

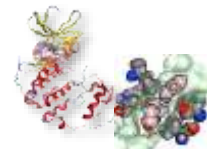
インフォマティクスとシミュレーションを 融合したインシリコスクリーニング



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？



insight review articles

Navigating chemical space for biology and medicine

Christopher Lipinski¹ & Andrew Hopkins²

¹Merck Global R&D, Green Laboratory, Kenilworth, New Jersey, USA (e-mail: chrolipin1_at_lipinski@greenpfmr.com)
²Merck Global R&D, Sandwick Laboratory, King's College London, London, UK (e-mail: andrew.hopkins@kcl.ac.uk)

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 10⁷ compounds) offer only a cursory

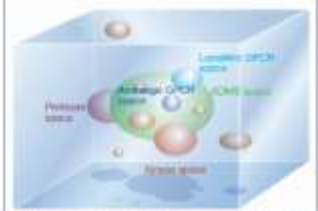
examination of all the possible organic compounds that comprise 'chemical space' (Fig. 1). Chemical space is for all practical purposes infinite and limited only by the chemist's imagination.

Not all biologically active compounds have the desired physicochemical properties to be a drug. A biologically active compound may be too lipophilic (poorly) 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 10²⁴ possible derivatives of n-heptane – if we use a list of only 150 substituents and consider mono- to 14-substituted heptanes²². 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 'stratigraphy', 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 properties²³. Given the vastness of chemical space, the challenge for chemical



Box 1 Figure The figure depicts a carbon replacement of the

100万 1兆 10⁶⁰

網羅的に計算すると **10⁶⁰** 個程度
存在し得ると推定されている。

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

ちなみに

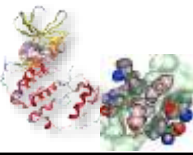
観測可能な宇宙にある「星」（恒星）の数
1000億個の銀河 * 1000億個の星 = 10²²

観測可能な宇宙に陽子をすきまなく埋めた場合の数
10⁶⁰

ただし実際にこれまで人間が天然・人工で

単離した化合物は **10⁸ ~ 10⁹** 個程度

医薬品のターゲットスペース



Human genes (~25,000)

Druggable genes
(~3,000)

Disease genes with
marketed drugs (215)

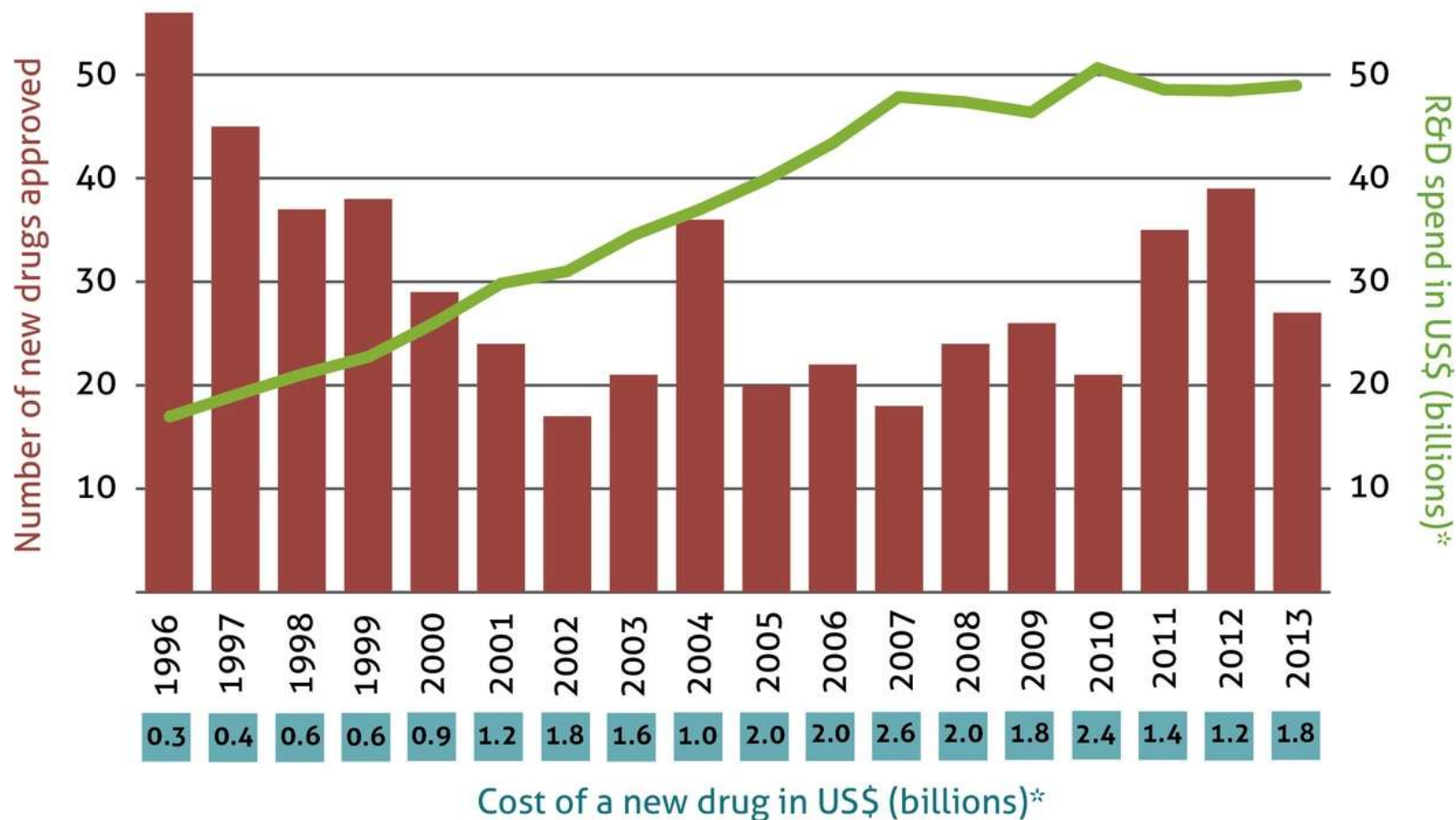
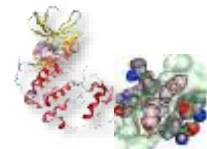
215

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。

医薬品の開発費(米国)



Data: USFDA, PhRMA

Akshat Rathi | theconversation.com

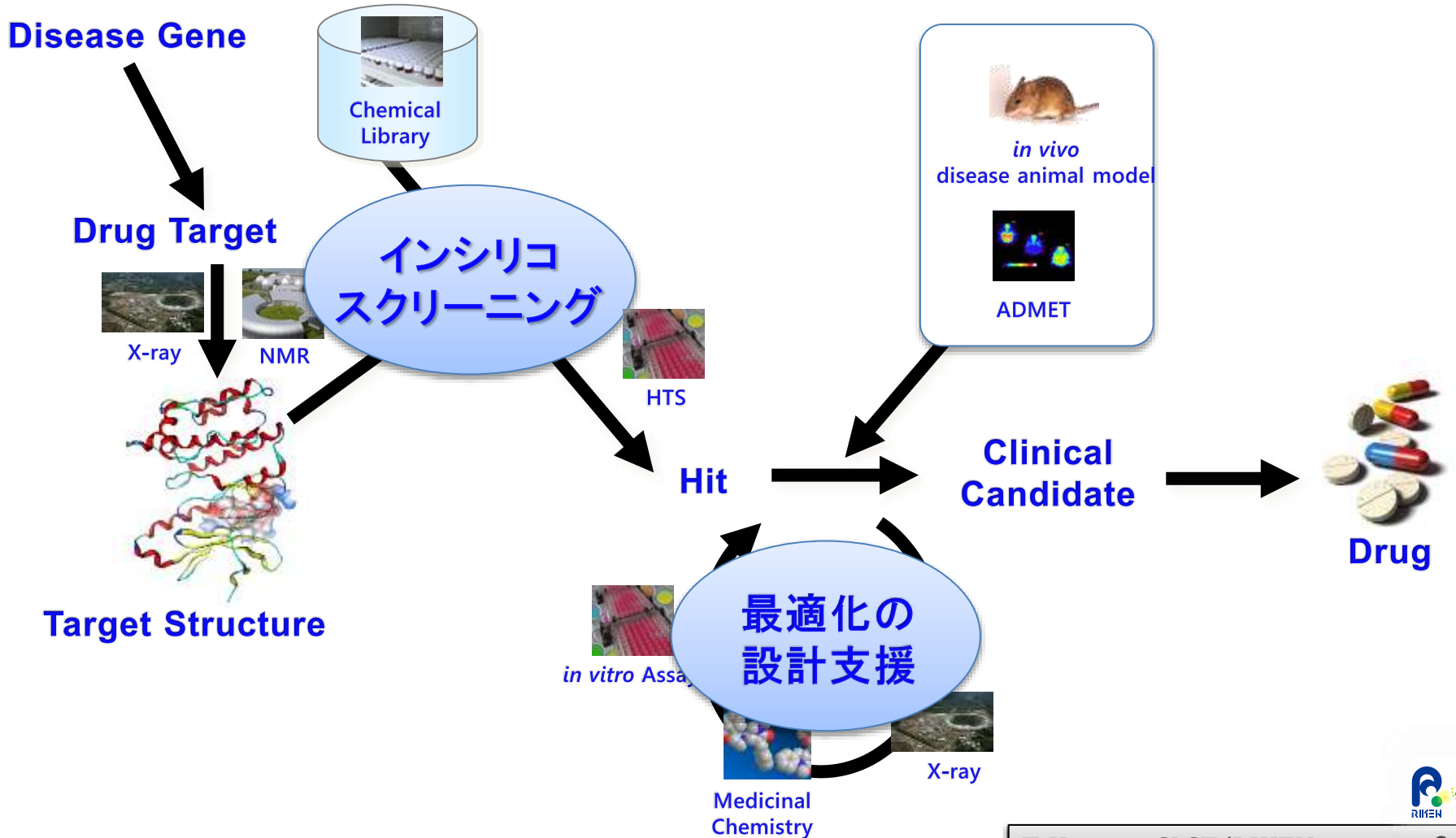
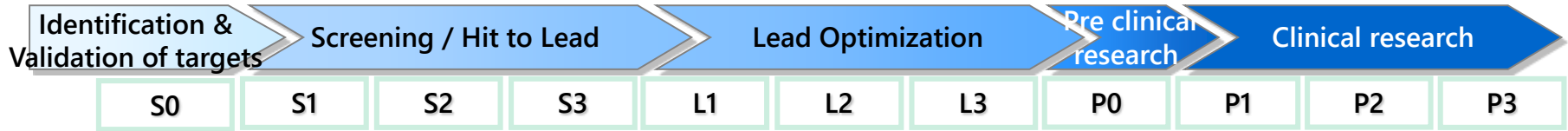
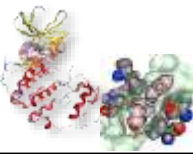
* New drug cost and R&D spend could be 30% higher if non-PhRMA members are included



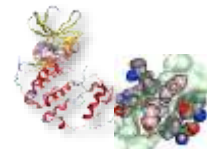
本当に薬なんて
できるの？

(© Dr. S. G. G. Hoffmann-La Roche 2003)

創薬におけるインシリコ役割



インシリコスクリーニング

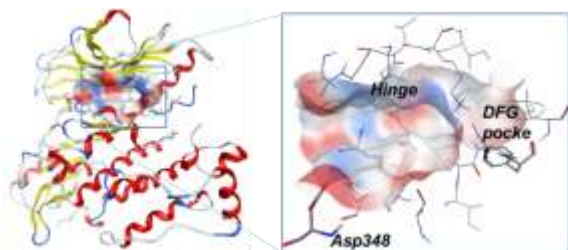


Simulation

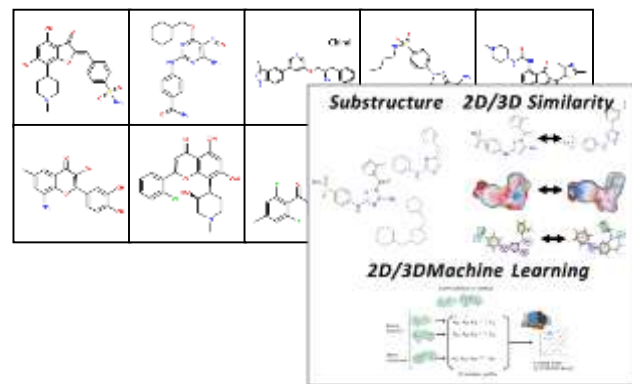
Infomatics

Docking

Similarity Search



Compounds with
Docking Poses
1,00 – 10,000



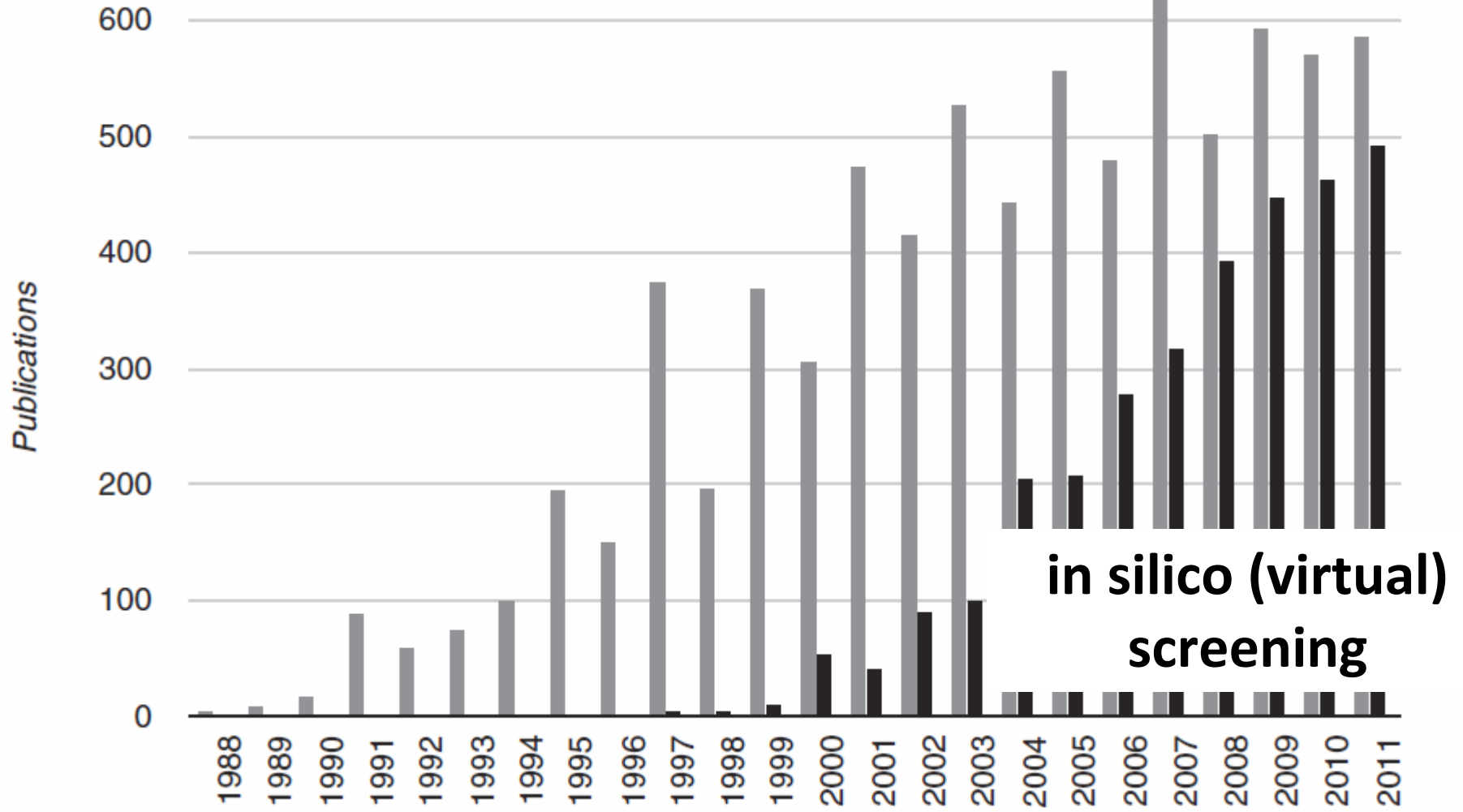
MD

QM

Machine Learning

Selected
Compounds
100-1,000

HTSは90年代、インシリコは2000年代から重要な創薬技術に



HTS, インシリコスクリーニングの論文数



Virtual screening



Virtual screening is a computational approach used to identify chemical structures that are predicted to have particular properties. For example, in the context of drug discovery, it may involve computationally searching large libraries of chemical structures to identify those structures that are most likely to bind to a drug target.

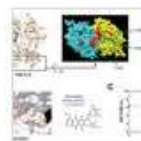
Latest Research and Reviews

Research | 16 September 2016 | [OPEN](#)

Identification of Doxorubicin as an Inhibitor of the IRE1 α -XBP1 Axis of the Unfolded Protein Response

Dadi Jiang, Connor Lynch [...] Albert C. Koong

Scientific Reports **6**, 33353

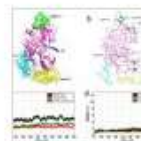


Research | 17 August 2016 | [OPEN](#)

Understanding the structural basis of substrate recognition by *Plasmodium falciparum* plasmepsin V to aid in the design of potent inhibitors

Rajiv K. Bedi, Chandan Patel [...] Prasenjit Bhaumik

Scientific Reports **6**, 31420

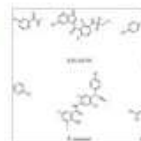


Research | 09 August 2016 | [OPEN](#)

Repositioning organohalogen drugs: a case study for identification of potent B-Raf V600E inhibitors via docking and bioassay

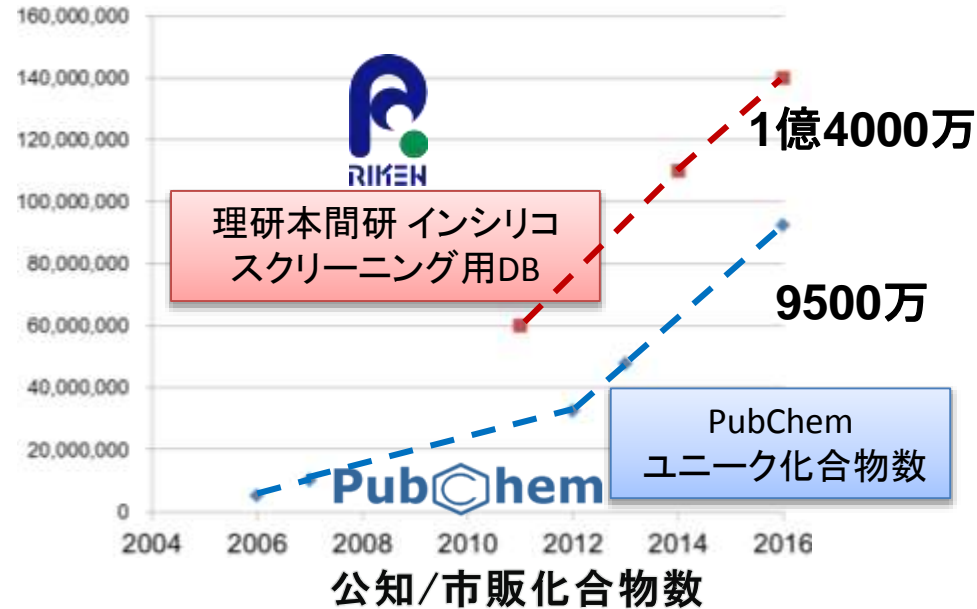
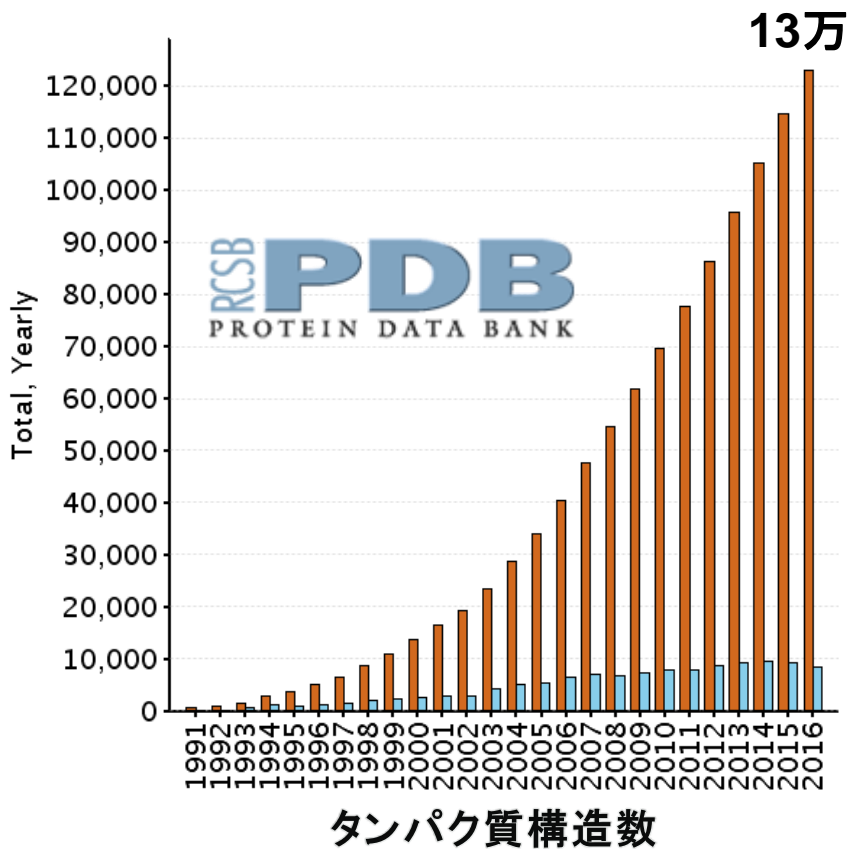
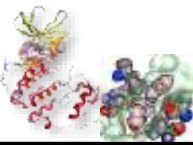
Yisu Li, Binbin Guo [...] Weiliang Zhu

