

化合物データベースの利用法

福西快文 2010/09/14

- (1) 多くの薬物スクリーニング結果があったとき、化合物データベースを用いて、ヒット化合物が出やすいであろう薬物スクリーニング結果を選び出す。
- (2) 化合物データベースを用いて、標的蛋白質のポケットの位置を予測し、かつそのポケットで開発しうる薬物の活性値を予測する。

myPresto : タンパク質等のモデリング、膜タンパク質等のシミュレーション、
薬物ドッキング・スクリーニングのためのプログラム群

LigandBox: 化合物データベース
蛋白質データベース

tplgene/tplgeneL : 分子(蛋白質、低分子)
の準備

cosgene : 構造探索エンジン

MTS/DSI : スクリーニング

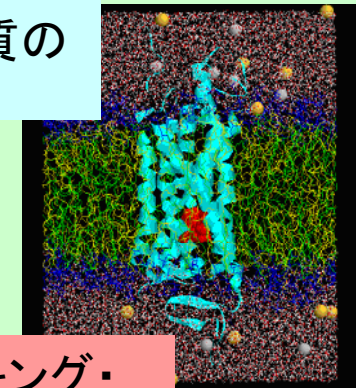
sievgene : ラフなドッキング

蛋白質などの
モデリング



構造探索

膜蛋白質の
計算



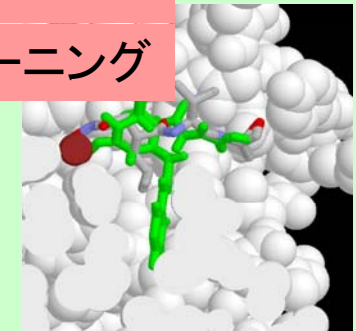
ツール

長時間シミュレーション

情報解析

薬物ドッキング・

薬物スクリーニング



tplgene / tplgeneL

トポロジー生成エンジン(蛋白質、低分子)

sievgene

ラフスクリーニングのためのドッキングエンジン

cosgene

MD計算による構造探索エンジン

各種ツール

cosgene の計算結果の解析ツール、et

1. 良い薬物スクリーニング結果 を Universal active probe (UAP) を使って選び出す。

How to select reliable screening
result among many screening
results?

Fukunishi, Yoshifumi; Ono, Kazuki; Orita, Masaya; Nakamura, Haruki,
Selection of in-silico drug screening result by using universal active probes
(UAPs), Journal of Chemical Information and Modeling, 2010, 50, 1233-1240.

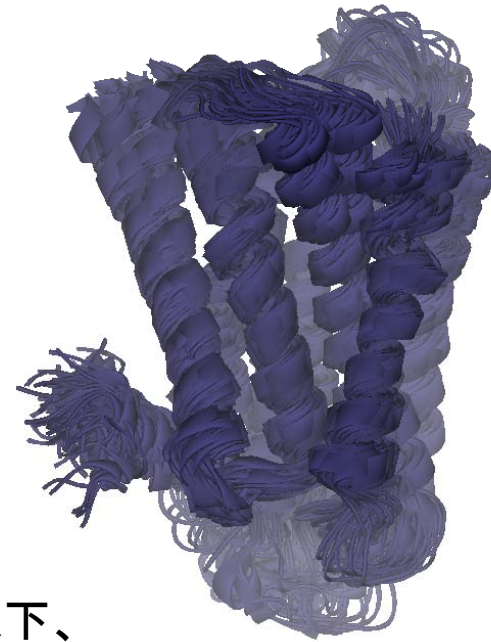
例:

- (1) Ensemble dockingでは多数のスクリーニング結果が得られる。
- (2) ある種のVirtual screening method (機械学習MTS法などの機械学習法)では、蛋白質構造が1つでも、複数のスクリーニング結果が生成しうる。

->「良い計算結果」を選ぶ必要がある。

多数のスクリーニング結果から、信頼性の高い計算結果を抽出する方法を開発する
—計算結果の信頼性尺度の開発—

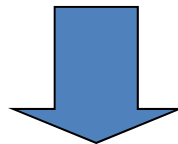
Structure-based in-silico drug screeningでは、標的タンパク質構造に結合する物質を探索する。しかし、標的蛋白質構造は、計算によって様々な形がサンプルされるし、現実、構造は揺らいでいる。そのため、計算に用いる構造の数だけ、スクリーニング結果が得られる(アンサンブルドッキング)。



データ提供
和田(富士通)、
酒匂(塩野義製薬)

