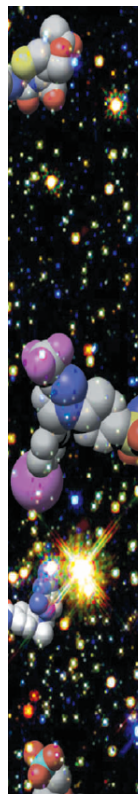
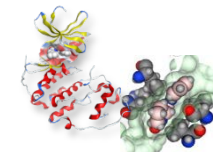


創薬における計算生命科学： インフォマティクスとシミュレーションを 融合したインシリコスクリーニング



理化学研究所
制御分子設計研究チーム
本間 光貴

低分子医薬品の候補は何個あるか



insight review articles

Navigating chemical space for biology and medicine

Christopher Lipinski¹ & Andrew Hopkins²

¹Pfizer Global R&D, Groton Laboratories, Eastern Point Road, Groton, Connecticut 06340, USA (e-mail: christopher_a_lipinski@groton.pfizer.com)

²Pfizer Global R&D, Sandwich Laboratories, Ramsgate Road, Sandwich, Kent CT13 9NJ, UK (e-mail: andrew.hopkins@pfizer.com)

Despite over a century of applying organic synthesis to the search for drugs, we are still far from even a cursory examination of the vast number of possible small molecules that could be created. Indeed, a thorough examination of all 'chemical space' is practically impossible. Given this, what are the best strategies for identifying small molecules that modulate biological targets? And how might such strategies differ, depending on whether the primary goal is to understand biological systems or to develop potential drugs?

The relationship between chemistry, biology and medicine has been remarkably productive over the past century, since Paul Ehrlich pioneered the idea of systematically searching for drugs. By screening just over 600 synthetic compounds, Ehrlich discovered arsphenamine (Salvarsan)¹, which greatly improved the treatment of syphilis. Researchers now routinely screen millions of compounds in the search for some that are biologically active. Yet even the compound files of the largest pharmaceutical companies (which typically contain approximately 10^6 compounds) offer only a cursory

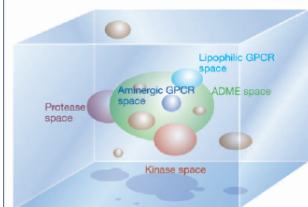
examination of all the possible organic compounds that comprise 'chemical space' (Box 1). Chemical space is for all practical purposes infinite and limited only by the chemist's imagination.

Not all biologically active compounds have the desired physicochemical properties to be a drug. A biologically active compound may be too lipophilic (greasy) to be orally absorbed, too polar to cross the gastrointestinal wall or may have too much vulnerable chemical functionality that can be attacked by metabolizing systems in the liver, and therefore not remain intact for long enough to have a useful *in vivo* effect.

Box 1

Chemical space

Chemical space can be viewed as being analogous to the cosmological universe in its vastness, with chemical compounds populating space instead of stars. For example, there are more than 10^{26} possible derivatives of *n*-hexane — if we use a list of only 150 substituents and consider mono- to 14-substituted hexanes². However, not all theoretically postulated compounds fall within the limits of what is synthetically feasible to produce, even with our current, extensive knowledge of organic chemistry. To navigate the vast diversity of chemical space, the concept of 'chemography', which is akin to a global positioning system, has been proposed. This involves mapping compounds onto coordinates of chemical descriptors of various physicochemical or topological properties^{3,4}. Given the vastness of chemical space, the challenge for chemical



Box 1 Figure The figure depicts a cartoon representation of the

100万 1兆 10^{60}

網羅的に計算すると 10^{60} 個程度は
あるらしいです
すごい数ですね！！

Lipinsky, Nature (2004), 432(16), 855

ちなみに

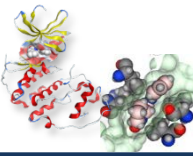
観測可能な宇宙にある「星」（恒星）の数
1000億個の銀河 * 1000億個の星 = 10^{22} に近いオーダー

観測可能な宇宙に陽子をすきまなく埋めた場合の数
 10^{60} に近いオーダー

ただし実際にこれまで人間が天然・人工で

入手した化合物は $10^7 \sim 10^8$ 個程度

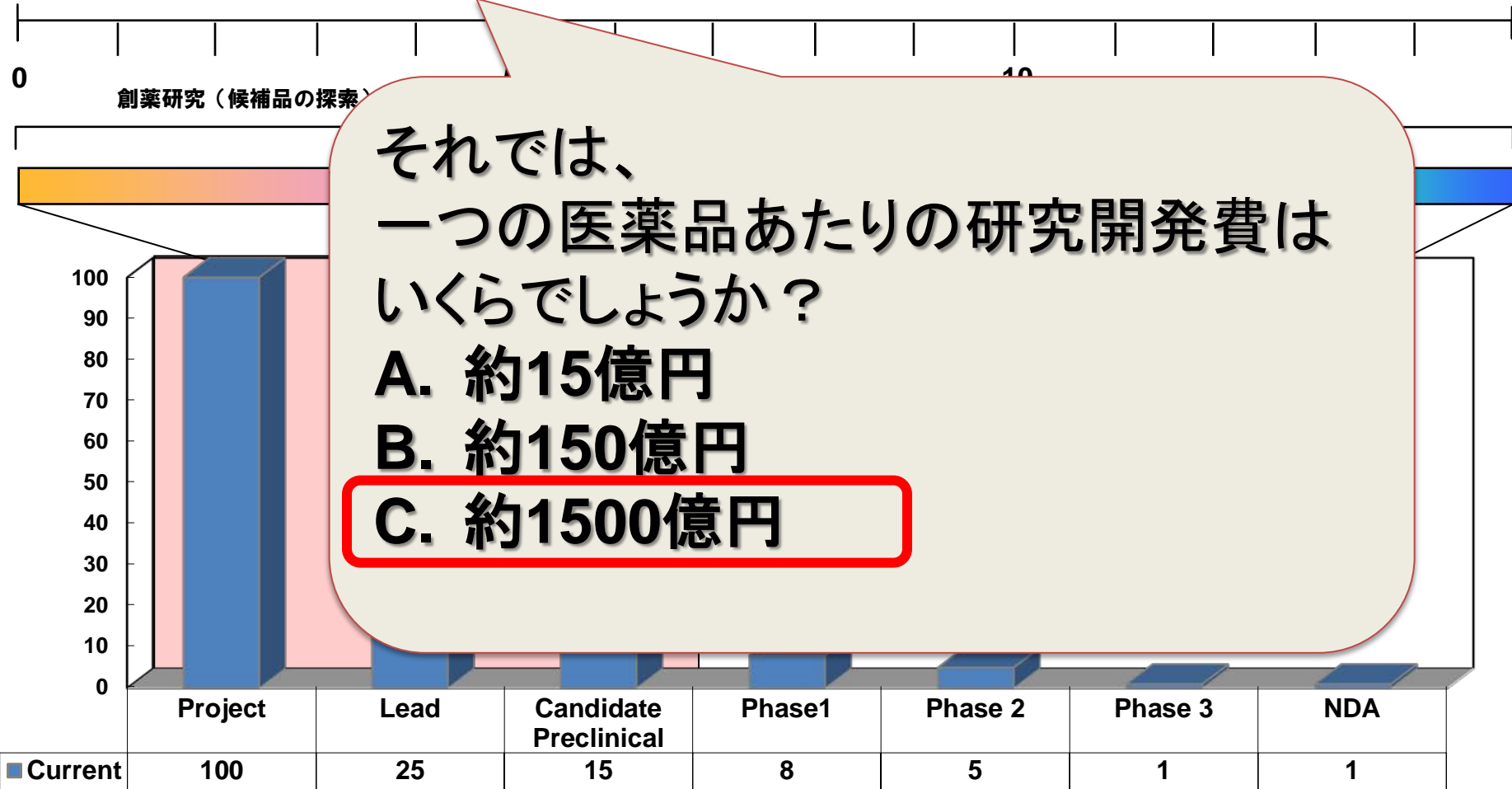
「創薬」のプロセスと成功率



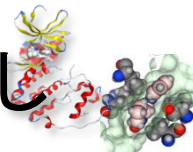
アイディア

12-15年 ○○億円 (多くの国で売れるブロックバスター狙いの場合、失敗込み)

医薬品

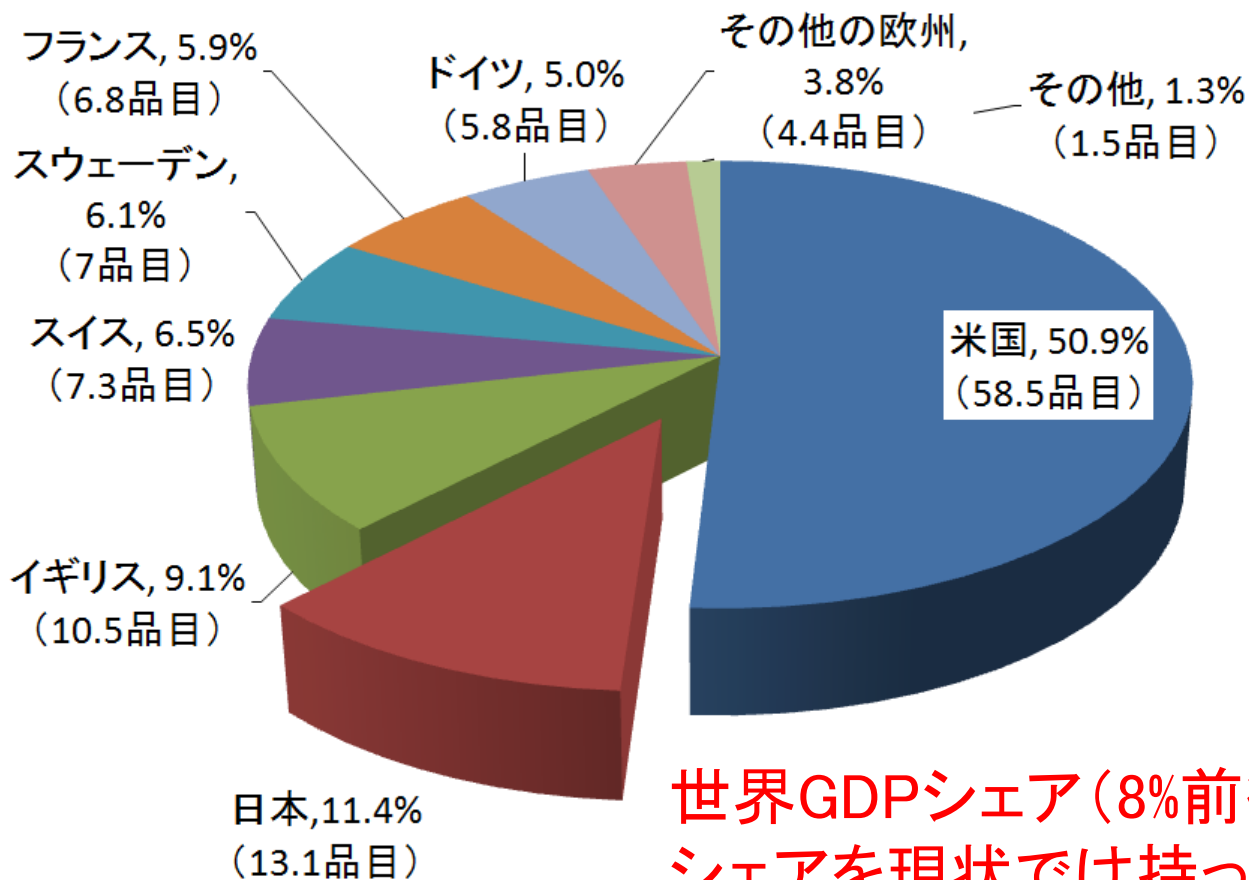


一般に開発候補品を出すまでが創薬「研究」の仕事である
ただし、開発候補品まで行っても、実際に薬になるのは15個のうち、たった1個。



国際市場における創薬シェア

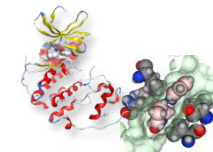
Best in Class 医薬品(115品目)の創出国



※同様に、ロケットや
戦闘機の開発なども
物差しになる

**世界GDPシェア(8%前後)以上の
シェアを現状では持っている**

凋落の兆し



国際市場における画期的新薬創出推移

Country	1970s		1980s		1990-2002	
U.S.A.	22.3	26.9%	38.3	34.5%	31.4	53.3%
Japan	14.0	16.9%	32.0	28.8%	5.1	8.6%
France	9.0	10.8%	8.8	8.0%	3.2	5.4%
Germany	8.0	9.6%	5.7	5.2%	1.6	2.7%
Italy	11.0	13.3%	6.0	5.4%	2.0	3.4%
Switzerland	6.0	7.2%	4.8	4.3%	2.6	4.3%
U.K.	3.0	3.6%	3.3	2.9%	4.0	6.8%
Others (in EU)	5.0	6.0%	9.5	8.5%	2.7	4.5%
Others	4.7	5.6%	2.7	2.4%	6.5	11.0%
Total	83	100%	111	100%	59	100%

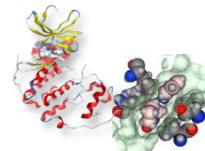
凋落の傾向

2010年問題

多くの80年代の日本発の医薬品が特許切れをむかえている

(医薬産業政策研究所 政策研ニュースNo.29 2010年1月)

インシリコスクリーニング・設計の流れ



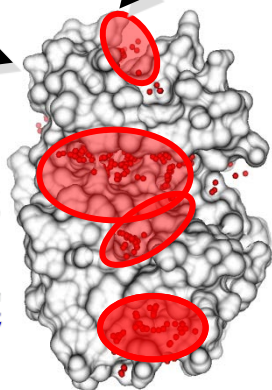
情報収集・分析



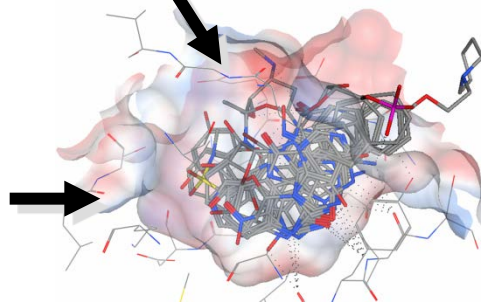
あるターゲットについての依頼が来る

1. ターゲットに関する情報収集

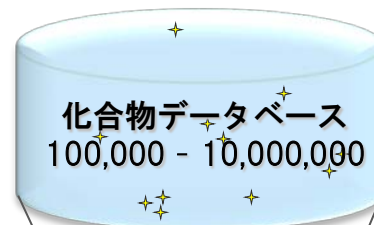
2. 薬物結合部位の分析と戦略の立案



3. ドッキング条件の検討

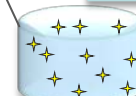


4. ドッキングによるインシリコスクリーニング



化合物データベース
100,000 - 10,000,000

ヒット探索

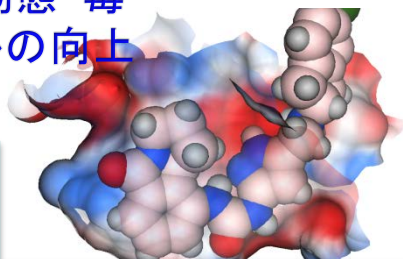


インシリコ選択化合物
100-1,000

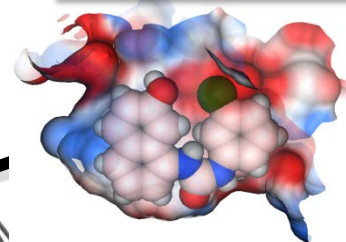
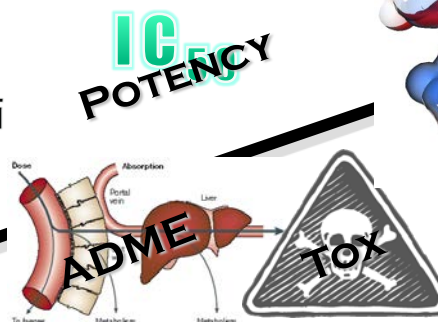
5. インシリコ設計による活性・体内動態・毒性プロファイルの向上

構造解析 → 設計 → 合成 → 評価

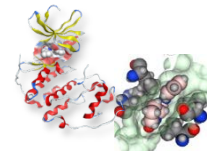
化合物の最適化



医薬候補品・ケミカルツール



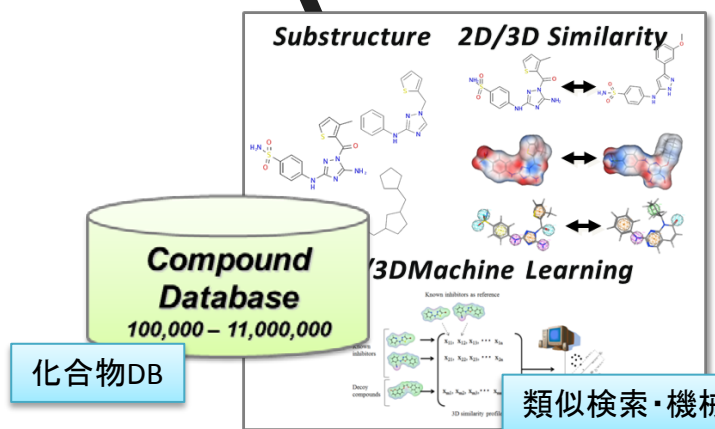
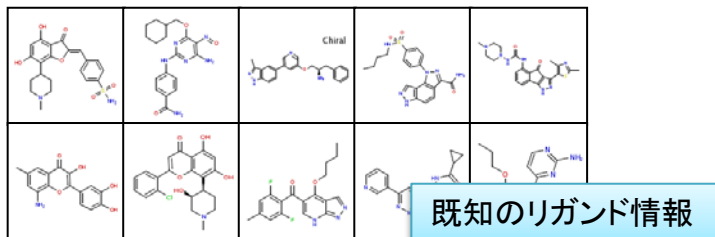
LBDD vs SBDD



LBDD

Ligand-based drug discovery(design)

既知のリガンド情報に基づいて
医薬品候補を探索・設計する方法

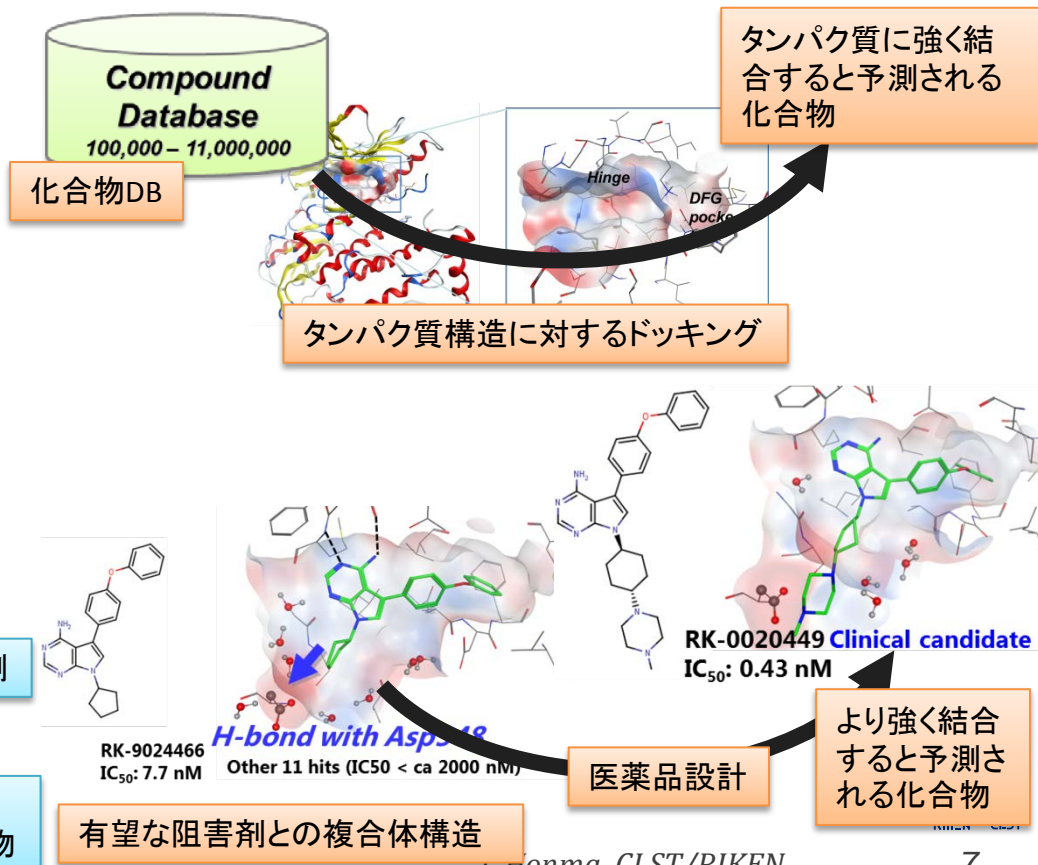


既知のリガンドと似ていることから、
同じ活性を持つ可能性も高い化合物

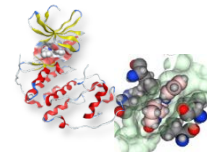
SBDD

Structure(receptor)-based drug discovery(design)

タンパク質構造に基づいて
医薬品候補を探索・設計する方法



LBDD vs SBDD in Pfizer



Sandwich, UK

必要性を感じない。
まずは、LBDD用
に進めてはどう？

10年前にファイザー
名古屋とファイザー
UKとの間で交わされ
た電話会議での一幕

Pfizerの化合物*のドッ
キング用前処理済み
データを皆で準備して
共有しませんか？

- **Christopher A. Lipinski**
(Rule of Five)
- **Andrew Hopkins**
(Ligand Efficiency, Bayesian Profiling)
- **Marcel De Groot**
(Ensemble Model)

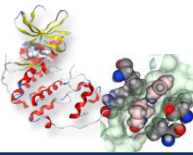
ファイザーUKには、リガンド情報の
大家がたくさん….



Nagoya, Japan

*実体サンプル数百万個、登録された反応で合成容易な数億個

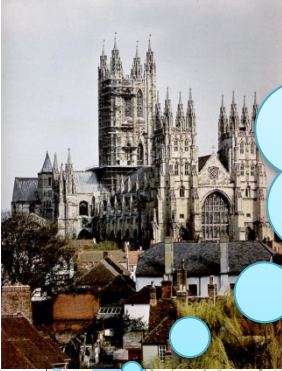
LBDD vs SBDD in Pfizer



SBDDは時間がかかる割には、効果が薄い。LBDDを優先的に進めるべき

ケースバイケースで両方使えば良い

我々は独自のドッキングソフトも開発しているし、成功例も多い。



Sandwich, UK



Chicago, US
Ex-Warner-Lambert



Nagoya, Japan



Groton, US
Original Pfizer

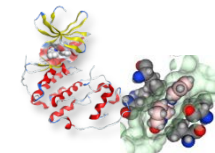


Boston, US



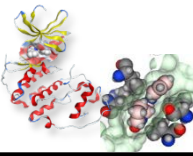
La Jolla, US
Ex-Agouron

LBDD vs SBDD



	利点	欠点
SBDD	<ul style="list-style-type: none">結合サイト、結合モードが異なる多様な阻害剤を得ることができる物理法則に従った精密な自由エネルギー予測が可能	<ul style="list-style-type: none">レベルにもよるが計算時間が非常に長い設定を決めるのに熟練を要する設定*を間違えた場合、完全な失敗に終わることもある <p>*: 不適切なタンパク質構造、不適切な計算方法</p>
LBDD (3D)	<ul style="list-style-type: none">計算時間が短いヒット率はむしろSBDDよりも高い2Dよりもやや異なる骨格も見つけ出せる	<ul style="list-style-type: none">元の情報(query)と同じ結合サイト、結合モードの化合物しかヒットしない
LBDD (2D)	<ul style="list-style-type: none">計算時間が非常に短いヒット率はむしろSBDDよりも高い設定がシンプルなので、失敗は少ない	<ul style="list-style-type: none">元の情報(query)に近い化合物しかヒットしない

Outline



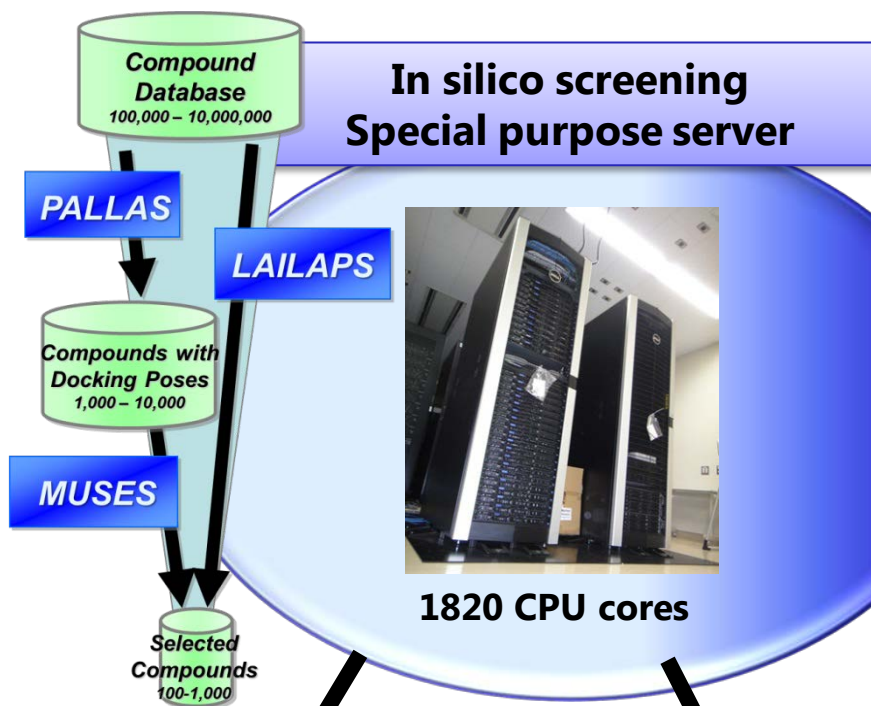
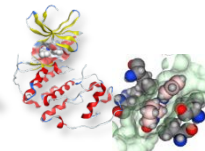
1. インシリコスクリーニングの技術開発

