある立体構造が自由エネルギーの極小点となっていれば、その立体構造は安定であるといえる（タンパク質フォールディングの熱力学的原理）。この自由エネルギー計算には分子動力学（Molecular Dynamics, MD）法[12][13][14]を用いることができる。分子動力学シミュレーションは、タンパク質や DNA などの生体高分子の構造、物性、機能相関の研究に有用なことが認められている[15]。本章では、本研究で用いた分子動力学の概説について述べる。

2.5.1 分子動力学概説

分子動力学計算は大きく分けて二つの方法に分けられる。1つは、古典分子動力学であり、もう1つは、第一原理分子動力学シミュレーションである。古典分子動力学とは、一般に「N 体多体系の状態の時間発展を、古典力学の運動方程式（Newton の運動方程式）の数値解を差分法により求めてシミュレーションすること」と定義することができる。一方、第一原理分子動力学とは、量子力学を組み合わせた方法で計算量が非常に多く、現在のスーパーコンピュータを用いても数百原子レベルでのシミュレーションが限界といった状況であり、タンパク質といった高分子に適用するには、現時点では困難である。そこで、本研究では古典分子動力学を用いることにした。以下、分子動力学法の概要について述べる。図 2.5 に分子動力学の大まかな手順を示す。

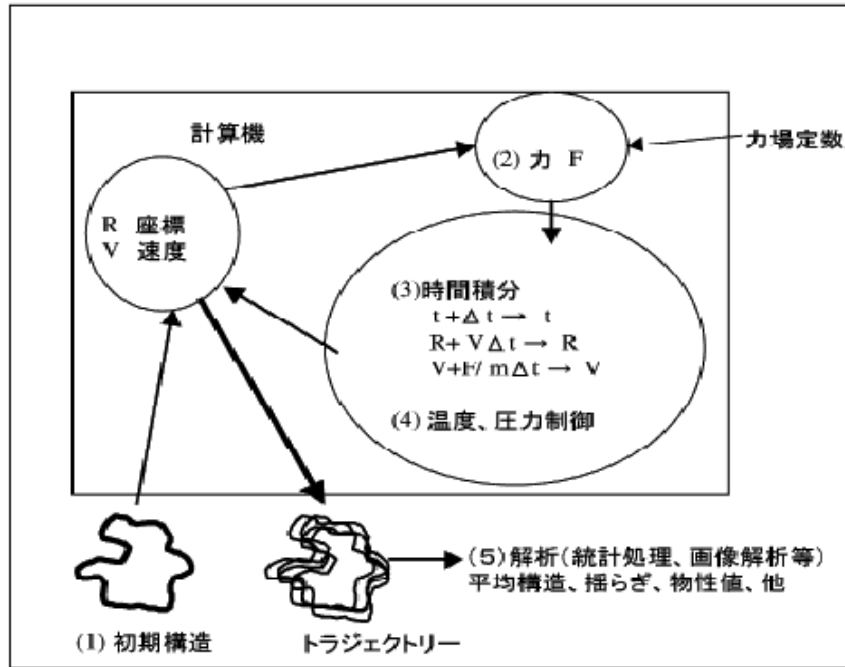


図 2.5 分子動力学の手順

分子動力学の手順は、上図に示したようになる。すなわち、

- (1) 初期構造（座標）を入力し，速度を発生させる。
- (2) その座標と力場定数を用いて力を計算する。
- (3) 時間刻み後の座標と速度を計算する。
- (4) 必要なら温度と圧力を制御する。
- (5) (2) ~ (4) の各段階を必要タイムステップ繰り返すことにより，座標，速度，エネルギーといった物理量の時間変化が得られる。
- (6) (5) で得た物理量をもとに，統計処理したり画像処理したりすることにより，その分子の構造や物性について考察をする。

以上が，一般的な分子動力学法の手順である。以下では，図中の各原理について，本研究で実装した機能について説明していく。

なお，ここで計算機に最初に入力すべきパラメータは，分子の初期構造と力場定数（ポテンシャル場）である。

2.5.2 運動方程式

古典分子動力学では、個々の原子はニュートンの運動方程式に従って他の粒子から力を受けながら運動する (図 2.6). N 個の粒子からなる質点系を対象とすることで分子集団の挙動をシミュレートすることができる. 系の i 番目の粒子の質量を m_i とすると, 運動方程式は,

$$\vec{F}_i = m_i \frac{d\vec{v}_i}{dt} = m_i \frac{d^2\vec{r}_i}{dt^2} \quad (2.1)$$

ここで, \vec{F}_i は粒子に働く力であり, \vec{v}_i, \vec{r}_i はそれぞれ粒子の速度, 位置である. 個々の原子に働く力は原子間の相互作用 (ポテンシャル関数 Φ の微分) より定められる.

$$\vec{F}_i = -\text{grad}\Phi(r) = \sum_{i=1, j \neq i}^N \frac{\partial \Phi(r)}{\partial r_{ij}} \frac{r_{ij}}{r_{ij}} \quad (2.2)$$

3次元空間の運動では, $3N$ 個の2階常微分方程式となる.

2.5.3 数値積分

微分方程式を数値積分するには, 微分方程式を差分方程式で近似する.

これは時間微分を

$$\frac{d\vec{r}}{dt} \approx \frac{1}{\Delta t} \{\vec{r}(t + \Delta t) - \vec{r}(t)\} \quad (2.3)$$

と近似することに相当する. この有限微小時間を用いた離散化した微分方程式を数値積分する方法として, 分子動力学では, ベルレの方法, 予測子—修正子法, ルンゲクッタの方法等が用いられる. この方法のうち, ベルレの方法は力

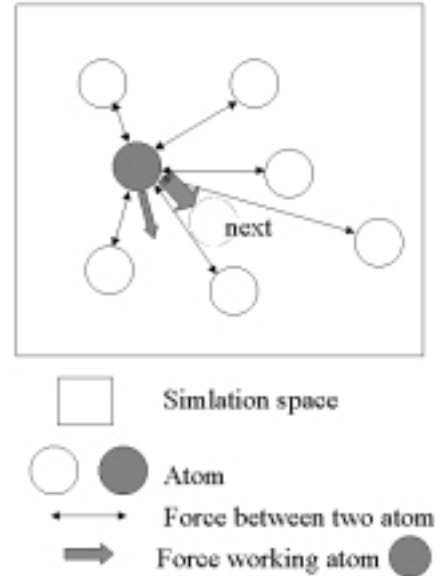


図 2.6 原子に働く力

の計算が一度だけであり，計算手順が簡便なため広く用いられている．ただし，他の方法に比べ計算精度が悪いため，有効な計算を行うには，時間刻みを十分小さくすることでシミュレーションの妥当性を検証する必要がある．本研究では，この以下の式で表される速度ベルレ法を用いた．

速度ベルレ法により位置，速度を計算する

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \Delta t \cdot \vec{v}_i(t) + (\Delta t)^2 \frac{\vec{F}_i(t)}{2m} \quad (2.4)$$

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{1}{2m} \Delta t (\vec{F}_i(t + \Delta t) + \vec{F}_i(t)) \quad (2.5)$$

2.5.4 境界条件

分子動力学で計算できる系は非常にミクロな部分系であるため，計算モデルの境界にはなんらかの境界条件が必要である．一般には周期境界条件が用いられることが多い．

周期境界条件

計算機の計算速度や記憶容量から一度に扱うことのできる粒子の数には限界がある．結晶内部のような多量の粒子の存在を仮定しなくてはならない場合は，空間が数十から数百個の原子の入った立方体の繰り返しで埋まっているとする周期的境界条件を用いる必要がある．ある粒子が境界を通過してシミュレーション領域から流出する場合，反対側の境界面を通過してそのままシミュレーション領域内に流入すると考える．周期境界条件の採用により，非常に小さなシミュレーション領域でも実験データを証明することがこれまでに明らかにされ，現在まで圧倒的に多く用いられてきた．しかし本研究で用いたモデルはごく小さな系であるため次の自由境界条件を採用した

自由境界条件

超微粒子やマイクロクラスターなどの比較的原子数の少ない系についてシミュレートするには，初期条件として原子を塊状に配置し，初速度を与える．これは，原子の個数が少ないため，原子の運動は初期条件に長く影響されるか

らであり，原子は境界条件のない自由境界のもとで，ゆっくりと平衡状態まで運動させる．

2.5.5 原子間ポテンシャル

原子に働く力を計算する際に，原子間相互作用であるポテンシャルエネルギーの微分を用いる．ポテンシャルエネルギーは原子間距離や原子間角度などの関数で表される．現在のところ，すべての物質のポテンシャルを統一的に表す関数はないため，対象とする物質により適切な関数（またはパラメータ）を選定するしかない．

DNA やタンパク質などの生体高分子用を扱うにはこれまでに何種類かのパラメータセットが開発・提案されている．調和型ポテンシャルとして以下にあげるポテンシャル（AMBER Potential Field）が代表的なものである[16]．一般に調和型ポテンシャルと呼ばれる 5 種類の各原子間のポテンシャルの和で表される（式 2.11）．個々のポテンシャルについて説明する．

Lennard-Jones Potential

これは，短距離相互作用で用いる 2 体間ポテンシャルで，次式で表される．

$$E^{LJ} = 4\epsilon \left\{ \left(\frac{\omega}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right\} \quad (2.6)$$

r ：原子間距離， ϵ ：ポテンシャル定数（エネルギーの深さを定義），

ω, σ ：各原子のファンデルワールス半径

クーロンポテンシャル

クーロンポテンシャルは，電荷をもつ分子間に働くポテンシャルである．分子 i と分子 j 間のクーロンポテンシャルは以下の式で決まる

$$E^{coulomb} = \frac{q_i q_j}{r_{ij}} \quad (2.7)$$

q ：各原子の電荷

直鎖高分子のポテンシャル

次の3つのポテンシャルは、原子間の共有結合に関係する力である。

ボンド結合 (C-C 結合ポテンシャル)

$$E_{ij}^{bond} = \frac{1}{2} k_{cc} (r_{i,i+1} - b_0)^2 \quad (2.8)$$

連続する2つのボンドによる結合角 (C-C-C 結合角ポテンシャル)

$$E_{ijk}^{angles} = \frac{1}{2} k_T (\cos\theta - \cos\theta_0)^2 \quad (2.9)$$

連続する3つのボンドによる2面角 (ねじれポテンシャル)

$$E_{ijkl}^{torsion} = \frac{1}{2} V \{1 - \cos[3(\phi - \phi_0)]\} \quad (2.10)$$

ここで、

k_{cc}, k_T : 力の定数, θ : 結合角, ϕ : 二面角, V : 定数(barrier height)

b_0 : 安定結合距離, θ_0 : 安定結合角, ϕ_0 : 安定二面角

生体分子での調和型ポテンシャル

生体分子の調和型ポテンシャルとして以下に示す代表的な Amber 型ポテンシャルがある。タンパク質, DNA といった分子のシミュレーションでは現在このポテンシャル型を用いることが多い。

$$\begin{aligned} \Phi = & \sum_{i < j} \sum_{i < j} 4\epsilon \left\{ \left(\frac{\omega}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right\} + \sum_{bond} \frac{1}{2} k_{cc} (r_{i,i+1} - b_0)^2 + \sum_{angles} \frac{1}{2} k_T (\cos\theta - \cos\theta_0)^2 \\ & + \sum_{torsion} \frac{1}{2} V \{1 - \cos[3(\phi - \phi_0)]\} + \sum_{i > j} \frac{q_i q_j}{r_{ij}} \end{aligned} \quad (2.11)$$

2.5.6 熱力学的物理量

熱力学的状態量の中で最も基本となる量は温度と圧力である。分子動力学においては、ここで示す温度や圧力を制御することで外界とのエネルギーの授与を考える

温度と圧力についての評価式を以下に示す。本研究では、ここに示す温度を熱力学的統計量として制御する温度制御 MD を実装した。

温度

温度は、粒子の熱運動と関係し、熱速度、すなわち粒子の速度から平均流速を引いた速度で定義される。系が停止している場合には、粒子の速度そのものが熱速度である。温度 T が与えられた系の場合、粒子の速度はマクスウェル分布に従った分布となる。この分布を用いて運動エネルギー $K(p)$ の集団平均を求めると、次のようになる。粒子の個数を N とし、ボルツマン定数を k とすると、

$$\langle K(p) \rangle = \left\langle \frac{1}{2m} \sum_{i=1}^N p_i^2 \right\rangle = 3N \frac{kT}{2} \quad (2.12)$$

ただし、 \mathbf{p} は運動量で、 p_1, p_2, \dots, p_N をまとめて表したものである。この式は、一自由度あたり $\frac{kT}{2}$ のエネルギーを有するというエネルギー等分配の法則からも容易に理解できる。そこで、小正準集団のように温度が変動する場合、温度 T は次式で定義される。

$$T = \frac{2}{3N} \cdot \frac{1}{k} \langle K \rangle \quad (2.13)$$

圧力

圧力 P はビリアル状態方程式として知られている以下の式で表される。

$$P = \frac{N}{V} kT + \frac{1}{3V} \left\langle \sum \sum \vec{r}_{ij} \cdot \vec{f}_{ij} \right\rangle \quad (2.14)$$

