

標的タンパク質の立体構造を用いたリガンド候補化合物の上限サイズの推定による化合物フィルタリング

柳澤 溪甫^{1,2,a)} 大上 雅史^{1,3} 石田 貴士^{1,2,3} 秋山 泰^{1,2,3}

概要：化合物群から薬剤候補化合物を選別するための計算手法としてタンパク質-化合物ドッキングがよく用いられるがこの手法は計算量が大きいため、一般に膨大な化合物ライブラリから対象となる化合物群をあらかじめ削減する前処理が必要である。化合物群の削減には Lipinski の法則などの物理化学的な特微量の閾値の適用や、機械学習などの手法が用いられてきたが、これらはドッキング計算と組み合わせることが考慮された手法ではなかった。そこで本研究では、ドッキング計算と組み合わせることを前提としたフィルタリングについて、(1) タンパク質のポケットサイズの推定手法を利用した化合物体積の上限の推定、(2) 少量のサンプル化合物を試験的にドッキング計算することによる化合物体積の上限の推定、の 2 つの手法を提案した。2 つの手法のうち、後者のサンプル化合物のドッキング計算を用いて閾値を決定する手法ではドッキング可能な化合物の取りこぼしを 0.5%未満に抑えつつ約 15%の化合物を評価対象から削減することができた。

キーワード：創薬支援、バーチャルスクリーニング、タンパク質-化合物ドッキング、化合物フィルタリング、タンパク質ポケット体積推定

Compound filtering by estimation of the candidate compound's upper limit size using target protein structure

KEISUKE YANAGISAWA^{1,2,a)} MASAHIKO OHUE^{1,3} TAKASHI ISHIDA^{1,2,3} YUTAKA AKIYAMA^{1,2,3}

Abstract: Protein-ligand docking is widely used as a computational method for drug candidate compound selection, however, the docking simulation is computationally expensive. Thus methods that filter out obviously unnecessary compounds prior to computationally expensive docking simulations have been proposed. In order to reduce the number of candidates, applying physicochemical thresholds like Lipinski's rule of five or filtering with machine learning are widely used. However, they are not considered to be combined with docking simulations. In this study, we proposed two methods presupposed to combine with docking simulation: (1) estimation of the upper limit of compound volume using a protein pocket size estimation method, and (2) estimation of the upper limit of compound volume by experimentally docking smaller amount of sample compounds. The sample compound docking method omitted 15% compounds overlooking only 0.5% dockable compounds.

Keywords: computational drug discovery, virtual screening, protein-ligand docking, compound filtering, protein pocket estimation

1. はじめに

創薬研究において、大量の化合物から新薬候補の化合物

¹ 東京工業大学 情報理工学院 情報工学系
Department of Computer Science, School of Computing,
Tokyo Institute of Technology
² 東京工業大学 情報生命博士教育院
Education Academy of Computational Life Sciences (ACLS),
Tokyo Institute of Technology
³ 東京工業大学 科学技術創成研究院 スマート創薬研究ユニット

Advanced Computational Drug Discovery Unit (ACDD), Institute of Innovative Research, Tokyo Institute of Technology
^{a)} yanagisawa@bi.c.titech.ac.jp

を発見することは「干草の山から針を探す」 (“Finding a needle in a haystack”) ようなものであるとえられており [1], 大量の化合物すべてについて *in vitro* 実験を行う前に計算機を用いてタンパク質との結合しやすさを評価するバーチャルスクリーニング (virtual screening, VS) が広く行われている [2]. このうち, タンパク質や化合物の三次元構造情報を用いた手法 (structure-based virtual screening, SBVS) は, 物理化学的に相互作用を考慮することができ, 既知の薬剤やタンパク質阻害剤を必要としないため注目されている. SBVS ではタンパク質や化合物の3次元構造情報が必要であるが, SBVS が広まるに伴ってデータベースも近年拡充されている. 例えば, タンパク質の三次元構造情報のデータベースである Protein Data Bank (PDB) には 2016 年末の時点で 12.5 万件余りの構造が登録されており, 1 年あたり約 10% 増加している [3]. 一方, 化合物のデータベースについても, 例えば ZINC データベース [4] には約 3,400 万件の化合物が登録されている.

