

既知の活性/非活性化合物のドッキング解析による バーチャルスクリーニングに適した タンパク質立体構造モデルの選択

和久井 直樹^{1,3,a)} 大上 雅史^{2,b)} 千葉 峻太郎^{3,c)} 石田 貴士^{2,3,d)} 岩崎 博史^{1,3,e)} 秋山 泰^{2,3,f)}

概要: Structure-based drug design (SBDD) において、ターゲットとなるタンパク質の立体構造データはスクリーニングの成否に大きな影響を及ぼす。しかし、既に構造解析されたタンパク質の立体構造は限られており、結晶構造は必ずしも生体内の構造を反映していないなどの問題点がある。本研究では、NMRを用いて解析された構造と X 線結晶構造解析を用いて解析された構造、さらに結晶構造を初期構造として分子動力学計算を行った際の構造モデルを用意し、既知活性化合物と非活性化合物に対してドッキングシミュレーションを行った。得られたドッキングスコアやドッキングスコアを分子量で正規化した値 (ポテンシー) によって活性化合物と非活性化合物の分離能を Receiver Operating Characteristics (ROC) 曲線で評価し、どのような構造モデルの場合に活性/非活性を識別することができるのかを考察した。最終的には NMR、X 線、あるいは構造予測技術で作成された初期構造から、分子動力学法等により多数の構造モデルを生成し、バーチャルスクリーニングに適したモデルを自動的に選択するシステムの構築を目指している。

1. 序論

創薬の上流プロセスにおいて、標的タンパク質の機能を阻害や活性化させる化合物 (ヒット/リード化合物) を見つけ出す必要がある。生物学的な実験のみに頼る場合、時間と費用の面で大きなコストがかかる。そのため、計算機を用いて化合物のライブラリから標的タンパク質に強く結合しそうな化合物を絞り込み、実験のヒット率の向上を試みる *in silico* screening が近年では盛んに行われている。*in silico* screening には大きく分けて 2 つのアプローチが存在している。1 つは structure-based drug design (SBDD) であり、もう 1 つは ligand-based drug design (LBDD) である。本研究では既知阻害剤とは異なる新しい母核を持つリード化合物の探索ができる等の理由から、標的の三次元構造に基づくドッキングシミュレーションのアプローチをとる SBDD に焦点を当てた。

SBDD では Protein Data Bank (PDB)[1] に登録されて

いる構造を用いて計算が行われることが多く、タンパク質の立体構造はスクリーニングの成否に大きな影響を及ぼす。しかし PDB に登録されている構造の約 9 割は X 線結晶構造解析法を用いて解析された結晶構造であり、結晶構造は必ずしも生体内での構造を反映していないなどの問題点がある。

一方、千葉・安尾ら [2] は、核磁気共鳴 (nuclear magnetic resonance, NMR) で解析された構造モデルのうち、既知活性化合物と非活性化合物を最もよく識別するモデルを選択する方法を検討しているが、NMR で構造解析できるタンパク質の大きさには技術的な限界が存在し、すべてのタンパク質を NMR で解析できるわけではない。そこで結晶構造に対して生体内での柔軟性を考慮するために分子動力学 (molecular dynamics, MD) 計算を用いて多数の構造アンサンブルを用意するという方法が考えられる。しかしながら、多数の構造アンサンブルの中からどの構造をスクリーニングに用いるとよいのかという指標は現在わずかしか存在していない [5], [6]。そこで本研究ではモデル選択に関わる指標の提案を目的とし、本稿ではその準備段階としてこれまでに得られた知見について報告を行う。具体的には、標的のタンパク質に対して、NMR を用いて得られた構造、X 線結晶構造解析で得られた構造、それを初期構造として MD 計算を実施して得た構造などを用意し、ドッキングソフトウェア Glide (Schrödinger LLC)[7],[8] を適用した計算