2. インシリコスクリーニング・設計の実例

3. 結合親和性予測／FMOの創薬活用

4. 体内動態・毒性の予測に関して

研究室で利用している計算機・データベース



K computer
Japanese governmental Supercomputer,
One of the fastest servers in the world
800K CPU cores, 10P FLOPS



RIKEN Wako RICC
RIKEN supercomputer
~30K CPU cores

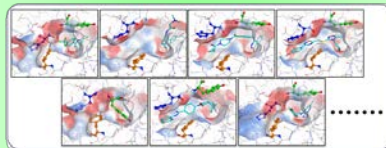
Compound Database

Preprocessed for in silico screening

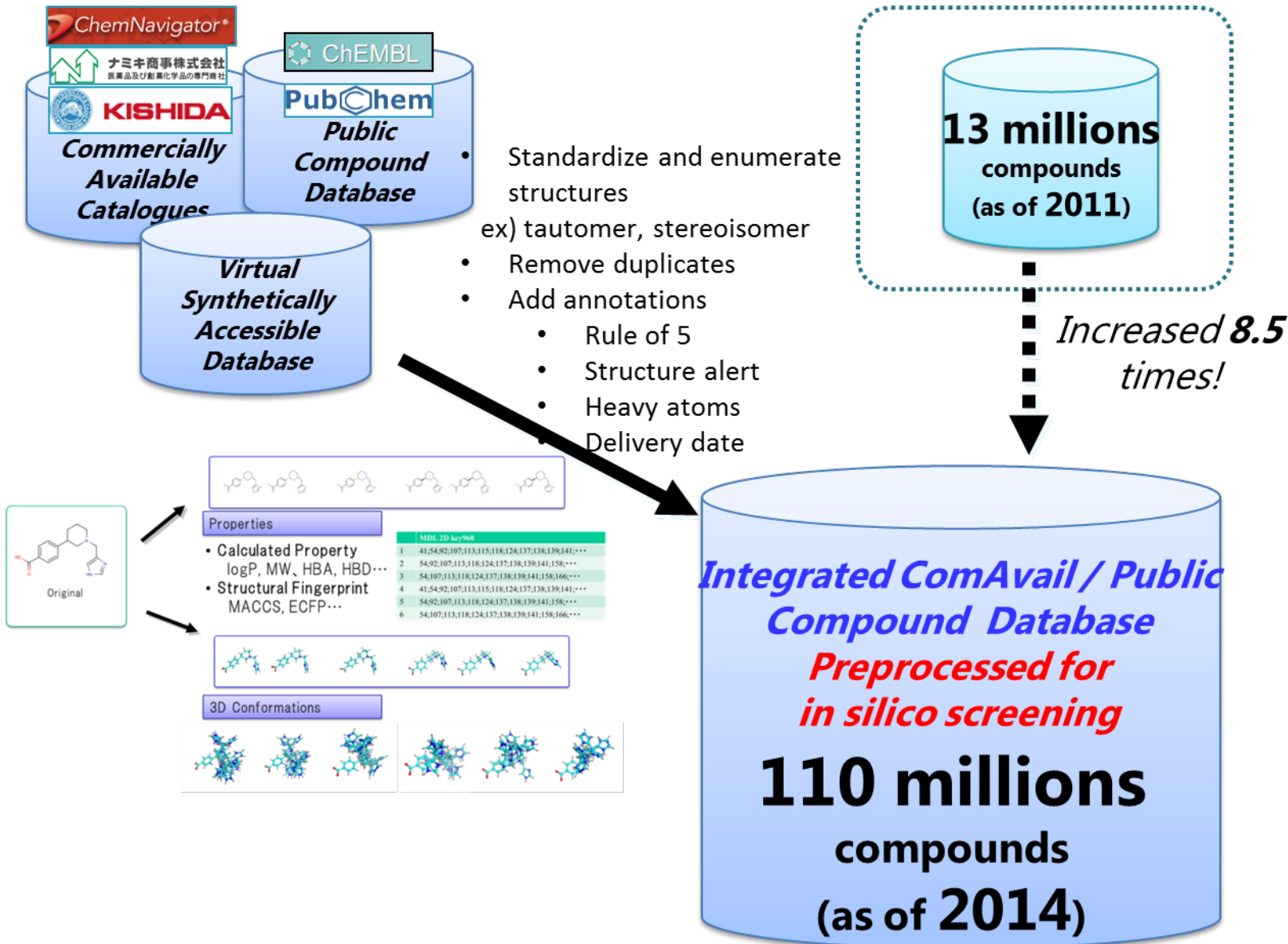
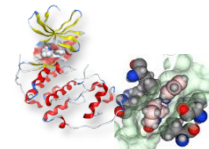
UT library: **200K cpds**

ComAvail, Literature: **110M cpds**

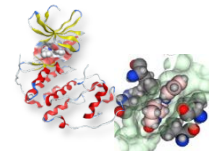
Protein Database



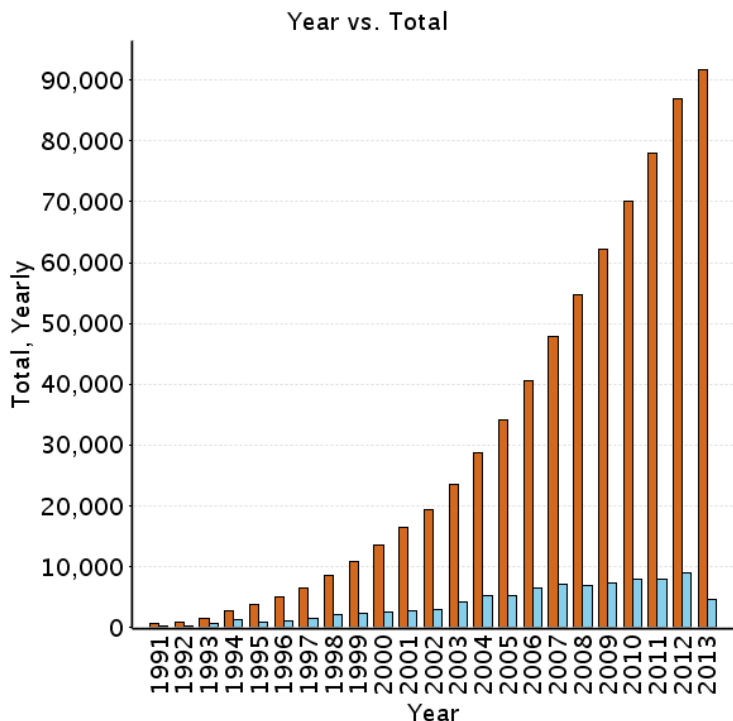
1億1千万化合物のインシリコスクリーニング用データベース



公共のタンパク質構造・アッセイ値データ



公開されている化合物・アッセイ結果データベース



- ChEMBL (2013.06.27)
 - Unique Compounds: 1,295,510
 - Assays (Targets): 9,844
 - Activities: 11,420,351

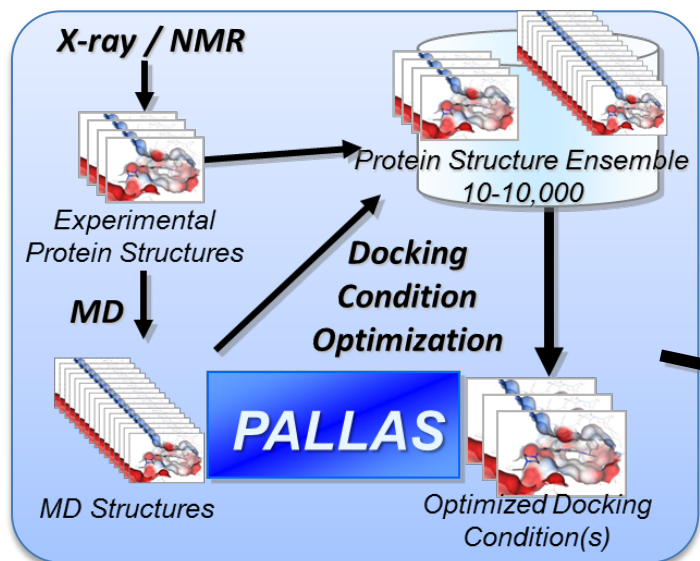


- PubChem (2013.06.27)
 - Unique Compounds: 47,659,498
 - Assays: 683943

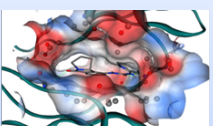
PDBに登録されたタンパク質構造数
2013年6月現在

<http://www.pdb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>

これらの双方のデータを最大限活用できる
手法の開発が必要！



- 1) *Bioorg. Med. Chem.*, 2012, 20(12), 3756
- 2) *J Mol Biol*, 2012, 417(3), 240



MUSES

Atom-based MM-GBVI energy descriptor

	VAL18Cγ	Lys23NC (100-200)
aPLIED	0	-6.621	
MMGBVI(w/o), GBVI(w/o)	-0.502	-0.356	
	0.040	6.116	

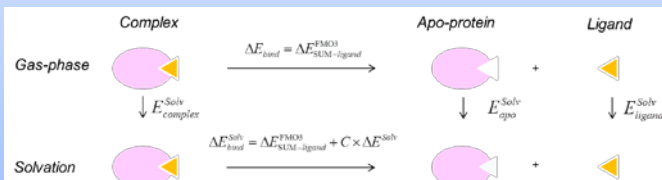
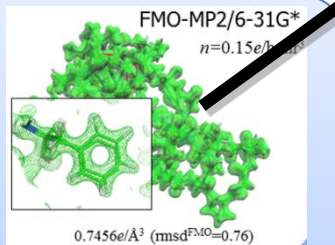
Machine Learning

Machine Learning Prediction Models

Protein-Ligand Interaction Analysis

Affinity Prediction

FMO-PBSA



J Chem Inf Model, 2010, 50(1), 170

Compound Database

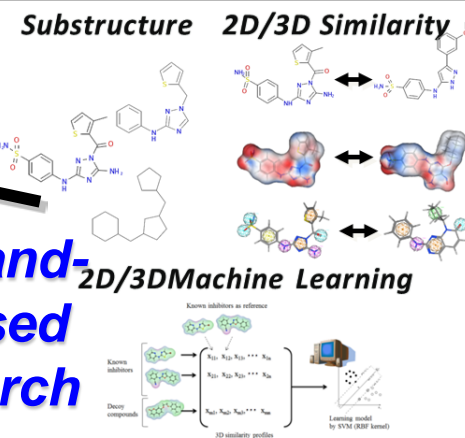
100,000 – 11,000,000

Docking

Compounds with Docking Poses

1,00 – 10,000

Ligand-based Search

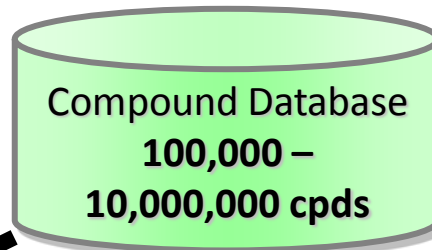
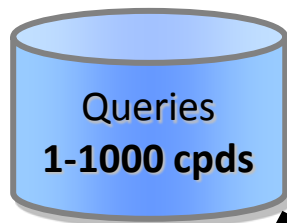
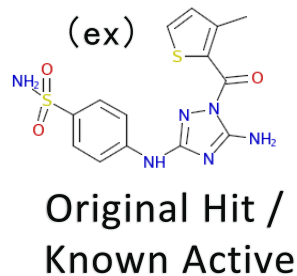
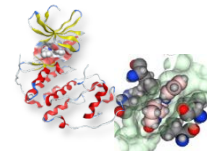


J Chem Inf Model, 2012, 52(4), 1015

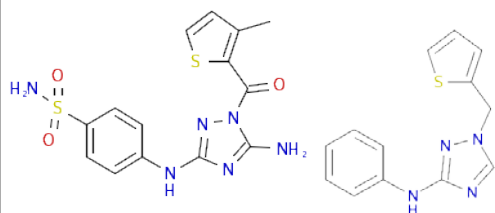
Selected Compounds

100-1,000

LAILAPSのしくみ

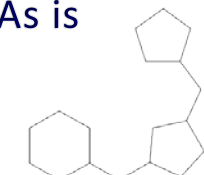


Substructure



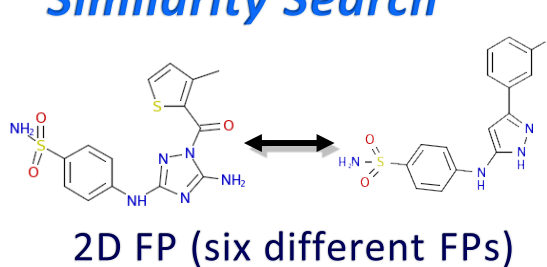
As is

Murcko
framework



Carbon Skelton
(in future)

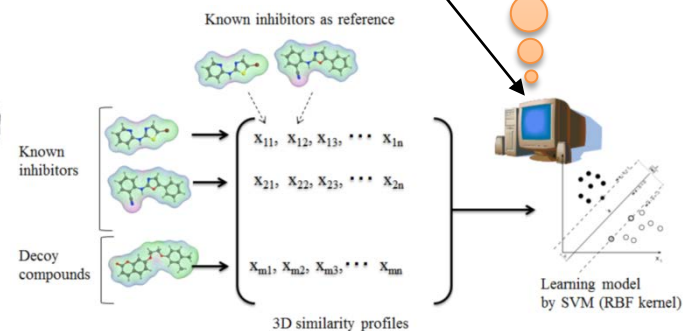
Similarity Search



3D shape and color
(ROCS)

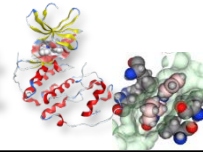
Machine Learning

Machine Learning
Sato, Honma,
J Chem Inf
Model (2012)



Machine Learning by 3D desc.
(in future)

Output (part of the scores are normalized)

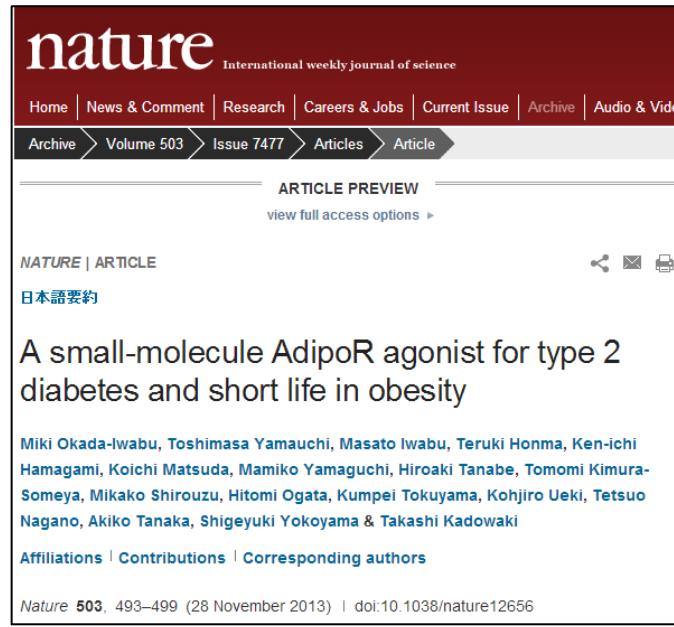
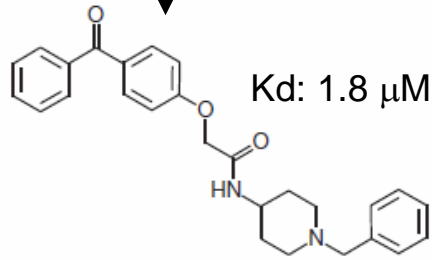
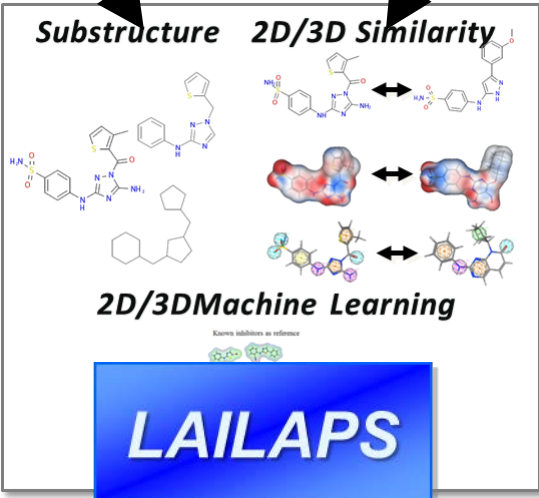


Adiponectin Receptor Agonist

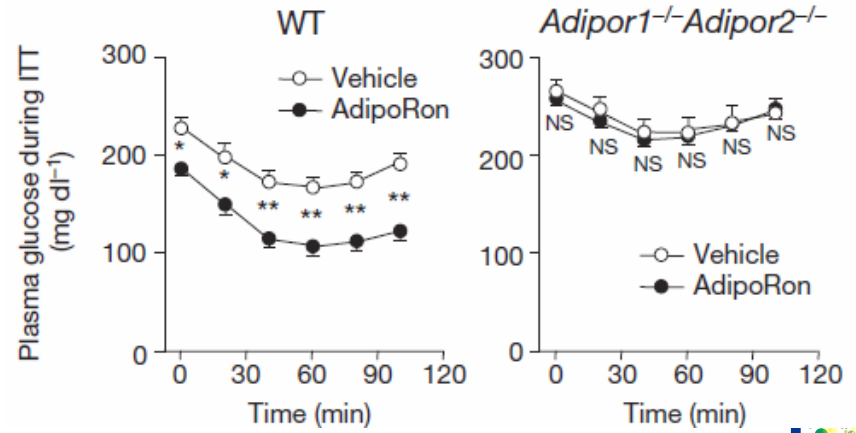
Target: **AdipoR**
7 TM Membrane receptor
Type 2 diabetes etc.

Known AMPK activator, Membrane protein ligands

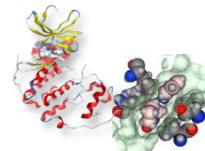
Compound Database
Preprocessed for in silico screening
Univ Tokyo library: 200K cpds
ComAvail, Literature: 3M cpds



Iwabu et al, Nature, 2013, 503, 493-499



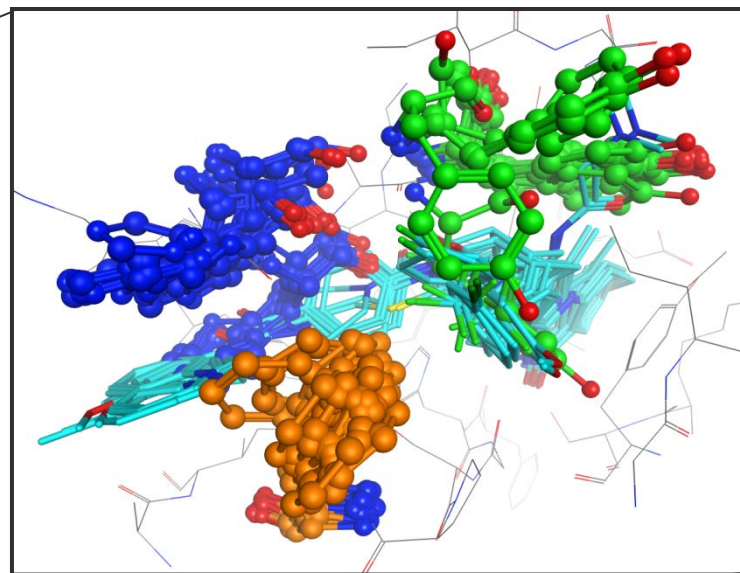
タンパク質の構造はフレキシブル



ターゲット

```
MHHHHHSSGVDLGTENLYFQSM LNTSGSGTI LIDLSPDDKEFQSV EEMQSTVREHRDGGHAGGI FNRY  
NILKI QKVCNKKLWERYTHR RKEVSEENHNHANERMLFHGSPFVNAI IHKGFDERHAY IGGMFGAGIY FA  
ENSSKSNQYVYVIGGGT GCPVHKDRSCYICHRQLLFCRVTLGKSF LQFSAMKMAHSPPGHHSVTGRPSVN  
GLALAEYVIYRGEQAYPEYLITYQIMRPEG
```

構造解析



拡大すると、主鎖だけではなく、3個の残基の側鎖が大きく動いており、リガンドと衝突するものも多い。

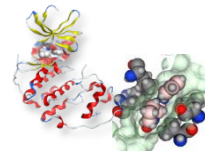
このような場合に、
どうしたら良いで
しょうか??



主鎖が大きく動いている

あるターゲットで26個の様々なリガンドが結合したX線構造

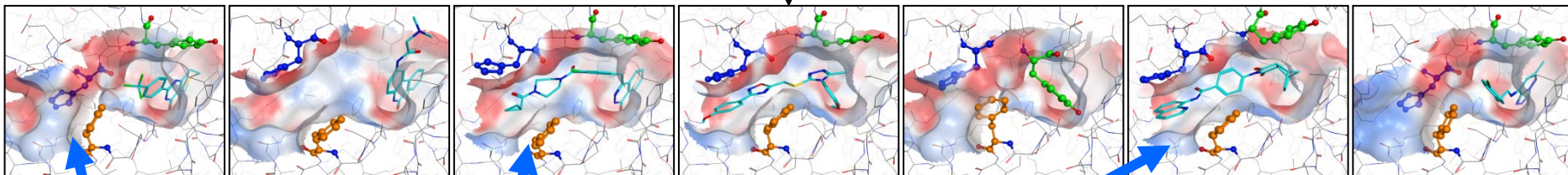
従来のインシリコスクリーニング



ターゲット

```
MHHHHHSSGVDLGTENLYFQSM LNTSGSGTI LIDLSPDDKEFQSV EEMQSTVREHRDGGHAGGIFNRY  
NILKI QKVCNKKLWERYTHRRKEVSEENHNHANERMLFHGSPFVNAI I HKGFDERHAY IGGMF GAGIYFA  
ENSSKSNQYVYVIGGGT GCPVHKDRSCYICHRQLLFCRVTLGKSF LQFSAMKMAHSPPGHHSVTGRPSVN  
GLALAEYVIYRGEQAYPEYLITYQIMRPEG
```

構造解析



最初のX線構造が
出たので、1個だ
け使って、早速
やってみよう

解像度が良かったり、
活性の高い阻
害剤が入っている
ものを使うべきか

うまく阻害
剤が取れ
ない...

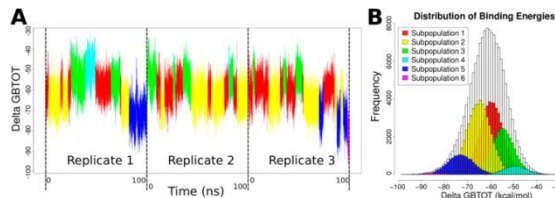
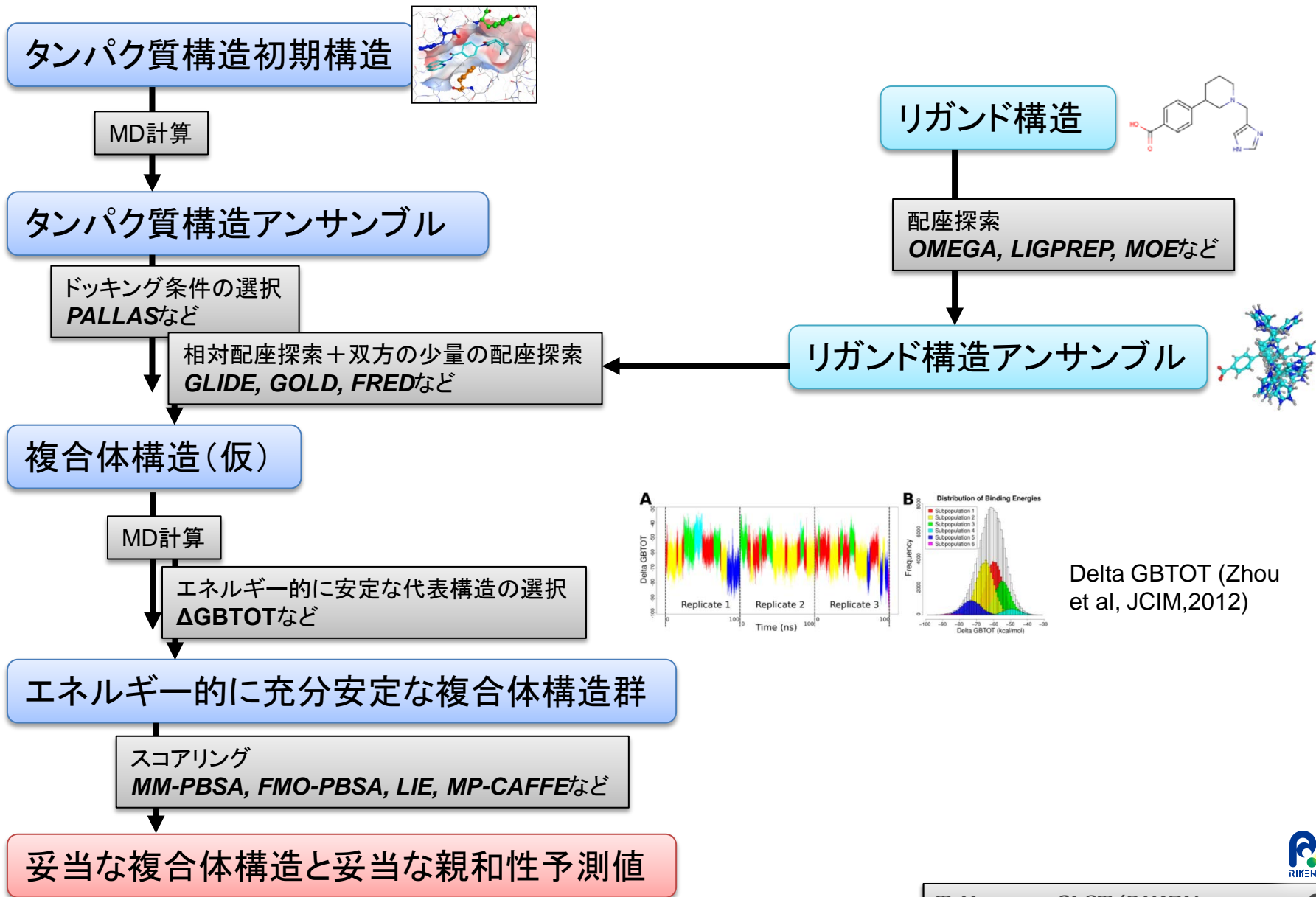
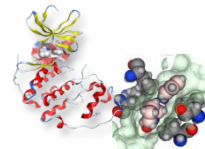
うまくいくか
は運しだい

結構な
頻度で...

