

HIV-1 プロテアーゼ, p38, CDK-2 複合体の立体構造を用いた ドッキングソフトウェアの能力評価

小田 彰史, 吉田 朋起, 岡安 愛, 神山由紀子, 高橋 央宜, 松崎 久夫

Computational Docking of HIV-1 protease, p38 and CDK-2 for Software Validation

Akifumi ODA, Tomoki YOSHIDA, Megumi OKAYASU, Yukiko KAMIYAMA, Ohgi TAKAHASHI,
and Hisao MATSUZAKI

(Received November 21, 2006)

In this study, we evaluated the docking accuracy of software ArgusLab, which is one of freely available computational docking programs both for commercial and academic users, by using experimentally determined protein-ligand complex structures. The complexes include one of three proteins, HIV-1 protease, p38 and CDK-2, as target proteins. Because these proteins are well known drug targets, the tests can be suitable for practical use. All three steps of computational docking, i.e. pose construction, pose selection, and virtual screening, were tested for 16 complexes. The results indicate that ArgusLab can generate reasonable complex structures for around 50% of protein-ligand systems, and ligand flexibility play important roles in docking accuracy. Although pose construction and pose selection were successfully carried out both for HIV-1 protease and for CDK-2 complexes, for binding free energy calculations, which is one of the important steps in virtual screening, the accuracy of results for HIV-1 protease complexes was better than those of kinase complexes. By using these results as reference, reliable docking calculations are expected to be carried out by ArgusLab.

Key words — Computer-Aided Drug Design, Structure-Based Drug Design, Computational Docking, Scoring Function, Virtual Screening

薬物標的となる生体高分子の立体構造を元にドッキングデザインを行う Structure-Based Drug Design (SBDD) において、標的タンパク質と低分子化合物との結合様式を推測することは非常に重要である。¹⁾ これまで生体高分子-リガンド複合体構造を計算機的に予測する、すなわち、計算機によるドッキングのためのプログラムは多数開発されているが、これらの多くは商用・学術用問わず有償で提供されている。非営利の学術研究に限り無料で使用できるソフトウェアも存在するものの、それらも商用での利用はやはり有料である。有償のソフトウェアは高い信頼性と手厚いサポート体制

を持っているものの、Computer-Aided Drug Design (CADD) の非専門家に対する教育や、創薬の初期段階で簡易的に SBDD の可能性を検討する目的で使用する場合には、その導入コストが無視できない。また大規模計算を行う場合、市販ソフトウェアでは計算規模に応じた数のライセンスを用意する必要があり、ライセンス数によって研究の進捗が支配されてしまう。そこで、無料でドッキングを行うことのできるソフトウェアが、CADD において重要な役割を果たすことになる。計算機によるドッキングのためのソフトウェアはほとんどが有償あるいはアカデミックフリーであるが、使用

権が無料で得られるドッキング関連ソフトウェアにArgusLab²⁾がある。

ArgusLabは元々分子モデリングのために開発されたソフトウェアであるが、バージョン4.0.1では生体高分子に対して低分子リガンドをドッキングする機能が導入されている。また、ドッキングにおいて重要な役割を果たすスコア関数については、フリーのスコア関数であるXScore³⁾を元に作られたAScoreを採用している。さらにこのソフトウェアはWindows上で使用でき、かつグラフィカルユーザーインターフェース(GUI)も整備されているため、容易に使用できるという点で非専門家の教育や簡易計算という目的に合致している。このようにArgusLabは無料でありながら操作性や動作環境の点でも優れているソフトウェアであるが、一方でドッキングの能力については十分に検討されていない。ドッキング機能について説明された文書はweb上²⁾に少数存在するものの、実際の計算手順に応じたドッキングの精度に関する議論や、ドッキングポーズに対するスコアリングの妥当性の議論などはほとんど存在しない。

そこで本研究では、ArgusLabのドッキング機能について、立体構造既知のタンパク質-リガンド複合体構造に対するテスト計算を通じて評価する。ドッキング研究を行う場合、クリアすべき段階は大きく分けて3つ存在する。⁴⁻⁶⁾ まず最初の段階では、低分子化合物を標的高分子にはめ込む作業が行われる。この段階では一般に複数のポーズが作成されることになるが、そのうちの最低1つが正しい配座・配向・位置となっている必要がある。次の段階として、得られた複数のポーズの中から、正しいポーズを選び出す作業に入る。この段階では結合自由エネルギーを推定するスコア関数が使用され、ポーズが順位付けされる。最後に、一つの標的高分子に対して複数の低分子化合物をドッキングさせ、どの化合物が最も薬になりやすいかを予測する。この作業を仮想的にスクリーニングを行うという意味でヴァーチャルスクリーニングと呼ぶ。これら3つの段階すべてが常に必要になるとは限らないものの、3つの段階すべてで良好な結果を与えるドッキングプログラムが望ましいことは言うまでもない。本研究では1型ヒト免疫

