

大規模計算機によるタンパク質間の結合自由エネルギー計算

池田 潤一^{1,3}, 大野 一樹⁴, 秋岡 明香^{1,5}, 折田 正弥⁴,
村岡 洋一², 野口 保^{2,3}, 関嶋 政和^{2,3}

¹ 早稲田大学 基幹理工学研究科 情報理工学専攻

² 早稲田大学 理工学術院

³ 産業技術総合研究所 生命情報工学研究センター

⁴ アステラス製薬株式会社

⁵ 電気通信大学 情報システム研究科 情報ネットワークシステム学専攻

概要

結合自由エネルギー情報は、タンパク質間の相互作用を知る上で非常に重要である。本研究では、汎用的な PC クラスタを用いて、タンパク質間の結合自由エネルギー値を求めるシステムを紹介する。このシステムは、結合自由エネルギーを求める際に Shirts のアルゴリズム、Jarzynski の等式を用いて、大規模分子動力学シミュレーションを行い、計算結果が定量比較可能にした。

Binding Free Energy Calculation between Proteins by Large-Scale Computer

Junichi Ikeda^{1,3}, Kazuki Ohno⁴, Sayaka Akioka^{1,5}, Masaya Orita⁴,
Muraoka Youichi², Tamotsu Noguchi^{2,3}, Masakazu Sekijima^{2,3}

¹ Department of Computer Science and Engineering,

Graduate School of Fundamental Science and Engineering, Waseda University

² Faculty of Science and Engineering, Waseda University

³ Computational Biology Reserch Center, Natinal Institute
of Advanced Industrial Science and Technology

⁴ Drug Discovery Research, Astellas Pharma Inc

⁵ Graduate School of Information Systems, The University of Electro-Communications

Abstract

Information on binding free energy is important for understanding interactions between proteins. In this paper, we introduce an evaluation system for binding free energy of proteins, which applies Shirts' algorithm on PC clusters. We also use Jarzynski's equation and Shirts' algorithm.

1 はじめに

様々な疾病に関与するタンパク質が、遺伝子解析などの手法により明らかになってきている。また、たんぱく 3000 のようなプロジェクトや、タンパク質立体構造予測技術の進歩により、多くのタンパク質の構造情報が得られるようになってきている。そ

こで、これらの構造を考慮した創薬に関する研究が進められている。創薬とは、疾病の原因になっているタンパク質の疾病に関与する機能を抑える化合物を発見することである。

そこでバイオインフォマティクスを用いてゲノム情報を利用して疾病に関連するタンパク質の情報を分析し、創薬に活用しようという試みが為されてい

る。タンパク質とリガンドの結合部位が求められた後に、タンパク質とリガンド間の結合自由エネルギー値を求めたり、リード最適化と呼ばれるより高い薬としての効果を得るためのリガンド改変を行った際の結合自由エネルギー値の変化を求めることが必要とされている。

タンパク質とリガンド間の結合自由エネルギー値は、MM-PBSA 法 [1] などを用いて求められてきた。MM-PBSA 法では、タンパク質 A,B 間の結合自由エネルギー (ΔG_{bind}) は次のように表される。 $\Delta G_{bind} = G(\text{複合体}) - G(\text{タンパク質 A 単独}) - G(\text{タンパク質 B 単独})$ それぞれの分子の自由エネルギー ($G(\text{分子})$) は次のように計算される。 $G(\text{分子}) = E_{MM} + G_{solv} - TS(\text{分子})$ 。 E_{MM} は真空中での分子力場エネルギー、 G_{solv} は水和自由エネルギー、 $TS(\text{分子})$ は分子についてのエントロピーである。この方法では、複合体、タンパク質 A,B それぞれに対して長時間の分子力学シミュレーションが必要とされる。計算機リソースの問題から、連続した数十ナノ秒の分子力学シミュレーションを実行するのが一般的であり、このような時間では初期構造依存性を始めとしてサンプリングが十分に行われないう問題がある。また、エントロピー値を基準振動解析を用いて求めるのは容易ではない。

