

簡易疎水性相互作用モデルによる タンパク質間ドッキング予測の高精度化

大上 雅史^{1,2} 石田 貴士¹ 秋山 泰¹

概要: タンパク質間ドッキング予測ソフトウェア MEGADOCK では、従来は形状相補性と静電相互作用の2つの効果を評価関数としていたが、本研究では新たに Atomic Contact Energy による疎水性相互作用モデルを提案し、MEGADOCK に追加した。MEGADOCK などの FFT を用いて計算されるグリッドベースのドッキング予測では、通常3つの効果を計算するために2回以上の相関関数計算を要するが、提案手法ではレセプターのみを考慮する新しい簡易型スコア関数によって、3つの効果を1回のFFT計算で同時に計算することが可能となり、高速性を損なわずに精度を向上させることに成功した。

キーワード: タンパク質間ドッキング, 疎水性相互作用, MEGADOCK

Improvement of the Protein-Protein Docking Prediction by Introducing a Simple Hydrophobic Interaction Model

OHUE MASAHITO^{1,2} ISHIDA TAKASHI¹ AKIYAMA YUTAKA¹

Abstract: In this study, we proposed a new hydrophobic interaction model which applied Atomic Contact Energy for our protein-protein docking software called MEGADOCK in which we previously used only two score terms, namely, shape complementarity and electrostatic interaction. Using the proposed score function, MEGADOCK can calculate three physico-chemical effects with only one correlation function. Therefore we succeeded improvement of accuracy without losing speed.

Keywords: Protein-Protein Docking, Hydrophobic Interaction, MEGADOCK

1. はじめに

タンパク質間ドッキング予測とは、単体のタンパク質構造を用いて、計算機によってタンパク質複合体の構造を予測する問題であり、生体分子間メカニズムの理解やタンパク質間相互作用の予測などに有用である。

タンパク質間ドッキング予測で最も有名な方法の1つに、3-D グリッド上でタンパク質の相互作用エネルギーを計算する Katchalski-Katzir アルゴリズム [1] がある。こ

れは複合体形成時の評価値（ドッキングスコア）をスコア関数同士の相関関数を用いて、高速フーリエ変換 (FFT) および逆変換 (IFFT) によって以下のように計算するものである。

$$\begin{aligned} S(\mathbf{t}) &= \sum_{\mathbf{v} \in \mathbb{N}^3} R(\mathbf{v})L(\mathbf{v} + \mathbf{t}) \\ &= \text{IFFT}[\text{FFT}[R(\mathbf{v})]^* \text{FFT}[L(\mathbf{v})]] \end{aligned}$$

ここで R, L は一方 (レセプター; Receptor) と他方 (リガンド; Ligand) のタンパク質のスコア関数であり、 $\mathbf{v}$ は 3-D グリッド空間 $\mathbb{N}^3$ 内のある座標を、 $\mathbf{t}$ はリガンドタンパク質の平行移動ベクトルを表す。

この方法を用いた実装は数多く存在し [2], [3], [4], 中でも ZDOCK [5], [6] は今日でも広く用いられているソフト

¹ 東京工業大学 大学院情報理工学研究科 計算工学専攻
Graduate School of Information Science and Engineering,
Tokyo Institute of Technology

² 日本学術振興会 特別研究員
Research Fellow of the Japan Society for the Promotion of
Science

ウェアである [7]。しかし、ZDOCK の用いる相互作用エネルギースコアは、精密であるが複数回の相関関数計算を行うため、例えば相互作用ネットワークを対象とするような多くの計算を行う際には現実的ではないという問題があった。

我々の開発している MEGADOCK [8] は ZDOCK などと同じく Katchalski-Katzir アルゴリズムを用いているが、独自の rPSC スコアと呼ばれる形状相補性スコア関数を用いることで、形状相補性と静電相互作用の効果を 1 つの相関関数で表現することに成功した。しかしタンパク質の相互作用時には疎水性相互作用が大きく働いており、さらなる精度の向上のためには疎水性相互作用の導入が不可欠であると考えられる。

本研究は MEGADOCK に疎水性相互作用を取り入れることを目的とする。特に、相互作用ネットワーク予測への応用を見据え、計算時間の大幅な増加を伴わずに性能が向上するスコア関数の設計を試みる。

