

L0023-14	より迅速な構造比較ツールの開発				
	氏名	矢島 龍	主査	黒田	副査

【背景・目的】

蛋白質構造比較ツールは立体構造分類のみならず、配列相同性が検出されない新規蛋白質の注釈付けといった面でも有用であり、また立体構造予測や機能予測への応用も期待されることより、その重要性は益々高まって来ている。しかしながら Protein Data Bank に登録されている立体構造データ件数が 5 万 5 千件を突破した現在では、各種既存構造比較ツールに於ける計算速度は十分に速いと言えなく、更に高速な計算が可能である立体構造比較ツールが求められている。これより本研究はより迅速な構造比較ツールの開発を目指す。

【アプローチ】

構造の比較に際して必要不可欠である構造記述子は現在数多く定義されている。しかしそれらに於ける構造比較性能を詳細に比較検討する研究は行われておらず、既存ツールのうちの多くが伝統的な構造記述子である “残基間ベクトル” を依然用いている。この際、より高速に算出可能で、且つより構造比較性能の高い構造記述子を用いたならば、より性能の高い構造比較ツールが構築可能であると考えられる。従って本研究では、まず構造記述子の比較検討からより高性能である構造記述子及び構造記述子の組み合わせを選出し、その構造記述子に適したアルゴリズムを適用する方法を採る。

【方法】

[データセット] 解像度 2.5 Å 以下で、CATH に於ける Topology、SCOP に於ける Fold レベルにて相同とされる蛋白質対をランダムに 100 個選出し、データセットとした。[構造記述子の性能比較] 各構造記述子それぞれ、また複数の組み合わせのそれぞれに於ける、1 層 DP アルゴリズムによるアライメントを算出し、その際のアライメント精度及び速度を性能として、比較検討を行った。[構築したツールの性能評価] 上記データセットに於ける GDT_TS, SAS, 計算時間を既存ツール RASH, FAST, STRUCTAL と比較することで評価した。

【構造記述子の性能比較結果】

多くの既存ツールにて使用されている “残基間距離・ベクトル” でない、“残基→重心ベクトル” が高い性能を持つ構造記述子であると示された。(Fig. 1) 構造記述子の組み合わせとしては、“残基→重心ベクトル” と “C-alpha density” の組み合わせが高い性能を示したが、単体に於ける結果を大きく超えるものではなかった。

【アルゴリズム】

上述の構造記述子に適したアルゴリズムとして、蛋白質内への複数の重心の構築、重心対のそれぞれについてのスコアマトリクスの作成、及び Dot-Matrix Filtering を介しての統合、を考案した。アライメントは、統合された 1 層のスコアマトリクスに於ける DP にて算出するものとした。この際、従来 DP では最大スコアからのみアライメントを算出するところを、極大スコアから複数のアライメントを算出し、GDT_TS 及び SAS を用いて選出するものとした。また従来 DP では不可能である両方の配列へのギャップ挿入を、閾値を設けることで可能とした。

【構築したツールに於ける結果・考察】

構築したツールに於ける各種パラメータを精製した結果、重心数を二つとした際に最も良い性能が得られた。これは、重心を増やすことによって生じる、相同でない重心対に於けるノイズを多く含むスコアマトリクスの影響であると考えられる。構築したツールのアライメント精度は各種既存ツールに対して 70~90%の性能ではあったが、しかし 5 倍以上もの速度を示した。(Fig. 2) これよりトータルとしては他の既存ツールと遜色のない性能であると言える。この際、“残基→重心” ベクトルという、既存ツールにて使用されていない構造記述子を用いていることから、改善の余地は多く残されている。例えば、構築する重心数を更に増やし、重心同士の相同性をアライメントに加味することで更なる精度向上が期待される。本研究は蛋白質構造比較分野に対して大きな一石を投じるものとなるだろう。

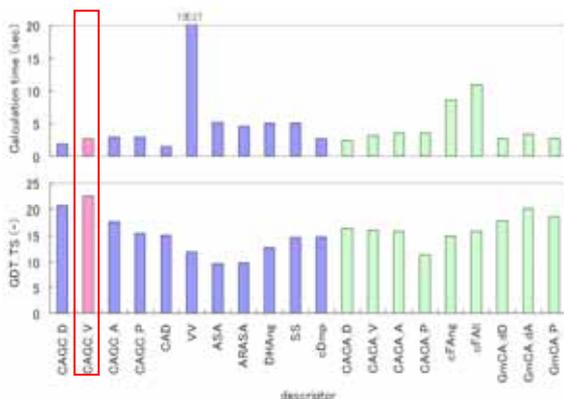


Fig. 1 構造記述子の性能比較

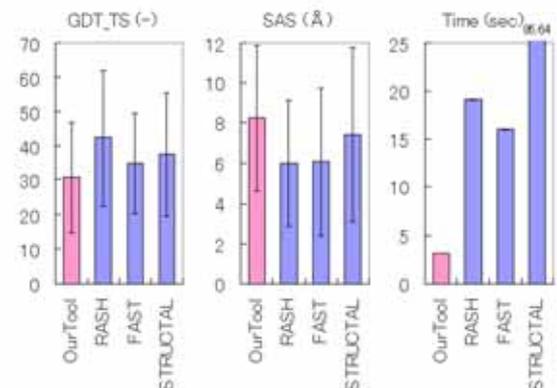


Fig. 2 構築したツールと既存ツールの性能比較