

オープンイノベーションコンテストによる IT 創薬

千葉峻太郎^{†1} 池田和由^{†2} 石田貴士^{†1} 大野一樹^{†3,1} 関嶋政和^{†1}

コンピュータを用いた創薬支援技術によるヒット化合物探索は対象とする系（標的分子）によって発揮される性能が異なることが知られている。本研究では一つの標的分子に対して様々な手法を適用し結果としてヒット化合物を得られるかどうかを調べるため、標的分子を阻害すると考えられる化合物の ID を化合物ライブラリから提案する形式の公開のコンテストを実施した。結果として全参加グループから提出された化合物のケミカルスペース上の多様性は単一のグループから提案された化合物群より広く、加えて得られたヒット化合物も多様であったことから、オープンイノベーションコンテストによる化合物探索は有効であることが示唆された。

Open innovation contest for IT drug discovery

SHUNTARO CHIBA^{†1} KAZUYOSHI IKEDA^{†2} TAKASHI ISHIDA^{†1}
KAZUKI OHNO^{†3,1} SEKIJIMA MASAKAZU^{†1}

Many computer-aided drug discovery methods have been developed with a view to identifying potent compounds that inhibit function of a target biomolecule. However, the performance of each method depends on a selected target biomolecule. In this study, we aimed at investigating whether collective knowledge can help obtain potent compounds against a target biomolecule by organizing a contest, where all the participants employ their own method to propose their compound IDs. Consequently, we found that the collection of chemical space of the potent compounds proposed based on their methods was larger than that from a single participant and the hit compounds were diverse, indicating the contest-based approach was effective to identify potent compounds.

1. はじめに

創薬には 12~14 年にもわたる期間とおおよそ 1000 億円のコストが必要とされている[1]。これを効率化するアプローチとしてコンピュータを用いた創薬支援技術 (IT 創薬) に関心が高まっている。IT による支援技術は毒性・代謝予測や構造活性相関による化合物の最適化など多岐にわたるが[2-4]、本研究では創薬の上流過程であるヒット化合物の探索プロセスに注目した。これまでに開発されてきた IT 創薬によるヒット化合物探索手法は系（標的分子）によってパフォーマンスが異なることがわかっており標準的な手法は存在しない[5]。そこで本研究では標的分子に対してして多様な手法を適用しヒット化合物を得られるか調査することを目的とした。

多様な手法を一研究者が習得して適用することは容易ではない。これを回避するために集合知的アプローチとして特定非営利活動法人 並列生物情報処理イニシアティブ (IPAB, Initiative for Parallel Bioinformatics) は「指定した標的分子の機能を阻害する化合物の探索をテーマとした公開コンテスト」を実施した。コンテストでは実際のヒット化合物探索をモデルとし、参加するグループは、購入可能な 240 万化合物から標的分子の活性を阻害すると考えられる候補

化合物 ID を優先順位と併せて提案することとした (各グループ 400 個)。全 11 グループの参加があり合計 4400 化合物が提案された。このうちそれぞれのグループの優先順位の高い化合物約 181 個、合計 1991 化合物の阻害活性アッセイを実施し高活性化合物に関しては 50%阻害濃度 (IC₅₀) 測定を実施しヒット化合物を決定した。

標的分子はチロシンキナーゼ Yes とし、このリン酸化活性をテーマとした。Yes は Src ファミリーに属する 9 種類のキナーゼのうちの一つである[6,7]。同じファミリーに属するキナーゼが抗がん剤の創薬ターゲットになっているため、Src ファミリーキナーゼの既知阻害剤は ChEMBL[8], BindingDB[9], DrugBank[10], and PubChem[11]などに豊富に登録されている。たとえば PubChem BioAssay (AID 686947) には同一実験条件による 858 個の化合物の IC₅₀ が登録されている。このため既知阻害剤に基づいたアプローチが可能である。また Yes 自体の立体構造は報告されていないが同じファミリーに属するキナーゼの構造は複数報告 (PDBID: 1Y57[13], 2SRC[14], 1FMK[15]) されているため、ホモロジーモデリングなどを用いた構造に基づくアプローチも可能であると考えられる。手法の選択や工夫に自由度が大きく多様なグループの参加を期待した。さらに Yes のリコンビナントタンパク質[16]やアッセイキット[17]が販売されているため実験を容易に開始できる。これらの理由で Yes を標的分子として選択した。