Scientific Reports **6**, 31074



現在では、
Nature及び関連誌にも毎月のようにvirtual screeningを活用した創薬事例が報告されている

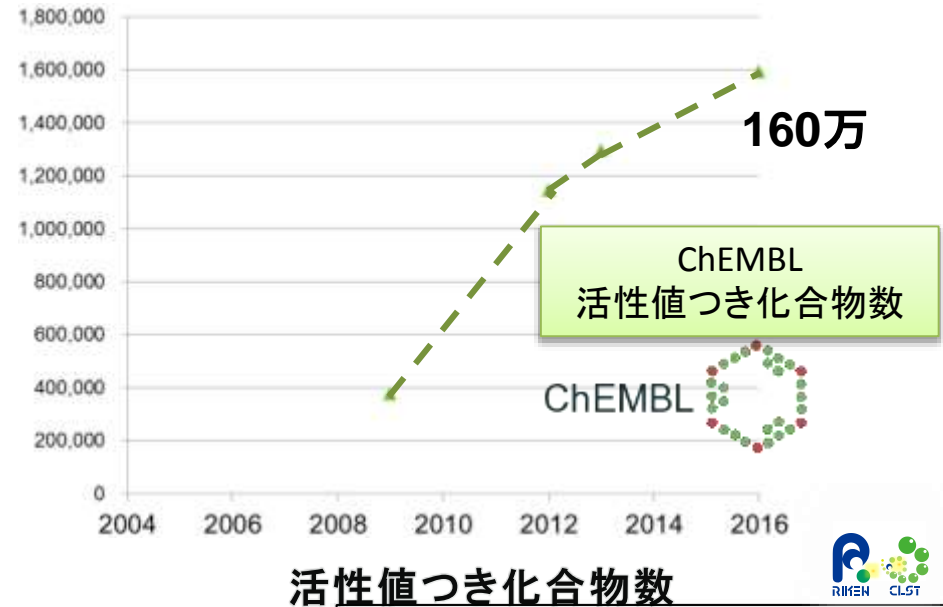
創薬情報のビッグデータ化



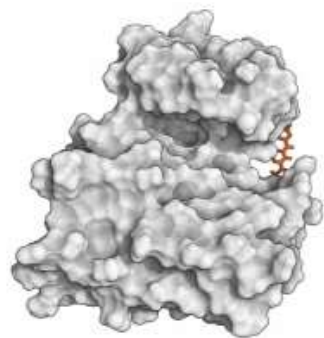
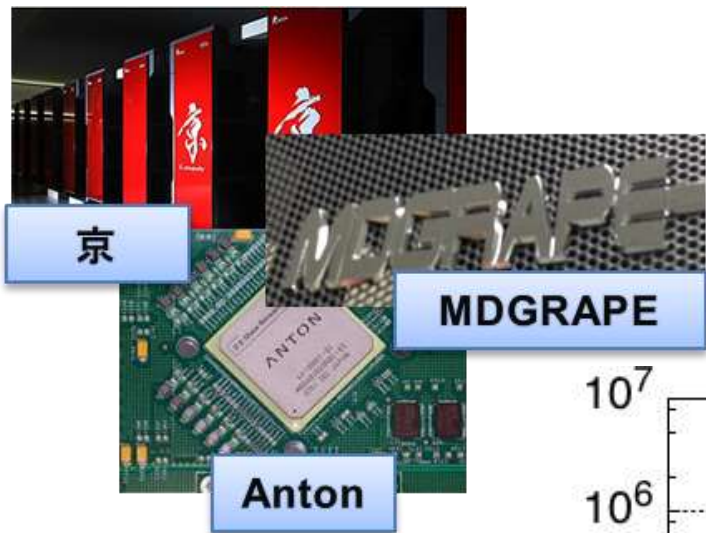
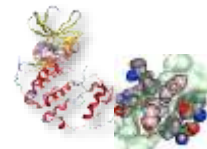
タンパク質構造、化合物情報ともに、
2005-10年から急増し、ビッグデータとして
解析できる環境が整いつつある



機械学習、シミュレーションへの活用

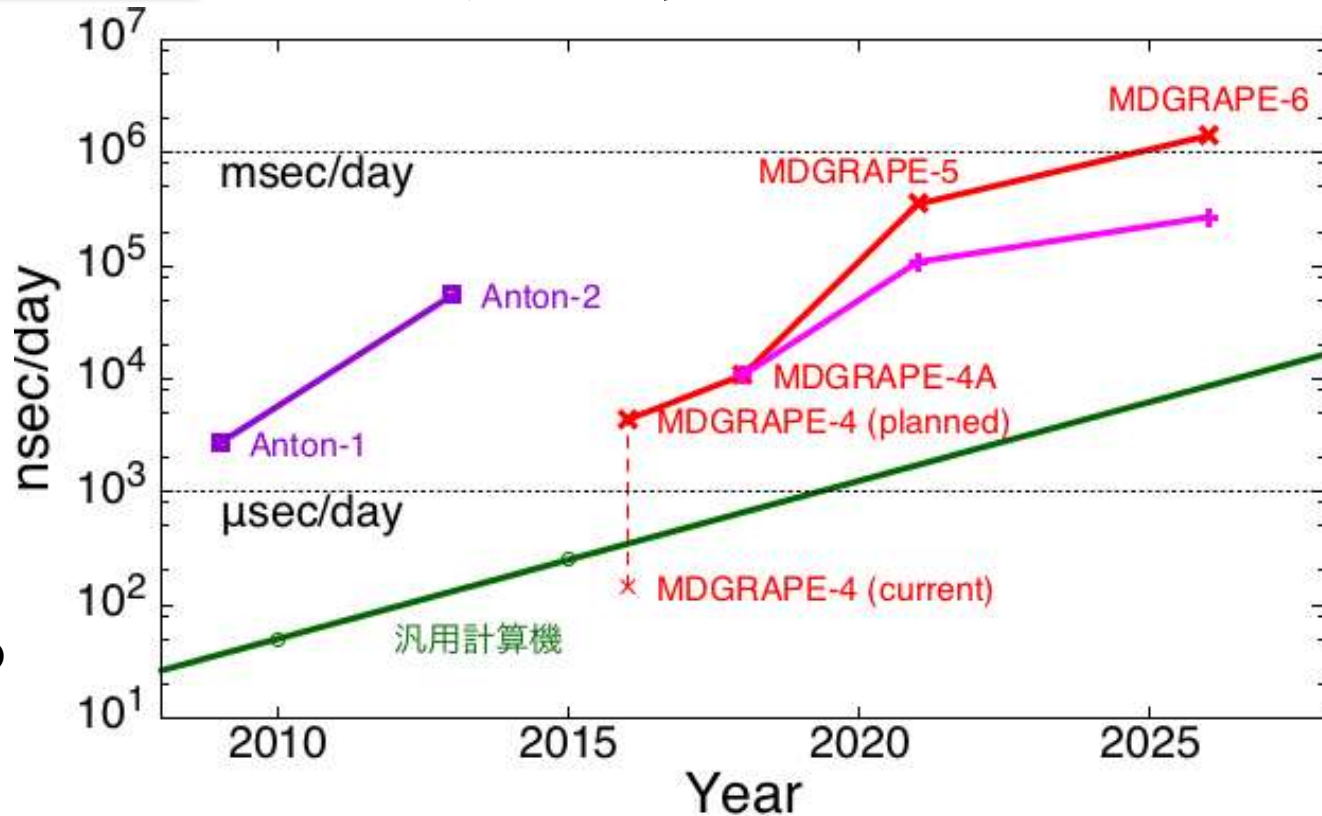


計算機の高性能化

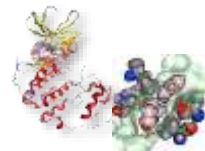


数10マイクロ秒スケールのシミュレーションが可能に！

分子動力学計算の性能のロードマップ



Informatics vs Simulation

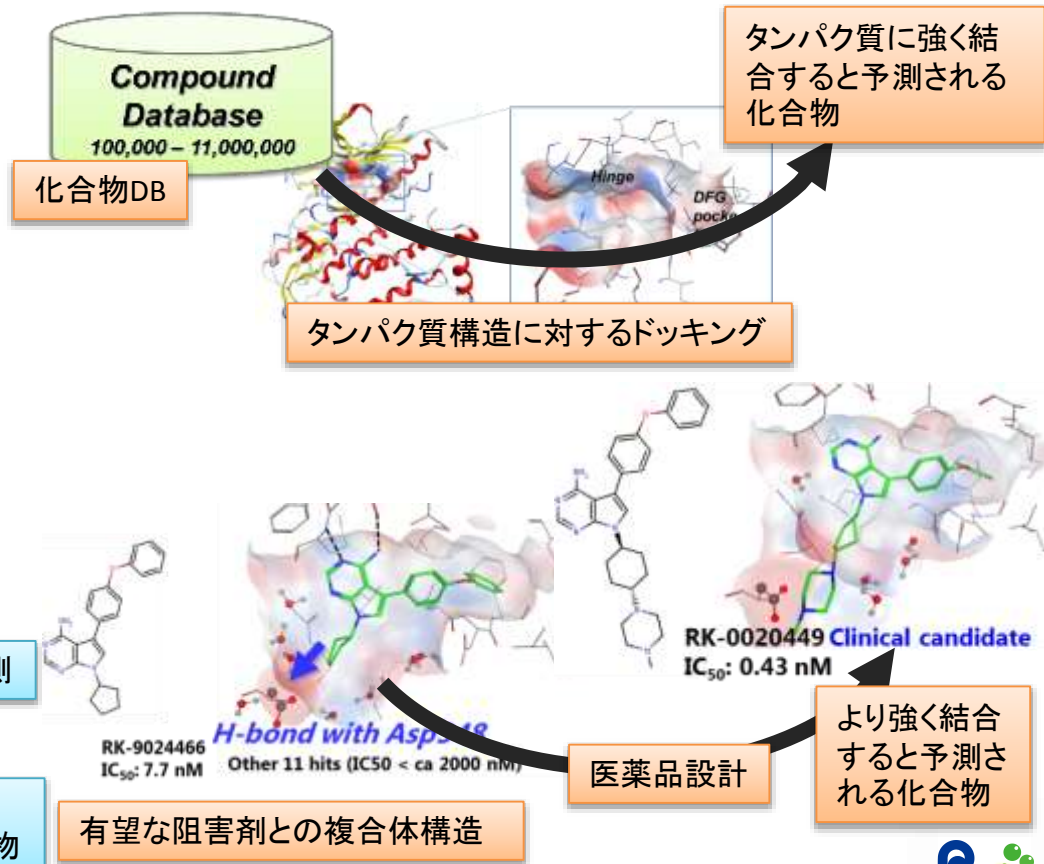
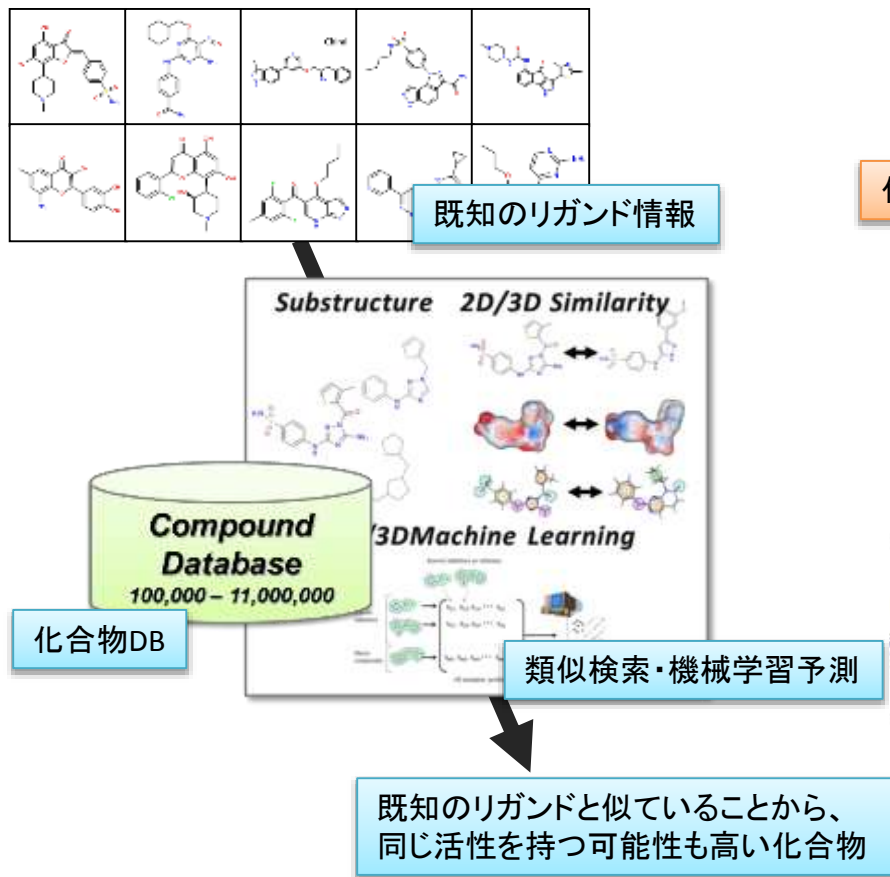


Informatics (LBDD)

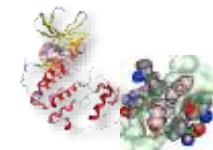
Ligand-based drug discovery(design)
リガンド情報に基づいた設計

Simulation (SBDD)

Structure-based drug discovery(design)
タンパク質構造に基づいた設計



Informatics vs Simulation in Pfizer



Sandwich, UK

- **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
世界的権威がたくさん…

必要性を感じない。
まずは、
Informatics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

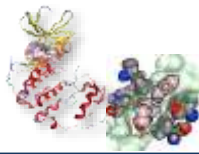
Pfizerの化合物*のドッキング用前処理済み
データを皆で準備して
共有しませんか？



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

Informatics vs Simulation in Pfizer



Simulationは時間がかかる割には、効果が薄い。Informatics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、Simulationの成功例も多い

Sandwich, UK

Chicago, US
Ex-Warner-Lambert

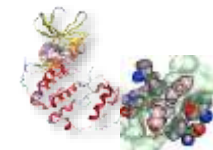
Nagoya, Japan

Groton, US
Original Pfizer

Boston, US

La Jolla, US
Ex-Agouron

Informatics vs Simulation



	利点	欠点
<i>Simulation</i>	<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>
<i>Informatics</i>	<ul style="list-style-type: none">計算時間が短いヒット率が安定している（失敗が少ない）	<ul style="list-style-type: none">元の情報と同じ結合サイト、結合モードの化合物しかヒットしない

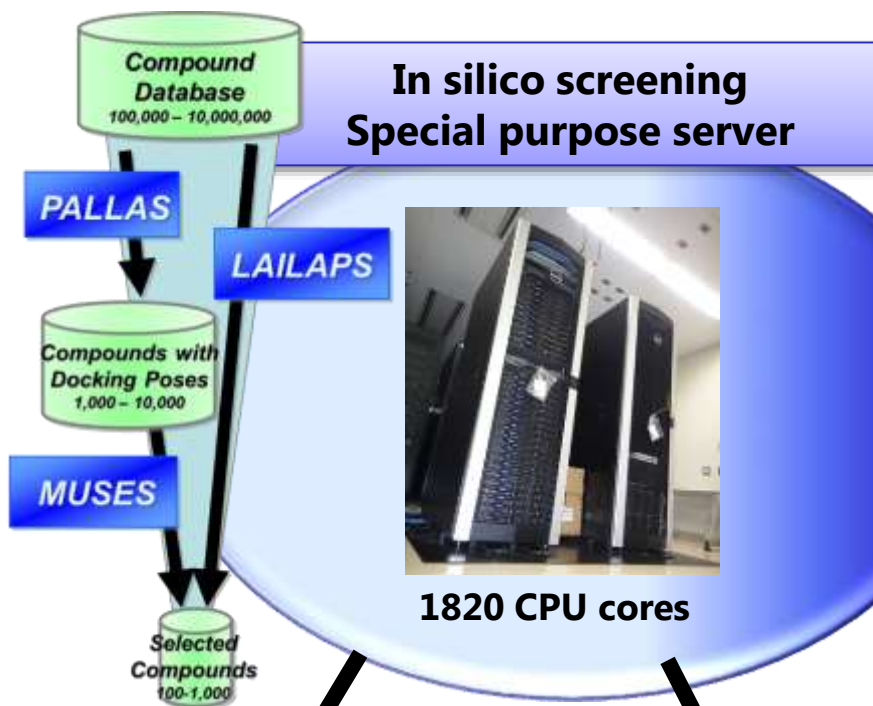
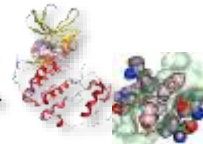


1. インシリコスクリーニングの技術開発

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

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

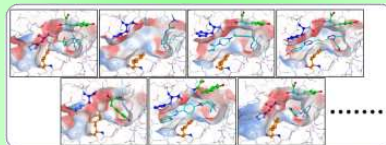
Compound Database

Preprocessed for in silico screening

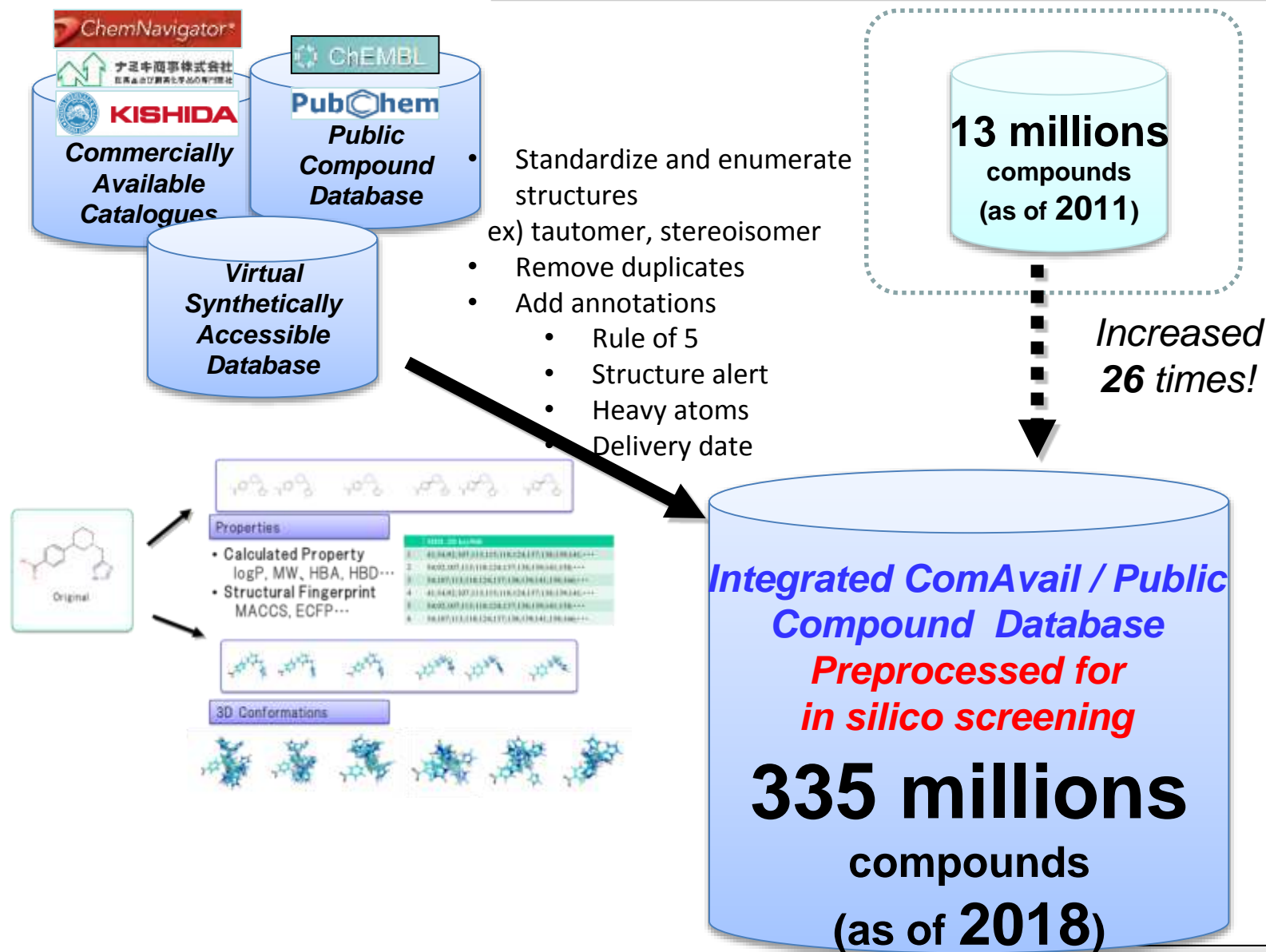
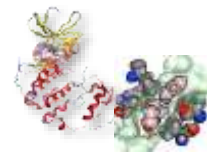
UT library: **200K cpds**

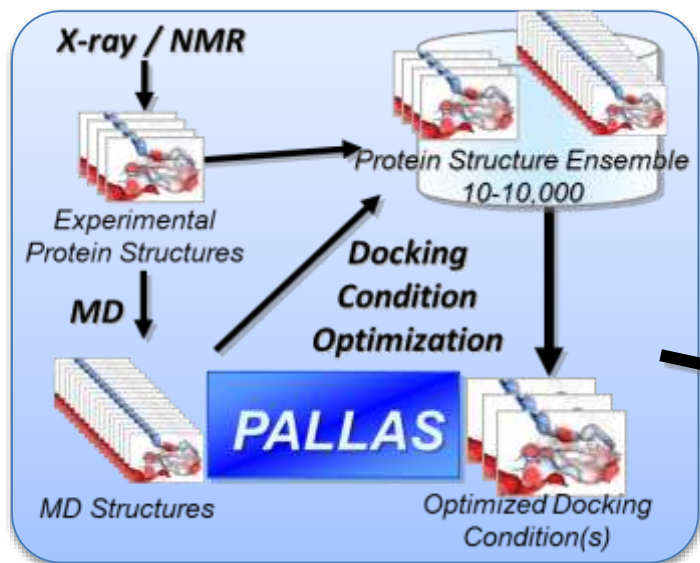
ComAvail, Literature: **146M cpds**

Protein Database

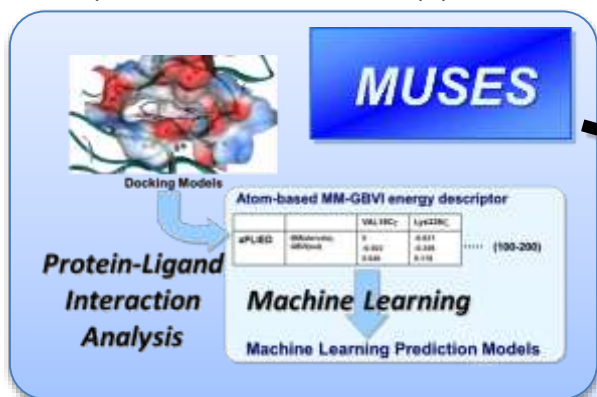


世界最大級(3億3500万化合物)の インシリコスクリーニング用データベース

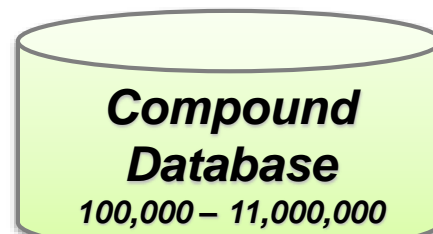
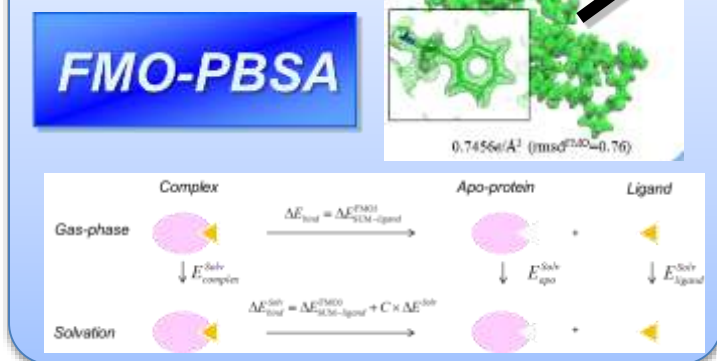




