

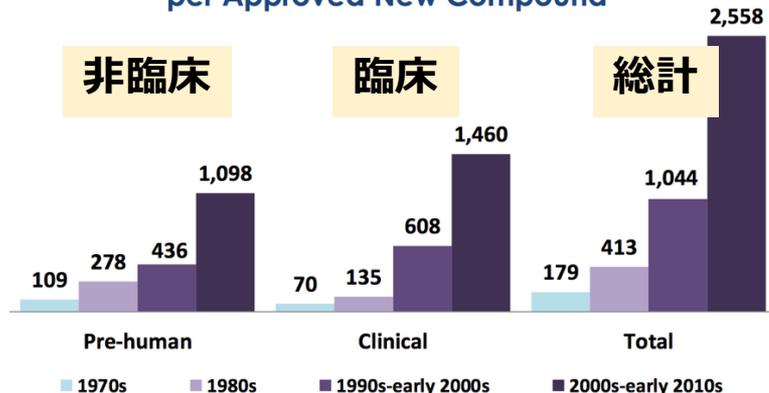
創薬の専門家の目を代替する AIの開発とその応用

東京工業大学 科学技術創成研究院
スマート創薬研究ユニット
ユニットリーダー 関嶋政和

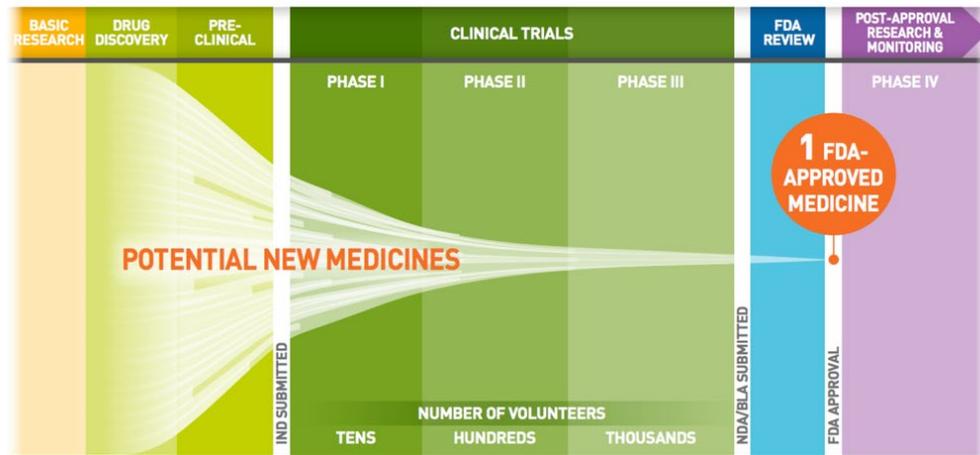
スマート創薬が必要な社会的背景

Growth in Capitalized R&D Costs per Approved New Compound

Millions of 2013 \$



Sources: 1970s, Hansen (1979); 1980s, DiMasi et al. (1991); 1990s-early 2000s, DiMasi et al. (2003); 2000s-early 2010s, Current Study



- 上市薬 1 案件あたり **3000億円**程度まで上昇 (1ドル=120円)
- 上市薬 1 案件あたり **10年**

- 従来、製薬企業で一貫して行われてきた創薬の限界
- 単純な標的分子は調べ尽くされてしまった
→ **生命・細胞システムの熟慮が必要な難しい標的分子が対象に**

スマート創薬とは

- 情報科学技術による予測（**IT創薬**）と生化学実験を**相互補完的**に用いることで効率的な創薬を実現
 - IT創薬：
 - 候補化合物探索のためのバーチャルスクリーニング
 - ディープラーニングなどの機械学習
- オープン開発による**共通基盤技術**の開発と利用



技術の背景：
顧みられない熱帯病治療薬探索

トリパノソーマ原虫が引き起こす疾病

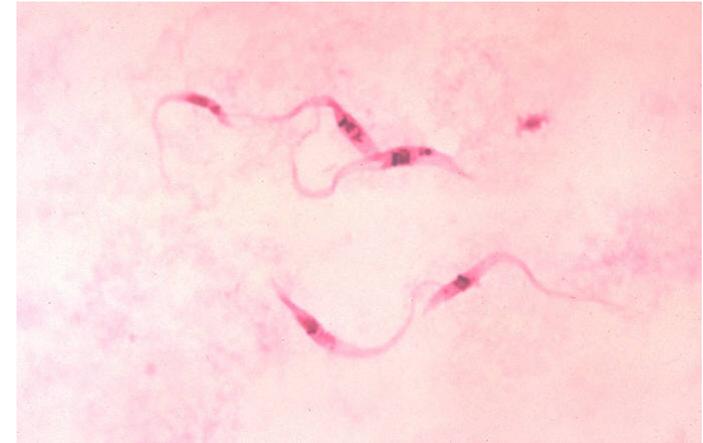
シャーガス病

寄生性の原虫である *Trypanosoma cruzi* (*T. cruzi*) によって引き起こされる感染症

- 米国南部や中南米などの約 20 カ国で影響を及ぼす病気⁽¹⁾
- 約 1500 万人が感染していると見積もられている⁽²⁾

1) Schmunis G.A., Yadon Z.E., *Acta. Trop.*, **2010**, 115, 14-21.

2) World Health Organization(TDR/GTC/09 2007) Reporte sobre la enfermedad de Chagas. Grupo de trabajo científico. Ginebra: World Health Organization.