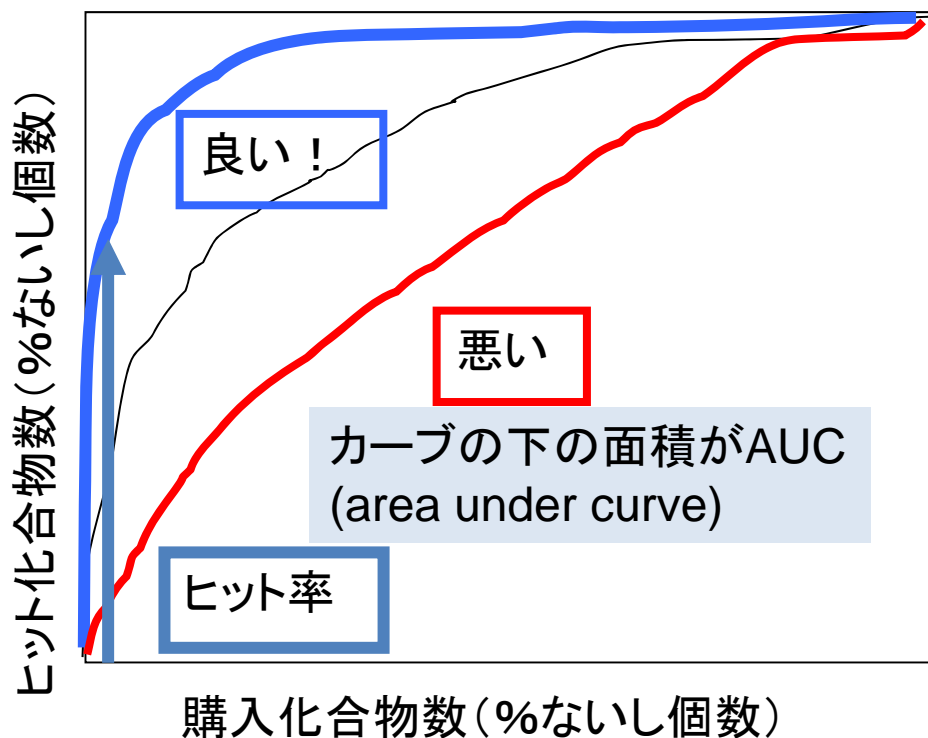
おおよそ、多数の構造に対し、

- ・極めて良好なスクリーニング結果が得られる場合は、10%以下、
- ・まあまあ良好なスクリーニング結果が得られる場合は40%、
- ・悪いスクリーニング結果が得られる場合が30%、
- ・極めて悪い(ランダムスクリーニングより悪い)場合が20%、
といったところである。



信頼性の高いスクリーニング結果を得る手法の開発とは別に、多数のスクリーニング結果から信頼できる結果を選ぶ手法が必要。

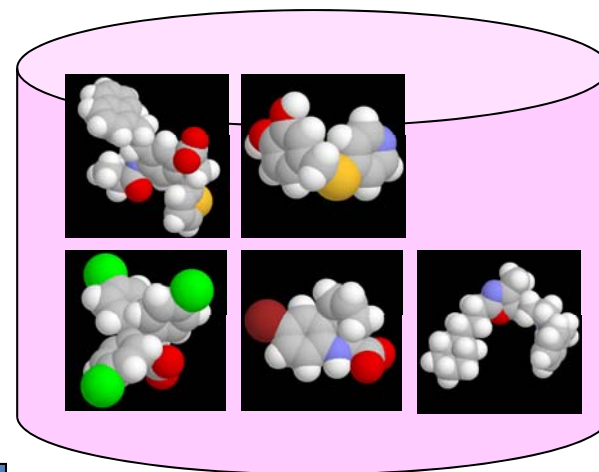
データベースエンリッチメントカーブ



薬物スクリーニングで順位づけした化合物を上位から何%購入したら、本来データベースに含まれていたヒット化合物を何%を見つけられたかを示す。

$$\text{ヒット率} = \text{ヒット数} / \text{購入化合物数} \times 100(\%)$$

0% (悪い) < AUC < 100% (良い).
AUC=50%=ランダム



薬物スクリーニング

group	file_name	mts	mts_score	SUPPLIER
c001	0006240-01	1	-2.0627	Ambinter
c001	0020411-01	3	-2.0674	Ambinter
c001	0013395-01	4	-2.0611	Ambinter
c001	0020413-01	5	-2.0663	Ambinter
c001	0013566-01	6	-2.0617	Ambinter
c001	0002135-01	7	-2.0683	Ambinter
c001	0020501-01	8	-2.0617	Ambinter
c001	0022138-01	9	-2.091	Ambinter
c001	0022583-01	10	-2.0709	Ambinter
c001	0021187-01	11	-2.0647	Ambinter
c002	0103033-01	2	0.2702	Aronis