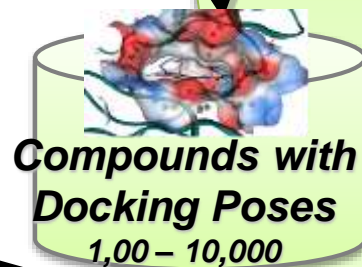
- 1) *Bioorg. Med. Chem.*, 2012, 20(12), 3756
- 2) *J Mol Biol*, 2012, 417(3), 240



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

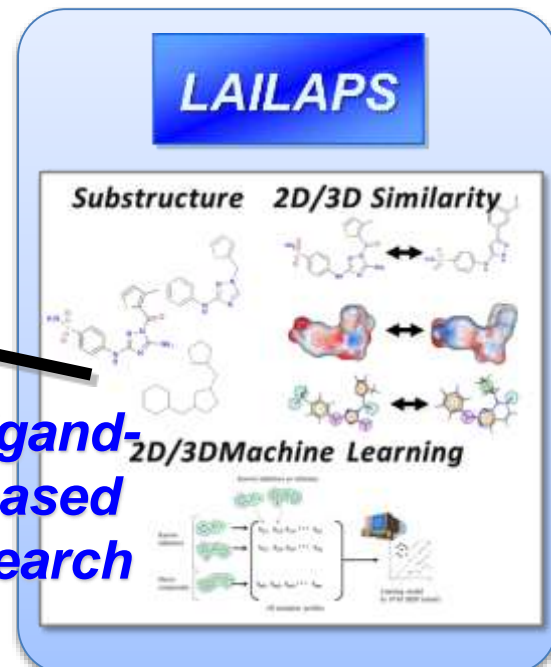


Docking

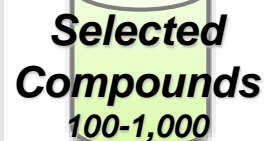


Affinity Prediction

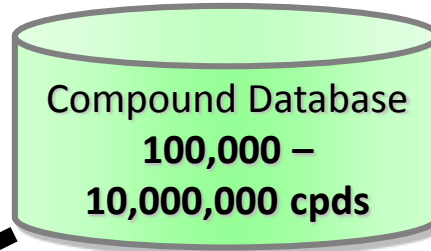
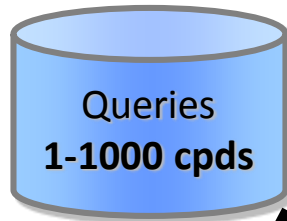
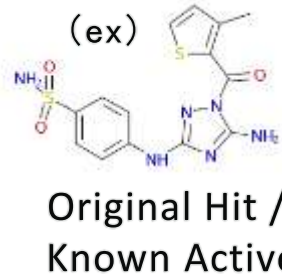
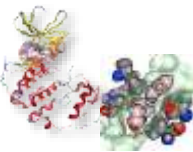
Ligand-based Search



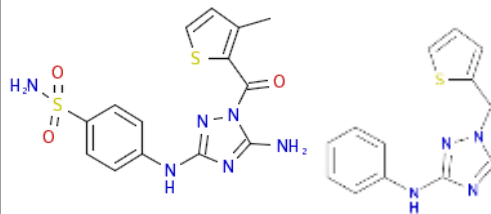
J Chem Inf Model, 2012, 52(4), 1015



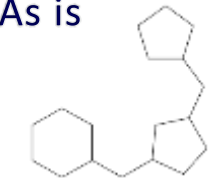
LAILAPSのしくみ



Substructure

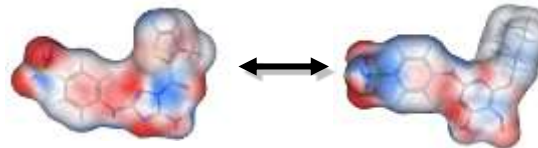
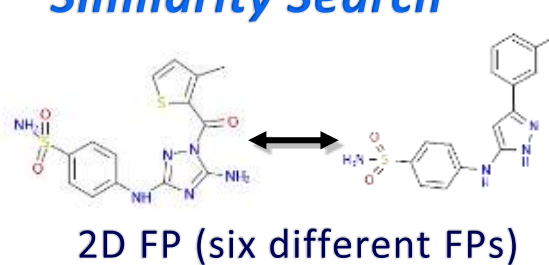


As is
Murcko
framework



Carbon Skelton
(in future)

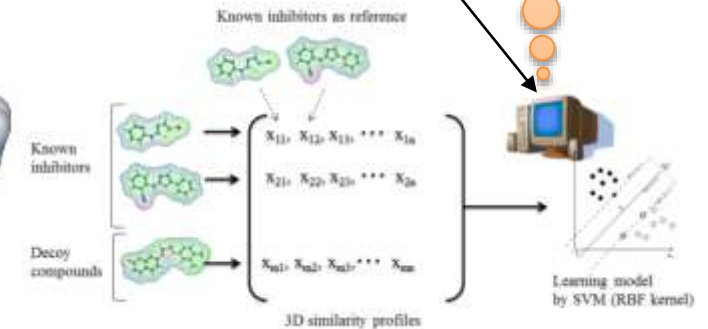
Similarity Search



3D shape and color
(ROCS)

Machine Learning

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



Machine Learning by 3D descs.
(in future)

Output (part of the scores are normalized)

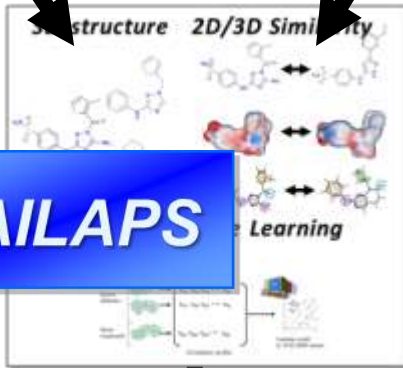
AdipoRonの発見と特許化



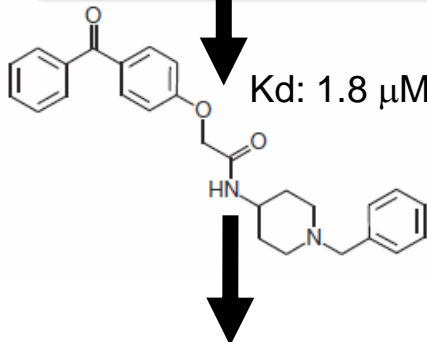
東大・門脇
先生との
共同研究

Known AMPK
activator, Membrane
protein ligands

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



LAILAPS



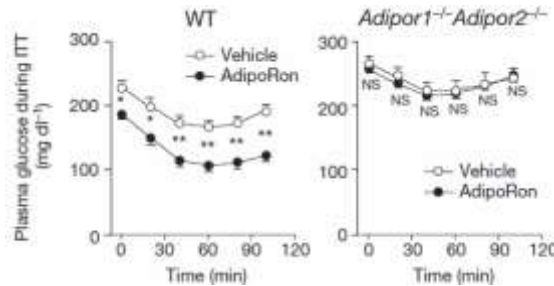
特許のクレーム化合物の設計



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



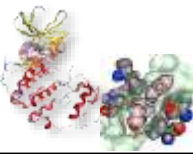
Iwabu Honma, et al,
ADIPONECTIN RECEPTOR-
ACTIVATING COMPOUND,
PCT/JP2014/076185



欧米メガファーマとの
共同研究・導出契約に結び付いた



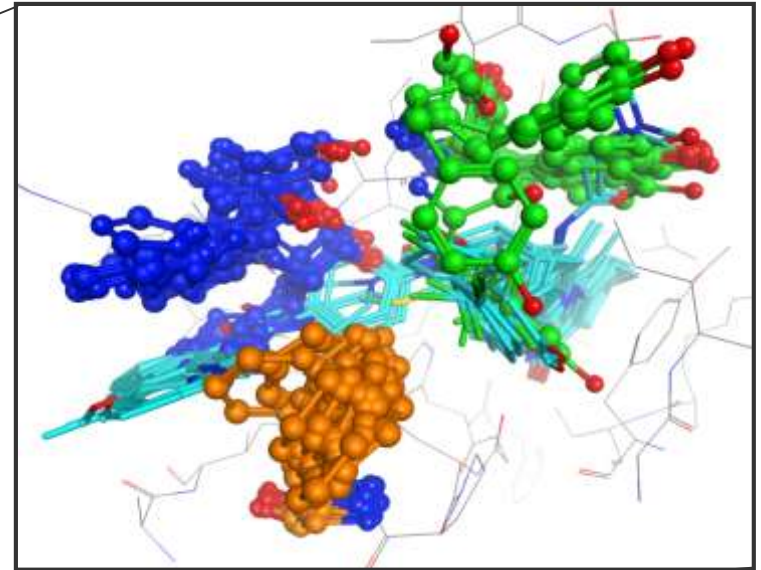
タンパク質の構造はフレキシブル



ターゲット

```
MHHHHHSSGVDLGTENLYPQSMNLNTSGSGTYLIDLSPDQKEFQSVVEEMQSTVPEHRDGGHAGGIFNRY  
NILKIQKVCNKLLMERYTHRRKEVSEENHNHANERMLFHGSPFVNAIIHRGFDEPHAYIGGMFGAGIYFA  
ENSSKSNQYVYIGGGTCCPVHKDRSCYICHRQLLFCRVTLGKSFLLQFSAMFMAHSPPGHHSVTGRPSVN  
GLALAEYVIYRGEQAYFEYLITYQIMRPEG
```

構造解析

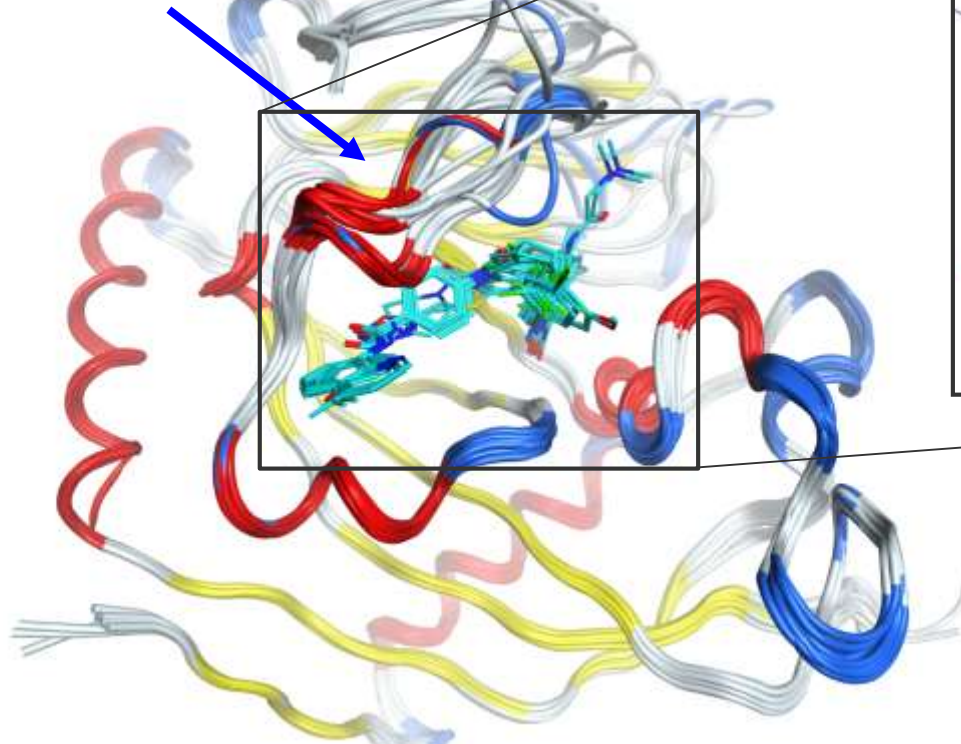


拡大すると、主鎖だけではなく、3個の残基の側鎖が大きく動いており、リガンドと衝突するものも多い。

このような場合に、
どうしたら良いで
しょうか？

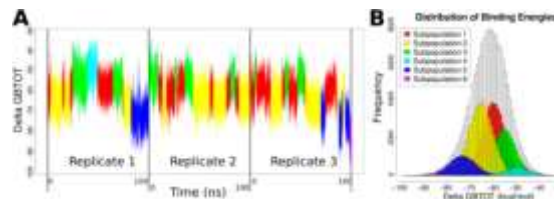
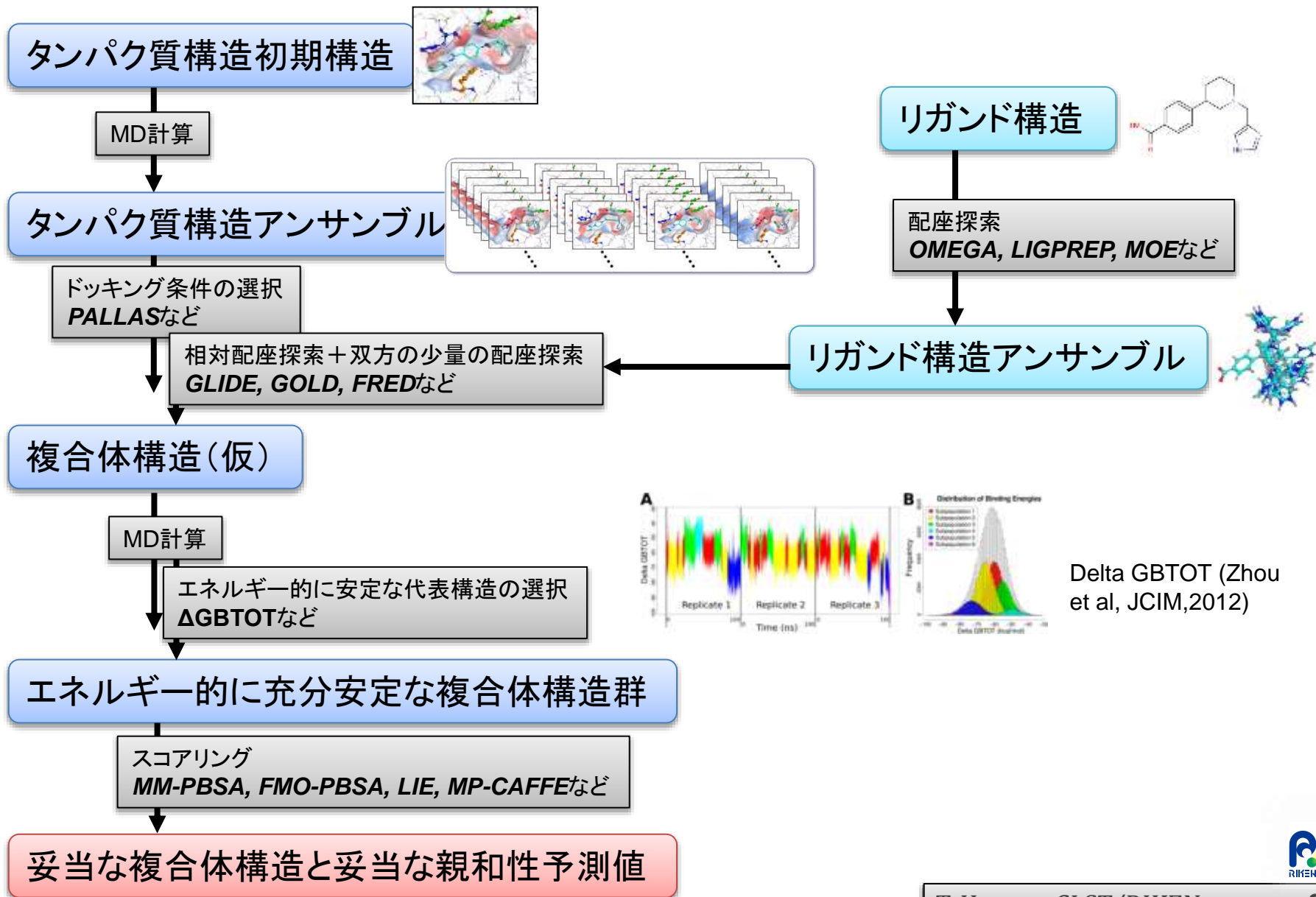
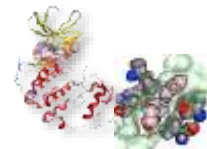


主鎖が大きく動いている



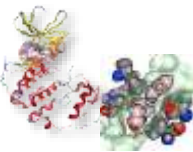
あるターゲットで26個の様々なリガンドが結合したX線構造

“妥当な”ドッキングに必要なこと



Delta GBTOT (Zhou et al, JCIIM, 2012)

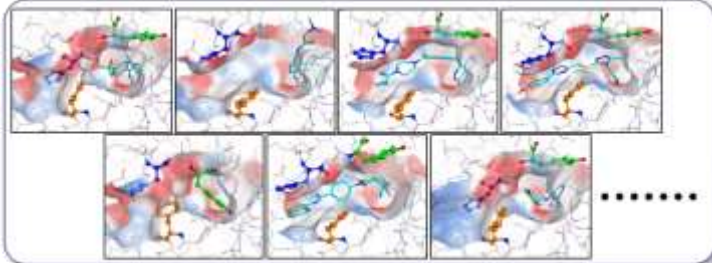
半自動ドッキング条件検討システム PALLAS



ターゲット

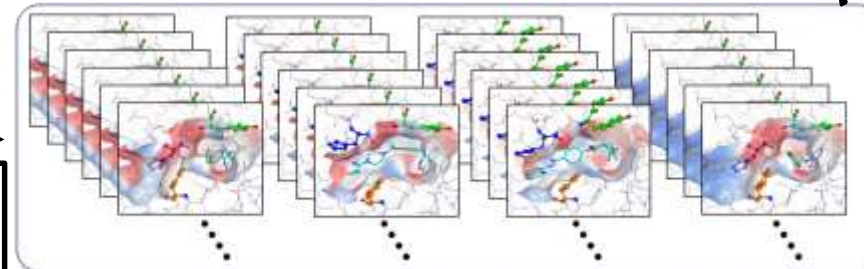
```
MHHHHHSSGVDLGTENLYPQSMNLNTSGSGYI LIDLSPDQKEPQSVVEEMQSTVPEHRDGGHAGGI FWRV  
NILKI QKVCNKLLMERYTHRRKEVSEENHNHANERMLFHGSPFVYNAI IHRGFDEPHAYIGGMFGAGIYFA  
ENSSKSNQYVYIGGGTCCPVHKDRSCYICHRQLLFRCVTLGKSFQFSAMFMAHSPDGHHSVTGRPSVN  
GLALAEYVYRGEQAYPEYLITTYQIMRPEG
```

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

RMSD values between 2-
most ligand poses and
docking poses are
calculated for all defined
ligands. Reproducibility =
of ligands with RMSD <= 2.0 Å
of all ligands

Enrichment Factor (EF)

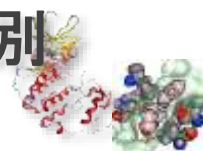
EF shows improvement of the hit rate by docking
compared with that of random selection.
% Enrichment Factor (EF)
= # of active compounds in top 1% score
of active compounds in 1% random selection

ROC score

Using ROC (Receiver Operating Characteristic) plot,
ROC-AUC (Area Under the Curve) is calculated.
ROC score (ROC-AUC)
= Area under the ROC curve
Maximum area of ROC plot (ideal selector)

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)

MUSES



X-ray Structures and Docking Results

good score and positive ex) -10.5 kcal/mol	good score, but negative Ex) -10.3 kcal/mol	bad score, but positive ex) -7.2 kcal/mol	bad score and negative >-7.0 kcal/mol

