

課題名 (タイトル) :

## フラグメントアセンブリ法と量子化学計算を用いた蛋白質立体構造予測

利用者氏名 : ○緒方 浩二\*, 中田 浩弥\*, シモンチニ デビッド\*\*

所属 : \*イノベーション推進センター 中村特別研究室

\*\*基幹研究所 Zhang 独立主幹研究ユニット

## 1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

蛋白質の立体構造は、蛋白質機能の解析や分子標的薬の探索など、様々な用途に用いられている。実験的に得られた蛋白質の構造は Protein Data Bank (PDB) に登録されており、internet を通じて容易に取得することが出来る。しかし、PDB に登録されている立体構造の数は生物種全体の蛋白質の数に比べて非常に小さなものである。従って、PDB に登録されていない構造が未知の蛋白質の構造を予測することは、非常に重要である。

従来、蛋白質立体構造予測の分野においては、フラグメントアセンブリ法などにより数百万の立体構造を生成し、その中から最適な構造を分子力場や統計学を応用した評価関数を用いて選択している。これらの方法は、タラメータの問題などから実際の X 線構造を再現するのは非常に困難である。そこで、量子化学計算を用いて精密に計算する必要があると考えられている。

我々は、蛋白質立体構造の予測に量子化学(QM)計算を応用する試みを行った。その最初のステップとして、Zhang 研究室で開発されたフラグメントアセンブリ法と Fragment Molecular Orbital (FMO)法による QM 計算を組合せる試みを行った。具体的には、構造既知の比較的小さなサイズの蛋白質の構造をフラグメントアセンブリ法により予測を行い、その中から適当に 100 個の構造を選び、それらの構造に対して QM 計算を用いてエネルギーの計算を行った。更に、得られたエネルギーと X 線構造とのずれを表す TM スコアの値の比較を行い、QM 計算のエネルギーの立体構造予測法への応用の可能性を探ることを行った。

## 2. 具体的な利用内容、計算方法

## 2.1 フラグメントアセンブリ法

フラグメントアセンブリ法は、フラグメントデータベースに登録されているフラグメントを繋ぎ合わせるにより構造未知の蛋白質の立体構造を予測する方法である。フラグメントデータベースは PDB に登録されている蛋白質立体構造を 10~20 残基毎の部分構造 (フラグメント) に分割抽出し、それらを集めたものである。非常に高速に数千万の構造を生成することが出来る。それらの構造は分子力場や統計的なパラメータを用いて評価を行い、最適な構造を見つけ出すことが出来る。

## 2.2 FMO によるエネルギー計算

FMO 法は分子を部分構造に分割して独立に計算し、そして各々の部分構造の計算結果を組み合わせることによって全エネルギー量を得る。その式は

$$E = \sum_I^N E_I + \sum_{I < J} (E_{IJ} - E_I - E_J + \Delta D_{IJ} V_{IJ})$$

ここで  $E_X$  ( $X=I, J, IJ$ ) は各部分構造毎の内部エネルギーを示しておりそれに加えて静電場の影響を補正することで合計のエネルギーを求めることが出来る。

本計算方法の優れた点は様々な物理的な影響、例えば溶媒効果、ファンデルワールス相互作用などを独立に追加していくことができることで、構造の安定性の予測を比較的少ない計算コストで行うことが出来る。

## 3. 結果

本研究では比較的小さなサイズで様々なフォールドを持つ 8 種類のタンパク質それぞれについて、100 個の異なる構造をフラグメントアセンブリ法で生成し、それらのエネルギーを計算し評価した。一つの例として 1scj の結果を示す。1scj は 2 つの  $\alpha$  ヘリッ

クスと 1 つの  $\beta$  シートを持っている :

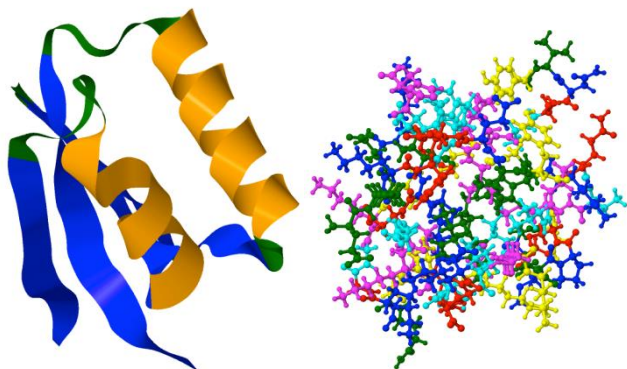


図 1 1scj の構造と計算で分割した部分構造

66 残基 (1054 原子) からなるタンパク質を 1 残基ごとに分割すること (図 1 : それぞれの色が別の分割をしめす) によって全エネルギーを計算することが出来た。その 100 種類の全エネルギーを図 2 に示す。