ここで、 $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ 、 \vec{f}_{ij} は粒子 j が粒子 i に及ぼす力、 V は体積である。

右辺第 1 項は、粒子の運動による寄与、第 2 項は粒子間力に起因する項である。

第 3 章

ユビキチンの MD シミュレーション

3.1 はじめに

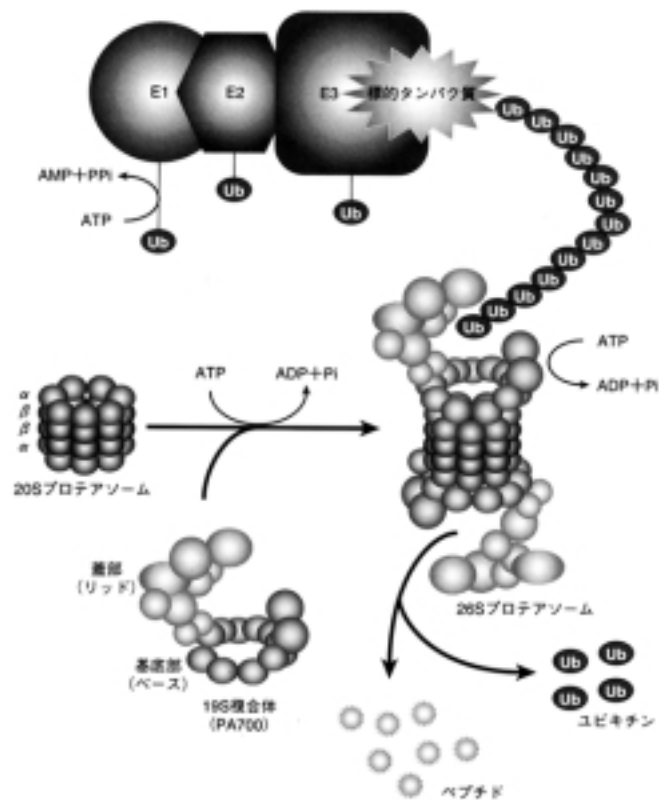
ユビキチンとプロテアソームはタンパク質分解システムにかかわる物質である。最近の研究では、がん、免疫疾患、炎症、脳疾患などといった病気では、タンパク質分解システムの異常が主な原因で起こっていることが明らかになってきている。ユビキチンはタンパク質の翻訳後修飾分子であり、ユビキチン鎖の付加はタンパク質分解の目印となる。ユビキチン分子の標的タンパクの分子レベルでの認識機構を解明することが求められている。

3.2 ユビキチンプロテアソーム系

3.2.1 タンパク質分解系

タンパク質分解には、プロテアソームがかかわっている。プロテアソームは、各細胞にありタンパク質の複合体が 6~7 個リング状にくっついた構造をしており、これがタンパク質を分解する役目を果たしている。このとき、決まったタンパク質だけを分解するのであるがそのときに目印となるものがユビキチンである。ユビキチンは分子量 8600 の小さなタンパク質である。このユビキチンが標的タンパク質に結合すること（ユビキチン化）により分解の指示をする（図 3.1） [17]。細胞内でユビキチン化されたタンパク質がエネルギー依存的に分解されることの意味は、第一には不要なタンパク質の細胞内からの除去である。ストレスや環境変化、または合成途上での間違いやアミノ酸類似物質の取り込みに由来する変性したタンパク質は速やかに除去される必要がある。第二に、

機能調整を行っているタンパク質を特定の時期に積極的に壊す「選択的分解」機構として作用することで、数々の機能制御に直接的にかかわることである。これは特に時間的、段階的な制御のステップを不可逆的にコントロールすることに適している。重要な点は、このようなユビキチンを経た選択的タンパク質分解を通して、制御されている細胞の機能が、転写、DNA複製と修復、細胞周期の制御、ストレス応答、多くの原ガン遺伝子産物の分解による増殖制御、タンパク質のトランスロケーション、免疫始動など、極めて多岐にわたっていることにある。ここで、標的タンパク質選択における最大の課題は、いかに特定のタンパク質を特定の時期に認識して分解するかの特異性とタイミングの分子認識機構の解明である。



真核生物の ATP プロテアーゼである 26S プロテアソームは、触媒ユニットである 20S プロテアソームと調節ユニットである 19S 複合体が ATP に依存的に分子集合した巨大な多成分複合体である。

図 3.1 ユビキチン-プロテアソーム系

出所、参考文献[17]

ユビキチン化は図 3.2 に示すような 3 段階で成り立っている。まず，第一段階では，ユビキチンに「E1」と呼ばれる活性化酵素が結合し，第二段階では，その酵素が付いたユビキチンに「E2」の結合酵素が結合する。そして第三段階では，その複合体に「E3」のリガーゼが付く。このようにしてユビキチンに 3 つの酵素が結合した複合体が形成される。ここで，タンパク質のユビキチン化修飾は，標的タンパク質に分解シグナルを付与することであり，一連の複合酵素反応で標的タンパク質の特定リシン(Lys)残基にイソペプチド結合し，ユビキチン相互でこの反応を繰り返して，複数のユビキチン分子が共有的に連鎖結合する。生じたポリユビキチン鎖は，真核生物の ATP 依存性プロテアーゼである 26S プロテアソームによって補足され，標的タンパク質は選択的かつ即効的に分解される。しかしここで，ユビキチンシステムがいかにして標的分子を的確に選別するかという原子レベルでの分子機構については明らかにされていない。ごく最近，SCF (Skp11-Cdc53-F-Box タンパク質) というタンパク質では，その構成因子である Cdc53/Cullin が Rub1/Nedd8 (ユビキチン様タンパク質) によって修飾されることが判明した[18]。

そこで，本研究では，ユビキチンとユビキチン様タンパク質 (Ubiquitin like protein, Ubl) についてその標的タンパク質の分子認識機構に着目する。次節で紹介するユビキチンとユビキチン様タンパク質のユビキチン化結合部位と標的タンパク質結合部位を含むポリペプチド鎖の主鎖の構造変化から標的分子認識機構における特徴を探ることを目的とする。

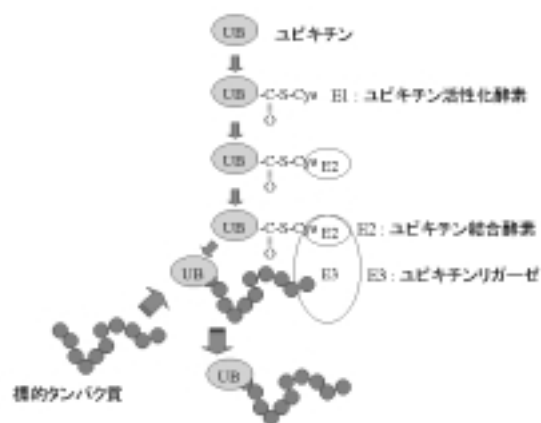


図 3.2 ユビキチン化のしくみ

3.2.2 ユビキチン様タンパク質修飾システム

ユビキチンと相同性のあるユビキチン様タンパク質 (Ubl) の存在は古くから知られていたが、その機能および意義についてはほとんど理解されていなかった。最近、多数の Ubl がさまざまなタンパク質と共有結合し機能制御にかかわることが明らかになりになった[1]。タンパク質と共有結合する Ubl はその C 末端に GG 配列を有し、ユビキチンと全体的な相同性を有する 1 型 Ubl と、ユビキチン様構造を有するが GG 配列をもたず、他のタンパク質の共有結合が認められない 2 型 Ubl とに分類される。ヒトの 1 型 Ubl は 4 種類 (SUMO-1, NEDD8, UCRP, MNF β /Faul) 知られている。本研究では、ユビキチンと 1 型 Ubl である SUMO-1[19]と NEDD8[20]の構造変化に着目する。図 3.3 に示すように、Ub の 48 残基目のリシンと 76 残基目の Gly がそれぞれユビキチン、標的タンパク質との結合部位である。Ubl についても相同性を示す図中丸印の部位が結合部位となる。そこで図 3.3 中の [] 部分で示した 3 つのタンパク質のポリペプチド主鎖 (ただし、SUMO-1 については、PDB で構造既知の 75 残基までを考慮) について温度制御 MD シミュレーションを行うにより、動的挙動と安定構造を求め、その構造の特徴を考察した。

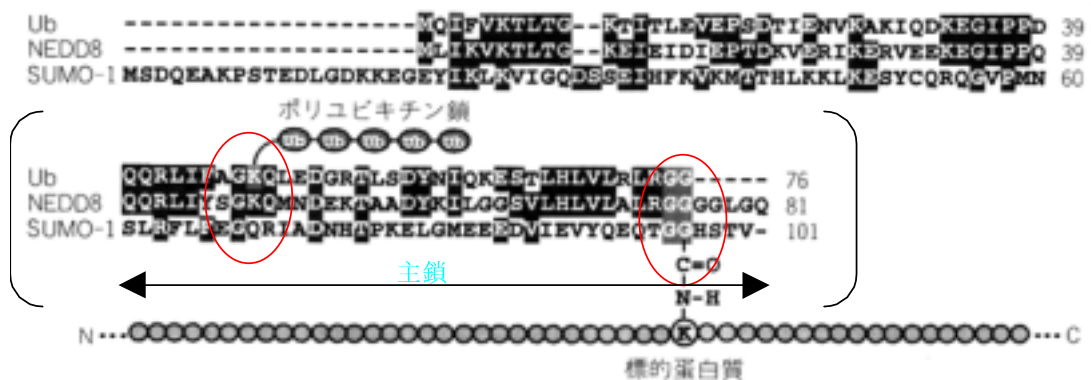


図 3.3 ユビキチンと Ubl (SUMO-1, NEDD8)

3.3 数値実験モデル

3.3.1 数値実験方法

本節では，分子動力学を用いた分子シミュレーションとして，前節のユビキチンとユビキチン様タンパク質を取り上げ，結合部位を 2 箇所持つ図 3.3 に示した残基の主鎖について着目し，各分子の最終安定構造を求めることを目的とした．ユビキチンについては，ポリユビキチン化の結合部位は，Ub の 48 残基目のリシン (Lys) であり，標的タンパク質と結合する部位は 76 残基目のグリシン (Gly) である．全体の構造変化とともにこの部位についてもどのように構造が変化するかについても考察を行う．

ユビキチン (Ubiquitin) とユビキチン様タンパク質 (SUMO-1 と NEDD8) の構造と，シミュレーションで扱う主鎖構造をそれぞれ図 3.4，図 3.5，図 3.6 に示す．それぞれ，左図が PDB で得られる全体構造であり，右図が上で述べた計算機実験で用いる残基の主鎖構造である．

数値実験方法としては，最初に系のエネルギーが一定に保持されることを NVE アンサンブル (後述) で確認したのちに，NVT アンサンブル (後述) に基づき，1 [ps] (10000 MD Time Step) 毎に系の温度 (エネルギー) を 100 [K] ずつ下げていくことにより，絶対 0 度でのエネルギー最小の安定構造を求めた．初期の設定温度は 400 [K] である．実際行ったシミュレーションのアルゴリズム，ポテンシャル関数，パラメータについて述べる．

Ubiquitin, SUMO-1, NEDD8 の構造



図 3.4 Ubiquitin (PDB Entry : 1D3Z)

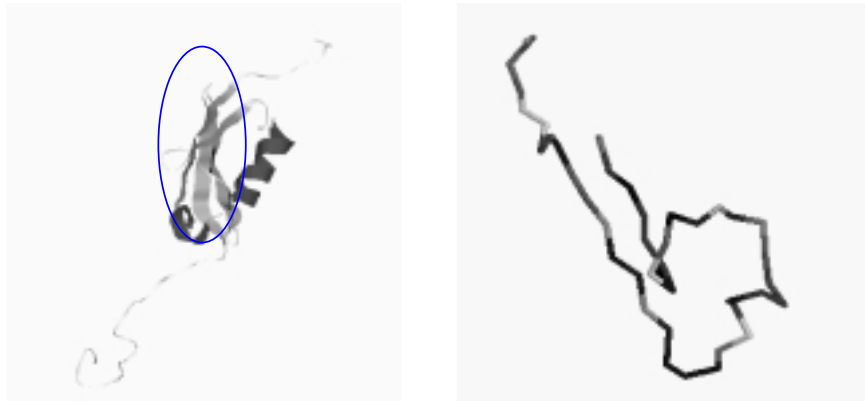


図 3.5 SUMO-1 (PDB Entry : 1A5R)