Trypanosoma cruzi

アフリカ睡眠病

シャーガス病と同様にトリパノソーマ原虫の寄生により引き起こされる疾病

細胞に寄生した原虫が体内のエネルギーを消費

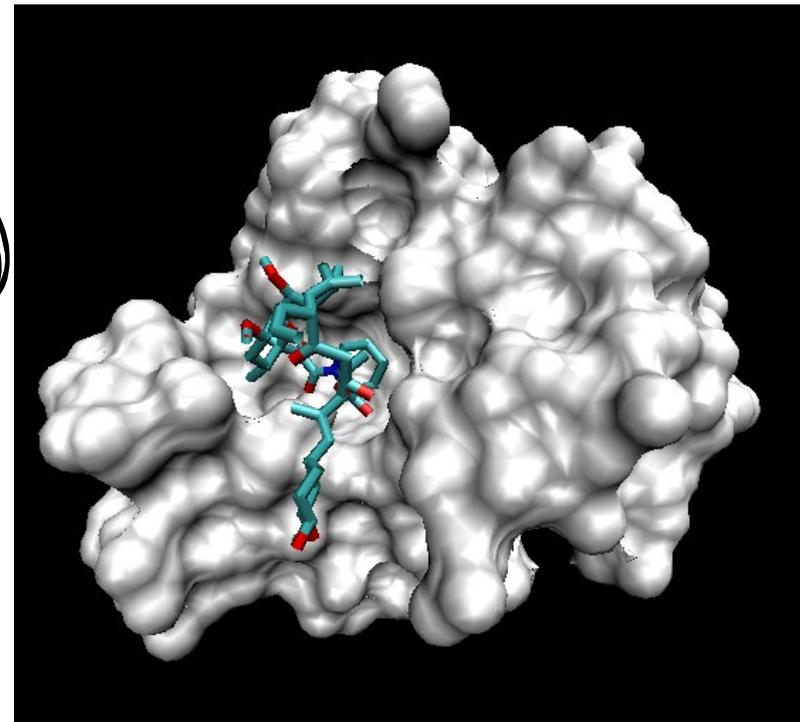
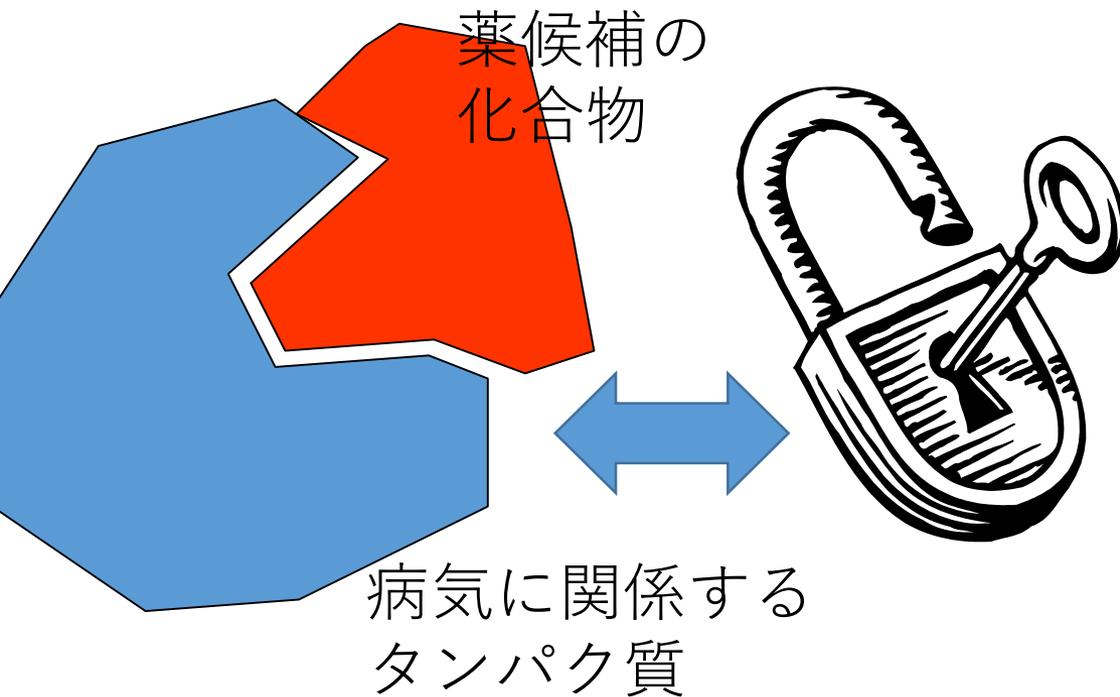


宿主が昏睡状態に陥る



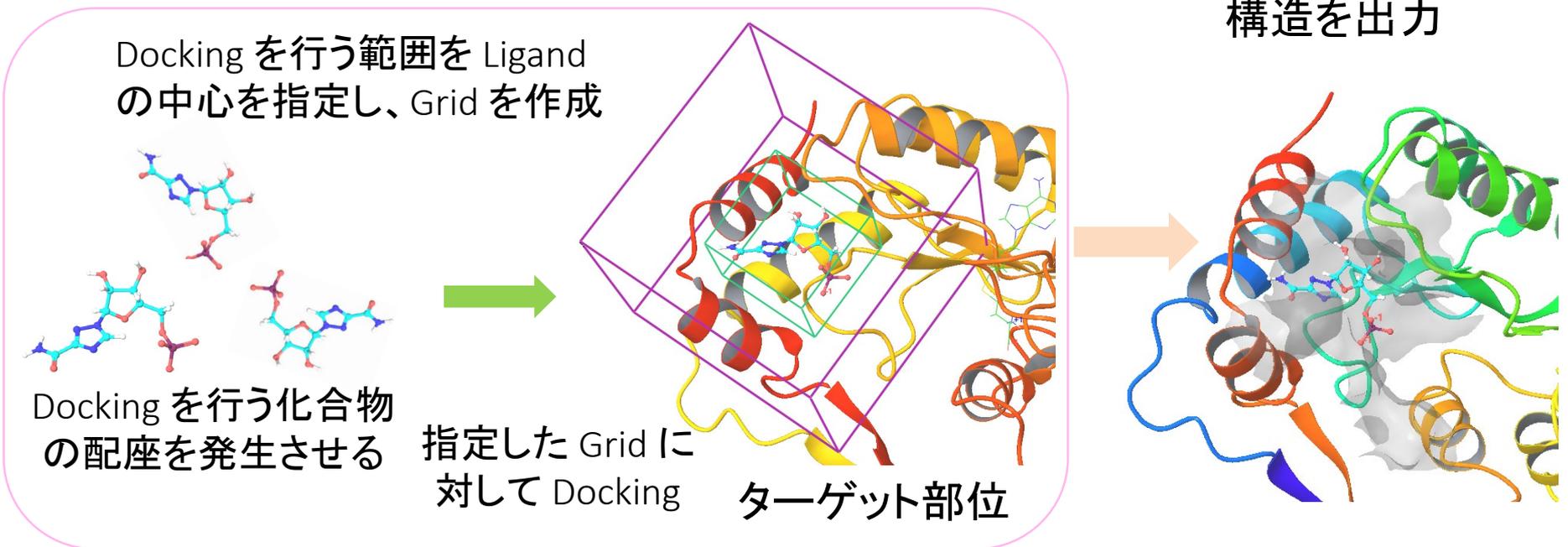
アフリカ睡眠病

タンパク質と薬候補化合物の関係は 鍵と鍵穴の関係



実際のタンパク質と薬の例

- 鍵候補を探し出すことを高速に実行する (免疫抑制剤 タクロリムス)



