

課題名 (タイトル) :

## 量子化学計算を用いた蛋白質の折り畳み予測

利用者氏名 : ○緒方 浩二\*、シモンチニ デビッド\*\*、青野 真士\*\*\*、中田 浩弥\*\*\*\*、若林 政光\*\*\*\*

所属 : \*社会知創成事業 イノベーション推進センター 特別研究室プログラム 中村特別研究室

\*\*基幹研究所 Zhang 独立主幹研究ユニット

\*\*\*基幹研究所 揺律機能研究チーム

\*\*\*\*東京工業大学大学院 生命理工学研究科 生体分子機能工学専攻

## 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。現在、約 82,000 個の蛋白質立体構造が Protein Data Bank (PDB) に登録されているが、重複を除いた蛋白質の数は約 25,000 種類となり約 3 分の 1 に減少する。この数は、生物種全体の蛋白質の数に比べると非常に小さく、PDB に登録されていない蛋白質が多く存在する。そのような場合、蛋白質立体構造の予測は非常に有効である。現在、様々な蛋白質予測法が考案されている。その一つに既知の構造データから有益な情報を抽出し、その情報を用いてモデルを構築方法がある。その代表としてホモロジーモデリング法が挙げられる。これらの方法の精度は、抽出された蛋白質構造情報の質に依存し、得られた情報が乏しい時には、実際の構造とは全く異なったものを得ることになる。それらの方法に対し、*ab initio* 法と呼ばれる既知の蛋白質の情報を用いることなく、物理化学的に定められた関数などを用いて立体構造を構築する方法がある。この方法は、PDB から有益な情報が得られない蛋白質に対し、その立体構造の予測を行うことが可能である。しかし、その精度は非常に低いのが現状である。この問題を解決する試みとして我々は、本研究においてはフラグメントアセンブリ法と量子化学計算(QM)によるスコア関数を組み合わせることにより、予測の精度を向上させることを行った。

また我々は、蛋白質の折り畳み問題は NP 問題であることから、それを解くための一つの方法として、充足可能性問題を解くためのアルゴリズムの解法の評価を行った。これは、単純な演算の繰り返しを逐次行うことを得意とする現在のコンピュータを利用して組み合わせ最適化問題や判定問題の解探索を行うことが可能

であるが、その際に、解に至るまでの時間が課題となる。クラス NP に属する問題はサイズ (変数の個数) の増加に伴い、解の候補の数が爆発的に増加し、全探索では実際的な時間で解に至ることが不可能となる。全探索の手続きとは異なる効率の良い解探索アルゴリズムの研究と開発が求められる。

本研究では、粘菌が光による外部フィードバックを受けながら巡回セールスマン問題の解に至る様子から着想を得て、クラス NP 完全に属する充足可能性問題 (Satisfiability Problem; SAT) を解くためのアルゴリズム (AmoebaSAT Algorithm) の構築を行った。

これらの問題を解決することは非常にインパクトが大きく有意義なことである。また、その方法の確立は、生物学、生化学、物理化学などの分野に於いて大きな進歩となり得ることが期待される。

## 2. 具体的な利用内容、計算方法

本研究においてはフラグメントアセンブリ法と量子化学(QM)計算によるエネルギーを組み合わせることにより、予測の精度を向上させることを目標とする。特に、フラグメントアセンブリ法から FMO 法による QM 計算まで、二つの方法がうまく組み合わせたり、予測したモデルから X 線構造に近い構造を抽出できるかどうかについて重点を置いた。これは、QM 計算によるエネルギーの値は蛋白質の構造の良し悪しに左右される。立体構造予測で得られた構造の多くは原子の結合長や結合角が実際の値と比べて異なっていたり、原子間の距離が短く衝突していたりと、予測された蛋白質の構造が歪んでいる場合が多く存在する。これらの歪がもたらす QM 計算でのエネルギーは非常に大きくなってしまう場合がある。従って、フラグメントアセンブリ法によって生成された構造が QM 計算の入力となり得るかどうかについての調査を行う必要がある。

ここで、フラグメントアセンブリ法とは、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。QM によるスコア関数は中村研究室において部分的に開発に参加した GAMESS に実装されている FMO 法を用いた (図 1)。

