

分布の異なる複数のドッキングソフトを用いたインシリコスクリーニング

岡田 正人* 塚本 雅登† 大和田 勇人‡ 青木 伸†

東京理科大学理工学研究科経営工学専攻* 同薬学部生命創薬科学科† 同理工学部経営工学科‡

1. はじめに

近年、コンピュータとソフトウェア双方の性能向上により創薬シミュレーションが可能となっている。特に、病気に関係するタンパク質(ターゲットタンパク)に対して効果のある化合物の探索が研究されている。この技術はインシリコスクリーニングと呼ばれる。

インシリコスクリーニングでは、タンパク質と化合物の結合実験をコンピュータ上でシミュレーションする。特に、タンパク質の構造を用いる手法をドッキングシミュレーションと呼ぶ。ドッキングシミュレーションを行うためのソフトは DOCK[1]や AutoDock[2]など多数開発されている。これらのソフトウェアは結合実験の結果から精度が高められており、ターゲットタンパクに対して結合力をもつことがわかっている化合物(既知リガンド)は高いスコアを出すようになっている。

しかし、ターゲットタンパクや化合物と、各ソフトウェアの間には相性が存在するため、単一のドッキングソフトを信頼するのは危険である。本研究では複数のドッキングソフトを用いてシミュレーションを行い、各スコアを統合して利用するコンセンサススコアリングを行うことで問題を解決する。ただし、各ドッキングソフトが出すスコアは分布が大きく異なる。本研究では各スコアに標準化処理を行うことで、計算が容易な数値を得る。

2. 使用ソフトウェア

本研究で用いたドッキングソフトは以下の3種類である。本研究ではシミュレーションによるコスト削減を考慮し、無料で用いることのできるソフトウェアのみを用いた。

- AutoDock Vina
- DOCK (version6)
- Accelrys Discovery Studio

AutoDockVina[3]は Autodock を元に、並列化による速度向上と精度向上のために作られている。DOCK は精度の良さから多くの研究論文で用いられている最も古いドッキングプログラムである。Discovery Studio[4]は商用ソフトウェアであるが、東京理科大学学内で利用することができる。本研究では Discovery Studio において動作が比較的高速な LibDock を用いる。

3. ランダムデータセット

複数のソフトウェアを用いたときの精度検証のため、本研究ではランダムな 1000 個の化合物をテストセットとし、ドッキングシミュレーションを行った。これらの化合物は化合物データベースの ChEMBL からランダムに取得している。

また、ターゲットタンパクにはガン細胞の発生と生存に関係する XIAP(X-linked Inhibitor of Apoptosis Protein) (PDBID:3HL5 由来)を用いる。

4. スコア変換

本研究で用いる各ドッキングソフトが出すスコアの分布はそれぞれ大きく異なる。Auto DockVina は -20~0 の間に分布するのに対し、DiscoveryStudio の分布は 0~200 である。そのため、コンセンサススコアリングを行うためには相互計算が可能な形に変換する必要がある[5]。本研究では各スコア分布の範囲を同一にし、計算や比較を可能にするために、各スコアに対して標準化処理を行う。

In silico screening using docking software with different distribution.

Masato OKADA*, Masato TSUKAMOTO†, Hayato OHWADA‡, Shin AOKI†

Department of Industrial Administration, Graduate School of Science and Technology, Tokyo University of Science*, Faculty of Pharmaceutical Sciences, Tokyo University of Science†, Department of Industrial Administration, Faculty of Science and Technology, Tokyo University of Science‡