¹ 東京工業大学 大学院生命理工学研究科

² 東京工業大学 大学院情報理工学研究科

³ 東京工業大学 情報生命博士教育院

a) nwakui@bio.titech.ac.jp

b) ohue@bi.cs.titech.ac.jp

c) chiba@acls.titech.ac.jp

d) ishida@cs.titech.ac.jp

e) hiwasaki@bio.titech.ac.jp

f) akiyama@cs.titech.ac.jp

結果からどのような構造モデルの時に既知活性化合物と非活性化合物を識別できるのかを報告する。

2. 方法

本研究の大まかな手順を図 1 に示す。以下にそれぞれのステップを詳しく記す。

本研究では提案手法の有効性を示すために今後さまざまなタンパク質に対して評価実験を進めていく予定であるが、本稿では、共著者の一部(千葉ら)が近年取り組んでいるデングウイルス (Dengue virus) のプロテアーゼ阻害薬の探索を例として説明する。

2.1 ターゲットタンパク質構造の準備

デングウイルスの成熟にはポリタンパク質の分解が必須であり、この必須の過程にプロテアーゼが関与していることから抗デングウイルス薬開発のターゲットタンパク質の候補としてデングウイルスプロテアーゼが考えられている [10], [11]。プロテアーゼは non-structural protein 2B (NS2B) と non-structural protein 3 (NS3) から構成されており、現在までに 15 種類の構造が解かれている [10], [11], [12], [13], [14], [15], [16], [17]。本稿では、Noble らによって X 線結晶構造解析が行われた NS2B-NS3 の構造 (PDB ID: 3U1I)[16] と Gibbs らによって NMR で構造解析が行われた NS3 の構造 (PDB ID: 2M9P)[17] を用いた。千葉らによって 3U1I を初期構造として行われた 140 ns の MD 計算結果 [3], [4] の 0–100 ns のトラジェクトリーのうち、5 ns ごとに 20 個の構造をサンプリングした。これらの構造に対して、Maestro (Schrödinger LLC) の Protein Preparation module を用いて構造の最適化を行った。続いて SiteMap module を用いて結合サイトの検索を行い、Receptor Grid Generation module を用いてドッキングに用いるグリッドの作成を行った。

2.2 既知活性化合物、非活性化合物の準備

アッセイなどにより既に活性があることがわかっている既知活性化合物を 32 種類用意した。また、既知活性化合物と同様の母核を持つもののアッセイでは活性が見られなかった非活性化合物を 3 種類用意した。これら 35 種類の化合物に対して、LigPrep module を用いてリガンドが取り得る解離状態などの初期構造を作成した。

2.3 ドッキングシミュレーション

ドッキングシミュレーションは Schrödinger 社の Glide (Maestro の Ligand Docking module) を用いて standard precision (SP)[7],[8] モードで行った。

2.4 モデルの評価

本稿では、各化合物に対するドッキングシミュレーション

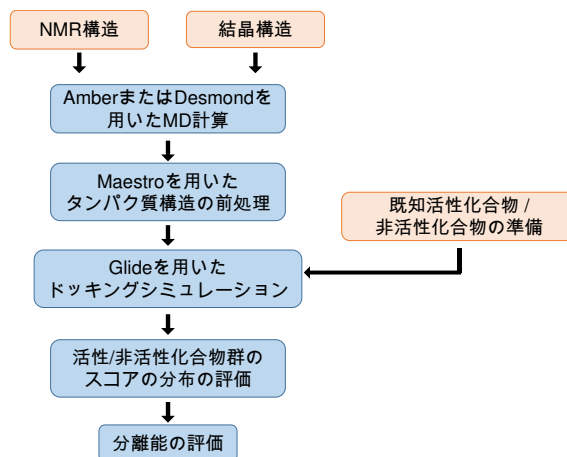


図 1 本研究の手順

ンのスコア (Glide スコア) を各化合物の分子量で割った、分子量当たりの Glide スコアをポテンシー (potency) と呼ぶ。ポテンシーから計算される Area Under the Receiver Operating Characteristics (AUROC) を各モデルの分離能を表す指標とし、各モデルの評価を行った。AUROC は対象の化合物の数を N 、活性化合物の数を n 、活性化合物 i の順位を r_i としたとき、

