平成15年度知能システム科学専攻修士論文

分子動力学シミュレーションにおける力場パラメー タの推定と解析

皆川 恵一

An Analysis of Estimated Force Field Parameters for

Molecular Dynamics

Keiichi Minakawa

提出年月日 2004 年 2 月 23 日
主查教官 山村 雅幸 教授
審查教官 新田 克己 教授
審查教官 小林 重信 教授

再提出時の変更箇所

- ページ番号の修正
- 英文要約の修正
- より明示的な参考文献の付加
- (第1章)セクション組みの変更
- (第1章)研究の目的を加筆
- (第5章)図中の誤りを修正
- •(第6章)本研究の最終的な目標に対する適用手法の妥当性の検討を加筆

分子動力学シミュレーションにおける力場パラメー タの推定と解析 皆川 恵一

An Analysis of Estimated Force Field Parameters for Molecular **Dynamics**

Abstract

In this paper, we estimated force field parameters for Molecular Dynamics(MD) and analyzed those estimated parameters. For MD Simulation, we focused on protein structures, and we used structure fitting, i.e. RMSD (root means square deviation), as a fitness function when we estimated those parameters. Analysis of those parameters were implemented by comparison trajectory which was resulted by MD simulation with original parameters, to that with estimated parameters. After that, we observed those differences of trajectory and discussed the usefulness of estimated parameters.

はじめに 1

1.1 研究の背景

分子動力学 (MD) シミュレーションは分子間相 互作用の動的な性質を解析するための手法として、 1970年頃から多くの研究に用いられてきた。近年 では、計算環境の性能が飛躍的に向上した事によ り、分子数がある程度大きくても MD シミュレー ションを適用する事が可能になりつつある。例え ば、タンパク質の構造解析や折り畳みのシミュレー ションなどの生体高分子化合物へも適用され始め ている。また、タンパク質の機能と構造の相関か ら、タンパク質の構造解析における MD シミュレー ションの重要性はさらに増している。

MD シミュレーションでは、シミュレーション 対象の分子を構成する原子に対してそれぞれ運動 方程式を立て、これを差分近似法などを用いて解 く。そして得られたトラジェクトリーを解析して、 興味ある分子の物性値や構造を得ようというもの ラメータであるのかは分らず、このパラメータを である。しかしながら、MD シミュレーションを 用いてシミュレーションされた構造やトラジェク 生体高分子に対して適用する際に、以下の二つの 問題が指摘されている。

1.1.1 初期構造への依存

まず、生体高分子などの複雑な系ではエネル ギーの極少状態が無限に存在するために、従来の カノニカル・アンサンブル (温度一定)の MD シ ミュレーションでは、それらのエネルギー極少状 態に留まってしまうという問題である。つまり、 シミュレーション対象の分子の初期構造に強く依 存してしまうのだ。この問題に対し岡本らは、統 計力学の知識を用いて「拡張アンサンブル法」と いう手法を提案し、問題解決を試みている[1]。

1.1.2 力場への信頼性

次に、シミュレーションで用いられる力場の精度 に対する信頼性の問題である。MD シミュレーショ ンでは分子間相互作用を汎用的なポテンシャル関 数の形で表現していて、有名なものに AMBER の 力場がある。この関数は原子ごとに異なる多くの パラメータを持っている。しかしながら、これら パラメータは生物学や物理化学の知識を用いて経 験的に決定されているため[2]、大域的に最適なパ トリー、動的な性質に対する信頼性が保証されて いるとは言えない。また、これらパラメータの値

は、ある化合物での実験値を再現するように決定 されているが、その値を他の化合物に対しても適 用することには妥当性に疑問が残る。つまり、汎 用的な関数である既存の力場に対し、画一的なパ ラメータで全てのシミュレーションを網羅してい る事は大きな疑問であり、シミュレーション対象 の化合物ごとに最適なパラメータが存在するはず である。