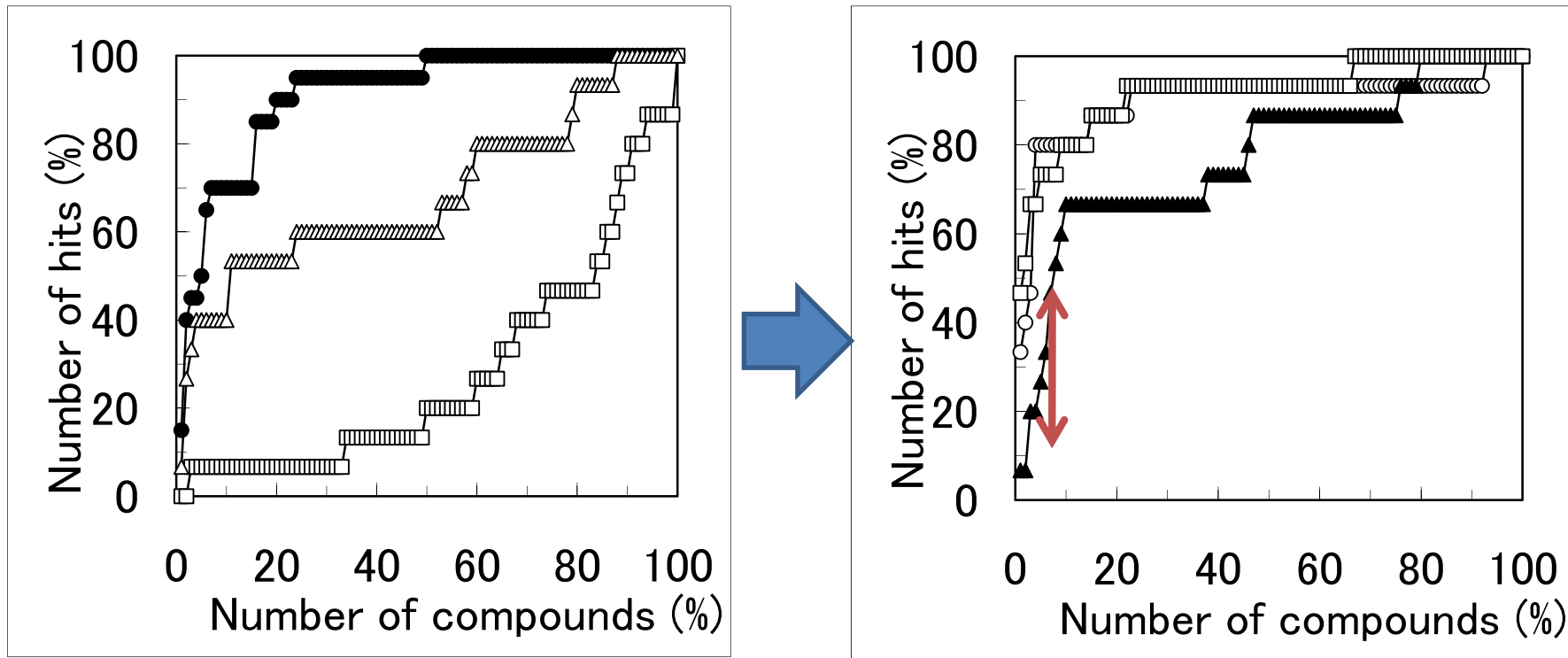
モデリングされた多数の構造に対する、スクリーニングテストでのM受容体でのAUC(q値)

q value table

	moe										prime
	1	2	3	4	5	6	7	8	9	10	
ini.	79.2	63.3	47.6	62.6	56.5	50.5	64.3	47.7	35.5	63.3	49.3
min.	54.3	48.0	50.1	65.6	64.4	69.2	60.7	47.7	51.6	71.9	62.3
2100 ps	76.6	69.4	61.8	59.1	56.1	60.0	60.1	45.7	65.0	73.4	59.2
2300 ps	66.7	63.2	35.2	57.2	50.2	49.2	82.8	34.9	66.7	48.7	60.3
2500 ps	68.8	71.1	51.9	66.2	64.8	36.2	75.6	54.7	65.8	63.9	59.7
2700 ps	48.8	60.3	44.5	62.7	65.8	47.5	63.0	61.2	60.9	54.1	58.0
2900 ps	74.2	64.8	56.7	62.2	71.9	56.9	53.6	57.2	62.4	37.3	50.7
3100 ps	87.7	65.4	70.9	57.2	60.1	63.7	67.3	53.3	53.0	58.6	46.3
3300 ps	73.7	68.8	57.9	52.6	56.8	59.1	65.2	54.3	41.2	64.6	60.3
3500 ps	64.7	40.4	36.8	57.8	58.0	58.2	57.1	53.2	55.7	49.0	60.6
3700 ps	76.9	61.9	48.9	56.1	57.5	37.3	70.0	60.9	66.2	32.5	55.2
3900 ps	75.8	64.3	54.6	45.0	52.2	49.3	69.6	47.3	63.3	44.9	60.4