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

--	--

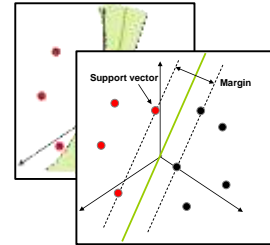
Training Set

Machine Learning by

- Support Vector Machine
- Random Forest

MUSES Models

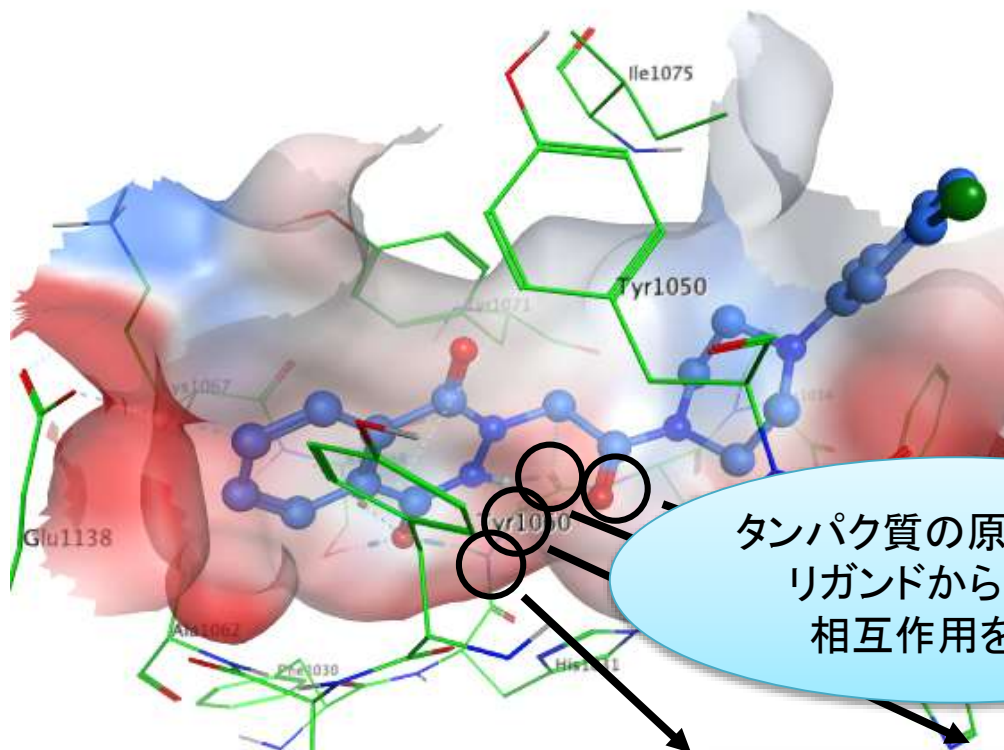
Validate Prediction Performance



--	--

Test Set

タンパク質-リガンド相互作用エネルギー記述子 *aPLIED*

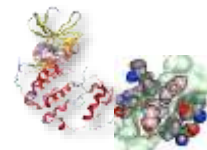


すべての結合サイトタンパク質原子に対して、その原子とリガンドとの間のMM-GBVIエネルギーを計算する

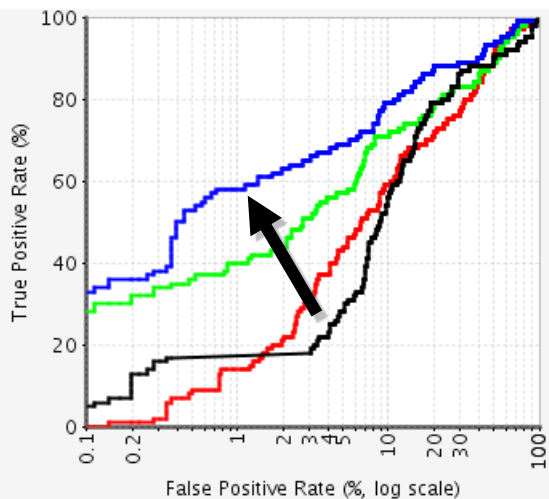
タンパク質の原子単位でリガンドから受ける相互作用を計算

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

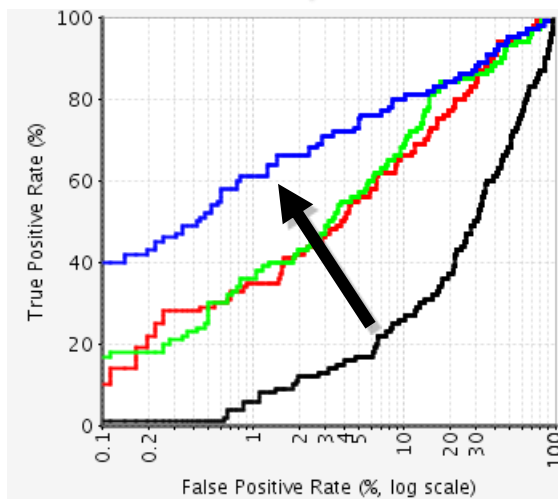
100個の活性化化合物を使った Support Vector Machine Model の ROC curve



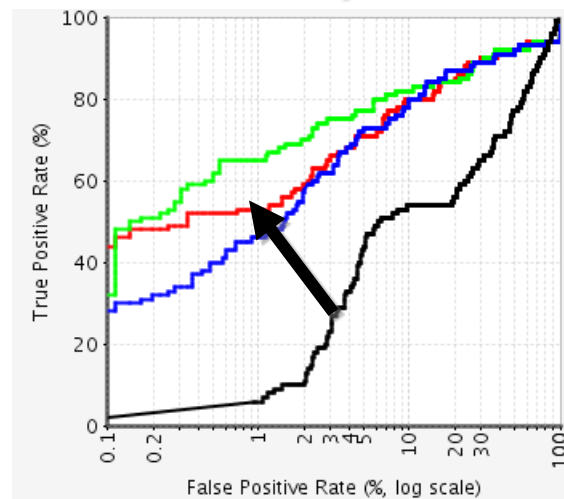
Cdk2



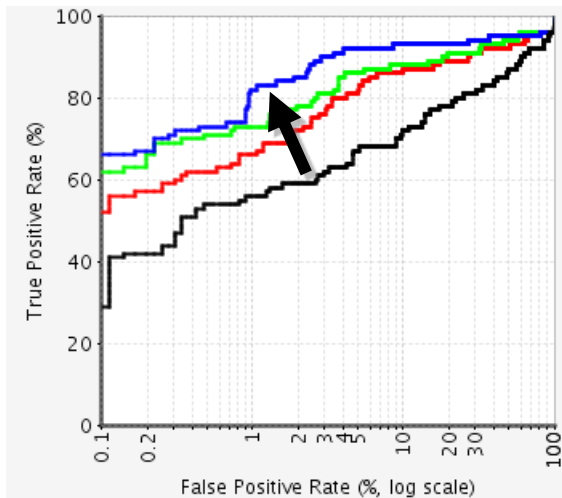
Cathepsin K



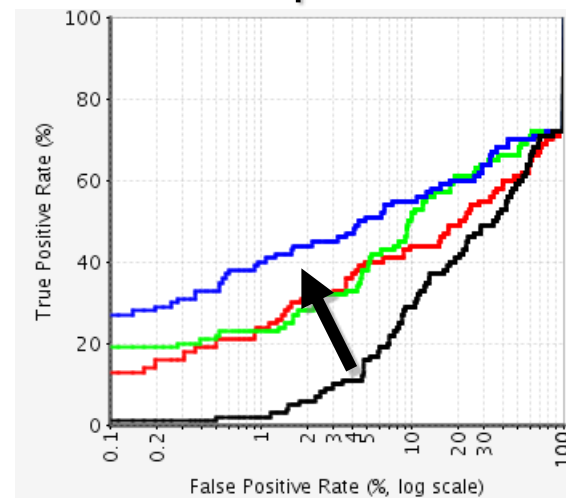
Carbonic Anhydrase 2



Thrombin

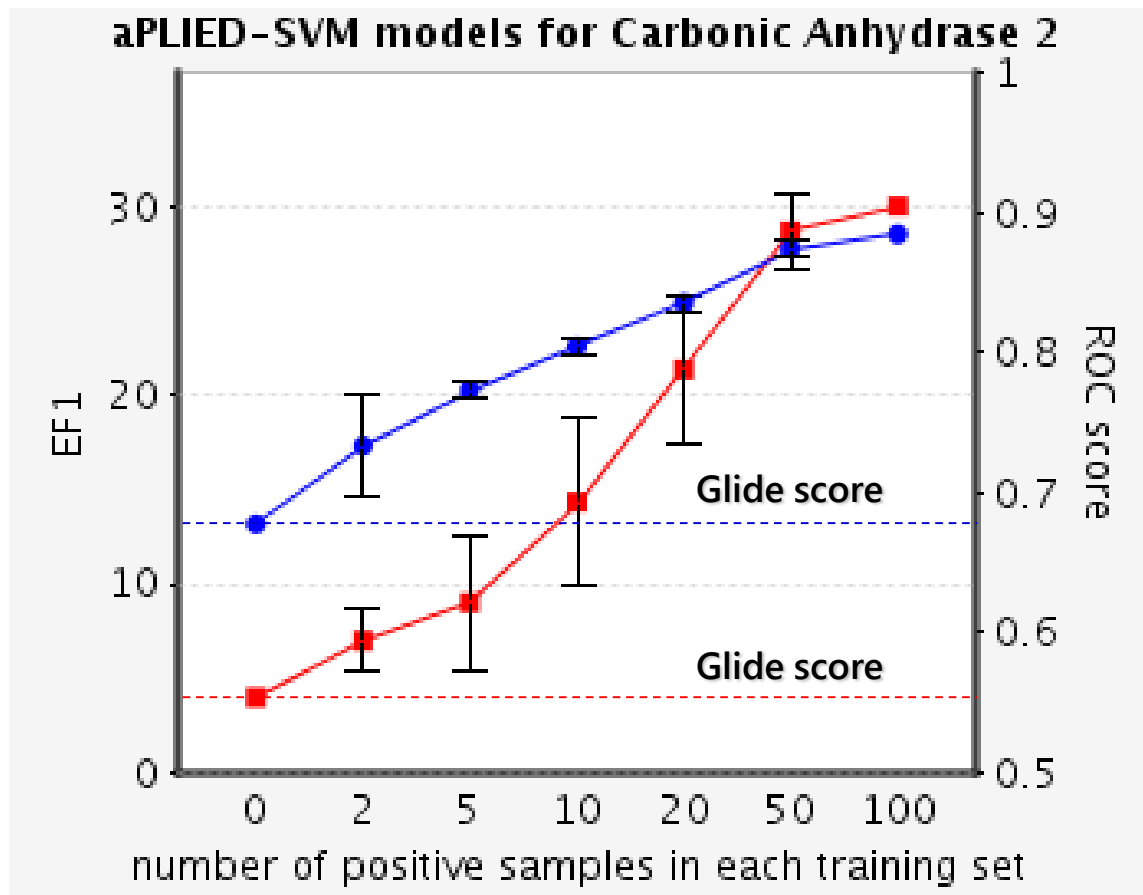
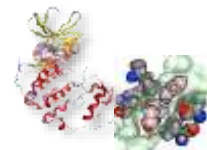


HIV-1 protease



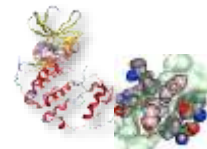
すべてのターゲットで **MUSES** モデルが **Glide score** を上回る結果を示した。

学習セットの活性化化合物の数と学習効果



■ EF1 Mean
■ ROC score Mean

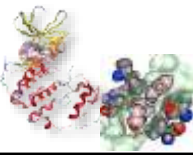
学習セットの既知活性化化合物数と
スクリーニング効率
(ドッキングポーズは最もスコアの高いもののみ利用)



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



Human genes (~25,000)

Druggable genes
(~3,000)

Disease genes with
marketed drugs (215)

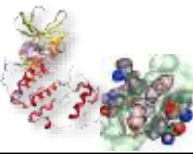
215

Disease genes
(~3,000)

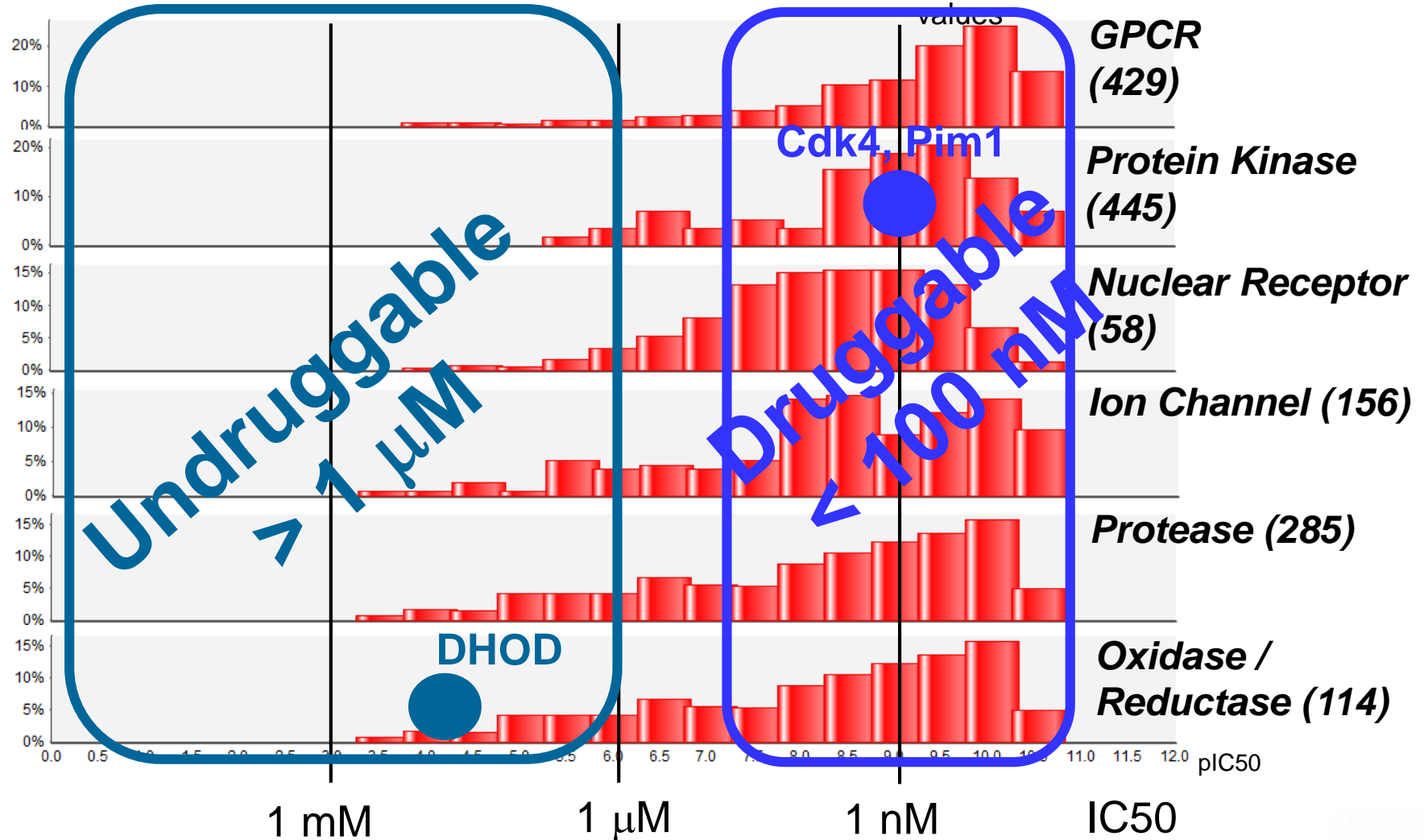
Disease genes with
Druggable domains
(~500)

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。

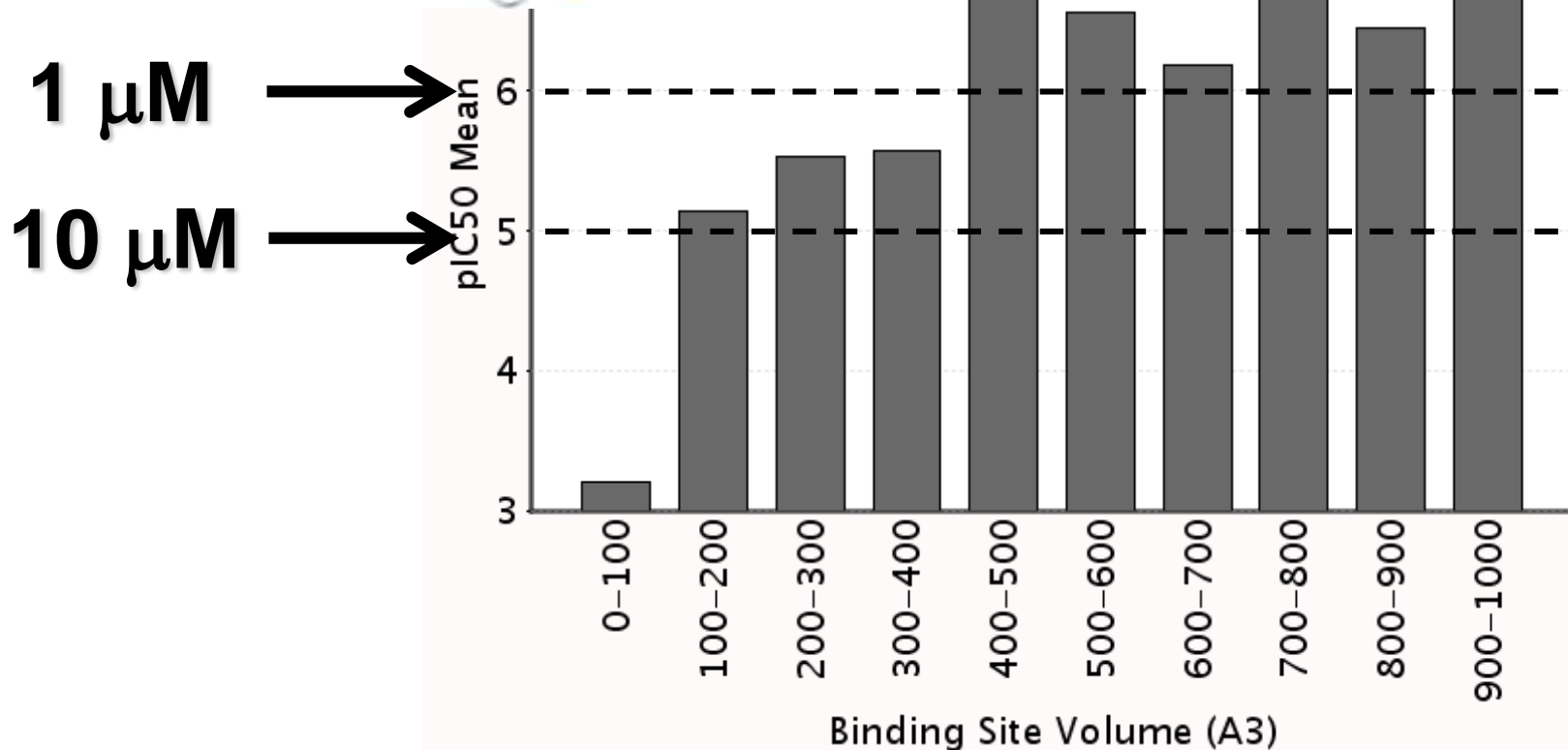
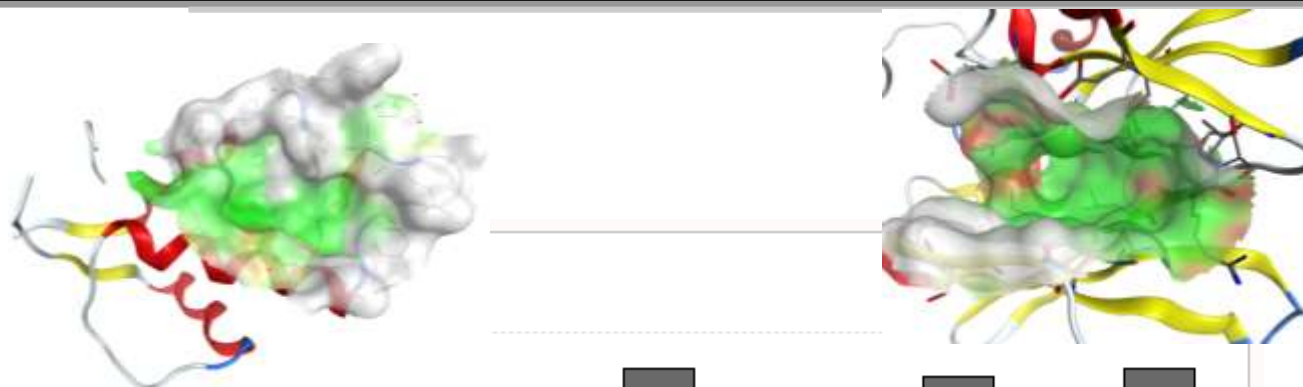
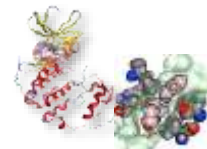
各ターゲットの最高到達活性の分布