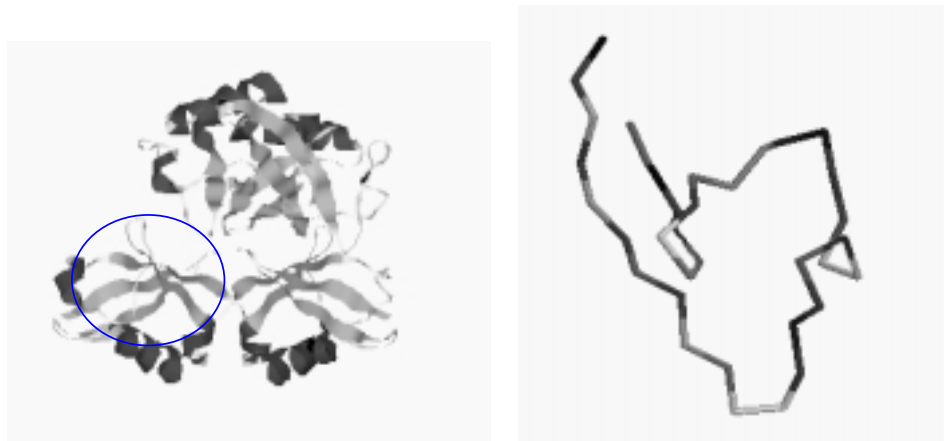


図 3.6 NEDD8 (PDB Entry : 1NDD)

3.3.2 数値実験で用いた式

運動方程式

運動方程式の時間発展による原子の位置，速度の計算法には 3 章で述べた速度ベルレ法を採用した．次に示す式で表される．

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \Delta t \cdot \vec{v}_i(t) + (\Delta t)^2 \frac{\vec{F}_i(t)}{2m} \quad (3.1)$$

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{1}{2m} \Delta t \left(\vec{F}_i(t + \Delta t) + \vec{F}_i(t) \right) \quad (3.2)$$

ポテンシャル関数

3 章で述べた調和型ポテンシャルを採用した．ただし，水素結合や他原子の電荷作用は考慮していないため，クーロンポテンシャルは導入していない．

$$\begin{aligned} \Phi = & \sum_{i < j} \sum 4\varepsilon \left\{ \left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right\} + \sum_{bond} \frac{1}{2} k_{cc} (r_{i,i+1} - b_0)^2 + \sum_{angles} \frac{1}{2} k_r (\cos \theta - \cos \theta_0)^2 \\ & + \sum_{torsion} \frac{1}{2} V \{ 1 - \cos[3(\phi - \phi_0)] \} \end{aligned} \quad (3.3)$$

アンサンブル

MD では，温度制御や圧力制御を行うことで様々な統計的集合（アンサンブル）を実現する．生体分子の MD では，NVE（粒子数・体積・エネルギー一定），NTV（粒子数・温度・体積一定），NTP（粒子数・温度・圧力一定）のいずれかが使用される．ここでは，本研究で用いた NVE，NTV アンサンブルについて説明する．

NVE アンサンブル（ミクロカノニカルアンサンブル）

孤立系の全エネルギーは，保存される．

$$H = \sum \frac{1}{2} mv^2 + U$$

温度制御や圧力制御を行わないと，このミクロカノニカル MD が実現する．外

界とのエネルギーの授与がないため、系のエネルギーが保存されることでシミュレーションの妥当性を検証する。実際には計算誤差があり、若干のエネルギーの変動は生じる。

NVT アンサンブル (カノニカルアンサンブル)

温度を一定にしたいときの最も簡単な方法に Woodcock の速度スケールリング法がある。温度 T と粒子の速度の関係は、

$$\frac{3NkT}{2} = \sum \frac{mv^2}{2}$$

の関係にあるので、数ステップ毎に速度のスケールリングを行う。本研究では、この方法に基づき温度制御を行った。現在では、この方法では計算精度が悪いことも挙げられており、能勢の方法と呼ばれる定温アルゴリズムも開発されている。

3.3.3 トrajekトリの解析

MD シミュレーションの結果からエネルギーや、分子構造、またそのダイナミクスなど様々な情報を得ることができる。本節で行うシミュレーションでは、平衡状態に到達してからの分子構造の揺らぎとして、根二乗平均揺らぎを計算した。

根二乗平均揺らぎ (RMSF, Root Mean Squares Fluctuation)

平衡状態に達してからの MD のトラjekトリの構造の揺らぎは、以下のような RMSF として計算することができる。

$$RMSF = \sqrt{\langle (R_j - \langle R_j \rangle)^2 \rangle}$$

本研究でのシミュレーションにおいては、基本単位を以下のように統一した。

| | | | |
|-------|------|------|----------|
| 空間距離 | [Å] | 結合角度 | [Radian] |
| エネルギー | [eV] | 温度 | [K] |

3.4 シミュレーションパラメータ

3.4.1 初期構造

図 3.3, 図 3.4, 図 3.5 の各右図のように PDB での構造から, 主鎖のみを取りだした構造を初期状態とした. それぞれ, 主鎖に含まれる原子数は, 37 個, 42 個, 35 個の炭素である.

3.4.2 数値実験でのシミュレーション時間

シミュレーション時間は, NVE アンサンブルでは, 1 ステップ 0.1[fs]で 5000 ステップすなわち 0.5[ps]行った. NVT アンサンブルでも同じく 1 ステップ 0.1[fs]で設定温度それぞれ 10000 ステップすなわち 1[ps]のオーダーで行った.

表 3-1 NVT アンサンブルでの計算条件

| 粒子の個数 | 計算ステップ [STEP] | 設定温度 [K] |
|-------------------------------|---------------|----------------------|
| 37(Ubi),42(SUMO), 35(NEDD) | 10000 | 400, 300, 200, 100,0 |

3.4.3 炭素原子の材料定数

数値実験で用いた炭素原子の材料定数を, 表 3.2 に示す.

また, 計算で使用する力に関する定数と炭素間での安定結合距離・角度を表 3.3 に示す[22].

表 3-2 材料定数

| 質量 | σ | ϵ/k [K] |
|------|----------|------------------|
| 12.0 | 3.923 | 72 |

表 3-3 定数

| 力の定数 Kcc | 力の定数 Ktt | 結合距離 [\AA] | 結合角 [rad] |
|----------|----------|-----------------------|-----------|
| 3.633720 | 1.349668 | 1.53 | 1.83 |

3.5 数値実験結果

数値実験の結果を示す。3.5.1 節でユビキチン分子について NVE アンサンブル条件で数値実験を行った結果を示す。3.5.2 節では各タンパク質について、温度制御 MD の数値実験を行った。その結果得られた主鎖最終安定構造と構造ゆらぎ (rms : RMSF, average distance は平均原子間距離) を示す。それぞれの制御温度でのエネルギー、温度、最遠原子間距離の値を示したグラフは、付録に付録図 1~15 として添付した。付録図は、それぞれのタンパク質において設定温度状態での 10000MD ステップ間の

(a) 系の全エネルギー

ここで、全エネルギーを total energy, 運動エネルギーを kinetic energy, ポテンシャルエネルギー (調和型ポテンシャル) を potential energy とした。

(b) 温度

(c) 最遠原子間距離 (分子全体の大きさの指標)

を示した。

3.5.1 Ubi の NVE アンサンブル

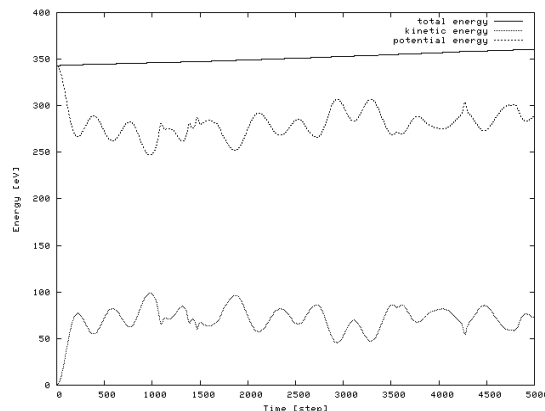


図 3.7 NVE アンサンブルでの Energy (Ubiquitin)

図 3.7 より、系の全エネルギー(total energy)は少しずつ増加しているがほぼ一様にエネルギーは保存されているため、MD シミュレーションが正常に行えると考えてよい。

3.5.2 各タンパク質の主鎖安定構造

数値実験結果 : ユビキチン (Ub)

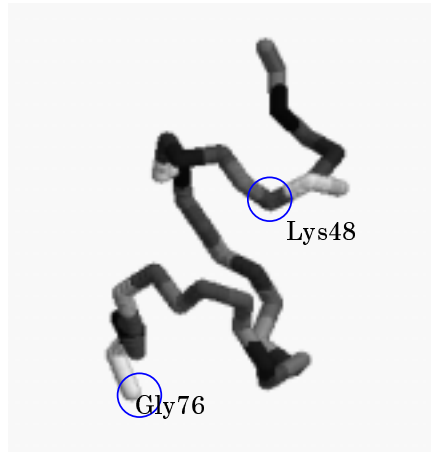


図 3.8 ユビキチン主鎖安定構造

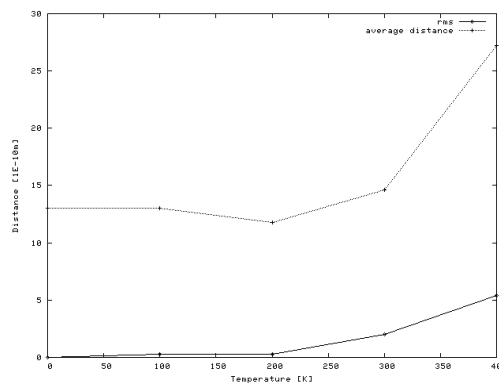


図 3.9 ユビキチン 構造揺らぎ

ユビキチンの最終安定構造は、図 3.7 に示すような、 α ヘリックス型になった。図 3.9 の根二乗平均をみると、400[K],300[K]では、揺らぎが大きくやわらかな構造をしていると推測できる。しかし、最終構造で得られた構造において Lys 残基は、ヘリックスの内側に入ってしまったため、本来の結合部位では結合しにくくなっているように見うけられる。本来 Lys 残基は、親水性アミノ酸であるため、 α ヘリックスの外側にでてくる確率が高い。この実験では、水溶媒や水素結合をいったことを考慮できていないためだと考えられる。

数値実験結果 : SUMO-1

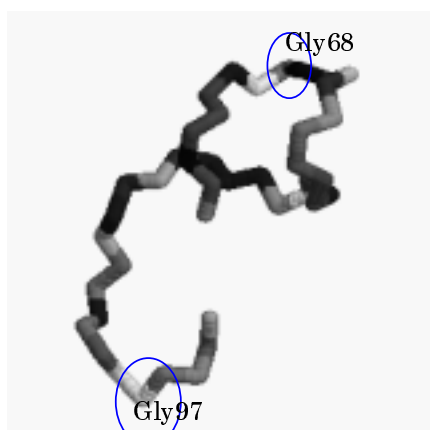


図 3.10 SUMO-1 主鎖安定構造

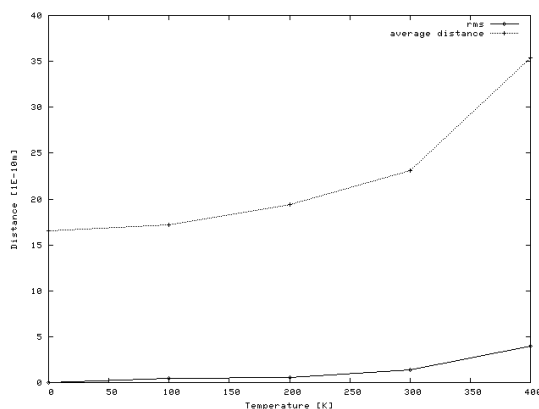


図 3.11 SUMO-1 構造揺らぎ

主鎖の安定構造は, 図 3.9 のような形を形成した. 付録図 6(a)で示されるように初期の構造が最も不安定 (高エネルギー) なものは, この分子であった. そして, 付録図 6(c)で示されるように最遠原子間距離が小さくなるのが最も遅かった, そのため図 3.11 の根二乗平均より 400[K]という比較的高温においても揺らぎは小さかった. 最終構造では, 68 残基目の Gly, 97 残基目の Gly とともに外部との接触面積の大きな場所に露出していた.