機械学習Multiple-target screening法を用いた場合。
スクリーニング計算結果は、揃ってくるが、それでも、エンリッチメントの立ち上がりにはばらつきがみられる。

COX-2を標的とした場合の、データベースエンリッチメントカーブ



機械学習法は、なんらかの方法で乱数系列などを使うため、結果が乱数系列に依存する。

Universal active probe (UAP)の導入による、スクリーニング結果の選択

MTSスクリーニングは、蛋白質構造が多数あれば、その数だけ結果を生成できる。

蛋白質構造は、MDシミュレーションでいくらでも生成できる。

問題は、どのスクリーニング結果を選ぶか、である。

標的蛋白質には結合しないはずの、drug-likeな化合物の集団を作成した(UAP)。これをスクリーニング計算に混ぜておく。

すると、標的に対する真の化合物がヒットする割合は、UAPが発見される割合に比例する結果となった。

つまり、標的の活性化化合物が未知であっても、UAPを目印に、もっともらしいスクリーニング結果を選ぶことができる、と考えられる。

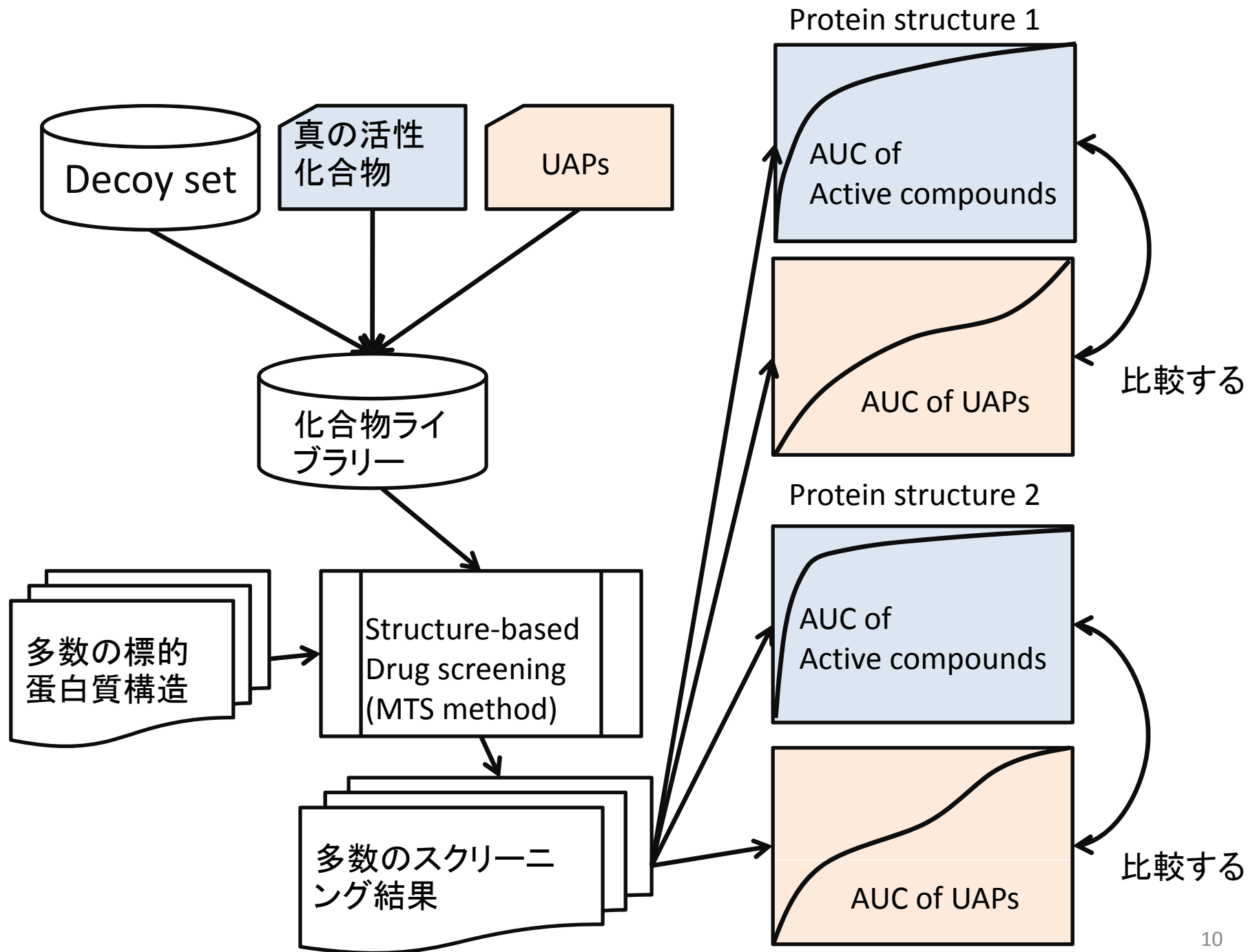
用いたデータセット

標的蛋白質： 水溶性の 11 蛋白質 (COX2, AMPC, fXa, Thrombin and etc).

UAP:

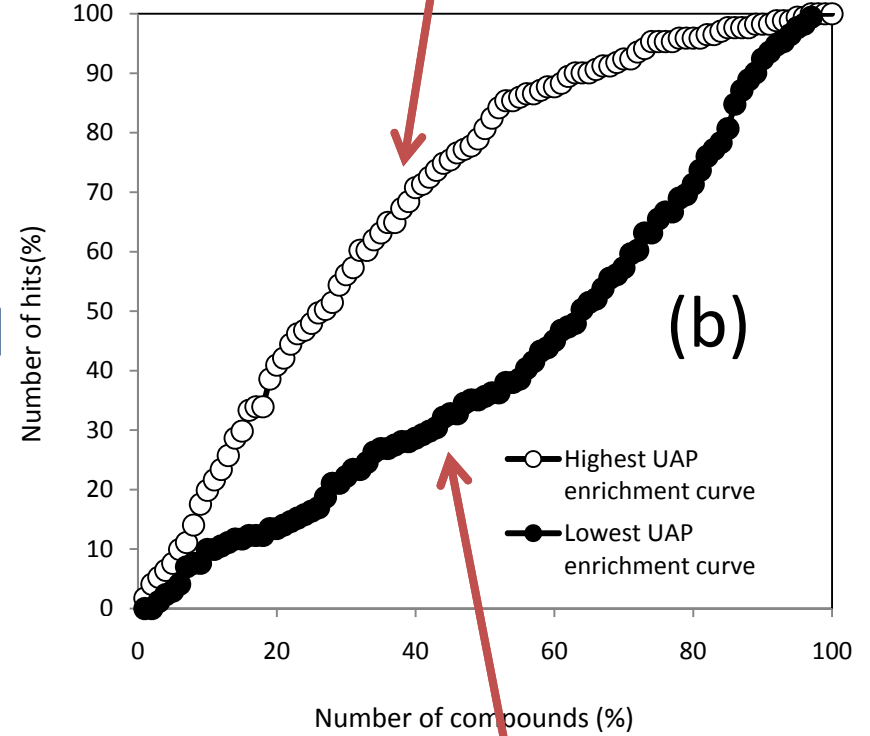
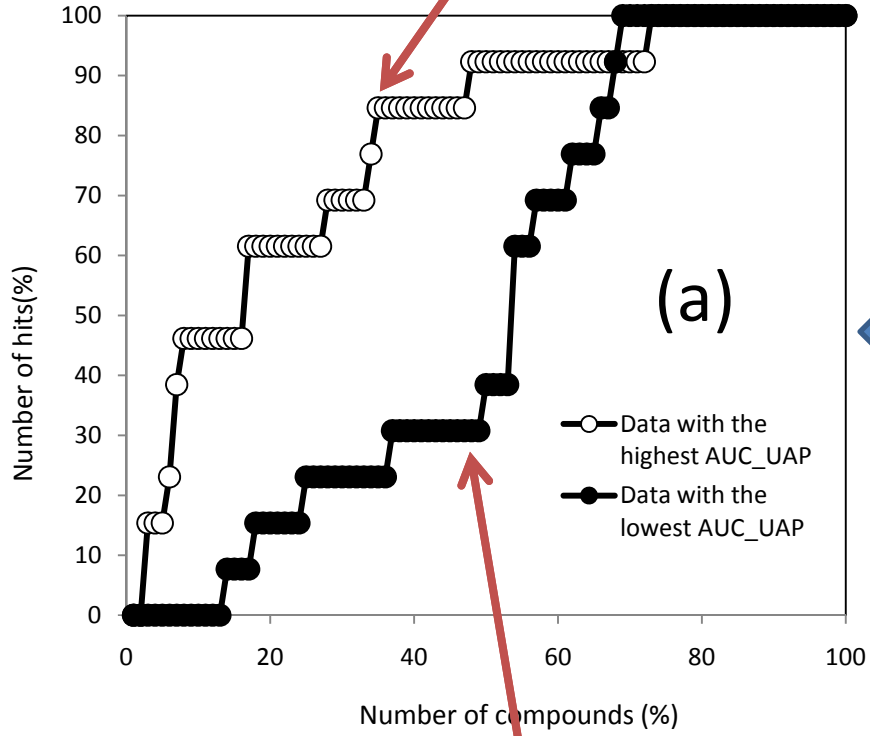
- (1) GPCRの既知リガンド
(約 100 compounds: UAP_GPCR),
- (2) DUDから集めた水溶性蛋白質の既知活性化合物
(約 100 compounds: UAP_DUD),
- (3) PDBにある蛋白質-化合物複合体のリガンド
(約 100 compounds: UAP_PDB).

