

# 相互作用エネルギーベクトルの距離を用いた タンパク質-リガンド間ドッキングのリランキング

安尾 信明<sup>†</sup> 関嶋 政和<sup>†,‡</sup>

東京工業大学 情報理工学研究所 計算工学専攻<sup>†</sup> 東京工業大学 学術国際情報センター<sup>‡</sup>

## 1. 序論

医薬品開発は我々の健康に関する重要な産業であるが、新規の医薬品を開発するためのコストは年々増大しており<sup>1</sup>、情報技術を用いた創薬のコストダウン手法の需要が高まっている。情報技術を用いた創薬における手法の一つにタンパク質-リガンド間ドッキングが存在する。これは、標的タンパク質に薬剤候補となる化合物を仮想的にドッキングし、最も安定な結合状態を探索するシミュレーション手法である。様々なドッキングのソフトウェアがこれまでに開発されているが、それらによる探索の精度は未だ十分ではない<sup>2</sup>。

この問題を解決するため、ドッキングの結果に既知の化合物の評価結果を取り入れリランキングを行い、精度を向上しようとする試みが存在している。例えば SIFt<sup>3</sup> や Pharm-IF<sup>4</sup> は、タンパク質-化合物間の相互作用の有無をその種類ごとにビット化した Interaction Fingerprint を作成し、その類似度で化合物の類似度を表現する手法である。しかし、これらの手法は相互作用の有無のみを用いて判断するため、相互作用の強さの情報が失われているという問題がある。同一種の相互作用であってもその強さはまったく同じではなく、また比較的強い相互作用である水素結合と比較的弱い相互作用である疎水性相互作用の有無が同じ重みで扱われることについても議論の余地がある。

本研究では、ドッキング中に得られる相互作用エネルギーを各残基ごとに並べた相互作用エネルギーベクトルを用い、それらの間の距離を用いたスコアである SIEVE-Score を提案する。また、ベンチマークセットである DUD-E を用いてリランキングの評価を行い、その精度について考察する。

## 2. 手法

### 2.1 手法の概要

本研究におけるリランキングでは、まず活性が既知の化合物と未知の化合物をすべてドッキングし、相互作用エネルギーベクトルを計算する。最終的なスコアは、化合物の相互作用エネルギーベクトル同士の距離から求まる SIEVE-Score (Similarity of Interaction Energy VEctor-Score) によって決定される。

### 2.2 SIEVE-Score

化合物  $l$  の SIEVE-Score  $S(l)$  は、

$$S(l) = \sum_{i \in \text{AUD}} w_i \min\left(\frac{1}{\text{dist}(l, i)}, 1\right)$$

$$w_i = \begin{cases} \frac{|D|}{|A|} & \text{if } i \in A \\ -1 & \text{if } i \in D \end{cases}$$

で定義される。ここで、 $A$  は活性のある化合物の集合、 $D$  は活性のない化合物の集合、 $w_i$  は化合物  $i$  の重み、 $\text{dist}$  は二つの化合物の相互作用エネルギーベクトル間のユークリッド距離である。距離が非常に近い場合に距離の逆数が発散してしまうことを防ぐため、距離が 1 未満の場合 1 としている。

化合物  $l$  が活性の有無が既知である全ての化合物と等しい距離にある場合は  $S(l) = 0$  となり、既知の活性化合物との距離が近く、不活性化合物との距離が遠いほど高いスコアとなる。

## 3. 実験

### 3.1 実験条件

比較実験では、ドッキングに Glide version 65013<sup>5</sup>、データセットに DUD-E<sup>6</sup> を用いた。DUD-E には 102 タンパク質が登録されており、各タンパク質について共結晶構造、活性のあるアクティブ化合物、活性がないと考えられるデコイ化合物が登録されている。今回の実験においては、Leave-one-out クロスバリデーションによりリランキング結果の ROC 曲線を求め、ROC 曲線下の面積である AUC を用いて比較を行った。比較対象はリランキング前の Glide SP モード及び既存手法の SIFt である。SIFt の実装は Schrödinger suite 内に含まれる interaction fingerprints version

Reranking of protein-ligand docking using the distance of interaction energy vector

Nobuaki Yasuo<sup>†</sup>, Masakazu Sekijima<sup>†, ‡</sup>

<sup>†</sup>Department of computer science,

Tokyo institute of technology

<sup>‡</sup>Global scientific information and computing center,

Tokyo institute of technology

47013 を用いた。SIFt によるリランキングは、活性のある化合物の中で最も Tanimoto 係数による類似度が高い化合物との類似度をスコアとした。

### 3.2 結果

DUD-E 中で、Glide SP モードと比較した際に最も AUC が向上及び低下したタンパク質の ROC 曲線を図 2, 各手法を比較した際に AUC が向上した標的タンパク質数を表 1 に示す。