数値実験結果 : NEDD8

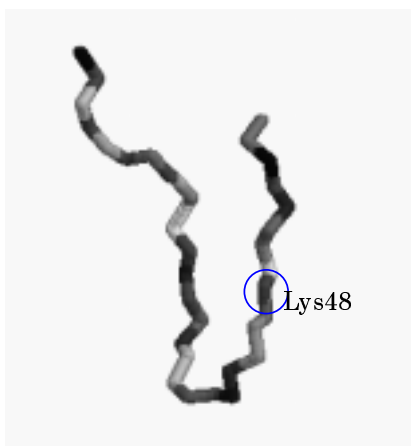


図 3.12 NEDD8 主鎖安定構造

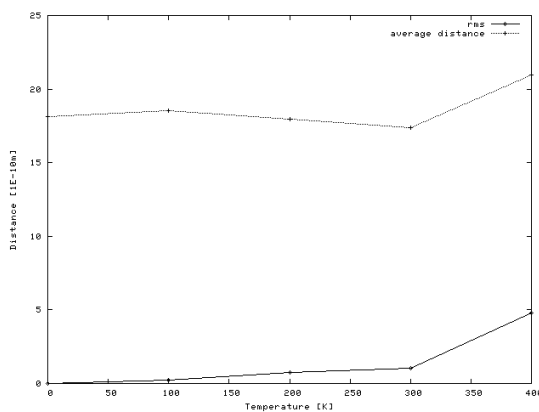


図 3.13 NEDD8 構造揺らぎ

NEDD8 の最終構造は、ユビキチンが α ヘリックス型であったのに対して、 β シート構造に似た構造をとった。初期構造でのエネルギーが最も小さく（付録図 11(a)）。トポロジーとしての構造変化が最も小さかったのがこの分子であったといえる。最終的には最も長軸方向に長い分子構造となった。

3.6 考察

この数値実験では、ユビキチンとユビキチン様タンパク質 (SUMO-1, NEDD8) について主鎖構造の安定構造を温度制御の MD で求めた. 400[K]から 1[ps]毎に系の温度 (エネルギー) を 100[K]ずつ冷却していくことにより, 絶対 0 度, すなわち自由エネルギー 0 の状態の安定構造を求めた.

数値実験前の前提として以下のようなことを考えた. 実際の細胞内では, Ubiquitin と Ubl はともにポリユビキチン化をおこし, 標的タンパク質と定まった相同な配列位置で結合する. そのため, 細胞内での他分子の分子認識機構は似ていると考え得る. それにより, シミュレーション前の予測では, 配列に相同性があり構造も似ているため, 同じようなフォールディング過程で折りたたまれ, 似た安定構造に行きつくと考えていた. その際には, 標的タンパクとポリユビキチンの結合部位の挙動について似たような挙動が示されるかも知れないと考えていた. しかし, 数値実験の結果は予想していた以上に初期値の違いに依存したためか, 最終的な構造は 3 つともトポロジーとして全く似つかない構造となった. 実際には, 自由エネルギーを徐冷していく速度が速すぎたことや, 室温環境といった温度での構造変化を長時間観測することで, 特異的な構造変化を考察することができた可能性もある.

本実験から以下のことが考えられる.

- ・ 炭素主鎖だけというように, モデルを単純化することで, 他原子に関してエネルギー安定性において大きな違いがでてきてしまうため, 予想に反した結果となってしまった.
- ・ α ヘリックス, β シートに似た構造が炭素主鎖だけでも生じることが有り得ることがわかった.

問題点として次のことについて詳しく検討していかなければならない.

- ・ アミノ酸残基の他の原子, 水素の付加, 水溶媒, 境界条件といったより厳密なモデル化

第 4 章

分子動力学の並列化

4.1 並列化の必要性

分子動力学計算は、分子内の各原子 1 個 1 個の運動を計算するシミュレーション法である。タンパク質レベルの巨大分子の MD シミュレーションは計算量が膨大で高価なスーパーコンピュータなどが必要とされてきた。そしてその計算量の多さから、これまでも並列計算機上での並列化手法が考えられてきた[23]。これまでは、CM-5 や T3E などの数百から数千プロセッサ要素をもつ商用並列計算機が広く使用されている。しかし近年になって、高速で安価な並列コンピュータとして PC クラスタが注目されている。そこで本研究では、これに着目し MPI ライブラリを用いることにより PC クラスタ上での MD 並列化も行った。ちなみに生体分子を扱う分子動力学法の計算能力とメモリ空間の大きさは、計算能力 10^{19} FLOPS, メモリ容量 10^{13} byte と推測されており並列計算を避けることは現時点では困難である[24]。

分子動力学では、原子間クーロンポテンシャルは遠距離間においては著しく小さくなるため、ある一定距離以上では分子間相互作用を考えなくてもよい(カットオフ)。このためたくさんの原子を扱うことができる。ただし、この場合にも原子間距離の計算回数は $N(N-1)$ 回となるため、ポテンシャルを計算しなくても距離計算の負荷は大きい。これを避けるために、粒子登録法と呼ばれる方法により、原子の近傍粒子とそうでない粒子を分類しすることにより、ステップごとの距離計算全体を減らすことができる[25]。

本章では、並列化手法の一つである粒子分割法を用いてシミュレーションの計算処理時間を効率化することを目的とし、本研究で開発してきた分子動力学プログラムの並列化を行った。また並列処理による計算能力の向上を原子数、タイムステップ数について評価するためのベンチマークを本研究室内の PC ク

ラスタで行った。その結果，16 プロセッサの PC クラスタ上においても，並列化効率 70%，速度向上率 3 倍という結果を得ることができた。

4.2 並列処理について

4.2.1 並列コンピューティングモデル

並列計算機のアーキテクチャーは，まず大きく 2 つに分けることができる[26]。SIMD (Single Instruction Multiple Data) と MIMD (Multiple Instruction Multiple Data) である。前者では，多くのプロセッサが局所的なメモリーを有し，すべてのプロセッサは完全な同期のもとでコントローラから送られる単一の命令を同時に実行する。そしてその結果は集約され一つの結果となる。後者では，プロセッサごとにコントローラが独立して配置さえているため，各プロセッサは異なった処理を行うことができ，SIMD アーキテクチャーよりもより柔軟な利用が可能となる。近年では，MIMD 型計算機が一般的である。一方，プログラムを組む上では，PC クラスタのような完全分散メモリー型の計算機では，SPMD (Single Program Multiple Data) と呼ばれる異なるデータを同一のプログラムで処理するといった方法が一般的にとられる。分子動力学では，これまでは，SIMD 型計算機を用いたデータパラレルなモデルでプログラムが組まれることが多かったが，本研究では，SPMD 型のモデルで並列プログラムを実装した。並列化には MPI (Message Passing Interface) ライブラリである MPICH[29]を用いた。

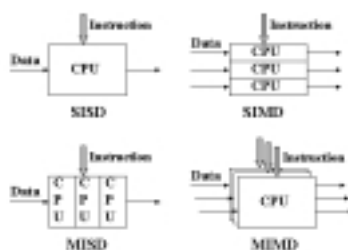


図 4.1 並列計算の方法

4.2.2 並列化効率

並列処理の効率 F は、一般に次式で表される。 P 個のプロセッサを用いたときの効率は 1 個のプロセッサで行ったときの逐次処理計算時間 T_1 を並列処理したときの計算時間 T_p と p の積で除したものである。ここで、逐次処理には最適なアルゴリズムを用いる必要がある。

$$F = P \times \frac{T_1 - T_p}{P} \quad (4.1)$$

並列処理の速度向上率を表すものとして Amdahl の法則がある。これは、並列処理の部分とそうでない部分の比率の関数として処理速度を表したもので、全体の処理速度が低速部分に大きく影響されることを示している。

$$R(f) = \frac{1}{(1-f) + f/P} \quad (4.2)$$

4.3 分子動力学の並列化

先に分子動力学の並列化の重要性は述べた。生体分子のような高分子を扱う並列分子動力学を行う際に考慮する問題としては

- ・ 一般的な並列計算機上での実行時間の定式化
- ・ プロセッサ要素ごとの計算負荷の分散
- ・ 分子内相互作用の並列化

のような点を考えなければならない。

並列計算機に分子動力学シミュレーションを実装する際には、いくつかの並列化手法の中からどの手法を選択するかが重要である。

分子動力学の並列化には、大きく分けて次の 2 通りの方法がある。

粒子分割法

分子の位置関係は考慮せず、分子の数を均等に各プロセッサに割り当てる

粒子分割法

シミュレーション空間をいくつかの領域に分割して、各プロセッサが担当する領域中に存在する分子をそのプロセッサに割り当てる

近年では、並列計算機を用いた巨大分子のシミュレーションでは、領域分割法が用いられることが多い。これは粒子分割法では、分子のすべての組み合わせについて距離を計算する必要があるため、相互作用の形式に依存せず、計算量が $O(N^2)$ である。また、全プロセッサ要素間でのデータの交換が必要であるため、プロセッサ間通信時間も大きくなる。そのため粒子分割法が用いられるのは、長距離相互作用の働く場合か、原子数が少数の場合が多い。これに対して領域分割法では、分散共有メモリ型計算機では、領域境界のセルのデータをプロセッサ要素間で交換することにより、プロセッサ間通信のオーバーヘッドを小さくでき、計算量を $O(N)$ にすることができる。ただし、領域分割法では分子の分布が領域全体で一様でなければ十分な性能がでないことも指摘されている。本研究ではタンパク質分子では空間に一様に分布しているとも限らないことを考慮し、プロセッサ要素ごとでの計算負荷の分散に焦点を絞り、粒子分割法を実装することで有効性を確かめた。

4.3.1 粒子分割法

粒子分割法は同数の原子づつを各プロセッサに割り当てる方法である。図 4.2 に直鎖分子を扱った場合の粒子分割法の概念を示す。本研究においても主鎖分子について先頭部分から均等の原子数を各プロセッサへ割り当て時間発展を計算する。この際、毎 MD ステップ毎に他の全プロセッサへ原子の位置と速度をブロードキャストして、データの通信を行う。

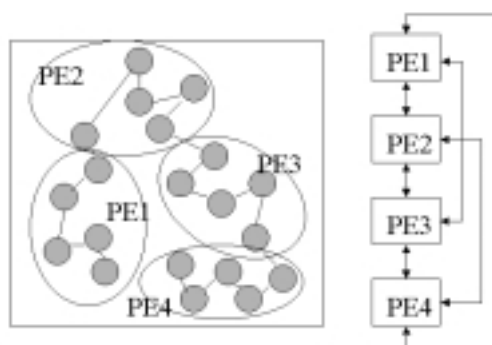


図 4.2 粒子分割法

4.4 ベンチマーク

本研究で開発した逐次処理（単一プロセッサ用）のプログラムと粒子分割法における並列化プログラムのベンチマークを行った。今回の実験では、完全分散メモリ型の PC クラスタで SPMD 型プログラムでの結果である。

ベンチマークの方法は、次の3つのことについて行った。

- (1) タイムステップを 1000 ステップと固定して、原子数を増加させたときの計算経過時間を計測した。
- (2) 原子数を 20 個と固定して、MD タイムステップを増加させたときの計算経過時間を計測した。
- (3) NAS Parallel Benchmark と呼ばれる並列コンピュータ用のベンチマークを用いて本研究室の PC クラスタ実行性能を計測した。浮動小数点演算に関する性能を評価する EP というクラスのベンチマークを行った。

ベンチマーク (1), (2) より MD プログラムの原子数の増加, MD ステップの増加に関しての並列化効率について考察を行う。

ベンチマーク (3) では、他の並列計算機と Alta Cluster との性能比を評価する。

4.4.1 用いたクラスタ構成

本研究で用いた PC クラスタ (Alta Cluster) の構成を示す。

表 4-1 クラスタ構成

| | |
|----------|--------------------------------|
| 構成 | 8 node |
| CPU | Dual PentiumIII 700 Mhz |
| DRAM | 868MB per node |
| OS | RedHat Linux 6.2 kernel.2.2.14 |
| ネットワーク構成 | Myrinet + 100Base Ethernet |
| 並列ライブラリ | MPICH |

4.4.2 原子数に関するベンチマーク

タイムステップを 1000 ステップに固定して、原子数を増加させたときの経過時間を測定した。その結果、図 4.3 のような結果が得られた。

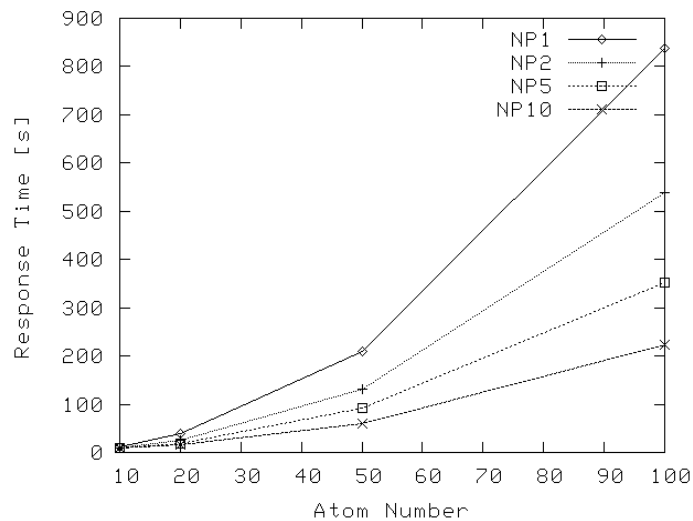


図 4.3 原子数に関する応答時間

式 4.1, 式 4.2 より、並列化率、速度向上率について以下のことが考察される。

- ・原子数 20 の時には、並列化率が 0.68 であり、速度向上は、16 プロセッサの場合には約 2.75 倍である。
- ・原子数 100 の時には、並列化率が 0.71 であり、速度向上は、16 プロセッサの場合には約 2.9 倍である。