Total 約 300 compounds.



真の活性化合物がスクリーニング結果の上位にくる

UAPがスクリーニング結果の上位にくる

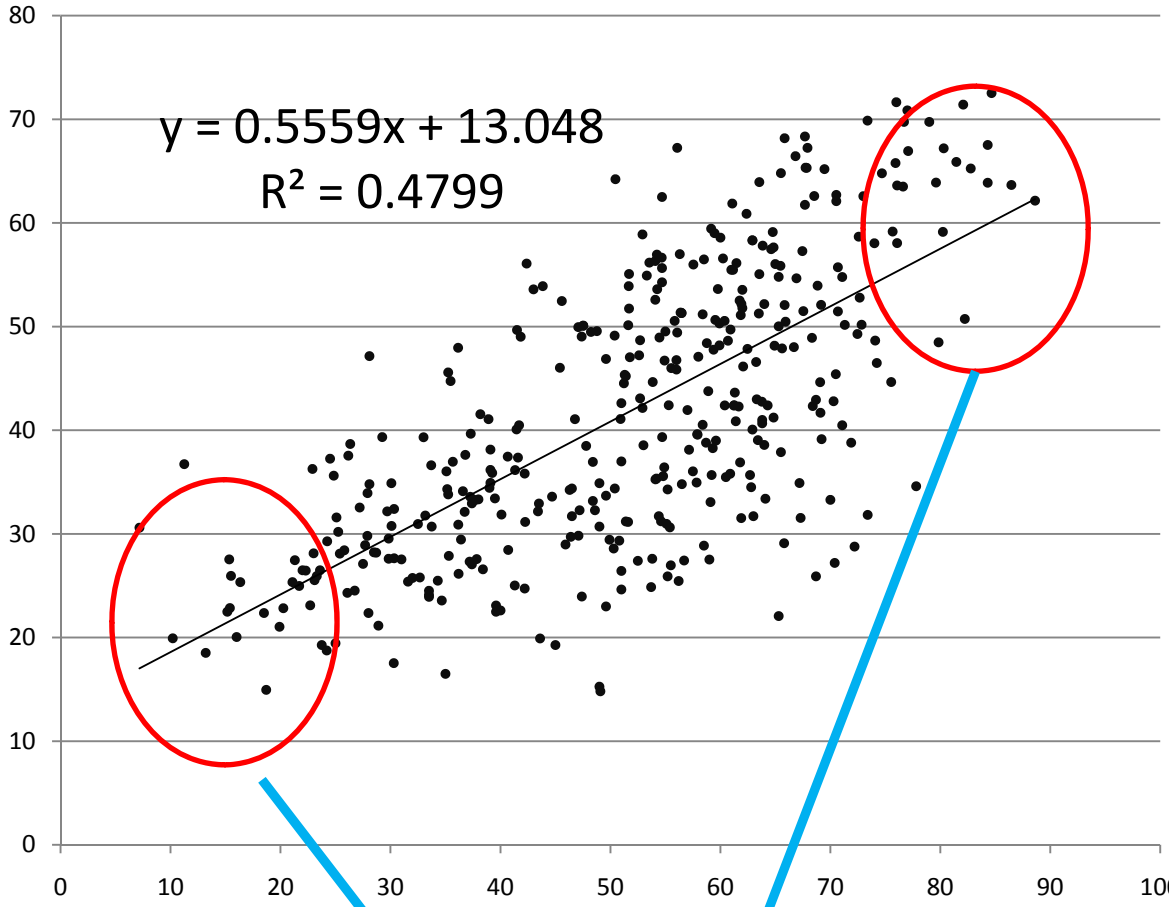


真の活性化合物がスクリーニング結果の下位にくる

UAPがスクリーニング結果の下位にくる

相関係数(R) = 0.69

UAPが見つかるヒット率
(AUC for UAP)