SBVS では, タンパク質と化合物との結合親和性を評価するためにタンパク質-化合物ドッキング計算 (以下, ドッキング計算) がよく用いられる [5] が, 計算量が大きいという問題を持っている. 例えば最も利用されているドッキングツールである AutoDock Vina[6] は 1 つのタンパク質-化合物ペアのドッキング計算に約 500 CPU コア秒を要する [7]. この計算量では 1 つのタンパク質と ZINC 全体の約 3,400 万化合物とのドッキングには 500 CPU コア年以上もの計算時間を要するため, データベース全体をドッキング計算することは事実上不可能といえる.

以上のことから, ドッキング計算を行う前に候補化合物を減らすためのフィルタリングが一般的に行われている. フィルタリング手法としては, 簡易的なドッキング計算を行う手法 [8], [9] や化合物の物理化学的な特徴量を利用した手法 [10], [11], 標的タンパク質に対する既知薬剤情報を用いた手法 [12], [13] などが提案・利用されている. しかし, これらの手法はいずれも「新薬候補として適している化合物を選別する」手法であり, その後に行うドッキング手法との組み合わせを念頭に置いたものではない. 一方で, タンパク質のポケットに入りきらない化合物はドッキング計算で結果が出力されず, 不適な化合物としてスクリーニング対象から除外されるため, タンパク質のポケットサイズをそれぞれ評価・比較することで不適な化合物をあらかじめフィルタリングすることが可能と考えられる. 特に, タンパク質のポケットが小さい場合にはフィルタリングにより多数の化合物を除去することができ, 無駄な計算を抑制することができる.

本研究ではタンパク質のポケットサイズの評価を通して化合物のフィルタリングを行う 2 つの手法を提案し, それぞれの手法の有用性について議論する.

2. 手法

2.1 化合物フィルタリング手法

1 章で述べたように, ドッキング計算はタンパク質の 3 次元構造情報をを利用してタンパク質のどの部位に化合物が結合しうるか探索する. 一般にタンパク質は表面にポケットのような形状を持ち, そこに化合物が入り込むため, タンパク質のポケットに比べて明らかに大きい化合物はドッキング計算するまでもなく不適である. したがってタンパク質のポケットサイズを推定することで, ドッキング結果が outputされる (ドッキング可能である) 化合物と, ポケットに入らずにドッキング結果が outputされない (ドッキング不可能である) 化合物を分別することが可能であると考えた. タンパク質のポケットサイズの推定手法として, 本研究では 2 つの手法を提案する.

- (1) タンパク質ポケットサイズ推定手法を利用した化合物体積の上限決定
- (2) スクリーニング対象化合物の一部を試験的にドッキング計算することによる化合物体積の上限決定
これら 2 つの手法について, 以下に詳細を記す.

2.1.1 手法 1 : タンパク質ポケットサイズ推定手法の利用

タンパク質のポケット予測は基質の結合部位予測や薬剤開発のしやすさを推定するために用いられており [14], [15], ポケット体積の推定に特化したツールとして POVME (POCKET Volume MEasurer)[16], [17] が存在する. 化合物の体積についても, Zhao らによって 2 次元構造から体積を推定する手法が提案されている [18] ため, タンパク質のポケット体積と化合物の体積を比較することで, ドッキング計算すべき化合物を選別する. 手順を以下にまとめる.

- (1) POVME[16], [17] を用いて標的タンパク質のポケット体積を推定する.
- (2) Zhao らの手法 [18] を用いて化合物の体積を推定する.
- (3) タンパク質のポケット体積を超える体積を持つ化合物を除外する.

この手法はタンパク質の立体構造情報から直接閾値を決定するため計算コストが小さく済むという長所がある. ただし, ポケット形状の推定に失敗すると実際のサイズと大きく異なる推定値が算出される可能性がある.

2.1.2 手法 2 : サンプル化合物群のドッキング結果の利用

本研究では, タンパク質のポケットにドッキングすることができる化合物の体積の上限を推定したい. そこで, スクリーニング対象の化合物群からサンプリングを行い, 実際にドッキングを行うことで化合物の体積の上限値を決定する. 具体的には, 以下の手順で化合物のフィルタリングを行う.