1.2 研究の目的

カ場への信頼性への取り組みとしては、Faller らが tetrahydrohuran 等の化合物に対して実験から 得られた物性値 (圧力や分子振動など)をうまく再 現するパラメータを、シンプレックス法 (Simplex Algorithm)を用いて推定を行っている [3]。また、 Wang らは Faller らと同様の実験を Genetic Algorithm(GA)を用いて行っている [4]。しかし、彼等 が推定に用いた実験データは常に手に入手可能で あるとは限らない。また推定されたパラメータが 最良である保証はなく、これらの値も他の化合物 に対して有効であるかは分からない。

近年、X 線回析実験や NMR(Nuclear Magnetic Resonance:核磁気共鳴) によって決定されたタンパ ク質などの構造データが、PDB(Protein Data Bank) に登録され、そのエントリー数は 20000 を超えて いる。

そこで、この力場の信頼性の問題に対して一石 を投じるために、本研究では、この誰もが入手可能 なタンパク質の MD シミュレーションを研究の対 象とし、PDB に登録されているタンパク質の立体 構造データが天然状態である、つまりエネルギー が最小の状態であると仮定し、この構造をよく再 現する力場パラメータをシンプレックス法を用い て自動的に決定し、得られた力場パラメータの有 用性を検討する。その上で、経験的なパラメータ、 画一的なパラメータという問題に対する本研究が 適用した手法の方向性について検討を行う。

1.3 論文の構成

第1章では本論文の背景と研究の目的について 述べた。第2章では取り扱う問題に対する知識と して力場パラメータについて述べ、力場パラメー タの値に対する MD シミュレーションの結果の依 存性について述べる。第3章では力場パラメータ の推定の詳細な手法について述べる。第4章では 得られた実験結果と考察を行い、第5章で本論文 をまとめる。

2 準備

2.1 分子動力学法

MDシミュレーションのソフトは、カリフォルニ ア大学の Amber やアクセルリス社の CHARm が 有名である。国内では産業総合技術研究所の古明 地氏が作成した PEACH も有名である。その中で も最も有名な AMBER で用いられている標準的な 力場を式 (2.1) に示す。

$$U(R) = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_{\theta} (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} (1 + \cos[n\phi - \gamma]) + \sum_{i < j}^{atoms} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \sum_{i < j}^{atoms} \frac{q_i q_j}{\epsilon R_{ij}}$$
(2.1)

式(2.1)は、シミュレーションを行うシステム内 の全粒子についての、分子内相互作用、分子間相 互作用をポテンシャルエネルギー関数の形で記述 されている[5]。各分子の運動方程式は、式(2.1) を距離で微分すれば得る事ができて、MD法では この運動方程式を差分近似法などによって、分子 の運動の時間発展をシミュレーションする。そし て興味ある物性値や分子同士の相互作用、エネル ギーや構造を観察する。 式(2.1)の最初の3項は分子数がNのN体の系 ではO(N)の計算量で済む。しかし、4項目のvan der Walls 力の項と5項目のクーロン力の項は、そ のまま計算を行うとO(N²)の計算量である。その ために、計算量を減らすための様々な工夫がなさ れている。例えば、どこまで遠くの粒子までを計 算に含めるかの cut-off 値の設定もその一つである [5]。また、その他では、溶媒の取り扱いにも工夫 がなされている。タンパク質などの化合物に対し て MD シミュレーションを行う場合は、一般に水 などの溶媒をシミュレーションする系に含める。 MD シミュレーションでの溶媒の取り扱いは Fig. 2.1に示した。



Fig. 2.1 MD における溶媒の取り扱い

真空中でのシミュレーションでは、(1)の孤立系に 示す形態でシミュレーションが行われる。しかし ながら、通常は溶質分子が真空に存在することは なく、真空で MD シミュレーションを行うと溶媒 分子などの環境の及ぼす影響の不足から、分子の 形状が壊れてしまったり、安定したトラジェクト リーを得る事が出来なかったりするが、システム 内に含める粒子数が最も少ないので計算時間は最 も少ない。しかし、MD シミュレーションでは通 常は(2)か(3)の溶媒の影響を考慮したモデルを用 いて行う事が多い。(2) は溶媒を流体として捉え て、連続体近似を用いて溶媒の影響をシステムに 含める。MD シミュレーションに用いられる連続 体近似の手法では GB(Genetic Born) モデルが有名 である。GBモデルでは(1)と比較すると計算時間 は大きくなるが、計算精度は格段によくなる[6]。