不全ウイルス(HIV-1)プロテアーゼ、p38マイトジェン活性化プロテインキナーゼ(p38)、2型サイクリン依存性キナーゼ(CDK-2)とリガンド分子との複合体を用いて、計算の各段階におけるArgusLabの能力評価を行った。これらのタンパク質はいずれも疾患と関連が深いと言われており、分子標的薬開発における薬物標的となっている。そのためこれらの系に対するドッキング計算は現実のSBDDでも頻繁に行われており、実際の使用に即したテストが行えるのではないかと考えている。これにより創薬研究におけるArgusLabの可能性および限界が明確になり、ソフトウェアの導入コストなしにドッキング計算を行いたい場合の指針になるのではないかと考えている。

方 法

HIV-1プロテアーゼ、p38およびCDK-2の立体構造はRSCB Protein DataBank(PDB)⁷⁾から入手した。使用した複合体のPDB IDは、HIV-1プロテアーゼは1HPV, 1HPX, 1HSG, 1HVS, 1QBU, p38は1A9U, 1BL6, 1BL7, 1BMK, 1DI9, 1KV1, 1KV2, CDK-2は1PF8, 1PXM, 1PXN, 1PXPである。また、リガンド低分子についてはArgusLabの水素原子付加機能を使用して水素を付加した。ドッキングの際の設定についてはドッキングエンジン、パラメータともデフォルトの値を使用した。実際の使用では高速性が重視されることを考えて、ドッキングの結果得られたポーズは上位10個のみを採用した。また、ドッキングの際には水分子は一切考慮せずに行った。これらの計算はPentium 4を搭載し、メインメモリが512MBのWindows XPマシンによって実行した。

計算結果の評価については、ドッキングの各段階での操作を想定して行った。本研究では(i)まず10個のポーズのうち1つでも妥当な構造があったかどうか、(ii)あったとすればその妥当な構造を上位に(特に1位に)ランク付けできたか、(iii)そして計算で予測した結合親和性が実験値と対応していたかどうか、を検討した。どのようなポーズが「妥当な構造」と呼ぶのかについては諸説あるが、本研究ではPDBに収載されている実

験構造と比較して root mean square deviation (RMSD: 計算構造と実験構造を比較したときの各原子の座標のずれを2乗し、平均を取ったものの平方根) が 3.0\AA 以下のポーズを妥当なポーズとして採用した。また、結合親和性の実験値のデータについては PDB Bind^{8,9)} から取得した。

結果および考察

まず、各標的タンパク質に対して、(i) 10個のポーズのうち1つでも妥当な構造があったかどうかの評価を行った結果を Fig. 1 に示す。また、16個の系すべてに対する結果もあわせて示す。この図に示したように、CDK-2では75%の系でドッキングに成功しているのに対して、p38では30%程度の系でしか成功していない。特に、1A9Uと1KV2の2つの系ではポーズが1つも得られないという結果になっている。これらの系は複数の環構造が単結合によって結ばれており、比較的配座自由度が高い。他の系においても総じて配座自由度の高い分子でドッキングに失敗する傾向があり、標的タンパク質ではなくリガンド側の要因でドッキングの正否が決定しているのではないかと考えられる。特にp38とCDK-2はいずれもキナーゼであり、この両者で成功率が著しく異なるという結果は、

リガンドの配座自由度の重要性を示唆している。ただし、HIV-1プロテアーゼでは1HVSや1QBUなどの系において巨大かつ自由度の高いリガンド分子のドッキングに成功しており、必ずしも自由度の高い分子ではドッキングできないというわけではない。また、この結果から、標的タンパク質の性質の影響を受けている可能性も否定できない。なお、テストセット全体での成功率はおよそ44%であり、決して高い値ではないものの高速ドッキングプログラムとしてある程度の有用性は保っているものと考えている。本研究ではポーズを10個に限定しているが、より多数のポーズを取りだしても計算コスト的にはあまり違いがないため、本プログラムが高速であることを考えれば出力数を増やすことも十分可能である。そうすることで、広範囲な系に対してより高い成功率が得られるのではないかと期待できる。

