

## ArgusLab を使用したタンパク質–リガンド複合体の結合自由エネルギー予測

小田 彰史, 内山 孝夫, 大城 信雄, 山崎 尚希, 高橋 央宜, 松崎 久夫

### Prediction of Protein-Ligand Binding Free Energy by Using ArgusLab

Akifumi ODA, Takao UCHIYAMA, Nobuo OHKI, Naoki YAMAZAKI, Ohgi TAKAHASHI, and Hisao MATSUZAKI

(Received November 21, 2007)

In this study, we evaluated the accuracy of predicted binding free energies between proteins and ligands calculated by using software ArgusLab, which is one of freely available computational docking programs both for commercial and academic users. For the evaluation, experimentally determined protein-ligand binding affinities were used. The complexes include one of three proteins, adenosine deaminase, carboxypeptidase A and cathepsin K, as target proteins. Because these proteins are well-known drug targets, the tests can be suitable for practical use. Test sets were composed of 33 complexes. The results indicate that ArgusLab can predict binding free energies for all three types of drug targets, quantitatively. 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, Binding Free Energy, Virtual Screening

薬物標的となる生体高分子の立体構造を元にドラッグデザインを行う Structure-Based Drug Design (SBDD) では、標的高分子と低分子化合物の複合体構造を予測し、その結合自由エネルギーを正確に見積もる必要がある。<sup>1)</sup> この見積もられた結合自由エネルギーの低い複合体ほど強固な結合を有していることになり、より強い薬物となりうる。従って、通常コンピュータで仮想的にスクリーニングを行うバーチャルスクリーニングにおいては、結合自由エネルギーの予測値の低い化合物を医薬候補化合物として選別する。そのため結合自由エネルギーの予測法には高い精度が要求されることになる。

これまで、生体高分子とリガンドの複合体を仮想的にドッキングし、バーチャルスクリーニングを行うためのプログラムは多数開発されている。これらの多くは有償で提供されており、多大な計算資源を要するものが多い。また、多くは UNIX 環境下で使用するソフトウェアが多く、実験と計算の両面が重要な役割を果たすバ

ーチャルスクリーニングにおいては、計算操作に熟練していない研究者にとって導入しにくいものとなっている。そのため、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、無料でドッキングを行うことのできるソフトウェアが Computer-Aided Drug Design (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 ArgusLab<sup>2)</sup> 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 XScore<sup>3)</sup> 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。また、バーチャルスクリーニングにおいては高速にドッキング計算が実行できる必要があるが、Windows 環境下で使用するソフトウェアでありながらドッキングの高速性も確保している。<sup>4)</sup> このように ArugsLab 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いくつかの文献<sup>4-6)</sup> および web 上<sup>2)</sup> に少数存在するだけに留まっている。

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。<sup>7-9)</sup> 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。すなわち、タンパク質ーリガンド複合体の構造候補を多数作成する段階である。複合体構造候補中のリガンドの座標 (配座・配向・位置のすべてを含めた情報) をポーズと呼ぶが、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。そしてそのうちの最低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<sup>10)</sup> が使用され、ポーズが順位付けされる。この段階では、配座・配向・位置は異なるものの、同じ低分子化合物の異なるポーズについて優先度を評価することになる。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バーチャルスクリーニングと呼ぶ。この段階ではほとんどの場合各化合物について単一のポーズのみを考慮し、異なる化合物間で優先順位をつける。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。

これまで我々は、ArgusLab のドッキング機能についていくつかの評価を行ってきた。内容としては、1型ヒト免疫不全ウィルスプロテアーゼ、p38 マイトジェン活性化プロテインキナーゼ、2型サイクリン依存性キナーゼといった薬物標的としてよく知られたタンパク質に対するリガンドドッキングの精度の評価や<sup>5)</sup>、ドッキングに用いるパラメータおよび設定を変更して、その影響を評価する<sup>4)</sup> といった、ドッキングの最初の2つの段階に主眼をおいた評価が主である。一方でバーチャルスクリーニングについての評価については少数のセットしか用意できず、実験値との相関係数もわずか10個のテストセットからしか求めていない。<sup>5)</sup>

そこで本研究では、ArgusLab のドッキング機能をバーチャルスクリーニングに使用した場合の、タンパク質ーリガンド間の結合自由エネルギー計算の精度を評価する。このテストには立体構造既知でかつ結合親和性についても実験値の得られている複合体を使用する。対象とする薬物標的はアデノシンデアミナーゼ (ADA)、カルボキシペプチダーゼ A (CPA)、カテプシン K (CK) の3種類である。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

ADA、CPA および CK の立体構造は RCSB Protein DataBank (PDB)<sup>10)</sup> から入手した。使用した複合体の PDB ID は Table 1 に記した。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

Table 1. Test sets.