また、(3)では(2)とは対照的に明示的に溶媒分子 をシステム内に含める。しかし、計算を行う事の 出来る溶媒分子の数は計算時間の観点から限られ てしまうため、実際には中央のセル内の分子だけ 計算を行い、セル間の溶媒分子の出入りを考慮し た周期境界条件(密度を一定に保つ等)を用いる事 でシステム全体の溶媒の影響を近似する。しかし ながら、(2)に比べると計算時間は大きいがシミュ レーション精度はよい事が知られている[7]。

2.2 力場パラメータ

式(2.1)には様々なパラメータが含まれ、これら は力場パラメータと呼ばれる。Amber では、それ らのパラメータをファイルに書き込んでパッケー ジに含めて配布されていて、例えば AMBER99 と 呼ばれるパラメータは、式(2.1)の第4項を式(2.2) の形に変形して、そのパラメータがファイルに書 き込まれている (Fig. 2.2)。

$$V(R) = 4\epsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^6 \right\}$$
(2.2)

atom name σ[Å]		ε [kcal/mol]	feature
0	1.6612	0.2100	OPLS
02	1.6612	0.2100	OPLS
OW	1.7683	0.1520	TIP3P water model
OH	1.7210	0.2104	OPLS
OS	1.6837	0.1700	OPLS ether
C*	1.9080	0.0860	Spellmeyer
CT	1.9080	0.1094	Spellmeyer
С	1.9080	0.0860	OPLS
N	1.8240	0.1700	OPLS

Fig. 2.2 AMBER99 の LJ パラメータの例

式 (2.2) は Lennard-Jones 6-12(LJ) ポテンシャル関 数と呼ばれ、その概観を Fig. 2.3に示す。

共有結合をしていない原子対は、近い距離だと 反発し、遠くなると小さな力で引き付けあう。LJ ポテンシャルは非共有結合力の1つである分子間 力 (van der Walls force)を表現するものである。12 乗の項が反発力を、6乗の項が引き付けあう力を表 現し、このポテンシャルのパラメータである ϵ と σ はそれぞれ、ポテンシャルの井戸の深さと VDW 半径を示す [8]。このポテンシャル関数は、MD法



Fig. 2.3 Lennard-Jones 6-12 potential

における分子間力を記述するためのポテンシャル 関数として最も一般的なものであり、分野を問わ ず広く使用されている。

しかしながら、これらの力場パラメータは中性 子回析や分子力学、量子力学や格子結晶解析など によってある程度の値を推定し、その後、実際の 化合物の実験から得られている様々な物性値を上 手く再現するように手動で調節される。例えば、 methyl-ethyl-etherやtetrahydrohuran等の化合物に 対して、実験で得られている物性値、具体的には 糖のパッカリングや立体構造、分子振動に関する データやエネルギー等に近い値を再現する事が出 来るようにtry-and-errorで調節されている[2]。

2.3 シミュレーション結果の依存性

カ場パラメータが MD シミュレーションによっ て得られる結果に及ぼす影響を調べるために、Protein G(PDB ID:1PGB) を用いて、炭素原子に関す る LJ パラメータを様々な値に変更して MD シミュ レーションを行った。この MD シミュレーション は、1PGB の構造を初期構造として 20[ps] のトラ ジェクトリーを得て、力場パラメータの値の違い の影響を調べる。炭素原子の既存のパラメータは、 Table 2.1に示す。

	$\sigma_C[\text{Å}]$	ϵ_C [kcal/mol]
original	1.908000	0.086000

Table 2.1 炭素原子の AMBER99

結果は Fig. 2.4から Fig. 2.5に示した。Fig. 2.4では VDW 半径を表す σ_C の値を 0.5 刻みで 0.0 ~2.0 に変更してシミュレーションを行い、PDB の構造との RMSD(Root Means Square Deviation: 平均二乗誤差)を計算しプロットした。 σ_c を変更 する事で、原子間距離に影響を与え、分子全体と しての構造が大きく変化してしまう事が分かる。 また、Fig. 2.5ではポテンシャルの井戸の深さの係 数である € の値を 0.01 刻みで 0.05~0.09 に変更 してシミュレーションを行い、同様に RMSD をプ ロットした。