この試みは 2014 年 1-3 月に第一回 [18,19]、2015 年 1-3 月に第二回[20]が開催された。二回の開催はほぼ同じスケジュールで実施された。すなわち 1 月中旬に標的分子の

†1 東京工業大学
Tokyo Institute of Technology
†2 株式会社レベルファイブ
Level Five Co. Ltd.
†3 株式会社カタリスト
Catalyst Inc.

公表や化合物ライブラリのダウンロードを開始し3月下旬まで化合物IDを募集した。参加するためには手法の提出・連絡先の明記・化合物が公開されることに同意する必要があった。本稿では第二回に関する結果を報告する。

2. コンテストの詳細

2.1 参加条件

連絡先を明記すること、化合物が公開されることに同意すること、化合物を選択した手法を提出すること。参加無料。参加グループは指定の化合物ライブラリから優先順位と併せて400化合物のIDを提出した。

2.2 コンテストの期間

2015年1月15日：標的タンパク質の公表，化合物ライブラリのダウンロード開始，化合物ID提出開始

3月20日：化合物ID提出締切り

6月12日：参加者へのアッセイ結果の周知

7月17日：コンテストの発表・講評会

2.3 標的タンパク質

チロシンキナーゼ Yes を標的としリン酸化活性を阻害する化合物の探索をテーマにした。ただしアッセイには BPS Biosciences から購入したリコンビナントタンパク質[16]を用いた。

2.4 化合物ライブラリ

240万化合物が収載された Enamine 社提供の High throughput screening (HTS) に用いられるライブラリを基に以下の編集を実施し作成した。まず ChEMBL と BindingDB に登録された以下の条件に該当する Src の既知阻害剤（重複を除き 3528 化合物）および ChEMBL に登録されたすべての化合物について標的分子特異性が低いとみなされる化合物（フリークエントヒッター 245 化合物）のリストを作成しこれに該当する化合物を除外した。この既知化合物リストは IPAB ウェブサイト[20]からダウンロード可能である。

2.4.1 既知阻害剤 (BindingDB)

Src をクエリに検索し「 $IC_{50} \leq 10\mu M$ 」または「 $K_i \leq 10\mu M$ 」または「 $K_d \leq 10\mu M$ 」に該当した 4778 化合物。

2.4.2 既知阻害剤 (ChEMBL)

src をクエリに検索し、「Inhibition rate のデータがある」または「Relation が Inhibition rate $\geq 30\%$ 」に該当する 1858 化合物。標的分子は Src ファミリーに含まれるものに限定した（表 1 参照）。

2.4.3 フリークエントヒッター

同一化合物が 5 種類以上の標的に対して「 $IC_{50} \leq 10\mu M$ 」または「 $K_i \leq 10\mu M$ 」または「 $K_d \leq 10\mu M$ 」に該当しかつドラッグライクでない（ドラッグライクの定義は表 2 参照）化合物。245 化合物が該当した。

2.5 アッセイ手法

すべてのアッセイは Bienta 社に委託された。Bienta では Promega 社の ADP-Glo kinase assay platform のスクリーニングキット[17]が poly (Glu-Tyr) を基質として使用された。まず、全 1991 化合物を対象としたプライマリアッセイ、プライマリアッセイで基準を超えた化合物の阻害活性有無の再確認（セカンダリアッセイ）で化合物の絞り込みを実施した。これら二つのアッセイでの最終濃度は Yes 5.5 nM, ATP 0.013 mM, 基質 0.2 mg/mL, 注目している化合物 10 μM となるように調製された。さらにセカンダリアッセイで高活性であった化合物に関しては IC_{50} 測定が実施された。

表 1 Src ファミリー既知阻害剤として化合物ライブラリから除外するために用いたリスト

Table 1 ChEMBL IDs in which compounds eliminated from compound library are shown.