クラスタのプロセッサ数を増やすことにより、

- ・64 プロセッサで、3.03 倍の速度向上率。
- ・128 プロセッサで、3.07 倍の速度向上率が期待できる。

これ以上、プロセッサ数を増やしても速度向上率の増加は現時点では期待できない。

4.4.3 Time Step に関してベンチマーク

原子数を 20 個に固定して、タイムステップを増加させたときの計算経過時間を測定した。その結果、図 4.4 のような結果を得た。

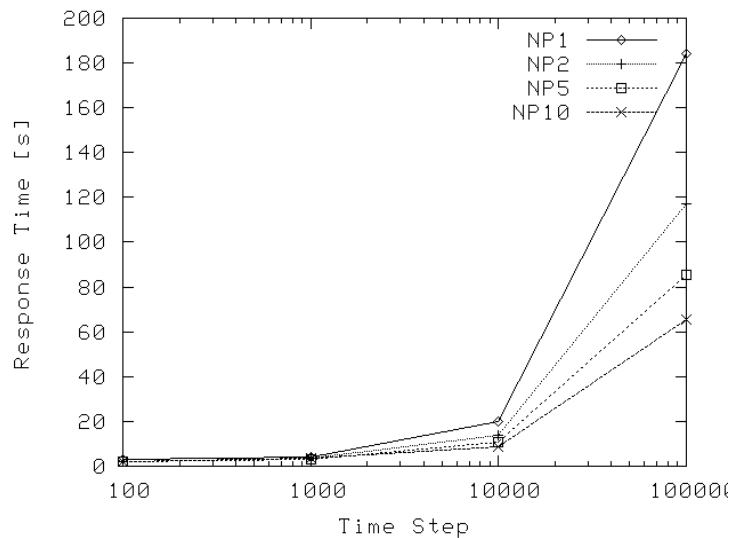


図 4.4 TimeStep に関しての応答時間

並列化率，速度向上率について以下のことが考察される。

- ステップ数 1000 の時には，並列化率が 0.23 であり，速度向上率は，16 プロセッサの場合には約 1.28 倍である。
- ステップ数 10000 の時には，並列化率が 0.72 であり，速度向上率は，16 プロセッサの場合には約 3.08 倍である。

並列化率が 0.72 のとき，クラスタのプロセッサ数を増やすことにより，

- 64 プロセッサで，3.45 倍の速度向上率。
- 128 プロセッサで，3.50 倍の速度向上率が期待できる。

プロセッサ数を増やして期待できる速度向上率は現時点では約 3.5 倍である。

4.4.4 NAS Parallel Benchmark

ここでは、本研究室で所有する PC クラスタ (Alta Cluster) が他の計算機と比較してどれほどの計算性能を持っているか評価した。性能評価には、NAS Parallel Benchmarks (NPB) を使用した。NPB は、NASA Ames Research Center で開発された並列コンピュータのためのベンチマークで、並列コンピュータの実行性能を知るうえで権威あるベンチマークの一つである [27]。NPB は、5つの Parallel Kernel Benchmarks と 3つの Parallel CFD(Computer Fluid Dynamics) Application Benchmarks から構成されている。本研究の MD シミュレーションでは、浮動少数点演算が多いため、浮動小数点演算に関する性能を評価する EP というクラスのベンチマークテストを行った。一般に、この EP がプロセッサ性能をあらわすものと評価されている。

1997 年時のスーパーコンピュータと PC クラスタについての性能評価は図 4.5 のようになっている。

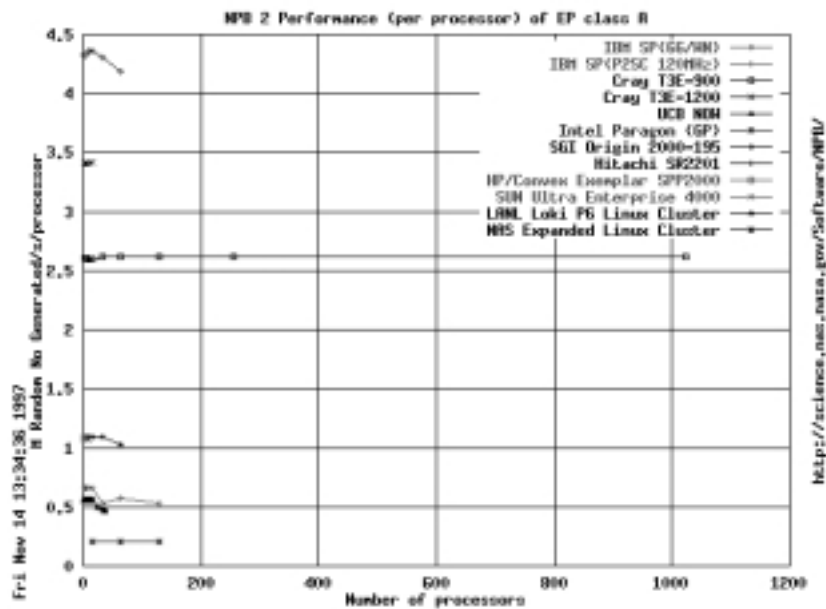


図 4.5 様々な計算機の NPB

本研究室の Alta Cluster でのベンチマークを以下に示す.

NPB EP(class A)を用いてベンチマークを行った. 比較のために, Cray T3E 1200 と NAS Expanded Linux Cluster のベンチマーク結果を示す.

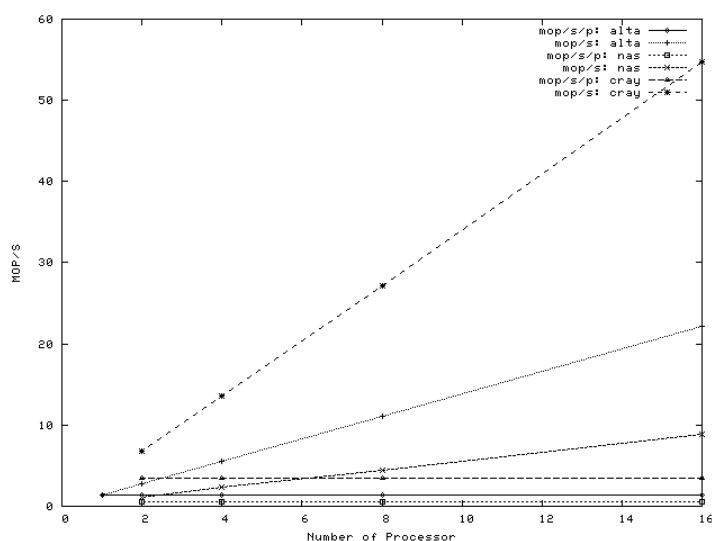


図 4.6 NPB(EP) on Alta Cluster

図 4.6 より Alta クラスタは 16 プロセッサの場合では, CrayT3E 1200 の 40% 程度の性能を示している. 実際には, CrayT3E のような超並列計算機は約 1000 個ものプロセッサを有する. タンパク質のような巨大分子を扱う場合には, より多くのプロセッサに負荷を分散させるのが当然有利ではある. しかし, 実際の分子動力学では並列化において, 多数のプロセッサを使用すればよいわけではない. MD 計算全体において並列化されていない部分が目立つようになるためである. これまでに, 並列化プロセッサ数 16 で約 90%の並列化効率が得られること, またプロセッサ数が 10 程度で計算するのが効率よいという報告もされている[28]. 16 プロセッサの場合には, Alta クラスタのような PC クラスタにおいても CrayT3E の 40%程度の性能が得られることから扱う原子数がそれほど多くない場合には, PC クラスタを使った場合でも, 十分な計算性能の向上が得られると考えられる.

4.5 考察

本章では、粒子分割法を実装して開発した並列分子動力学プログラムについて、原子数、タイムステップに関して計算時間を計測し、並列化効率と速度向上率を求めた。その結果以下のことがわかった。

原子数についてベンチマーク結果を見ると、原子の数がある程度大きくなると並列化効率が 0.7 であった。この場合には、速度向上率としては逐次処理の約 3 倍の高速化が期待できる。しかし、現行のプログラムでは 64 プロセッサを使用してもたった 3 倍の高速化しかできていない。原因としては、以下のことが考えられる。

- ・全プロセッサへのブロードキャストを毎ステップ行っており同期をとっているため、プロセッサ間通信時間が非常に長くなるため、計算処理速度がでないものと考えられる。

また、タイムステップに関しての結果図 4.4 より、ステップ数が少ないときには、負荷を分散させても並列化率が低く、速度向上率についても 2 倍も望めなかった。これはステップ数が小さかったため、ポテンシャルや力を計算する処理時間よりも、プロセッサ間通信が遅延しボトルネックになっていると考えられる。

今回、開発したプログラムでは、MPI ライブラリの中でも Broadcast や Gather と言ったプロセッサ間通信量の多い関数を使用し、MD ステップ毎に他のプロセッサへのデータの受け渡しを多量に行ったため上述の平均化効率しか得られなかった。しかし、今後はプロセッサ間通信を効率的に行えるようなデータ構造で粒子分割法を実装することや、ポテンシャル計算においてもより効率的に他のプロセッサに割り当てるといった方法を導入することでさらなる並列化効率を実現できると考えられる。

第 5 章

ま と め

本研究では、ユビキチンとユビキチン様タンパク質についてポリペプチド主鎖のフォールディングに関して分子動力学シミュレーションプログラムの開発とシミュレーションを行い、以下のような知見を得ることができた。

3 章では、計算機実験により、Ubl と Ubl(SUMO-1,NEDD8)の主鎖について安定構造を求めた。Ub と Ubl のように配列相同性を持ち構造も似ていれば、同じようなフォールディング過程で折りたたまれ、似た安定構造にたどりつくと考えられる。その際には、標的タンパクとポリユビキチンの結合部位の挙動について似たような挙動が示されるかも知れないと考えていた。しかし、計算機実験の結果、主鎖のみしか考慮していないため予想していた以上に最終的な構造は3つともトポロジーとして全く似つかない構造となった。トポロジーの似た初期構造であったが、MD シミュレーションの結果は、初期値の違いに大きく依存し、MD で求めた安定構造ではトポロジーの類似は認められなかった。炭素主鎖だけというように、モデルを単純化しすぎたことで、他原子に関してエネルギー安定性において大きな違いができてしまうため、予想に反した結果となってしまった。また今回の数値実験では、系のエネルギーの高い初期構造からたった 5[ps]で絶対 0 度まで冷却するという条件にも問題があったことも考えられる。初期構造でのエネルギーの極小化や、室温環境下（細胞内環境下）の温度での構造揺らぎを観測し、揺らぎが一様になってからの構造について挙動を考察することで特異的な特徴を観測できた可能性も考えられる。

4 章では、タンパク質のような巨大分子シミュレーションを行うために、粒子分割法を用いた分子動力学の並列化を行った。そして、PC クラスタ上でその計算

時間の速度向上性について考察を行った。その結果、粒子分割法による並列化で、完全分散メモリ型 PC クラスタにおいても並列化効率 70%、速度向上率 3 倍といった性能向上を示した。これは決して高効率な値ではないが、最も単純な MPI 通信関数のみを使って実装していることを考えると、今後改善の余地はたくさんあると考えられる。これまでの大型計算機に頼った方法ではなく、今後は PC クラスタ上での分子動力学シミュレーションもより現実的になってきていると考えられる。

最後に、今後の課題として以下のことが挙げられる。

- ・ タンパク質分子のシミュレーションを行ううえで、精度を向上させるためには、他の原子種や、水素の付加、水溶媒といったモデル化、境界条件の考慮が必要不可欠である。
- ・ 並列化に関しては、領域分割法の実装や動的な負荷分散でさらに並列化効率を向上させることが必要である。

謝 辞

本研究は、遺伝子知識システム論講座において小長谷明彦教授の御指導のもとで行われたものです。二年間に渡って常に適切かつ有意義な御指導，ご助言を頂きました小長谷明彦教授に感謝の意を表し，心から御礼申し上げます。

同講座の佐藤賢二教授には，本研究を進めるにあたり有益なご助言，ご示唆を頂きました。ここに記して感謝の意を述べます。

同講座の高橋勝利助手，Xavier Defago 助手にも研究に関し有益なご助言をたくさん頂きました。ここに感謝の意を述べます。また，副テーマを通じて本多卓也教授，前一樹助手にも研究に関してたくさんのご助言を頂き，大変お世話になりました。ここに感謝の意を述べます。

遺伝子知識システム論講座のみなさまには，常日頃から共に研究を行う仲間として，有意義な研究室生活を送ることができ大変感謝しています。また，田中芳幸さんには，研究その他において数多くの有意義なコメントを頂きました。

ここに感謝の意を述べます。最後になりますが，この2年間に渡り公私に関わらず，至るところで有意義なアドバイスを頂き，またよき手本をたくさん見せていただきました博士後期課程の金大心先輩に心から感謝致します。