$$AUROC = 1 - \frac{\sum_{i=1}^n r_i}{n(N-n)} + \frac{n+1}{2(N-n)}$$

で計算される [9]。この AUROC 値が高いほど、活性化合物と非活性化合物の分離が良好であると判断できる。

3. 結果

3.1 NMR の各モデル間での活性/非活性化合物の分離能の比較

2M9P-NMR Model01 から 2M9P-NMR Model10 (以降、単に NMR-01~NMR-10 と呼称する) の Glide スコアを用いた場合とポテンシーを用いた場合のそれぞれの AUROC 値を表 1 に示す。NMR-07 のアロステリックサイトを対象としてドッキングシミュレーションを行った場合にポテンシーを用いた時の AUROC 値が 0.771 と 10 モデル中で最も良いことがわかり、この結果は千葉・安尾ら [2] と一致した。

3.2 結晶構造を初期構造とした MD 計算で得た構造間での活性/非活性化合物の分離能の比較

結晶構造 3U1I を初期構造として MD 計算から得られた 20 個のモデル (以降 MD-01~MD-20 と呼称する) の AUROC 値を表 1 に示す。各モデルをポテンシーを用いた時の AUROC 値と比較すると、MD-14 のアロステリックサイトと MD-17 のアロステリックサイトの AUROC 値が 0.823 と初期構造を含めた 21 個のモデルの中で最も良いことがわかった。MD-14 のアロステリックサイトにおける

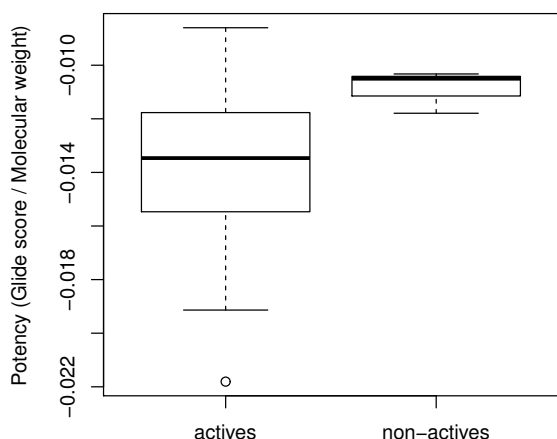


図 2 MD-14 のアロステリックサイトにおけるポテンシーの箱ひげ図。箱の下端が第 1 四分位点、箱の上端が第 3 四分位点、箱の中の太線が中央値、箱の下にある線が最小値、箱の上にある線が最大値を表している。

活性/非活性化合物のポテンシーの分布を図 2 に示す。

4. 考察

表 1 において、本稿で導入した Glide スコアを分子量で補正したポテンシーを用いることで、活性/非活性化合物の分離能を高める傾向が見られ、本研究の目的を達成する上で好ましいのではないかと思慮される。

また、10 個の NMR のモデルの中で AUROC 値が最も良かった NMR-07 のアロステリックサイトに関して、最も値が悪かった NMR-05 のアロステリックサイトと比較すると、NMR-05 のアロステリックサイトでは NS2B ドメインの SER34 から SER38 にかけてのループ (ループ NS2B) と NS3 ドメインの THR179 から GLY182 にかけてのループ (ループ NS3) の距離が 10 Å 程度であるのに対し、NMR-07 では 15 Å 程度と離れていることがわかった (図 3)。アロステリックサイトを構成するこれらのループの距離が離れていることでアロステリックサイトの体積が確保され、分離能に影響しているのではないかと考えられる。同様に、MD-14 と MD-17 においてもループ NS2B とループ NS3 の距離を調べてみたところ、それぞれ 14 Å 程度離れていることがわかった (図 3)。結晶構造を初期構造として MD 計算をしたにもかかわらず、NMR で解析された構造と良く似た部分構造を持つ構造が MD 計算から得られたことは大変興味深い。