を変更する事で、原子間の距離を 一定に保つための安定力が変わり、安定したトラ ジェクトリーが得られ難くなってしまう事が分か る。よって力場パラメータの値を変更すると、MD シミュレーションによって得られるトラジェクト リーが大きく変わってしまう事が観察できる。ま た、各シミュレーションでの 20[ps] を経過した時 点でのポテンシャルエネルギーの値は、 σ_{C} の値 の変更により 100[kcal/mol] 程度、 *c* の値の変更 により 50[kcal/mol] 程度のずれが生じていた。



Fig. 2.4 RMSD (*σc* を変更)



Fig. 2.5 RMSD (*c* を変更)

3 力場パラメータの推定

3.1 シンプレックス法

カ場パラメータの推定では、最適化の手法とし てシンプレックス法を選択した。推定中で行われる 評価計算にMDシミュレーションを行っているが、 このシミュレーションが必要とする計算時間が非 常に大きいため、局所最適化の手法の一つである シンプレックス法を用いた。また、シンプレックス 法は初期集団であるシンプレックスの配置に注意 をすれば、最適化が発散してしまう恐れが少ない。 シンプレックス法の詳細については、付録.Aにて 後述する。本研究ではシンプレックス法における 各イタレーションごとのシンプレックスの持つ評 価値の分散の値が 0.001 以下になったら収束とみ なす。

3.2 MD シミュレーション

シンプレックス法の各評価値の計算では、それ ぞれのシンプレックスが持っている力場パラメー タを用いて MD シミュレーションが行われる。MD シミュレーションは以下の手順で行われる。

(1) 溶質分子を固定したエネルギー最小化計算

(2) 本計算

溶質分子を固定したエネルギー最小化計算とは、 溶質分子が固定され溶媒分子だけが自由な状態に あるために、溶質分子の隙間に溶媒分子を行き渡 らせる等の効果があり、MDシミュレーションで は常套的な手段である [5]。このエネルギー最小 化計算を行った後で20[ps]の本計算を行い、評価 値の計算を1[ps]ごとに実行する。また溶媒に関 しては、計算時間の節約とシミュレーション精度 の2つの観点から考えGBモデルを選択した。

3.3 評価関数

シンプレックスの評価値は、それぞれのシンプ レックスが持つ力場パラメータを用いた MD シミュ レーションの結果から計算される。MD シミュレー ションで分子の時間発展のデータ、つまりトラジェ クトリーが得られる。そして、MD シミュレーショ ンを行っている最中に、0.2[*ps*] 刻みでその時点で の構造を抽出し、PDB のオリジナルの構造との原 子間の RMSD を計算する (式 (3.1))。

$$RMSD = \sqrt{\sum_{i \in atoms} (Ref_i - Target_i)}$$
(3.1)

参照するトラジェクトリーは、計算中のシミュレー ション中のその時点より 10[ps]分の過去のトラジェ クトリーを用いている。そして、式(3.1)より計算 される RMSD 値から計算される分散が 0.05 以下に なったら計算を打ち切り、過去 10[ps]の RMSD 値 の平均値を評価値とする。つまり、PDB から得た オリジナルの構造に近い構造でエネルギーが収束 をしていた方が評価値が高くなる。参照するトラ ジェクトリーを過去 10[ps] に制限しているのは、X 線回析によって得られた構造を用いて MD シミュ レーションを行うと、溶媒を付加した影響で開始 直後は構造が揺らいでしまう。この構造の揺らぎ を評価値の計算に含めずに、エネルギーの収束し ている状態での構造を比較するためである。