せっちな人

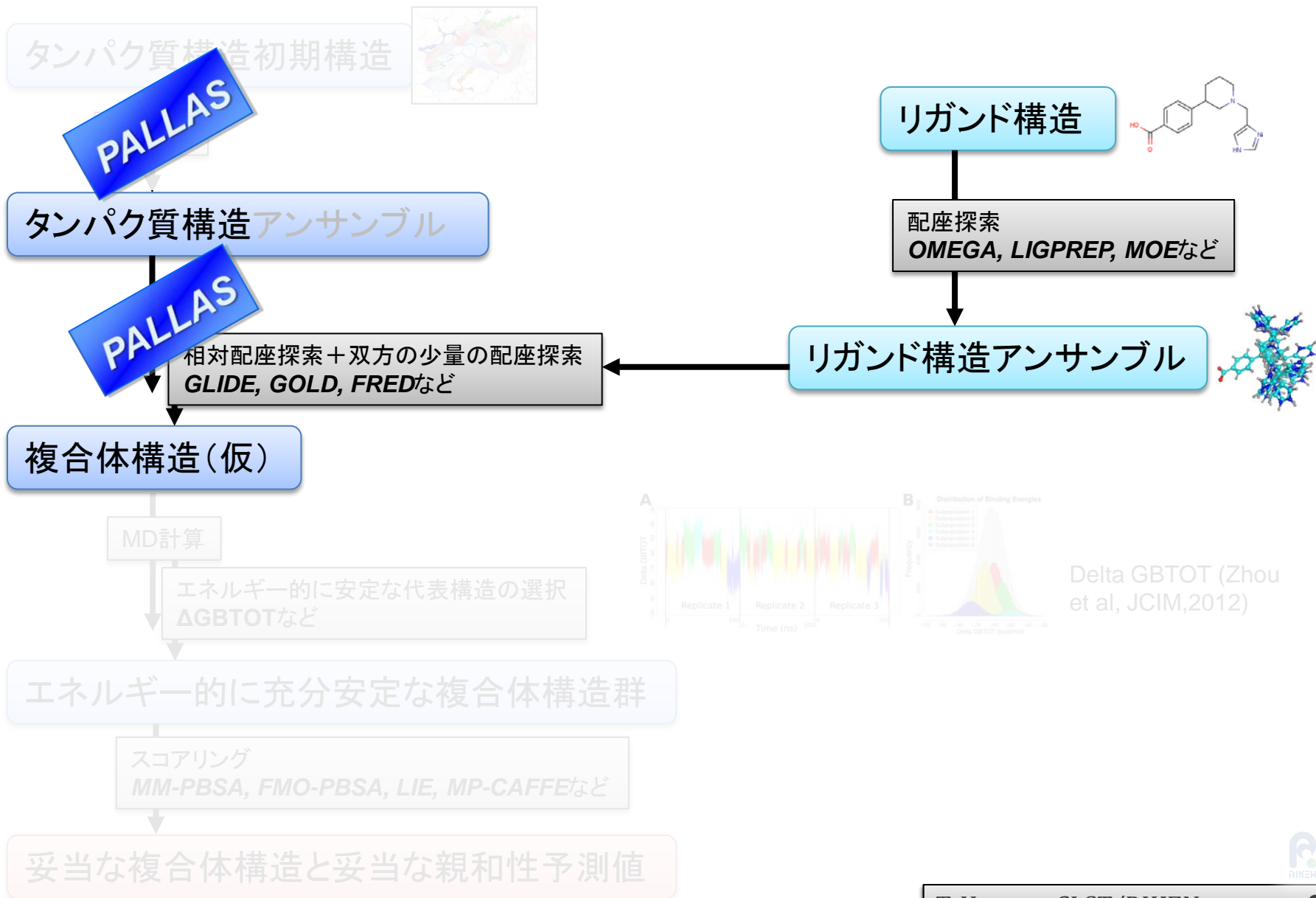
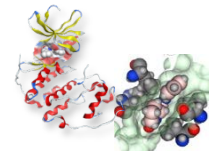
やや慎重な人

“妥当な”ドッキングに必要なこと

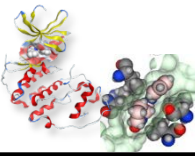


Delta GBTOT (Zhou et al, JCIIM, 2012)

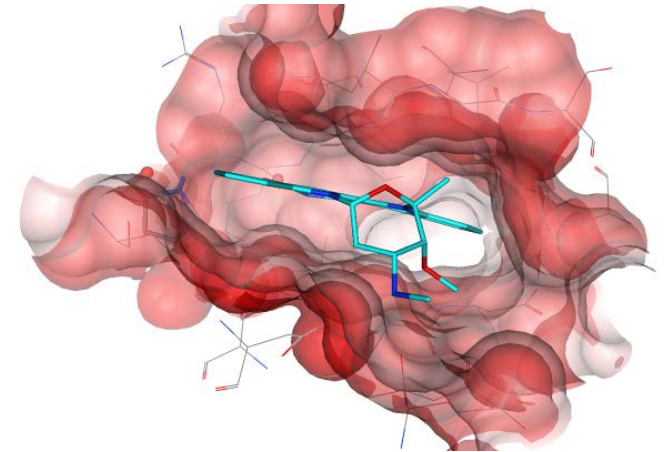
“妥当な”ドッキングに必要なこと



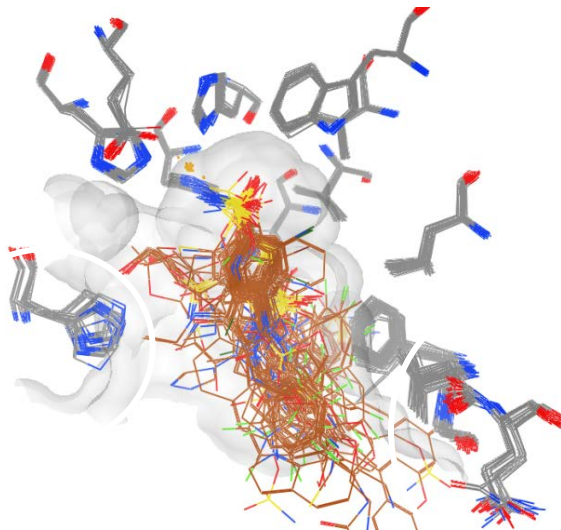
ドッキングの条件



- タンパク質構造 (配座)
- ドッキングアルゴリズム
- van der Waals表面のスケーリング
- 結合サイトの領域の定義
- 必須ファーマコフォアの拘束条件
- ドッキングポーズのスコア関数



van der Waals表面のスケーリング

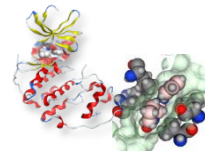


タンパク質の配座

ドッキングアルゴリズム

- リガンド配座
 - 最初から用意
 - タンパク質中で側鎖の配座を探索
- 初期ポーズの発生
 - ファーマコフォアのPotential Energy Gridへの配置
 - 形状相補性による配置
- 初期ポーズから最終ポーズへの探索・最適化
 - 力場上を利用した最適化
 - taboo / genetic algorithmによるポーズ探索

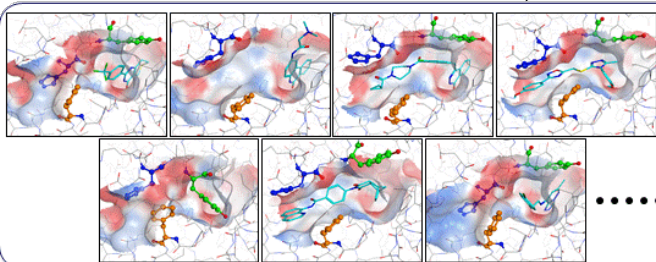
半自動ドッキング条件検討システム PALLAS



ターゲット

```
MHHHHHSSGVDLGTENLYFQSM LNTSSGGTILIDLSPDDKEFQSV EEMQSTVREHRDGGHAGGIFNRY
NILKI QKVCNKKLWERYTHRRKEVSEENHNHANERMLFHGSPFVNAI I HKGFDERHAY I GGMFGAGIYFA
ENSSKSNQYVYVIGGGT GCPVHKDRSCYICHRQLLFCRVTLGKSLFQFSAMKMAHSPPGHHSVTGRPSVN
GLALAEYVIYRGEQAYPEYLIT YQIMRPEG
```

しっかりと検証のステップを踏もう！



X線・NMR構造群

十分な構造数がない場合には、MDシミュレーションを行う

MDスナップショット

タンパク質構造アンサンブル

20-1000構造程度の構造を用いて「網羅的に」ドッキング条件の検証を行う

ドッキング結果をX線ポーズ再現率、EF, ROCスコア等で評価。2-3種類の組み合わせも検討する。

ターゲットまたは類縁タンパク質の既知阻害剤*

Reproducibility of X-ray ligand poses

RMSE values between X-ray ligand poses and docking poses are calculated for all docked ligands.

Reproducibility = $\frac{\# \text{ of ligands with RMSE} \leq 2.0 \text{ \AA}}{\# \text{ of all ligands}}$

Enrichment Factor (EF)

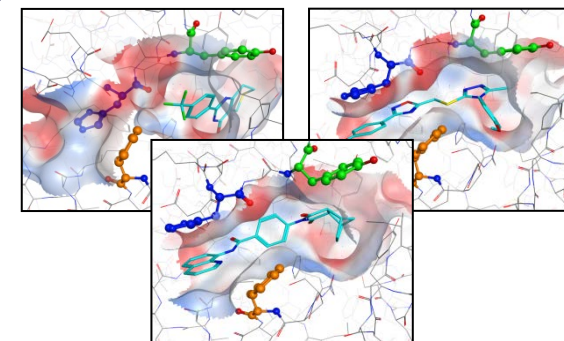
EF shows improvement of the hit rate by docking compared with that of random selection.

X% Enrichment Factor (EF_{X%}) = $\frac{\# \text{ of active compounds in top X\% score}}{\# \text{ of active compounds in X\% random selection}}$

ROC score

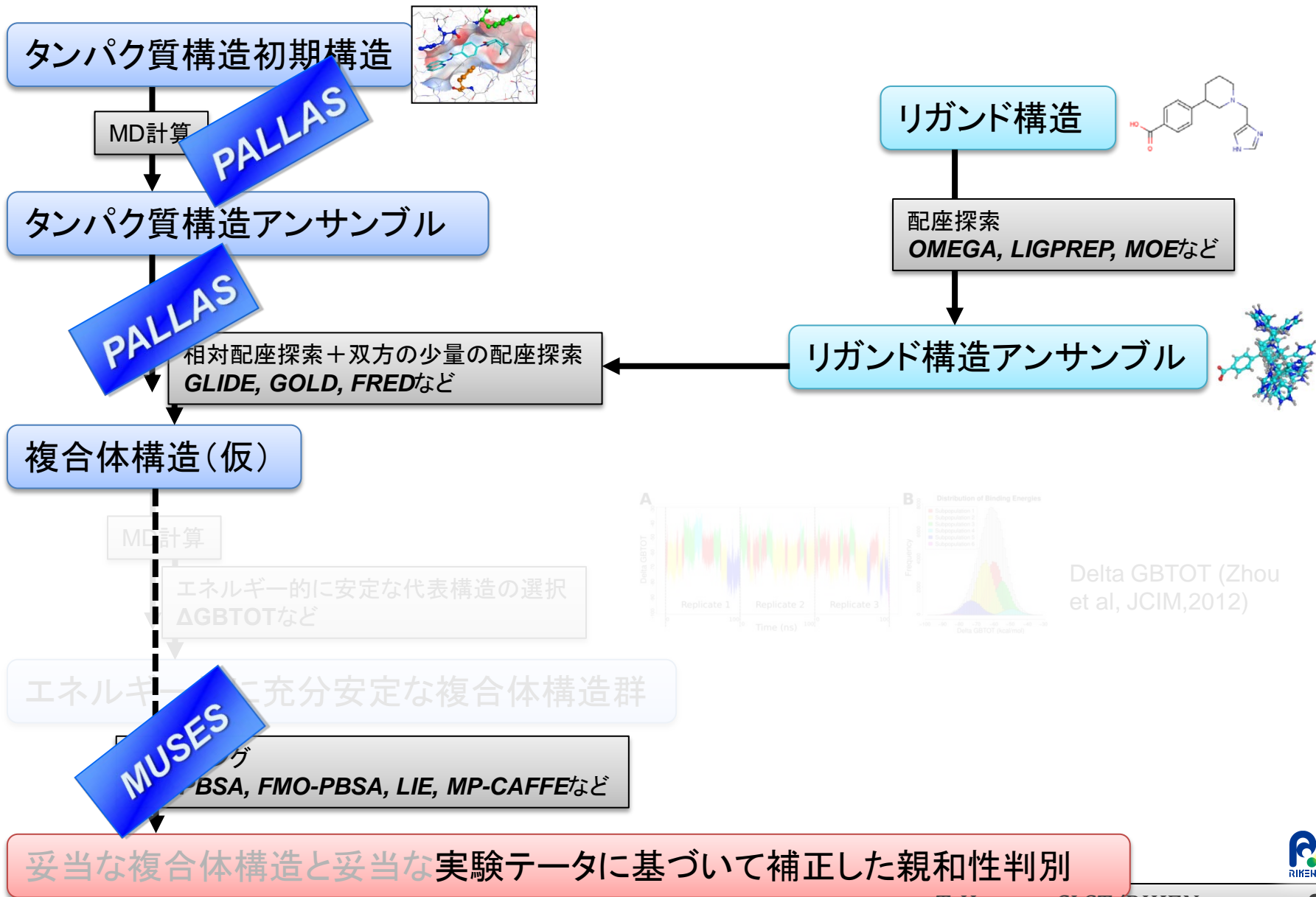
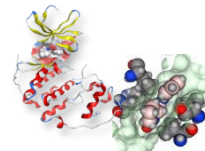
Using ROC (Receiver operating characteristic) plot, ROC AUC (ROC score) is calculated.

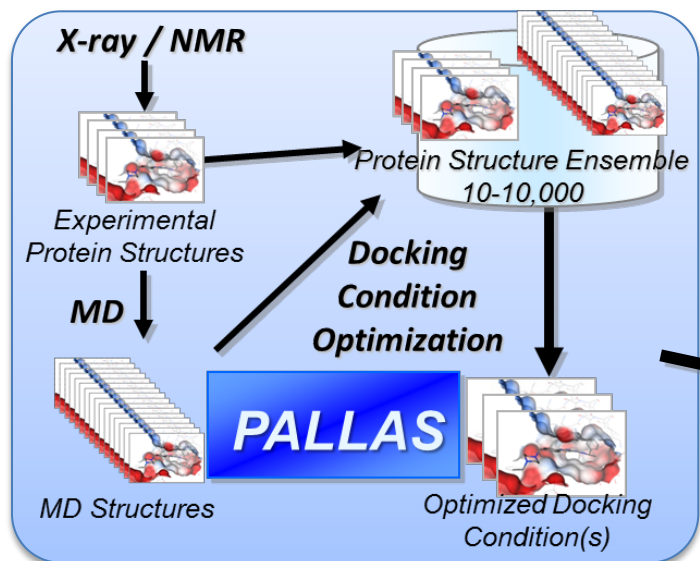
ROC score (ROC AUC) = $\frac{\text{Area under the ROC curve}}{\text{Maximum area of ROC plot (ideal outcome)}}$



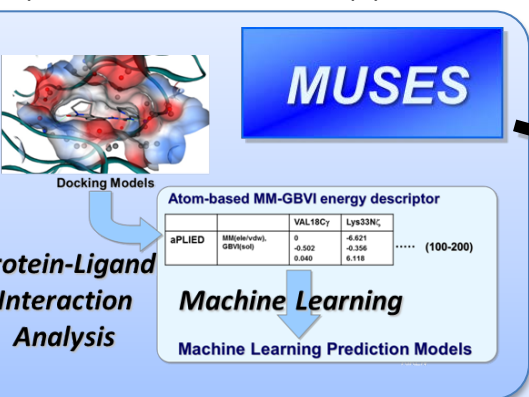
選ばれた最適ドッキング条件 (タンパク質構造及びドッキング設定)

“妥当な”ドッキングに必要なこと

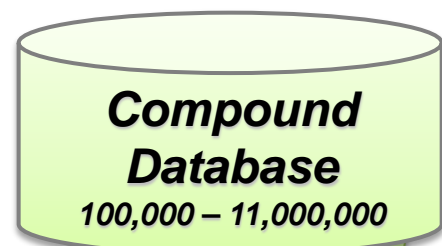
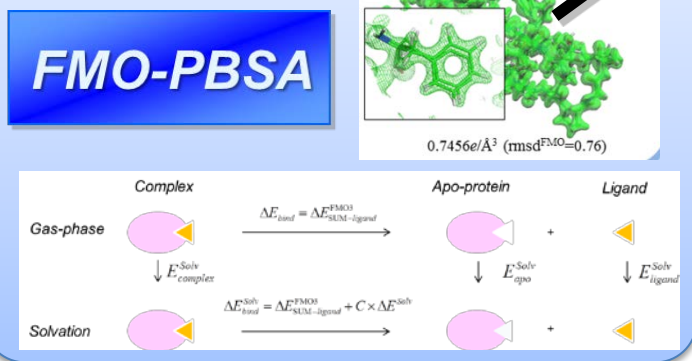




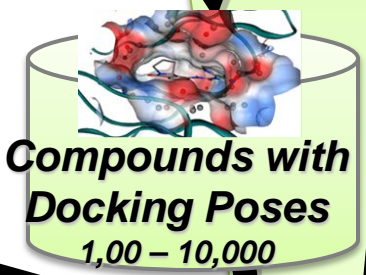
- 1) *Bioorg. Med. Chem.*, 2012, 20(12), 3756
- 2) *J Mol Biol*, 2012, 417(3), 240



J Chem Inf Model, 2010, 50(1), 170

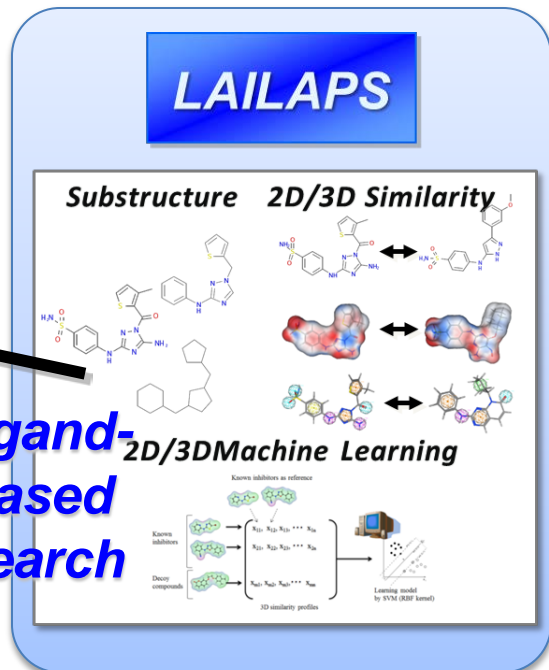


Docking

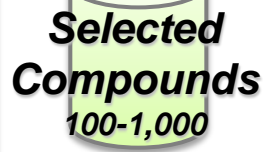


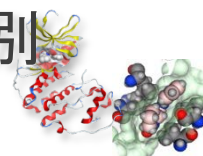
Affinity Prediction

Ligand-based Search



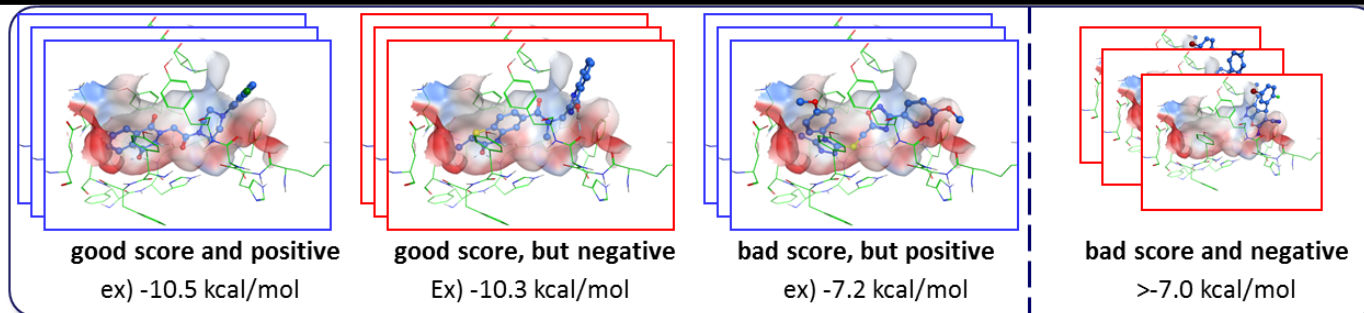
J Chem Inf Model, 2012, 52(4), 1015



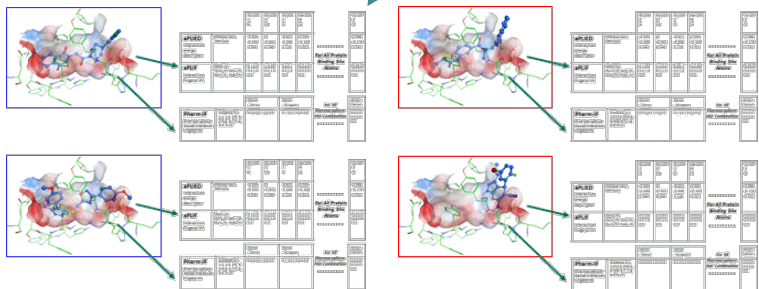


MUSES

X-ray Structures and Docking Results



Analysis of Interactions

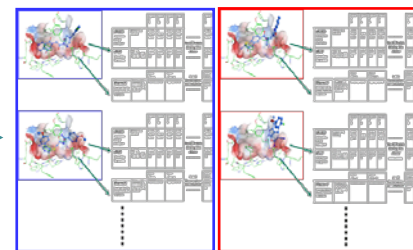


Calculate Interaction Descriptors for All POSITIVES

Calculate Interaction Descriptors for All NEGATIVES

Protein-Ligand Interaction Data of POSITIVES and NEGATIVES

Divide into "Training" and "Test" Sets



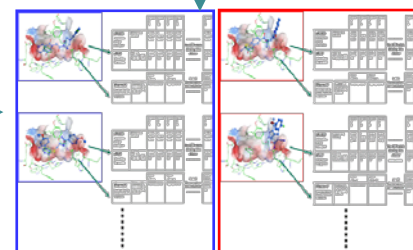
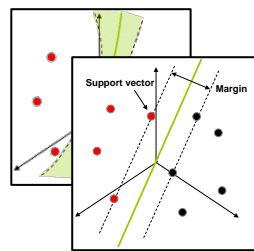
Training Set

Machine Learning by

- Support Vector Machine
- Random Forest

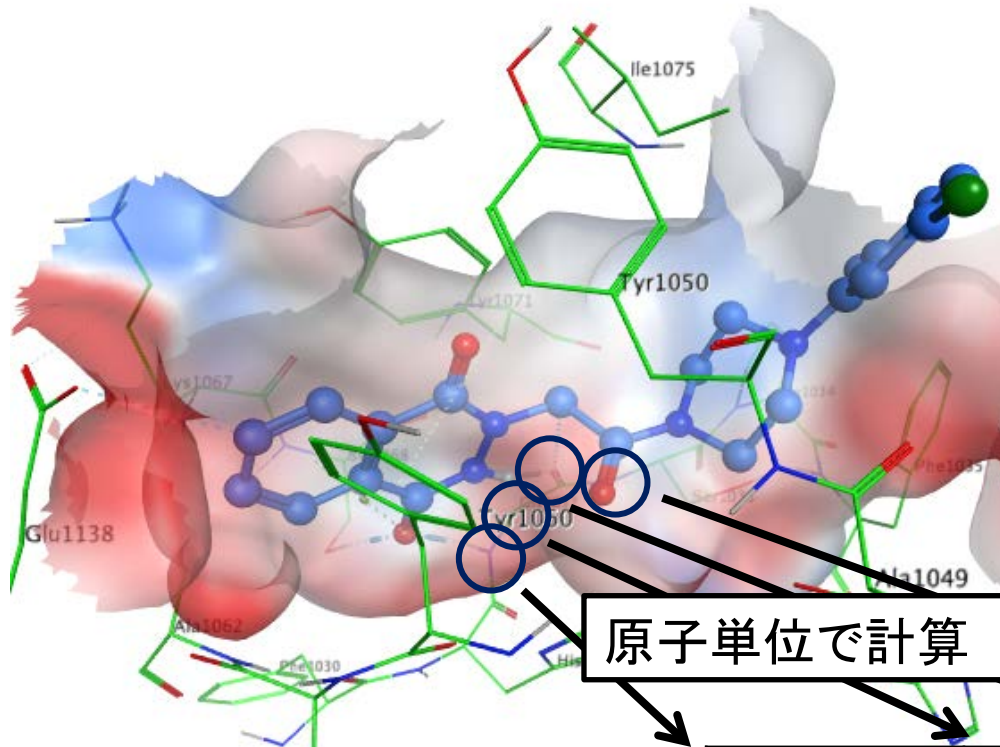
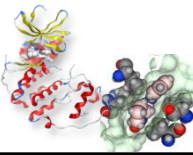
MUSES Models