5. 結論

本稿では既知活性化合物と非活性化合物を識別できる構造モデルの選択を目的とし、ドッキングシミュレーションの結果に基づいて Deng ウイルスプロテアーゼのモデル選

表 1 デングウイルスプロテアーゼの様々な立体構造モデルに対する活性/非活性化合物群のドッキング成績の比較。Glide スコアを用いた場合と Glide スコアを各化合物の分子量で補正したポテンシーを用いた場合の 2 種類のドッキング成績 (AUROC 値) を以下にまとめた。

モデル	Glide スコアでの AUROC		ポテンシーでの AUROC	
	活性 サイト	アロステリック サイト	活性 サイト	アロステリック サイト
NMR-01	0.167	0.250	0.396	0.510
NMR-02	0.344	0.406	0.531	0.677
NMR-03	0.250	0.094	0.479	0.406
NMR-04	0.073	0.344	0.354	0.448
NMR-05	0.354	0.167	0.656	0.350
NMR-06	0.167	0.417	0.479	0.479
NMR-07	0.125	0.604	0.375	0.771
NMR-08	0.135	-*	0.479	-*
NMR-09	0.344	0.562	0.510	0.625
NMR-10	0.344	0.115	0.583	0.438
X-ray	0.365	0.344	0.562	0.531
MD-01	0.417	0.333	0.500	0.573
MD-02	0.383	0.479	0.483	0.635
MD-03	0.423	0.302	0.462	0.583
MD-04	0.217	0.260	0.350	0.542
MD-05	-*	0.396	-*	0.458
MD-06	0.385	0.448	0.519	0.552
MD-07	0.427	0.208	0.583	0.646
MD-08	0.173	0.354	0.288	0.656
MD-09	0.620	0.365	0.640	0.583
MD-10	0.354	0.615	0.396	0.698
MD-11	-*	0.448	-*	0.635
MD-12	0.217	0.406	0.400	0.573
MD-13	0.115	0.604	0.365	0.656
MD-14	0.104	0.604	0.396	0.823
MD-15	0.385	0.323	0.552	0.635
MD-16	0.271	0.333	0.396	0.604
MD-17	0.510	0.656	0.573	0.823
MD-18	0.135	0.354	0.375	0.625
MD-19	0.717	0.677	0.717	0.708
MD-20	0.312	0.615	0.500	0.729

* 数値のない欄は SiteMap が結合サイトとして認識しなかったため、ドッキング計算結果が存在しない。

択を試みた。Glide スコアを分子量で補正したポテンシーを用いて各モデルに対して AUROC を計算したところ、化合物の識別能が改善する傾向が見られた。今後、より多くの対象タンパク質について吟味する必要がある。

6. 今後の課題

本研究で扱った非活性化合物は活性化合物と構造が良く似ており、微細な構造の違いを識別する目的で準備したものの、数が少なく 3 種類しか準備することができなかった。今後はより多くの非活性化合物を収集する予定である。また、たまたま既知活性化合物と非活性化合物をうまく識別することができる overfitting が起きている構造モデルを選択することを避ける方法を開発する必要がある。これは、まずモデルの識別能の良否を確率などの枠組みで定

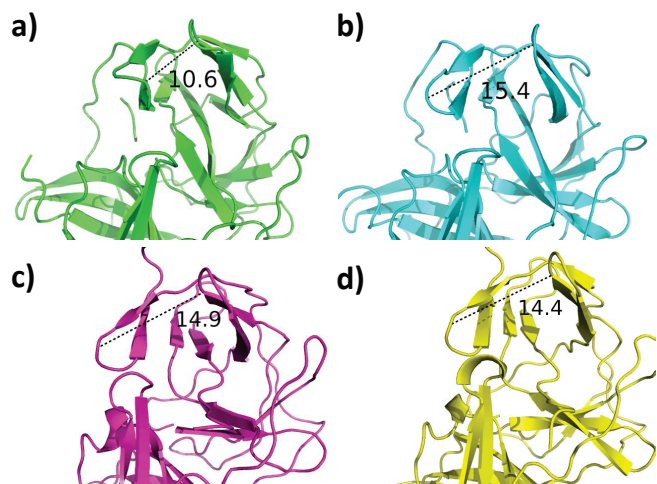


図 3 NS2B ループと NS3 ループの距離 (Å)。a) NMR-05、b) NMR-07、c) MD-14 と d) MD-17 における NS2B ループと NS3 ループの距離を示した。