この問題を解決するために Shirts のアルゴリズムを採用し、汎用的な並列計算機においてより高速で正確に結合自由エネルギーを求めるシステムを開発する。

2 結合自由エネルギー計算方法

2.1 タンパク質・リガンドの結合自由エネルギー

結合自由エネルギーとは、タンパク質とリガンド分子が水中で結合した状態と、タンパク質とリガンド分子がそれぞれ別に水中に溶けているとき二つの異なった平衡状態の間の自由エネルギー差である。

リード化合物探索は、創薬にとって重要なプロセスであり、タンパク質とリガンド間の結合自由エネルギーを求めることは非常に重要である。

現在、タンパク質とリガンドとの結合自由エネルギーを求めるさまざまな計算方法 [2] がいくつか提案されているが、物理論理に基づき生化学的実験によって求まる値に近い値を求めることは難しいとされる [3]。

この計算を実現するアルゴリズムについては、大規模な分子動力学法 [4] を用いることによって網羅的にサンプリングすることが可能になり、より高い精度で計算することが可能である。

2.2 Jarzynski の等式

分子動力学シミュレーションから得られるポテンシャルエネルギー値から仕事量を得るために、Jarzynski の等式 [5] を利用する。Jarzynski の等式とは、非平衡下での総仕事量と平衡状態での自由エネルギー差の間に成り立つ。2つの状態の間の自由エネルギーの差とそれらの間の遷移に必要な仕事の有限時間でのアンサンブル平均との間の関係式である。Jarzynski の等式は、 $\exp(-\Delta G/k_B T) = \langle \exp(-W/k_B T) \rangle$ で表わされる。この等式が成り立つのは準静的過程の場合である。この等式により、任意の非平衡状態下で系になされた仕事から、平衡状態における自由エネルギーとして得ることができる。

2.3 Shirts のアルゴリズムを用いた計算手法

Shirts が考案したアルゴリズム [7] を用いる。タンパク質とリガンド間に働く相互作用の力であるクーロン力とファン・デル・ワールス力の二つの力について考える。それぞれの力の寄与を求めるために、 λ を λ^C と λ^{LJ} とに分ける。タンパク質とリガンドが結合していない状態 ($\lambda^C = 0, \lambda^{LJ} = 0$)、結合している状態 ($\lambda^C = 1, \lambda^{LJ} = 1$) とする。クーロン力の寄与を調べる状態 ($\lambda^C = 1 \rightarrow 0, \lambda^{LJ} = 1$)、ファン・デル・ワールス力の寄与を調べる状態 ($\lambda^C = 1, \lambda^{LJ} = 0 \rightarrow 1$) のシミュレーションを行う。この λ の変化の刻みの間隔を小さく調節することにより、高い精度で結合自由エネルギーを求めることができる。

この相互作用が働いている状態において、パラメータ λ を 0 から 1 で変化させる。この一つ一つの変化の前後の状態の差、つまりポテンシャルエネルギー差を求める。 $W = U(\lambda_j, x) - U(\lambda_i, x)$ で表わされる。この式により遷移に必要な仕事量が定まる。分子動力学法を用いる場合、Shirts のアルゴリズムは分子同士に働く二力であるクーロン力とファン・デル・ワールス力の二つがどの程度寄与しているかを求めることができる。これによって、タンパク質とリガンド間におけるクーロン力とファン・デル・ワールス力それぞれが寄与する自由エネルギー $\Delta G_{Complex}^C$ と $\Delta G_{Complex}^{LJ}$ が求まる。

3 実験

3.1 対象となる系

本研究における計算対象として、Protein Data Bank(PDB)の2JUPを用いる。

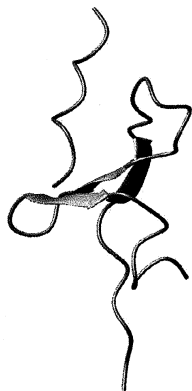


図 1: 対象となる系 (2JUP)