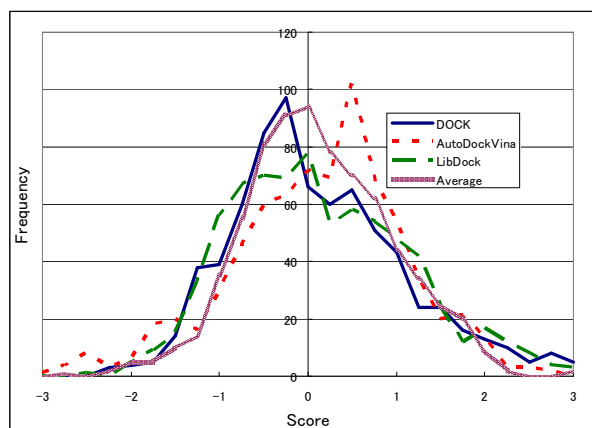


図1 標準化

図1は前述の1000化合物に対して各ドッキングソフトを用いてシミュレーションを行い、得られたスコアについて標準化処理を行ったときのスコア分布を示している。実験では780の化合物についてスコアを得ることができた。Averageは各化合物における3種のスコアの平均値である。図のように、各ソフトウェアから得られるスコアの分布を同じ範囲内に変換できている。

5. コンセンサススコアリング

上記の標準化で得られた新しいスコアを用いて各化合物の結合スコアをランク付けする。ランク付け時において重要となるのは、既知リガンドが上位に位置づけられるかどうかである。

本研究ではコンセンサススコアリングで、複数スコアのなかで最も悪いスコアを取り入れる、最大値(MAX)手法を用いる。これは、既知リガンドはどのドッキングソフトでも高いスコアが取れるという性質を利用している。また、この手法を用いたときに上位に検出される化合物は、すべてのドッキングソフトで高いスコアを得られているため、既知リガンドと同様にタンパク質に作用する可能性が高いといえる。

6. 実験結果

ランダムに取得した1000の化合物に対し、コンセンサススコアリングを行った結果を図2に示す。図中の横軸はテストデータの処理数、縦軸は既知リガンドの検出数である。既知リガンドとして4種の化合物を用いる。これらはXIAPとの活性が報告されている。既知リガンドが他の化合物に対して上位に検出できているほど、実験に近い、精度の高いシミュレーションができているといえる。

図では各ドッキングソフトを単独で用いた場

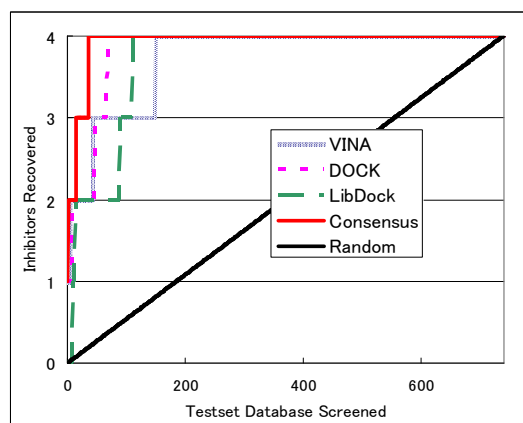


図2 コンセンサススコアリング結果

合とコンセンサススコアリングを行った場合を比較している。各ソフトウェア単独では100位前後ですべての既知リガンドを検出しているのに対し、コンセンサススコアリングの場合には36位で検出できている。以上のデータから、本研究の手法によりシミュレーション精度を向上させることができている。

7. おわりに

本研究では複数のドッキングソフトを用いたインシリコスクリーニングに焦点をあてた。本研究から、複数のスコアを組み合わせることにより精度を向上することが確認できた。

現在はこれらの手法を用いて大規模なスクリーニングを行っている。対象として400万化合物を含む商用データベースと、タンパク質XIAPを用いている。そのため、ドッキングシミュレーションの精度向上と速度向上が必要である。

参考文献

- [1] Ewing, T. J.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* 1997, 18, 1175–1189.
- [2] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* 1999, 20, 1639–1662.
- [3] Trott, O. and Olson, A.J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* 31, 455–461.
- [4] Discovery Studio, 2.1; Accelrys Inc.: San Diego, CA 92121, U.S.A., 2008.
- [5] Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J. Chem. Inf. Model.* 2006, 46, 380–391.