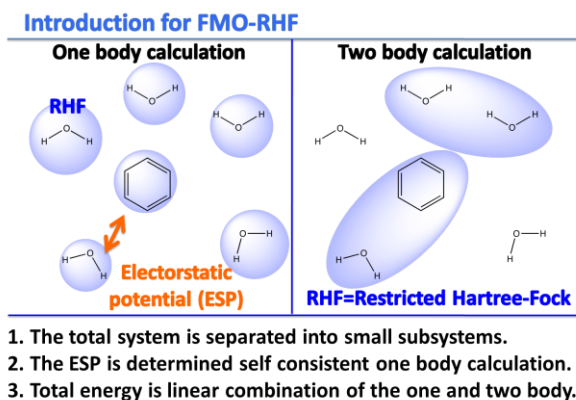


図 1 FMO 法の概略

FMO 法による量子化学計算は、従来の方法よりも計算時間が少なく、蛋白質などの大きな分子にも応用することが出来るのが利点である。これら二つの方法の組合せにより、今までの分子力場(MM)を基に考案されたスコアよりも天然の構造に近いものを選択できることが期待される。

以上の事を考慮して、比較的小さな蛋白質である Acylphosphatase-1 を標的とし QM 計算によって得られたエネルギーと Global Distance Test (GDT) スコア、更に、予測によって得られた構造のずれと QM 計算で得られたエネルギーの相関関係などについて解析を行った。Acylphosphatase-1 は 94 残基の蛋白質で X線構造が Protein Data Bank (PDB) に登録されている (PDB-id:2vh7)。その構造は $\alpha$ -ヘリックスと $\beta$ -シートから成り (図 2)、更に、蛋白質の中に埋もれている水がなく、テストを行う際に非常に都合がいいことから今回の標的蛋白質とした。

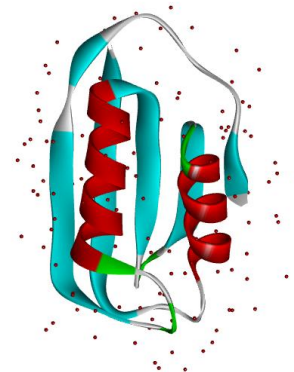


図 2 Acylphosphatase-1 の構造

今回行ったテストでは Acylphosphatase-1 のアミノ酸配列からその立体構造を予測し、GDT スコアと rmsd の相関や GDT スコアと QM エネルギー、QM エネルギーと rmsd の相関などを観察し、フラグメントアセンブリ法と QM エネルギーのマッチングに関して議論を行う。

また、NP 問題に関しては、SAT ライブラリ公開されている難易度の高い問題に焦点を当て、問題サイズ (変数の個数)  $N=20, 50, 75, 100, 125, 150$  の場合のそれぞれにつき 100 例題 (計 600 題) において解に至るまでの平均ステップ数と平均時間を統計によって見積もった。この際、比較の対象として、以前より提唱されていた効率の良いと言われている WalkSAT Algorithm との比較を行った。

### 3. 結果

最初に、GDT スコアと X 線構造との rmsd の関係を図 3 に示す。

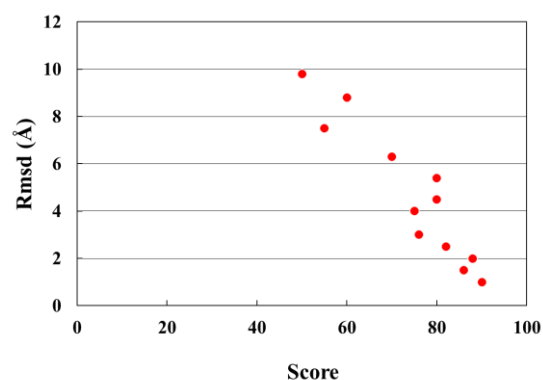


図 3 GDT スコアと rmsd の関係

グラフから、GDT スコアは rmsd の値に相関があることが分かる。これは、GDT スコアは X 線構造を参照して求めるので、rmsd と相関があることは明らかであ