ChEMBL ID	標的分子
CHEMBL4223	Tyrosine-protein kinase FRK
CHEMBL3234	Tyrosine-protein kinase HCK
CHEMBL3905	Tyrosine-protein kinase Lyn
CHEMBL2250	Tyrosine-protein kinase BLK
CHEMBL258	Tyrosine-protein kinase LCK
CHEMBL4454	Tyrosine-protein kinase FGR
CHEMBL5703	Tyrosine-protein kinase Srms
CHEMBL1841	Tyrosine-protein kinase FYN
CHEMBL267	Tyrosine-protein kinase SRC
CHEMBL2073	Tyrosine-protein kinase YES

表 2 ドラッグライクネスの定義

Table 2 Definition of druglikeness

プロパティ ^a	範囲	プロパティの説明
QPlogPo/w	-2.0 - 6.5	オクタノール / 水分配係数の予測値
QPlogS	-6.5 - 0.5	溶解度の予測値
#rtvFG	0 - 2	反応性の高い官能基の個数

a) これらの値は Canvas Version 2.2.013[21,22]を用いて計算された。

2.5.1 プライマリアッセイ

冷凍保存されていた化合物を DMSO に溶解し、25 枚の 384 well プレートに 4 well 分ずつ割り振り、阻害率測定を 4 回（4 well 分）行い、それぞれの化合物について、4 回のアッセイ結果の平均値をとり、プレートごとに異なる以

下の基準でバリデーションに進める化合物を選択した。この基準は 384 well プレートのよう測定に用いる溶液（最終容量 20 μL ）が少ないためプレートごとの実験誤差が大きくなりやすいこのようなアッセイにおいて一般的なものの一つである。

「その化合物の阻害率 > その化合物が測定されているプレートのポジティブコントロール・ネガティブコントロールを除いた阻害率の平均値 + その標準偏差 $\times 3$ (プレートクライテリア)」かつ「その化合物の阻害率 $\geq 25\%$ 」

2.5.2 セカンダリアッセイ

上述の基準を満たした 68 化合物が冷凍保存されていた状態から新しく溶解され、1 枚のプレート上で 4 well 分ずつ用いてアッセイされた。4 回のアッセイの平均値を阻害率とした。

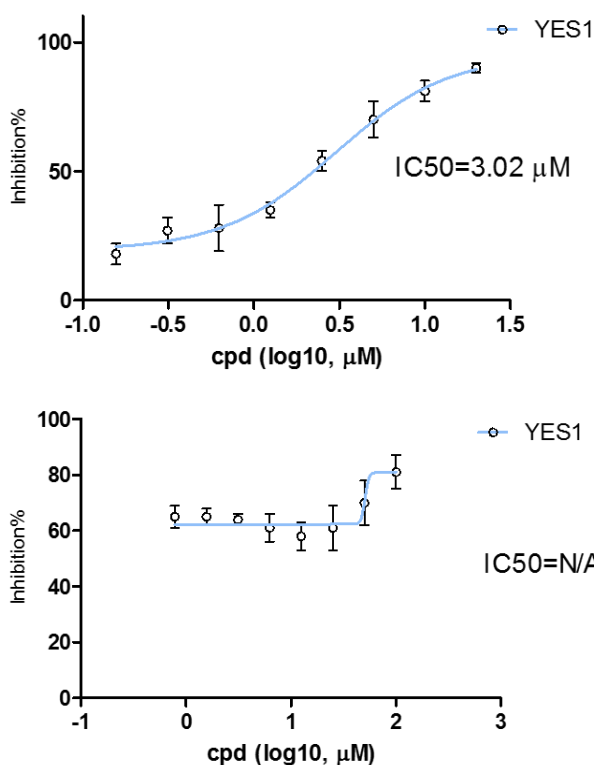


図 1 用量反応曲線の例。ヒット化合物の IC_{50} の典型例(上)、用量反応性がみられなかった場合(下)。

Figure 1 Typical examples of dose-response curves. Hit compound (Upper), compound without dose-response relationship (Lower).

