

タンパク質リガンド相互作用実験データのインシリコにおける再現

市川貴章*, 梅山秀明**, 岩館満雄**

In-Silico screening based on the protein ligand interaction of experimental data

Takaaki ICHIKAWA*, Hideaki UMEYAMA**, Mitsuo IWADATE**

Abstract

Prediction accuracy of ChooseLD which was a technique of the in In-Silico screening based on a three-dimensional structure was significantly improved by considering hydrophobic interaction.

Specifically, the hydrophobic interaction evaluation function Hc index which optimized the coefficient was included. It followed that an effect of the hydrophobic interaction was shown to be remarkable as a physical parameter. The effectiveness of the fusion of an empirical parameter and the physical parameter are shown simultaneously.

1. 緒言

近年、より多くのターゲットタンパク質が生化学実験によって決定される[1] [2]. その結果、製薬会社、研究機関の間でターゲットタンパク質を阻害するような化合物の探索競争が激化している[3] [4]. 実験による化合物のスクリーニングによって、ターゲットタンパク質を阻害する化合物が数多く見つかった。しかし、実験によるスクリーニングは非常に費用がかかるため、コンピューターを利用したインシリコスクリーニングがのぞましい。

インシリコスクリーニングにおけるターゲットタンパク質の立体構造に関しては、X線結晶構造解析、NMR等の実験構造が多数登録されている Protein Data Bank (PDB) [5]に登録されている場合、その構造を用いるか、エネルギー極小化、分子動力学計算により複数の水溶液構造の構築最適化等が行われたあと、その構造をターゲットタンパク質の原子座標として用いる。一方、立体構造が登録されていない場合は、その立体構造をホモロジーモデリング法[6]に代表されるタンパク質立体構造予測法で立体構造を予測する必要がある。いずれの場合も実験によるタンパク質立体構造解析により得られたデータを用いている。

また、近年 PDB へ登録されるタンパク質リガンド複合体構造も増加しており、一つのファミリータンパク質内において、異なる原子のリガンドを含んだ複数のX線構造が存在することも多く[1] [2], これらの情報を用いてドッキングソフトで予測された複合体構造がそれらの情報を満たしているかを評価する試みが多数行われている。

これらは既存のドッキングソフトを用いて仮想化合物ライブラリーから、ターゲットタンパク質に阻害剤候補化合物をドッキングし、タンパク質リガンド複合体の構造を予測した後、ターゲットタンパク質の阻害活性試験で Hit Compound をより多く選択するために、既知のタンパク質リガンド複合体の構造からタンパク質とリガンド間の距離、古典物理学的エネルギーの計算などを行い、相互

* 中央大学大学院理工学研究科生命科学専攻 〒112-8551 東京都文京区春日1-13-27

** 中央大学理工学部生命科学科 〒112-8551 東京都文京区春日1-13-27

作用情報を抽出し再評価を行うといったものである[7] [8].

今回用いた ChooseLD (CHOOse biological information Semi-Empilically on the Ligand Doking) 法[9]では、PDB に登録されている相互作用既知のタンパク質ーリガンド複合体の生物化学的な情報から効率的に有効な情報を選び出しドッキングを行い、かつ、Hit Compound を多数検出できるが、ホモロジーモデリング法に代表されるタンパク質立体構造予測法で構造を得た場合、リガンド結合部位に変異の入ったターゲットタンパク質やターゲットタンパク質と立体構造の類似したタンパク質を参照した際に、疎水性相互作用の観点からみて、通常高く評価してはいけないドッキング結果を高く評価することが考えられる。

本研究では、ChooseLD のドッキングにおけるタンパク質ーリガンド複合体相互作用の評価関数 FPAScore に梅山秀明、赤羽健司らによって考案された疎水性相互作用評価関数 Hc (hydrophobic correlation) index [10] を導入することで精度向上を図った。

さらに、近年創薬研究にとって有用な大規模データベースの公開も進んでいる。その中の一つ ChEMBL [11] は、医薬品及び開発化合物の Structure Activity Relationship (SAR) 情報のデータベースである。化学及び生物学の専門家によって収集・整理された質の高い SAR 情報が無償で提供されている。この ChEMBL に登録されているターゲットの Ki の活性情報と相関をとることで、疎水性相互作用評価関数 Hc index を含めた ChooseLD 法が、オリジナルの ChooseLD 法と比較して正確に化合物候補を探索できるかを検討した。