- (1) スクリーニング対象の化合物群から一定数 (1,000 件など) をランダムサンプリングし, サンプル化合物群

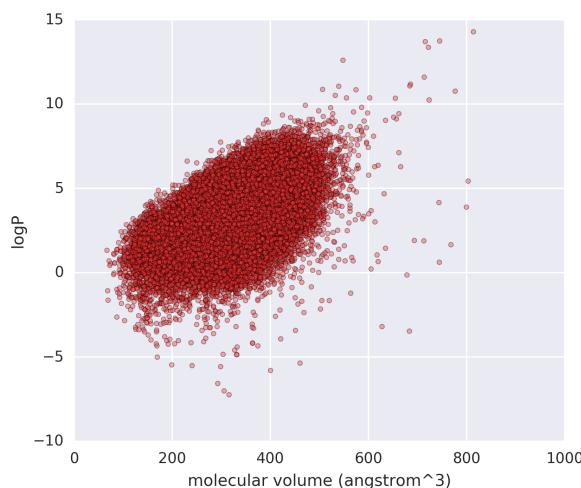


図 1 テスト用化合物群の分布. logP は脂溶性を示す値である.

を作成する.

- (2) サンプル化合物群のドッキング計算を行う.
- (3) サンプル化合物群の中でドッキングが可能であった化合物の体積を Zhao らの手法 [18] を用いて推定する.
- (4) ドッキング可能であった化合物を一定の割合 (90%など) 包含するような体積の値を閾値とする.
- (5) 閾値を超える体積を持つ化合物を除去する.

この手法ではサンプル化合物群のドッキング計算に時間を要するものの、ポケットサイズ推定に比べて高精度な閾値の決定を行うことが期待できる。また、サンプル化合物群が化合物全体に対して例えば 5% 以下であるなど、十分に小さい場合には全体の計算コストに与える影響は限定的である。

3. 評価実験

まず 2 つの提案手法の評価実験に共通して利用するテスト用のデータセットについて説明し、続いて個々の実験の詳細について述べる。

3.1 テスト用化合物群

ZINC all purchasable サブセットに登録されている化合物 22,724,825 件から 100,000 件の化合物をランダムサンプリングし、それぞれ Glide SP モード [19] を用いたドッキング計算を行った。このテスト用化合物群の体積-logP 分布を図 1 に示す。なお、最小化合物の体積は 66.92 \AA^3 、最大化合物の体積は 813.72 \AA^3 である。

3.2 標的タンパク質群

1 章で述べた通り、タンパク質によってポケットサイズは大きく異なり、本研究で提案するフィルタリング手法はポケットの小さなタンパク質ほど閾値が小さく、化合物フィルタリングの効果が大きい。このため、以下の手順を

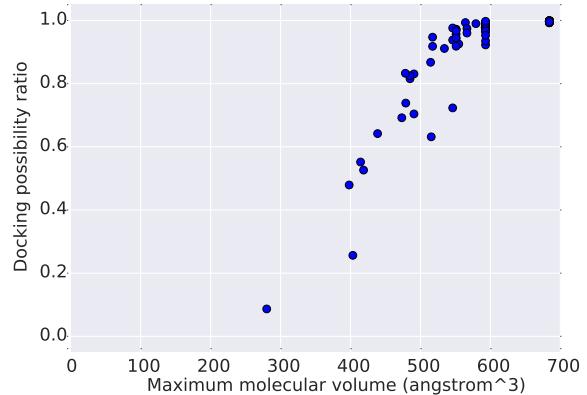


図 2 DUD-E 全タンパク質に対する 10,000 化合物ドッキング結果.
各点が 1 つのタンパク質を示す.

用いてドッキング計算のベンチマークデータセットである DUD-E[20] に登録されている 102 種のタンパク質から 11 種を選択した。

- (1) ZINC all purchasable サブセットからランダムサンプリングした化合物 10,000 件を 102 タンパク質へドッキング
- (2) ドッキング可能な化合物が全体の 80% 以下であるようなタンパク質構造を選択

DUD-E に登録されている 102 タンパク質それぞれのドッキング可能な最大体積およびドッキング可能率を図 2 に、選択された 11 種のタンパク質を表 1 に示す。タンパク質選択に用いた 10,000 件、およびテスト用化合物群 100,000 件、いずれのデータセットでも約 45% がドッキング不可能な化合物であり、この値がフィルタリングを行った場合の理想的な化合物削減率となる。

