

タンパク質間ドッキングによる複合体構造予測

Protein Quaternary Structure Prediction with Protein-Protein Docking

大上 雅史

東京工業大学 大学院情報理工学研究科 計算工学専攻 秋山研究室 博士課程 3年

Masahito Ohue

Department of Computer Science, Graduate School of Information Science and Engineering,
Tokyo Institute of Technology, Japan.

Email: ohue@bi.cs.titech.ac.jp

URL: <http://www.bi.cs.titech.ac.jp/~ohue/>

1 背景

細胞内に存在するタンパク質の相互作用関係(protein-protein interactions)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とされ、実験・計算問わず広く研究されている。タンパク質間相互作用研究に欠かせないのが複合体構造に関する情報であるが、複合体構造を予測する計算技術として**タンパク質ドッキング**(protein-protein docking)という手法が盛んに研究されてきた。タンパク質ドッキングは、実験的に得られた単体の結晶構造からスタートし、それらが形成する複合体の構造を予測する手法である。タンパク質複合体を予測するための探索空間は非常に膨大であるため、複合体構造予測は通常、1)候補構造のサンプリング→2)候補構造の絞り込み→3)構造リファインメント、といった複数のステージから成ることが多い[1]。本ハンズオンセミナーでは、複合体構造予測の初期ステージとして、1)候補構造のサンプリングに使われる**剛体ドッキング**(rigid-body docking)について概説する。剛体ドッキングを行うツールはいくつか存在するが、我々の開発する MEGADOCK[2]は予測精度を可能な限り維持しつつ、種々の計算の高速化や MPI/OpenMP によるハイブリッド並列化によって、通常のノート PC から「京」や TSUBAME などのスーパーコンピュータに至るまでスケラブルな利用が可能である。

2 ハンズオンセミナー内容

本ハンズオンセミナーでは、MEGADOCK による剛体ドッキングを行い、1兆を超える可能な複合体構造空間から数百～数千の複合体候補構造(decoy)をサンプリングし、その結果を可視化することを実践する。また、得られた剛体ドッキングの結果に対して、ZRANK[3]を用いることで、その結果を改善することができることを実際に確認し、タンパク質ドッキング予測がどのように行われるかを知ることが目的とする。当セミナーに関する情報は次の URL に掲載したので、参考にしてほしい。 <http://www.bi.cs.titech.ac.jp/~ohue/bps2013>

3 MEGADOCK 概説

MEGADOCK の剛体ドッキングについて説明する。MEGADOCK の処理では、各タンパク質 A, B の原子配置や種別によって、その 3 次元複素離散関数 $A(\mathbf{v}), B(\mathbf{v})$ を定義し、これらの相関関数

$$C(\mathbf{t}) = \Re \left[\sum_{\mathbf{v} \in \mathbb{N}^3} A(\mathbf{v}) B(\mathbf{v} + \mathbf{t}) \right]$$

によって定義される評価関数 $C(\mathbf{t})$ の値を用いて最適な平行移動 \mathbf{t} の探索を行う。探索は、タンパク質 B の各回転パターンごとに繰り返す。なお、このタンパク質 A は便宜上 **レセプター** と、B は **リガンド** と呼ばれることが多い。 $C(\mathbf{t})$ は相関関数の形式をしているため、高速フーリエ変換 FFT とその逆変換 IFT を用いて計算することができ、計算量を $O(N^6)$ から $O(N^3 \log N)$ に削減することができる。具体的な離散関数の内容、その他の処理の内容や、並列計算の技法などについては、論文[2, 4]を参照されたい。

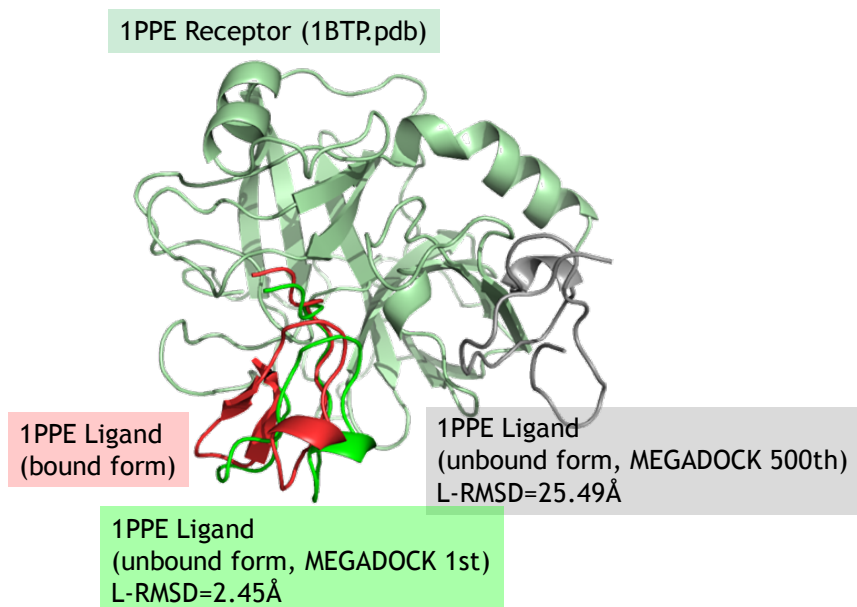
4 操作手順

- ・チュートリアル用のファイルをダウンロード・解凍・ディレクトリに移動

```
$ wget http://www.bi.cs.titech.ac.jp/~ohue/bps2013/handson2013.tgz
$ tar xzvf handson2013.tgz
$ cd tutorial
```

以降の操作手順は以下を参照のこと。

http://www.bi.cs.titech.ac.jp/~ohue/bps2013/handson_ohue.pdf



PDB ID: 1PPE 複合体のドッキング例。MEGADOCK による予測 1 位と 500 位の decoy をそれぞれ示した。 (`$ pymol tutorial/results/pymol/pymolsession_1PPE.pse` で描画可)

参考文献

- [1] Vakser IA. Low-resolution structural modeling of protein interactome, *Curr. Opin. Struct. Biol.*, **23**(2):198-205, 2013.
- [2] Ohue M, *et al.* MEGADOCK: An all-to-all protein-protein interaction prediction system using tertiary structure data, *Prot. Pept. Lett.*, 2013. (in press)
- [3] Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function, *Proteins*, **67**(4):1078-1086, 2007.
- [4] Ohue M, *et al.* Improvement of the Protein-Protein Docking Prediction by Introducing a Simple Hydrophobic Interaction Model: an Application to Interaction Pathway Analysis, *Lec. Notes Comput. Sci.*, **7632**:178-187, 2012.
- [5] Pierce BG, *et al.* Accelerating protein docking in ZDOCK using an advanced 3D convolution library, *PLoS ONE*, **6**(9):e24657, 2011.

----- MEMO -----