AUC value

Bottom	avg	Top	avg
10	31.79	10	76.43
20	33.00	20	73.41
30	34.90	30	72.47
50	36.19	50	69.36

真の化合物が見つかるヒット率
(AUC for true ligands)

真のヒット化合物のAUCと、UAPのAUCの相関

標的タンパク質はCOX2, AMPC, fXa, Thrombinなど11種類の水溶性タンパク質。
UAPは、GPCRリガンド、DUDの活性化化合物、蛋白質ーリガンド複合体PDBのリガンド
など、それぞれ100種類程度の化合物。

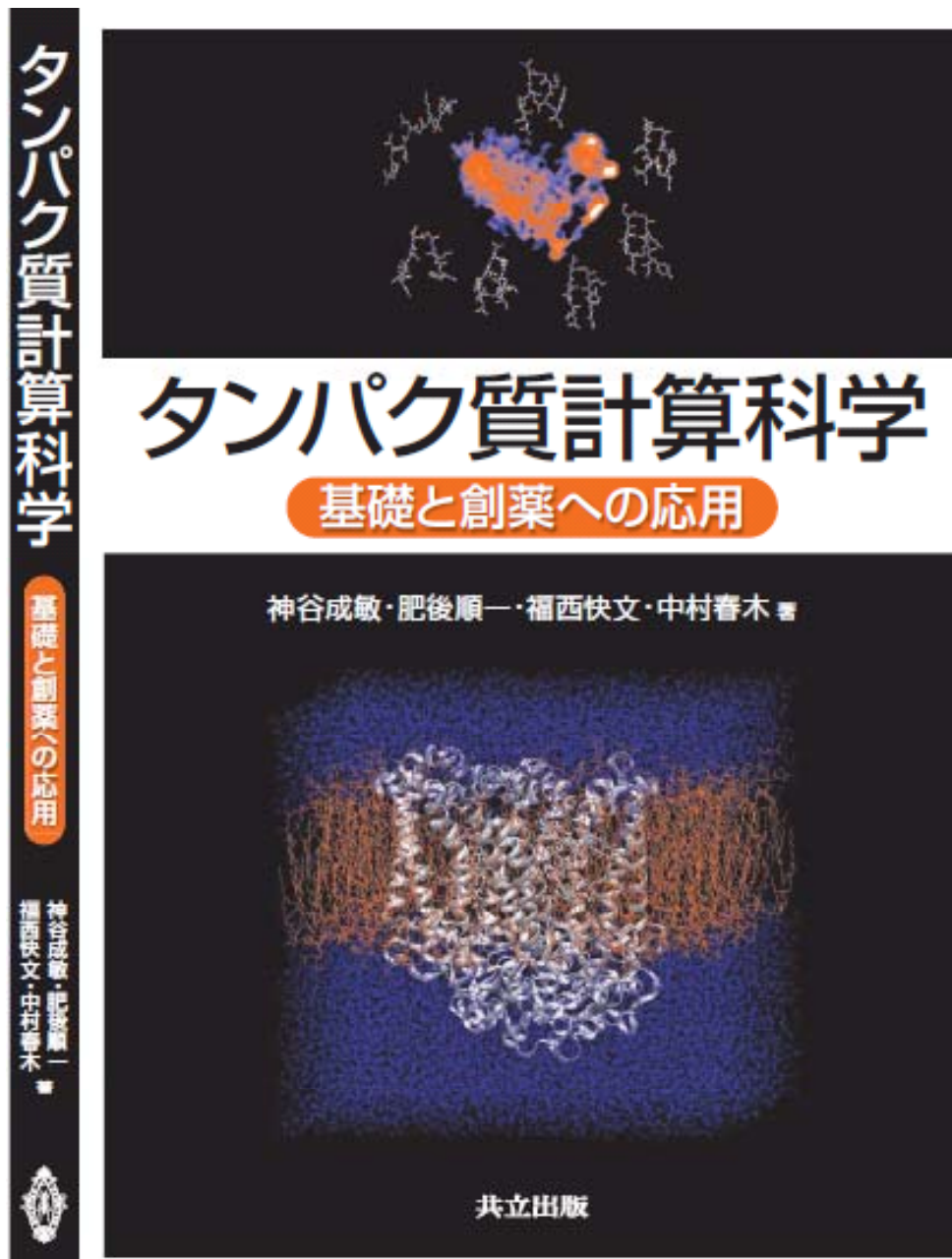
3つのUAPと、データベースに関する相関係数 (R) (GPCR, DUD, PDB)
3種類のUAPの和 (UAP_avg) が良い結果を与える。