2. 方法

2-1. 概要

ChooseLDはPDBに登録されている相互作用既知のタンパク質—リガンド複合体から生物化学的な情報を抽出し、ドッキングを行うソフトウェアである。相互作用が既知のリガンドから部分的な結合自由エネルギーを保持した部品である fingerprint を元に未知のリガンド構造をドッキングする。

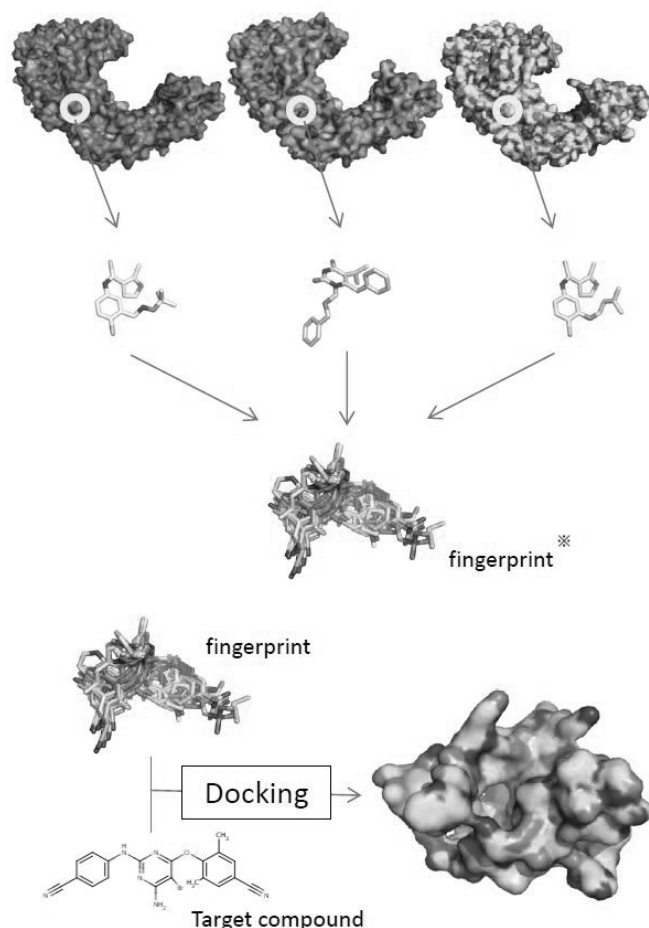


Fig.1 ChooseLD法を使ったドッキングシミュレーションの流れ

ChooseLD法を使ったドッキングシミュレーションの流れの概要を Fig.1 に示す。ChooseLD法ではターゲットタンパク質と立体構造の類似したタンパク質を参照し、それぞれのタンパク質から同じような箇所に結合しているリガンドを選択する。これにより相互作用が既知のリガンドから部分的な結合自由エネルギーを保持した部品である fingerprint を構築する。この fingerprint の情報を元に化合物をドッキングする。つまり、ターゲットタンパク質と似たような局所構造をもつタンパク質のリガンド情報を元にドッキングを行えば、ターゲットタンパク質に作用する医薬品化合物を探索することができる。しかし、その参照する局所構造に変異が入っていた場合、もしくはターゲットタンパク質のリガンド結合部位に変異が入っていた場合、ターゲットタンパク質側のリガンド結合部位が疎水的な結合部位である部分に親水的なリガンドをドッキングすることで、間違ったタンパク質—リガンド複合体構造を高く評価する、またはその逆も同様に行ってしまうことから Hc Index の導入を図った。

2-2. Hf Index

Hc index は元々、タンパク質—化合物間の疎水的、親水的な相互作用の状態を把握するために作られた指標であり、Hf (hydrophobic field-effect) index より求められる。まずは Hf index から順に説明していく。