ADA		CPA		CK	
PDB ID	pK <sub>a</sub>	PDB ID	pK <sub>a</sub>	PDB ID	pK <sub>a</sub>
1ADD	6.745	<b>1CBX</b>	6.347	<b>1AYU</b>	9.155
1FKW	5.046	<b>1CPS</b>	6.658	1BGO	8.456
1NDV	5.921	1F57	5.638	<b>1NL6</b>	11.319
1NDW	5.229	1HDU	5.824	<b>1NLJ</b>	9.796
1NDY	6.167	1IY7	6.194	1Q6K	7.585
<b>1NDZ</b>	8.114	2CTC	3.886	<b>1SNK</b>	9.602
<b>1O5R</b>	8.125	<b>6CPA</b>	11.523	1TU6	9.081
<b>1QXL</b>	7.959	<b>7CPA</b>	13.959	1U9V	8.222
1UML	7.523	<b>8CPA</b>	9.149	1U9X	8.155
1V7A	7.886			1YK7	7.721
1V79	7.886			<b>1YT7</b>	9.886
<b>1WXZ</b>	8.066				
<b>2E1W</b>	7.959				

ジン, パラメータともデフォルトの値を使用した. また, バーチャルスクリーニング等の高速計算における操作を想定して, ドッキングの際には水分子は一切考慮せずに行った. 計算の際には, PDB より得た複合体構造よりリガンドの構造を抜き出し, 抜き出したリガンドを抜き出された標的分子に対してドッキングさせる, いわゆる再ドッキング問題としてテストを行った. ポーズについては, 通常バーチャルスクリーニングでは最上位のポーズのみを考慮することから, 結合自由エネルギーが最も低かったポーズを採用し, この結合自由エネルギーをリガンドの予測結合自由エネルギー値として採用した. 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, 計算で予測した結合親和性と実験値とを比較することによって検討した. 結合親和性の実験値のデータについては PDB Bind<sup>11,12)</sup> から取得した pK<sub>a</sub> 値を使用した. この値についても Table 1 に記載した. 実験値との比較法としては, 相関係数をとることによって行った. また, 計算によって結合親和性が高いと評価された化合物が, 実際に他の化合物と比較して結合親和性の実験値が高かったかどうか, 存在比をとって比較も行った. 具

体的には, 結合親和性の実験値が上位 5 位以内の化合物が, 計算によって上位 5 位以内と認識された化合物の中にいくつ入っているかを数え上げた. 実験的に得られた結合親和性が上位 5 位以内であった化合物は Table 1 に太字で示した.

### 結果および考察

まず, Table 2 にドッキングによって得られたポーズのうち, 第 1 位だったものの結合自由エネルギーの計算値を示した. この表では, 実験的に得られた結合親和性が上位 5 位以内である複合体の PDB ID を太字にしており, 計算による予測値が上位 5 位以内であった複合体については自由エネルギーの値を太字で記している. この表に記したとおり, CPA の 1IY7 については結合自由エネルギーとして非常に絶対値の大きな正の値が与えられている. 他の値と比較して, これは正しく自由エネルギー計算が実行できたとは言い難い値であり, 以後の考察ではこの値を除外して行う. 実際, 開発者自身が ArgusLab のドッキングエンジンである ArgusDock にバグが残されている可能性について言及しており<sup>2)</sup>, これはその一つである可能性が高い. 我々がこれまでに行ってきた研究<sup>4,5)</sup>と本研究を通じて結合自由エネルギー計算において異常値が出た

のはこの 1IY7 のみであるため、バグの起こる条件等を特定することは現段階では困難である。今後の研究で同様のバグが現れた場合に、さらなる検討を行う予定である。

Table 2 に PDB ID を太字で示した実験値が上位 5 位以内だった化合物と、数値を太字で示した計算による結合自由エネルギーが上位 5 位以内であった化合物とを比較すると、ADA では計算値上位 5 位以内に実験値上位 5 位以内が 2 化合物、CPA では 5 化合物すべて、CK では 3 化合物含まれていた。これはそれぞれ 40 %, 100 %, 60 % の確率である。バーチャルスクリーニングの効率については、濃縮率 (enrichment factor<sup>13)</sup>) という値を使用する。濃縮率は以下の式で得られる。

$$E_f = \frac{N_{\text{experimental}}}{N_{\text{pickup}}} \bigg/ \frac{N_{\text{active}}}{N_{\text{all}}} \quad (1)$$

ここで、 $E_f$  が濃縮率、 $N_{\text{experimental}}$  は計算によりピックアップした中 (本研究では計算での上位 5 化合物中) での「強活性化化合物」の数、 $N_{\text{pickup}}$  はピックアップした系の数 (本研究では 5)、 $N_{\text{active}}$  は系全体の中での「強活性化化合物」の数