2. 従来手法

ここでは従来手法として、我々がこれまでに開発した MEGADOCK 2.5 のドッキングスコアの概要について説明する。MEGADOCK 2.5 は、 $R(\mathbf{v})$ と $L(\mathbf{v})$ を形状相補性項 G と静電相互作用項 E の複素数として、

$$R(\mathbf{v}) = G_R(\mathbf{v}) + iE_R(\mathbf{v})$$

$$L(\mathbf{v}) = G_L(\mathbf{v}) + iw_eE_L(\mathbf{v})$$

と表され、スコア S は最終的に実部のみを取り出すことで計算される。 w_e は静電相互作用項の重みパラメータである。 G には rPSC スコア [8] が用いられ、それぞれ以下のような定義となっている。なお、規定されていない領域の値は 0 である。

$$G_R(\mathbf{v}) = \begin{cases} \text{半径 } r \text{ 内の原子数} & (\text{外部}) \\ -27 & (\text{コア, 表面}) \end{cases}$$

$$G_L(\mathbf{v}) = \begin{cases} 1 & (\text{表面}) \\ 2 & (\text{コア}) \end{cases}$$

ここで $r = 3.6\text{\AA} + \text{原子のファン・デル・ワールス半径}$ である。

3. 提案手法

提案手法ではアミノ酸の各原子について与えられる non-pairwise 型の Atomic Contact Energy (ACE) スコア [9] を利用して、レセプターの空間ボクセルに周辺原子の ACE スコアを加える形でレセプター表面の疎水性要素を導入する。

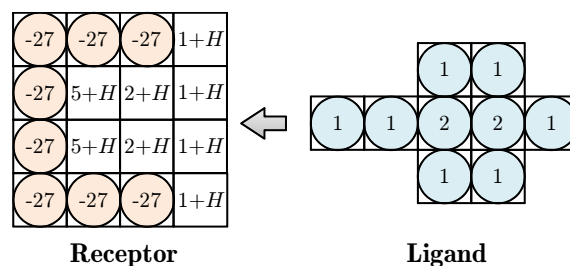


図 1 提案手法の離散関数 $G'_R(\mathbf{v})$ と $G_L(\mathbf{v})$ の模式図。ここでは簡単のため 2 次元で表している。

Fig. 1 Proposed scoring model $G'_R(\mathbf{v})$ and $G_L(\mathbf{v})$. The model consists of 3-D voxels, but here the model is shown by two dimensions for simplicity.

$$G'_R(\mathbf{v}) = G_R(\mathbf{v}) + w_h H_R(\mathbf{v})$$

$$H_R(\mathbf{v}) = \begin{cases} \text{半径 } r \text{ 内の ACE 和} & (\text{外部}) \\ 0 & (\text{コア, 表面}) \end{cases}$$

w_h は疎水性相互作用項の重みパラメータである。図 1 に、提案手法の模式図を示す。レセプターの表面近傍の空間とリガンド表面が重なると $G_R(\mathbf{v}) + w_h H_R(\mathbf{v})$ のスコアが得られるようになっている。リガンドコアの値が 2 となっているのはコア衝突時のペナルティを大きく (-54) する目的によるものである。位置によってはリガンドコアによって $2 \times \{G_R(\mathbf{v}) + w_h H_R(\mathbf{v})\}$ が得られてしまうことも想定されるが、高スコアを得られる凹みにリガンドコアが入るような状況下では、リガンド表面とレセプターとの衝突によってペナルティ (-27) が課されるため、大きな問題にはならないものと考えられる。

4. 実験と結果

Protein-protein docking benchmark 4.0 [10] の全 176 タンパク質ペアを用いて提案手法の性能評価を行った。このデータセットは、複合体の結晶構造を切り離して構成した bound データセットと、それぞれが単体の結晶構造である unbound データセットを含むが、本研究では両方を対象とした。ただし、 w_e と w_h のパラメータの決定には unbound データに対するオーバーフィッティングを避ける目的で bound データセットのみを用い、事前実験の結果 $w_e = 1.15, w_h = 0.6$ と決定した。比較対象には MEGADOCK 2.5, ZDOCK 2.3, ZDOCK 3.0 を使用した。なお、MEGADOCK 2.5 のパラメータは $w_e = 1, w_h = 0$ に相当する。

評価指標として、Average Hit Count と Success Rate を定義する。各手法はドッキングスコアに基づいて 3600 個の複合体候補構造を順位付きで出力する。Average Hit Count は、候補構造のうち準正解構造 (正解構造と Ligand-RMSD が $5\text{\AA}$ 以下のもの) が、上位 i 個の候補構造中に何個現れたかをカウントして平均したものである。また

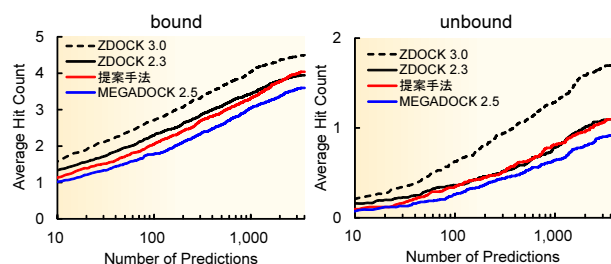


図 2 Benchmark 4.0 の全データに各手法を適用したときの Average Hit Count の結果.