Hf index は化合物がタンパク質から受ける疎水性、親水性の強さを表す指標である。Hf index を求めるにはまず化合物の表面を細分した微小面(パッチ)にし、パッチ単位で式(1)の計算を行う。

$$Hf_j = \sum_k f_k \cdot \varphi \quad (1)$$

$$\varphi = \exp\left\{\beta \cdot (r_{jk} - R)^2\right\} \quad (2)$$

Hf_j は j 番目のパッチにおける Hf index を意味している。 f_k はタンパク質の k 番目の原子の単位面積あたりの移行自由エネルギーである。ここでの移行自由エネルギーとは、ある分子が水相から気相へ移行する際の自由エネルギーの変化量である。この移行自由エネルギーはその分子の疎水性、親水性と深い関わりがあることが Ben-Naim によって既に述べられている[12]。移行自由エネルギーを数値化した f 値は正の値をとるほど強い親水性を、負の値をとるほど強い疎水性を示し、先行研究によって官能基レベルで既に求められている[10] (Table.1)。

Table.1 タンパク質の官能基に対する f 値

| 置換基 | $f(\text{cal}/(\text{mol} \cdot \text{\AA}^2))$ |
|-------------------------|---|
| Guanidinium | 19.30 |
| -SH | -24.10 |
| -S- | 0.71 |
| Imidazolium | 1.27 |
| Indolyl | -12.56 |
| $-\text{NH}_3^+$ | 45.28 |
| $-\text{C}_6\text{H}_5$ | -12.88 |
| $-\text{CONH}_2$ | 11.30 |
| $-\text{COO}^-$ | 18.63 |
| -OH(aliphatic) | 11.26 |
| -OH(aromatic) | 15.78 |
| Hydrocarbon | -20.87 |
| Backbone amide | 29.34 |

φ は距離依存の関数であり、原子 k による疎水的な影響が距離によって減衰していくことを考慮するためのものである。 φ の求め方は式(2)の通りである。 r_{jk} はパッチ j の面中心とタンパク質原子 k の原子表面との距離である。ここで、 β は定数 -0.1312 であり、これは Pratt-Chandler の理論から2つの分子の間に水1分子を挟んだ準安定状態[13]を再現できるようとした値である。 R はタンパク質中の原子の半径を意味し ($r_{jk}-R$) で化合物とタンパク質の原子表面同士の距離を示す。このように Hf index はタンパク質の各原子の疎水的、親水的な強さに距離による重みをかけて総和することで、そのパッチがタンパク質から受ける影響を表した指数である。

2-3. Hc Index

Hc index は Hf index を用い、式(3)で求めることができる。

$$H_c = \sum_j f_j \cdot Hf_j \cdot S_j \quad (3)$$

S_j はパッチjの面積を表している. このようにHc indexは各パッチがタンパク質から受ける疎水, 親水的な影響(Hf_j)にそのパッチ自身の f_j 値とパッチの面積 S_j をかけたものの総和である. これにより, 化合物とタンパク質全体の疎水的, 親水的な相関を算出することができる. しかし, 化合物の f 値は先行研究によって求められているものの, 実際にインシリコスクリーニングにおいて使用するには, 非常にデータ数が少ない. このことから今回, 化合物のパッチ自身の f 値を求める方法を開発した.

2-4. f-value

Fig.2は水-n-オクタノール系の分配係数 $\log P$ と単位面積あたりの移行自由エネルギー f 値とで相関をとったものである.