3.2 実験環境

本研究では、産業技術総合研究所の Magi クラスタ (PentiumIII 933MHz × 1040) を使用する。コンパイラとして Intel Fortran Compiler(バージョン 9.1.052) を用いた。分子動力学シミュレーションプログラムある Tinker[8] に Shirts のアルゴリズムを適用したプログラムを用いて行う。

シミュレーションで用いるポテンシャル関数には AMBER の ff99 を用いた。温度は 300K とした。

Jarzynski の等式、Shirts のアルゴリズムを用いる手法では、クーロン力とファンデルワールス力に関して λ を係数とする独立のシミュレーションとして扱うことが出来る。本研究においても、PC クラスタ上で網羅的に実行している。

今研究で λ の設定は、規則的な間隔で λ を設定した。

以上の条件で、Magi cluster を用いて

1. 50 ピコ秒の分子動力学シミュレーション
分子動力学シミュレーションを 33 通り実行した。
2. 1 ナノ秒の分子動力学シミュレーション
分子動力学シミュレーションを 33 通り実行した。
3. 初期速度をランダムに振り 50 ピコ秒シミュレーションを 20 パターン
50 ピコ秒 × 33 通り × 20 パターンのシミュレ

ーションを、並列計算機で分子動力学シミュレーション実行し、それぞれの λ におけるポテンシャルエネルギーの平均を求めた。

以上のように、3つの実験を行い、シミュレーションの時間による差異、それぞれの λ におけるポテンシャルエネルギーの平均値をとることによる差異を調べた。

それぞれの結果を比較し、シミュレーション結果の安定性や分子動力学の実行時間による結果の違いを考察する。

3.3 50 ピコ秒におけるシミュレーションの結果と考察

得られたポテンシャルエネルギーの値は図 2、図 3 のようにのようになった。

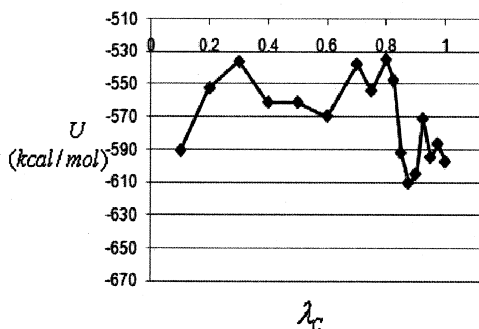


図 2: 50 ピコ秒における λ_C の変化によるポテンシャルエネルギー値

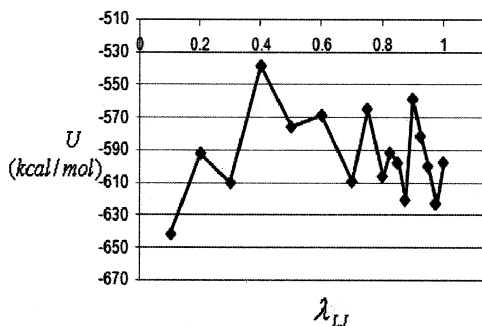


図 3: 50 ピコ秒における λ_{LJ} の変化によるポテンシャルエネルギー値
求められたポテンシャルエネルギーより Shirts のアルゴリズムを適応し計算する。

$$\Delta G = \Delta G_{Complex}^C + \Delta G_{Complex}^{LJ} = -82.244[kcal/mol]$$