3.3 実験 1: タンパク質ポケットサイズ推定手法の利用

本研究で用いた POVME のバージョンは 2.0.1 であり、標的タンパク質と共に結晶構造を形成していたリガンド化合物の中心をポケットサイズ推定の中心に指定した。また、ContiguousPocketSeedBox を化合物の中心座標を用いて設定することで、指定した部分と連結していないポケット候補部分を除外した。パラメータはポケット判定の間隔を示す GridSpacing を 0.5 \AA に設定し、それ以外はデフォルト値としている。

3.4 実験 2: サンプル化合物群のドッキング結果の利用

サンプル化合物群としてテスト用データセット 100,000 件の中から 1,000 件のランダムサンプリングを行い、Glide SP モードを用いたドッキング計算を行った。また、ドッキング可能なサンプル化合物群 80%, 90%, 95%, 100% を包含するような化合物体積の閾値を求め、それぞれ評価を行った。

表 1 使用した 11 種の標的タンパク質

タンパク質名	10,000 化合物		テスト用化合物群	
	ドッキング可能化合物率	ドッキング最大体積	ドッキング可能化合物率	ドッキング最大体積
prgr	74.0%	478.99 Å ³	74.3%	458.22 Å ³
lkha4	72.4%	546.08 Å ³	72.6%	562.25 Å ³
comt	70.5%	490.70 Å ³	71.0%	536.33 Å ³
pgh1	69.3%	473.33 Å ³	69.1%	512.46 Å ³
fgfr1	64.5%	462.30 Å ³	63.8%	472.59 Å ³
pgh2	63.6%	515.42 Å ³	63.5%	499.20 Å ³
mcr	55.6%	414.37 Å ³	55.8%	450.15 Å ³
andr	53.0%	418.67 Å ³	53.0%	446.25 Å ³
aofb	48.2%	398.09 Å ³	47.4%	417.23 Å ³
cah2	25.8%	403.34 Å ³	25.2%	415.43 Å ³
sahh	8.7%	280.19 Å ³	8.2%	315.78 Å ³

表 2 フィルタリング計算時間

フィルタリング手法	11 タンパク質 計算時間 [sec.]		
	平均	最大	最小
ポケットサイズ推定	37.88	109.91	6.88
サンプルドッキング	469.85	733.36	129.31

3.5 計算機環境

本研究では、東京工業大学の TSUBAME 2.5 の Thin ノードを利用した。Thin ノードには 6 コアの Intel Xeon X5670 CPU が 2 個、メモリが 54 GB 搭載されている。

3.6 評価方法

速度評価では、time コマンドを用いて実行時間を測定した。それぞれの評価には 3 回の独立な実時間計測を行い、その中央値を用いた。なお、POVME によるポケットサイズ推定、Glide SP モードによるドッキング計算はいずれも 12 並列で計算を行っている。一方精度評価では提案した 2 つの手法に加え、比較対象として Lipinski の法則のうち化合物サイズの制限を示すルールである分子量の制限（分子量 ≤ 500）を適用した場合のフィルタリング結果を示す。

4. 実験結果

4.1 フィルタリングの計算時間

表 2 にフィルタリングに要した時間を示す。ポケットサイズ推定手法は、サンプル化合物群のドッキング計算によるフィルタリング手法に比べ 10 倍以上高速である。ただし、サンプル化合物はバーチャルスクリーニングの対象となる化合物から取得しているため、評価すべき残りの対象化合物が減少する。この効果を考慮したバーチャルスクリーニング全体の計算時間は 5.1 節で述べる。

4.2 フィルタリング精度

本研究で提案した 2 つの手法、および比較手法である Lipinski の法則に基づく分子量閾値の適用を行った時の精

度を表 3 に示す。Lipinski の法則に基づく分子量の閾値では約 3% の化合物削減にとどまっており、十分なフィルタリングが行えていない一方、サンプル化合物群のドッキングによるフィルタリングを用いると高精度に化合物を削減することができる。この手法ではドッキング可能な化合物の保持率を指定でき、利用者の要望に応じて閾値を変更できるという利点もある。

