

タンパク質ドッキング計算結果の可視化とその解析

山本 航平[†] 大上 雅史^{††} 松崎 由理^{††} 石田 貴士^{††} 秋山 泰^{†††}

[†]東京工業大学 工学部情報工学科 ^{††}東京工業大学 大学院情報理工学研究科 計算工学専攻

1 序論

タンパク質は複数のアミノ酸から構成されており、生体内において様々な役割を担っている。タンパク質は単独で機能をもつことは少なく、その多くは他のタンパク質と相互作用することで機能を実現する。従ってタンパク質間の相互作用関係を理解することは、病因の解明や薬剤の設計において重要である。タンパク質同士が相互作用するかどうかを実験的に検出するには多くのコストがかかっていたが、近年の計算機性能の向上に伴って、計算機を用いて相互作用するかどうかの予測を行うことが可能になってきた。そこで我々は構造ドッキングに基づいて計算機上でタンパク質間相互作用予測を行うシステム“MEGADOCK”を開発した[1]。

MEGADOCKはタンパク質立体構造を利用して、複数のタンパク質間の相互作用関係を網羅的かつ高速に予測することを目的としたソフトウェアである。MEGADOCKは、まず形状の相補性などに基づいたドッキング計算を行い、2000個以上の相互作用構造候補(デコイ、図1)を出力する。次に出力されたデコイに対し、統計解析などの後処理を行い相互作用するかどうかの最終的な予測を行う。

ドッキング計算では、2つのタンパク質(レセプター、リガンド)について、レセプターを固定し、その周囲をリガンドが移動し形状相補性などに基づいたスコアが最大となる位置の探索を行う。これをリガンドの向きを変えて繰り返し行うことで、大量のデコイが出力される(図1)。

ドッキング計算の出力を用いて行われる最終的な予測については、構造クラスタリングを利用した手法[2]などが提案されているが、予測精度が充分とは言えない。従って新しい最終的な予測手法の開発が望まれているが、大量のデコイ群から意味のある情報を取り出

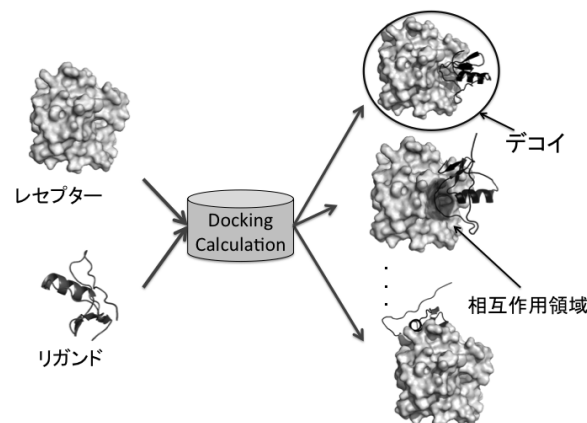


図1: ドッキング計算の概要

すことは困難なため、それが新たな予測手法を開発する障害となっている。

そこで本研究では、MEGADOCKのドッキング計算により大量に出力される計算結果を可視化する手法の開発を行った。大量の出力結果を1つの情報にまとめて可視化することにより、新たな知見を得ることができ、予測精度の向上に役立つことが期待される。

2 可視化手法の提案と考察

2.1 可視化手法

リガンドがレセプター周辺にどのように分布しているのか。またレセプター表面について考えたとき、どの領域がリガンドとの相互作用に使われ、またどのような頻度で使われるかを可視化することを目的とし、次の2手法を提案する。

- リガンドの重心の分布をレセプターの周囲に表示する。
- リガンドが相互作用しているレセプターの領域に色を付ける。

2.2 リガンド重心を利用した可視化

リガンドがどのように分布しているかを詳細に知ることが目的に、リガンドの重心をレセプター周辺にプロットし、可視化を行った。実際に表示させた結果を図2に示す。

この手法により、リガンドの位置分布を容易に理解できるようになった。しかしリガンドがレセプターのどの領域と相互作用しているかを理解する点で問題があった。

Visualization and Analysis of Protein Docking Calculation Results

[†] Kohei YAMAMOTO(k.yamamoto@bi.cs.titech.ac.jp)

^{††} Masahito OHUE(ohue@bi.cs.titech.ac.jp)

^{††} Yuri MATSUZAKI(y.matsuzaki@bi.cs.titech.ac.jp)

^{††} Takashi ISHIDA(t.ishida@bi.cs.titech.ac.jp)

^{†††} Yutaka AKIYAMA(akiyama@cs.titech.ac.jp)

Department of Computer Science, Tokyo Institute of Technology (†)

Graduate School of Information Science and Engineering, Tokyo Institute of Technology (††)

Ookayama 2-12-1-W8-76, Meguro-ku, Tokyo, 152-8550 Japan.