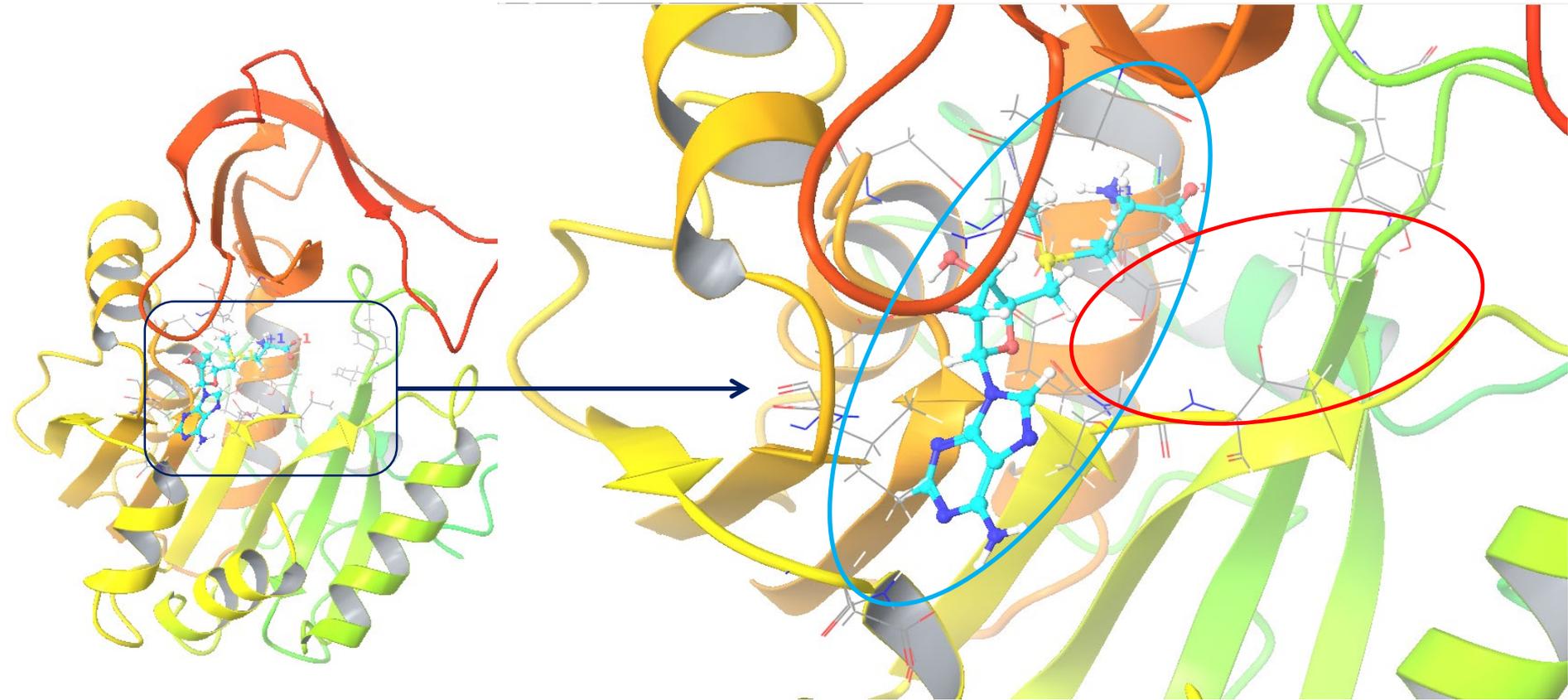
Docking Score ¹⁾

$$\Delta G_{\text{bind}} = C_{\text{lipo-lipo}} \sum f(r_{lr}) + C_{\text{hbond-neutral-neutral}} \sum g(\Delta r) h(\Delta \alpha) + C_{\text{hbond-neutral-charged}} \sum g(\Delta r) h(\Delta \alpha) + C_{\text{hbond-charged-charged}} \sum g(\Delta r) h(\Delta \alpha) + C_{\text{max-metal-ion}} \sum f(r_{lm}) + C_{\text{rotb}} H_{\text{rotb}} + C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms}$$