る。

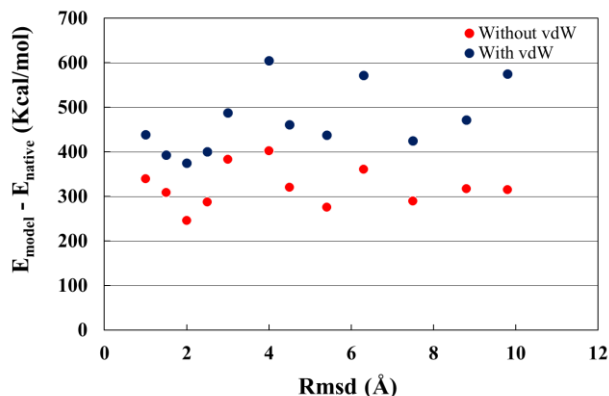


図 4 rmsd と QM エネルギーの関係

図 4 は rmsd と QM エネルギーの関係を示している。このグラフに於いて、native の構造の QM エネルギーが一番小さな値を示していることが分かる。また、rmsd の値が小さいものは native 構造からエネルギーの差も小さくなっていることが分かる。またグラフから、QM エネルギーと rmsd の間には相関あることがわかる。このグラフを詳細に観察すると、rmsd が 2 Å 以下のものに関しては rmsd の値と QM エネルギーの間には相関関係が崩れている部分がある。しかし、native 構造が最小のエネルギーを示していることから、これらの結果は、構造の歪が QM 計算に与えた影響であると考えられる。

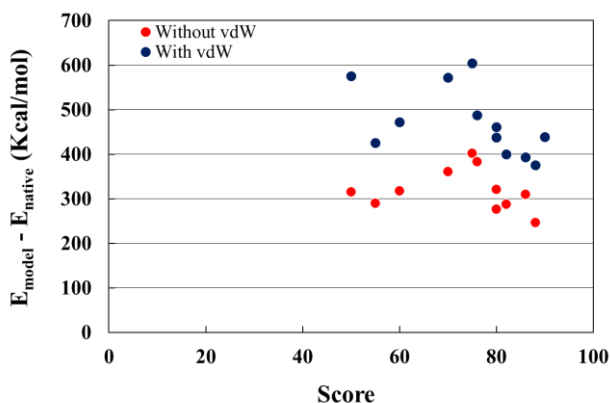


図 5 GDT スコアと QM エネルギーの関係

図 5 は QM エネルギーと GDT スコアの関係を示している。このグラフから、GDT スコアと QM エネルギーの間に相関があることが分かる。

NP 問題に関しては、AmoebaSAT の効率の良さが明確になった。N=20 では、平均ステップ数、平均時間ともに WalkSAT の方が優れているが、平均ステップで

は N=50 で、平均時間では N=75 で互角となり、その後、問題サイズが大きくなるほどに、AmoebaSAT の有効さが際立つ結果となった。

#### 4. まとめ

今回我々は、Zhang 研究室で開発されたフラグメントアセンブリ法による蛋白質立体構造予測法と中村特別研究室に於いて部分的に開発に携わっている FMO 法を組合せることを行った。フラグメントアセンブリ法により予測された構造は、QM エネルギーを少し大きくするような歪が存在すること。この歪をとることにより rmsd と QM エネルギーの間により明確な相関関係が見出されることと考えられる。しかし、今回テストに用いた 12 個のモデルに関して、FMO 法により蛋白質全体のエネルギーを求めることが出来た。このことは、フラグメントアセンブリ法から QM 計算までの一連の流れがスムーズに実行できたことを意味する。従って、今回の試みは成功したと言える。

#### 5. 今後の計画・展望

今回のテストに於いては比較的小さなサイズの蛋白質を標的としたが、今後、もう少し大きな蛋白質に応用することを考えている。また、様々なフォールドを持つ蛋白質に関して応用し、rmsd と QM エネルギーの相関関係や QM によるエネルギーとフラグメントアセンブリ法で用いている予測に用いているスコア（統計量）の相関などを調べ、その結果をスコアに反映することを考えている。

また、クラス NP に含まれる問題は SAT と相互に変換可能であるから、タンパク質フォールディング問題と結びつけることを考えている。