参 考 文 献

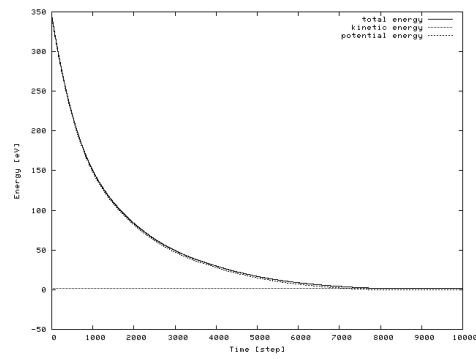
- [1] Tanaka,K.Suzuki,T.,Chiba.T,Mol.Cells,8,pp503-512,1998
- [2] Berman,H.M.,Westbrook,J.,Feng,Z.,Gilliland,G.,Bhat,T.N.,Weissig,D., Shindyalov,L.N.,and Bourne.PE."The Protein data Bank",Nucleic Acids Res,28,235-242,2000
- [3] Murzin,A.G.,Brenner,S.E.,Hubbard,T.,andClothia,C."SCOP:a structural classification of proteins database for the investigation of sequences and structures",J Mol Biolo,247,536-540,1995
- [4] Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M, CATH- A Hierarchic Classification of Protein Domain Structures. Structure. Vol 5. No 8. p.1093-1108,1997
- [5] L. Holm, C. Sander Nucl. Acids Res. 26, 316-319,1998
- [6] <http://www.ncbi.nlm.nih.gov/Entrez/>
- [7] G.P. Tsurupa, T. Ikura, T. Makio and K. Kuwajima: Refolding Kinetics of Staphylococcal Nuclease and Its Mutants in the Presence of the Chaperonin GroEL. J. Mol. Biol. 277, pp733-745,1998
- [8] Anfinsen,C.B."Principles that govern the folding of protein chains",Science,181,223-230,1973
- [9] Onuchic,J.N.et al.,"Protein folding funnels ,the nature of the transition state ansemble.Folding&design,1,441-450,1996
- [10] Dill,K.A.,Chan,H.s:From Levinthal to pathways to funnels.Nature Struct.Biol.,4,10-19,1997
- [11] 中村春木, タンパク質のかたちと物性, 共立出版, pp91-92 , 1998
- [12] 河村雄行, パソコン分子シミュレーション, 海文堂, pp1-6, 1990
- [13] 上田顕, コンピュータシミュレーション, 朝倉書店 , pp10-43, 1990
- [14] 神山新一, 佐藤明, 分子動力学シミュレーション, 朝倉書店, pp15-26, 1997

- [15] Komeiji, Fast and accurate molecular dynamics simulation of a DNA binding Protein using special-purpose computer, *J.Comp.com.*,18,1546-1563,1997
- [16] Soren Toxvaerd, Comment on constrained molecular dynamics of macromolecules, *J.Chem.phys.*87, pp6164-6143, 1987
- [17] 田中啓二, 生命活動を解く鍵 ユビキチン-プロテアソームシステム, *細胞工学*, vol.18,No.5,pp610-614,1999
- [18] Patton,E.E.,Willems,A.R.,Tyers,M.,*Trends Genet.*,14,pp236-243,1998.
- [19] Saitoh,H.,Pu,R.T.,Dasso,M. *Trends Biochem.Sci.*,22,pp374-376,1997
- [20] Kumar,S.,Yoshida,Y.Noda,M.*Biochem.Biophys.Res.Commun.*,195,pp393-399,1993
- [21] 千葉智樹, 逢坂文夫, ユビキチン様蛋白質とその蛋白質修飾システム, *蛋白質 核酸 酵素*, vol.44, 5, pp744-751,1999
- [22] Stephan L. Mayo, A Generic Force Field for Molecular Simulations, *J.Chem.phys.*94, pp8897-8909, 1990
- [23] 西村健, 分子動力学シミュレーションの並列化における分割法の一考察, *ハイパフォーマンスコンピューティング*, pp61-66, 1999
- [24] Emerging Computational Opportunity: (Requirements for Solution in 1 week) <http://www.cise.nsf.gov/asc/PaulYoung/sld008.htm>
- [25] 高井 昌彰, 江丸裕教, 山本強, 動的粒子集団分割による分子動力学法の並列処理, *電気情報通信学会誌*, vol.J-82-D-I,No.6,pp711-717,1999
- [26] 飯塚肇, 緑川博子, 並列プログラミング入門, 丸善, pp3-9,1999
- [27] <http://www.nas.nasa.gov/Software/NPB/>, NASA Ames Research Center
- [28] Komeiji,Y.,Uebayashi,M.,Molecular dynamics simulation of biological molecule(1) *Methods*, *J. Chem. Software* 6, pp1-36,2000
- [29] MPICH, 米国アルゴンヌ国立研究所,<http://www.mcs.anl.gov/mpi>

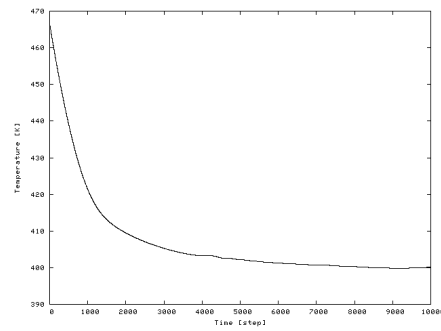
付 録

NVT アンサンブルでの結果

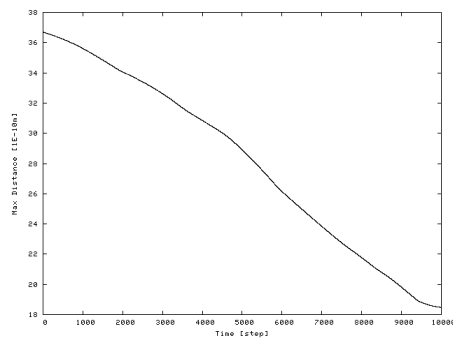
Ubiquitin：設定温度 400[K]のときの結果



(a) Energy



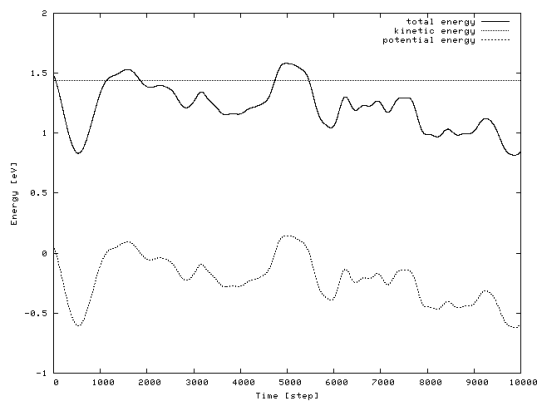
(b) Temperature



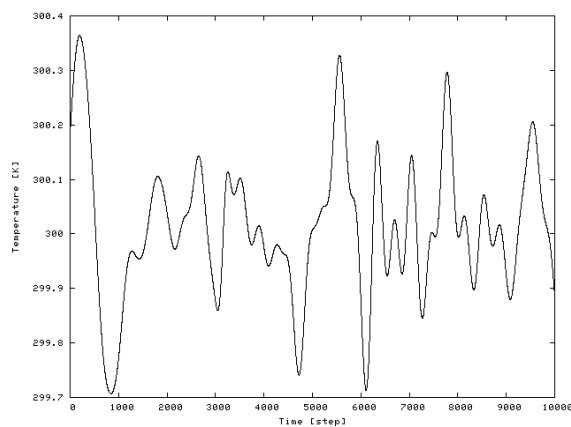
(c) Maximum Distance

付録図 1 Ubiquitin at 400 [K]

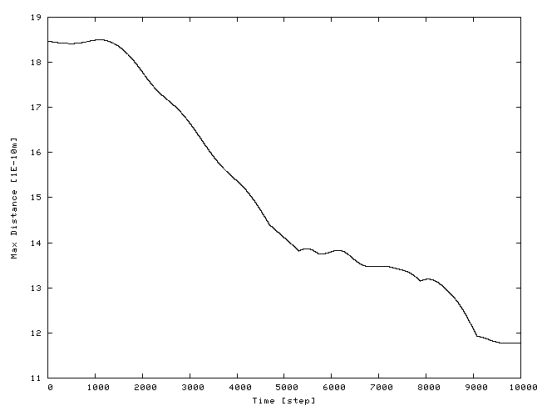
Ubiquitin : 設定温度 300[K]のときの結果



(a) Energy



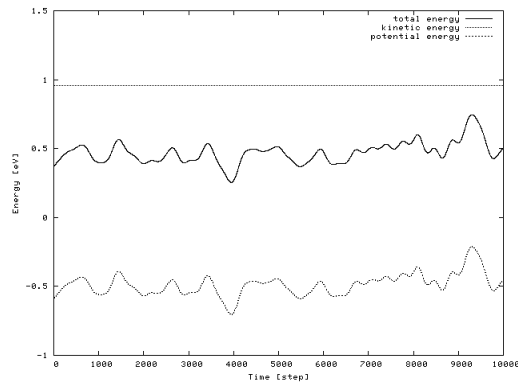
(b) Temperature



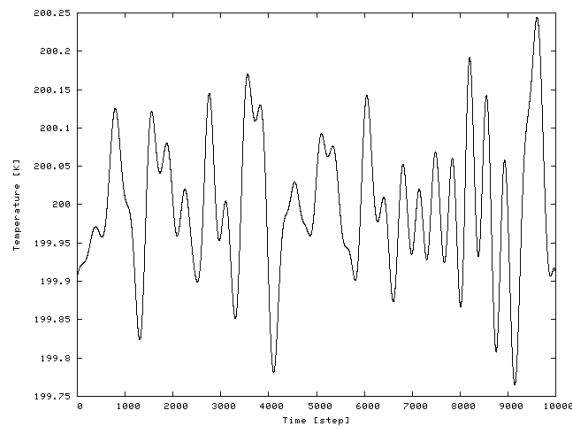
(c) Maximum Distance

付録図 2 Ubiquitin at 300 [K]

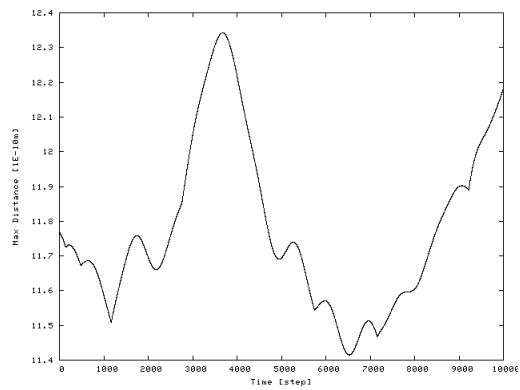
Ubiquitin: 設定温度 200[K]のときの結果



(a) Energy



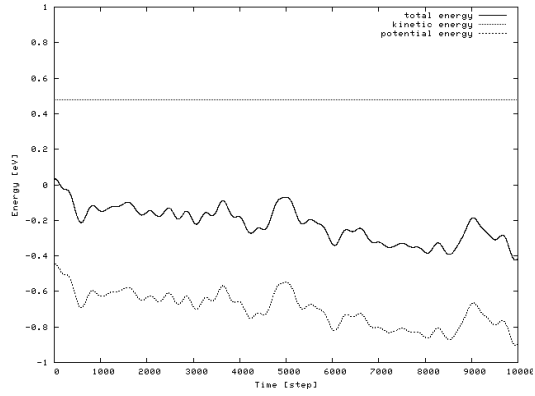
(b) Temperature



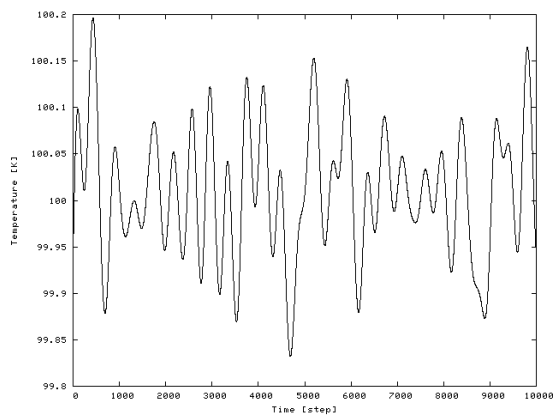
(c) Maximum Distance

付録図 3 Ubiquitin at 200 [K]