3.4 推定を行うパラメータ

Amber 力場には、多数の力場パラメータが存在 している事は前述した。しかし、全てのパラメー タに対して最適化を行う事は、計算時間の大きさ から現実的ではない。そこで、本研究で推定を行 う力場パラメータは、Lennard-Jones 6-12 の ϵ_{ij} と σ_{ij} を選択し、その中で炭素原子のパラメータ ϵ_{c} と σ_{c} についての推定を行った。推定を行わないパラ メータには AMBER99(ファイル名=leaprc.ff99) を 選択した。この力場は、GB を用いたシミュレー ションに適しているものである [5]。また、最適 化を行うパラメータには計算時間の節約の観点か ら、その初期値の既存のパラメータの近傍に乱数 を振って得られたものとした。これは既存の力場 パラメータから余り遠くないところに最適解が存 在している事を仮定しているが、過去の研究など がこのパラメータを用いてシミュレーションを行っ てきた事や、前章での予備実験より、大きく異なる パラメータを与えて MD シミュレーションを行っ ても、構造が壊れてしまったり、安定したトラジェ クトリーが得られなかったりする事が分かってい るためである。

3.5 推定を行うタンパク質

パラメータ推定を行うタンパク質は PDB から、 Fig. 3.1に示す3つのエントリーを選択した。これ



Fig. 3.1 パラメータ推定に使用するタンパク質

ら全ての立体構造は、X線回析によって決定され ている。6PTIは2つの α ヘリックスと β シートを 持つ膵臓トリプトシン阻害剤である。1PGBは4 本からなる β シートと1つの α ヘリックスを持つ 免疫グロブリン結合タンパク質である。

4 実験結果

4.1 シンプレックス法の評価

シンプレックス法の収束状況を確認するために、 各イタレーションごとの最良のシンプレックスの 評価値を Fig. 4.1に、各イタレーションごとの評 価値の分散を Fig. 4.2示す。



Fig. 4.1 最良のシンプレックスの評価値



Fig. 4.2 シンプレックスの評価値の分散

Fig. 4.2で 1PGB、6PTI のタンパク質についての 推定が途中で終了しているのは、シンプレックス の持つ評価値の分散が 0.001 以下となり収束した 事を示している。1AJJ は Fig. 4.1より推定が進ん でいる事が確認できるが、50 回のイタレーション 後もシンプレックスの分散の値が 0.001 以下とな ることはなかった。

4.2 推定されたパラメータ

推定結果として得られたパラメータの一覧をTable 4.1に示す。

PDB Id	$\sigma_C[\text{\AA}]$	ϵ_C [kcal/mol]	fitness	original fitness
original	1.908000	0.086000		
1AJJ	1.304959	0.055135	2.284790	3.680939
1PGB	1.904152	0.082200	2.293704	2.817471
6PTI	1.855800	0.085171	1.502860	1.553851

Table 4.1 推定されたパラメータと評価値

1PGB のパラメータが最も既存の値に近い値が 推定され、既存のパラメータよりも評価値が約 0.6[Å] ほど PDB の構造に近い状態で安定状態に 達したことが分かる。 *cc* は約 0.004 [Å] 下がってい て、これは既存のパラメータよりも原子間の距離 を小さくする方向に最適化が進んだ事が分かり、 また σ_C は約 0.003[kcal/mol] だけ既存のパラメー タよりも低い値の方が安定したトラジェクトリー を得る事が出来たという事を表す。

6PTIでは、 ϵ_C が約0.05[Å]の低い値が、 σ_C が 約0.005[kcal/mol] だけ低い値が推定された。しか しながら、6PTIの評価値は0.05[Å] しか下がらず、 既存のパラメータより僅かに PDB に近い構造で 安定状態に達したトラジェクトリーを得る事が出 来たといえる。

1AJJは、他の2つのタンパク質と比較すると既 存のパラメータの値から最も遠い値が推定され、 既存のパラメータとの差異は他のタンパク質に比 ~~1桁違うものであり、 ϵ_C が約0.5[Å] ほど低い値 が、 σ_C が約 0.03[kcal/mol] ほど低い値であった。

考察 5

3つのタンパク質で、全ての推定されたパラメー タの値が下がった (Fig. 5.1)。



では結晶構造に比べ、比較的自由に動き回る事が 出来る。そのため、パラメータ推定では分子同士 を互いに束縛し、結晶構造に近い構造を再現する ために、 σ_c のパラメータを小さくして、より原子 間を近づける方向に最適化が進んだ可能性がある。