3.4 1 ナノ秒において平均をとったシミュレーション結果と考察

得られたポテンシャルエネルギーの値は図4、図5のようにのようになった。

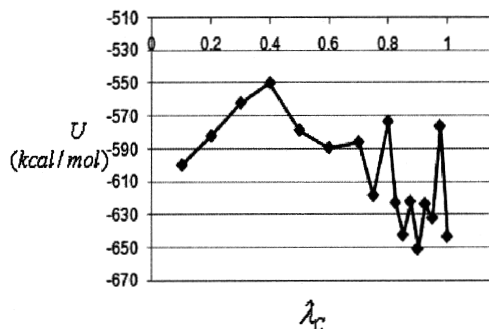


図4: 1 ナノにおける λ_C の変化によるポテンシャルエネルギー値の平均

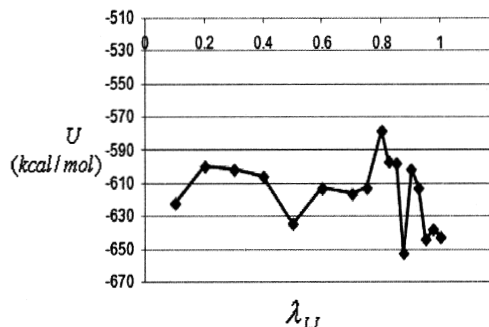


図5: 1 ナノにおける λ_{LJ} の変化によるポテンシャルエネルギー値の平均

求められたポテンシャルエネルギーより Shirts のアルゴリズムを適応し計算する。

$$\Delta G = \Delta G_{Complex}^C + \Delta G_{Complex}^{LJ} = -118.894[kcal/mol]$$

3.5 50 ピコ秒において平均をとったシミュレーション結果と考察

得られたポテンシャルエネルギーの値は図6、図7のようにのようになった。

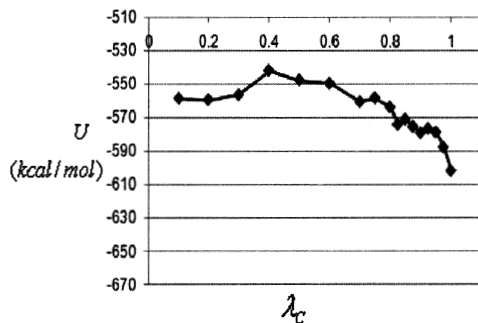


図6: 50 ピコにおける λ_C を変化によるポテンシャルエネルギー値の平均

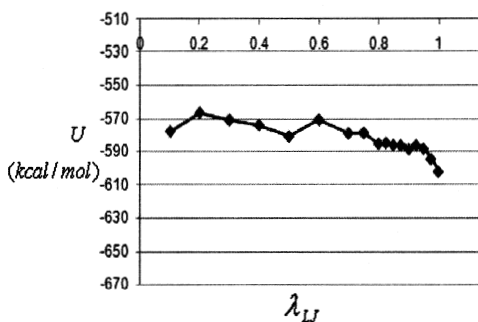


図7: 50 ピコにおける λ_{LJ} を変化によるポテンシャルエネルギー値の平均

求められたポテンシャルエネルギーより Shirts のアルゴリズムを適応し計算する。

$$\Delta G = \Delta G_{Complex}^C + \Delta G_{Complex}^{LJ} = -19.121[kcal/mol]$$

4 まとめと考察

結果から分かるようにポテンシャルエネルギー値にはばらつきが大きく見られる。しかし、50 ピコ秒のシミュレーションより1 ナノ秒のシミュレーションのほうがばらつきが少ない。 λ を大きくするに従い、ポテンシャルエネルギー値は小さくなる傾向が読み取れる。

λ_C を変化させたときのポテンシャルエネルギー値(図2、図3)、 λ_{LJ} を変化させたときのポテンシャルエネルギー値(図4、図5)はともに値に一定性がない。仕事量はポテンシャルエネルギー値の差から求めるため、ポテンシャルエネルギー値に一定性がないと仕事量の値が安定しない。自由エネルギーを求めるときは、Jarzynski の等式を利用するため、結合自由エネルギーはばらつきがあるポテンシャルエ