2.5.3 IC_{50} 測定

セカンダリアッセイの阻害率 $\geq 45\%$ が確認された 16 化合物について IC_{50} が測定され最終的な活性値が決定された。 IC_{50} は化合物濃度を 100 μM から二倍希釈しながら 6 well 用いて測定した阻害率に基づいて決定された。測定濃度範囲の下限は必要に応じて 0.06 μM まで拡張された。

2.6 ヒット化合物の定義

$\text{IC}_{50} \leq 10 \mu\text{M}$ を満たす化合物をヒット化合物と定義した。

3. 結果と考察

3.1 参加グループ

大学生・ポストドクター・製薬業界の個人参加・ベンチャー企業など合計 11 グループから参加があった。

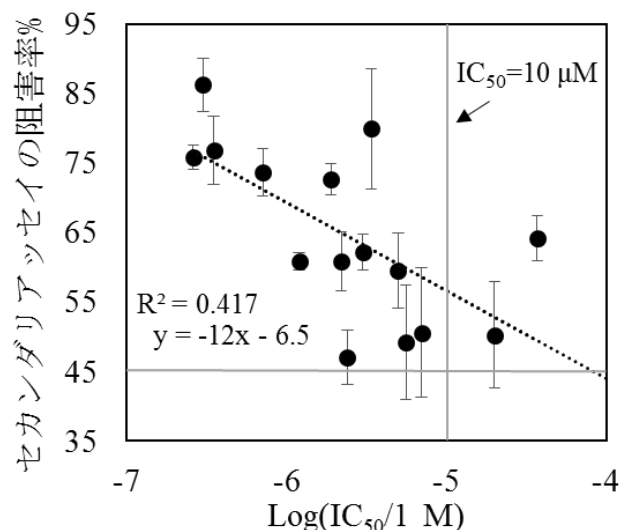


図 2 IC_{50} とセカンダリアッセイの阻害率。エラーバーはセカンダリアッセイの標準偏差。 $\text{IC}_{50} = 10 \mu\text{M}$ と阻害率 = 45% に垂直と水平の線を引いている。

Figure 2 Relationship between $\text{Log}(\text{IC}_{50}/1 \text{ M})$ vs inhibition rate in secondary assay. Lines on $\text{IC}_{50} = 10 \mu\text{M}$ and on inhibition rate = 45% are shown.

3.2 ヒット化合物数

用量反応曲線の例を図 1 に示した。 IC_{50} 測定によって用量反応性がみられない化合物を取り除くことができた。ヒット化合物の定義を満たした化合物は 13 個得られ、最終的なヒット率は 0.65% であった。

まずグループから提出された化合物群をランダムサンプリングと比較するために、化合物の優先順位順にアッセイ数を増やしていく場合の累積ヒット化合物数を調べた。結果としてヒット化合物は優先順位が高い領域に多く現れることが分かった。全アッセイ化合物の優先順位の上位 30% までに全ヒットの 77% が得られた。このことは提出された化合物群はランダムサンプリングと比較し、活性化化合物を効率的に濃縮して含んでいたことを示している。

次に IC_{50} 測定に進めるための基準「セカンダリアッセイの阻害率 $\geq 45\%$ 」がどの程度適当だったか調べるために、図 2 に $\text{Log}(\text{IC}_{50}/1 \text{ M})$ に対してセカンダリアッセイの阻害率