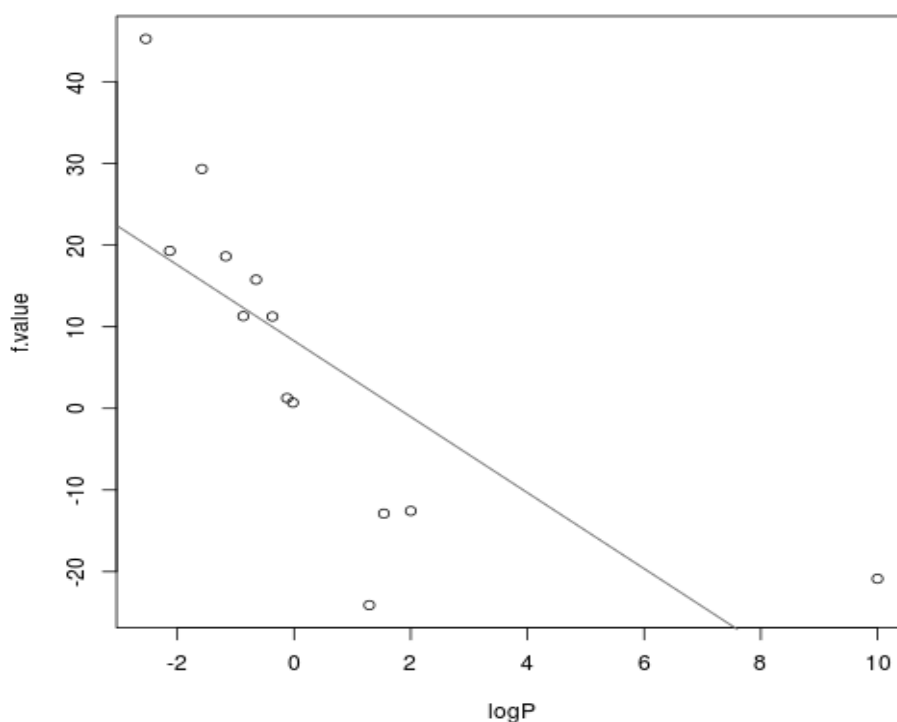


Fig.2 分配係数 $\log P$ 及び単位面積あたりの移行自由エネルギー f 値

$\log P$ と f 値の間に負の相関がみられることから分配係数 $\log P$ に負をとり, 化合物原子ごとの f 値とした.

$$f_j = -\log P \quad (4)$$

また $\log P$ については加算性が成り立つという仮定のもとで, Crippenの方法[14]で求められる $\log P$ を化合物一原子あたりの $\log P$ とした. Crippenの方法では, $\log P$ を求めたい原子の隣接原子のパターンが, あらかじめ用意されたSMART形式のパターンの中でどれに一致するかを判断し, 一致したパターンの $\log P$ を一原子ごとの $\log P$ とするという方法である. 以下に例を示す.

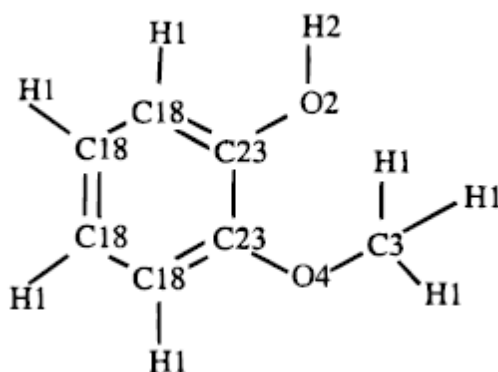


Table.2 化合物の一原子あたりの logP

| Type | SMARTS | logP |
|------|--|---------|
| C3 | '[CH3] [(N, O, P, S, F, Cl, Br, I)]' | -0.2035 |
| C18 | '[cH]' | 0.1581 |
| C23 | '[c] (:a) (:a)-O' | 0.5437 |
| H1 | '[#1] [#6]', '[#1] [#1]' | 0.1230 |
| H2 | '[#1]O[CX4]', '[#1]Oc', '[#1]O! (C, N, O, S)', '[#1] [!C, N, O]' | -0.2677 |
| O2 | '[OH]', '[OH2]' | -0.2893 |
| O4 | '[O] (A)a', '[O] (a)a' | -0.4195 |

上図の化合物において、同じ芳香環を構成する炭素原子であってもヒドロキシ基と結合する炭素原子 (type:C23) の方が、logP が高くなる。このように隣接原子によって一原子あたりの logP を求め、その値に符号をかけることで化合物の f 値として計算を行った。この方法を取り入れ、最終的に求める Hc Index は以下の式(5)となる。

$$H_c = \sum_j -\log P \cdot Hf_j \cdot S_j \quad (5)$$

2-5. 係数 k の基準値の決定

ChooseLD 法により算出される FPAScore と Hc Index では系が異なるため、FPAScore に係数 k をかけた Hc Index を加算したものを、疎水性相互作用評価関数 Hc index を含めた ChooseLD 法で求められる最終的なスコアとした。以下にその式を示す。(式(6))