5. 考察

5.1 バーチャルスクリーニングへの適用

4 章の結果およびテスト用化合物群 100,000 件のドッキング計算の時間から、各フィルタリング手法をバーチャルスクリーニングに適用した場合の計算時間および精度の推定を表 4 に示す。Lipinski の法則の分子量閾値ではバーチャルスクリーニング全体の計算量がほとんど変化していない。また、ポケットサイズ推定手法を用いるよりもサンプル化合物ドッキングによる手法を用いることで、取りこぼしを示すドッキング可能な化合物の削減率が少ないままバーチャルスクリーニングの実行時間を抑えることができる。

5.2 タンパク質ポケットサイズ推定結果

4.2 節の結果から、ポケットサイズ推定手法を用いると 11 タンパク質の平均として多数のドッキング可能な化合物が削減される結果が得られた。この問題についてタンパク質ごとに結果を確認したところ、

- (1) ポケットが浅く、ドッキングに利用できる空間もポケットではないと判定するケース（例：comt タンパク質、図 3）
 - (2) 本来ポケットと判定されてほしい空間以外もポケットと判定するケース（例：lkha4 タンパク質、図 4）
- の 2 つのケースがあることがわかった。(1) のケースはドッキング可能な化合物の平均削減率を高め、(2) のケースはドッキング不可能な化合物の平均削減率を下げている。こ

表 3 フィルタリング精度

フィルタリング手法	テスト用化合物群 平均削減率		
	ドッキング可能化合物	ドッキング不可能化合物	全化合物
ポケットサイズ推定	33.45%	46.75%	40.69%
80%	19.37%	78.17%	46.63%
サンプル	10.02%	67.22%	37.39%
ドッキング	5.59%	57.88%	31.27%
100%	0.35%	26.01%	15.63%
Lipinski の法則 (分子量)	0.22%	7.08%	2.94%

表 4 テスト用化合物群 100,000 件を用いたバーチャルスクリーニング性能評価

フィルタリング手法	11 タンパク質平均 推定計算時間 [sec.]			ドッキング可能 化合物削減率
	フィルタリング	ドッキング計算時間	合計	
ポケットサイズ推定	37.88	27,866.58	27,904.46	33.5%
80%	469.85	24,824.93	25,294.78	19.4%
サンプル	469.85	29,122.90	29,592.74	10.0%
ドッキング	469.85	31,969.60	32,439.45	5.6%
100%	469.85	39,244.51	39,714.36	0.4%
Lipinski の法則 (分子量)	-	45,603.27	45,603.27	0.2%
(フィルタリングなし)	-	46,984.61	46,984.61	-

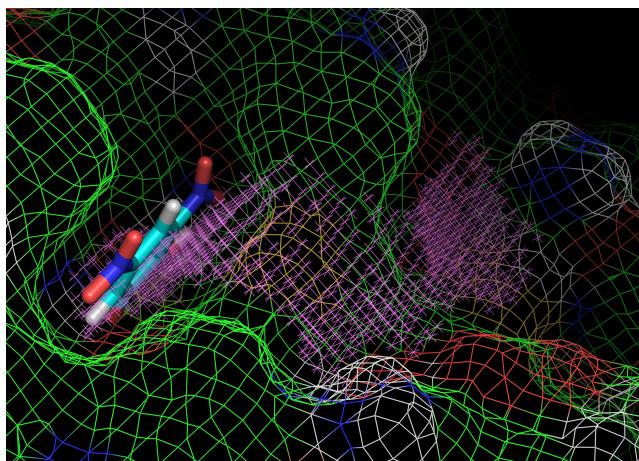


図 3 comt タンパク質に対するポケットサイズ推定結果。タンパク質（メッシュ）と共に結晶を生成している化合物、ポケットと推定された領域（紫色の点群、 173.75 \AA^3 ）が表示されている。

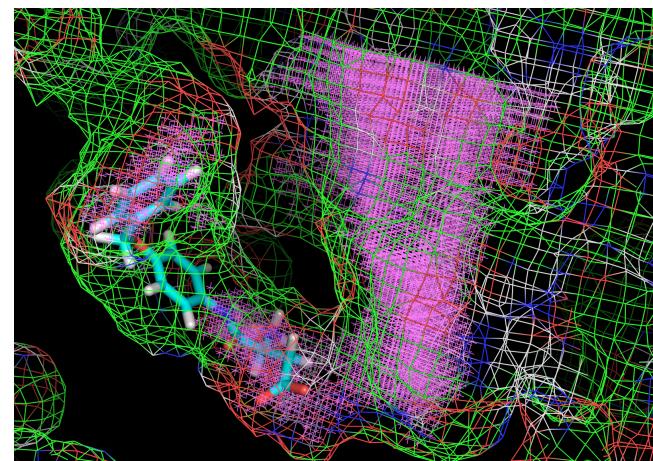


図 4 lkha4 タンパク質に対するポケットサイズ推定結果。タンパク質（メッシュ）と共に結晶を生成している化合物、ポケットと推定された領域（紫色の点群、 781.63 \AA^3 ）が表示されている。