量的に評価できるかが問題となる。また、全くの偶然でも既存データに対して良く説明できるモデルが得られることがあるが、これはいわゆる多重検定の問題になるため、適切な補正をどのようにするべきかの検討も必要となる。今後、さらにこれらの点についても検討を進めていく予定である。

謝辞 本研究を遂行するにあたり、デングウイルスプロテアーゼに関する助言をいただき、2.2 節に記載した化合物データの利用を許可していただいた東京工業大学 学術国際情報センター 関嶋 政和准教授と、先行研究 [2] の内容について助言をいただいた東京工業大学 大学院情報理工学研究科 計算工学専攻 安尾 信明氏に深く感謝する。本研究の一部は、日本学術振興会 科研費 基盤研究 (A) (24240044) の支援によって行われた。

参考文献

- [1] Berman, HM, *et al.* “The protein data bank.” *Nucleic Acids Res* 28.1 (2000): 235–242. <http://www.rcsb.org/pdb/>.
- [2] 千葉峻太郎, 安尾信明, 東京工業大学, personal communications.
- [3] Chiba, S, *et al.* “分子動力学シミュレーションで探る阻害剤結合によって誘起される dengue virus プロテアーゼの特徴的な構造.” 熱帯医学会, P2-20, 2013.
- [4] Chiba, S, *et al.* “Discovery of dengue virus protease inhibitors and their inhibition mechanism through docking simulation.” 熱帯医学会, P-05-07, 2014.
- [5] Sato, M, *et al.* “Extended Template-Based Modeling and Evaluation Method Using Consensus of Binding Mode of GPCRs for Virtual Screening.” *J Chem Inf Model* 54.11 (2014): 3153–3161.
- [6] Rataj, K, *et al.* “Impact of Template Choice on Homology Model Efficiency in Virtual Screening.” *J Chem Inf Model* 54.6 (2014): 1661–1668.
- [7] Friesner, RA, *et al.* “Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of

- docking accuracy.” *J Med Chem* 47.7 (2004): 1739–1749.
- [8] Halgren, TA, *et al.* “Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening.” *J Med Chem* 47.7 (2004): 1750–1759.
- [9] Zhao, W, *et al.* “A statistical framework to evaluate virtual screening.” *BMC Bioinform* 10.1 (2009): 225.
- [10] Luo, D, *et al.* “Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein.” *EMBO J* 27.23 (2008): 3209–3219.
- [11] Luo, D, *et al.* “Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications.” *J Biol Chem* 285.24 (2010): 18817–18827.
- [12] Luo, D, *et al.* “Crystal structure of the NS3 protease-helicase from dengue virus.” *J Virol* 82.1 (2008): 173–183.
- [13] Yildiz, M, *et al.* “Allosteric inhibition of the NS2B-NS3 protease from dengue virus.” *ACS Chem Biol* 8.12 (2013): 2744–2752.
- [14] Chandramouli, S, *et al.* “Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family.” *J Virol* 84.6 (2010): 3059–3067.
- [15] Erbel, P, *et al.* “Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus.” *Nat Struct Mol Biol* 13.4 (2006): 372–373.
- [16] Noble, CG, *et al.* “Ligand-bound structures of the dengue virus protease reveal the active conformation.” *J Virol* 86.1 (2012): 438–446.
- [17] Gibbs, A, *et al.* “NMR structure of an inhibitor bound dengue NS3 protease provides new insights into the NS2B NS3 ligand interactions.” To be published.