

高速計算機システムによる蛋白質フォールディングの研究

東京薬科大学生命科学部

○肥後 順一

Protein folding by a high performance computing system

Junichi Higo

School of Life Science, Tokyo University of Pharmacy and Life Science.

The project developed a computer system, MD-engine system, for powerful conformational sampling of proteins. We also intended to apply the computer system to molecular docking problem. The established MD-engine system showed a high performance in computing a large biological system, where all of the atoms were explicitly treated. Now, we can obtain the free-energy landscape of long peptides of about 30 residues in explicit water. During the project, we developed a novel force-field. A docking simulation between a protein and a small molecule was done in explicit water, where the two molecules were located apart to each other at the beginning of the simulation.

1. はじめに

蛋白質の立体構造は、そのアミノ酸配列が与えられたら一意的に決まる。さらに「構造活性相関」という専門用語が示すように、立体構造が直接的にしる間接的にしる生理学的な機能を決定する。従って、アミノ酸配列から立体構造を計算科学的に予測するという課題は、生命科学の分野で長く研究されてきた。その結果、立体構造既知の蛋白質からヒントを得ての構造予測は、ある程度の成功を収めている。しかし、蛋白質の折れ畳み (folding) のメカニズムを根本的に (つまり物理化学的に) 解明する取り組みは、まだ成功していない。その理由は、ポリペプチドのとりうる立体構造の多様性が、その鎖の長さに対して指数関数的に増大するという問題が一つ、もう一つは、原子間の相互作用を記述する力場の不正確さである。我々は、この課題の解決に向けて、まず高速の計算機システム (MD-engine system) の開発と効率的な立体構造探索法の開発を目指し、プロジェクトを出発させた。また、力場の不正確さの問題を解決するために、新たな力場の開発も視野に入れた。

ここで言う立体構造探索法とは、ポリペプチド鎖の多様な立体構造を計算機を用いて作成し、それら個々の立体構造の持つ熱力学的な重み (つまり、その立体構造がどれくらいの確率で出現するか) を算出する方法である。それによって、ポリペプチド鎖の自由エネルギー地形が描ける。自由エネルギー地形が得られると、鎖の折れ畳みの道筋が議論できる。代表研究者 (肥後) は、このプロジェクトが始まる以前から、マルチカノニカル分子動力学法という構造探索法の開発に携わってきた [1,2]。このプロジェクト開始時点では、水中で (つまり水を近似せずにそのまま取り扱うという立場で) 約 10 残基のペプチドの計算が可能になっていたが [1]、それが当時の世界記録であった。北村と山岸 (および株式会社富士ゼロックス) は、プロジェクトが始まる以前から共同で専用計算機の開発に取り組んできた。たくさんの原子から

なる生体高分子系の中に働く複雑な原子間相互作用を高速かつ正確に計算するのは、分子シミュレーションの技術において本質的に重要である。プロジェクト開始時点では多数の演算ボードを一つのシャーシに詰め込むシステムを作成し一定の成果を得ていたが、プロジェクトでは MD-engine ボードを PC に内蔵しそれを並列化することで計算効率を上げることを目標とした。

一個の蛋白質の折れ畳みの研究ができるということは、分子会合(ドッキング)も研究できることを意味する。なぜなら分子間相互作用の計算は、分子を構成している多数の原子の間の相互作用を計算することに等しく、それが正確かつ高速に計算できれば分子会合を効率的に研究できるからである。

2. 研究開発の成果概要

三年間の研究期間で、研究は大きく進んだ。高速計算機システムは、理論演算速度の 99 % 以上の実効性能が実現した。また、並列化効率も高く、72 CPU の並列化効率は原子数 6×10^4 の系の場合、1 CPU の計算速度との比で 71.89 倍(99.8%)、原子数 1×10^4 の系の場合、67.81 倍(94.1%)である(括弧内は MD-engine システムが受け持つ非結合相互作用計算のみの場合)。

ポリペプチドの立体構造探索では、マルチカノニカル分子動力学法を適用することで、現在約 30 残基のポリペプチド鎖(水分子の中に置かれている)が、1-2 ヶ月で計算できるようになった。水を厳密に考慮して構造探索することで得られる自由エネルギー地形は、水を何らかの簡便な方法で近似する方法よりも厳密な計算結果が得られるのはもちろんである。しかし、その一方計算時間が膨大になる。約 30 残基のポリペプチド鎖を 1-2 ヶ月で計算できるという成果は世界的にみてトップレベルにあると、我々は考えている。

その研究過程で、構造二面性を示すカメロン配列の自由エネルギー地形を求めた[3]。この配列は、周囲の環境に適応して α 構造や β 構造をとる。得られた自由エネルギー地形から、この配列は α 構造と β 構造の両方を安定構造として内包しており、それ以外の立体構造は自由エネルギー的に不利であることもわかった。このことから、配列の周囲のわずかの環境変化によって、 α 構造と β 構造のどちらが優勢かが変化するであろうと推測できる。

現在、我々は約 30 残基のポリペプチド鎖(より詳しいことは、講演で述べる)を研究している。研究の過程で、様々な力場を試してみた。その結果、既存の力場ではこのポリペプチド鎖の立体構造は再現できないことが示唆された。そこで、我々は新たに力場を作成し[4]、実験結果に近いものを再現できるようになった。力場の精度の問題は、このプロジェクトが開始される前から指摘されている[1,5]。現在使われている力場は α ヘリックス(または β シート)の形成を必要以上に促す傾向があり、アミノ酸配列に応じて立体構造を導き出せるバランスのとれたものがなかった。