これらの問題は 3 次元立体構造を確認し、妥当なポケット推定が行われているか確認することで回避が可能である。しかし、複数のタンパク質を対象とする場合には一律に手法を適用することができないという問題があり、自動化できないことによるデメリットは大きい。

5.3 化合物体積フィルタリングを行う妥当性の評価

本研究で提案した手法は化合物の体積を利用しているが、ドッキングが可能かどうかの推定に化合物の体積が十分に有用な情報であるか評価を行う必要がある。そこで、体積の小さな化合物から順番にドッキング可能な化合物であると判定した場合の ROC 曲線を作成した（図 5, 6）。図 5 で示した andr タンパク質では ROC 曲線下面積 (AUROC)

が高く、化合物体積でドッキング可能な化合物とそうでない化合物をよく分離することができているが、図 6 で示した lkha4 タンパク質に関しては AUROC が低く、体積情報だけでは十分にドッキング結果を出力できる化合物と出力できない化合物を分けることができていない。

5.4 新たな特徴量の追加

5.3 節の結果より、用いる特徴量には改善の余地が存在している。体積だけではポケットに合わせた形状にならない化合物を排除できないことが原因として考えられ、その場合化合物の内部自由度（化合物が持つ回転可能な結合数）は新たな特徴量として有用である。体積のみでは十分に化合物を弁別できなかったタンパク質 lkha4 について、化合