(本研究では 5)、 $N_{\text{all}}$  は系全体の数 (ADA では 13, CPA では 9, CK では 11) である。これを計算すると、ADA, CPA, CK でそれぞれ 1.04, 1.8, 1.32 にあたる。今回の研究では結合自由エネルギー予測を調査しているため、濃縮率を計算できるほどのデータ数はないものの、それでも濃縮率がすべて 1.0 を越える結果となっている。これはバーチャルスクリーニングによって探索効率を改善できることを意味しており、これらの系に対して ArgusLab が有効に機能する可能性を示唆している。特に CPA では結合親和性の高かった化合物すべてを計算によって選別できており、ArgusLab によるバーチャルスクリーニングが有効に利用できる系であることがわかる。また、ADA や CK で上位 5 位以内に計算値が出なかった化合物においても、おおむね上位に位置していた。実際のドラッグデザインにおいては「上位  $x$  位以内の化合物」と機械的に決めるわけではなく、実験者の目で有望と思われる化合物を上位から順に見ていくのが一般的であるため、親和性の高い化合物が上位にあることで成功率が上昇するのではないかと期待できる。

Table 2. Calculated binding free energies.

ADA		CPA		CK	
PDB ID	$\Delta G/\text{kcal mol}^{-1}$	PDB ID	$\Delta G/\text{kcal mol}^{-1}$	PDB ID	$\Delta G/\text{kcal mol}^{-1}$
1ADD	-7.47	<b>1CBX</b>	<b>-12.19</b>	<b>1AYU</b>	<b>-9.96</b>
1FKW	-6.92	<b>1CPS</b>	<b>-10.71</b>	1BGO	<b>-10.64</b>
1NDV	-9.84	1F57	-8.62	<b>1NL6</b>	-9.52
1NDW	-10.04	1HDU	-10.07	<b>1NLJ</b>	<b>-9.84</b>
1NDY	<b>-11.93</b>	1IY7	208.79	1Q6K	-7.87
<b>1NDZ</b>	<b>-11.93</b>	2CTC	-10.02	<b>1SNK</b>	<b>-10.37</b>
<b>1O5R</b>	-11.22	<b>6CPA</b>	<b>-13.84</b>	1TU6	-9.79
<b>1QXL</b>	<b>-12.44</b>	<b>7CPA</b>	<b>-15.74</b>	1U9V	-8.64
1UML	<b>-11.86</b>	<b>8CPA</b>	<b>-12.77</b>	1U9X	-7.92
1V7A	<b>-11.69</b>			1YK7	-9.74
1V79	-10.49			<b>1YT7</b>	<b>-10.86</b>
<b>1WXZ</b>	-11.67				
<b>2E1W</b>	-11.42				

Table 3. Correlation coefficients between experimentally obtained  $\text{pK}_a$  values and computational binding free energies.