Validate Prediction Performance



Test Set

タンパク質-リガンド相互作用エネルギー記述子 *aPLIED*

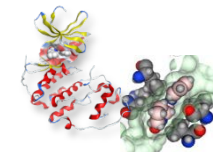


すべての結合サイトタンパク質原子に対して、その原子とリガンドとの間のMM-GBVIエネルギーを計算する

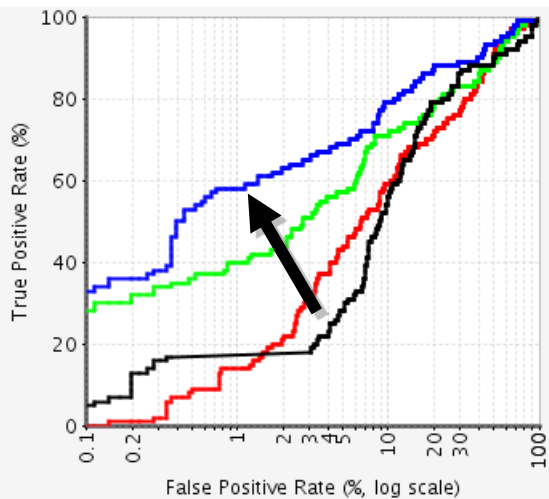
原子単位で計算

		Gly 1032 N	Gly 1032 Ca	Gly 1032 O	Ser 1036 N		Tyr 1071 O
<i>aPLIED</i> atom-based P-L Interaction energy descriptor	MM(ele)	-4.534	0	-6.621	0.034			0.098
	MM(vdw)	-0.435	-0.502	-0.356	-0.103			-0.176
	GBVI(sol)	3.040	0.040	3.118	0.021			0.032

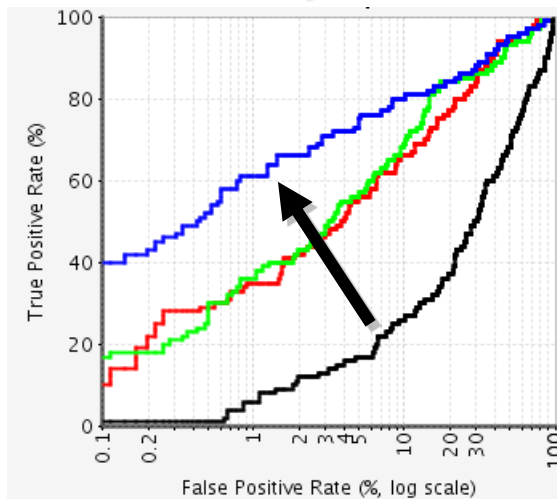
100個の活性化化合物を使った Support Vector Machine Model の ROC curve



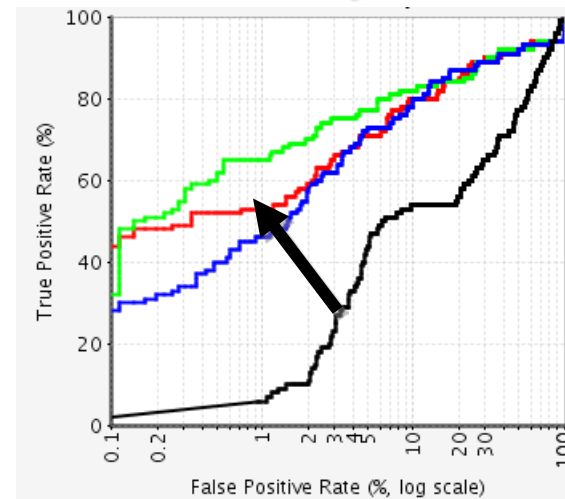
Cdk2



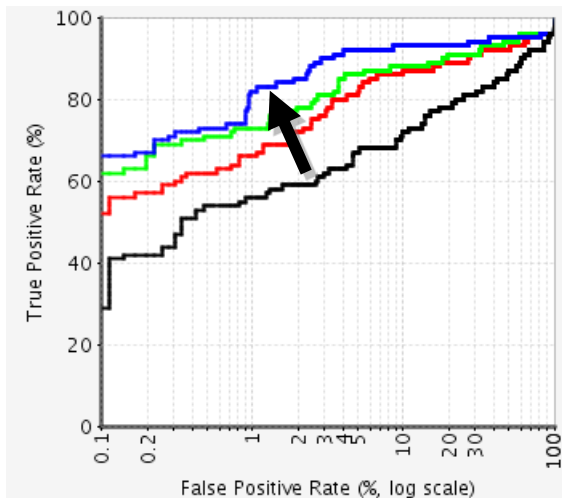
Cathepsin K



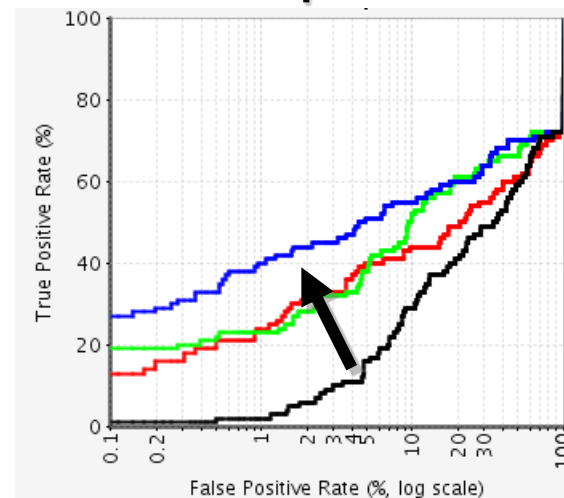
Carbonic Anhydrase 2



Thrombin



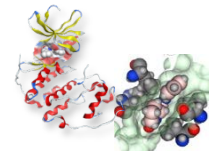
HIV-1 protease



- Glide score
- aPLIED-SVM
- aPLIF-SVM
- PLIF-SVM

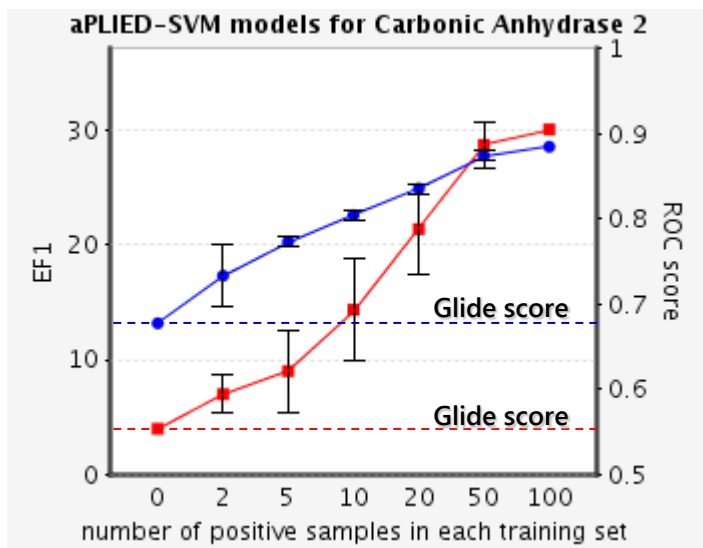
すべてのターゲットで **MUSES** モデルが **Glide score** を上回る結果を示した。

学習セットの活性化化合物の数と学習効果



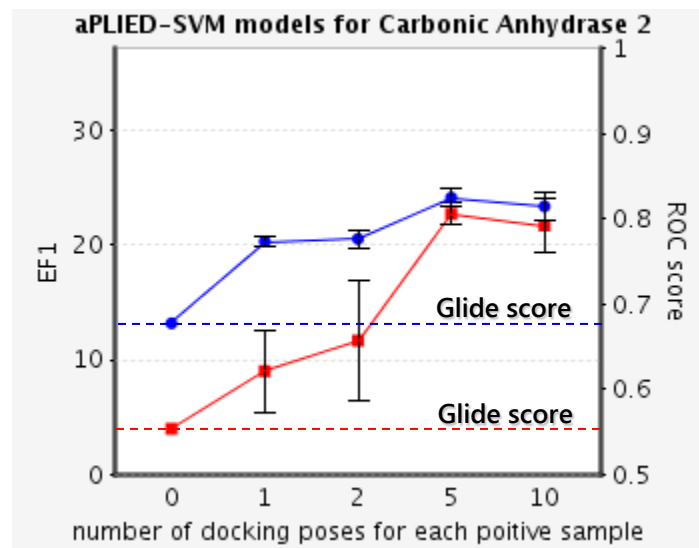
学習セットの既知活性化化合物数とスクリーニング効率

(ドッキングポーズは最もスコアの高いもののみ利用)



学習セットの既知活性化化合物あたりポーズ数とスクリーニング効率

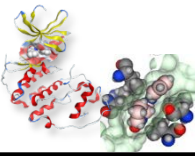
(既知活性化化合物数は5に固定)



- EF1 Mean
- ROC score Mean

- 基本的に学習セット中の既知活性化化合物数を増やすほどスクリーニング効率は向上する
- 2化合物でも有意にEF1, ROC scoreが向上し、50化合物程度でピークに達する。
- 各活性化化合物のドッキングポーズ数もトップだけではなく複数与えた方が良かった。5個程度でピークに達する。

Outline



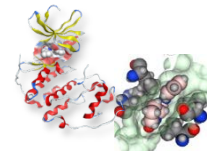
1. インシリコスクリーニングの技術開発

2. インシリコスクリーニング・設計の実例

3. 結合親和性予測／FMOの創薬活用

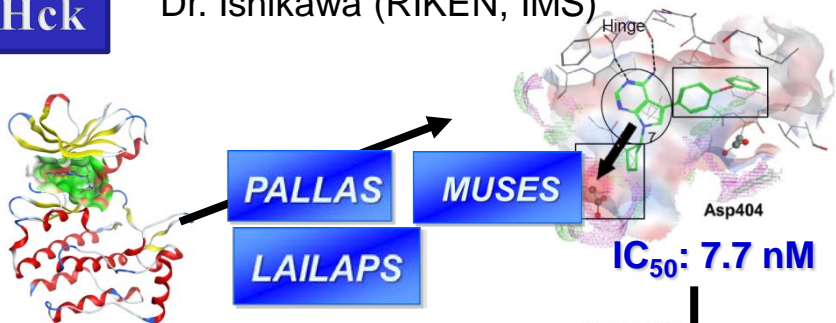
4. 体内動態・毒性の予測に関して

主なインシリコスクリーニングの結果



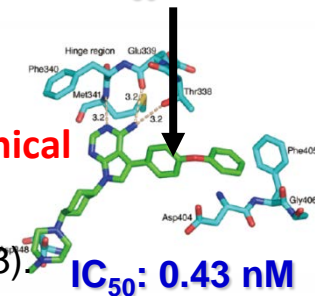
Hck

Dr. Ishikawa (RIKEN, IMS)



Promising chemical series was identified. Now we found a clinical candidate from the hit

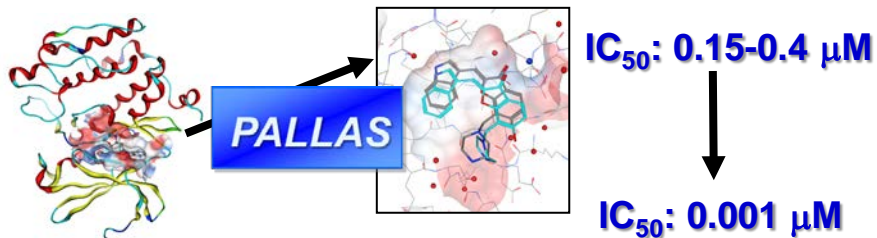
Sci Trans Med, 5:181ra52, (2013)



Pim1

Dr. Tanaka (RIKEN CLST) & Prof. Nagano (Univ. Tokyo)

Promising chemical series was identified.



J Med Chem, 55(11), 5151, (2012)

J Mol Bio, 417(3), 240, (2012)

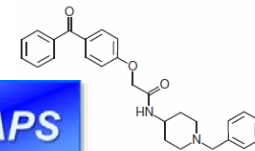
AdipoR

Prof. Kadowaki (Univ. Tokyo)

The first AdipoR agonist with in vivo efficacy

Nature, 503, 493-499, (2013).

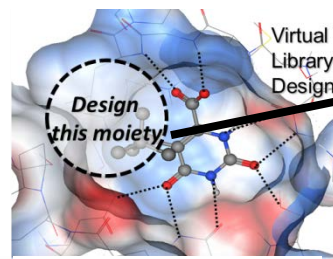
LAILAPS



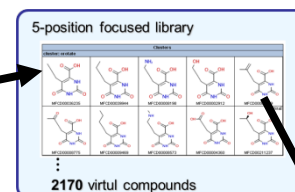
DHOD

Prof. Kita (Univ. Tokyo)

Inhibitor design for "undruggable" target

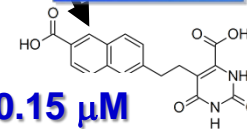


IC₅₀: 30-100 μM



PALLAS

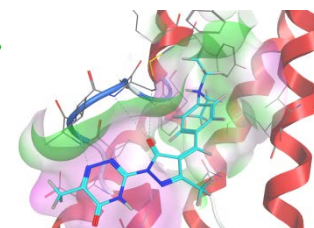
MUSES



DOCK2

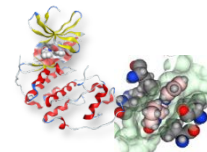
Dr. Fukui (Kyushu Univ.)

Difficult PPI target



IC₅₀: 5 μM





Human genes (~25,000)

Druggable genes
(~3,000)

Disease genes with
marketed drugs (215)

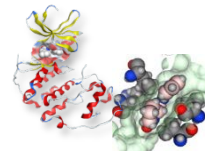
215

Disease genes
(~3,000)

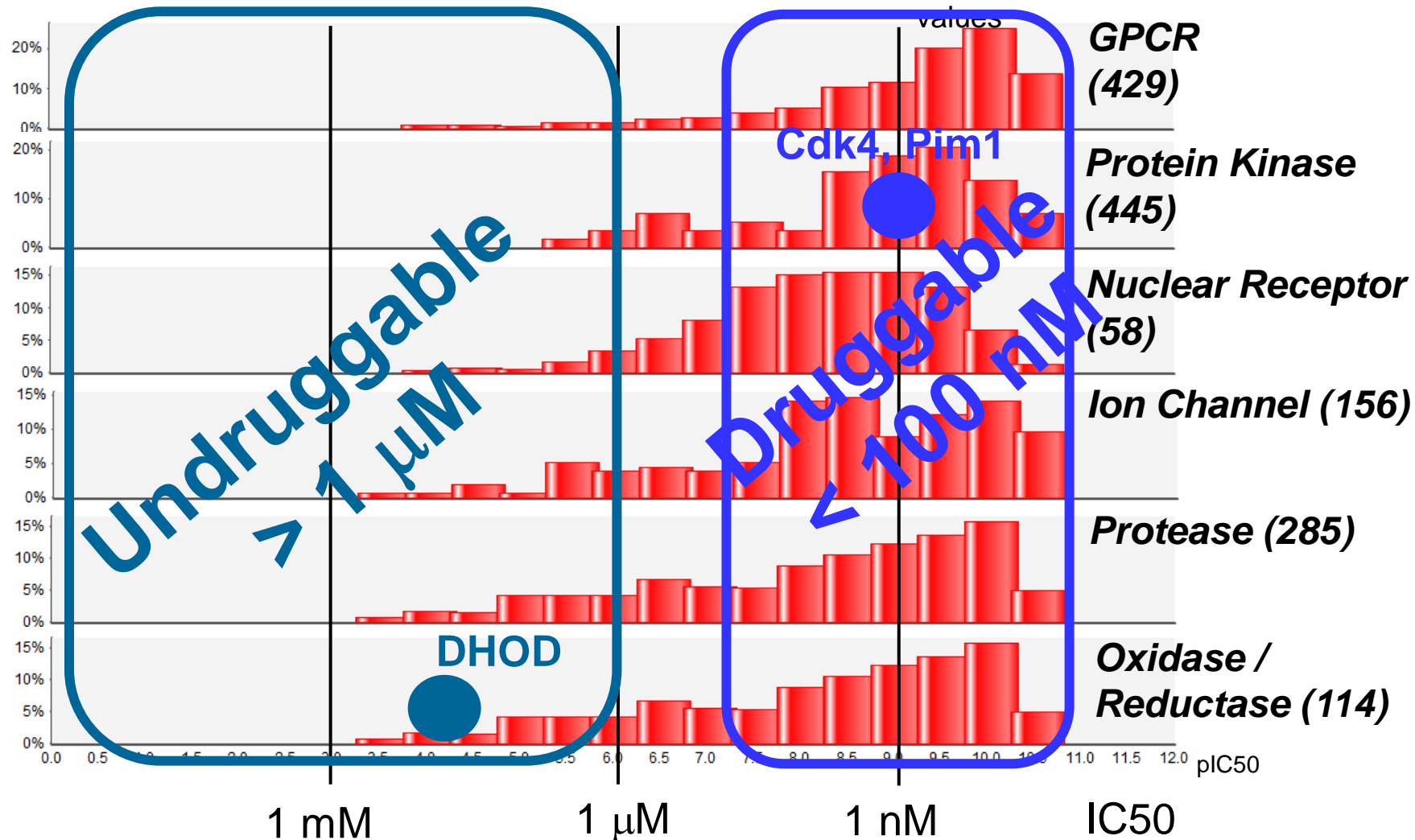
Disease genes with
Druggable domains
(~500)

従来の方法での**Druggable genes** (低分子で創薬可能なターゲット) は枯渇しつつある。インシリコなどの手法を使って、難易度の高いターゲットへのチャレンジが重要となる。

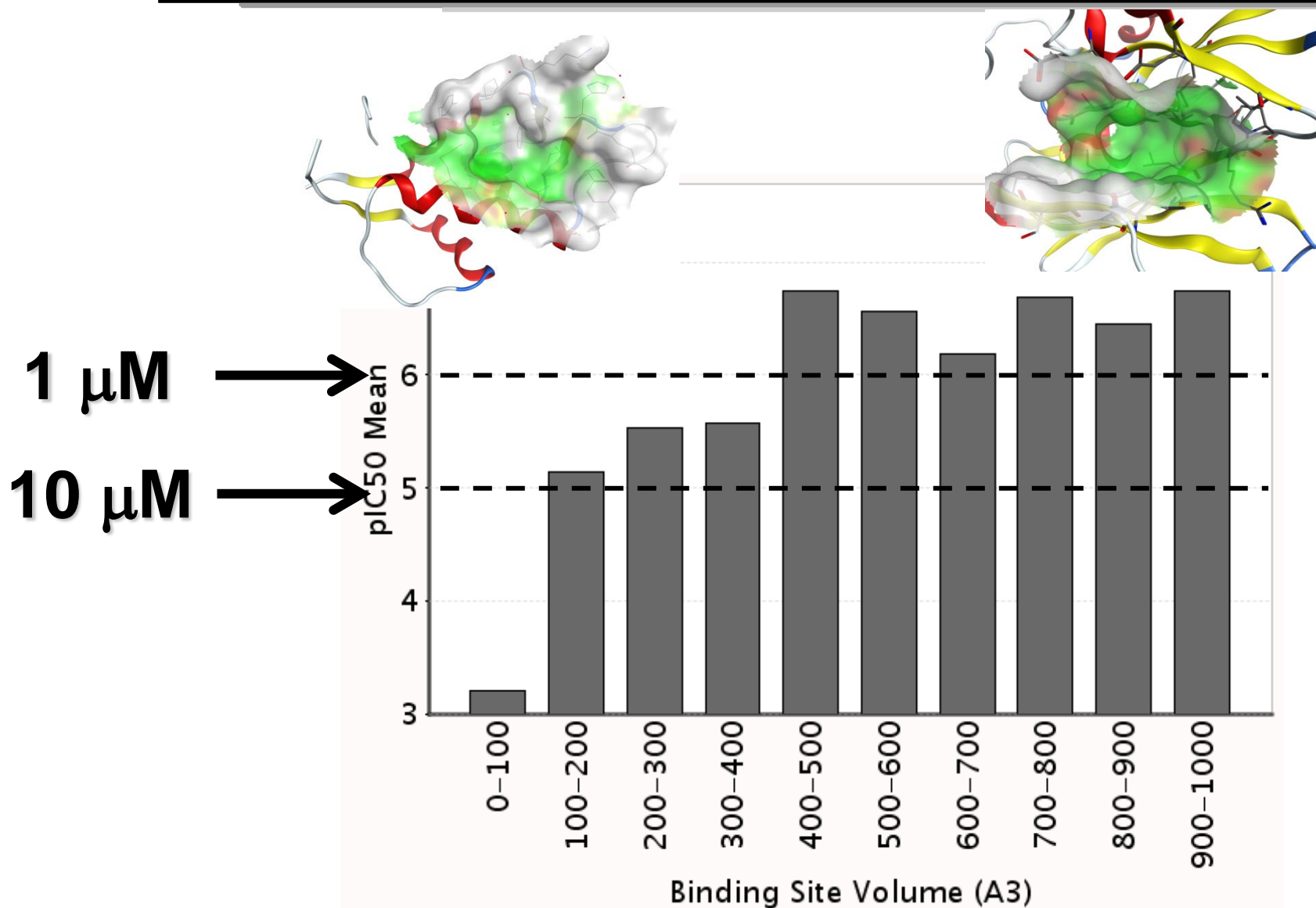
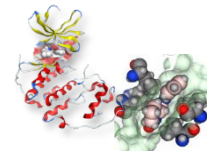
各ターゲットの最高到達活性の分布



Honma et al, analysed based on ChEMBL >600,000 IC50 values



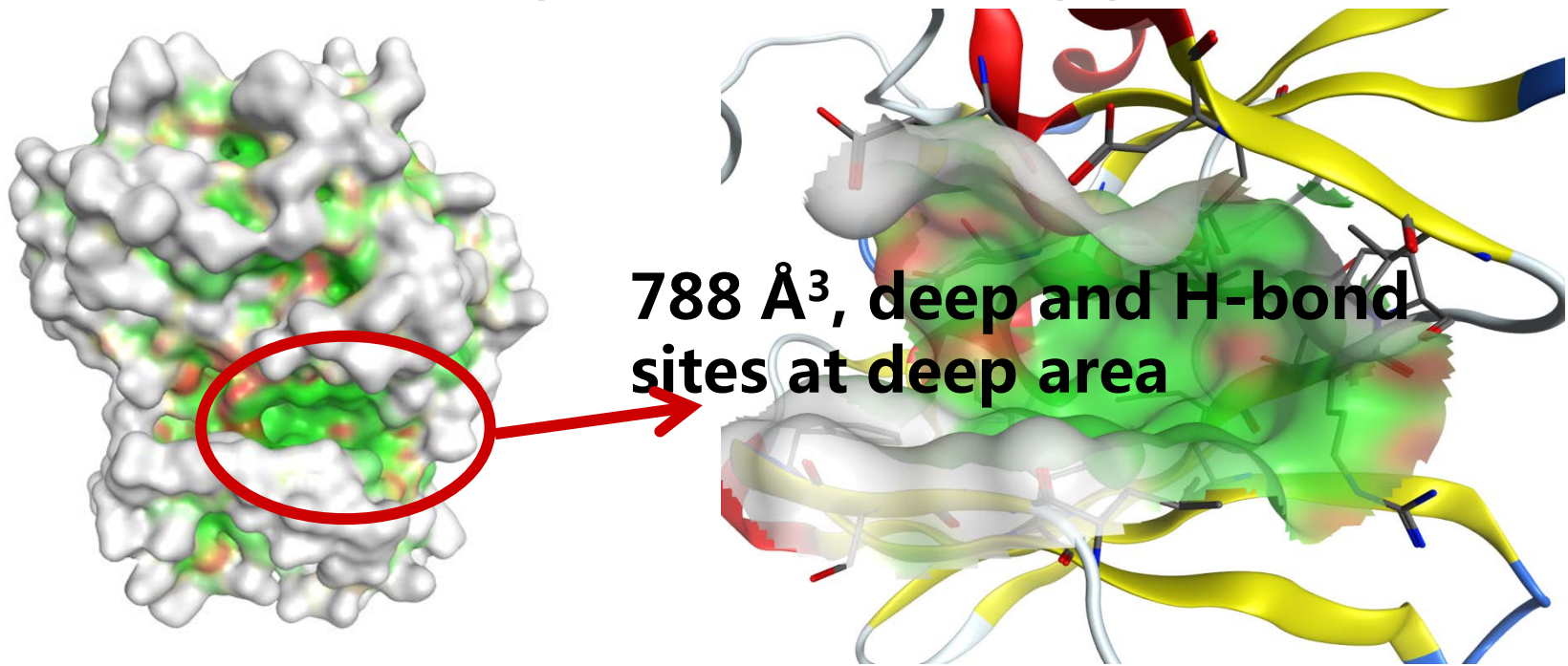
結合サイトの大きさと活性の強さ



This figure is based on ca. 3,000 complex data with IC50 values in PDBbind

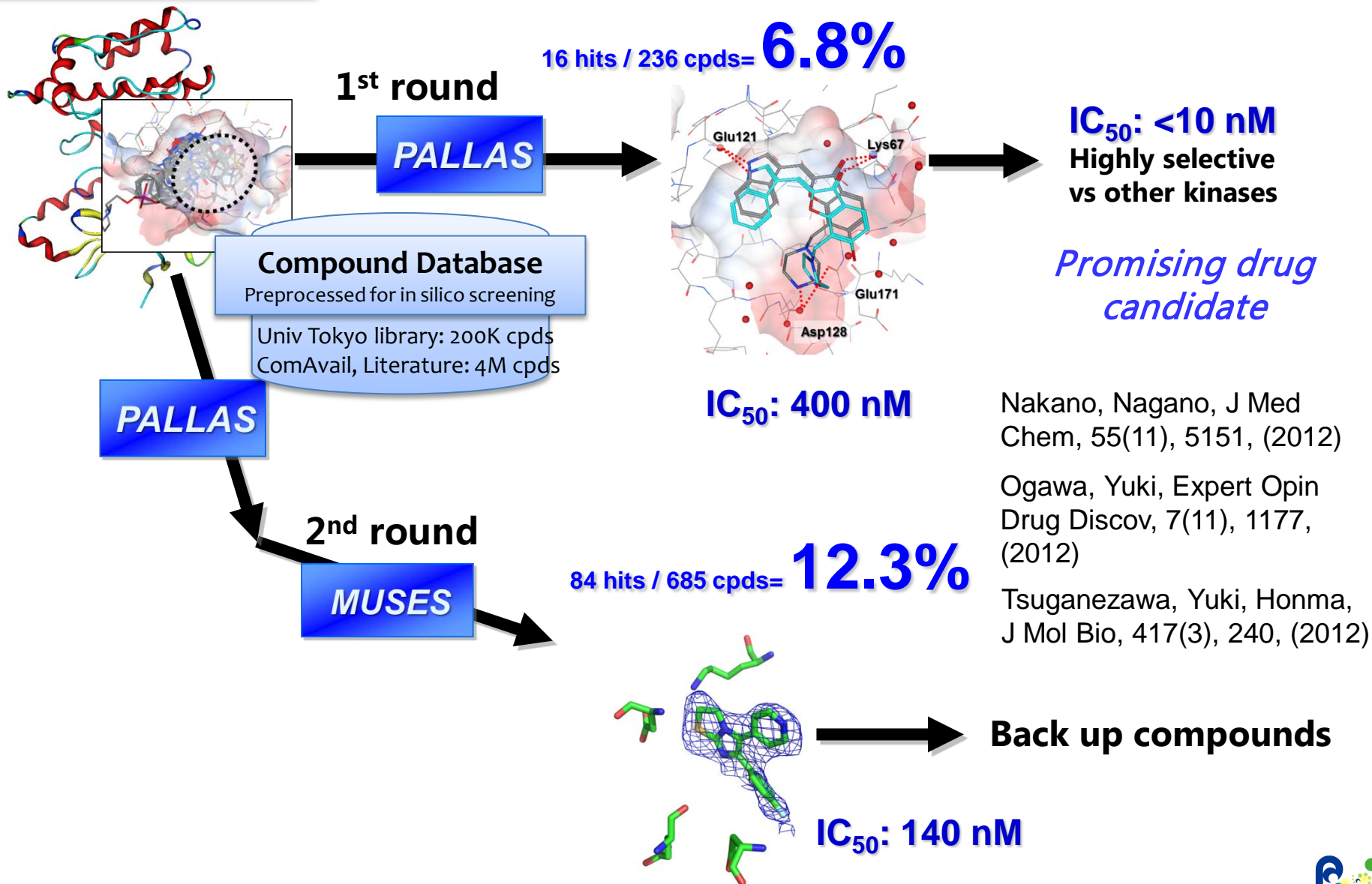
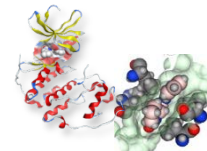
Category 1

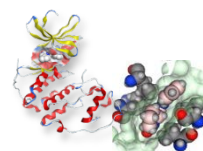
Is the following protein druggable ?