$$\text{Total Score} = \text{FPAScore} + k \times \text{Hc Index} \quad (6)$$

この係数 k の基準となる値を決定するため、ChEMBL (ChEMBL20) に登録されている全ターゲット配列を対象に、立体構造と FP Library が構築でき、さらに同一 Assay の化合物数が十分にあるターゲットを求める。次にこれらのターゲットに対し、ChooseLD を実行し FPAScore を求め、さらに ChooseLD が出す FPAScore が最も高い構造に対し、Hc Index を求める。最後にこれらの各スコアがとる値の分散を求め、その比を係数 k の基準値とした。FPAScore は 565.5944 ± 415.2728 、Hc Index は 0.3959617 ± 12.76401 となり、分散の幅が大きく異なるので、分散の比に相当する $k=32.5$ を中心に k の値を変えながら、ChEMBL (ChEMBL20) に登録されている実験値 (K_i) との相関をとり、最適な係数 k を求めた。

3. 結果・考察

3-1. 最適な係数kの決定

ChEMBL (ChEMBL20)に登録されている全5,569ターゲットの内, LigandFAMSで構築できたターゲット数は4,348ターゲットであった. さらにこの中で, 計算したスコアと実験値とで相関をとるために必要な Assay の情報が存在し, かつ K_i でデータがあるターゲット数は596ターゲットであった. この内 ChooseLD が正常に実行できたターゲットは273ターゲットとなった. ここで得られた各ターゲットの情報を Assay ID 毎にまとめた. これは同じ化合物でも Assay 毎に実験条件が変わり, それにより K_i などの値も変わってくるためである. 実際に精度が上がったかどうかを確認するため, K_i と元の ChooseLD プログラムにて算出したスコアとの相関, K_i と Hc index を含めた ChooseLD プログラムにて算出したスコアとの相関の2通りを行った. 相関方法に関しては各 Assay 毎にスピアマンの順位相関をとった.

K_i と元の ChooseLD プログラムにて算出したスコアとの相関, K_i と Hc index を含めた ChooseLD プログラムにて算出したスコアとの相関の比較を行うため『順位相関係数 ρ に差がない』という帰無仮説を立て, t 検定を行った. 以下, 各 k における検定結果を Table.3 に示す.

Table.3 各kにおける検定結果

| k | 両側検定 | 片側検定(ρ =less) | 片側検定(ρ =greater) |
|------|--------|---------------------|------------------------|
| 27.5 | 0.5078 | 0.7461 | 0.2539 |
| 28.5 | 0.2288 | 0.8856 | 0.1144 |
| 29.5 | 0.463 | 0.7685 | 0.2315 |
| 30.5 | 0.0285 | 0.9858 | 0.0143 |
| 31.5 | 0.1175 | 0.9413 | 0.0588 |
| 32.5 | 0.2554 | 0.8723 | 0.1277 |
| 33.5 | 0.0893 | 0.9553 | 0.0447 |
| 34.5 | 0.1765 | 0.9117 | 0.0883 |
| 35.5 | 0.1602 | 0.9199 | 0.0801 |
| 36.5 | 0.2673 | 0.8663 | 0.1337 |

Table.3 の検定結果より, 最適な k は30.5である. 元々の ChooseLD での相関係数をオリジナル, 疎水性相互作用評価関数 Hc index を含めた ChooseLD での相関係数を改良後と表し, k=30.5における検

定結果について説明する.

①両側検定

帰無仮説を『オリジナルと改良後に差がない』とし検定を行い, 以下の結果になった.

p-value=0.02849と低い値であり, 帰無仮説を棄却することが妥当であると判断できる. ゆえに, オリジナルと改良後に意に差があると証明された.

②片側検定

片側検定を行いオリジナルと改良後の値の大きさを比較した.

1. 帰無仮説を『オリジナルが改良後より小さくない』とし検定を行い, p-value=0.9858と非常に高い値であり, 0.9858の確率でこの仮定が採択されることが示された. 帰無仮説を『オリジナルが改良後より大きくない』とし検定を行い, p-value=0.01425と非常に低い値であった.

