

## N 分木アルゴリズムを用いた標的タンパク質のペプチド結合部位探索

○中川 卓也、増田 尚之、石飛 秀斗、後藤 仁志

豊橋技術科学大学大学院（〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1）

**【緒言】** 抗体医薬やペプチド医薬の開発において、タンパク質上のペプチドが結合する（エピトープ様）部位を特定することが重要である。しかし、ペプチドは低分子医薬化合物に比べ官能基が多く、また配座も柔軟であるため、通常のドッキング法の適用が難しい。そこで我々は、任意のタンパク質表面とそこにあるペプチドが相互作用したときの状態を適切な精度で高速に評価でき、かつ、タンパク質表面全体を網羅的に評価できるシステムの開発を進めている[1]。

**【方法】** 本研究では、評価基準となる粗視化 4 体ポテンシャル[2]の改良と、これを用いたペプチド結合部位の網羅的探索法を開発している。粗視化 4 体ポテンシャルとは、アミノ酸残基を粗視化した 1 つの残基点とし、タンパク質全体を近接する残基点を頂点とする四面体の集合体とみなし、各四面体を構成する残基パターンに対して予め天然のタンパク質立体構造から統計的に求めた標準出現確率を 4 体ポテンシャルの評価スコアとする。そして、タンパク質全体を構成する四面体の評価スコアの合計値からそのタンパク質の立体構造を評価する。本研究では、新たな標準タンパク質立体構造データベース[3]を構築し（標準 DB）、粗視化 4 体ポテンシャルを標準 DB 内のタンパク質表面と結合性ペプチド間の 4 面体から再構成し、いくつかの新しい 4 体ポテンシャルによる評価スコアからペプチドの結合力を評価する。また、タンパク質表面上からペプチド結合部位を探索するため、タンパク質全体を覆う溶媒接触面解析アルゴリズムを応用し、網羅的に探索点を置く新しい手法を開発する。従来は逐次高いスコアを探索しながら、結合部位を求めるエリート戦略[4]を用いていたが、指定されたペプチド配列が最高評価スコアを示す結合部位を見逃さないようにするため、コンピュータ囲碁への応用で高く評価されているモンテカルロ木探索アルゴリズム[5]を導入し、並列分散処理技術を適用して高速化を図る。

実装するモンテカルロ木探索法は、ペプチド配列の残基を探索開始残基点（親節点）とし、そこから指定距離内にある探索点から乱数を用いて N 点選択し、ペプチド配列上でその残基と連続している次の残基点（子節点）として生成する。尚、エリート戦略では、高いスコアが得られる上位 N 点を選択することから、分木数  $N=1$  は分木探索ではなく、逐次最大スコアが得られるように残基点を決定すること同等である。次に子節点は親節点として同様のプロセスを繰り返すことで木構造を構築する。最初の探索開始残基点（ルート節点）から末端の子節点までを辿ることによって結合部位探索が終了する。分木数の増加に伴い、より多くの探索点を評価対象とするため、結果としてより網羅的な探索が可能になることが分かっている。

**【結果】** 標的タンパク質であるカテプシン E に対して、8 残基の活性阻害ペプチドの結合部位探索を行った。分木数 N におけるエリート戦略法の高速度化率を Fig. 1 に示す。これより、各ルート節点の探索をマルチスレッド処理することで、分木数が小さくとも高い高速度化率が得られることを確認した。分木数 N ( $=1\sim 8$ ) における探索時間を Fig. 2 に示す。シングルスレッドに対して 12 スレッドのマルチスレッド処理により、探索時間を短くすることができた。さらに、分木数 N ( $=1$

～8) の探索結果から得られた最高スコアと、上位 10 位までの平均スコアを Fig. 3 に示す。分木数 5 以上の最高スコアは同じ値になったのに対し、上位 10 位の平均スコアは分木数の増加に伴って高くなった。このことから、分木数を増やすことで高スコア結合部位をより多く見つけられることが分かった。次に、全 10 種類の活性阻害ペプチドに対して分木数 5 で探索を行った。各ペプチドの最大スコアを示す結合部位は異なるが (Fig. 4)、どれも比較的近い、同じ領域内にあることがわかった。発表では、新たに実装したモンテカルロ木探索の結果について述べる。