UAP	Average Correlation coefficient (R)	R for lig0 decoy	R for c001	R for DUD
UAP_GPCR	0.6295	0.6981	0.5295	0.5834
UAP_DUD	0.5151	0.6173	0.3622	0.4477
UAP_PDB	0.4165	0.7468	0.6161	0.6347
UAP_avg	0.6608	0.7297	0.5723	0.6072



UAPを化合物データベースに混ぜておいて、UAPが化合物スクリーニング結果の上位にランクされる結果を選べば良い。

参考図書



共立出版
発売中
3900円(税抜き)

NEDO講座での内容をほぼ全てカバーし、付録CD-ROMには、演習で用いたテキスト及び演習用のデータを含む。

蛋白質構造と計算が主だが、

薬物ドッキング
薬物スクリーニング

をまとめて記載した邦書としては
おそらく初めて。

文献

- Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D.M. Predicting ligand binding to proteins by affinity fingerprinting. *Chemistry & Biology* **1995**, *2*, 107-118.
- Vigers, G.P.A.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J. Med. Chem.* **2004**, *47*, 80-89.
- “Similarity among receptor pockets and among compounds: Analysis and application to *in silico* ligand screening”, Y. Fukunishi, Y. Mikami, H. Nakamura, *Journal of Molecular Graphics and Modelling*, *24*, 34-45 (2005).
- “Classification of chemical compounds by protein-compound docking for use in designing a focused library”, Y. Fukunishi, Y. Mikami, K. Takedomi, M. Yamanouchi, H. Shima, H. Nakamura, *Journal of Medicinal Chemistry*, *49*, 523-533 (2006).
- “Multiple target screening method for robust and accurate *in silico* screening”, Y. Fukunishi, Y. Mikami, S. Kubota, H. Nakamura, *Journal of Molecular Graphics and Modelling*, *25*, 61-70 (2005).
- “A virtual active compound produced from the negative image of a ligand-binding pocket, and its application to *in-silico* drug screening”, Y. Fukunishi, S. Kubota, C. Kanai, H. Nakamura, *Journal of Computer-Aided Mol Design*, *20*: 237-248 (2006)
- “Finding ligands for G-protein coupled receptors based on the protein-compound affinity matrix”, Y. Fukunishi, S. Kubota, H. Nakamura, *Journal of Molecular Graphics and Modelling*, (2007) *25*:633-643
- “Noise reduction method for molecular interaction energy: application to *in silico* drug screening and *in silico* target protein screening”, Y. Fukunishi, S. Kubota, H. Nakamura, *Journal of Chemical Information and Modeling* *46*: 2071-2084 (2006)
- “An Efficient *in Silico* Screening Method Based on the Protein-Compound Affinity Matrix and Its Application to the Design of a Focused Library for Cytochrome P450 (CYP) Ligands””, Y. Fukunishi, S. Hojo, H. Nakamura, *Journal of Chemical Information and Modeling* (2006) *46*:2610-2622
- Yoshifumi Fukunishi, Structural ensemble in computational drug screening, *Expert Opin. Drug Metab. Toxicol*, vol 6, No 7, 1-15 (2010)

文献(日本語)

- 「ドッキングシュミレーションは創薬の標準手段」福西快文、最新医学 62 卷9月増刊号(2007)
- 「膜タンパク質の結晶化技術の新展開及び創薬バリューチェーンの紹介」井上豪、安達宏昭、村上聡、高野和文、松村浩由、森勇介、福西快文、中村春木、木下誉富、仲西功、奥野恭史、南方聖司、下条真司、坂田恒昭、YAKUGAKU ZASSHI 128(4) 497-505 (2008)
- 福西快文、中村春木、”タンパク質構造に指南された活性化合物評価—近年の薬物スクリーニング—”, 実験医学増刊号、2009, 27, 151-158.
- 創薬バリューチェーン編著「コンピューターで薬を創ろう」 化学同人 2009年6月25日
- 福西快文、中村春木、”タンパク質構造と活性化合物評価“、バイオテクノロジージャーナル(羊土社)1-2月号、122-125 (2007)
- 「創薬の効率を飛躍的に高めた化合物スクリーニング計算」シンセシオロジー, vol.2, No.1;60-68 (200)
http://www.aist.go.jp/synthesiology/vol02_01/vol02_01_p60_p68.pdf