以上の結果より, $k=30.5$ において疎水性相互作用評価関数 Hc index を含めることが ChooseLD の予測精度が向上に最も効果的であると結論した.

4. まとめ

疎水性相互作用を考慮することにより, 立体構造に基づいたインシリコスクリーニングの手法である ChooseLD の予測精度の向上の傾向が認められた. 具体的には係数の最適化を行った疎水性相互作用評価関数 Hc index を含めた. 物理パラメータとして疎水性相互作用の効果が顕著であることが示された結果となった. 経験的パラメータと物理パラメータの融合の有効性も網羅的に示せた.

5. 参考文献

- [1] Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, Dickerson SH, A Unique Structure for Epidermal Growth Factor Receptor Bound to GW572016 (Lapatinib): Relationships among Protein Conformation, Inhibitor Off-Rate, and Receptor Activity in Tumor Cells, *Cancer Res.*, 64, 6652-6659 (2004)
- [2] Stamos J, Sliwkowski MX, Eigenbrot C, Structure of the Epidermal Growth Factor Receptor Kinase Domain Alone and in Complex with a 4-Anilinoquinazoline Inhibitor, *J. Biol. Chem.*, 277, 46265-46272 (2002)
- [3] Nakamura K, Yamamoto A, Kamishohara M, Takahashi K, Taguchi E, Miura T, Kubo K, Shibuya M, Isoe T, KRN633: A selective inhibitor of vascular endothelial growth factor receptor-2 tyrosine kinase that suppresses tumor angiogenesis and growth, *Mol. Cancer Ther.*, 3, 1639-1649 (2004)
- [4] Nakamura K, Taguchi E, Miura T, Yamamoto A, Takahashi K, Bichat F, Guilbaud N, Hasegawa K, Kubo K, Fujiwara Y, Suzuki R, Kubo K, Shibuya M, Isoe T, KRN951, a Highly Potent Inhibitor of Vascular Endothelial Growth Factor Receptor Tyrosine Kinases, Has Antitumor Activities and Affects Functional Vascular Properties, *Cancer Res.* 66, 9134-9142 (2006)
- [5] Umeyama H, Iwadata M, FAMS and FAMSBASE for protein structure, *Curr. Protoc. Bioinformatics*, Chapter 5:Unit5.2. (2004)
- [6] Westbrook J, Feng Z, Chen L, Yang H, Berman HM, The Protein Data Bank and structural genomics, *Nucleic Acids Res.* 31, 489-491 (2003)
- [7] Muralia S, Hojob S, Tsujishitaa H, Nakamurac H, Fukunishi Y, In-silicodrug screening method based on

- the protein compound affinity matrix using the factor selection technique, *Eur. J. of Med. Chem.*, 42, 966-976 (2007)
- [8] Deng Z, Chuaqui C, Singh J, Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions, *J. Med. Chem.* 47, 337-344 (2004)
- [9] Takaya D, Takeda-Shitaka M, Terashi G, Kanou K, Iwadate M, Umeyama H, Bioinformatics based Ligand-Docking and in-silico screening, *Chem. Pharm. Bull.*, 56, 742-744 (2008)
- [10] Kenji Akahane, Yasuo Nagano, and Hideaki Umeyama, Hydrophobic Effect on the Protein-Ligand Interaction; Hydrophobic Field-Effect Index and Hydrophobic Correlation Index, *Chem. Pharm. Bull.*, 37, 86-92 (1989)
- [11] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP, ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, 40, D1100-D1107 (2012)
- [12] Ben-Naim A, Statistical Mechanical Study of Hydrophobic Interaction. III. Generalization and Further Applications, *J. Chem. Phys.*, 57, 5257-5265 (1972)
- [13] Pratt LR, Chandler D, *J. Chem. Phys.*, Theory of the hydrophobic effect, 67,3683-3704 (1977).
- [14] Scott A. Wildman and Gordon M. Crippen, Prediction of Physicochemical Parameters by Atomic Contributions, *J. Chem. Inf. Comput. Sci.*, 39, 868-873 (1999)