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



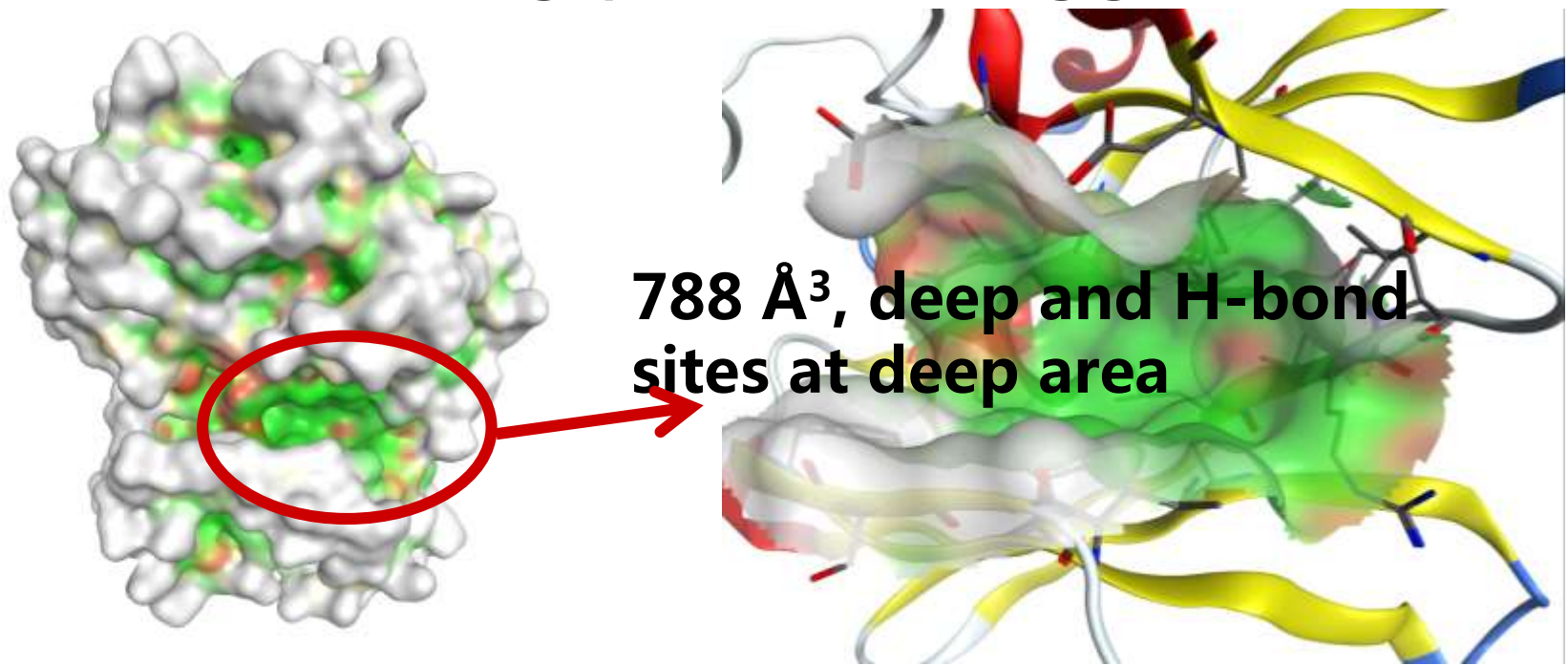
結合サイトの大きさと活性の強さ



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

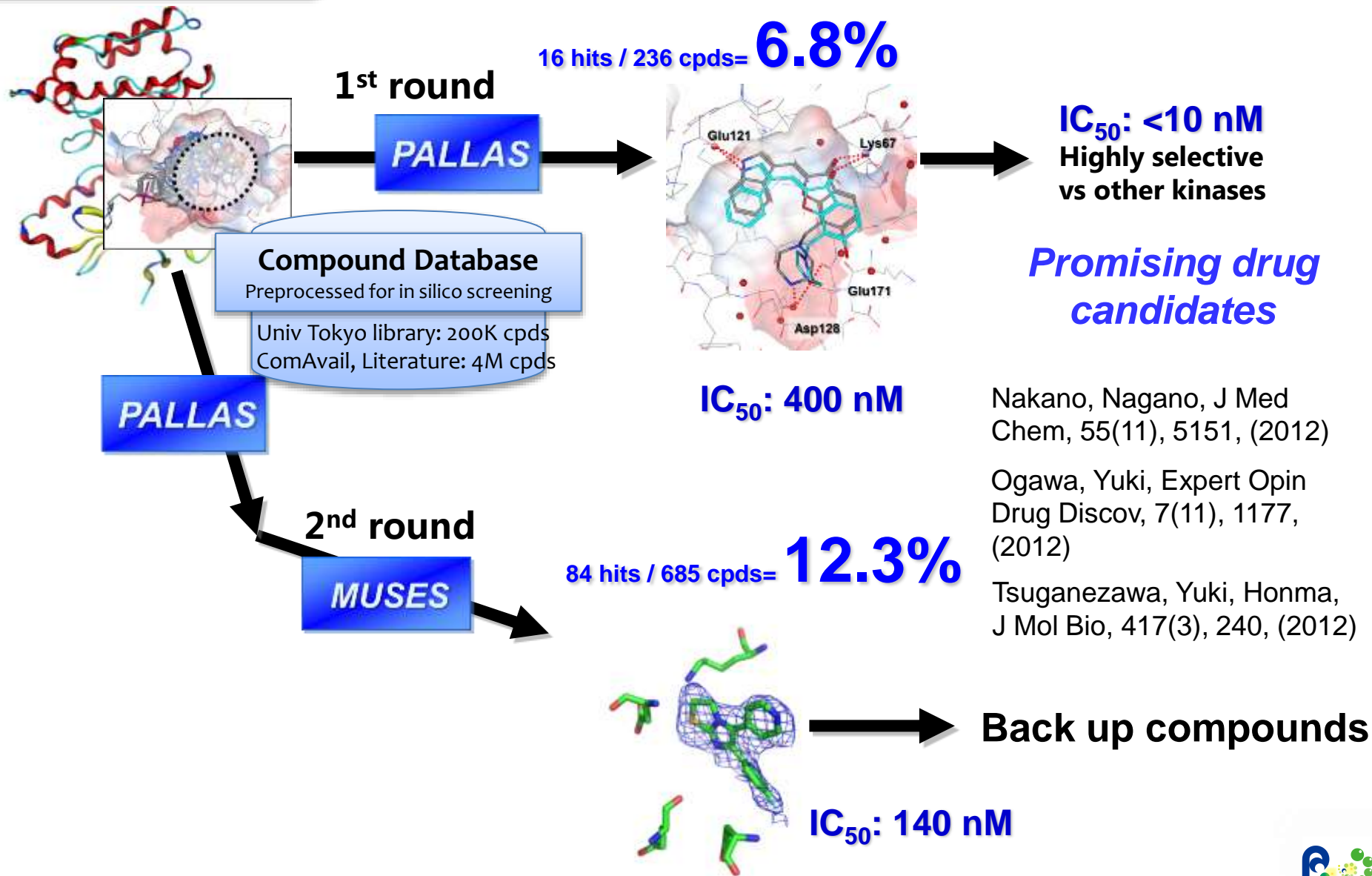
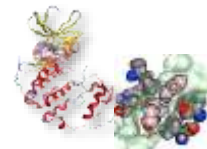
Category 1

Is the following protein druggable ?



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

Pim1 Inhibitors



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

Ogawa, Yuki, Expert Opin Drug Discov, 7(11), 1177, (2012)

Tsuganezawa, Yuki, Honma, J Mol Bio, 417(3), 240, (2012)

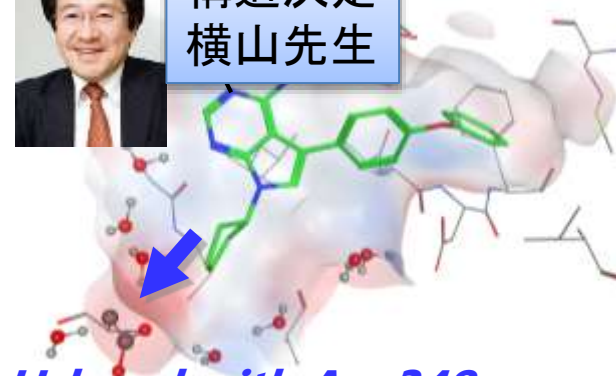
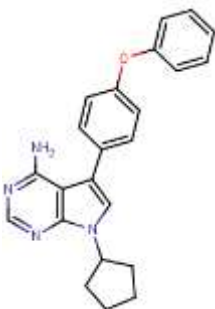
白血病治療薬HCK阻害剤の発見



理研・石川
先生との
共同研究



構造決定
横山先生



H-bond with Asp348

RK-9024466
IC₅₀: 7.7 nM

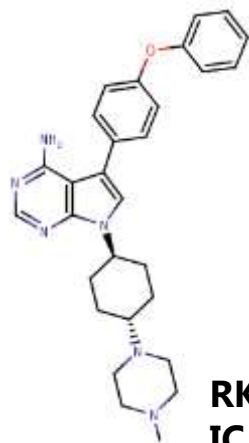
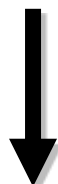
Other 11 hits (IC₅₀ < ca 2000 nM)



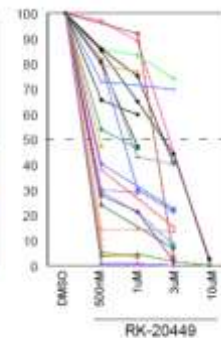
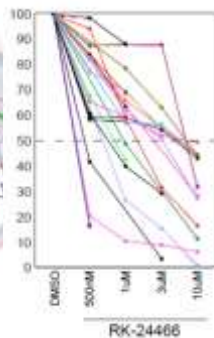
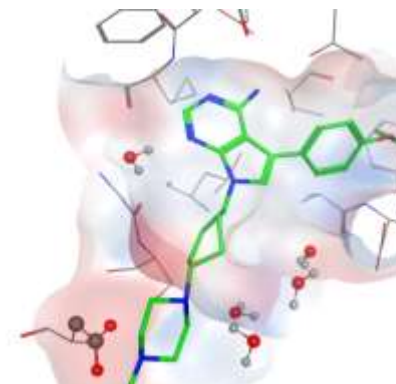
SHARE RESEARCH ARTICLE ACUTE MYELOID LEUKEMIA
A Pyrrolo-Pyrimidine Derivative Targets Human Primary AML Stem Cells in Vivo
Yuriko Saito¹, Hitomi Yuki², Mitsuo Kuratani², Yoshinobu Hashizume¹, Shinsuke Takagi^{1,2}, Teruki Honma¹, Akiko Tanaka¹, Mikako Shirouzu¹, Junko Mikuni¹, Noriko Handa¹, Ituko Ogihara¹, Akiko Sone¹

Saito, Yuki, Honma, Ishikawa, et al,
Sci Trans Med, 5:181ra52, 2013.

現在、米国にベンチャーを
設立し、医師主導臨床研究
に向けて準備中



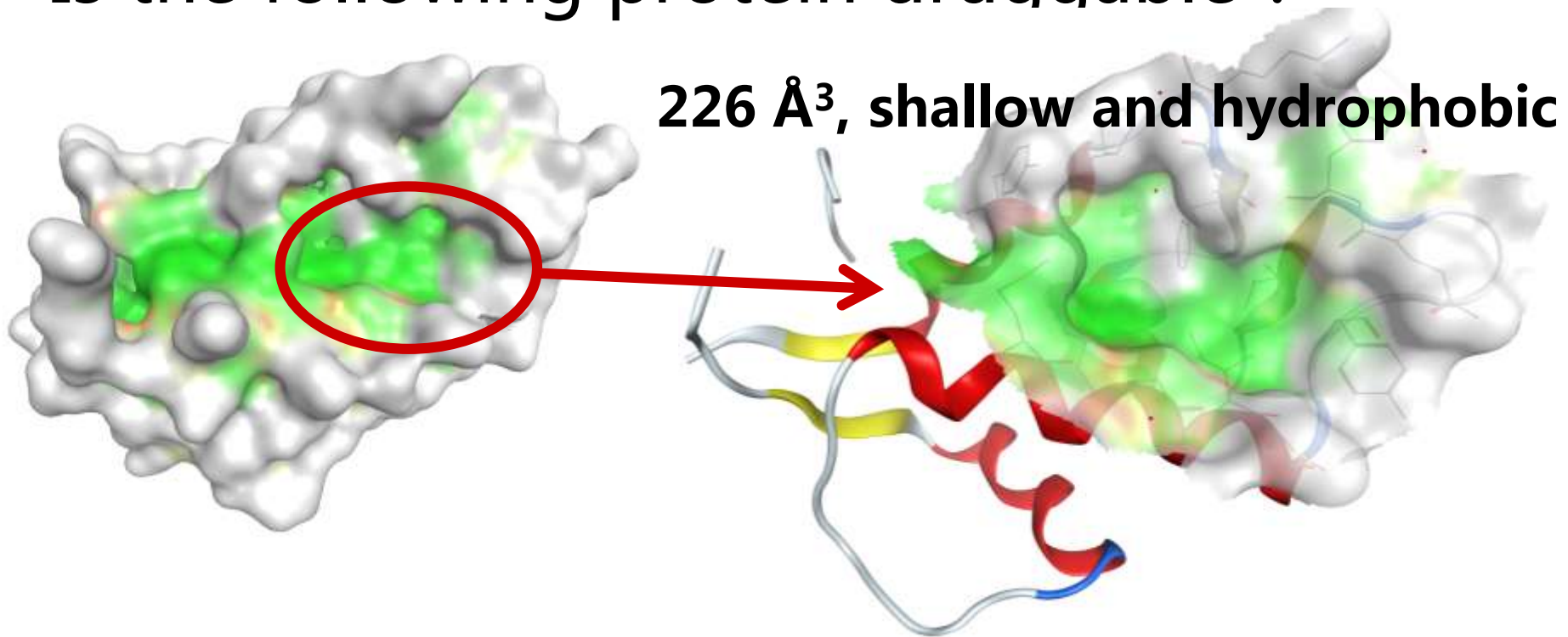
RK-0020449 **Clinical candidate**
IC₅₀: 0.43 nM



Saito, Ishikawa, et al,
AGENT FOR TREATING OR CONTROLLING
RECURRENCE OF ACUTE MYELOGENOUS
LEUKEMIA, PCT/JP2013/070436

Category 2

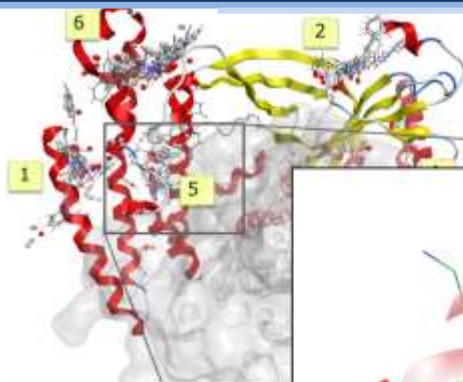
Is the following protein druggable ?



免疫疾患治療薬DOCK1/2阻害剤の発見

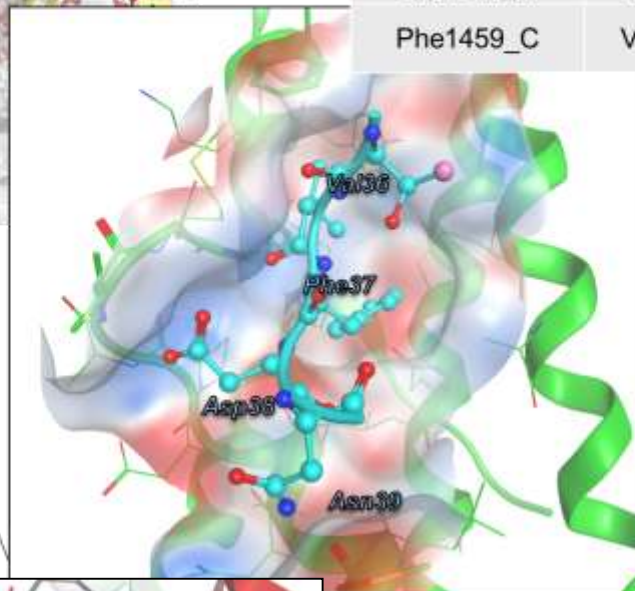


九大・福井
先生との
共同研究



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

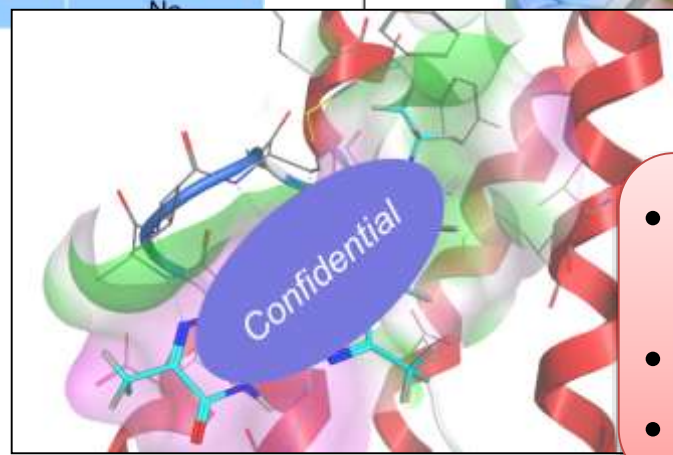
DOCK2側の結合サイト候補の分析



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

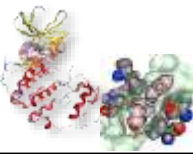
4 hits / 584 cpds = **0.7%**

IC₅₀: 5 μM



- 動物モデル薬効を示し、論文準備中
- 特許出願済み (PCT移行中)
- AMED-LEAPに採択

主なインシリコスクリーニングの結果



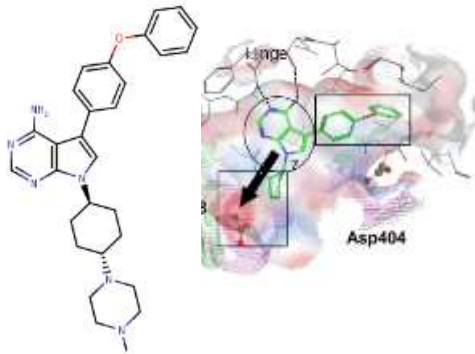
前臨床段階



理研・石川先生**
HCK
急性骨髄性白血病



A Pyrrolo-Pyrimidine Derivative Targets Human Primary AML Stem Cells in Vivo

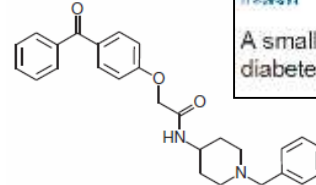


現在、米国で
臨床試験準備中

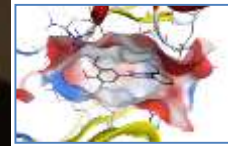
知財取得



東大・門脇先生*, **
AdipoR
糖尿病



東大・宮園先生**
ALK2
FOP



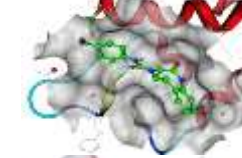
九大・福井先生*, **
DOCK1/2
がん、自己免疫疾患



東大・長野先生*
Pim1
急性骨髄性白血病



理研・田仲先生*
MdmX
がん



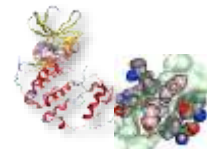
がん研・清宮先生
Tankyrase
がん



創薬シード ツール

RAMP2**
Tob1**
Tau**
DHOD*, **
GK*, **
TAO*, **
c-Cbl-pTyr*
Cathepsin K*
p70S6K*
Atg4B*
ATC*
AMPK*
ACC2*
CLK1
H-PGDS
G9a
MTH1
Ras
Ras-Raf
LSD1
Sult2B

Outline

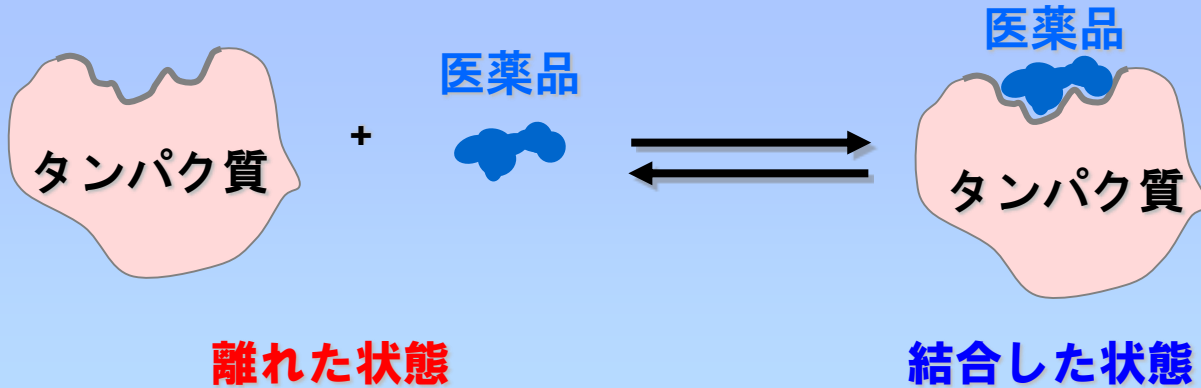
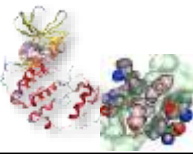


1. インシリコスクリーニングの技術開発

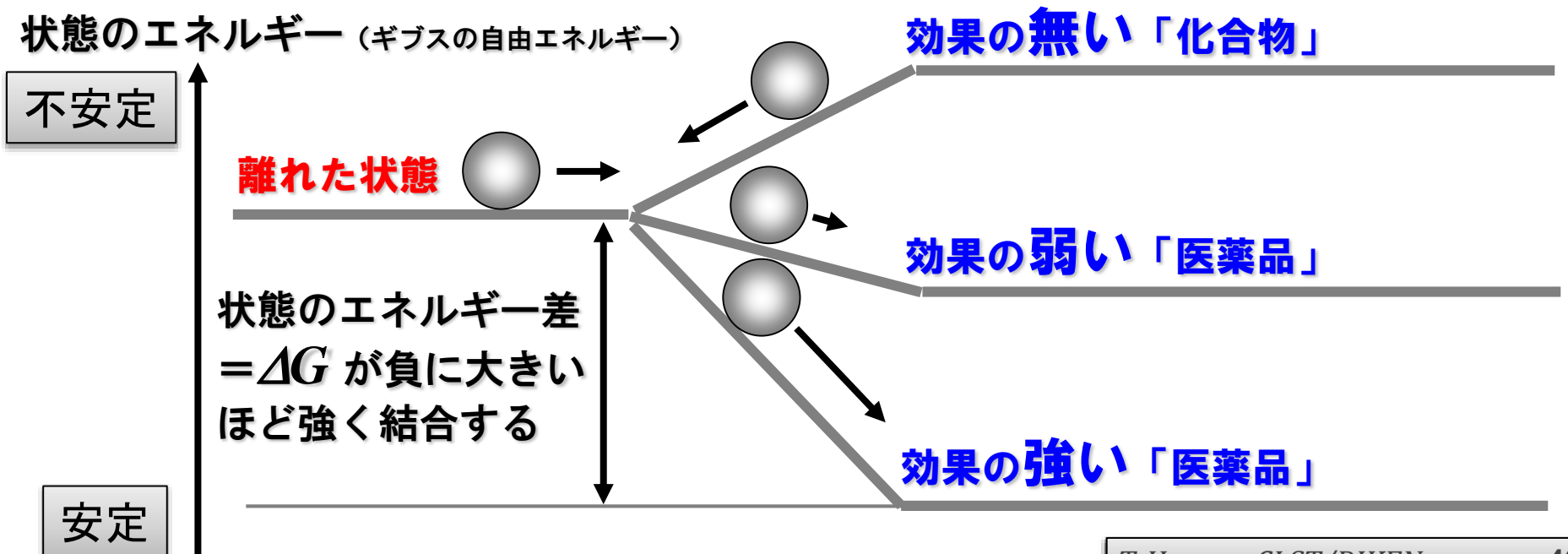
2. インシリコスクリーニング・設計の実例

3. 結合親和性予測／FMOの創薬活用

医薬品の結合・解離と状態のエネルギー



※タンパク質の周りは主に水に囲まれている

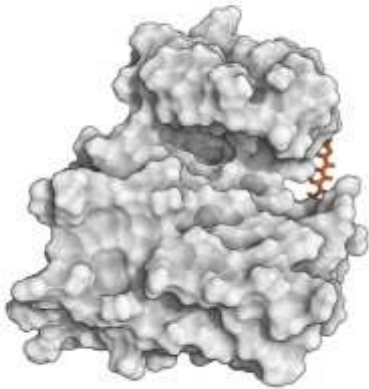


エンタルピー項とエントロピー項

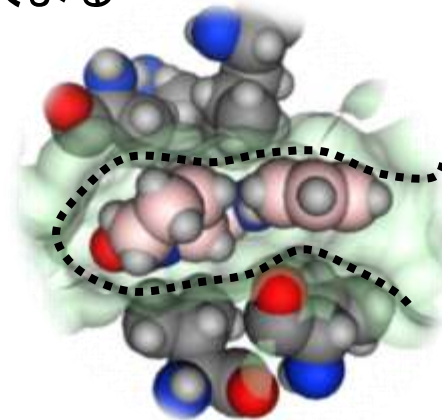


$$\Delta G = \Delta H - T\Delta S$$

状態のエネルギー差 ΔG はエンタルピー差 ΔH とエントロピー差 ΔS の兼ね合いによって決まる



エントロピー項（自由度を好む）



エンタルピー項（結合を好む）

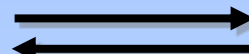
自由に動ける方が安定

自由に動ける方が安定

タンパク質

くすり

+



くすり

結合すると安定する
...