また、1AJJ は既存のパラメータから特に大き くはずれたものが推定された。この理由として考 えられるものとしては、PDBから得た1AJJの構 造データは、X線回析の精度を示す解像度が他の タンパク質の構造データに比べて低い。特に解像 度のレンジ (Resolution Range) が他のタンパク質 が10[Å] 以内であるのに対して、1AJJ は15[Å] 以 内である。そのため、間違った構造に対してパラ メータの最適化を行った結果が得られているとい う可能性がある。

また、1AJJでは、その構造内にαヘリックスと β シートの2次構造を持っている割合が低い(Table パラメータ推定では評価基準を PDB から得た X 5.1)。α ヘリックスは約4残基で1段の螺旋を形成 線回析による結晶構造とした。結晶構造では分子し、その螺旋間は水素結合によって構造が安定化 が規則的に配置され、分子同士が互いに束縛し合 される。また、βシートは1本の真っ直ぐに伸び うため分子が自由に動く事が出来ないが、溶液中 たストランドが複数本が集まって形成し、それら

PDB ID	残基数	α ヘリックスとβシートを形成する残基数	比率 [%]
1AJJ	37	6	16.22
1PGB	56	46	82.14
6PTI	58	31	53.45

Table 5.1 タンパク質全体を形成する残基から見た2次構造を構成する残基の比率

ストランド間に同様に水素結合が作られ、ストラ ンパク質の立体構造をよく再現するように推定を ンド間の構造を安定化させる。それら2次構造以 外の部分の非繰返し構造は、溶液中において不規 則な動きやすい構造である [9]。そのため、溶液中 での不安定な構造を安定化するために、σc を下げ 原子間の距離を小さくする方向に最適化が進み、 それに伴ってエネルギー障壁を低くするために €C が下がった可能性がある。さらに、1AJJではその 2次構造以外の部分の割合が高いために、結晶構 造と溶液中での構造に大きな違いが現れ、それが より大きく 1AJJ のパラメータを押し下げたと考 える事が出来る。逆に 1PGB や 6PTI では、構造 中にそれら2次構造が占める割合が高い。そのた めに溶液中での構造が、1AJJに比べ安定していて パラメータの値が低い方向に最適化が進む事はな かったと言える。本研究では、X線回析によって 得られた結晶構造が溶液中の安定な構造に近い構 造であるという仮定をおいたが、このようにαへ リックスやβシート以外の構造が多くを占める構 造である場合は、必ずしも仮定がうまく機能する 訳ではないという事を示すものかもしれない。

まとめ 6

本研究では、AMBER 力場のように汎用的な関 数において、シミュレーションの対象とする分子 ごとに異なる力場パラメータを割り当てるべきで はないかという考え方に基づいて、3つのタンパ ク質の力場パラメータの推定を行った。パラメー タ推定では、LJ ポテンシャルの炭素原子に関する ϵ_{C} と σ_{C} の2つのパラメータを、PDB から得たタ

行った。また、その際の評価値の算出方法は、既 存の実験値を評価値計算の基準とするものではな く、タンパク質の構造を評価値の計算基準とする という新たな考え方であった。

本研究より得られた結果では、用いたタンパク 質のうち1AJJと1PGBで既存のパラメータよりも PDB から得た構造に、より近い構造を再現するパ ラメータの値を推定する事ができた。また、推定 されたパラメータの値がタンパク質ごとに異なっ ていた、という結果は「シミュレーション対象の 分子ごとに異なる力場パラメータ」という考え方 が間違っていない事を後押しするものであった。

しかしながら、結晶構造を参照構造として、溶 液中での MD シミュレーションによってパラメー タ推定を行うと、既存のパラメータの値よりも低 く値が推定される事が分かった。また、α ヘリッ クスやβシートの2次構造と、それら以外の部分 の構造の安定性の違いから、タンパク質中の α へ リックスやβシートの2次構造を形成する残基の 割合が低いタンパク質では、その構造が溶液中に おいて不安定になる。この事より、構造からパラ メータ推定を行って得られる結果は、構造を安定 化させる2次構造に大きく依存してしまう可能性 がある事が分かった。これは、推定されたパラメー タの値が、本質的に正しく推定されているか疑問 を生むものである。