1) Richard A. F., Jay L. B., *et al.*, *J. Med. Chem.* **2004**, 47, 1739-1749

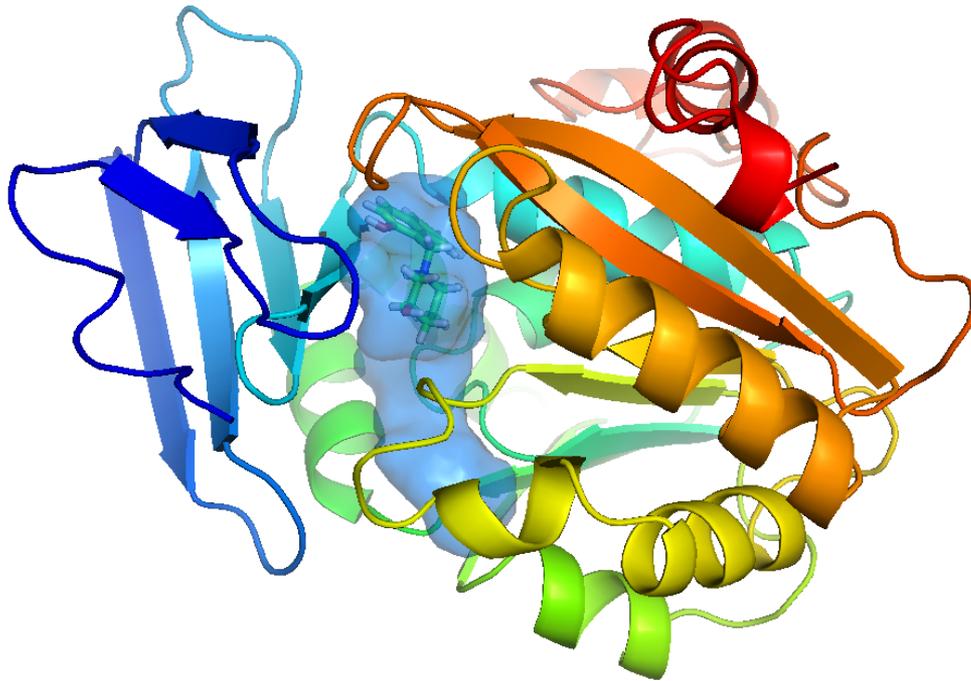
Docking の標的部位

T. cruzi spermidine synthase PDB ID : 3BWC

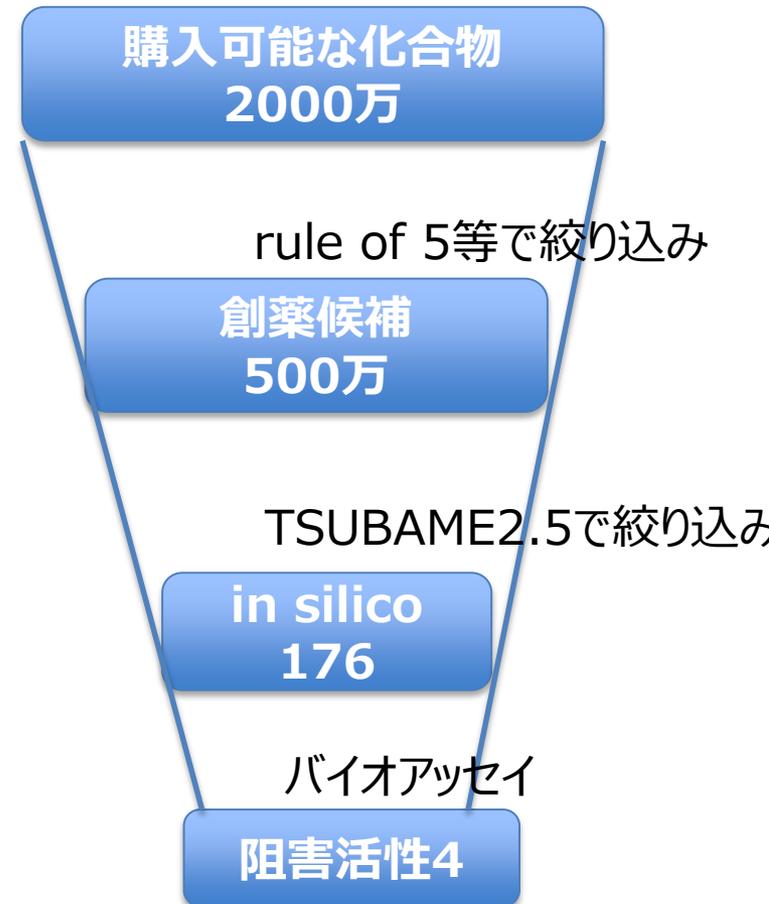


- 活性中心の Cavity には、ポリアミン供与体の dec-SAM が結合する青枠の部位と基質である putrescine が結合する赤枠の部位が存在する
— putrescine 結合部位を標的とする

in silico シミュレーションとそのアッセイ結果



TSUBAMEを用いたin silicoで計算し、
バイオアッセイで実際に阻害活性があると
確認された化合物のドッキングモード



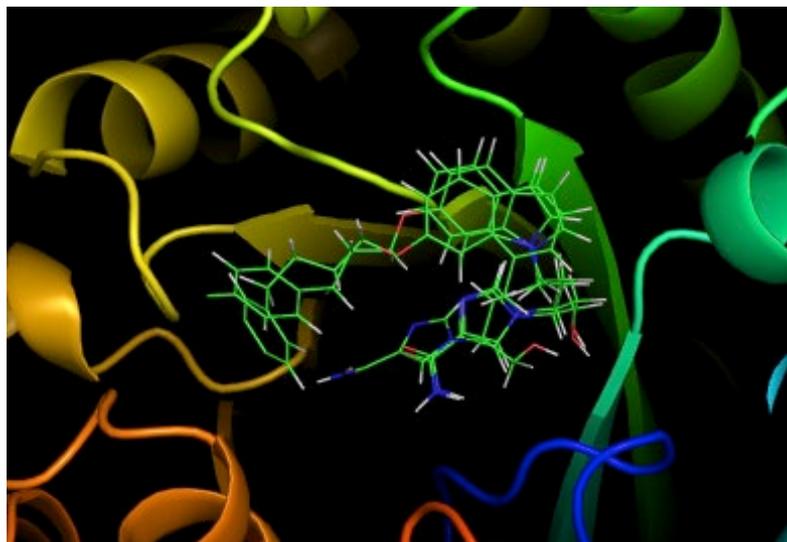
さらに、結晶構造解析で
阻害活性を持つ化合物と
標的蛋白質間の結合構造を確認

PDBID:5B1S

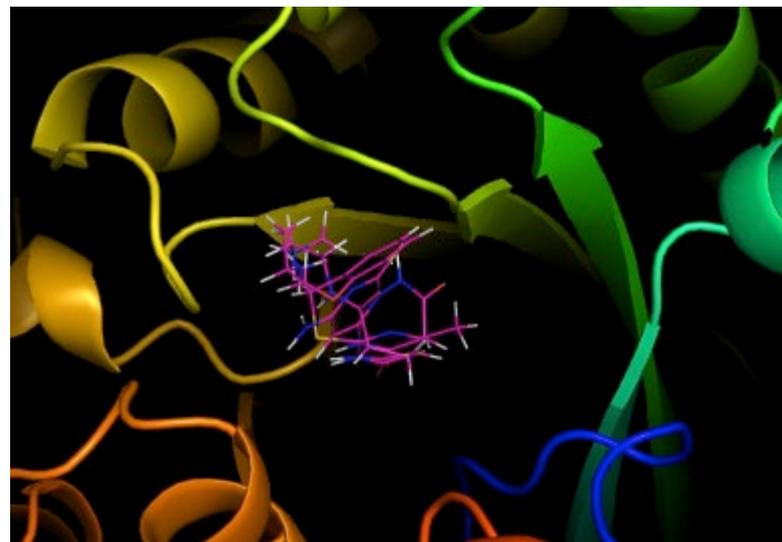
この研究から明らかになったこと

- **in silico**のシミュレーションで予測したヒット化合物は、ヒット率がHTS (High Throughput Screening)よりは高かった。
- しかし、ドッキングシミュレーションの評価関数の評価が良かったにも関わらず、ヒットしないものも多かった。