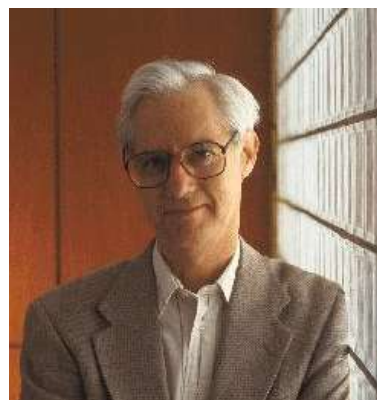
タンパク質

離れた状態

結合した状態

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



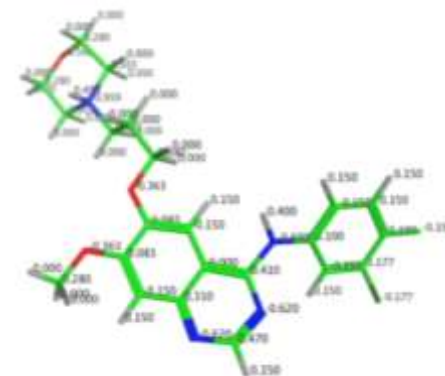
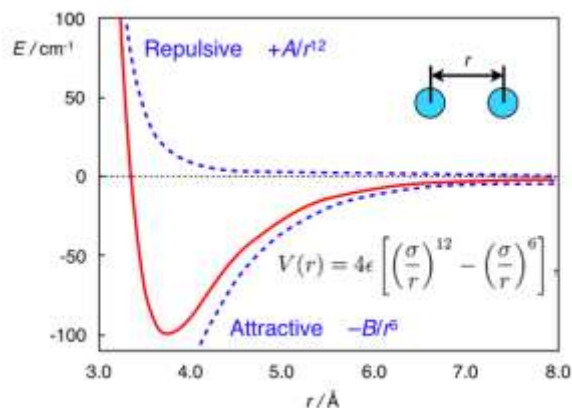
Kuntz, JMB, 269 (1982)

Van der Waals force
(Lenard-Jones approximation)

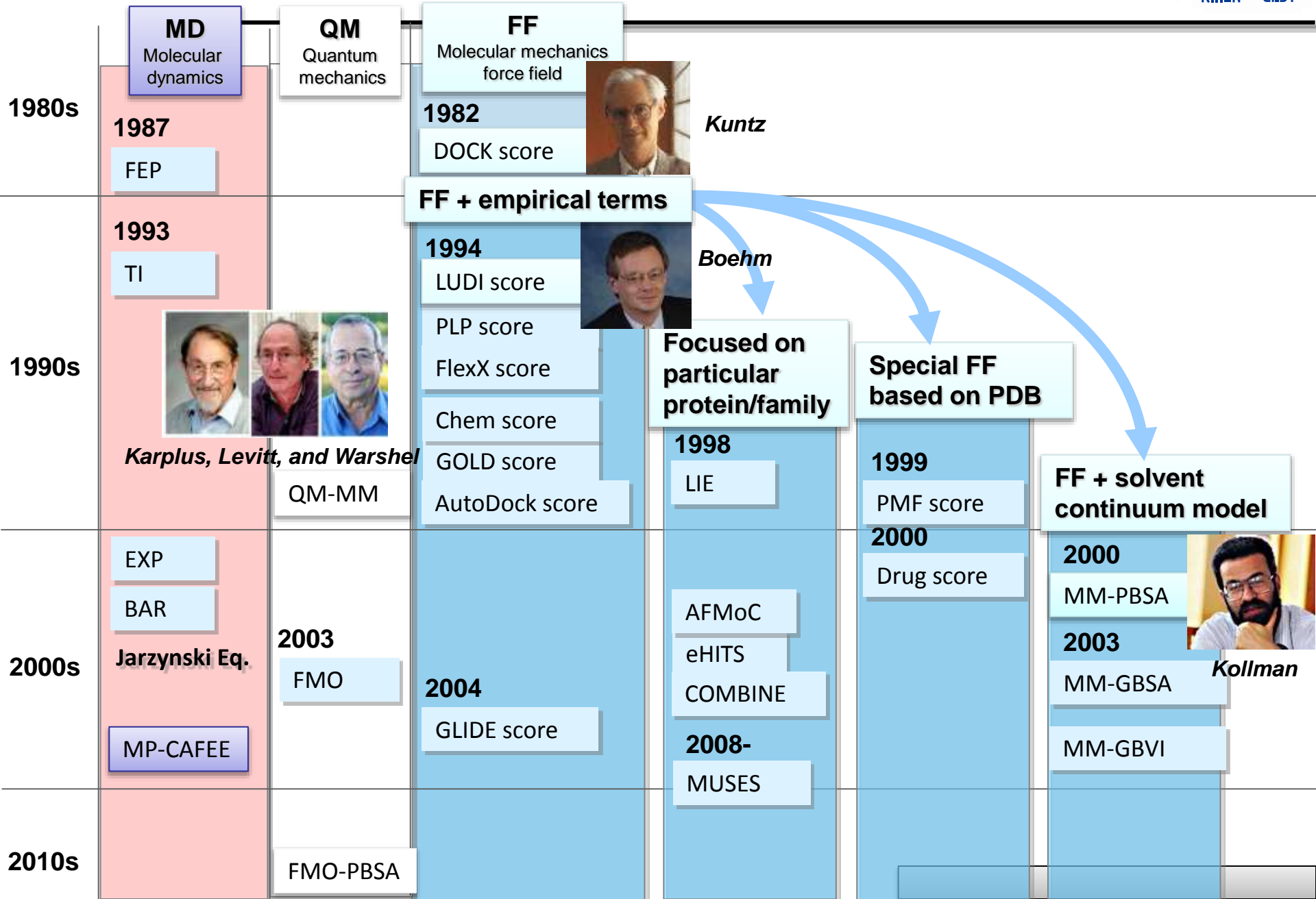
Electrostatic interaction
(Coulomb's equation)

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

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



MMベースの経験的結合親和性予測の限界

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

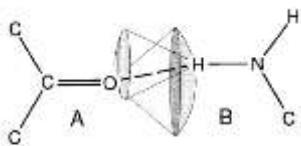
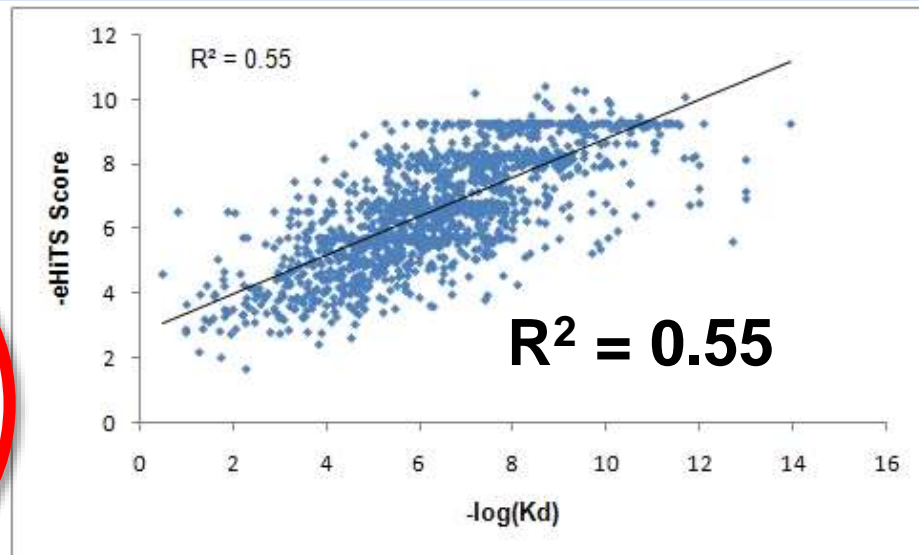


Figure 2. Conditions for the formation of intermolecular hydrogen bond between the carbonyl oxygen and the nitrogen. The interaction centers are the oxygen and the hydrogen atom forming the hydrogen bond. They have to fall mutually on the surrounding interaction surfaces.

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

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

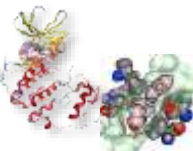
R² = 0.2-0.5程度

平均でIC₅₀値にして10倍から100倍の誤差が

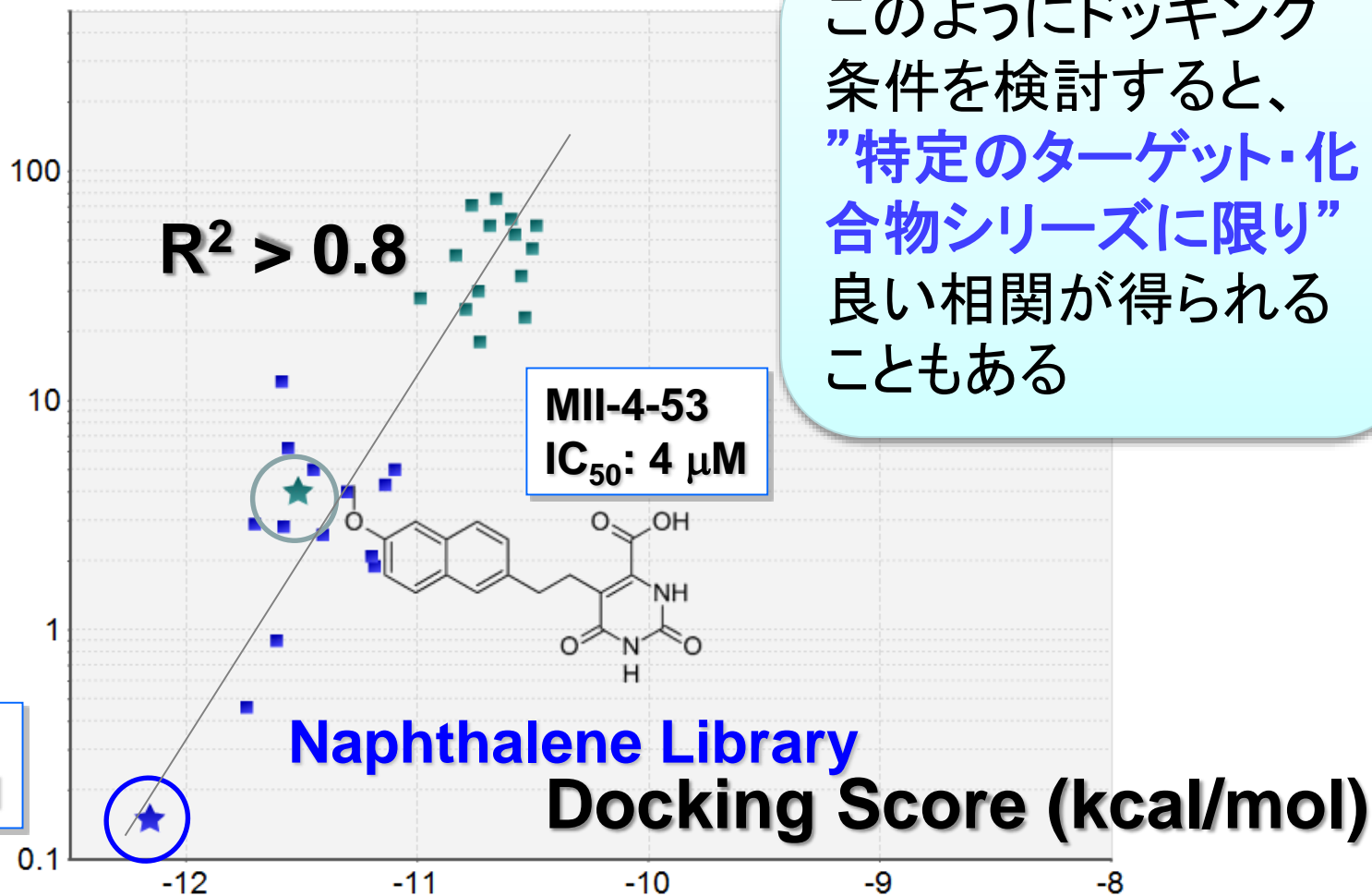
[eHiTS 2009 Binding Affinity technical note.:](http://www.simbiosys.ca/ehits/eHiTS_2009_binding.pdf)

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

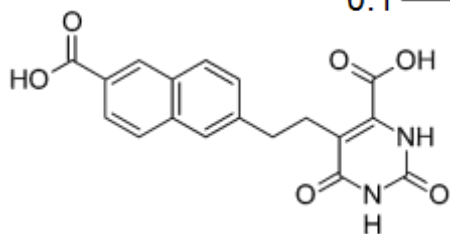
仮想化合物ライブラリーの構築による設計



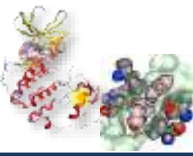
IC₅₀ (μM)



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

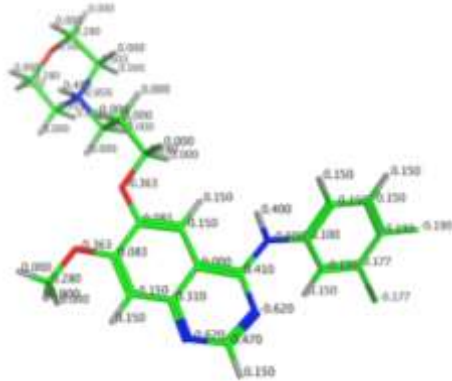


MM(FF) とQMの違い

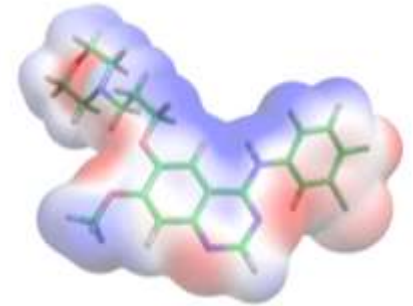


Charge

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。

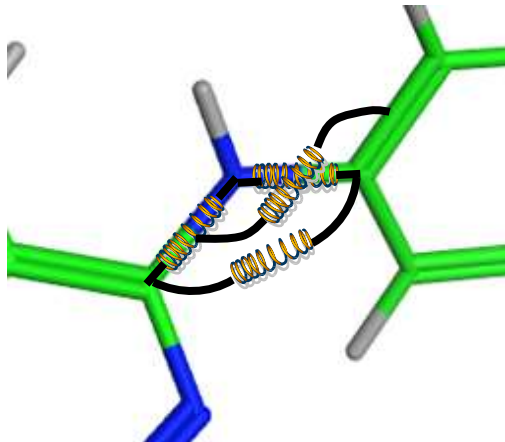


QM(quantum mechanics)では、Schrodinger方程式で記述される電子密度に従って、電荷の粗密が連続的に表現される。



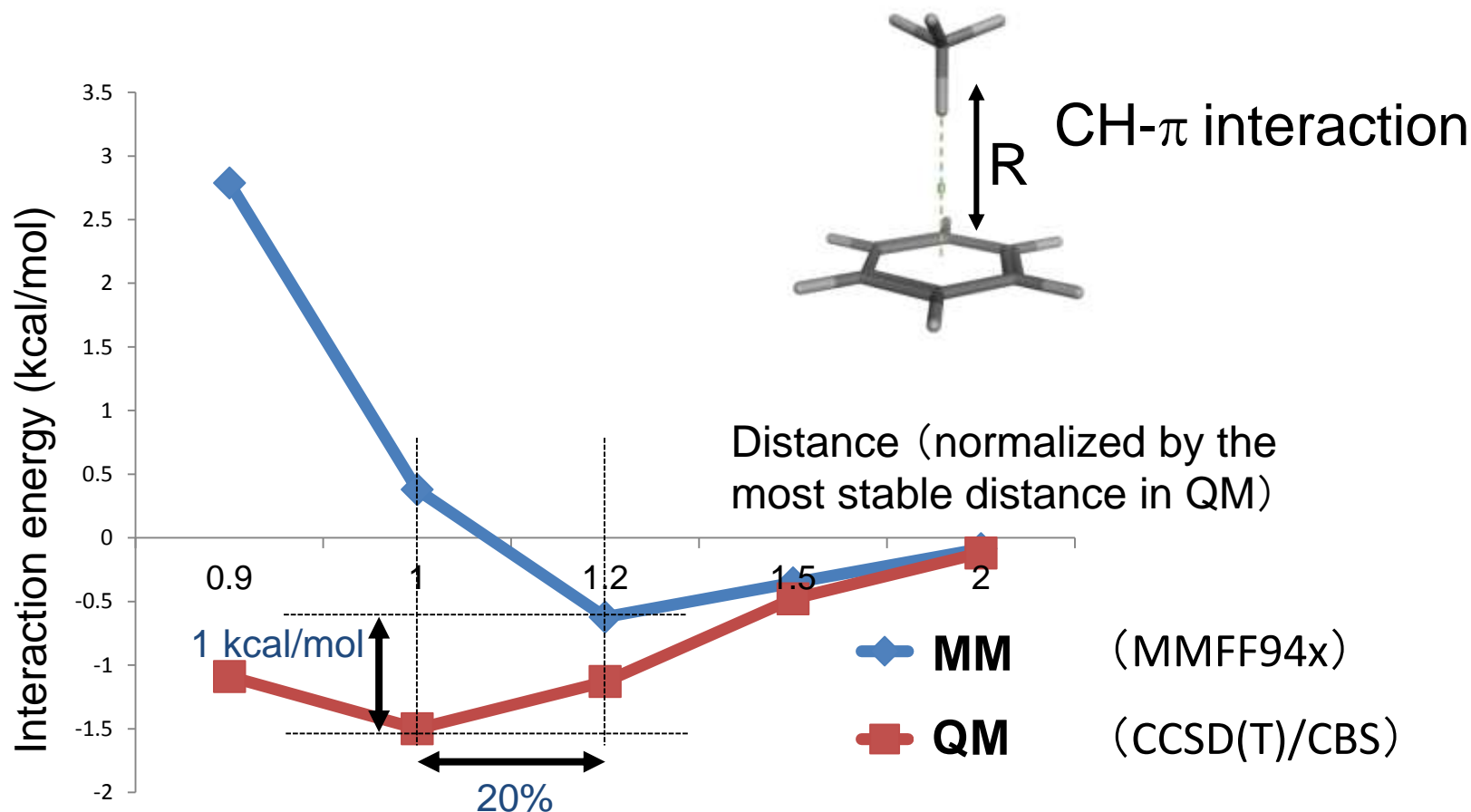
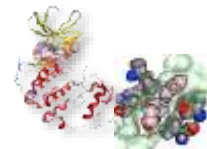
Strain energy

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。

CH- π 相互作用のMMとQMの違い



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

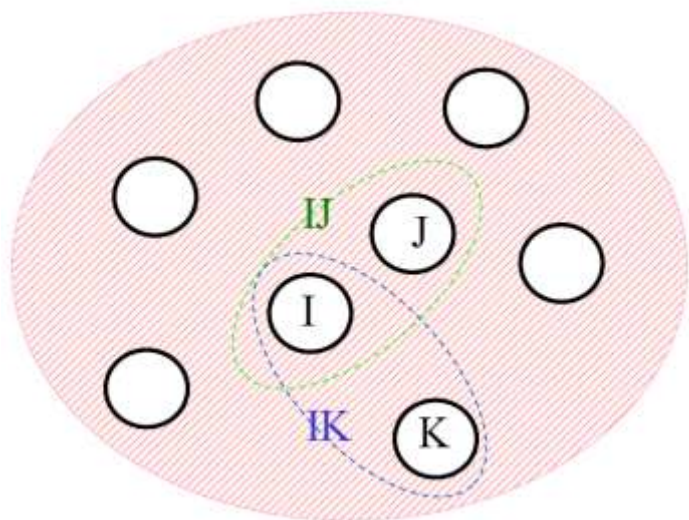
Hobza, Accounts of Chemical Research, 663, 2012
<http://www.begdb.com>

タンパク質に対するQM計算法 FMO



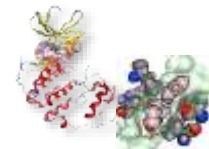
1999年、北浦らは、全く新しい、大きな系に対する効率的な量子力学計算法であるFMO(fragment molecular orbital)法を提案した。

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。



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

FMO創薬コンソーシアム(2014年設立)



「京」利用課題: HPCIを活用したFMO創薬プラットフォームの構築

FMO創薬コンソーシアム

産学官連携で実用的なFMO創薬技術を開発する



大学・研究所 **9機関**
製薬企業 **16社**
IT企業 **2社**

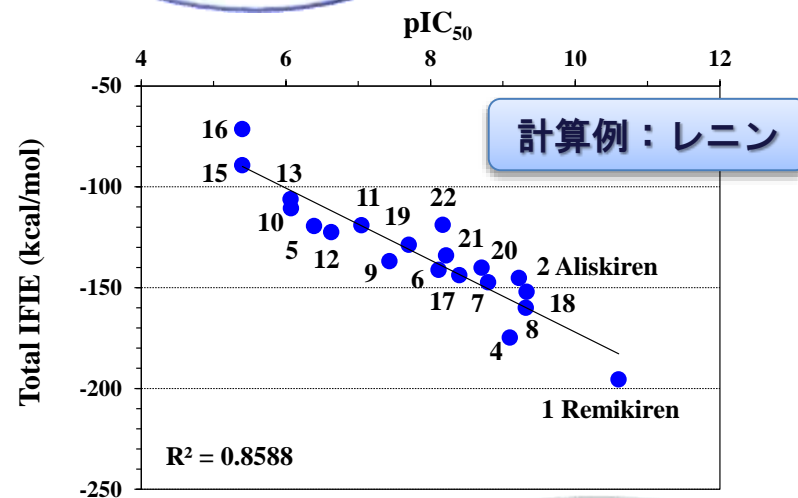


代表: 福澤 薫

K computer



	キナーゼWG	200構造	開発WG
	プロテアーゼWG	200構造	
	核内受容体WG	300構造	先行事例WG
	タンパー-タンパク相互作用WG	200構造	
			100構造