Fig. 2 Average Hit Count for all test cases of Docking Benchmark 4.0.

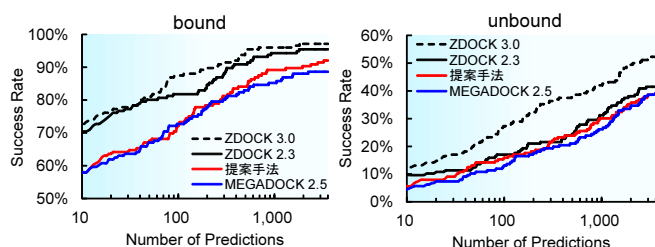


図 3 Benchmark 4.0 の全データに各手法を適用したときの Success Rate の結果.

Fig. 3 Success Rate for all test cases of Docking Benchmark 4.0.

表 1 平均計算時間 (分)

Table 1 The average of calculation time (min).

提案手法	MEGADOCK 2.5	ZDOCK 2.3	ZDOCK 3.0
14.2	14.1	53.6	124.6

Success Rate は、上位 i 個の候補構造中に準正解構造が含まれている場合を Success と定義し、全複合体における Success の割合を表すものである。

図 2 に Average Hit Count の結果を示す。図 2 より、bound・unbound とともに提案手法は MEGADOCK 2.5 に比べて性能が向上していることが分かる。また、ZDOCK 2.3 に近い精度が得られていることも見て取れる。一方で依然として ZDOCK 3.0 とでは予測性能に開きがあり、改善が必要であることも分かる。

図 3 に Success Rate の結果を示す。図 3 から MEGADOCK 2.5 に比べて提案手法の性能が向上していることが分かるが、この指標では bound データセットにおいて ZDOCK 2.3 と同程度の差が生じていることが確認された。しかし、unbound データセットにおいては提案手法と ZDOCK 2.3 との間の差はほぼ見られないということが分かる。

各手法の計算時間を表 1 に示す。計算時間の測定には東京工業大学学術国際情報センター TSUBAME2 Thin ノード (Intel Xeon processor 2.93 GHz (6 cores $\times$ 2), 32 GB RAM) の 1 core を用いた。提案手法は MEGADOCK 2.5 とほぼ変わらない計算時間を維持、ZDOCK 2.3 に比べて 3.8 倍、ZDOCK 3.0 の 8.8 倍の計算速度であることを確認した。

5. おわりに

本研究では、従来の形状相補性と静電相互作用によるタンパク質間ドッキングソフトウェア MEGADOCK に対し、レセプターのみを考慮した簡易型の疎水性相互作用スコアを導入した。その結果、MEGADOCK の計算速度を維持しつつ精度の向上に成功し、ZDOCK 2.3 に拮抗する精度が得られた。今後の課題として、リガンド側の原子の性質も考慮した疎水性相互作用モデルを検討することや、本研究の提案手法をタンパク質間相互作用予測に用いることが挙げられる。

謝辞

本研究は、科研費 (特別研究員奨励費 23-8750)、文部科学省 最先端・高性能汎用スーパーコンピュータの開発利用「次世代生命体統合シミュレーションソフトウェアの研究開発」、文部科学省 博士課程教育リーディングプログラム「情報生命博士教育院」の支援を受けて行われたものである。

参考文献

- [1] Katchalski-Katzir, E., Shariv, I., Eisenstein, M., *et al.* Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA*, 89, 2195-2199, 1992.
- [2] Gabb, H.A., Jackson, R.M., Sternberg, M.J.E. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*, 272(1), 106-120, 1997.
- [3] Mandell, J.G., Roberts, V.A, Pique, M.E., *et al.* Protein docking using continuum electrostatics and geometric fit. *Protein engineering*, 14(2), 105-113, 2001.
- [4] Cheng, T.M.-K., Blundell, T.L., Fernandez-Recio, J. py-Dock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins*, 68(2), 503-515, 2007
- [5] Chen, R., Li, L., Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, 52(1), 80-87, 2003.
- [6] Mintseris, J., Pierce, B., Wiehe, K., *et al.* Integrating statistical pair potentials into protein complex prediction. *Proteins*, 69(3), 511-520, 2007.
- [7] Fleishman, S.J., Whitehead, T.A., Strauch, E.-M., *et al.* Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *Journal of Molecular Biology*, 414(2), 289-302, 2011.
- [8] 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 数理モデル化と応用, 3(3), 91-106, 2010.
- [9] Zhang, C., Vasmatzis, G., Cornette, J.L., DeLisi, C. Determination of atomic desolvation energies from the structures of crystallized proteins. *Journal of Molecular Biology*, 267(3), 707-726, 1997.
- [10] Hwang, H., Vreven, T., Janin, J., Weng, Z. Protein-protein docking benchmark version 4.0. *Proteins*, 78(15), 3111-3114, 2010.