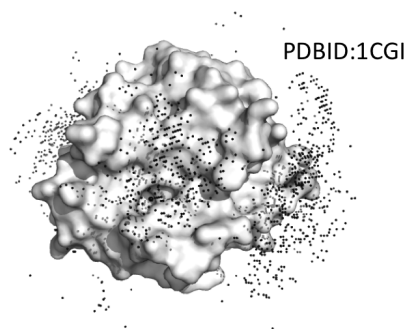


図 2: デコイ重心で可視化

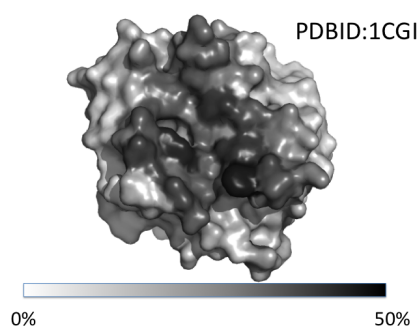


図 3: 相互作用領域を可視化

2.3 相互作用領域を可視化

レセプターの相互作用領域と、それが使用されている頻度をより詳細に可視化する手法を提案する。タンパク質を構成するアミノ酸残基を単位として相互作用領域に色付けを行うことで相互作用領域を表現し、さらに相互作用領域が使用されている頻度に応じた濃淡を付ける。ここで“相互作用領域”と“相互作用頻度”を以下に定義する。

リガンドのアミノ酸残基との距離が閾値以下のレセプターのアミノ酸残基群を**相互作用領域**と呼ぶ(図1)。残基間の距離は残基を構成する原子同士の間で計算するものとし、原子間距離最小の値を残基間の距離とする。

レセプターの各相互作用領域のアミノ酸残基について、何個のデコイがそのアミノ酸残基を介した相互作用をしているか計算する。計算で得た値の、デコイ全体に対する割合を**相互作用頻度**と呼ぶ。

相互作用頻度は大きくなるほど色が濃くなるよう表示させることとし、その値が50%を超えると最も濃い色が表示されるようにした。相互作用頻度に応じて相互作用領域に色付けを行ったレセプターを図3に示す。

この手法は2.2節の手法と比べ、相互作用領域を直感的に理解することが可能になり、相互作用頻度もある程度把握できるようになった。しかし複数の相互作用結果を定量的に評価する点で課題がある。

3 ドッキング計算結果への適用

前章に示した手法を、分子構造ビューワであるPyMOL[3]を用いて実際に可視化した。PyMOLはスクリプト言語であるPythonで記述されており、ユーザ自身がPythonによって機能拡張を行えるという特徴を持つ。本研究ではこのPythonによる拡張を利用した実装を行った。

可視化の対象として、protein-protein docking benchmark 2.0[4]の44個のレセプターと44個のリガンドによるペアに対して、 $44 \times 44 = 1936$ 通りの網羅的な相互作用予測をMEGADOCKによって行った結果を用いた。現在は2.3節に示した手法による1936通りの可視化を完了した。今後は1つのレセプターに対して44個のリガンドにおけるデコイ群がどのように分布しているかをより定量的に調べるため、2.3節に示した相互作用頻度による相関係数やtanimoto係数を用いた解析を予定している。これらの係数の値から定量的な相互作用の違いを見い出せることが期待され、予測精度の改善に繋がることが予想される。

4 結論

タンパク質相互作用予測ソフトウェア“MEGADOCK”の出力結果を可視化する手法を提案し、分子構造ビューワPyMOLを利用して描画するシステムを開発した。また本手法を、一般的なベンチマークデータセットに対するMEGADOCKの結果に適用した。今後の課題として、可視化手法をより定量的な評価のために改良することや、解析結果を活かして実際のタンパク質間相互作用予測手法に組み入れることが挙げられる。

参考文献

- [1] 大上雅史, 松崎由理, 松崎裕介, 佐藤智之, 秋山 泰, MEGADOCK: 立体構造情報からの網羅的タンパク質間相互作用予測とそのシステム生物学への応用, 情報処理学会論文誌, 3(3): 91-106, 2010.
- [2] Matsuzaki, Yuri, Matsuzaki, Yusuke, Sato, T. and Akiyama, Y., *In silico* screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis, *Journal of bioinformatics and computational biology*, 7(6), 991-1012, 2009.
- [3] DeLano, W.L., The PyMOL Molecular Graphics System, <http://www.pymol.org>.
- [4] Mintseris, J., Wiehe, K., Pierce, B., Anderson, R., Chen, R., Janin, J. and Weng, Z., Protein-Protein Docking Benchmark 2.0: an update, *Proteins*, 60(2), 214-216, 2005.