同じくらい評価関数の評価が良かった化合物



活性のある3化合物のドッキング例



活性のない3化合物のドッキング例

Visual Inspectionの応用例

近年、visual inspectionにCNN(Convolutional Neural Network)⁷⁾を用いる手法が開発されている

・コンクリートの修繕

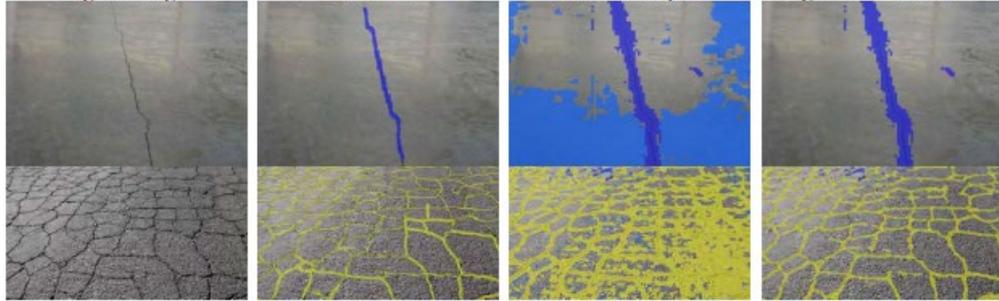
・コードのバグ

Original Image

True Labels

Classifier output

Segmenter + Classifier



診断結果 詳細情報

D:\data\exam\exam\ai\src\main\java\sample\Bo\Fox\Fox002\FX002SearchUserInfoBoObject.java

| No. | 検点 | OK | NG | OK率 |
|-----|-------|----|----|------|
| 1 | 検数度 | 2 | 1 | 0.67 |
| 2 | 可検性 | 2 | 1 | 0.67 |
| 3 | コスト密度 | 2 | 1 | 0.67 |

```

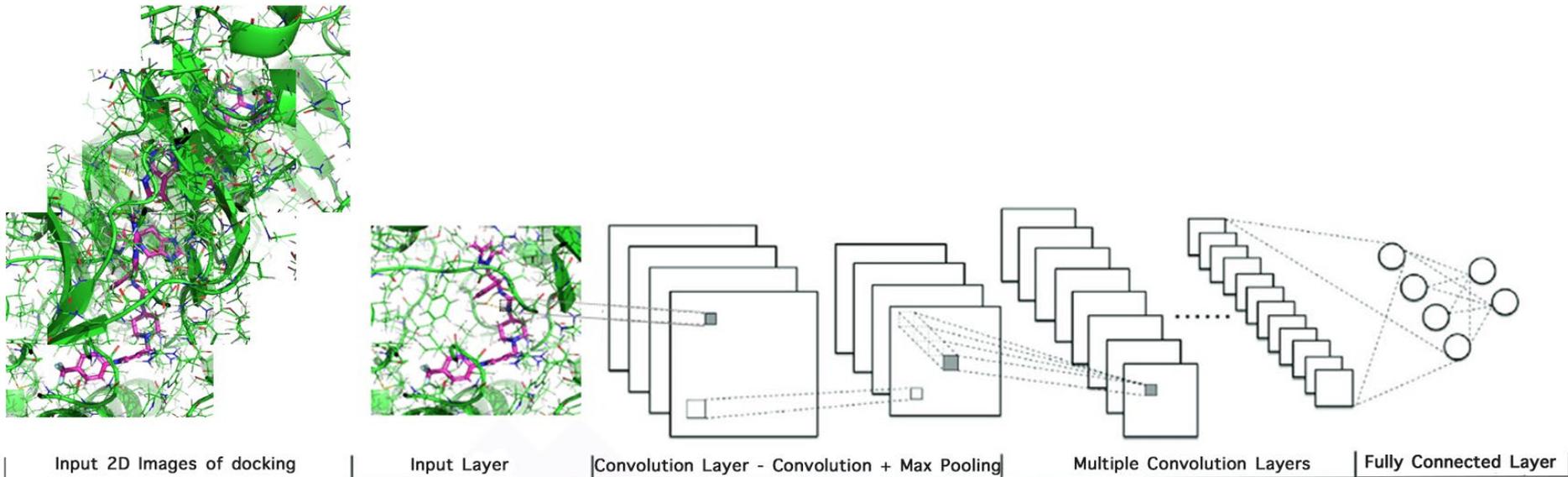
行番号 ソース内容
40 protected final void doProcess(SessionHandle sh, Map<String, Object> req) {
41
42     MUserInfoDaoImplMsg msg = createInstance(sh, MUserInfoDaoImplMsg.class);
43
44     msg.put(MUserInfoDaoImplMsgユーザID, req.get("userId"));
45     msg.put(MUserInfoDaoImplMsgユーザ名, req.get("userName"));
46     msg.put(MUserInfoDaoImplMsg電簿, req.get("req"));
47
48     if (StringUtil.hasValue((String) req.get("sortSelected")) && StringUtil.hasValue((String) req.get("sortDirection"))) {
49         String[] select = ObjectConvertUtil.convertToString((String) req.get("sortSelected")).split(",");
50         List<String> selectList = Arrays.asList(select);
51
52         String[] direct = ObjectConvertUtil.convertToString((String) req.get("sortDirection")).split(",");
53         List<String> directList = Arrays.asList(direct);
54
55         if (selectList.size() == directList.size()) {
56             msg.put(MUserInfoDaoImplMsgソート項目リスト, selectList);
57             msg.put(MUserInfoDaoImplMsgソート条件リスト, directList);
58         }
59     }
60
61     List<MUserInfoDaoImplMsg> list = this.daoimpl.execute(sh, msg);
62
63     if (list == null || list.isEmpty()) {
64         return;
65     }

```

- ・Alex K. et al., 2012.
- ・V. Hoskere. et al., 2017.
- ・<https://tech.nikkeibp.co.jp/it/atcl/column/14/346926/122501258/>

本技術のポイント

- **ディープラーニングを用いたvisual inspectionの自動化**
 - ドッキングを画像化し、CNNを用い学習する
 - メディシナルケミスト(創薬化学者)の属人的な勘と経験を大規模なデータからのディープラーニングで代替する