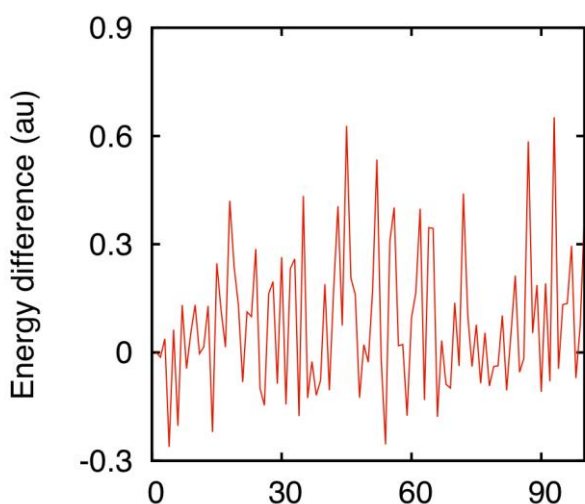


図 2 100 種類の構造に対するエネルギー

TM score とロゼッタスコアとの散布図を図 3 に示す。また、TM score と QM energy の散布図を図 4 に示す。ここで、TM score は X 線とモデルの rmsd に相関があり、TM score が小さいときはモデルの rmsd の値も小さいという関係がある。図 3 から TM score とロゼッタスコアの間には相関関係は観察されないことが分かる。

一方、TM score と QM エネルギーとの間では、ロゼッタスコアよりは強い相関が観察され、スコアが改善されたことが観察された。

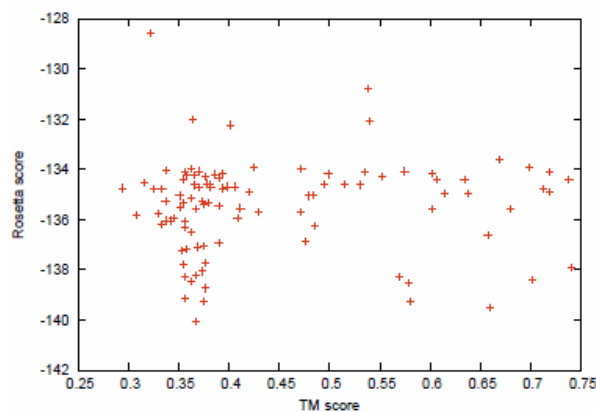


図 3 TM score とロゼッタスコアの散布図

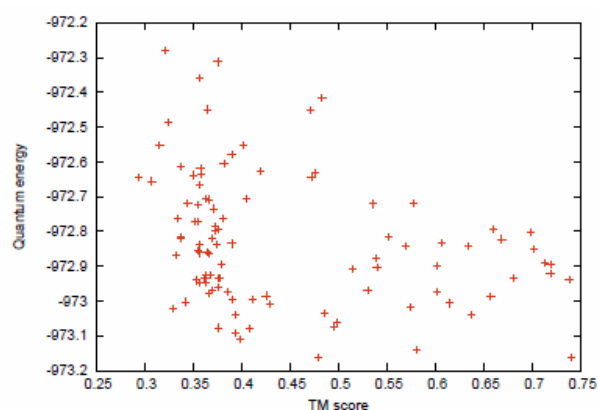


図 4 TM score と QM energy の散布図

しかし、その改善は当初予定したものよりも小さく、QM 計算をもう少し改善する必要があると考えられる。

#### 4. まとめ

今回の計算によって、8 種のタンパク質それぞれに対して系 1 0 0 種の全原子量子化学計算に成功し、X 線の構造からの差と全エネルギーの関係性を調べた。そこでは従来の統計的なスコアとエネルギーの評価と比較して量子化学計算を用いたときは微小ではあるが精度の向上が得られた。同時に精度の向上のためには各分割間の相互作用エネルギーを計算することと、分散力を考慮することが重要であることが本研究によって明らかになった。

#### 5. 今後の計画・展望

上記で示した方法の問題点として、統計的に得られたスコアと比べて、量子化学によって計算したエネルギーは当初期待していたほどの精度が得られな

かった。このれは、主に次の 3 つの原因が考えられる。(1)構造の歪に対して、量子化学計算がとても敏感に反応してエネルギーのばらつきが生じたため。

(2) 計算コストの関係で、溶媒の効果を無視したため、(3)エントロピー効果を取り入れていないため。

こうした効果はそれぞれのタンパク質毎に異なり、そのため系に依存して計算精度に非常に大きな差が発生していると考えられる。今後は誤差の原因と成り得る要因を一つ一つ調べることで、幅広いタンパク質の構造とエネルギーの関係を評価可能な解析方法を探索していくことを目指す。

尚、上記で示した方法と結果は現在専門誌に投稿するための準備を行っている。