	ADA	CPA	CK
correlation coefficient	-0.664	-0.920	-0.490

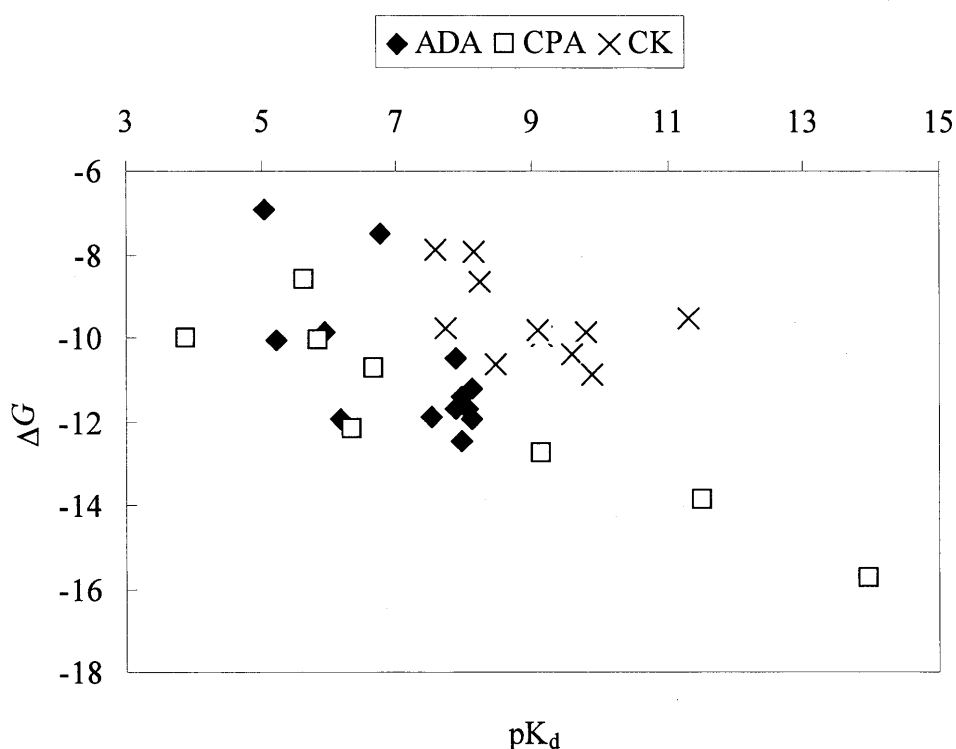


Fig. 1. Calculated binding free energies ( $\Delta G$ ) vs experimentally observed  $pK_a$  values.

Table 3 に、結合親和性の実験値と予測された結合自由エネルギーとの相関係数を記した。ここでは標的分子ごとに結果をまとめている。また、Fig. 1 に  $pK_a$  の実験値と結合自由エネルギーの計算値との分布図を示した。ここでは、 $pK_a$  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を意味している。一方で結合自由エネルギーは低いほど安定であるため、 $pK_a$  と結合自由エネルギーとの間には負の相関があることになる。すなわち、相関係数としては負値でかつ絶対値が大きい値であることが望ましい。Table 3 によると、ADA、CPA、CK の 3 つの標的すべてに対して負の相関係数が得られている。これは、今回扱った系においては結合自由エネルギー予測が実験値をうまく再現するように実行できたことを意味している。特に、CPA については絶対値が 0.9 を越えており、非常によく結合自由エネルギーを再現している。CPA のリガンドについては比較的活性部位内に収まっているリガンドが多く、そのため結合自由エネルギー計算で妥当な値が出ているのでは

ないかとも考えられる。それに対して CK では比較的相関係数の絶対値が低い値となっているが、この理由としてはシステインとの間の共有結合性の相互作用を正しく評価できていない可能性があることや、1BGO などのようにリガンドが巨大なために活性部位から外に出てしまい、それが正しく評価できていない可能性などが考えられる。しかしながら、CK においても相関係数の絶対値は 0.5 に近い値となっており、そのような評価の難しいと思われる系に対しても ArgusLab が定性的には妥当な結合自由エネルギーを出していることがわかる。実際、Wang らによる 3 種類の標的分子を用いた検討<sup>14)</sup> では相関係数が最大で 0.823、最低で 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、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。

ここまで述べた結果から、ArgusLab による結合自由エネルギー計算は少なくとも定性的には実験的に得られた結合親和性を再現すること

がわかる。従って、ある程度の注意は必要であるものの、ArgusLabはバーチャルスクリーニングにおいて有用なツールとなりうるのではないかと考えられる。操作性がよいというArgusLabの利点を生かして、実験と計算を有機的に組み合わせることで実験結果を計算にフィードバックするなど、実験と計算両方の長所を生かして利用することが期待できる。

本研究では、ArgusLabをそのままバ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。以前の研究<sup>4)</sup>では、ArgusLabの種々の設定を変更した場合のポーズ構築・ポーズ選択に対する影響を検討した。そこで今後、これら設定の変更が結合自由エネルギー計算に対して与える影響についても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., Okayasu M., Kamiyama Y., Yoshida T., Takahashi O., Matsuzaki H., *Bull. Chem. Soc. Jpn.*, **80**, 1920-1925 (2007).
- 5) 小田彰史, 吉田朋起, 岡安 愛, 神山由紀子, 高橋央宜, 松崎久夫, *東北薬科大学研究誌*, **53**, 93-97 (2006).
- 6) Joy S., Nair P. S., Hariharan R., Pillai M. R., *In Silico Biol.*, **6**, 0053 (2006).
- 7) Oda A., Hirono S., *Kinzoairyo*, **25**, 21-28 (2005).
- 8) Oda A., Tsuchida K., Takakura T., Yamaotsu N., Hirono S., *J. Chem. Inf. Model.*, **46**, 380-391 (2006).
- 9) Oda A., Yamaotsu N., Hirono S., *SAR News*, **11**, 13-17 (2006).
- 10) 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).
- 11) Wang R., Fang X., Lu Y., Wang S., *J. Med. Chem.*, **47**, 2977-2980 (2004).
- 12) Wang R., Fang X., Wang S., <<http://sw16.im.med.umich.edu/databases/pdbbind/index.jsp>>, PDBBind Database.
- 13) Bender A. Glen R. C., *J. Chem. Inf. Model.*, **45**, 1369-1375 (2005).
- 14) Wang R., Lu Y., Fang X., Wang S., *J. Chem. Inf. Comput. Sci.*, **44**, 2114-2125 (2004).