画像生成 3D circular augmentation

Algorithm 2 3D sphere augmentation

Require: $angles > 0$

for $i \leftarrow 0$ to $angles - 1$ do

$$X \leftarrow \frac{360 \times i}{angles}$$

for $j \leftarrow 0$ to $angles - 1$ do

$$Y \leftarrow \frac{360 \times j}{angles}$$

(X, Y) の方向から撮影

end for

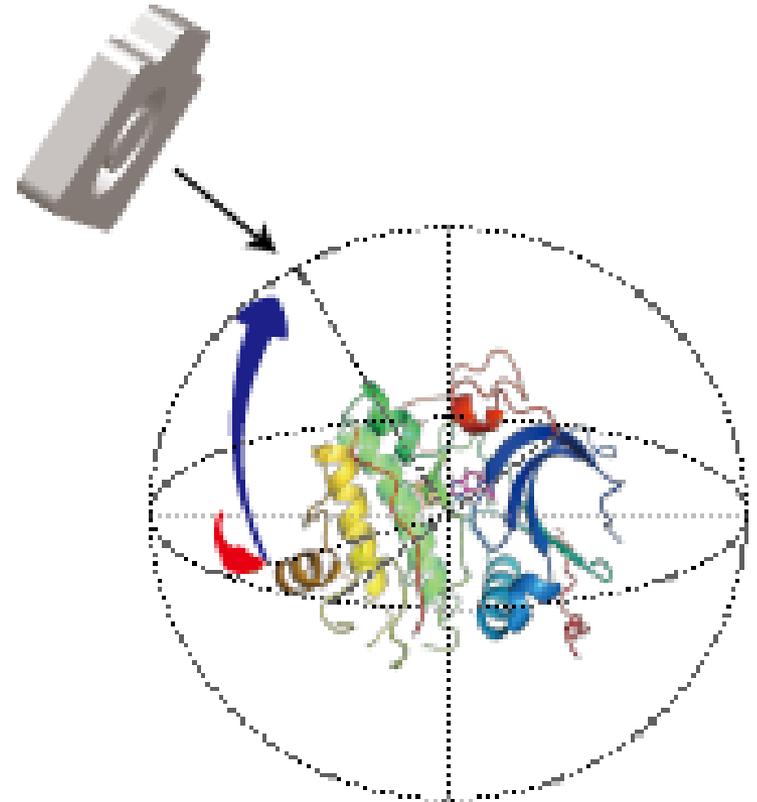
end for

オプション: 水素結合強調表示の有無

49枚: $angles=7$

81枚: $angles=9$

・距離 r 、ズームは各々の標的タンパクごとに設定する**必要がない**



活性に影響した部分を評価可能に

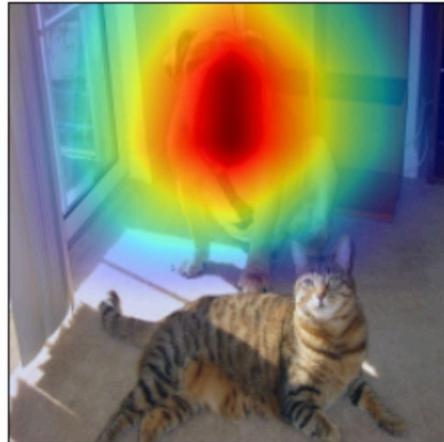
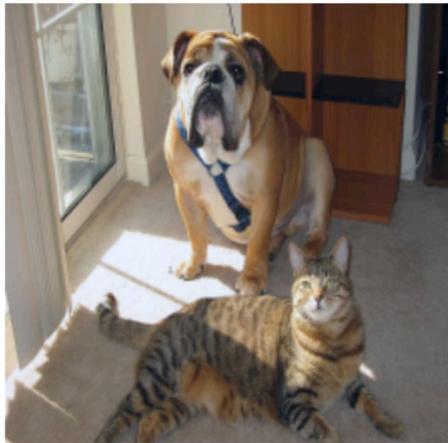
GradCam

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right)$$

画像のどの部分が、活性に影響をしたのかを評価することができる

例：犬を識別している



Selvaraju, R R. et al., 2016.

本技術の性能

| protein | EF1% | | Glide |
|---------|-------------|-------------|-------|
| | sphere_81 | sphere_H_81 | |
| akt1 | 53 | 50.4 | 10.3 |
| ampc | 37.5 | 37.5 | 6.1 |
| cp3a4 | 12.4 | 12.4 | 6 |
| cxcr4 | 50.4 | 58.8 | 14.7 |
| gcr | 56.6 | 43.4 | 20.9 |
| hivpr | 54.5 | 52.6 | 17.8 |
| hivrt | 49 | 46.8 | 20.4 |
| kif11 | 56 | 56 | 44.8 |

本技術は従来のドッキングシミュレーションの性能を大幅に上回る

特許申請済 出願番号 特願2019-015086

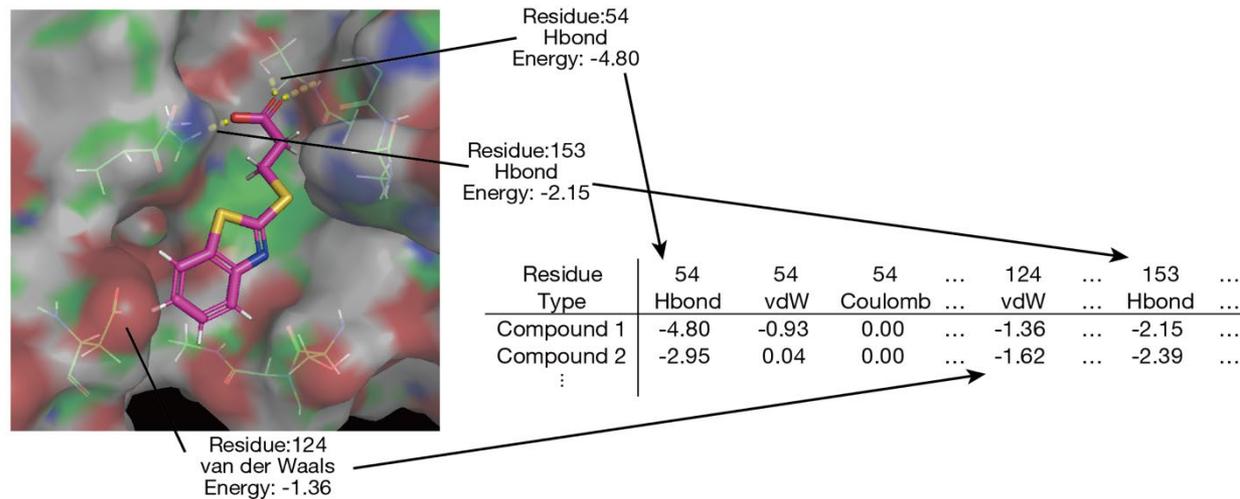
関嶋研のその他の創薬に関する技術

SIEVE-Score (Similarity of Interaction Energy Vector-Score)