本研究の最終的な目標は、経験的なパラメータ、 画一的なパラメータという問題に取り組む事であ る。そこで本研究では、この問題に対しての布石 となるものであった。しかしながら、自動的に個々 のタンパク質ごとに異なるパラメータの値を推定 する事は出来たが、上で示唆した可能性により、 その値は信頼性に欠くものであったため、今後は 本研究で適用した手法を改善していく必要がある と思われる。

今後の展望として、上述の適用手法の改善のた めに、本研究で示唆した可能性をより深く検討す る、つまり様々な2次構造を持つタンパク質を用 いた検討が必要であると考えられる。本研究にお いて用いたタンパク質は、SCOP(Structural Classification of Proteins) による2次構造に着目した構 造分類において、1AJJと6PTIがSmall Proteins、 1PGB は $\alpha + \beta$ のタンパク質である[10]。しかし、 SCOP による2次構造の分類では、これら以外に all α 、all β 、 α/β など、他に数種類ある。これらの 2 次構造の違いによって推定される力場パラメー タの特徴や、溶液中における構造のより深い安定 性の議論を行ってみたい。

謝辞

本研究を行うにあたり終止、多大なる御指導な らびに御教示を頂きました山村雅幸教授に深く感 謝の意を表します。また、本研究を進める上で様々 な相談をして頂いたり、貴重な御意見を頂戴致し ました山村研究室の皆様に深く感謝致します。最 後に学生生活を支えてくれた両親や友人に心から 感謝します。

参考文献

- [1] 岡崎進, 岡本裕幸. 生体系のコンピュータシ ミュレーション. 化学同人出版, 2002.
- [2] Scott J.Weiner, Peter A.Kollman, David A.Case, U.Chandra Singh, Caterina Chio, Ciuliano Alagona, Profeta Salvatore, Jr., and Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and pro-

teins. J.Am.Chem.Soc., Vol. 106, No. 3, p. 765, 1984.

- [3] Roland Faller, Heiko Schmitz, Oliver Biermann, and Florian Müller-Plathe. Automatic parameterization of force fields for liquids by simplex optimization. *J.Comp.Chem.*, Vol. 20, No. 10, p. 1009, 1999.
- [4] Junmei Wang and Peter A.Kollman. Automatic parameterizaton of force field by systematic search and genetic algorithm. *J.Comp.Chem.*, Vol. 22, No. 12, p. 1219, 2001.
- [5] David A.Case and so on. AMBER7 Users' Manual. University of California, 2002.
- [6] Vickie Tsui and David A.Case. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J.Am.Chem.Soc.*, Vol. 122, p. 2489, 2000.
- [7] Y Komeiji. Review: Molecular dynamics simulation of biological molecules (1) methods. *J.Chem.Software*, Vol. 6, p. 1, 2000.
- [8] 千原秀昭, 稲葉章(訳). 物理化学要論. 東京化学同人, 1994.
- [9] 田宮信雄,村松正実,八木達彦,吉田浩(訳).ヴォート生化学(上).東京化学同人,1992.
- [10] Alexey G.Murzin, Loredana Lo Conte, Antonina Andreeva, Dave Howorth, Narlett G.Ailey, Steven E.Brenner, Tim J.P.Hubbard, and Cyrus Chothia. Scop: a structural classificatioin of proteins database for the investigation of sequence and structures. *J.Mol.Biol.*, Vol. 247, p. 536, 1995.