をプロットした。破線で示した最小二乗法による回帰直線とその決定係数 ($R^2=0.417$) から、これらが相関関係を有することがわかる。阻害率が 65%以上になると IC_{50} が 10 μM 以上の化合物は得られなかった。一方 65%以下では阻害率の実験誤差 (4 well 分の測定値の標準偏差) が大きくなる傾向があり、今回アッセイされなかった阻害率 45%以下の化合物でも IC_{50} が 10 μM 以下の化合物が得られる可能性があることが判明した。 IC_{50} 測定が実施されなかった 45%以下の化合物の阻害率を良い順にいくつか挙げると $37.1 \pm 4.2\%$, $36.4\% \pm 5.7$, $27.9 \pm 8.5\%$ であった。これらの化合物やさらに阻害率の低かった化合物 IC_{50} 測定は今後取り組まれる必要があるかもしれない。しかし IC_{50} が 1 μM を下回る高活性化合物が得られる期待はそれほど高くないと思われる。

3.3 化合物の多様性

創薬の初期段階でヒット化合物が得られてもその後のプロセスで脱落することが多い。このため展開の余地を確保するため多様な骨格を有する化合物を探索し多様なヒット化合物を得ることが望ましい[1,23]。ここでは提出された化合物や最終的なヒット化合物の構造にどの程度の多様性があるか調べた。

3.3.1 提出化合物の多様性

ケミカルスペース上の各グループから提出された化合物およびすべてのグループの化合物の位置を解析すると、各グループは異なる領域を探索し全体として広い領域を探索できていたことがわかった。ヒット化合物の位置は探索領域全体から見ると狭い部分に集中していた。結果としてヒット化合物の領域から比較的遠い領域から集中して提案したグループもあったがこのことはこれらのグループの手法が劣っているということの意味しない。むしろほかの手法が提案しなかった領域を提案したため全体として探索空間の拡大に寄与したといえる。

3.3.2 ヒット化合物の多様性

ヒット化合物同士の構造類似度 (MACCS fingerprint [24] を用いて Tanimoto 係数に基づく) を計算し、異なるグループから提案された異なるヒット化合物の平均類似度と同一グループ内の平均類似度を比較した。前者の平均類似度 (0.44) は後者と比較して (0.55) 小さかった。これは、異なる手法を適用することで広いケミカルスペースの探索が可能となるだけでなく、結果としても多様なヒット化合物を得られることを示唆している。

4. 結論

本研究ではコンピュータによるヒット化合物探索を行う場合に多様な手法を単一の標的分子に適用することが有効かどうかを調べるためチロシンキナーゼ Yes を標的分子とした公開のコンテストを実施した。参加した 11 グループか

ら提出された化合物 ID の阻害活性を実際に測定し 13 個のヒット化合物を得た。ヒット化合物は化合物 ID 提出時に参加グループが指定していた優先順位の高い領域に濃縮されていたことから、本研究の手法がランダムサンプリングと比較して妥当であることが示された。全提案化合物のケミカルスペース上の領域は単一グループから提案された化合物の領域より多様であったこと、加えて、見つかったヒット化合物の多様性は単一のグループから提案されるより多数のグループから提案される方が広がったことから集合知 (本研究ではコンテスト形式) によるアプローチは有効であることが示された。

謝辞 本活動は以下の団体の協力を得て特定非営利活動法人 並列生物情報処理イニシアティブ(IPAB, Initiative for Parallel Bioinformatics)により実施された。謹んで感謝の意を表す。一般財団法人高度情報科学技術研究機構、一般社団法人バイオ産業情報化コンソーシアム、株式会社 H P C テック、シュレーディング株式会社、株式会社情報数理バイオ、ダッソー・システムズ・バイオピア株式会社、ディスクヴァリソース株式会社、株式会社データダイレクト・ネットワークス・ジャパン、デル株式会社、ナミキ商事株式会社、日本電気株式会社、三井情報株式会社、株式会社リジット、株式会社リバネス、株式会社レベルファイブ、経済産業省、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO)、日本製薬工業協会、一般社団法人情報処理学会、特定非営利活動法人日本バイオインフォマティクス学会、特定非営利活動法人 情報計算化学生物学会 (CBI 学会)、PC クラスタコンソーシアム、株式会社科学新聞社、株式会社日刊工業新聞社、国立大学法人東京工業大学。