FMO計算を実施した複合体構造数

平成27年度
500
複合体

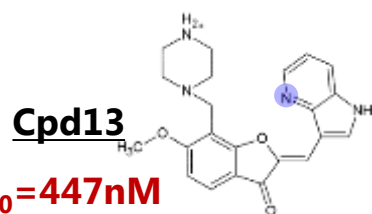
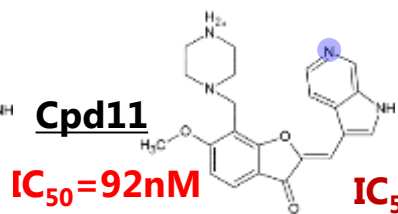
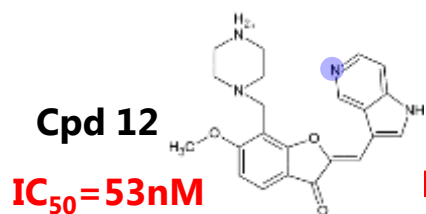
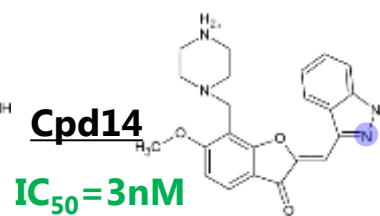
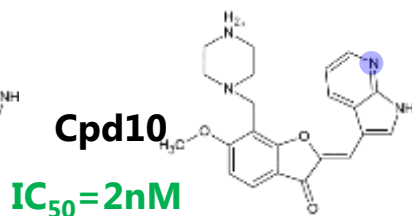
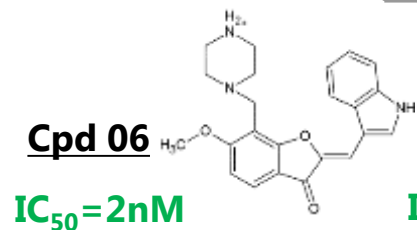
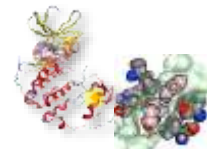
平成28年度
1,000
複合体

平成29年度
2,000
複合体(予定)

IFIE-DB
(IFIE-database)

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

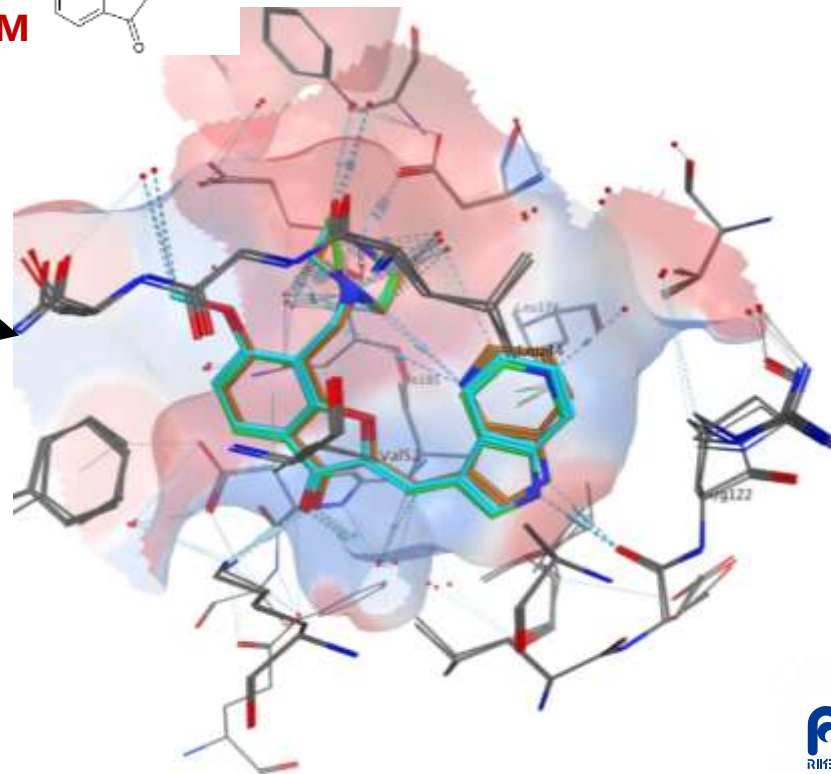
Activity Cliffの例 Pim1阻害剤



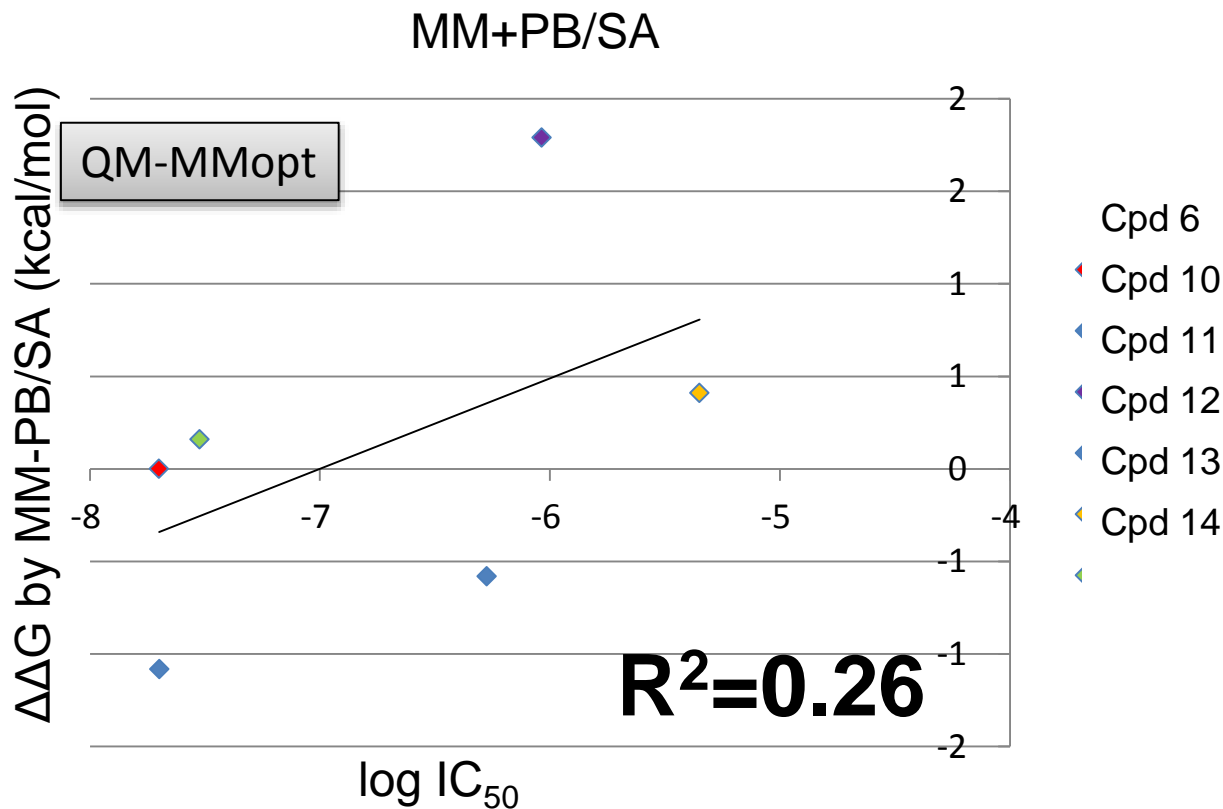
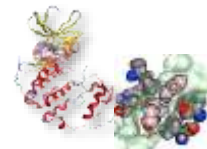
J. Med. Chem. **55** (2012) 5151-5164.

一原子異なるだけで、 IC_{50} 値が、2-200倍異なる！

上記のうち、4化合物（下線）のX線構造は、ほとんど同じ！

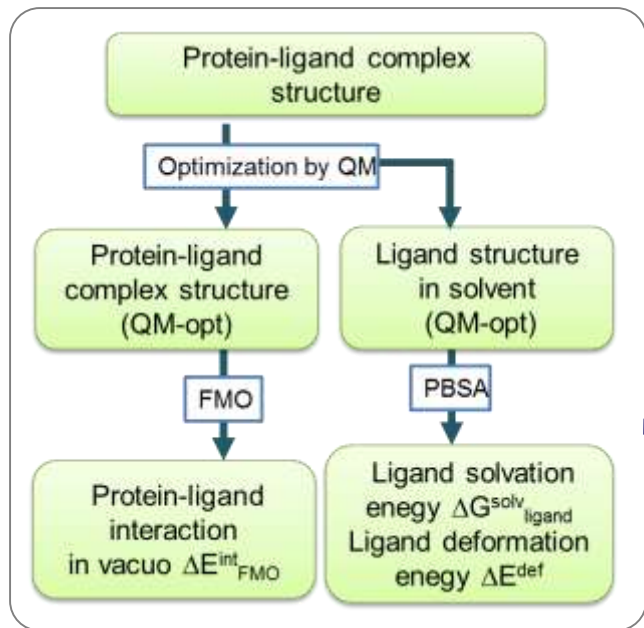
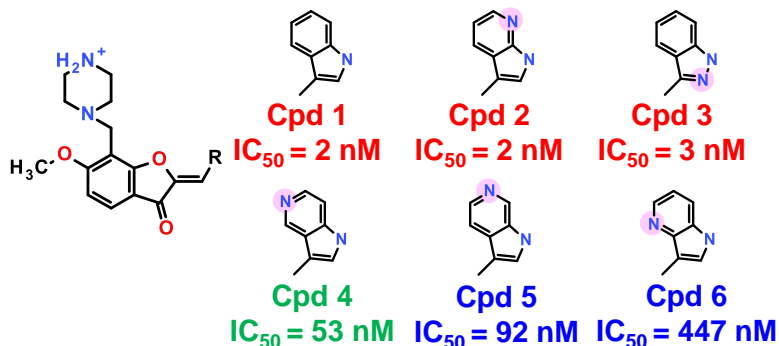
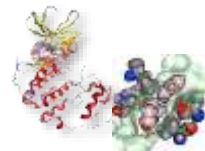


MM-PBSAで予測できる？

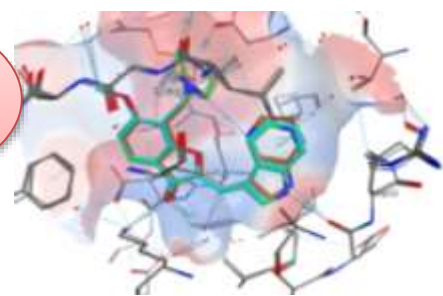


MM-PBSAでは、若干の傾向はあるものの、相関は非常に悪い
ちなみに、計算に使う構造としてMMoptとQM-Mmoptを使った場
合は、 $R^2=0.26$ よりもさらに悪い

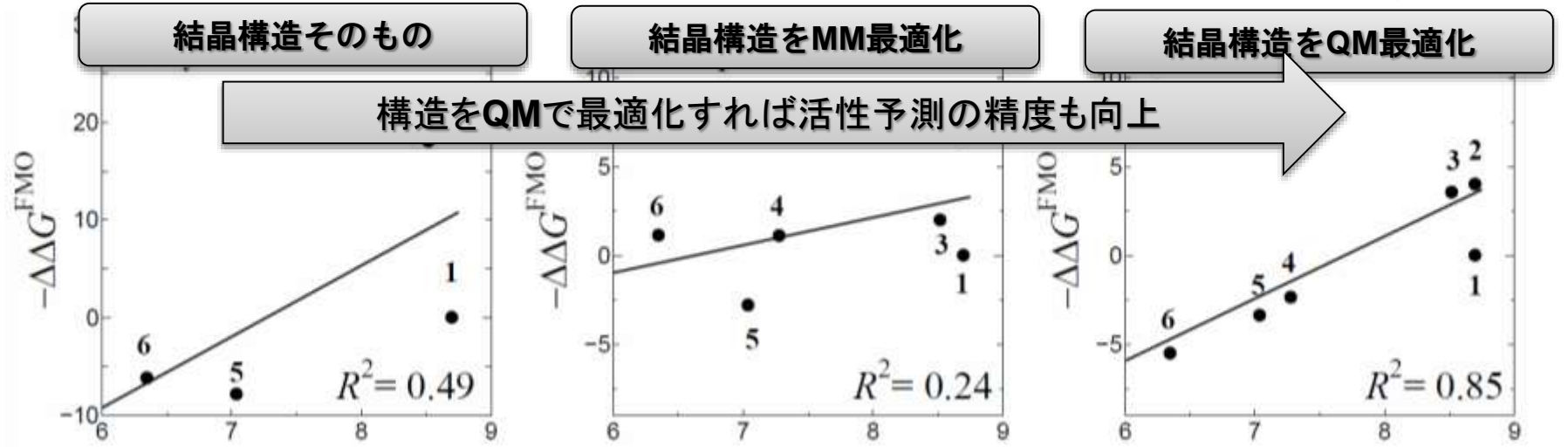
FMO+MM-PBSAによる 高難度の活性予測と結晶構造のクオリティー



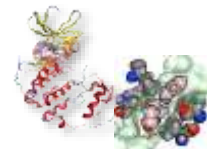
原子1個の違いで活性が200倍!



従来のMMだけでは $R^2=0.26$ だったが、
 $R^2=0.85$ まで劇的に向上

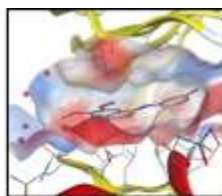


FMOの自動計算プロトコール



Automated search
of activity values

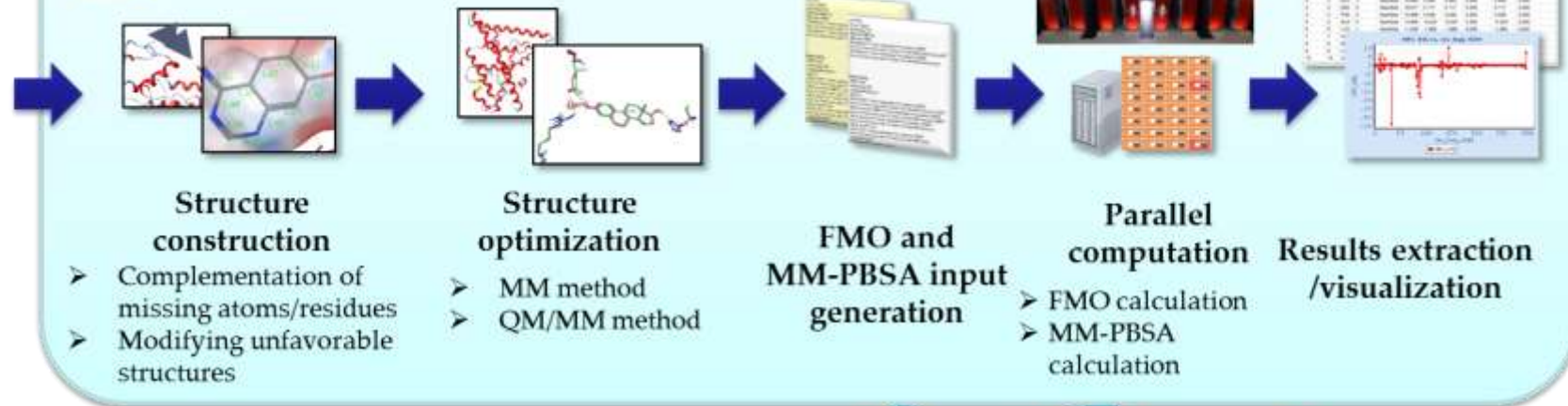
- IC50, Ki, Kd
- PDB with activity data



X-ray structures
from PDB
or docked ones



Automation by Pipeline pilot



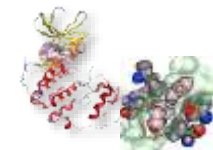
Automated analysis
with Excel macro etc.



**Construction of
IFIE-database**

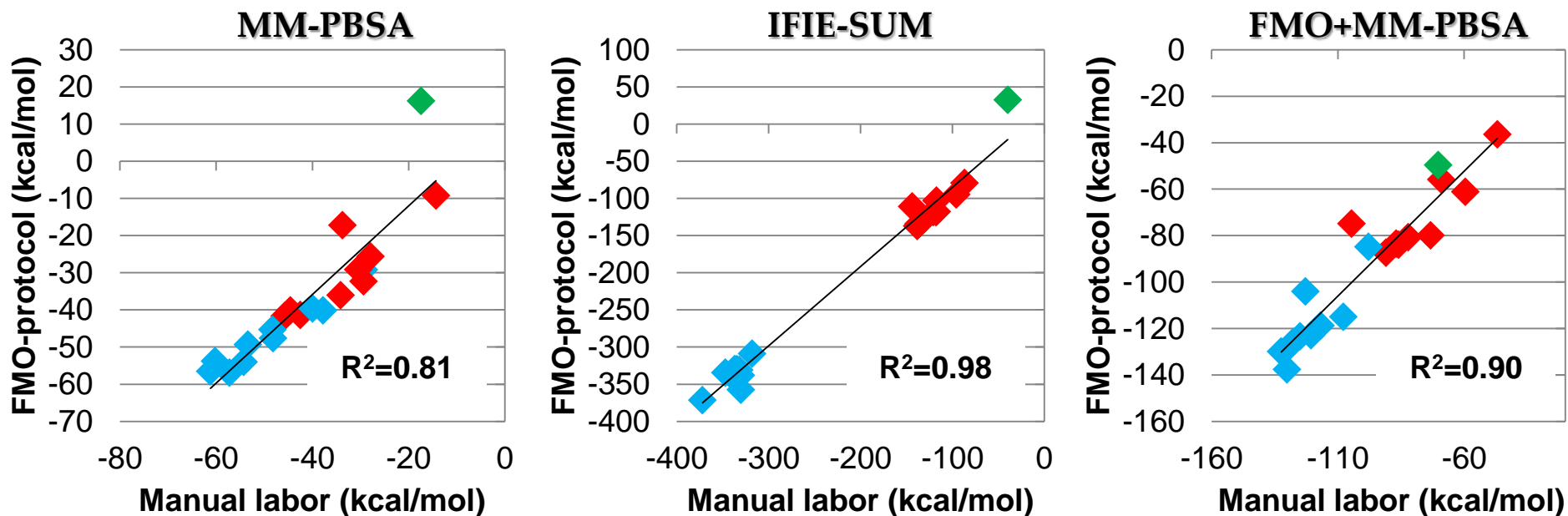
- IFIE/PIEDA
- VISCANA
- NPA charge
- Used structures
- Affinity data

自動計算プロトコルと手動の比較



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

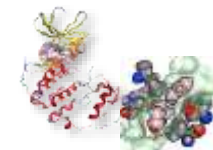
Coefficient of determination between FMO-protocol and manual labor data [Anzaki's data]



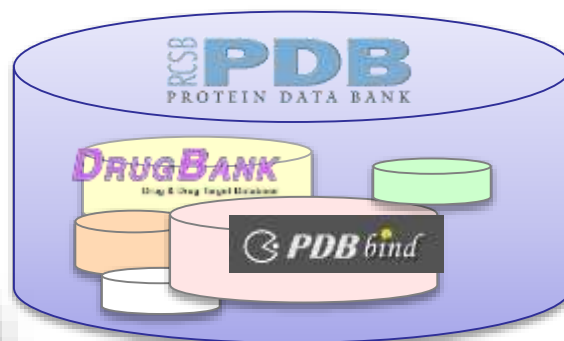
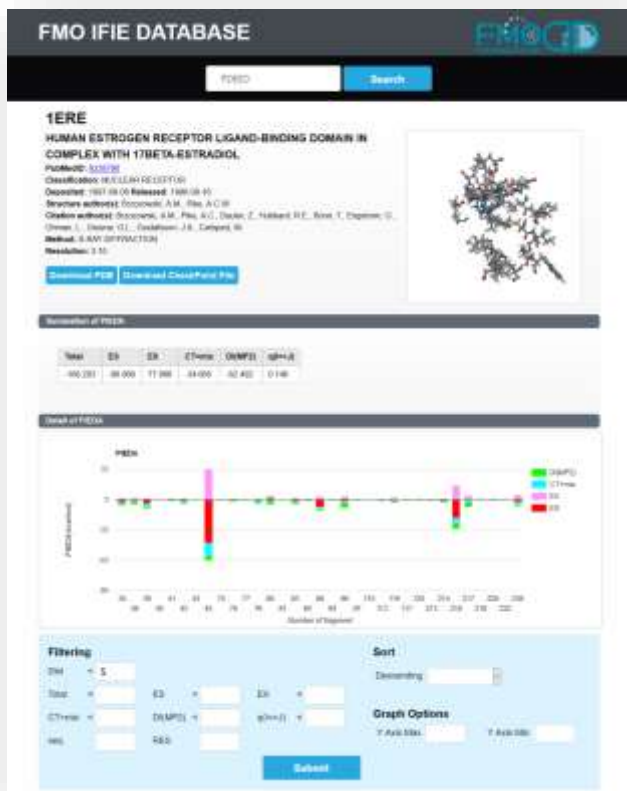
➤ **Good agreement between FMO-protocol and manual labor data.**

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

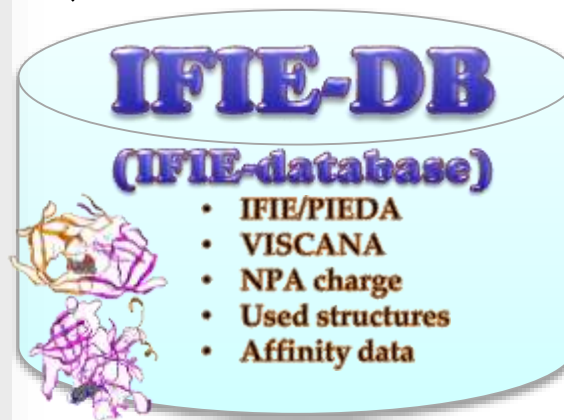
FMO IFIE データベース



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



<http://drugdesign.riken.jp/>

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)