蛋白質の各アミノ酸残基と化合物間の相互作用エネルギーの値を直接ベクトルとして用いて解析する
(以後相互作用エネルギーベクトルと呼ぶ)

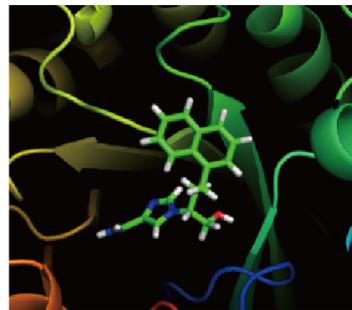
特徴量：化合物と各アミノ酸残基間の相互作用エネルギー
ファンデルワールス、静電相互作用、水素結合の3つに分解
次元数：3×(ドッキングに関連する残基数)

学習手法：Random Forest



提案手法のフロー

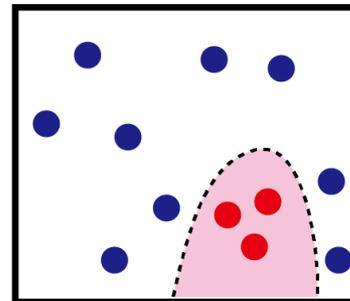
1. ドッキングを行う
2. 蛋白質の各残基と化合物の相互作用エネルギーを抽出
3. 活性既知の化合物でRandom Forestの予測モデルを作成
4. 活性未知の化合物の活性有無を予測し、再ランキング



Protein-ligand
docking

| Cpd No. | P101 vdW | P101 Coul | P101 hbond |
|---------|-------------|--------------|---------------|
| 1 | -0.15 | -3.1 | 0 |
| 2 | -2.5 | -4.0 | -0.03 |
| 3 | -2.8 | -2.5 | -0.03 |

Interaction
energy vectors
extraction



Prediction by
Random Forest

SIEVE-Score
0.08 ③
0.91 ①
0.75 ②
⋮

Re-ranking



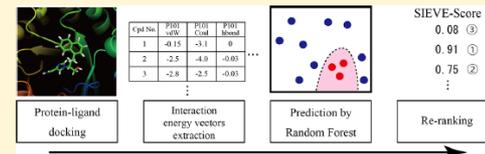
Improved Method of Structure-Based Virtual Screening via Interaction-Energy-Based Learning

Nobuaki Yasuo^{†,*,†,‡,¶,||} and Masakazu Sekijima^{†,*,†,||}

[†]Department of Computer Science, Tokyo Institute of Technology, 4259-J3-23, Nagatsuta-cho, Midori-ku, Yokohama, Japan

[‡]Advanced Computational Drug Discovery Unit, Tokyo Institute of Technology, 4259-J3-23, Nagatsuta-cho, Midori-ku, Yokohama, Japan

Supporting Information



ABSTRACT: Virtual screening is a promising method for obtaining novel hit compounds in drug discovery. It aims to enrich potentially active compounds from a large chemical library for further biological experiments. However, the accuracy of current virtual screening methods is insufficient. In this study, we develop a new virtual screening method named Similarity of Interaction Energy Vector Score (SIEVE-Score), in which protein–ligand interaction energies are extracted to represent docking poses for machine learning. SIEVE-Score offers substantial improvements compared to other state-of-the-art virtual screening methods, namely, other machine-learning-based scoring functions, interaction fingerprints, and docking software, for the enrichment factor 1% results on the Directory of Useful Decoys, Enhanced (DUD-E). The screening results are also human-interpretable in the form of important interactions for distinguishing between active and inactive compounds. The source code is available at <https://github.com/sekijima-lab/SIEVE-Score>.

INTRODUCTION

The demand for computer-aided drug discovery has been rapidly increasing since the average cost per new drug developed reached 2.6 billion dollars.¹ To find active compounds for target proteins, high-throughput screening has been used for the screening of chemical libraries.² However, the cost of such compound screening can be drastically reduced if the compound libraries can be effectively filtered.

Virtual screening is a method of key importance for decreasing the cost and increasing the hit rate of screening.^{3,4} In virtual screening, compounds are computationally filtered out by their predicted activities, and the compounds identified as promising candidates are designated for evaluation through biological experiments. Recently, competitive virtual screening contests have been also held, such as *in silico* drug discovery contest, to evaluate various virtual screening methods for specific target molecules in practical situations.^{5,6}

Methods of virtual screening can be categorized into two types, namely, structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS). The structures of target proteins are used in SBVS,^{7,8} while information on known inhibitors is used in LBVS.⁹

One widely used method in SBVS is protein–ligand docking, which simulates protein–ligand binding based on the interaction energy. Many applications exist for protein–ligand docking, such as DOCK,^{10,11} Glide,¹² GOLD,¹³ and AutoDock Vina.¹⁴ In each of these applications, the algorithm starts by generating ligand conformations and subsequently optimizes their orientations and angles to minimize their binding energies with the target protein using a scoring function. The results are the coordinates of the atoms in the ligand and the corresponding interaction energy as the “docking score”. In basic docking-based virtual screening, compounds are ranked by their docking scores.

One advantage of SBVS is that novel compounds are more likely to be obtained than they are in LBVS.⁵ This is because SBVS is based on physical interactions, while LBVS is based on the similarity/dissimilarity of known active/inactive compounds. Another advantage is the ability to perform interaction analysis using the docked structures, which provides knowledge of protein–ligand binding¹⁵ to understand the affinity¹⁶ and selectivity^{17,18} of the compounds. These analyses can also

