

IBM ハイパフォーマンス・コンピューティング フォーラム 2000 報告 ~ その 1 ~

The Report About IBM HPC FORUM 2000 (part1)

角 美智子

Michiko SUMI

Abstract: This is the report about "IBM HPC FORUM 2000". Scientists at IBM are tackling the grand challenge of understanding the process of protein folding, and are building an incredibly fast computer named "Blue Gene". And especially I report about the Protein Folding Problem.

1 はじめに

昨月大阪にて行われた「IBM HPC FOURUM 2000」に参加した。ここでは「Blue Gene Progect」についての発表が行われた。現在、IBM の科学者達はタンパク質折りたたみのプロセスを理解する壮大なプロジェクトに取り組んでいる。本発表では、そのプロジェクトについて報告したい。

2 なぜ強力なスーパーコンピュータの開発に取り組むのか

スーパーコンピュータが時代遅れとされる現在にもかかわらず、なぜ IBM の技術者達は「Blue Gene」とよばれる、今日世界の最も速いコンピュータより 500 倍強力なスーパーコンピュータの開発に力を注ぐのだろうか。その理由として以下のようなことが挙げられる。

- 楽しいから
IBM の技術者達はそれらの開発が得意でもあり、またそれが楽しいから開発を進めている。
- コンピュータ設計に役立つ見識が得られるから
このようなプロジェクトを打ち出すことで、それに向けてチップや大規模なパラレルシステムの開発が進み、新しいコンピュータ技術を得ることができる。
- 重大な科学の問題があるから
コンピュータシミュレーションによって解ける重大な科学の問題がある。その問題を解くことは、科学や医学の発展につながり、人類に恩恵を与えることになる。解析に 10 年から 15 年かかる問題を今後もっと高速化していかなければならない。
- ムーアの法則は永遠ではないから
IBM の技術者達はムーアの法則は永遠に成り立つわけではないと考えている。

3 タンパク質のフォールディング

タンパク質は 20 種類のアミノ酸が鎖状に連結して作られる。Fig.1 のようにアミノ酸は、1 個の炭素原子にアミノ基 (-NH₂)、カルボキシル基 (-COOH)、水素原子 (-H)、さらに側鎖 (-R) が結合した構造を持つ。側鎖には、様々な分子構造があり、これがアミノ酸の性質を決めている。

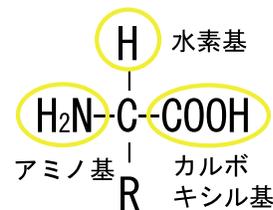


Fig. 1 アミノ基

そしてタンパク質は、アミノ酸が 2,3 個結合したのから 1000 個以上結合したのまで長さも様々である。このようアミノ酸の異なる組合せは、何万ものタンパク質を形成する。

アミノ酸配列が、熱力学的作用により折りたたまれると複雑な構造と機能を持つタンパク質となる。タンパク質は天然構造状態では、熱力学的に最も安定した状態に折りたたまれる、すなわち自由エネルギーを最小化するように折りたたまれる。このようなタンパク質の折りたたみの過程を「フォールディング」と呼ぶ。またタンパク質の折りたたみの経路はちょうど漏斗のような自由エネルギーの関数空間を考えるので「フォールディング・ファネル」と呼ばれている。タンパク質の初期状態の位置から自由エネルギーが最小の状態になるまでに通った構造の経路がフォールディングの経路となるが、その経路は無数にある。またこの関数空間は滑らかではなく、溶融状態のような中間状態に陥ることがある。

4 分子動力学

IBM では、分子動力学を用いたタンパク質のフォールディングシミュレーションを行っており、タンパク質を構成する各原子およびタンパク質をとりまく水分子の間の原子間相互作用エネルギー関数を式 (1) と定義し、1~10 フェムト (10⁻¹⁵) 秒単位でシミュレーションしている。

$$m_i \frac{d^2 x}{dt^2} = \sum_j F_{ij} \quad (1)$$

F_{ij} の i, j は、原子数を表している。また、 F_{ij} は、原子間のエネルギーを示しており、そのエネルギーは Table.1 に示したように、2 種類に分類することができる。short range は、近距離の原子間に働くストレッチング、曲げ、ねじれの力のエネルギー、long range は、離れた原子間に働くファンデルワールス力とクーロン力によるエネルギーである。

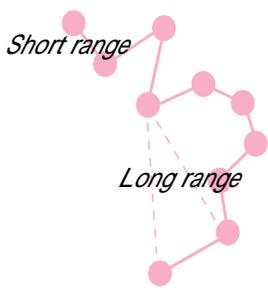


Fig. 2 原子間に働く力

Table 1 力の分類

short range	ストレッチング, 曲げ, ねじれ
long range	ファンデルワールス力, クーロン力

short range は、隣接する原子間のみのものであることから、 $O(N)$ で影響力があまりない。long range は、すべての原子が互いに作用するため、 $O(N^2)$ となり影響力が大きい。

5 解析時間と演算能力

Blue Gene の演算能力は Peteflops であるが、それがタンパク質のフォールディングにどのように使われているのだろうか。中規模のタンパク質を1年で解析するには、1秒あたり Petaflop の演算能力が必要となる。300個のアミノ酸が結合したタンパク質とそれをとりまく水分子をあわせると32000個の原子数となる。Fig.2に示したように、1ステップに必要な演算能力は 1.5×10^{11} である。全解析に必要なステップ数が 2×10^{11} であることから、全解析には 3×10^{22} FLOPs の演算能力が必

要となる。ここで解析時間を1年つまり 3×10^7 秒とすると1秒あたり1Petaflopの演算能力が必要となる。

Table 2 中規模タンパク質の解析時間と演算能力

Discription	Count
Atoms	~32,000
Force evaluations/time step	10^9
FLOPs/time step	1.5×10^{11}
Total time step	2×10^{11}
FLOPs/simulation	3×10^{22}
Execution time	3×10^7
Required FLOPs	$\sim 1 \times 10^{15}$

Fig.3は、タンパク質の規模とステップ数の関係を表した表である。

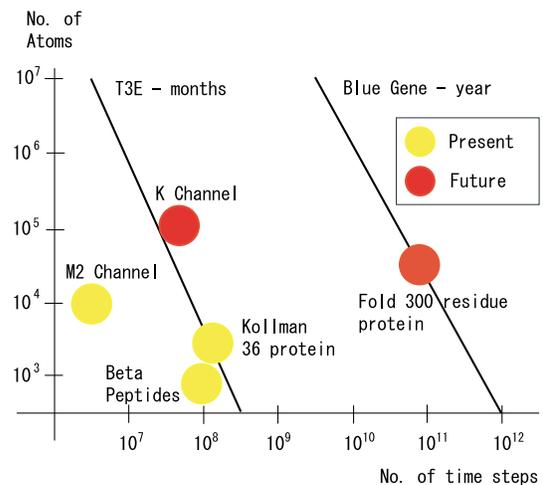


Fig. 3 Time Window into Biophysics

現在は、少ない原子数でステップ数が小さい問題にしか対応していないが、将来は300残基のたんぱく質をBlue Geneを用いて解くことを目標としている。

また、モデルやアルゴリズムをより現実に近いものにすることが目標とされている。

参考文献

- 『IBM HPC FORUM 2000 配布資料』
(日本IBM, 2000)
- 小長谷明彦『遺伝子とコンピュータ』
(共立出版, 2000)