Target: **Pim1**
Ser/Thr family kinase
Acute Myelogenous Leukemia

Pim1 Inhibitors

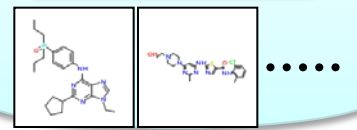




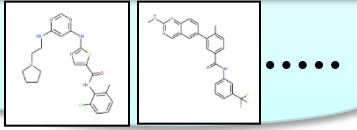
Target: **HCK**
 Src family kinase
 Acute Myelogenous Leukemia

HCK Inhibitors

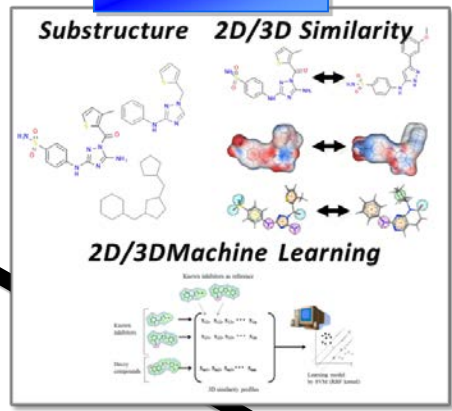
**HCK known
36 inhibitors**



**Src family known
141 inhibitors**

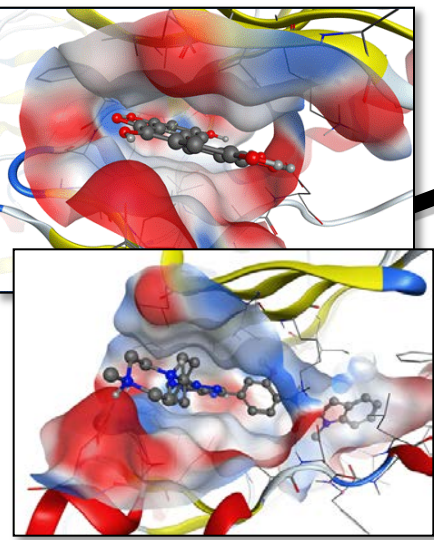


LAILAPS

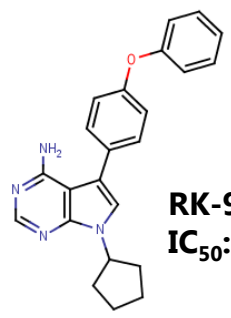


PALLAS

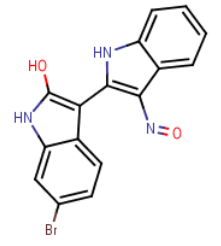
72 docking conditions of 12 X-ray structures were validated. PDB IDs (**2C01.B** and **2HCK.A**) were selected considering EF, ROC and X-ray reproducibility.



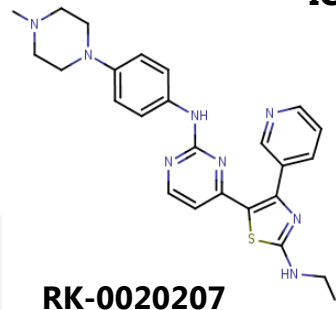
Compound Database
 Preprocessed for in silico screening
 Univ Tokyo library: 200K cpds
 ComAvail, Literature: 100M cpds



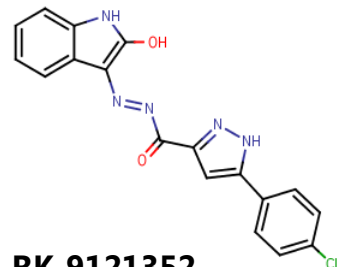
RK-9024466
 IC_{50} : 7.7 nM



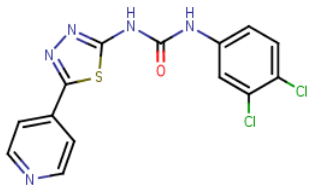
RK-9140696
 IC_{50} : 500 nM



RK-0020207
 IC_{50} : 395 nM



RK-9121352
 IC_{50} : 3800 nM

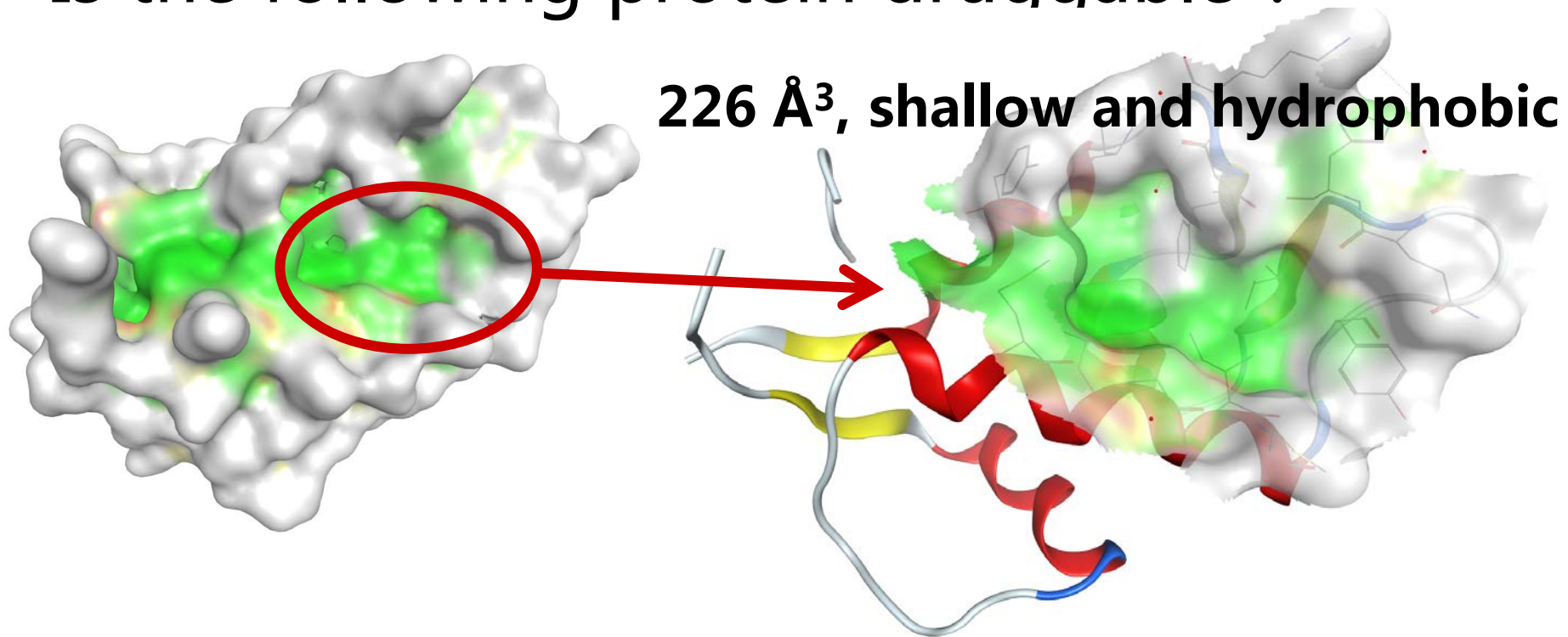


RK-9043895
 IC_{50} : 1100 nM

Other 11 hits (IC_{50} < ca 2000 nM)

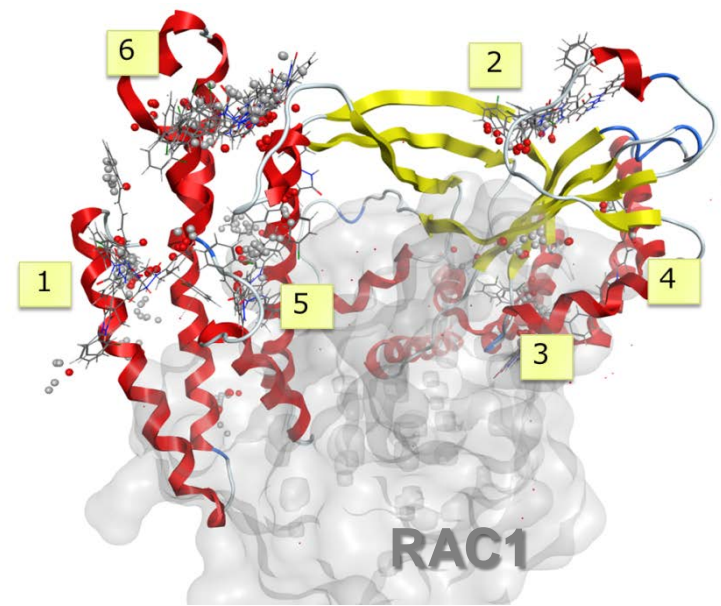
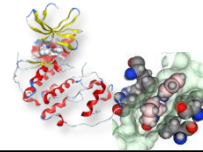
Category 2

Is the following protein druggable ?



Target: **DOCK1/2**
PPI
Cancer/Autoimmune drug

DOCK1/2 Inhibitors

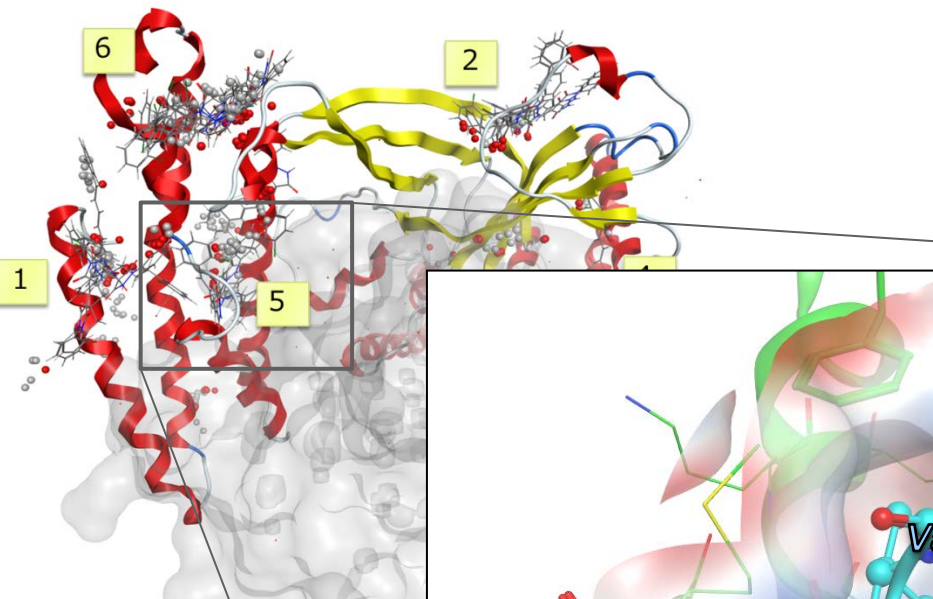
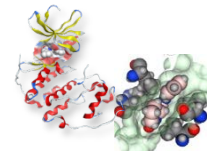


DOCK2側の結合サイト候補の分析

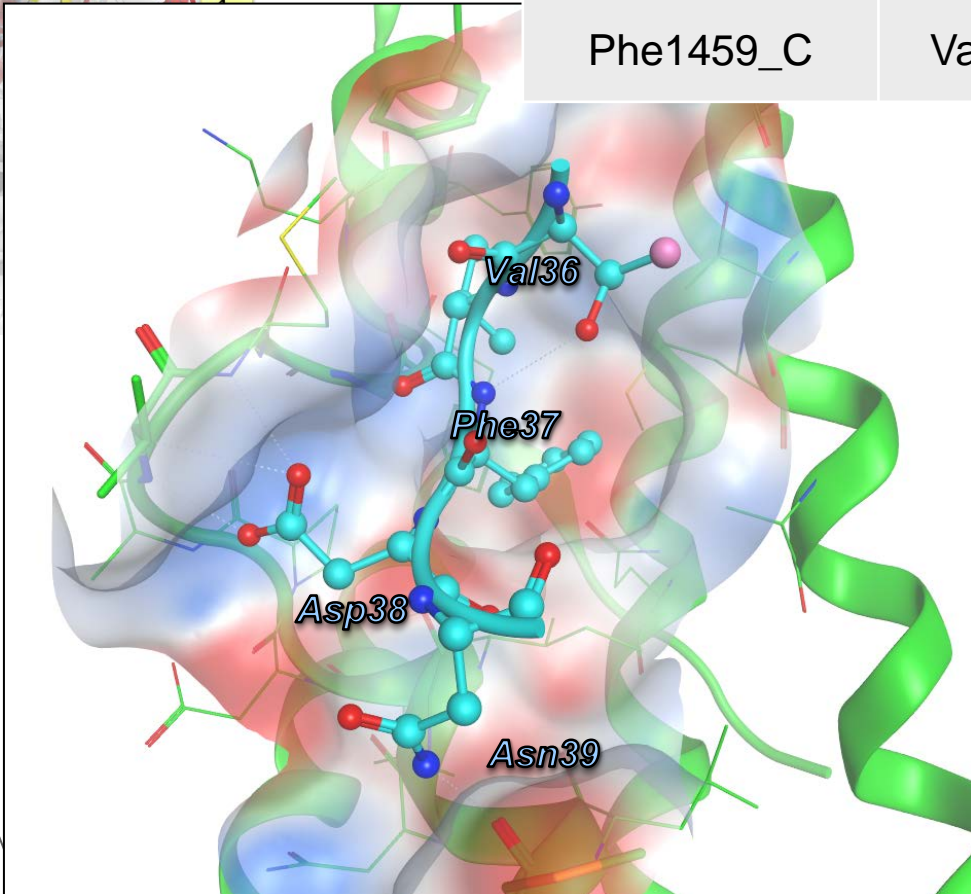
Site ID	Mean Docking score	Competitive w/ RAC1
1	-3.61	No
2	-3.92	No
3	-3.78	Partially yes
4	-5.25	No
5	-4.73	Yes
6	-4.75	No

Target: **DOCK1/2**
PPI
Cancer/Autoimmune drug

DOCK1/2 Inhibitors



DOCK2	RAC1-pep4	Type
Val1539_N or Met1540_N	Asp38_O	H-bond
Gly1541_N	Val36_O	H-bond
Phe1459_C	Val36_C	Hydrophobic contact



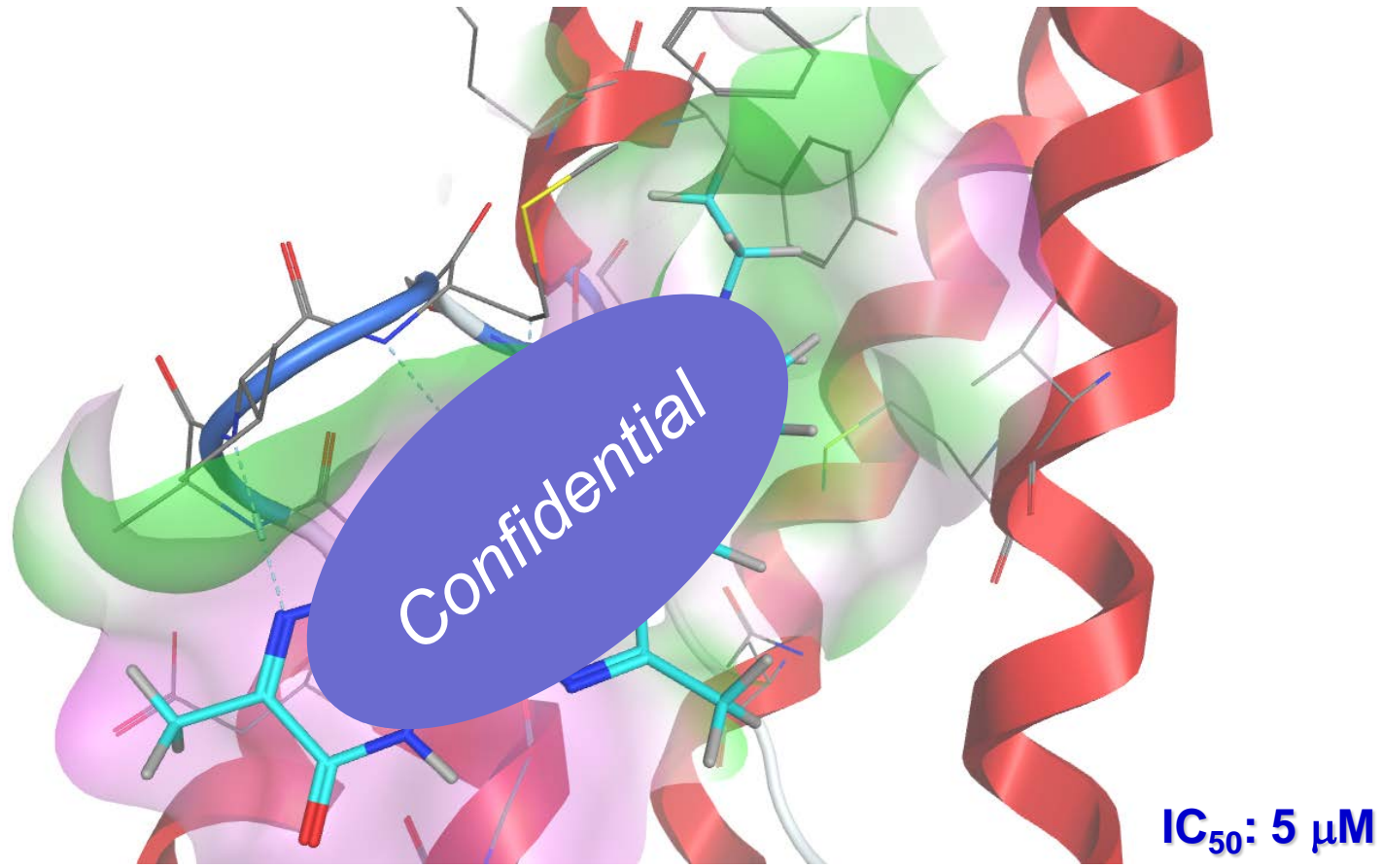
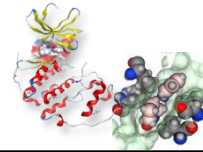
2つのX線構造を両方使い、RAC1との相互作用を再現する化合物584個を選んだ

4 hits / 584 cpds = **0.7%**

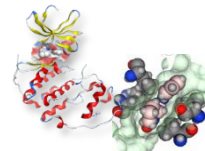


Target: **DOCK1/2**
PPI
Cancer/Autoimmune drug

DOCK1/2 Inhibitors

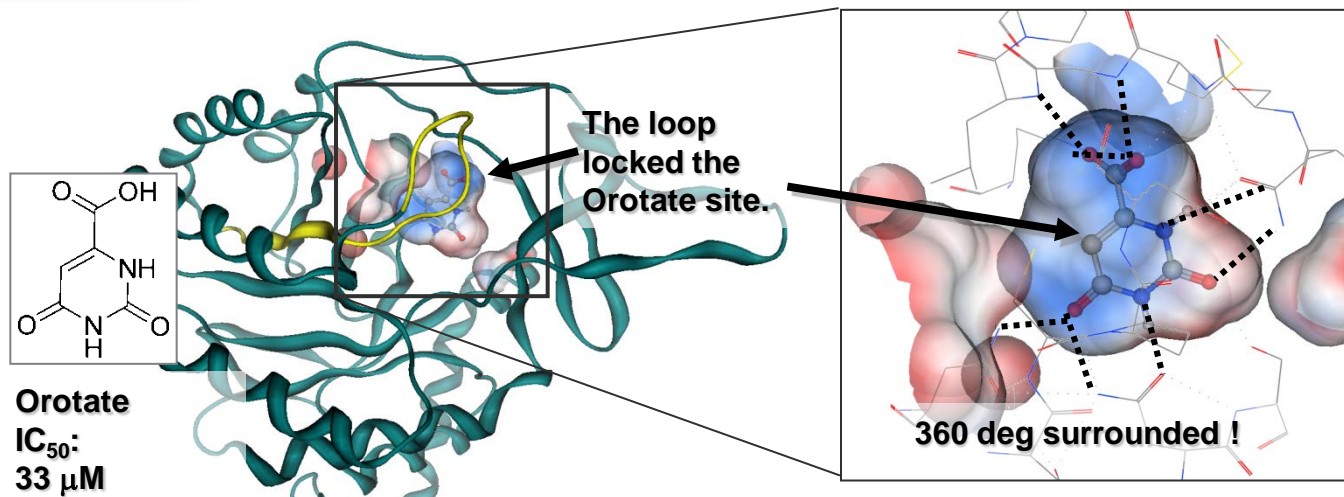


LAILAPSも含めて、8個の骨格の違う阻害剤を得たので、次のラウンドとして、それらの情報を利用したPALLAS/MUSESを実施できる。



Target: **DHOD**
 Oxidase/Reductase
 AntiTripanosoma drug

DHOD Inhibitors

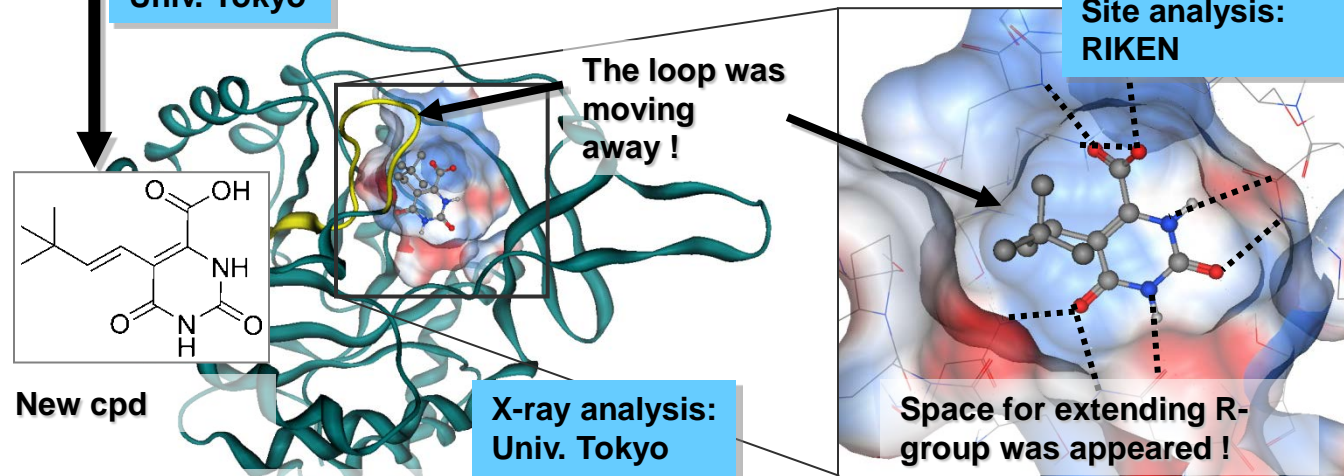


Site volume : 178 \AA^3

Closed form DHOD

Design:
 RIKEN

Synthesis :
 Univ. Tokyo

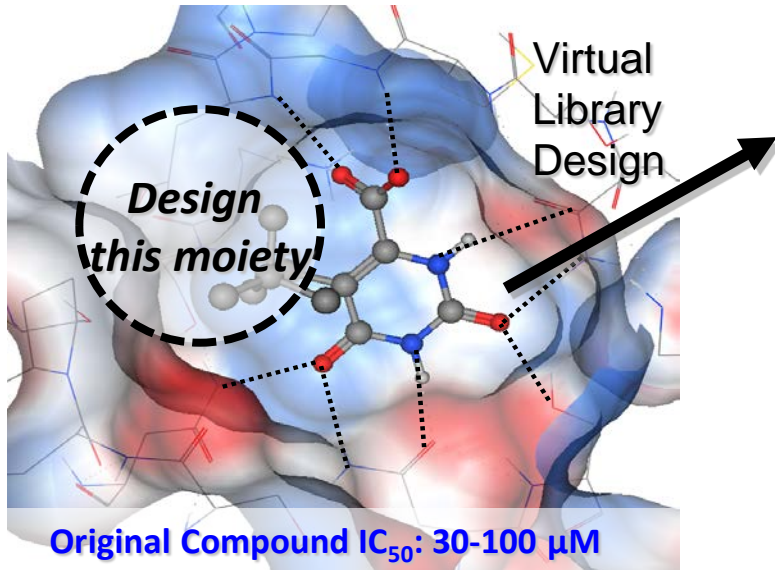
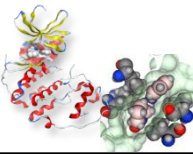