表 1. AUC が向上した標的タンパク質数

	SIEVE	SP	SIFt
SIEVE-Score	---	86	99
Glide SP	16	---	77
SIFt	3	25	---

各行が手法, 各列が比較対象を表す。

SIEVE-Score, Glide SP モード, SIFt の AUC の平均値はそれぞれ 0.861, 0.776, 0.745, 標準偏差はそれぞれ 0.111, 0.112, 0.106 であった。SIFt は Glide SP モードよりも精度が低下する標的タンパク質が多い一方, SIEVE-Score は Glide SP モード, SIFt のいずれと比較しても高い精度の向上が見られた。

### 4. 考察

SIEVE-Score が Glide SP モードに比べて精度が向上した理由は SIFt に比べて精度が向上した理由であるが, 一つは相互作用エネルギーベクトルに相互作用の強さの情報が考慮されている点, もう一つは最も類似度が高い化合物一つではなく全体の傾向が考慮されている点であり, これらの改善点によって精度の向上が達成されてい

ると考えられる。

本研究では, タンパク質-リガンド間ドッキングの精度を向上させるためのリランキング手法として SIEVE-Score を新たに提案した。SIEVE-Score は, DUD-E データセットに含まれる 102 タンパク質のうち Glide SP モードに対しては 86 タンパク質で, SIFt に対しては 99 タンパク質で精度が向上した。しかしながら, 未だ 16 タンパク質では精度の低下がみられるため, 相互作用エネルギーそのものではなく, 各相互作用に重みをつけるなど, より高精度のスコアを開発することが今後の課題となる。

### 参考文献

- [1] Mullard, A.: New drugs cost US \$2.6 billion to develop, *Nature Reviews Drug Discovery*, **13**:12, 877 (2014).
- [2] Lionta, E., Spyrou, G., Vassilatis, D. K. and Cournia, Z.: Structure-based virtual screening for drug discovery: Principles, applications and recent advances, *Current topics in medicinal chemistry*, **14**:16, 1923 (2014).
- [3] Deng, Z., Chuaqui, C. and Singh, J.: Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions, *Journal of medicinal chemistry*, **47**:2, 337-344 (2004).
- [4] Sato, T., Honma, T. and Yokoyama, S.: Combining machine learning and pharmacophore-based interaction fingerprint for *in silico* screening, *Journal of chemical information and modeling*, **50**:1, 170-185 (2009).
- [5] Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., *et al.*: Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *Journal of medicinal chemistry*, **47**:7, 1739-1749 (2004).
- [6] Mysinger, M. M., Carchia, M., Irwin, J. J. and Shoichet, B. K.: Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking, *Journal of medicinal chemistry*, **55**:14, 6582- 6594 (2012).

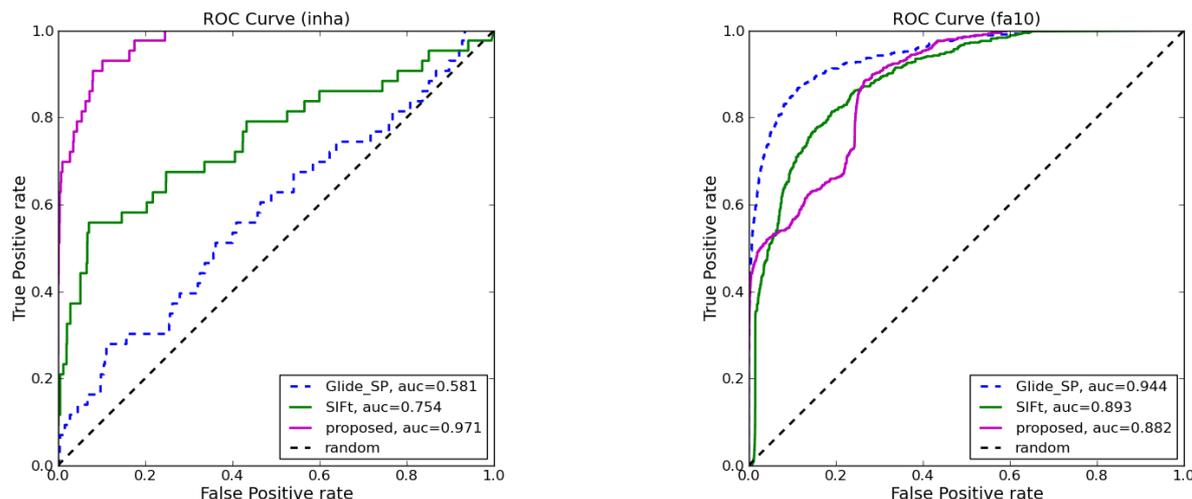


図 2: DUD-E における ROC 曲線の比較

紫実線: SIEVE-Score, 緑実線: SIFt, 青点線: Glide SP モード, 黒点線: ランダム予測。  
標的タンパク質: 左: Enoyl-[acyl-carrier-protein] reductase (inha), 右: Coagulation factor X (fa10)。