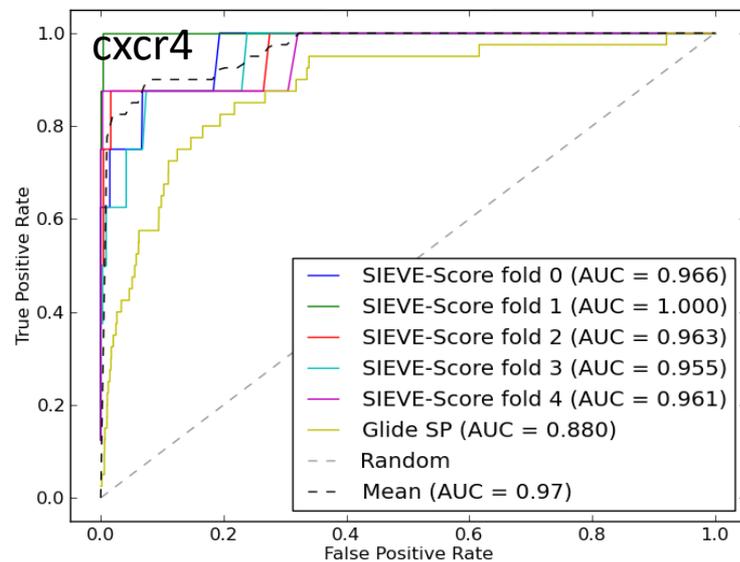
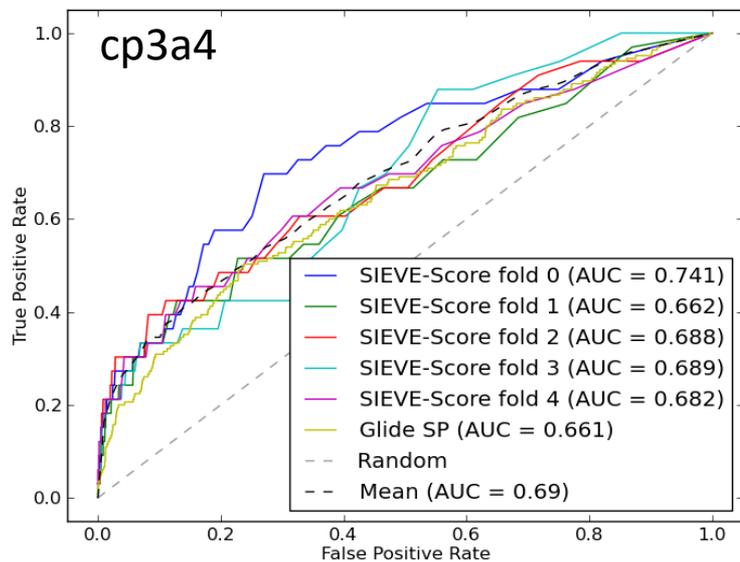
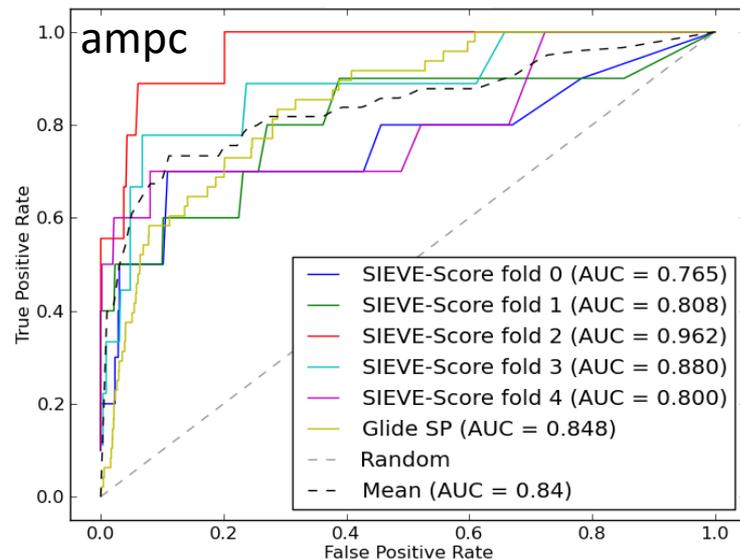
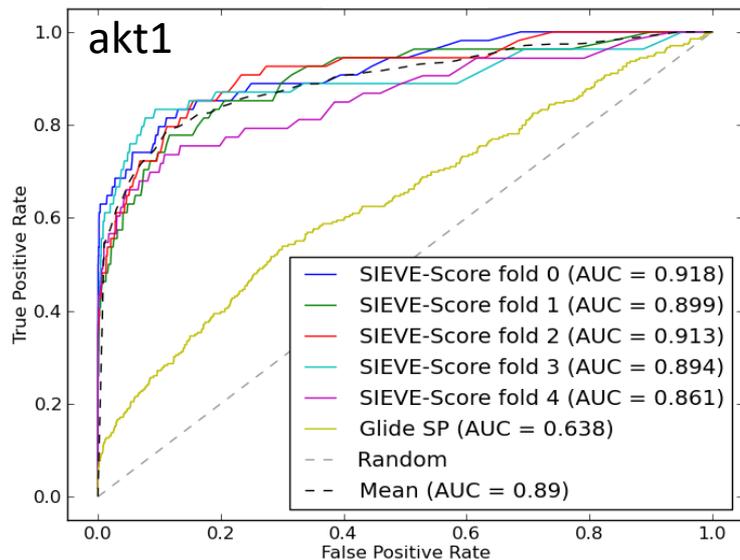
Special Issue: Machine Learning in Drug Discovery

Received: October 1, 2018

Published: February 26, 2019

結果：ROC曲線

- 黄線: Glide SP
- 他実線: SIEVE-Score, 5-fold
- - - 黒点線: 5-foldの平均



関嶋研の技術のまとめ

- 製薬企業との共同研究に基づいたin silico手法の開発
- メディシナルケミスト（創薬化学者）の目を代替する技術の開発
- ドッキングシミュレーションからエネルギーベクトルを抽出し、機械学習する方法も開発
- 本スライドでは発表していないが、蛋白質とリガンドの結合構造を直接3DCNNで学習する技術も利用可能
- 上記3技術はそれぞれ対立するものでなく、目的に応じて使い分けることで、効率的な創薬研究が実施可能

企業への期待

- 製薬企業との共同研究を希望
- 創薬の効率化を考えている製薬企業には、本技術の導入が有効と思われる

本技術に関する知的財産権

- 発明の名称 : 立体構造判定装置、立体構造判定方法、立体構造の判別器学習装置、立体構造の判別器学習方法及びプログラム
- 出願番号 : 特願2019-015086
- 出願人 : 東京工業大学
- 発明者 : 関嶋政和、依田洸、安尾信明

お問い合わせ先

東京工業大学
研究・産学連携本部

T E L 0 3 - 5 7 3 4 - 2 4 4 5

F A X 0 3 - 5 7 3 4 - 2 4 8 2

E-mail sangaku@sangaku.titech.ac.jp