Site volume : 694 \AA^3

Open form DHOD

X-ray analysis:
 Univ. Tokyo

仮想化合物ライブラリーの構築による設計



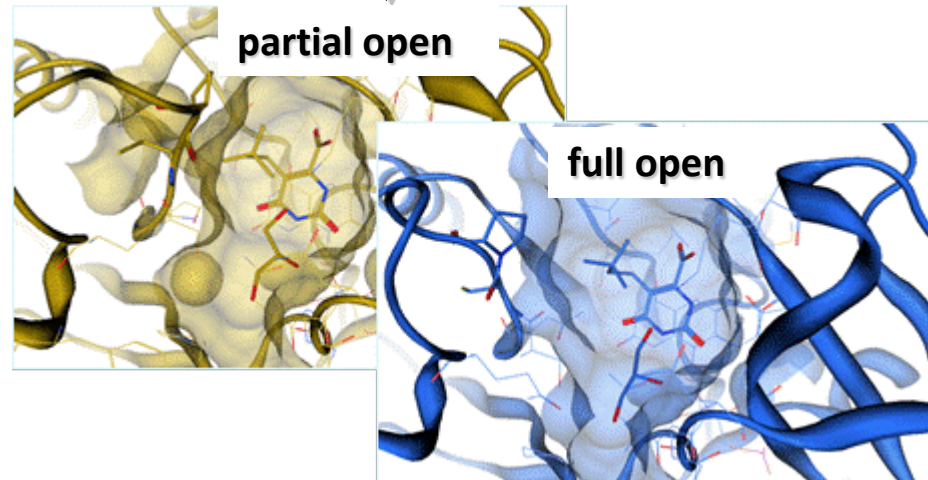
5-position focused library

Clusters				
cluster: orotate				
 MFCD00036235	 MFCD00039944	 MFCD00008198	 MFCD00002912	 MFCD00008599
 MFCD00008775	 MFCD00009469	 MFCD00008573	 MFCD00004360	 MFCD000211237

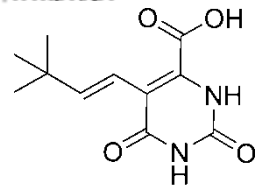
2170 virtual compounds

PALLAS / MUSES

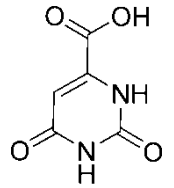
Docking



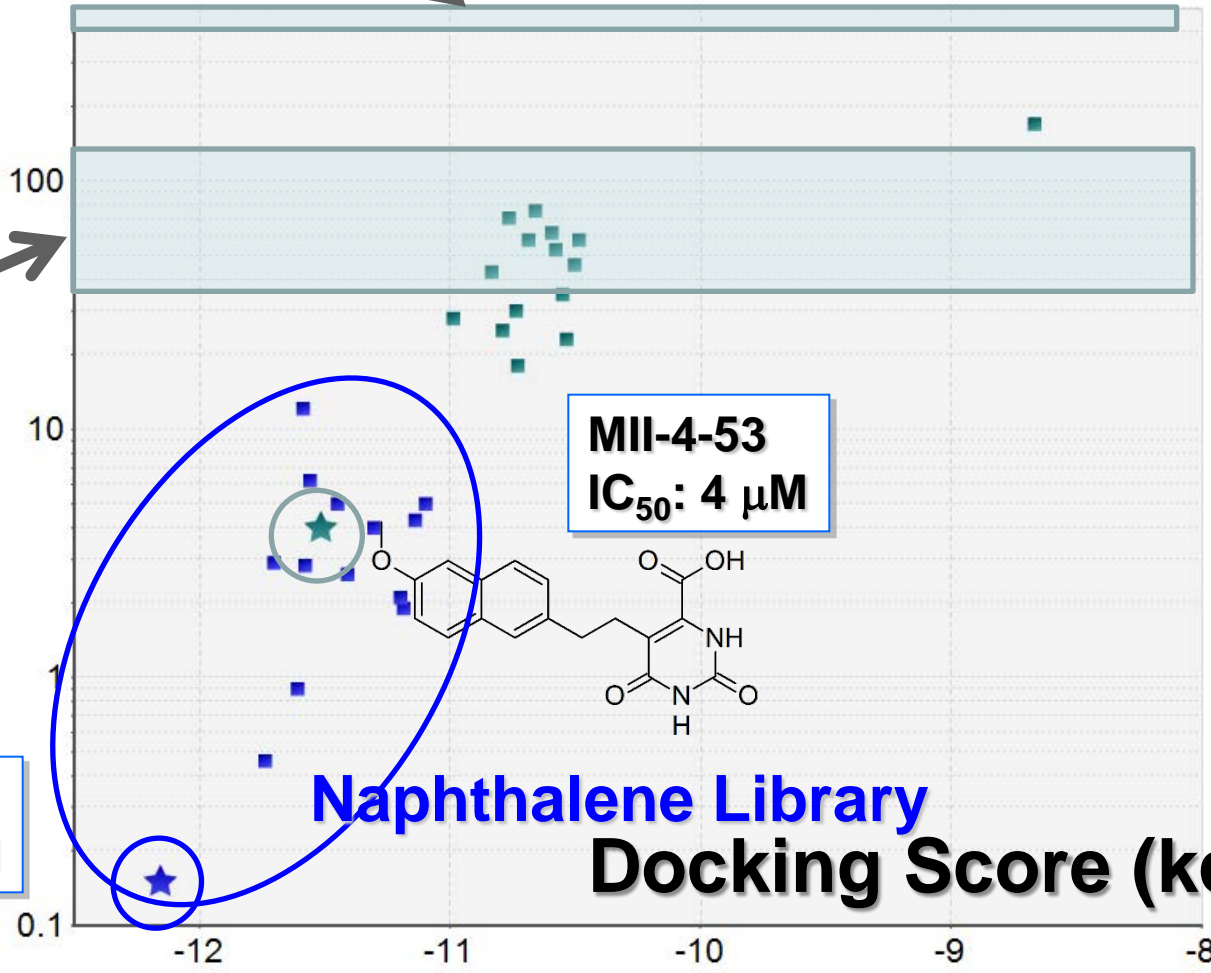
Synthesized Orotate-mimic inhibitor
IC₅₀: >400 μM



IC₅₀ (μM)



Orotate
IC₅₀: 33-135 μM

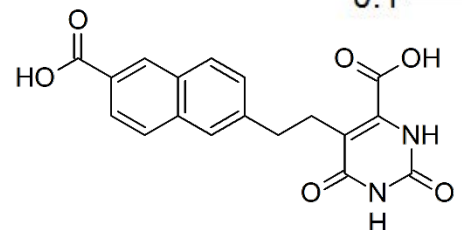


MII-4-53
IC₅₀: 4 μM

Naphthalene Library

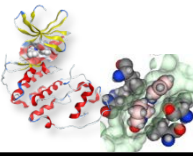
Docking Score (kcal/mol)

MII-6-040
IC₅₀: 0.15 μM



300倍活性の高い化合物が得られた

Outline



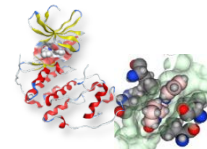
1. インシリコスクリーニングの技術開発

2. インシリコスクリーニング・設計の実例

3. 結合親和性予測／FMOの創薬活用

4. 体内動態・毒性の予測に関して

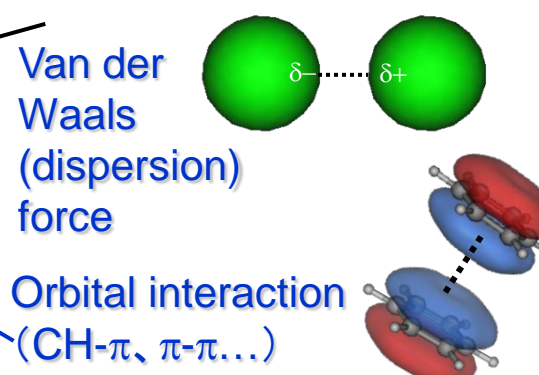
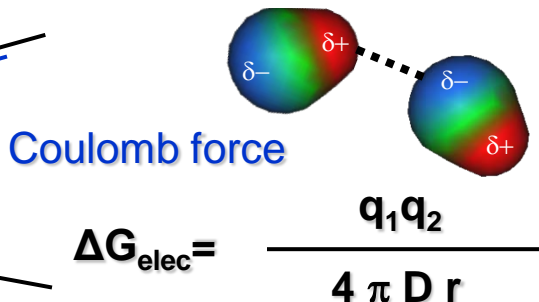
ΔGの内訳と活性向上への潜在能力



$$\Delta G \text{ (kcal/mol)} = - RT \ln K_d$$

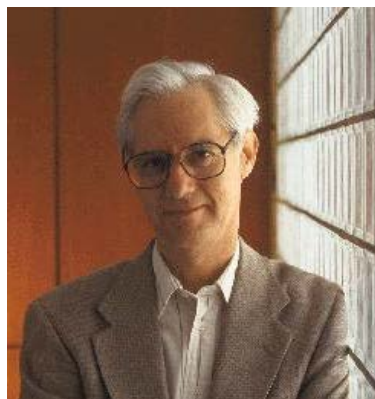
$$= -1.4 \log K_d = \Delta H - T\Delta S$$

		example	ΔG (kcal/mol)	Enhancement of IC ₅₀ (-fold)
ΔH enthalpy	Electrostatic	Hydrogen bond	0.6–1.8	~2-10
	Coulomb force	Ionic bond	2.4–4.8	~10-1000
ΔS entropy	Hydrophobic	methyl-methyl	1.4	~10
	Van der Waals (dispersion) force	phenyl-phenyl	3.4	~100
	Orbital interaction (CH-π, π-π...)	Rotation of a single rotatable bond is restricted	0.5–1.2	~2-10
	Entropy change of solvent			
	Entropy change of ligand			
	Entropy change of protein			



初期のスコア カ場ベースのDOCK

最初のscoring function 1982
年にKuntzらによって発表され
た最初のauto docking
softwareであるDOCKの
scoring function
分子力場の分子間相互作用寄
与をほぼそのまま利用



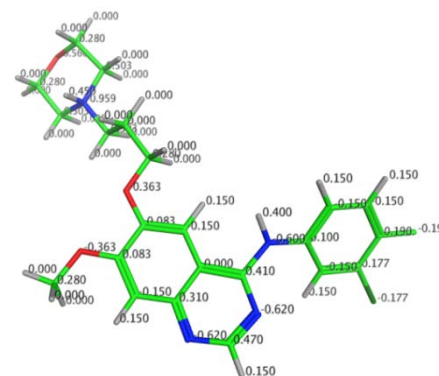
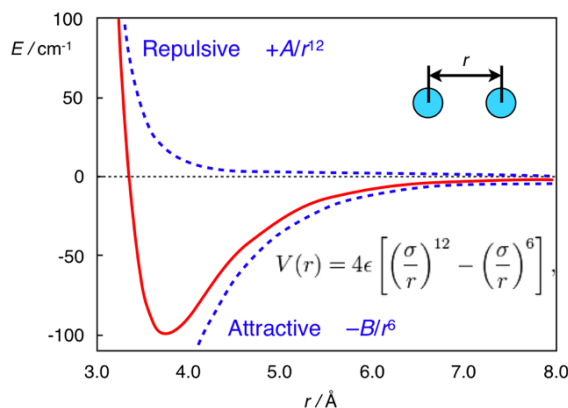
Kuntz, JMB, 269 (1982)

Van der Waals force
(Lenard-Jones approximation)

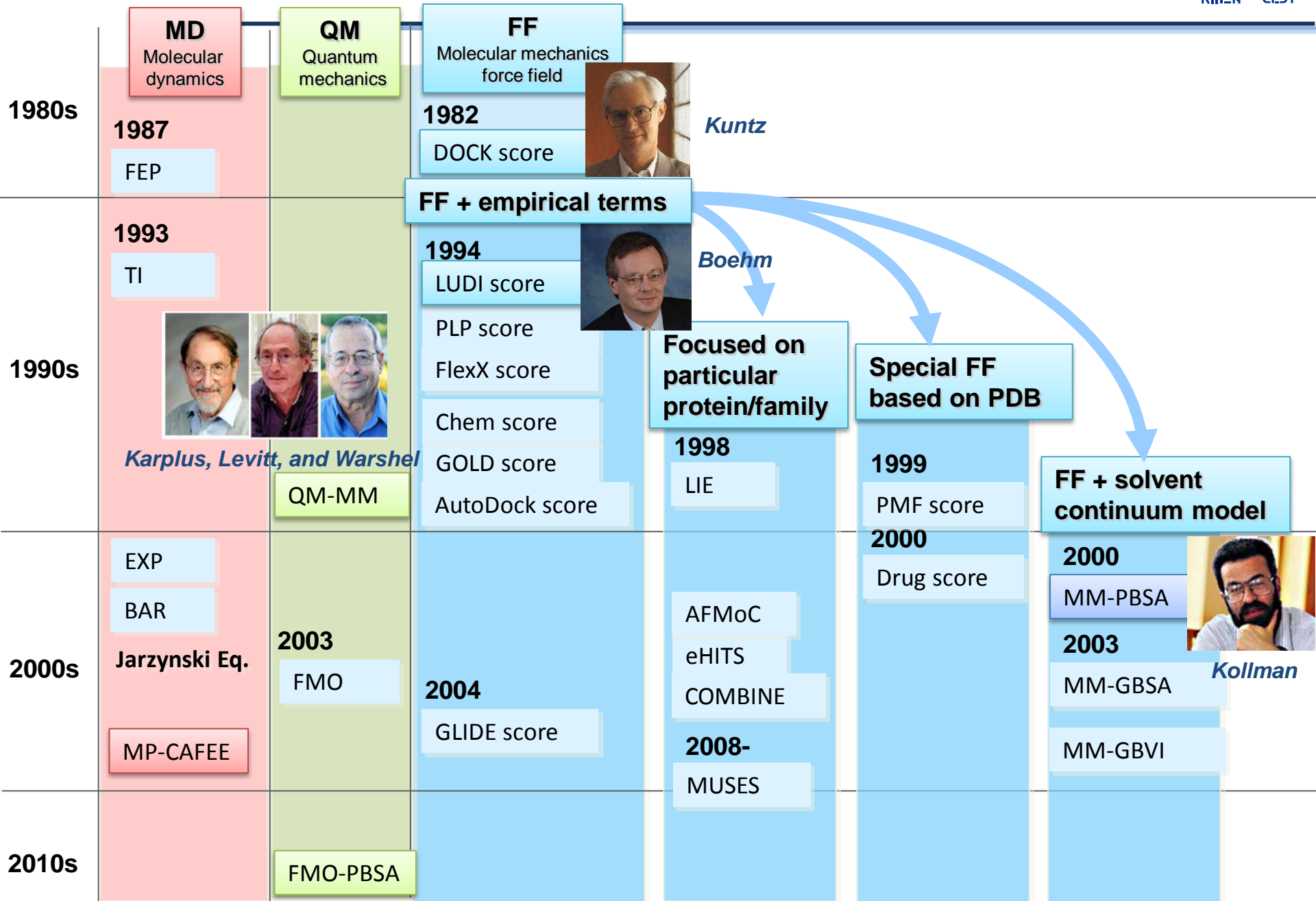
Electrostatic interaction
(Coulomb's equation)

$$\text{DOCK score} = \sum \sum (A / r_{ij}^{12} - B / r_{ij}^6) + 332 q_i q_j / D r_{ij}$$

r_{ij} : distance between
atom i and atom j
A,B: constants determined
by atom type
 q_i, q_j : partial charge of
atom i and atom j
D: dielectric constant



スコア関数の歴史



力場ベースの経験的スコアの限界

Empirical Force Field-based Docking Score (Glide, GOLD, ...)

	ΔH	ΔS	Force field	Conf. sampling
Protein	Fair	Not Considered	Fair	Fair / Bad
Ligand	Fair	Bad	Fair / Bad	Bad
Solvent	Bad / Not Considered	Bad	Fair	Bad / Not Considered

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

Glide Score

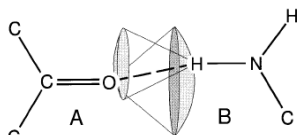
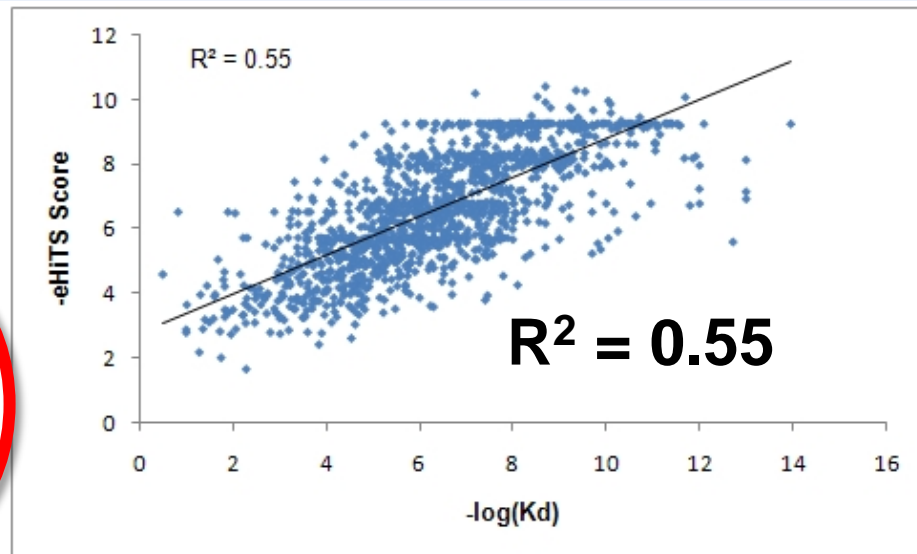


Figure 2. Condition for the formation of interactions: a hydrogen bond between the carbonyl oxygen and the nitrogen. The interaction centers are the oxygen and the hydrogen atom forming the hydrogen bond. They have to fall mutually on the surrounding interaction surfaces.

$$f^*(\Delta R) = \begin{cases} 0 & \Delta R > 0.6 \text{ \AA} \\ 1 - \frac{\Delta R - 0.2}{0.4} & 0.2 \text{ \AA} < \Delta R \leq 0.6 \text{ \AA} \\ 1 & -0.2 \text{ \AA} < \Delta R \leq 0.2 \text{ \AA} \\ 1 - \frac{-\Delta R - 0.2}{0.4} & -0.6 \text{ \AA} < \Delta R \leq -0.2 \text{ \AA} \\ \frac{\Delta R + 0.6}{0.2} & \Delta R \leq -0.6 \text{ \AA} \end{cases}$$

Judgement of forming a HB (FlexX)

Strength of HB (FlexX, distance)



PDB bind収載の800化合物のX線構造 + IC₅₀ (Kd/Ki) があるデータを使った検証

Scoring function	N	R _p	SD	ME
eHiTS 2009	800	0.64	1.34	1.06
X-Score::HMScore	800	0.57	1.82	1.42
DrugScore::Pair	800	0.47	1.94	1.51
AI-Score+XFurcate	799	0.46	1.96	1.54
FlexX Score(Opt)	799	0.43	1.99	1.55
GOLD::ChemScore	762	0.45	1.96	1.52
GOLD::GoldScore	772	0.37	2.06	1.63

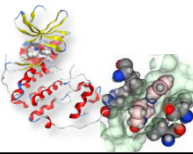
R² = 0.2-0.5程度

平均でIC₅₀値にして10倍から100倍の誤差が

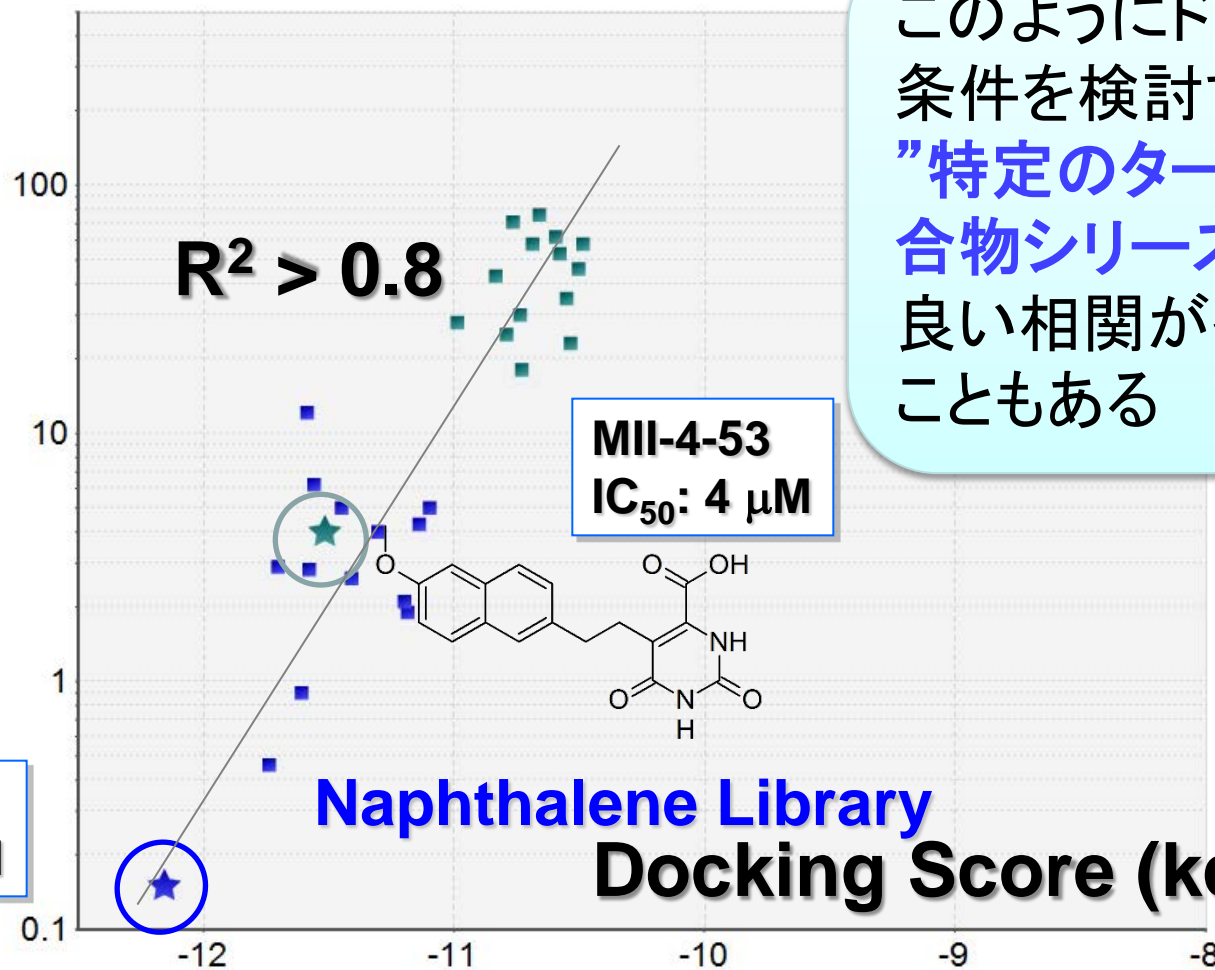
[eHiTS 2009 Binding Affinity technical note.:](http://www.simbiosys.ca/ehits/eHiTS_2009_binding.pdf)

http://www.simbiosys.ca/ehits/eHiTS_2009_binding.pdf

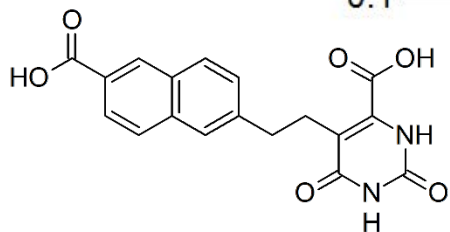
仮想化合物ライブラリーの構築による設計



IC₅₀ (μM)



このようにドッキング条件を検討すると、“特定のターゲット・化合物シリーズに限り”良い相関が得られることもある



より精度の高いスコア方法(1) MM-PB/SA

Empirical Force Field-based Docking Score (Glide, GOLD, ...)

	ΔH	ΔS	Force field	Conf. Sampling
Protein	Fair	Not Considered	Fair	Fair / Bad
Ligand	Fair	Bad	Fair / Bad	Bad
Solvent	Bad / Not Considered	Bad	Fair	Bad / Not Considered

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

Glide Score

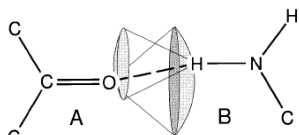


Figure 2. Condition for the formation of interactions: a hydrogen bond between the carbonyl oxygen and the nitrogen. The interaction centers are the oxygen and the hydrogen atom forming the hydrogen bond. They have to fall mutually on the surrounding interaction surfaces.