エネルギー値に大きく依存してしまう。結合自由エネルギーはそれぞれの λ_C 、 λ_{LJ} で求まるポテンシャルエネルギーの値の中で最も変化にとんだ値に依存する傾向がある。ゆえに結合自由エネルギーが定まらなかったと考えられる。

その原因として、 λ の間隔が広くポテンシャルエネルギー値の差が大きくなった点、水分子が入ったシミュレーションではなかった点、シミュレーション時間が十分ではなかった点が挙げられる。

次に、50 ピコ秒のシミュレーションを初期値を変えることで複数実行し、特定の λ におけるポテンシャルエネルギー平均値をとった場合には、 λ_C を変化させたときのポテンシャルエネルギー値 (図 6)、 λ_{LJ} を変化させたときのポテンシャルエネルギー値 (図 7) は、 λ を大きくすれば、ポテンシャルエネルギーはなだらかに小さくなることが示された。本研究では、ポテンシャルエネルギー値が安定したため、ばらつきのあるポテンシャルエネルギーに依存しない結合自由エネルギーを求めることができた。

以上よりシミュレーション時間を長くするより、初期速度を変え複数実行し平均をとるほうが、ポテンシャルエネルギーが安定して求められることがわかった。

5 最後に

本研究では、汎用的なクラスタを用いて特定のタンパク質とリガンドの結合自由エネルギーを高い精度で求めるために、Shirts のアルゴリズムを分子動力学シミュレーションプログラムに適用するシステムの開発を行った。シミュレーションの流れや算出値の正確さを確認するために、比較的残基数の少ないタンパク質を対象としたが、今後は残基数の多いものを含めて、表面電荷の与える影響なども解析を進めていく。また、アルゴリズムの改良、対象とするタンパク質やリガンドに対する λ 変数設定の自動最適化、などの汎用性を高めていきたいと考えている。

本研究では汎用的な PC クラスタを用いたが、将来的には分散コンピューティング環境を組み合わせたオープンな創薬環境を実現することで、PC クラスタにおける計算以上に大規模で効率的なシステムの開発を目指している。また、そのようなシステムは、一般の製薬企業では扱われにくい Neglected Disease に関するタンパク質の創薬支援に適していると考えている。

参考文献

- [1] H. Schafer, A.E. Mark, and W.F. van Gunsteren, "Absolute entropies from molecular dynamics simulation trajectories", *J. Chem. Phys.*, 113, pp.7809-7817, 2000.
- [2] H. Fujitani, Y. Tanida, M.Ito, G.Jayachandran, C.D.Snow, M.R.Shirts, E.J.Sorin, and V.S.Pande, "Direct calculation of the binding free energies of FKBP ligands", *J. Chem. Phys.*, 123, 2005.
- [3] S.P Brown, and S.W Muchmore, "High-Throughput Calculation of Protein-Ligand Binding Affinities: Modification and Adaptation of the MM-PBSA Protocol to Enterprise Grid Computing", *J. Chem. Phys.*, 46, pp.999-1005, 2006.
- [4] E. Osawa and Y. Kataoka, Molecular dynamics method and Monte Carlo method, Kodansha.
- [5] C.Jarzynski, Lett, "Nonequilibrium Equality for Free Energy Differences", *Phys. Rev. Lett.*, 78, pp.2690-2693, 1997.
- [6] Y. Tanida, M. Ito, and H. Fujitani, "Calculation of absolute free energy of binding for theophylline and its analogs to RNA aptamer using nonequilibrium work values", *J. Chem. Phys.*, 337, pp.135-143, 2008.
- [7] M.R. Shirts and V.S. Pande, "Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration", *J. Chem. Phys.*, 122, pp.144107-144116, 2005.
- [8] J.W Ponder and F.M Richards, "An Efficient Newton-Like Method for Molecular Mechanics Energy Minimization of Large Molecules", *J. Comput. Chem.*, 8, pp.1016-1026, 1987.