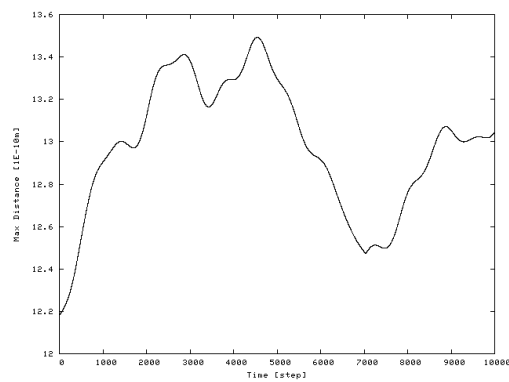
Ubiquitin : 設定温度 100[K]のときの結果



(a) Energy



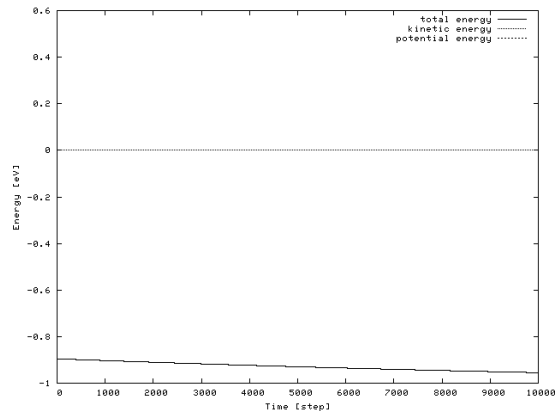
(b) Temperature



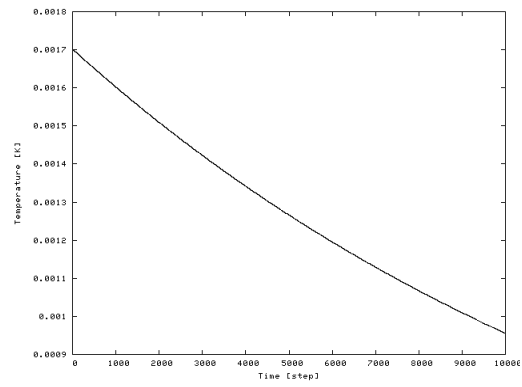
(c) Maximum Distance

付録図 4 Ubiquitin at 100 [K]

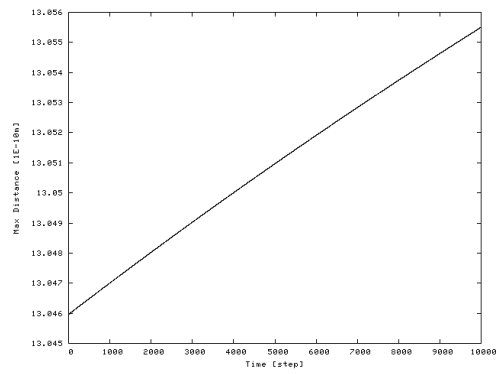
Ubiquitin : 設定温度 0[K]のときの結果



(a) Energy



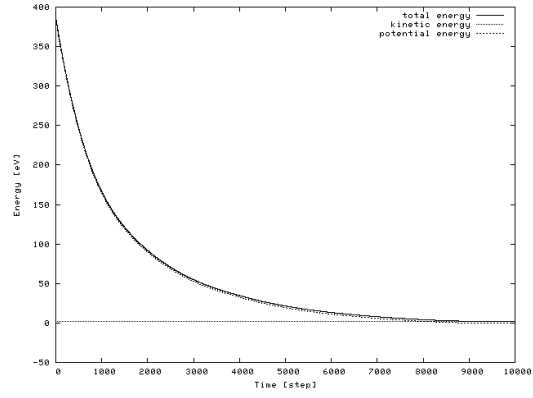
(b) Temperature



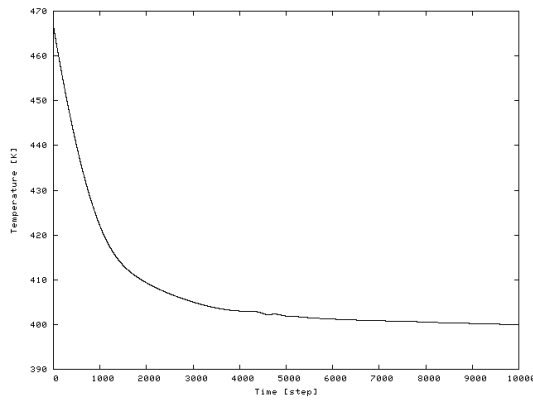
(c) Maximum Distance

付録図 5 Ubiquitin at 0 [K]

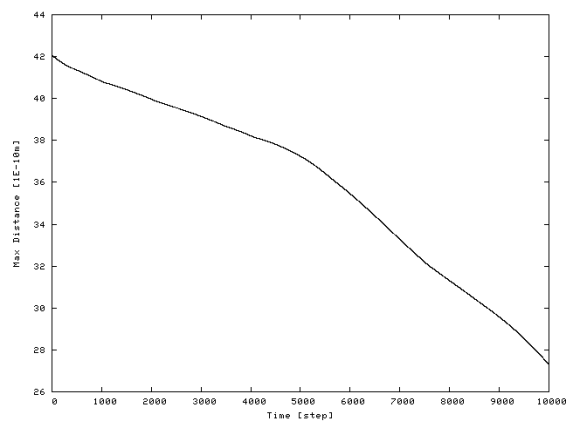
SUMO-1 : 設定温度 400[K]のときの結果



(a) Energy



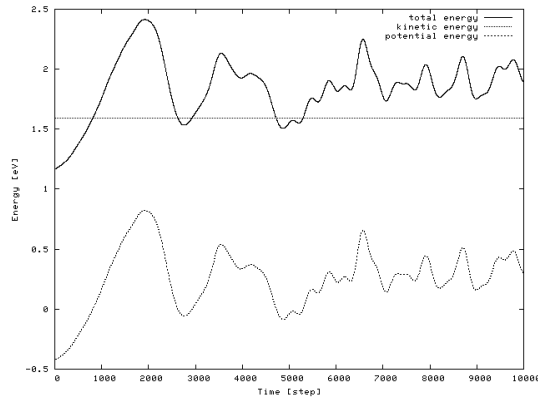
(b) Temperature



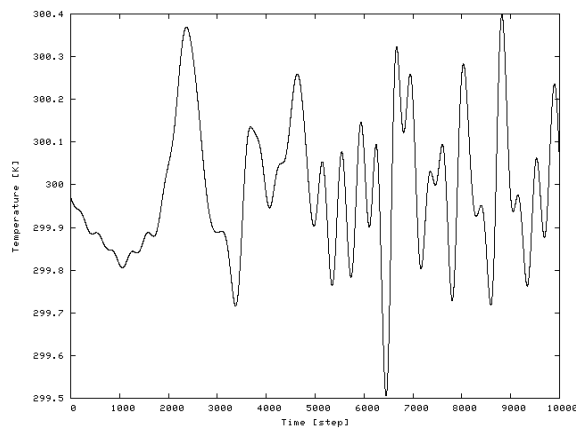
(c) Maximum Distance

付録図 6 SUMO-1 at 400 [K]

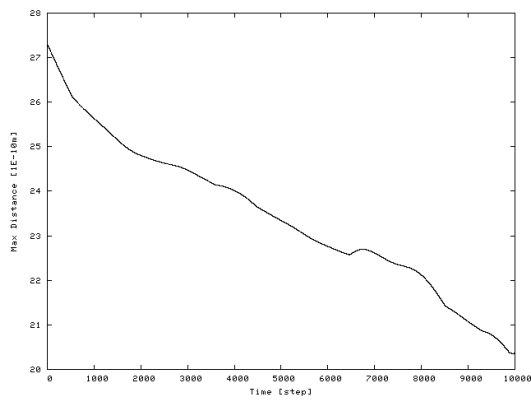
SUMO-1 : 設定温度 300[K]のときの結果



(a) Energy



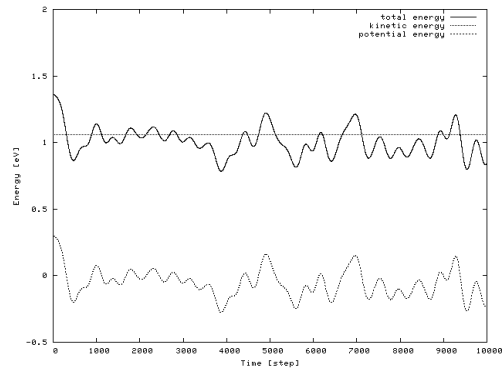
(b) Temperature



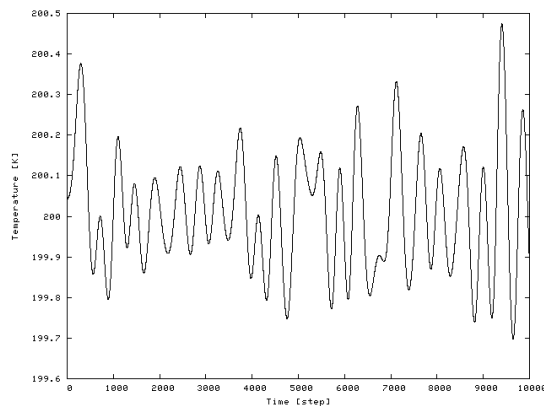
(c) Maximum Distance

付録図 7 SUMO-1 at 300[K]

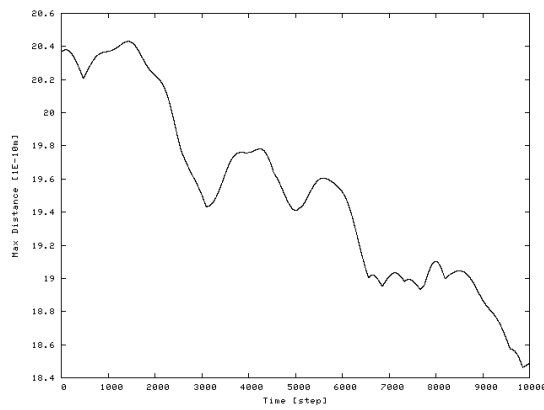
SUMO-1 : 設定温度 200[K]のときの結果



(a) Energy



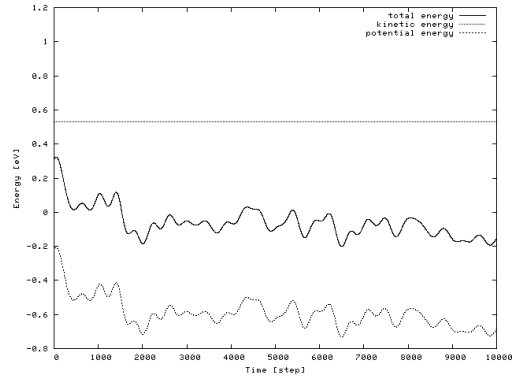
(b) Temperature



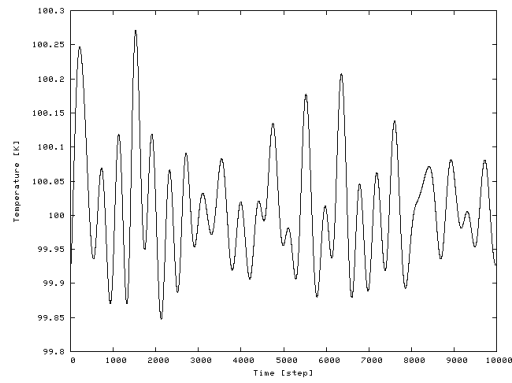
(c) Maximum Distance

付録図 8 SUMO-1 at 200[K]

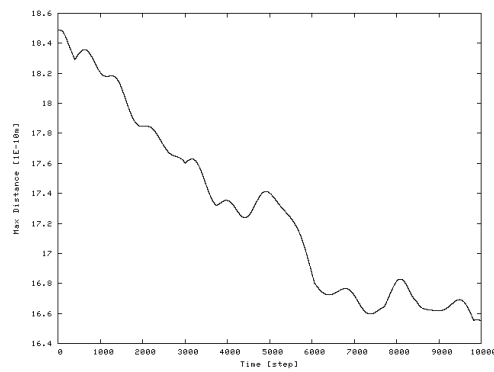
SUMO-1 : 設定温度 100[K]のときの結果



(a) Energy



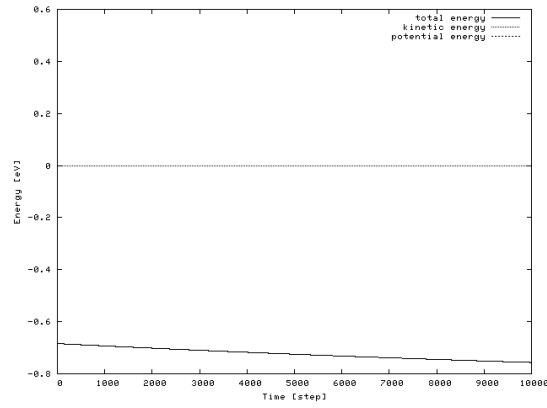
(b) Temperature



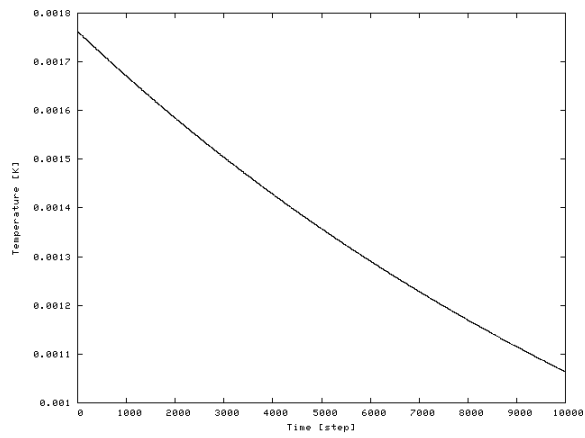
(c) Maximum Distance

付録図 9 SUMO-1 at 100[K]