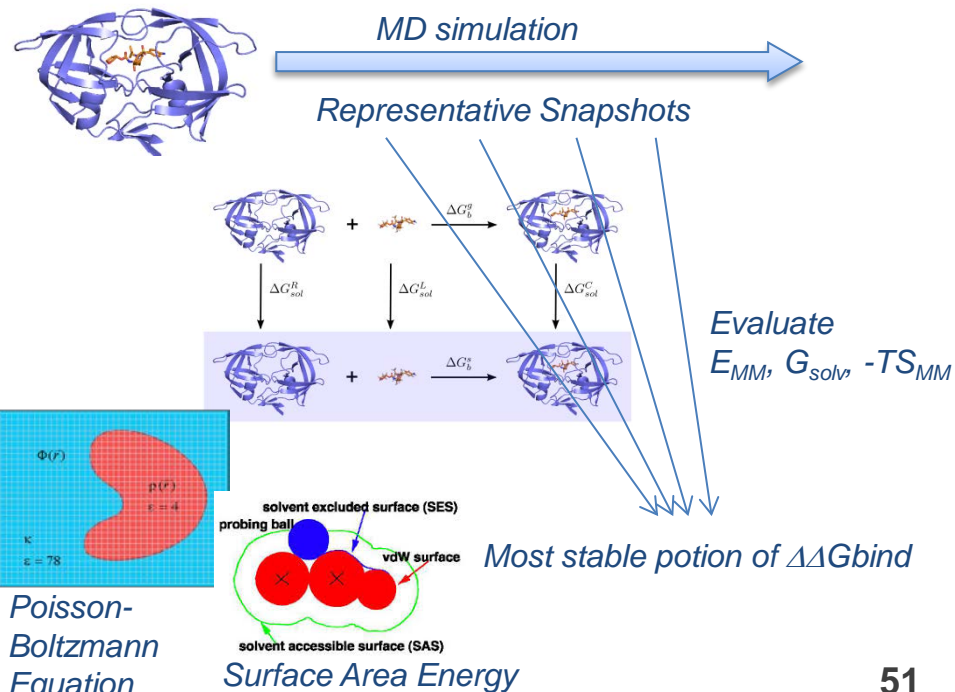
$$f^*(\Delta R) = \begin{cases} 0 & \Delta R > 0.6 \text{ \AA} \\ 1 - \frac{\Delta R - 0.2}{0.4} & 0.2 \text{ \AA} < \Delta R \leq 0.6 \text{ \AA} \\ 1 & -0.2 \text{ \AA} < \Delta R \leq 0.2 \text{ \AA} \\ 1 - \frac{-\Delta R - 0.2}{0.4} & -0.6 \text{ \AA} < \Delta R \leq -0.2 \text{ \AA} \\ \frac{\Delta R + 0.6}{0.2} & \Delta R \leq -0.6 \text{ \AA} \end{cases}$$

Judgement of forming a HB

Strength of HB (distance)

Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PB/SA)

	ΔH	ΔS	Force field	Conf. Sampling
Protein	Fair	Not Considered	Fair	Fair / Bad
Ligand	Fair	Fair / Bad	Fair / Bad	Fair / Bad
Solvent	Fair PB	Fair / Bad SA	Fair	Fair / Bad



MP-CAFEE と FMO 相互補完的關係

Massively Parallel Computation of Absolute binding Free energy with well-Equilibrated states (MP-CAFEE)

	ΔH	ΔS	Force field	Conf. Sampling
Protein	Fair	Fair	Fair	Good / Fair
Ligand	Fair	Fair	Fair / Bad	Good / Fair
Solvent	Fair	Fair	Fair	Good / Fair

MP-CAFEEは、シミュレーションが収束したかを妥当に判断できれば、原理的には、すべてのエネルギー項をかなり適切に評価できる。

一部の力場パラメータに不十分な点がある。

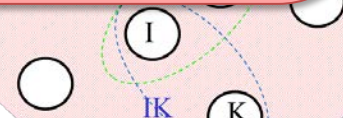
Fujitani et al, JCP (2005)

従来のFMOが力を発揮できるのは、原理的には、エントロピーや溶媒効果、初期構造の問題がない場合

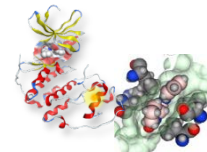
	Good	Not Considered	Good	Not Considered
Ligand	Good	Not Considered	Good	Not Considered
Solvent	Bad	Not Considered	Good / Not Considered	Not Considered

- ・計算対象の化合物群は似ていて、エントロピーや溶媒効果は相殺できる
- ・計算対象の化合物と良く似た化合物の複合体結晶構造が解かれている。

QM region

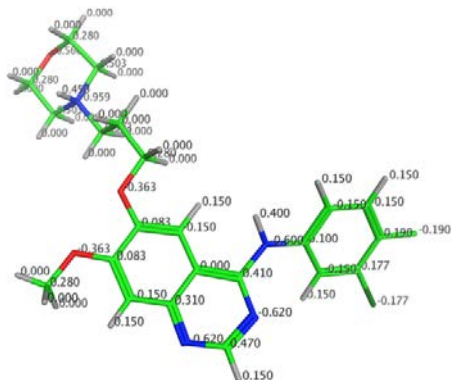


MM(FF) とQMの違い

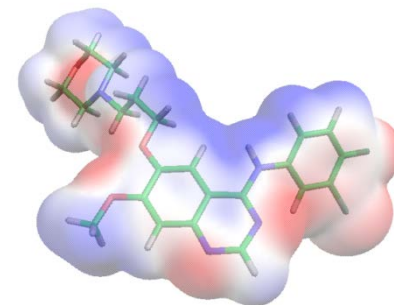


Charge

MM(molecular mechanics, called force field (FF)とも呼ばれる) は、各原子の中心のみに電荷を割り付けている。

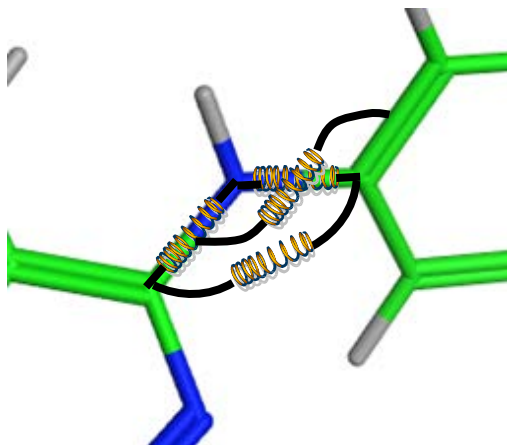


QM(quantum mechanics)では、Schrodinger方程式で記述される電子密度に従って、電荷の粗密が連続的に表現される。

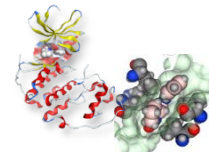


Strain energy

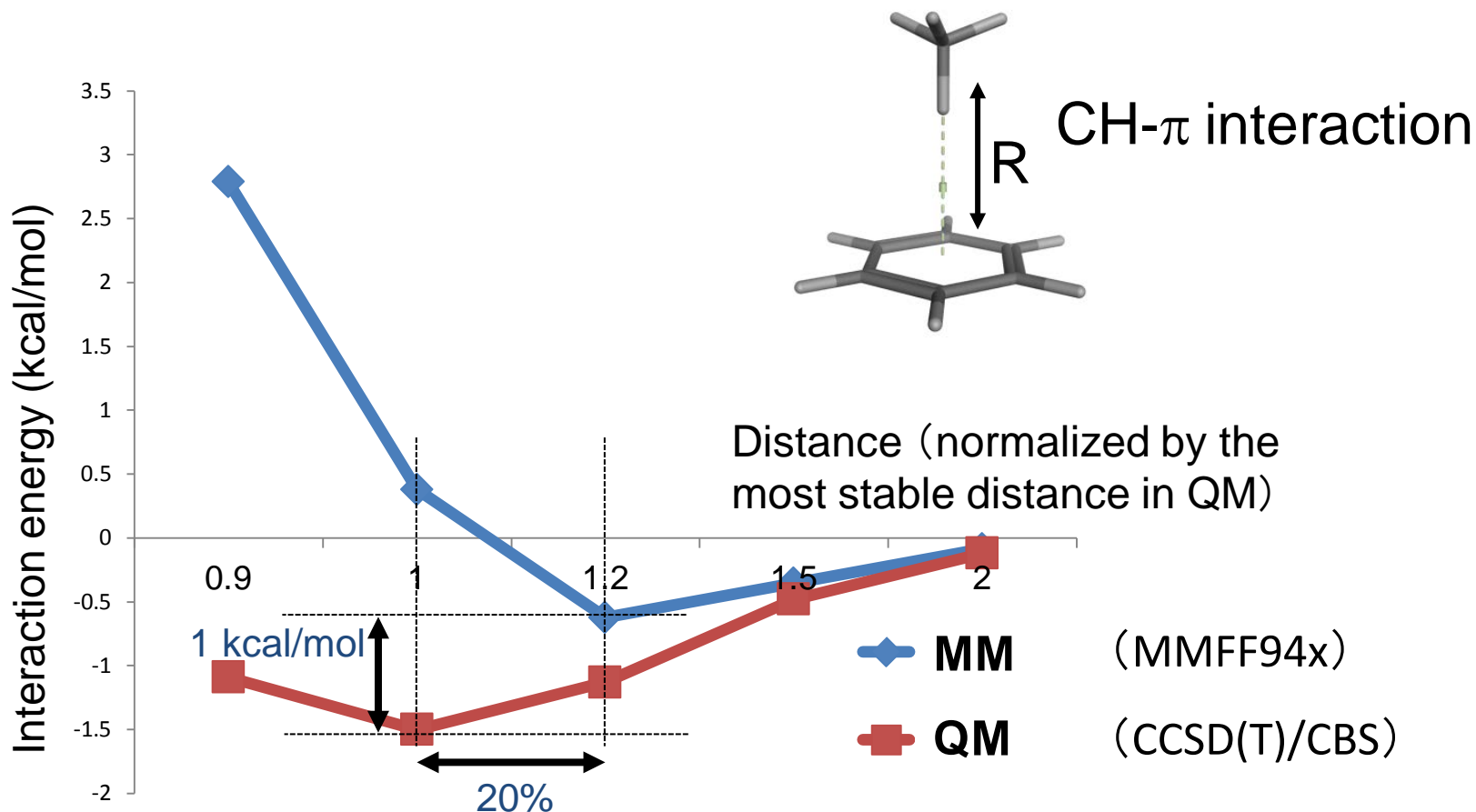
MMでは、配座の歪エネルギーは、フックの法則のバネ近似で表現される。



QMでは、Schrodinger方程式に基づいて、軌道の重なりによって、各配座が安定か不安定か記述される。



Ex) CH- π interaction of MM and QM



20% of distance error and 1 kcal/mol of interaction energy error were observed in MM

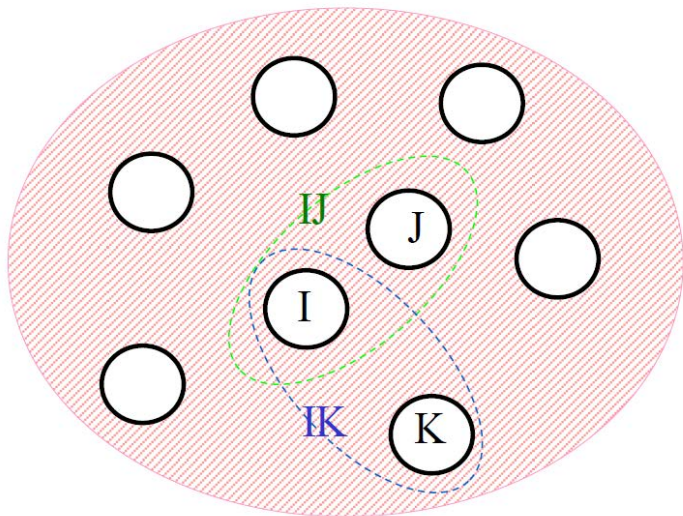
Hobza, Accounts of Chemical Research, 663, 2012
<http://www.begdb.com>

タンパク質に対するQM計算法 2) FMO



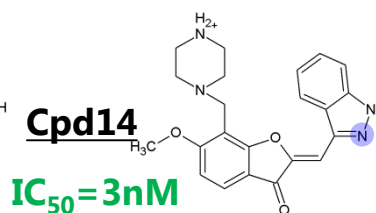
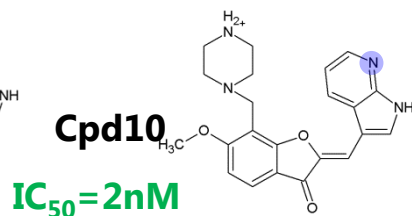
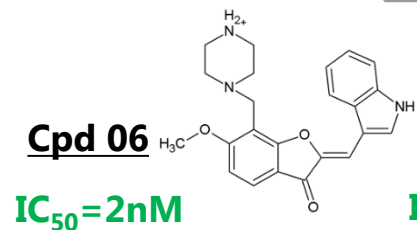
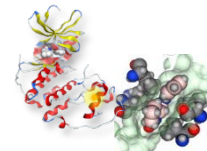
1999年、北浦らは、全く新しい、大きな系に対する効率的な量子力学計算法であるFMO(fragment molecular orbital)法を提案した。

FMO法は、タンパク質分子をアミノ酸単位に分割し、それらのアミノ酸単独と隣接したアミノ酸の組のエネルギーを組み合わせることにより、精度良く計算値を再現できる。

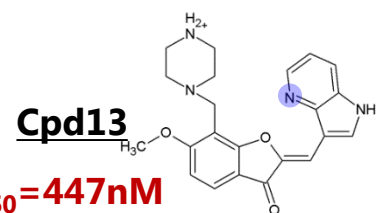
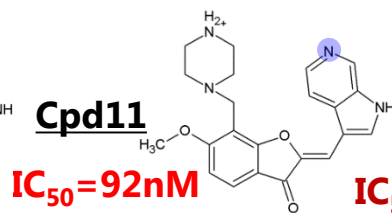
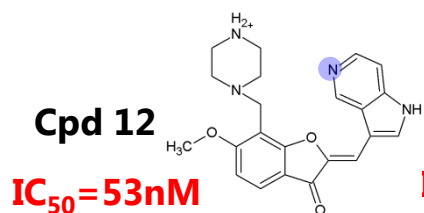


$$E = \sum_I E'_I + \sum_{I>J} \Delta\tilde{E}_{IJ}$$

Activity Cliffの例 Pim1阻害剤

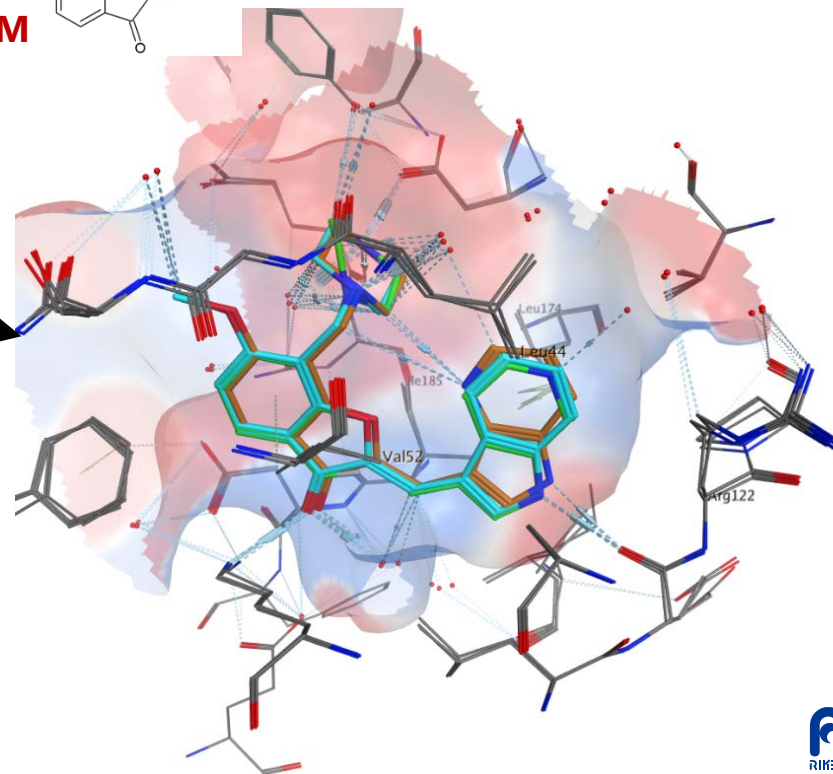


一原子異なるだけで、
 IC_{50} 値が、2-200倍
異なる！

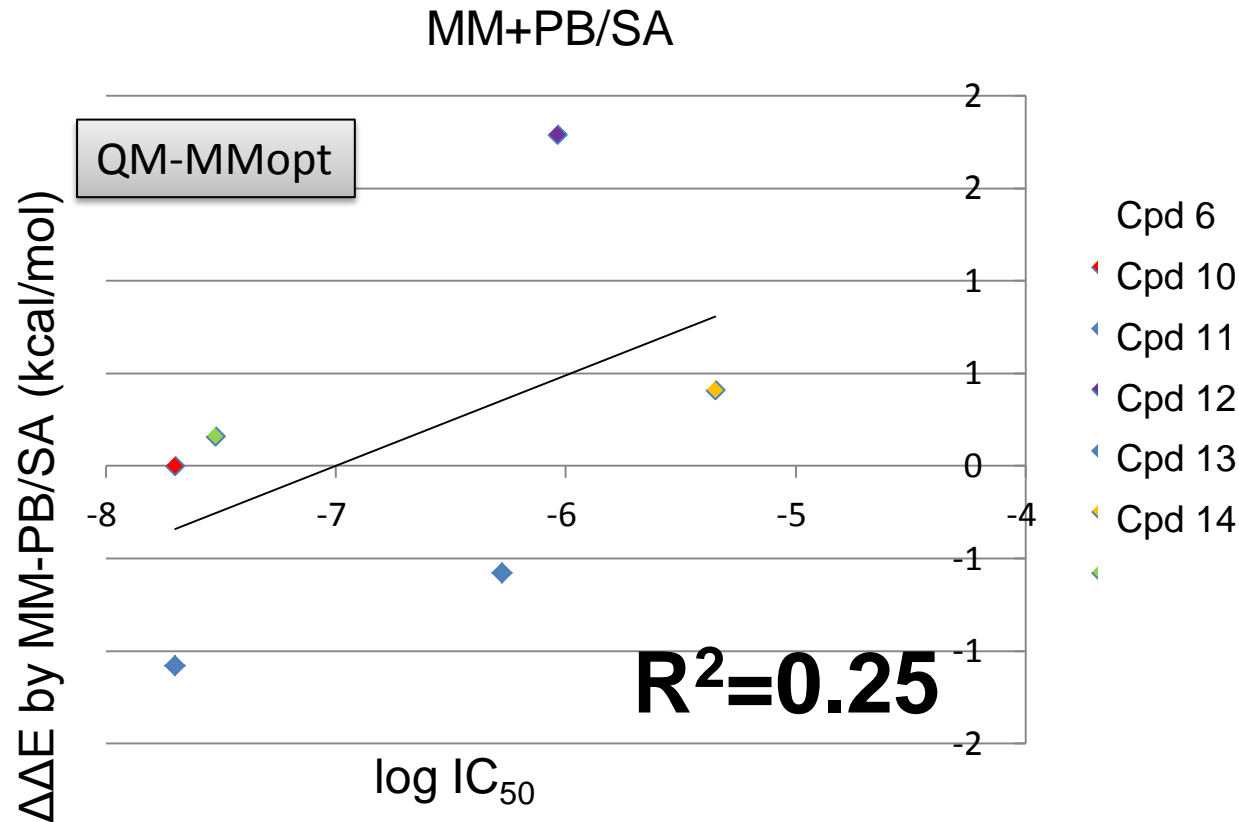
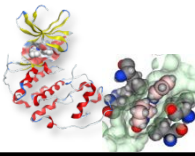


J. Med. Chem. **55** (2012) 5151-5164.

上記のうち、4化合物
(下線)のX線構造は、
ほとんど同じ！



MM-PBSAで予測できる？



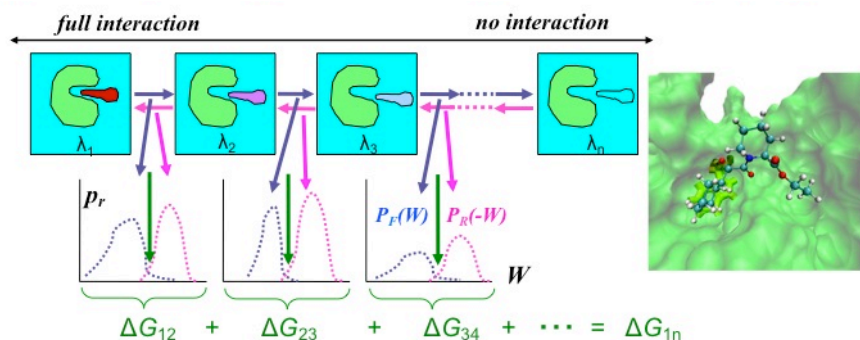
MM-PBSAでは、若干の傾向はあるものの、相関は非常に悪い
ちなみに、計算に使う構造としてMMoptとQM-Mmoptを使った場合は、 $R^2=0.25$ よりもさらに悪い

FMOからFMO-PB/SAへの展開

Massively Parallel Computation of Absolute binding Free energy with well-Equilibrated states (MP-CAFREE)

	ΔH	ΔS	Force field	Conf. Sampling
Protein	Fair	Fair	Fair	Good
Ligand	Fair	Fair	Fair / Bad	Good
Solvent	Fair	Fair	Fair	Good

Alchemical methods for calculating absolute Kd values



Nonequilibrium equality (Jarzynski, 1997)

$$\Delta G = -RT \ln \left\langle \exp \left(-\frac{W}{RT} \right) \right\rangle \quad \therefore K_d = \exp \left(\frac{\Delta G}{RT} \right)$$

Fujitani et al, JCP (2005)

QM/MM (Quantum Mechanics/Molecular Mechanics), Fragment Molecular Orbital (FMO) Method

	ΔH	ΔS	Force field	Conf. Sampling
Protein	Good	Not Considered	Good	Not Considered
Ligand	Good	Not Considered	Good	Not Considered
Solvent	Bad	Not Considered	Good / Not Considered	Not Considered



Fragment Molecular Orbital Poisson-Boltzmann Surface Area (FMO-PB/SA)

	ΔH	ΔS	Force field	Conf. Sampling
Protein	Good	Not Considered	Good	Fair / Bad
Ligand	Good	Fair / Bad	Good	Fair / Bad
Solvent	Good / Fair	Fair / Bad	Good / Fair	Fair / Bad

New Technology: FMO+MM-PBSA

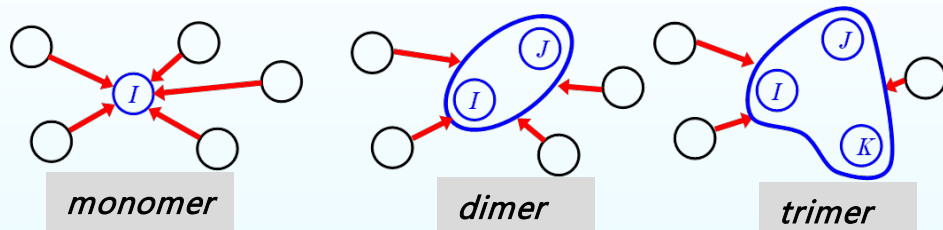


日大・福澤
先生との
共同研究

Fragment Molecular Orbital (FMO):

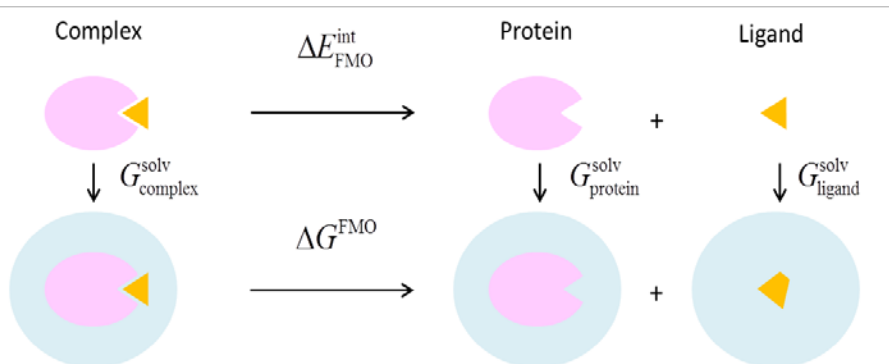
Quantum Mechanics Method for Protein

Kitaura, Chem. Phys. Lett. **523** (2012) 128.