参考文献

- 1) Paul, S. M. et al.: How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*, Vol.9, pp.203-214 (2010).
- 2) Egan, W. J. et al.: Prediction of drug absorption using multivariate statistics. *J Med Chem* Vol.43, pp.3867-3877 (2000).
- 3) Jorgensen, W. L et al.: Prediction of drug solubility from structure. *Adv Drug Deliv Rev* Vol.54, pp.355-366 (2002).
- 4) Sliwoski, G. et al.: Computational methods in drug discovery. *Pharmacol Rev* Vol.66, pp.334-395 (2014).
- 5) Morgan, S. et al.: The cost of drug development: a systematic review. *Health Policy* Vol.100, pp.4-17 (2011).
- 6) Manning, G. et al.: The Protein Kinase Complement of the Human Genome. *Science*, Vol.298 pp.1912-1934 (2002)
- 7) Manning, G. et al.: Evolution of protein kinase signaling from yeast to man, *Trends Biochem Sci*, Vol.27, pp.514-520 (2002)
- 8) Gaulton, A. et al.: ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res*, Vol.40, pp.D1100-D1107 (2012).
- 9) Liu, T., Lin et al.: BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities, *Nucleic Acids Res* Vol.35, pp.D198-D201 (2007).
- 10) Knox, C. et al.: DrugBank 3.0: a comprehensive resource for

- 'omics' research on drugs. *Nucleic Acids Res*, Vol.39, pp.D1035-1041, (2011).
- 11) Li, Q. et al.: PubChem as a public resource for drug discovery, *Drug Discov Today* Vol.15, pp.1052-1057 (2010)
- 12) Patel, P. R. et al.: Identification of potent Yes1 kinase inhibitors using a library screening approach. *Bioorg Med Chem Lett*, Vol.23, pp.4398-4403 (2013)
- 13) Cowan-Jacob, S. W. et al.: The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation, *Structure* Vol.13, pp.861-871 (2005).
- 14) Xu, W. et al.: Crystal structures of c-Src reveal features of its autoinhibitory mechanism, *Mol Cell* Vol.3, pp.629-638 (1999)
- 15) Xu, W. et al.: Three-dimensional structure of the tyrosine kinase c-Src, *Nature* Vol.385, pp.595-602 (1997)
- 16) NCBI reference sequence: NP_005424.1
- 17) Promega Technical Manual for ADP-Glo™ Kinase Assay (Fitchburg, WI, USA. Catalog number: V9102)
- 18) IPAB コンテスト : 「コンピュータで薬のタネを創る」
<http://www.ipab.org/eventschedule/contest/contest> 2015年8月10日閲覧
- 19) 論文投稿中
- 20) 第2回 IPAB コンテスト : 「コンピュータで薬のタネを創る2」
<http://www.ipab.org/eventschedule/contest/contest2> 2015年8月10日閲覧
- 21) Duan, J. et al.: Analysis and Comparison of 2D Fingerprints: Insights into Database Screening Performance Using Eight Fingerprint Methods, *J Molec Graph Model*, Vol.29, pp.157-170 (2010).
- 22) Sastry, M. et al.: Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments, *J Chem Inf Model*, Vol.50, pp.771-784 (2010).
- 23) Owens, P. K. et al.: A decade of innovation in pharmaceutical R&D: the Chorus model, *Nat Rev Drug Discov*, Vol.14, pp.17-28, (2015).
- 24) Durant, J. L. et al.: Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comp Sci* Vol.42, pp.1273-1280 (2002).