分子ドッキングでは、ある蛋白質とそのリガンドである小分子を約 30 \AA 離して配置した状態から(二つの分子は十分な水分子で取り囲まれている)分子動力学シミュレーションを繰り返し実行し、会合の様子を観察した(論文準備中)。注目すべきは、分子間に満たされた水分子の配向が秩序化し、それが二分子の間に間接的な相互作用を誘起することを見いだしたことである。これは、以前代表研究者が理想的な系での計算機実験から予測していたことと一致する [6]。現実に存在する系でも、ドッキングと水分子配向の秩序化が関係していることを示せたのは重要であった。

3. まとめ

ここ数年、蛋白質の **folding simulation** の分野は競争が激化した。特にどれだけ長いポリペプチド鎖の計算を行ったかが重視されている。その競争の中でなおざりにされている感があるのは、力場の正確さである。例えばヘリックス含有率の高い構造を導きたい場合(つまり実験構造がヘリックスを豊富に含む場合)そうなる傾向の強い力場を使う、といったことがなされている。鎖が長くなればなるほど計算は困難になるが、その困難さをバイアスのかかった力場を使うことで切り抜けようとしているのかもしれない。長いポリペプチド鎖で力場の精度をきちんと評価し、それを高めていく努力が必要である。

競争が激化するこの分野で、蛋白質構造探索法としては(分子研岡本博士の研究グループが蛋白質へ初めて応用した)レプリカ交換アルゴリズムが優勢になりつつある。この方法は操作が簡単で、多くの研究グループが利用している。一方、マルチカノニカル分子動力学法は、系に対する統計力学的な洞察が必要で、習熟するのに時間がかかる。しかし、いったん習熟すると効率的で正確な構造探索が可能になる。どちらの方法に未来があるのかまだ分からないが、我々はこの方向で研究を進めていく予定である。

競争の激化を背景に、水の取り扱いで簡便な方法(水をあらわに計算に取り込まず、かわりにその効果を近似的に計算する方法)がよく使われるようになった。それにより膨大な計算手法を省こうというのがその目的である。しかし、当然ながらそれによって計算結果の精度が落ちる。我々は、計算時間はかかっても厳密な計算結果を求めてきた。前述した水分子配向の秩序化は、水分子を簡便な計算手法で置き換えることが困難であることを示している。我々は、その意味でも、水をあらわに取り扱う方向で研究を進めている。

JST BIRD のサポートのもと、我々の研究が大きく進展した。特に蛋白質の折れ畳みの研究では、30残基程度の系の自由エネルギー地形が、厳密に水分子を考慮しつつ 1~2 ヶ月で計算可能になった。上述したように、**folding simulation** の分野は競争が激化し、その中で計算量を減らすための工夫が提案されてきている。我々の方法は、計算量は大きくても精度を重視している。興味深いのは、最も愚直に研究を進めているはずの我々の研究が、世界の主要な研究グループと肩を並べているということである。今後、40~50 残基の系へと進む予定である。

4. 研究開発実施体制

代表研究者 肥後順一(東京薬科大学生命科学部)

研究開発題目

(1) フォールディングアルゴリズムの開発とシミュレーション実行

グループリーダー 肥後順一(東京薬科大学生命科学部)

(2) 蛋白質の変性過程の研究

グループリーダー 山岸明彦(東京薬科大学生命科学部)

(3) パラレル化促進とドッキングシミュレーション

グループリーダー 北村一泰(大正製薬研究システム部)

5. 参考文献

- [1] J. Higo, N. Ito, M. Kuroda, S. Ono, N. Nakajima, and H. Nakamura. “Energy landscape of a peptide consisting of α -helix, 3_{10} -helix, β -turn, β -hairpin, and other disordered conformations.” *Protein Science* **10**, 1160-1171 (2001).
- [2] J. Higo, O. V. Galzitskaya, S. Ono, and H. Nakamura. “Energy landscape of a β -hairpin peptide in explicit water studied by multicanonical molecular dynamics.” *Chem. Phys. Lett.* **337**, 169-175 (2001).
- [3] K. Ikeda and J. Higo, “Free-energy landscape of a chameleon sequence in explicit water and its inherent α/β bifacial property” *Protein Science*, **12**, 2542-2548 (2003).
- [4] N. Kamiya, Y. S. Watanabe, S. Ono, J. Higo, “AMBER-based hybrid force field for conformational sampling of polypeptides.” *Chem. Phys. Lett.* in press.
- [5] S. Ono, N. Nakajima, J. Higo, and H. Nakamura, "Peptide free energy profile is strongly dependent on the force field: Comparison of C96 and AMBER95." *J. Comp. Chem.*, **21**, 748-762 (2000).
- [6] J. Higo, M. Sasai, H. Shirai, H. Nakamura, and T. Kugimiya. “Large vortex-like structure of dipole field in computer models of liquid water and dipole-bridge between biomolecules” *Pro. Natl. Acad. Sci. USA*. **98**, 5961-6964 (2001).