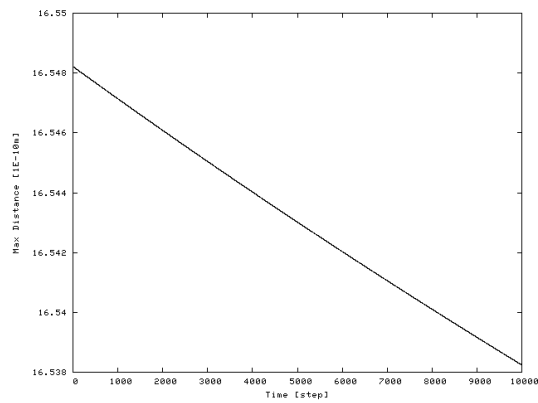
SUMO-1 : 設定温度 0[K]のときの結果



(a) Energy



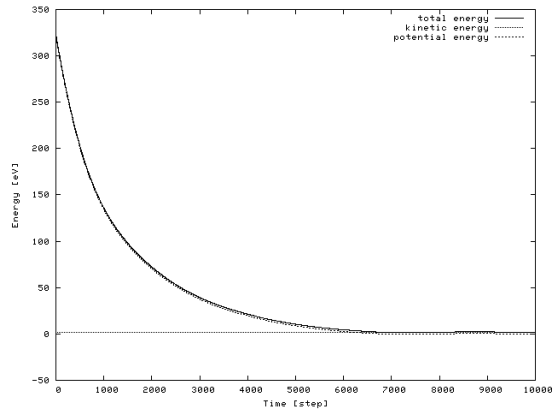
(b) Temperature



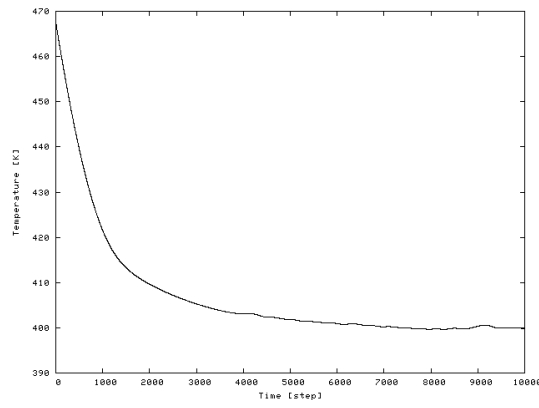
(c) Maximum Distance

付録図 10 SUMO-1 at 0[K]

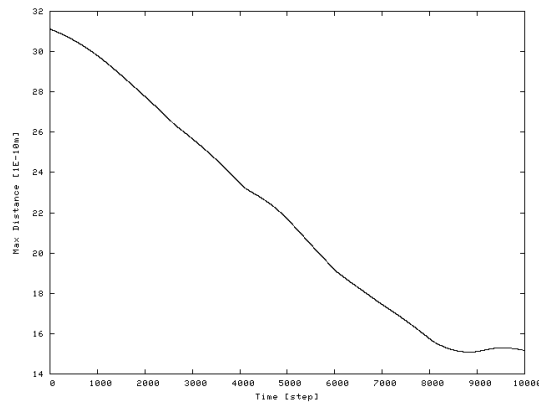
NEDD8 : 設定温度 400[K]のときの結果



(a) Energy



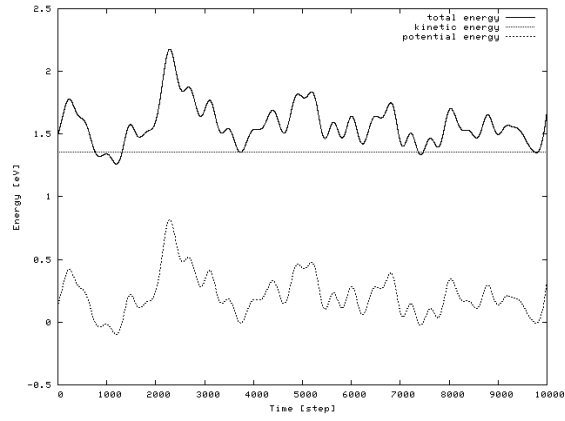
(b) Temperature



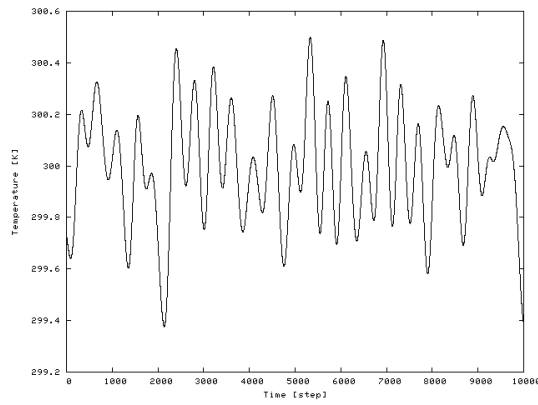
(c) Maximum Distance

付録図 11 NEDD8 at 400 [K]

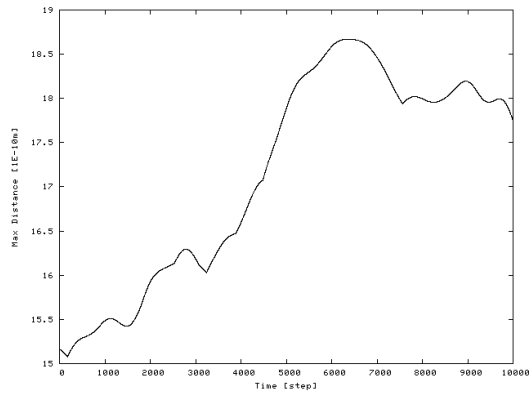
NEDD8 : 設定温度 300[K]のときの結果



(a) Energy



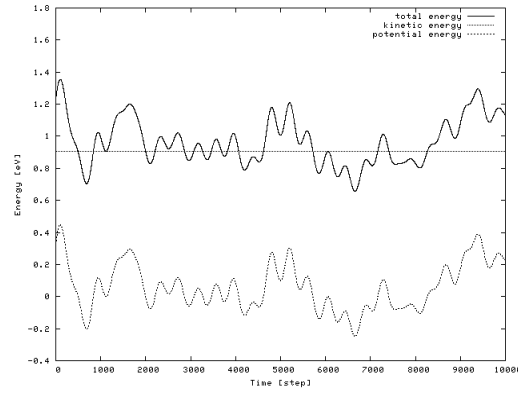
(b) Temperature



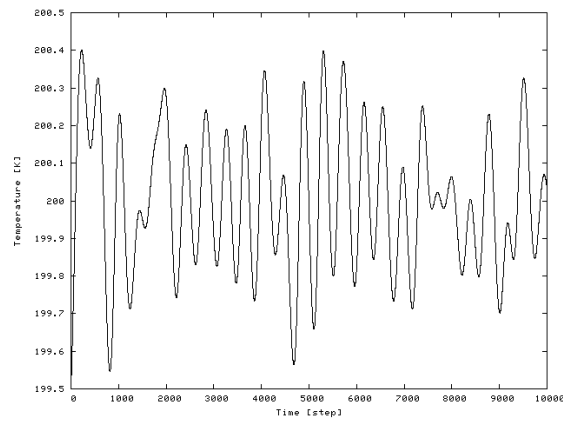
(c) Maximum Distance

付録図 12 NEDD8 at 300 [K]

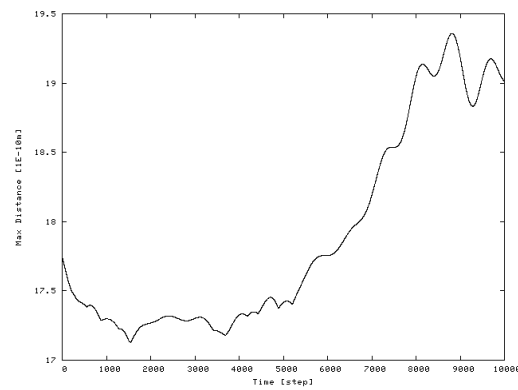
NEDD8 : 設定温度 200[K]のときの結果



(a) Energy



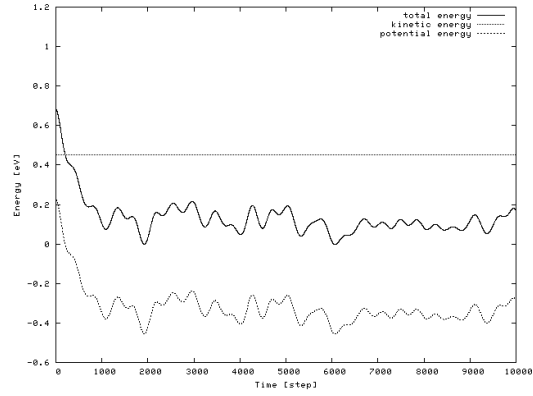
(b) Temperature



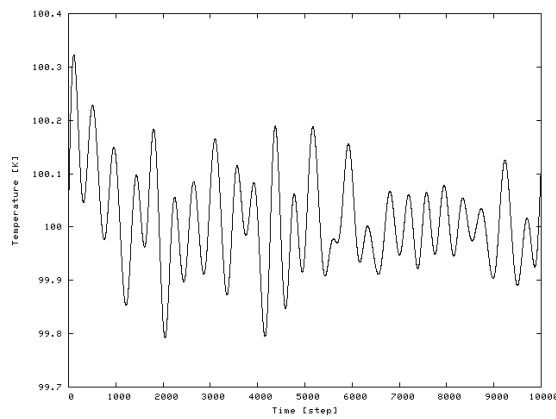
(c) Maximum Distance

付録図 13 NEDD8 at 200 [K]

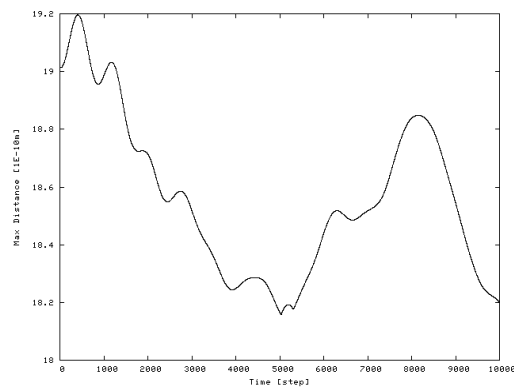
NEDD8 : 設定温度 100[K]のときの結果



(a) Energy



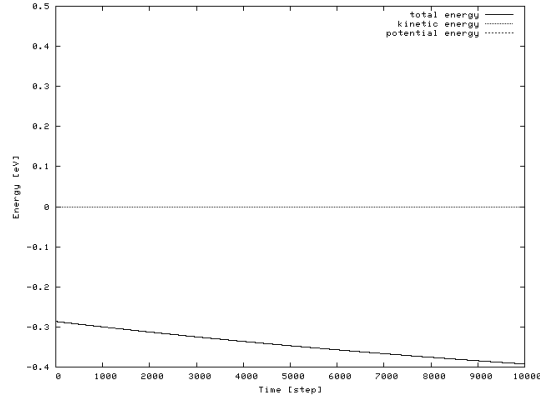
(b) Temperature



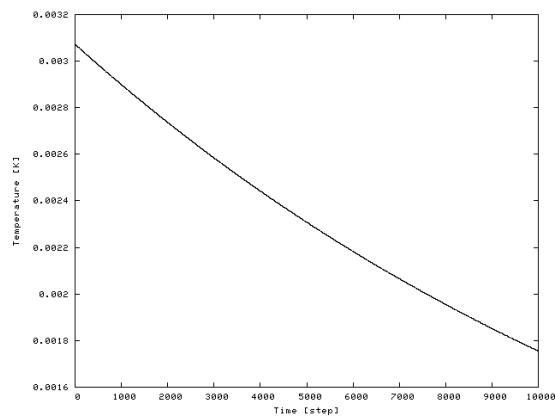
(c) Maximum Distance

付録図 14 NEDD8 at 100 [K]

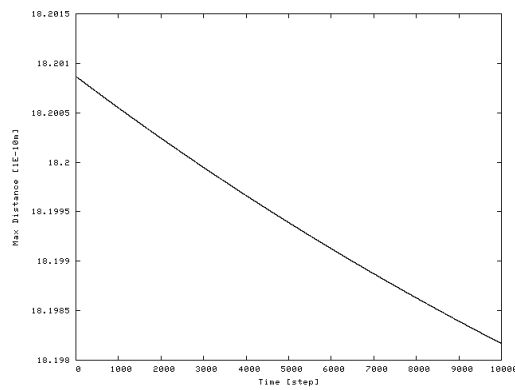
NEDD8 : 設定温度 0[K]のときの結果



(a) Energy



(b) Temperature



(c) Maximum Distance

付録図 15 NEDD8 at 0 [K]