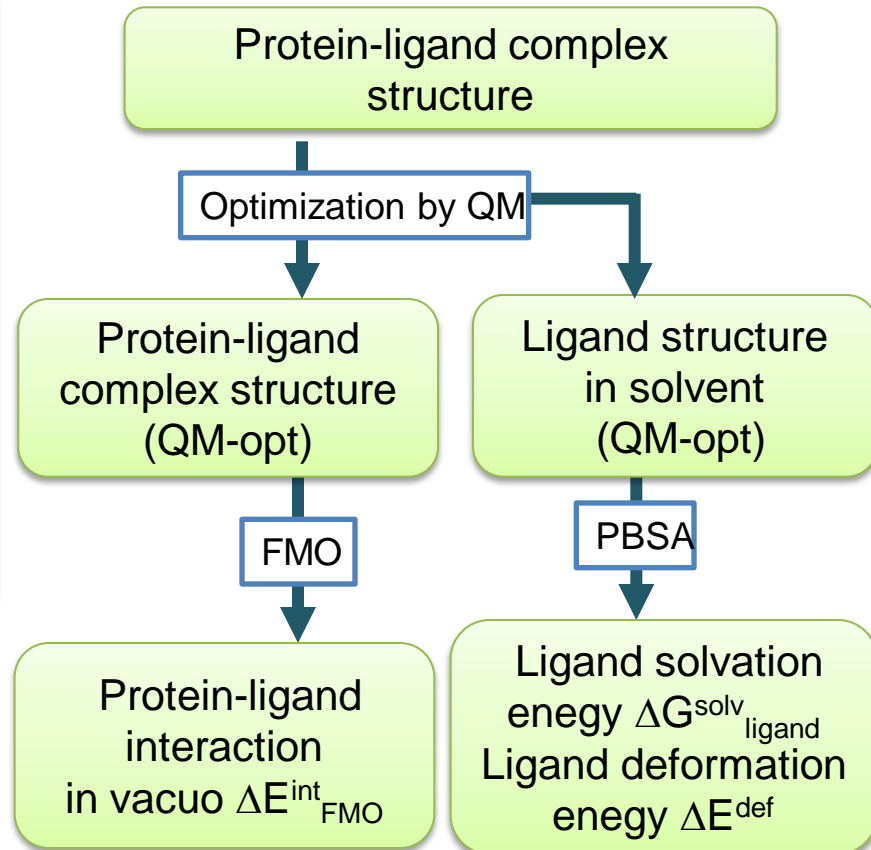


$$E_{\text{total}} = \sum_I E'_I + \sum_{I>J} \Delta \tilde{E}_{IJ} + \sum_{I>J>K} \Delta \tilde{E}_{IJK}$$

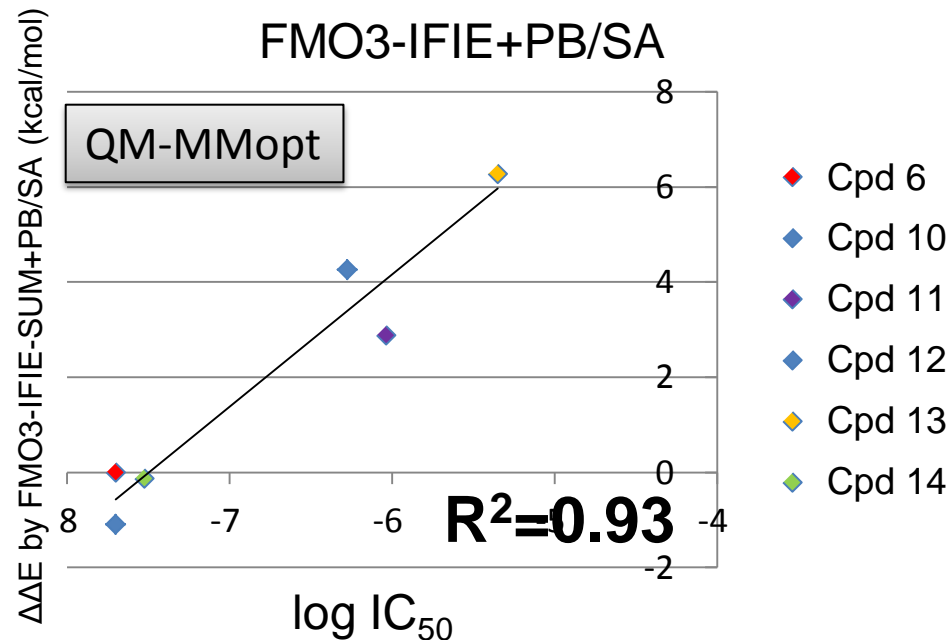
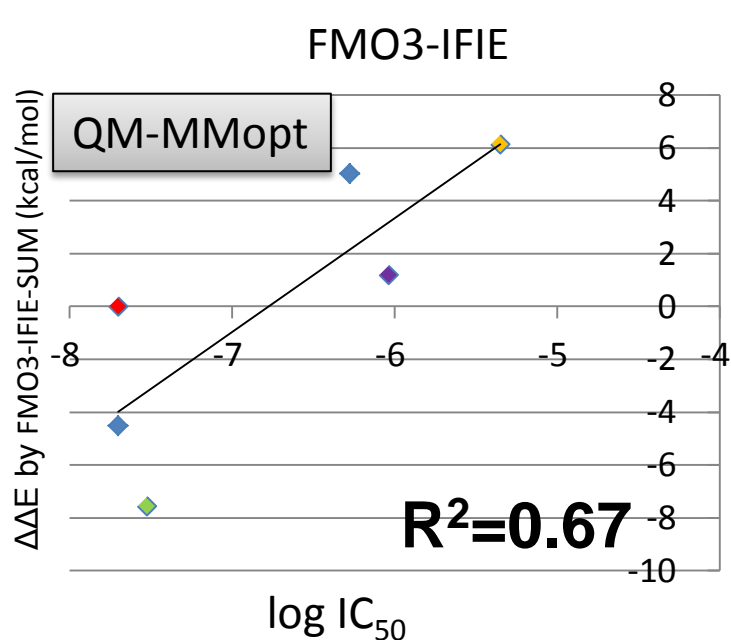
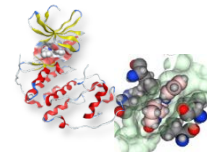
FMO2
FMO3



$$\Delta G^{\text{FMO}} = \Delta E_{\text{FMO}}^{\text{Int}} + \Delta E^{\text{def}} + C \times (\Delta G_{\text{complex}}^{\text{solv}} - \Delta G_{\text{protein}}^{\text{solv}} - \Delta G_{\text{ligand}}^{\text{solv}})$$

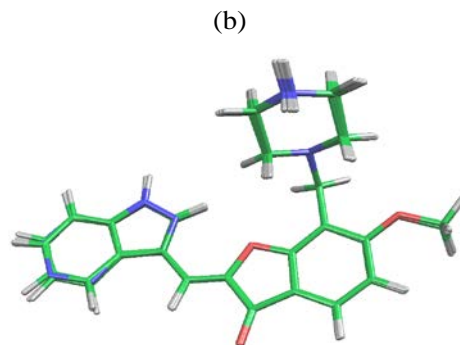
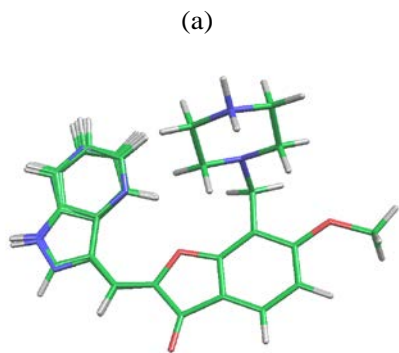


PBSA項の寄与

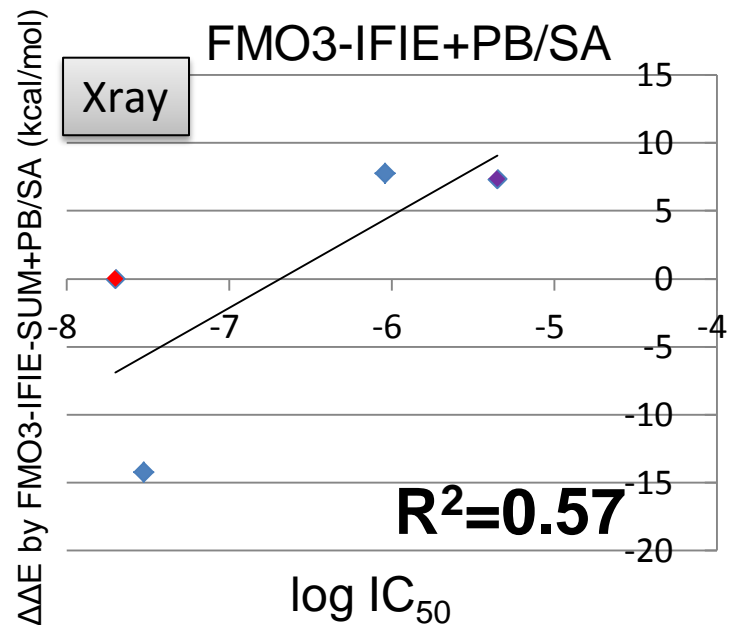


結合状態と解離状態の溶媒和の違いを考慮することは重要

現在、FMOで溶媒項(PB)を計算している

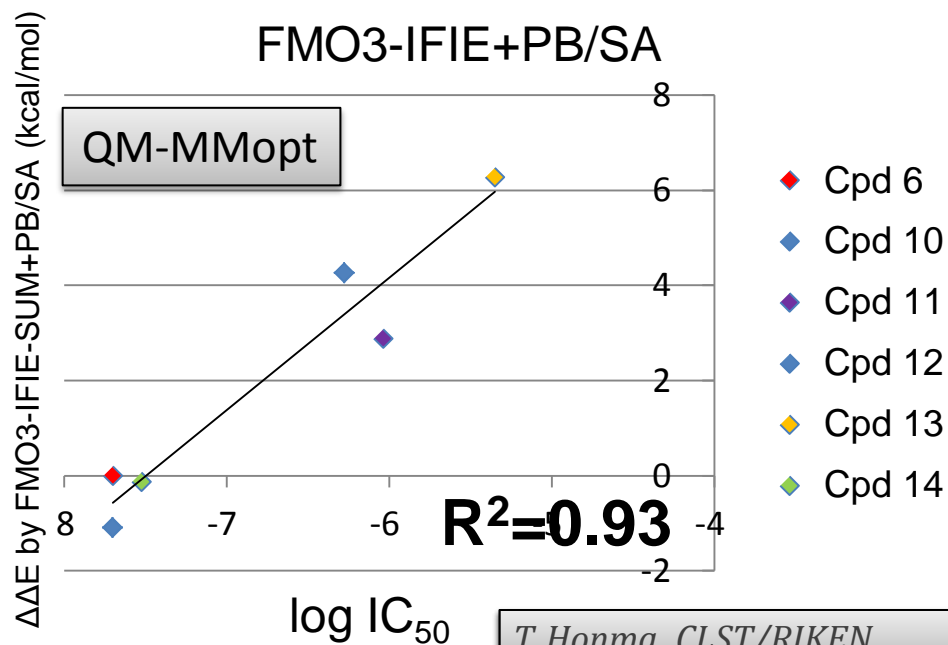
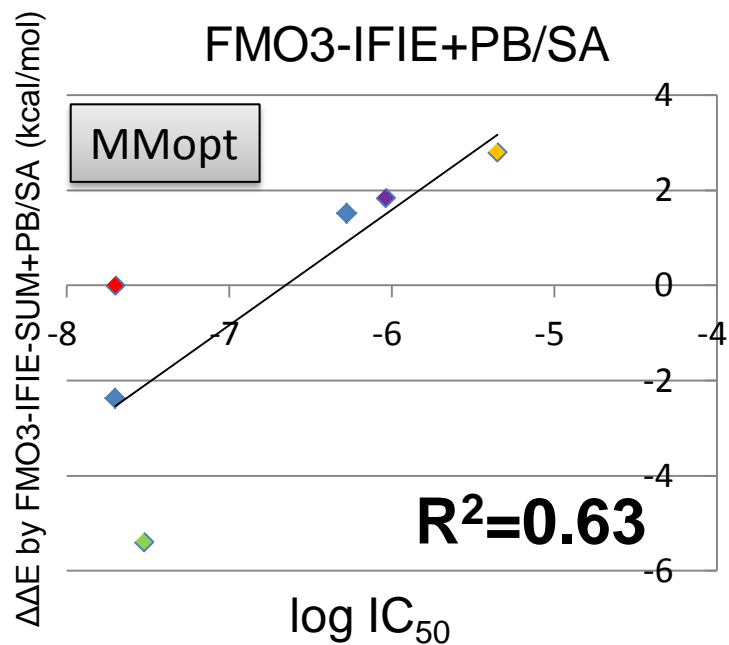


FMO+MM-PBSAの予測結果

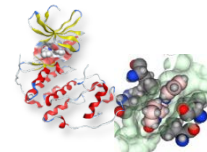


- ◆ Cpd 6
- ◆ Cpd 11
- ◆ Cpd 13
- ◆ Cpd 14

構造精密化のレベルを
上げていくとともに相関
係数が改善！

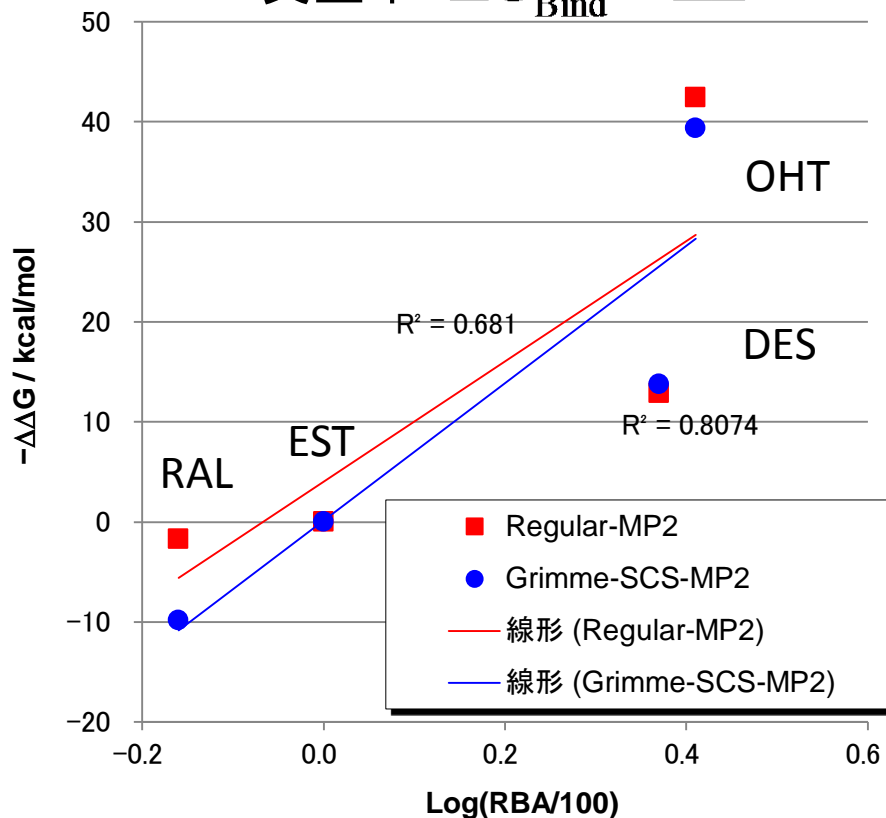


エストロゲン受容体に対するFMO-PBSA

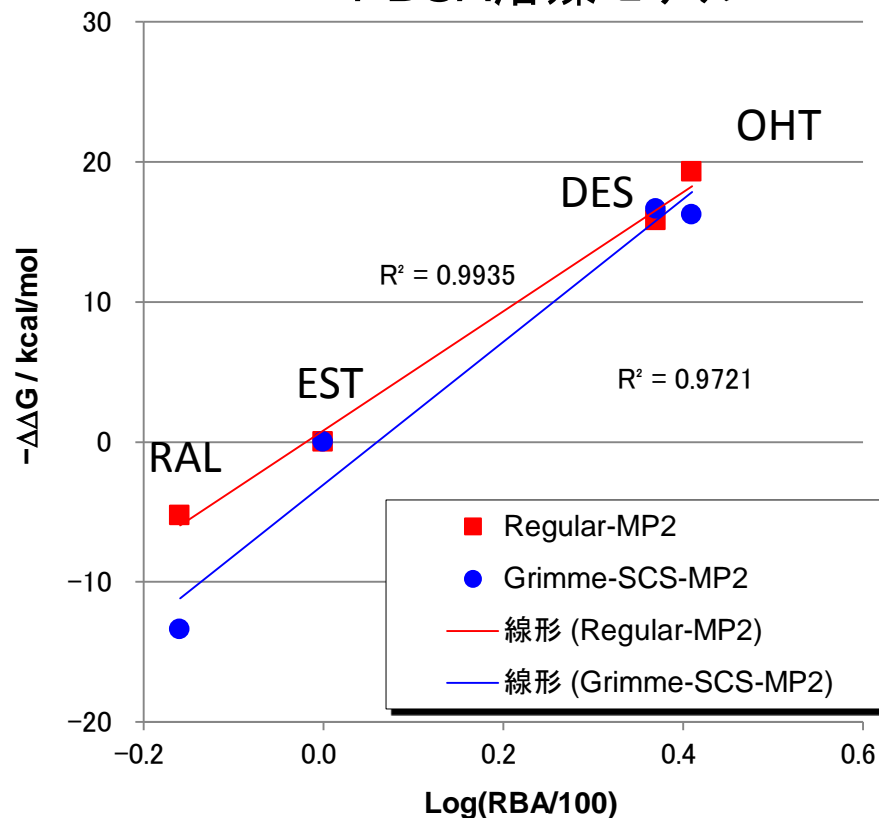


$$\Delta\Delta G = \Delta G_{\text{Bind}}^{\text{Ligand}} - \Delta G_{\text{Bind}}^{\text{EST}}$$

真空中 $\Delta G_{\text{Bind}} = \Delta E$



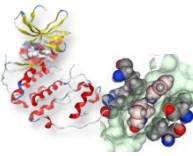
PBSA溶媒モデル



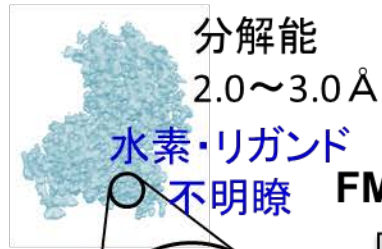
課題

- 誘電率・原子半径の選択
- ソルバーの高速化(収束性、並列化)
- ...

FMO法によるX線構造決定法の開発



X線電子密度

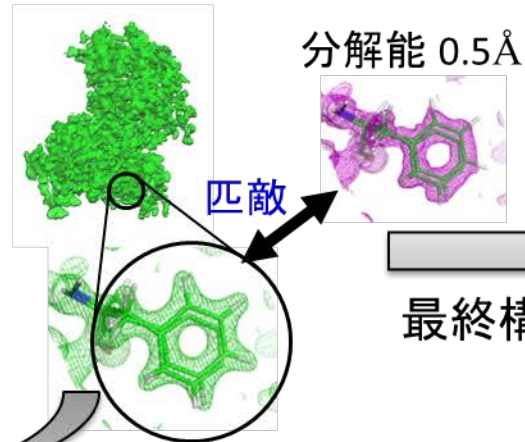


$2\rho_o - \rho_{FMO}$

FMO法に基づく構造決定

モデル最適化

FMO電子密度

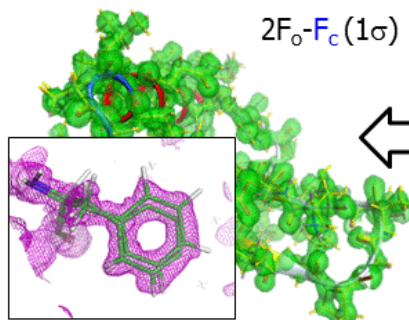


最終構造

超分解能構造

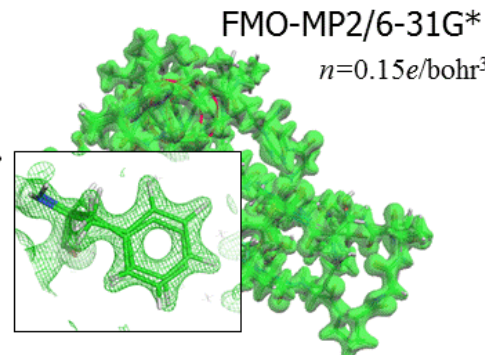


X-ray electron density



$0.7456e/\text{\AA}^3$ (rmsd^{X-ray}=1.00)

FMO-based electron density



$0.7456e/\text{\AA}^3$ (rmsd^{FMO}=0.76)



代表幹事： 福澤 薫（星薬科大学）



副代表幹事： 田中 成典（神戸大学）



副代表幹事： 本間 光貴（理研）

参加機関

- 製薬企業 12社
- IT企業 1社
- アカデミア 7機関

研究内容

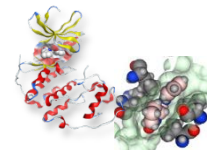
- FMO-PBSA計算のプロトコルの作成と自動化
- 精度の大規模検証
- FMOによるX線構造決定法の開発



京利用のHPCIの申請が採択されたので、**500個以上の複合体の計算**を行って、精度検証や手法の改善を開始した。

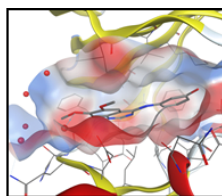
製薬企業が現場で使える技術としてFMO-PBSAのプロトコルを公開できるようにしたい。

FMOの自動計算プロトコール



Automated search
of activity values

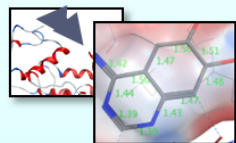
- IC50, Ki, Kd
- PDB with activity data



X-ray structures
from PDB
or docked ones

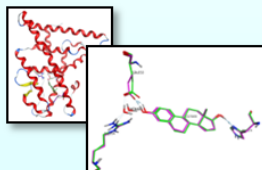


Automation by Pipeline pilot



Structure
construction

- Complementation of missing atoms/residues
- Modifying unfavorable structures



Structure
optimization

- MM method
- QM/MM method



FMO and
MM-PBSA input
generation

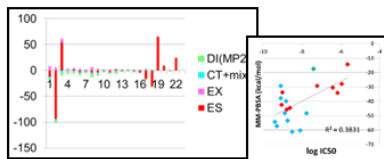


Parallel
computation

- FMO calculation
- MM-PBSA calculation



Results extraction
/visualization



Automated analysis
with Excel macro etc.

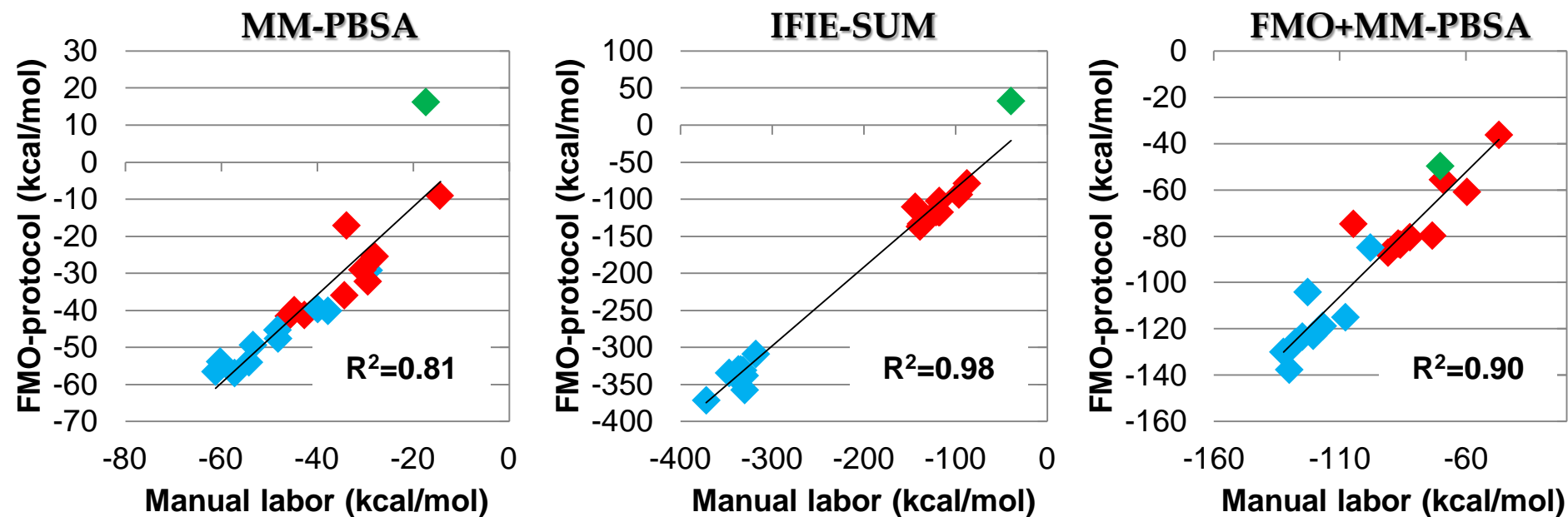
Construction of
IFIE-database

- IFIE/PIEDA
- VISCANA
- NPA charge
- Used structures
- Affinity data

Validation of FMO-protocol data

Estrogen receptor α (38 structures with 22 types of compound)

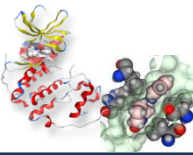
Coefficient of determination between FMO-protocol and manual labor data [Anzaki's data]



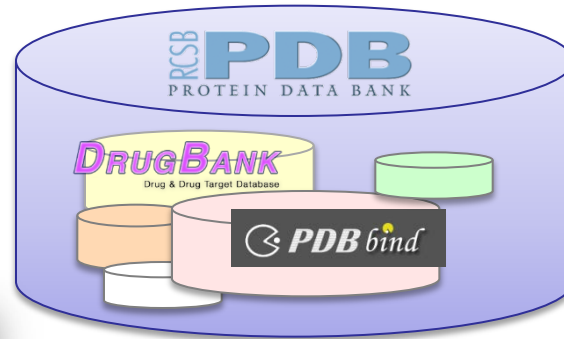
➤ **Good agreement between FMO-protocol and manual labor data.**

◆ **Negative charged ligand**
◆ **Neutral ligand**
◆ **Positive charged ligand**

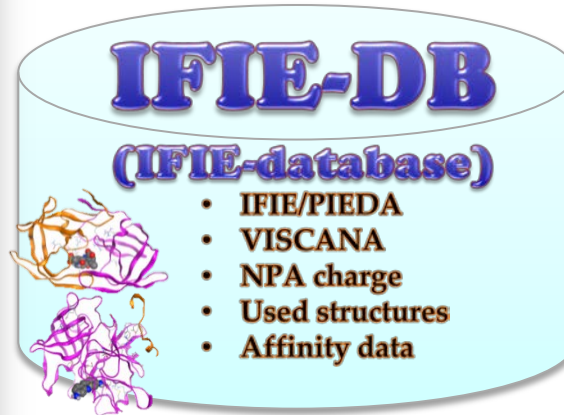
FMO IFIE database



PDBの全複合体データのFMO計算結果を収載するデータベースを計画している。現在、基本設計が終わったところ。来年度、世界中の研究者が使えるように公開する。



Semi-automated
FMO calculation
protocol
& Data of FMO DD



主な用途

FMO計算結果の解析
(IC50との相関、
PIEDAなど)

FMO力場開発の
基礎データ

医薬品設計への利用
(新しい置換基の
提案など)

CBIジャーナルなど
への投稿の際に利用
(PDBのデポジットに
近い感覚)

FMO IFIE DATABASE

1ERE
HUMAN ESTROGEN RECEPTOR LIGAND-BINDING DOMAIN IN COMPLEX WITH 17BETA-ESTRADIOL
PubMedID: 9338790
Classification: NUCLEAR RECEPTOR
Deposited: 1997-05-09 Released: 1998-05-18
Structure author(s): Bircanowski, A.M., Pike, A.C., Dauder, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A., Carlquist, M.
Method: X-RAY DIFFRACTION
Resolution: 3.10

Download PDB Download CheckPoint File

Summary of PIEDA

Total	ES	EX	CT+mix	D(MPZ)	q(=>J)
-108,283	-88,985	77,868	-34,965	-67,402	0,145

Detail of PIEDA

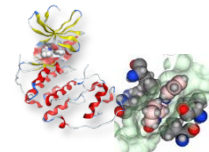
Filtering
Dist: \$
Total: ES, EX, CT+mix, D(MPZ), q(=>J)
seq, RES

Sort
Descending

Graph Options
Y Axis Max, Y Axis Min

Submit

Outline



1. インシリコスクリーニングの技術開発

2. インシリコスクリーニング・設計の実例

3. 結合親和性予測／FMOの創薬活用

4. 体内動態・毒性の予測に関して