次に、Table 1に(ii) 妥当な構造を上位にランク付けできたかの結果を示す。ここでは順位付けの成功率と、妥当なポーズの順位の平均値を示している。成功率については、ドッキングに成功した系(i)の検討で「成功」と判断された系の数を分母に、順位1位のポーズが妥当な構造だった系の数を分子に記している。また、順位の平均値については、それぞれの系で得られた妥当なポ-

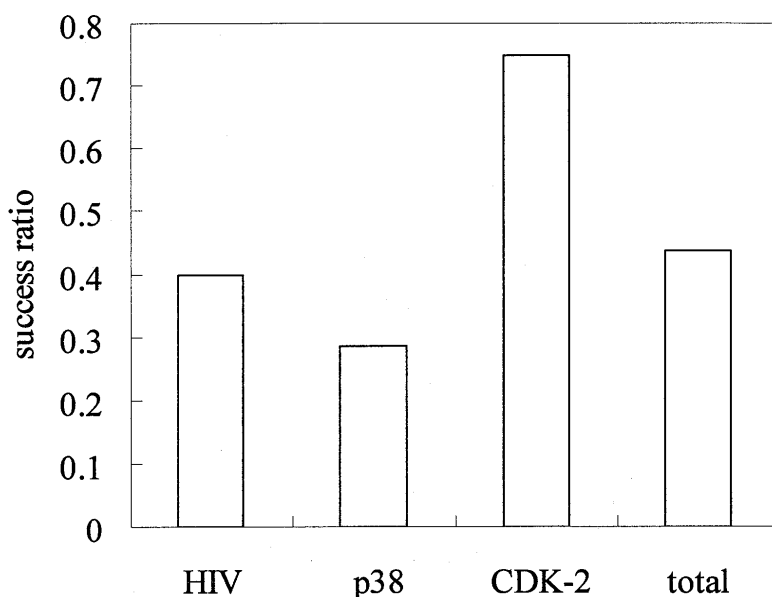


Fig. 1. Success ratio of docking for each target protein and total ratio.

Table 1. Results of ranking.

	HIV-1 protease	p38	CDK-2	total
Succession ratio	1/2	0/2	2/3	3/7
average of rank	5.00	7.00	1.67	4.14

(Success ratio = (number of complexes for which top ranked poses were reasonable) / (number of all successfully docked complexes)).

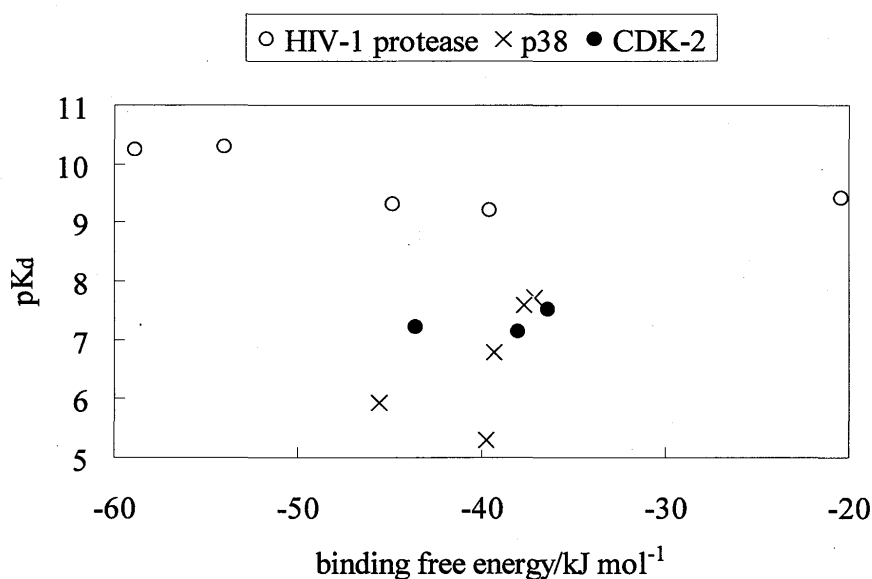
ズのうち、順位が最上位のものだけを取り出し、その順位の平均値を算出した。ここに示したように、(i) の検討で芳しい結果の得られなかった p38 については成功率、平均順位ともに最も悪い結果となっており、順位付けについてもうまくいっていないことがわかる。一方母数が少ないため、より詳細な検討が必要ではあるものの、HIV-1 プロテアーゼおよび CDK-2 については他のスコア関数と比較してそれほど遜色のない値となっており、¹⁰⁾ 注意深く使用することで妥当な順位付けが可能となるのではないかと考えられる。なお、こちらについてはポーズ数を増加させても結果が変化しないため、より良好な結果を得るためにはポーズ数以外の要素を調整する必要がある。