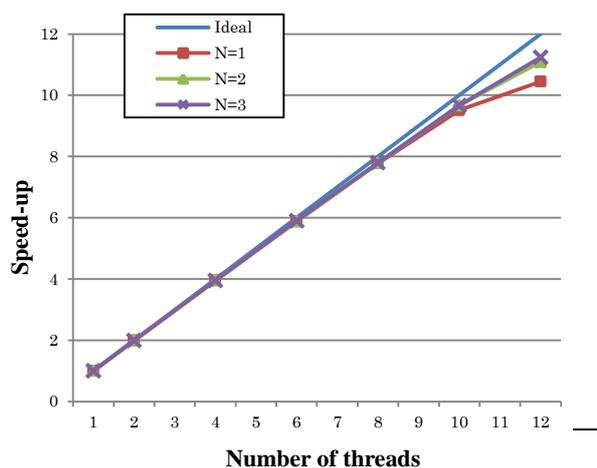


Fig. 1 Performance of multithreaded searches for peptide binding site

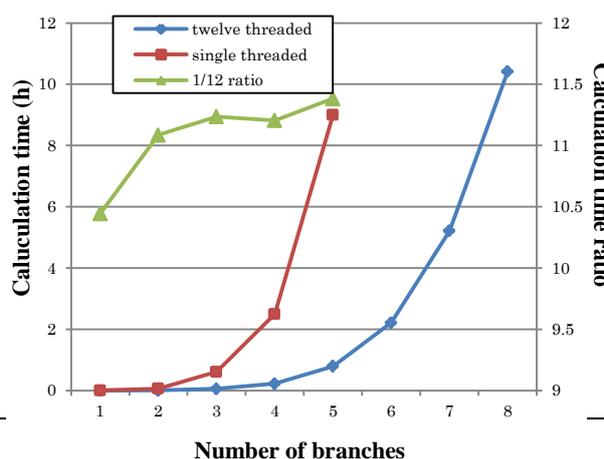


Fig. 2 Comparison of wall times in a single threaded and twelve-threaded searches

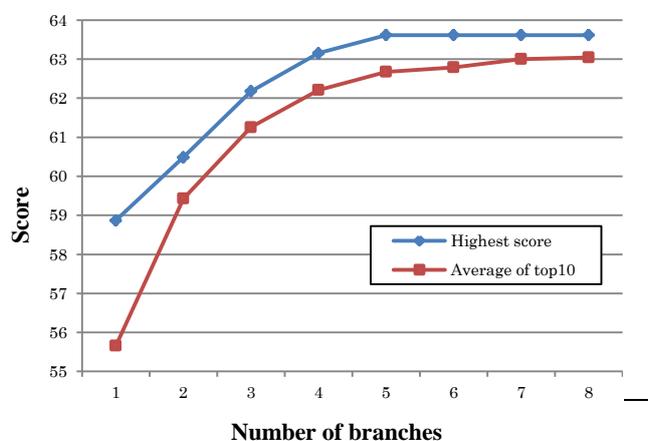


Fig. 3 Changes in highest and top 10 average scores depending on the number of branches.

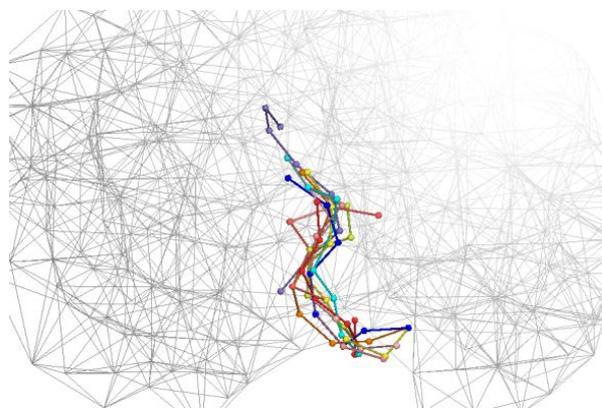


Fig. 4 Highest scored binding sites of ten peptides with different sequences

## 参考文献

- [1] T. Aita, K. Nishigaki, Y. Husimi, *Comput. Bio. Chem.*, 34, 53-62 (20)
- [2] B. Krishnamoorthy, A. Tropsha, *Bioinformatics*, 19, 1540-1548 (2003).
- [3] 石飛秀斗, 中川卓也, 孫林玉, 増田尚之, 後藤仁志, 相田拓洋, 西垣功一, 第 35 回情報化学討論会, P19(2012)
- [4] 増田尚之, 孫林玉, 相田拓洋, 西垣功一, 後藤仁志, *SCCJ2011 春季年会*, 1006 (2011).
- [5] 美添一樹, 石川裕, *IPSI SIG Technical Report*, 2010-HPC-126, 27(2010)