付録.A シンプレックス法

シンプレックス法は、*d* 次元の最適化を行う際 に*d*+1個のシンプレックス(単体)を考える。そ れらシンプレックスに対して以下の3つの変形を 適宜に用いる事で、評価値が低い方向に向かって *d*+1次元の空間の探索を進める。

反射

最も高い評価値を持つ点 *x_i* が残りの d 個の 点が作る超平面に対称に移動する (Fig. 付 録.A.1(a))。

$$x'_{i} = \frac{2}{d} \sum_{j=1}^{d+1} x_{j} - \left(\frac{2}{d} + 1\right) x_{i} \qquad (\text{fd} \text{\poundsA.1)}$$

拡大

最も高い評価値を持つ点 x_i が残りの d 個の 点が作る超平面の法線に従って移動する (Fig. 付録.A.1(b))。

$$x'_{i} = \frac{1-\lambda}{d} \sum_{j=1}^{d+1} x_{j} - \left(\frac{1-\lambda}{d} + 1\right) x_{i} \quad (\text{fr} \mathfrak{K}.A.2)$$

縮小

最も低い評価値を持つ点 *x_j* に向かって残り の d 個の点が移動する (Fig. 付録.A.1(c))。

$$x_i' = \frac{1}{2}(x_i + x_j), \forall i \neq j \qquad (\textit{ft} \textbf{\mathbb{R}.A.3})$$



Fig. 付録.A.1 シンプレックスの変形. (a) 反射. (b) 拡大. (c) 縮小

具体的には、まず初期のシンプレックスの評価値 を計算し、以下の手順を繰り返して用いる。

評価値が最も高いシンプレックスを選択し、
 このシンプレックスに対し反射が適用される。

- (2) 反射が適用されたシンプレックスが今だ評価 値が最も高い場合は $\lambda = 2.0$ の拡大が適用し、 それでも評価値が最も高い場合は $\lambda = 0.5$ の 拡大を適用する。
- (3) 今だ選択されたシンプレックスが評価値が最 も高い場合は、縮小を適用する。

収束するまで、この手順は繰り返し行われる。

付録.B 評価基準の検討

パラメータ推定に用いた評価値の算出方法につ いての検討を行うための追加実験を行ったので、 その結果を示す。

本研究では、シミュレーション直後の構造の揺ら ぎを除いて、エネルギーが平衡状態に達している 時点での構造と PDB の構造間の RMSD 値を評価 値と考えていた。そこで、Fig. 付録.B.1にパラメー タ推定の際と同様に20[ps]のMDシミュレーショ ンを、推定されたパラメータと既存のパラメータを 用いた結果を示した。さらに、推定されたパラメー タがエネルギーの平衡状態で正しく評価されてい たかの判断するために、より長時間の200[ps]のシ ミュレーションを行った結果を Fig. 付録.B.2に示 した。各タンパク質の推定された最良のパラメー タの評価値は、全て最後の 10[ps](10~20[ps])の RMSD 値の平均値であった。そこで、これらの結果 から 10~20[ps] での RMSD 値の標準偏差と 10~ 200[ps] での RMSD 値の標準偏差を算出し、Table 付録.B.1に示した。Table 付録.B.1では2つの異な

PDB Id	SD(10~20[ps])	SD(10~200[ps])
1AJJ	0.401	0.369
1PGB	0.178	0.396
6PTI	0.200	0.336

 Table 付録.B.1
 RMSD 値の標準偏差 (単位は [Å])

 る期間での標本数が異なるために直接的に比較を



Fig. 付録.B.1 推定されたパラメータと既存のパラメータの 20[ps] のシミュレーション結果の比較



Fig. 付録.B.2 推定されたパラメータと既存のパラメータの 200[ps] のシミュレーション結果の比較

行うことは難しいが、1AJJ、1PGB、6PTIにおい て2つの異なる期間での標準偏差の値の違いが0.1 ~0.2程度である。これは、10~20[ps]において参 照したデータが10~200[ps]において参照できる データと同じようなばらつきを持っていることを 表す。つまり、10~20[ps]においての評価計算が 大きくは間違っていなかった事が分かる。

しかしながら、6PTIでは、40~100[ps] 付近で、 α ヘリックスやβシートなどの2次構造を取って いない非繰返し構造の部分の変形が行われ、10~ 40[ps] 付近の構造とは異なる構造でエネルギーの 平衡状態になっているのが分かる。よって6PTIに おいてはエネルギーの平衡状態で評価計算を行う ためには、100[ps] 以上で行った方がよい事がわ かった。そのため、パラメータ推定を行う場合に は、個々のタンパク質によって構造の安定性が異 なるために、タンパク質ごとに異なる評価値の算 出方法を検討する必要があった。