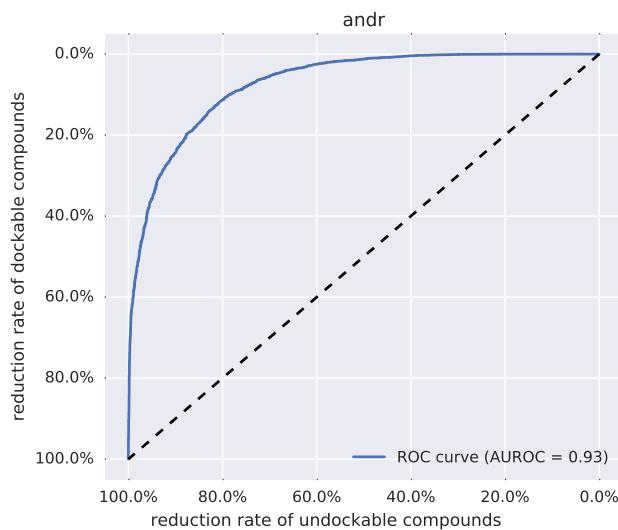


図 5 andr タンパク質に対して、体積で化合物のフィルタリングを行った場合の ROC 曲線

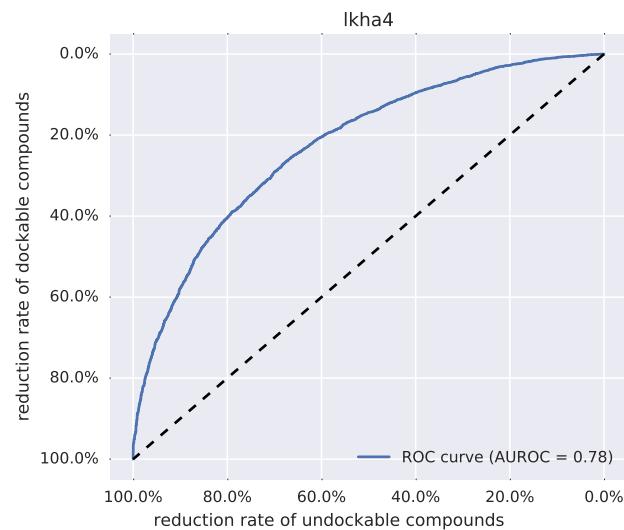


図 6 lkha4 タンパク質に対して、体積で化合物のフィルタリングを行った場合の ROC 曲線

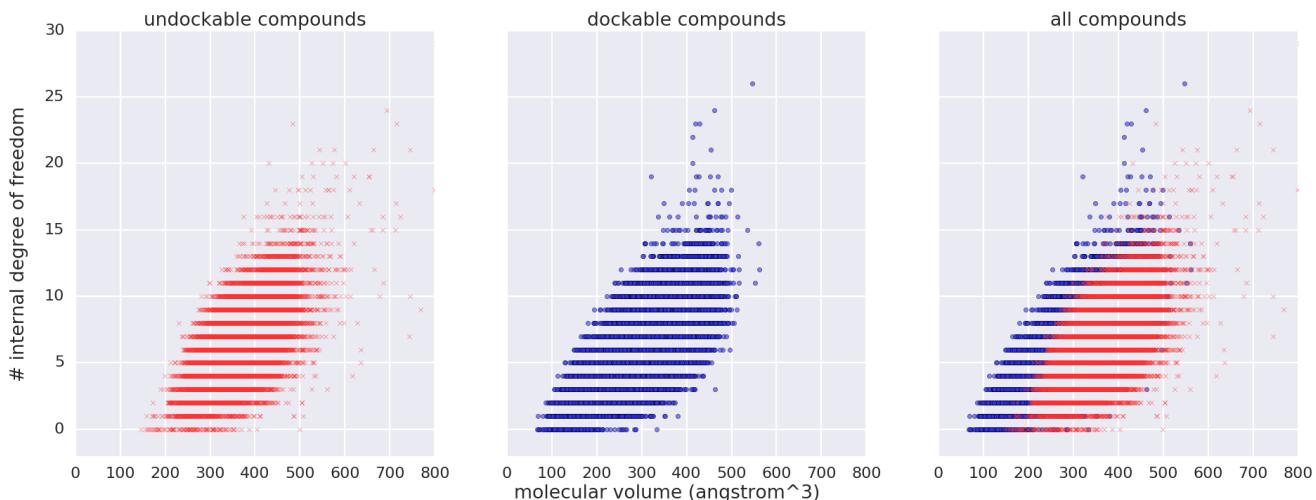


図 7 テスト用化合物群の化合物体積と化合物内部自由度の分布。青は lkha4 タンパク質にドッキング可能な、赤は lkha4 タンパク質にドッキング不可能な化合物を表す。

物の体積と化合物の内部自由度を用いて散布図を作成すると（図 7），化合物の内部自由度もドッキング可能か否かに関係しているという結果が得られている。このため，化合物の体積と内部自由度から化合物の「ドッキング可能性関数」を作成することで，フィルタリング精度が向上すると考えられる。

6. 結論

本研究では，標的タンパク質の立体構造を用いた 2 種類の化合物フィルタリング手法を提案・検討した。

(1) タンパク質ポケットサイズ推定手法の利用

タンパク質のポケットサイズを推定する手法として POVME を利用した。高速に化合物体積の閾値を決定することができるが，ポケットサイズの推定値が大きく誤っているケースがあり，推定結果の妥当性を確認

する必要がある。

(2) サンプル化合物群を利用した化合物サイズ推定

一部の化合物のドッキング結果を用いて閾値の決定を行うことで，ドッキング可能な化合物の削減率を調節しながらのフィルタリングを実現した。例えば，0.5%未満のドッキング可能な化合物の削減率で 15% の高速化を達成した。

今後の課題としては，以下のことが考えられる。

- 5.4 節より，化合物の体積のみでなく化合物の内部自由度を特徴量として用いることで，より高精度に化合物のドッキング可能性を評価することができると考えられる。
- 上記の手法のさらなる拡張として，多数の特徴量を導入し，機械学習を用いたドッキング可能性の評価をすることが考えられる。この場合，タンパク質ごとに有

用な特徴量が異なる場合にも対応できると期待される。

- 本研究では、ドッキング可能か否かで化合物を分離したが、本手法を有効的に適用できるタンパク質はDUD-E 102 タンパク質中 11 タンパク質と限定的である。一方バーチャルスクリーニングではドッキング計算で得られたスコアが上位であるような数%～1%以下の化合物を選別して *in vitro* 実験に移行するため、スコアが上位に入るかどうかを判別することで大幅な計算量の削減が期待される。ただし、スコアはタンパク質ポケットの表面電荷など物理化学的特徴にも左右されるため、本研究で用いた体積などの情報以上の特徴量が必要となる。