最後に、Table 2 および Fig. 2 に、(iii) 計算で予測した結合親和性が実験値と対応していたかどうかについて評価を行った結果を示す。Table 2 には

Table 2. Correlation coefficients between experimentally obtained pK_d and computational binding free energy

	HIV-1 protease	p38
correlation coefficient	-0.706	0.638

pK_d の実験値と結合自由エネルギーの計算値との相関係数を、Fig. 2 には pK_d の実験値と結合自由エネルギーの計算値の分布図を記載した。結合自由エネルギーの計算値については、順位 1 位のポーズに対して得られた値を使用している。ここで、 pK_d は解離定数の対数の負値であるため、大きいほど結合が強いことになる。一方で結合自由エネルギーは小さいほど強い結合であるため、相関係数としては負値であることが望ましい。Table 2 を見ると、HIV-1 プロテアーゼでは相関係数が -0.7 を越えており、非常に良好な結果が得られたと判

Fig. 2. Experimentally observed pK_d vs calculated binding free energy.

断できる。また, Wangらによる HIV-1 プロテアーゼに対する市販ソフトウェアを用いた検討¹¹⁾では最大でも相関係数が 0.555 であったことを考慮すると, テストセットの構成が異なるため単純には比較できないものの, HIV-1 プロテアーゼに関しては ArgusLab は市販ソフトウェアと遜色ない結果が得られるのではないかと考えられる。一方で p38 では相関係数が正の値になっており, この系に対しては結合自由エネルギーの計算に AScore を使用するの望ましくないことが示唆される。これについては, HIV-1 プロテアーゼでは複雑なリガンド分子でもドッキングに成功し, キナーゼでは配座自由度の低いリガンドでしかドッキングに成功しなかったという (i) の結果と関連している可能性が考えられる。また, すべての計算結果をまとめて示した Fig. 2 を見ると, HIV-1 プロテアーゼの系のほうがキナーゼの系よりも実験的に得られた結合親和性も高いことがわかる。これが (i) および (iii) での結果につながっている可能性もあり, それについては現在検討中である。なお, CDK-2 についてはサンプル数が少ないため相関係数に関する検討を行わなかったが, サンプル数を追加して検討する予定である。

ここまで述べた結果から, ArgusLab によるドッキングは系によって妥当な解を与える場合と与えない場合があることがわかる。他のドッキングプログラムやスコア関数と比較して特に劣るといわけではないものの,¹⁰⁾ 十分な注意が必要であろう。この点は ArgusLab に限らず計算機によるドッキングに常に付随する問題であるため, CADD を行う際には実験と計算を有機的に組み合わせるなど, 実験と計算両方の長所を利用した運用が必要であろうと考えている。

本研究では, ArgusLab をそのままドッキングお

よびヴァーチャルスクリーニングに使用した場合の有用性について検討した。しかしながら前述のように ArgusLab には調整すべきパラメータが複数あり, これらを適切に設定することでさらに良好な結果が得られる可能性がある。今後, これらの設定を変更した場合の影響について検討する予定である。

REFERENCES

- 1) Kitchen D. B., Decornez H., Furr J. R., Bajorath J., *Nat. Rev. Drug Discovery*, **3**, 935-949 (2004).
- 2) <<http://www.planaria-software.com/>>, ArgusLab 4.0.1, Planaria Software LLC, Seattle, WA, 2004.
- 3) Wang R., Lai L., Wang S., *J. Comput.-Aided Mol. Des.*, **16**, 11-26 (2002).
- 4) Oda A., Hirono S., *Kinozairyo*, **25**, 21-28 (2005).
- 5) Oda A., Tsuchida K., Takakura T., Yamaotsu N., Hirono S., *J. Chem. Inf. Model.*, **46**, 380-391 (2006).
- 6) Oda A., Yamaotsu N., Hirono S., *SAR News*, **11**, 13-17 (2006).
- 7) Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E., *Nucleic Acids Res.*, **28**, 235-242 (2000).
- 8) Wang R., Fang X., Lu Y., Wang S., *J. Med. Chem.*, **47**, 2977-2980 (2004).
- 9) Wang R., Fang X., Wang S., <<http://sw16.im.med.umich.edu/databases/pdbbind/index.jsp>>, PDBBind Database.
- 10) Wang R., Lu Y., Wang S., *J. Med. Chem.*, **46**, 2287-2303 (2003).
- 11) Wang R., Lu Y., Fang X., Wang S., *J. Chem. Inf. Comput. Sci.*, **44**, 2114-2125 (2004).