謝辞 本研究の一部は文部科学省 博士課程教育リーディングプログラム 東京工業大学「情報生命博士教育院」, JSPS 科研費 基盤研究 (A) (24240044), JST CREST 「EBD: 次世代の年々タバコ処理に向けたエクストリームビッグデータの基盤技術」, および JST 研究成果展開事業 世界に誇る地域発研究開発・実証拠点 (リサーチコンプレックス) 推進プログラム 「世界に誇る社会システムと技術の革新で新産業を創る Wellbeing Research Campus “Tonomachi”」の支援を受けて行われた。

参考文献

- [1] Klon A.E. *et al.*: “Finding more needles in the haystack: a simple and efficient method for improving high-throughput docking results”, *J. Med. Chem.* 47, 2743–2749, 2004.
- [2] Sliwoski G. *et al.*: “Computational methods in drug discovery”, *Pharmacol. Rev.* 66, 334–395, 2014.
- [3] Rose P.W. *et al.*: “The RCSB Protein Data Bank: views of structural biology for basic and applied research and education”, *Nucleic Acids Res.* 43, D345–D356, 2015.
- [4] Irwin J.J. *et al.*: “ZINC: a free tool to discover chemistry for biology”, *J. Chem. Inf. Model.* 52, 1757–1768, 2012.
- [5] Meng X. *et al.*: “Molecular docking: a powerful approach for structure-based drug discovery”, *Curr. Comput. Aided Drug Des.* 7, 146–157, 2011.
- [6] Trott O. *et al.*: “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading”, *J. Comput. Chem.* 31, 455–461, 2010.
- [7] Drwal M.N. *et al.*: “Combination of ligand- and structure-based methods in virtual screening”, *Drug Discov. Today Technol.* 10, e395–e401, 2013.
- [8] Niimivehmas S.P. *et al.*: “Ultrafast protein structure-based virtual screening with Panther”, *J. Comput.-Aided Mol. Des.* 29, 989–1006, 2015.
- [9] 柳澤 溪甫 他: “フラグメント分割に基づく超高速化合物プレスクリーニング手法 ESPRESSO”, 情報処理学会研究報告 2016-BIO-46, 1–7, 2016.
- [10] Lipinski C.A. *et al.*: “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings”, *Adv. Drug Deliv. Rev.* 23, 3–25, 1997.
- [11] Bickerton G.R. *et al.*: “Quantifying the chemical beauty of drugs”, *Nat. Chem.* 4, 90–98, 2012.
- [12] Nicolaes G.A.F. *et al.*: “Rational design of small molecules targeting the C2 domain of coagulation factor VIII”, *Blood* 123, 113–120, 2014.
- [13] Gabrielsen M. *et al.*: “Identification of novel serotonin transporter compounds by virtual screening”, *J. Chem. Inf. Model.* 54, 933–943, 2014.
- [14] Halgren T.A.: “Identifying and characterizing binding sites and assessing druggability”, *J. Chem. Inf. Model.* 49, 377–389, 2009.
- [15] Guilloux V.L. *et al.*: “Fpocket: An open source platform for ligand pocket detection”, *BMC Bioinform.* 10, 168, 2009.
- [16] Durrant J.D. *et al.*: “POVME: An algorithm for measuring binding-pocket volumes”, *J. Mol. Graph. Model.* 29, 773–776, 2012.
- [17] Durrant J.D. *et al.*: “POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics”, *J. Chem. Theory Comput.* 10, 5047–5056, 2014.
- [18] Zhao Y.H. *et al.*: “Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds”, *J. Org. Chem.* 68, 7368–7373, 2003.
- [19] Friesner R.A. *et al.*: “Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy”, *J. Med. Chem.* 47, 1739–1749, 2004.
- [20] Mysinger M.M. *et al.*: “Directory of Useful Decoys, Enhanced (DUD-E): better ligands and decoys for better benchmarking”, *J. Med. Chem.* 55, 6582–6594, 2012.