Outline

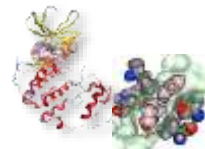


1. インシリコスクリーニングの技術開発

2. インシリコスクリーニング・設計の実例

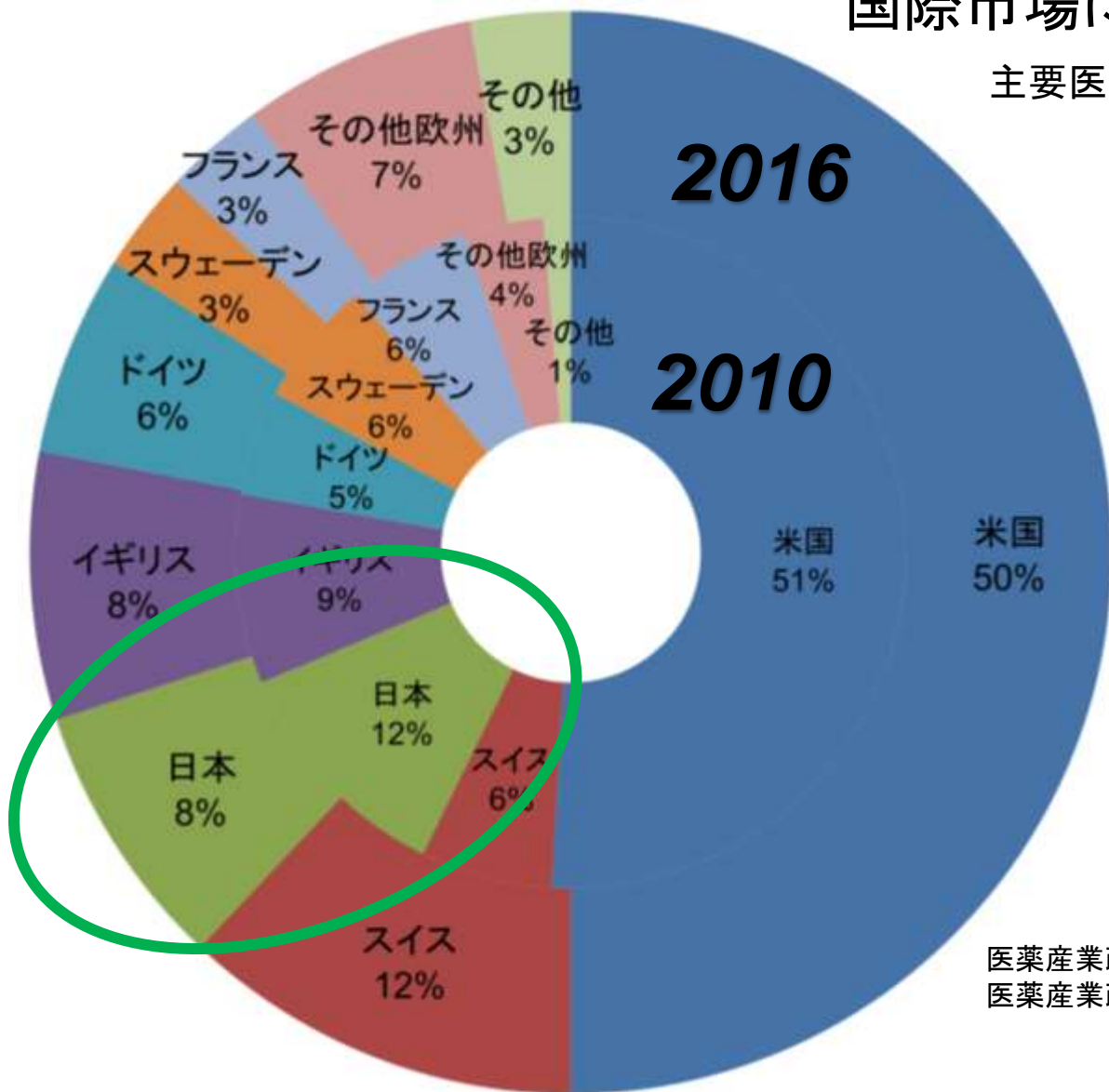
3. 結合親和性予測／FMOの創薬活用

時間があれば...最近の話題



国際市場における創薬シェア

主要医薬品(約100品目)の創出国



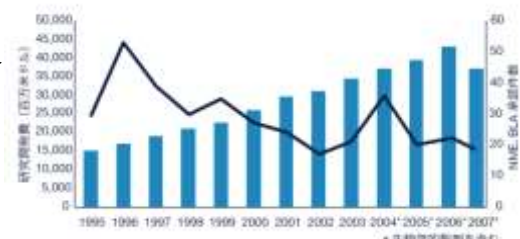
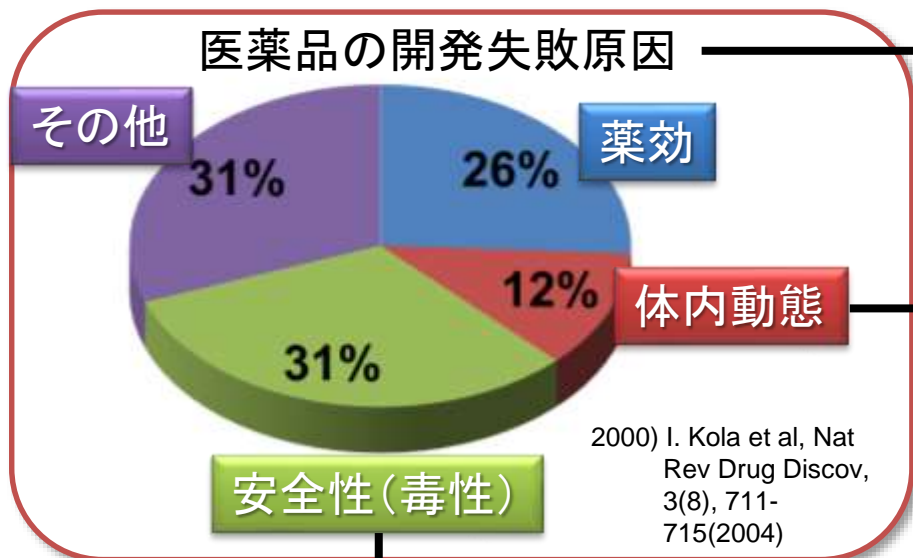
やや凋落しつつありますが、

日本は世界GDP
シェア(名目で6%前後)以上のシェアを
現状では持っている

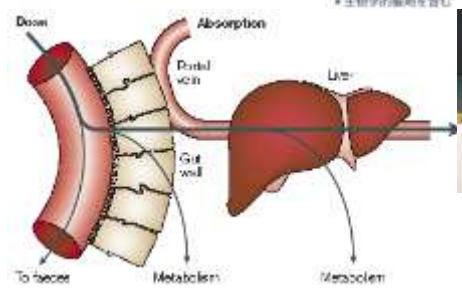
医薬産業政策研究所 政策研ニュースNo.29 2010年1月
医薬産業政策研究所 政策研ニュースNo.47 2016年3月

AMED創薬支援推進事業 「創薬支援インフォマティクス構築」

＝ 体内動態と安全性(毒性) 予測システムの構築



これらの原因が開発費 (1医薬品あたり2000億円)の高騰の原因



水口賢司 (医薬基盤研)

心毒性

毒性が原因での開発失敗のうち、心毒性と肝毒性で合わせて80%を占める

肝毒性



本間光貴 (理研)



山田弘 (医薬基盤研)

以下の予測システムを機械学習とシミュレーションを用いて開発

- 代謝酵素(CYP)に対する安定性
- 中枢移行性
- 薬物動態(血中濃度の推移)

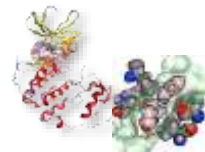
以下の予測システムを機械学習とシミュレーションを用いて開発

- 心毒性の原因タンパク質(hERG)への親和性
- hERGへのドッキングに基づく設計

肝毒性に関するバイオマーカーを同定し以下の予測システムを開発

- 肝臓由来iPS細胞のバイオマーカー実測値による肝毒性予測

人工知能オープンイノベーション (LINC)



ライフ系とIT系企業による 産学AIコンソーシアム設立



代表：奥野教授
(京大)

産学コンソーシアムによる
データ共有、モデル共有、AI共同研究開発

京大・理研・医薬基盤研・東大等(7機関)ライフ系
企業(35社)
IT系企業(27社)
その他(3)とで産学コンソーシアムを設立
参加者約300名



予防・先制医療

WG1. 未病・先制医療

- 健康診断データによる発症予測
- マイクロバイオーム・オミクスデータ解析
- デジタルヘルス

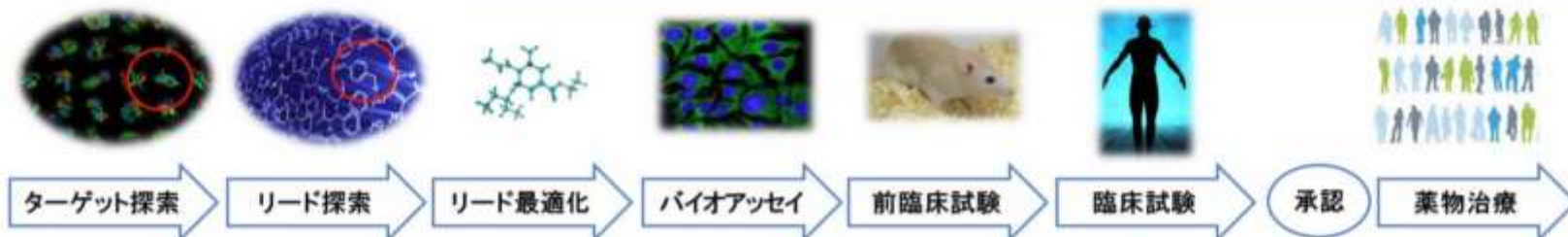
メディシナルケミストリー

WG4. 分子シミュレーション

- タンパク質立体構造・ターゲットペリテディ予測
- AIによるドッキング計算高度化
- MDトラジェクトリー解析
- 深層学習を用いた高精度分子力場

WG5. メドケム・分子設計・ADMET

- 合成経路予測
- 深層学習を用いた化学構造式生成モデル
- 深層学習を用いた化合物特性予測法
- QSAR/QSPR/ in vitro ADMET予測



WG3. 創薬テーマ創出

- 文献情報を基にした研究者探索
- 標的分子探索
- ドラッグリポジショニング

WG6. トランスレーショナルリサーチ

- 非臨床データからのヒトADMET予測
- 疾患メカニズム解明・ブリッジング予測

WG2. 臨床・診断

- がんゲノム医療におけるAI活用
- 微小流体シミュによる細胞分離
- 病理画像からの所見抽出
- 電子カルテによる患者層別化

WG7. バイオリジクス・製剤・ロボティクス

- バイオリジクス関連AI
- 結晶形・製剤関連AI
- 調剤ロボティクス

WG8. 治験・市販後・メディカルアフェアーズ

- AIによる治験の効率化
- 有害事象の情報基盤
- くすり相談 (コールセンター/製品Q&A)
- アウトカムリサーチ・医療技術評価

バイオメディカル・基礎から臨床への開発プロセス

診断・治療

WG9. 知識ベース・NLP

知識ベース / 計算機サーバー

WG10. AI基盤

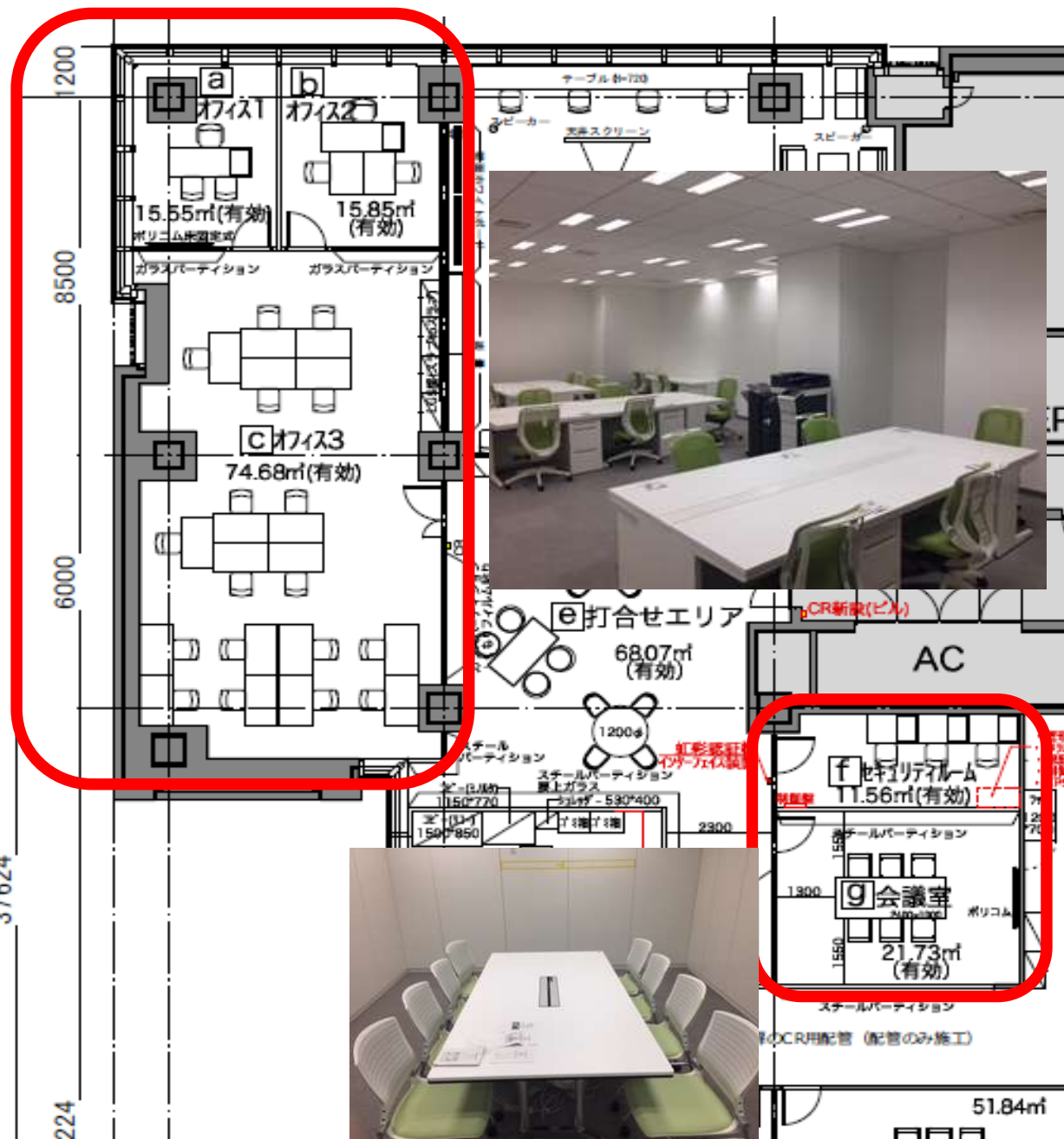
理研医薬プロセス最適化Gの横浜拠点

中央研究棟5階で1スパンを整備。
11月以降利用可能

スペースが足りない場合、隣の部屋も
利用できる。(白水部門長了承済み)



理研 医科学イノベーションハブ推進プログラムの東京拠点



〒103-0027 東京都中央区
日本橋1-4-1 日本橋一丁目
三井ビルディング 15階
東京連絡事務所
Tel: 03-3271-7101 / Fax: 03-3271-7100

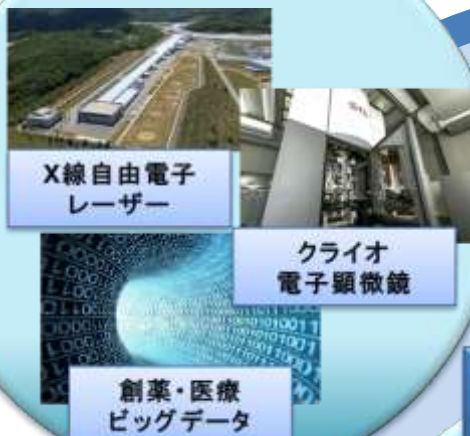
9月中旬から利用可能

in silico創薬の将来

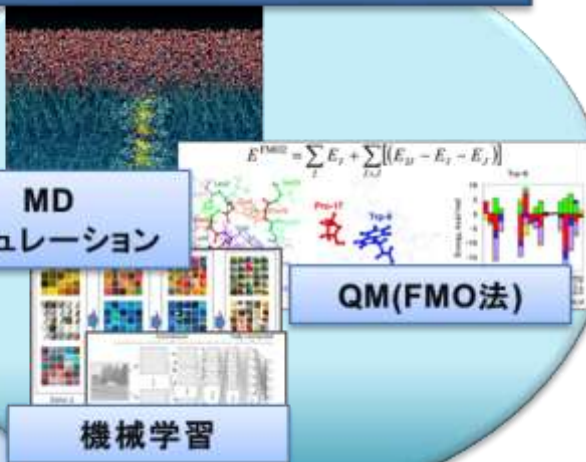
生体分子シミュレーション、構造生物学、ビッグデータの連携からアカデミア創薬へ

生体分子シミュレーション・
統計解析

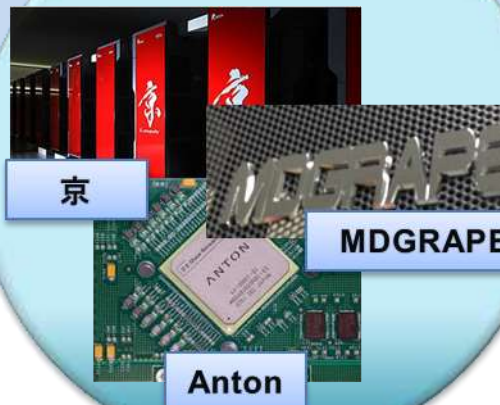
新時代の実験技術・
ビッグデータ



MD
シミュレーション



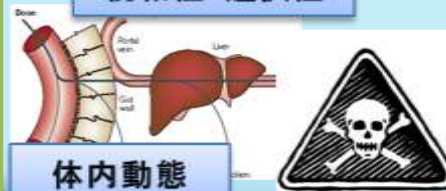
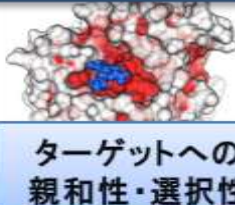
スパコン・
MD専用機



より多くの
創薬現場のニーズを
満たすための課題

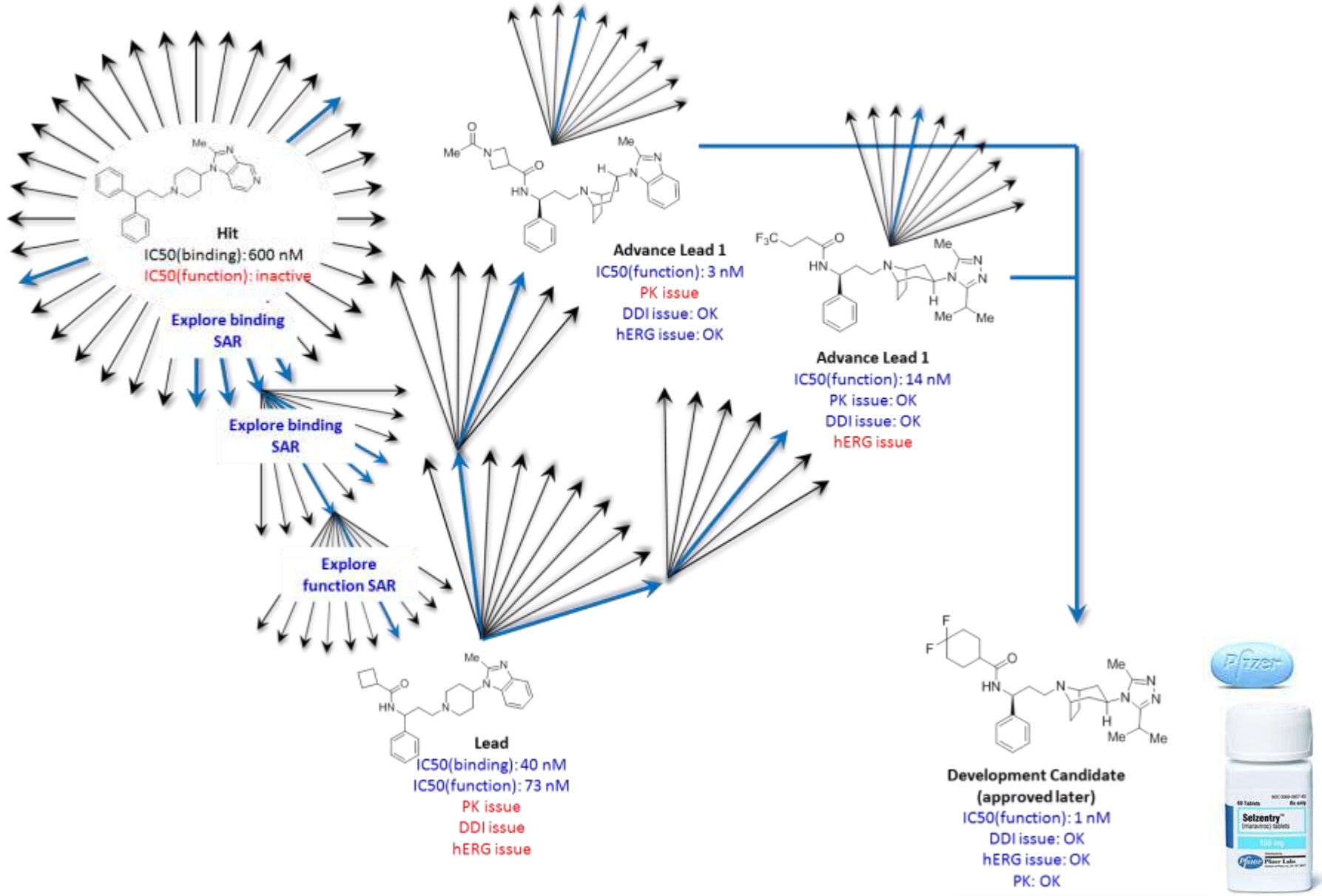
人工知能創薬への
昇華が可能か？

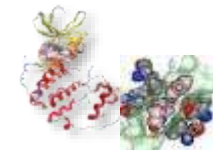
創薬に必要な
項目の予測



現状の成功例：
得意なところでは
成功するが、苦
手な対象も多い

抗HIV薬 Selzentryのヒットから開発候補品まで





苦手な予測対象

柔軟なタンパク質、巨大な複合体タンパク質、ポケットが浅い・小さい、中分子・高分子医薬品の設計

体内動態・毒性

- 原因タンパク質の構造解析が遅れている
- 同じ実験条件のデータを大量に収集しにくい

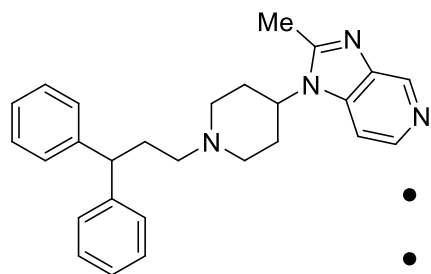
スピード

タンパク質構造解析や大規模シミュレーションによる予測では、結果が出るまでに数か月以上かかる場合もある。

総合的な判断

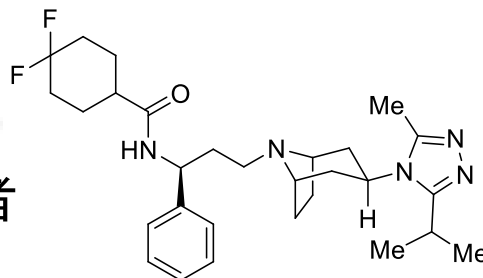
単一の予測はできても複数の同時最適化という難しい課題への回答を出せない。

抗HIV薬 Selzentryのヒットから開発候補品まで



ヒット

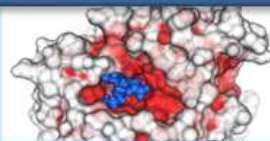
- 5-10人の創薬化学者
- 2-4年
- 1000個以上の化合物



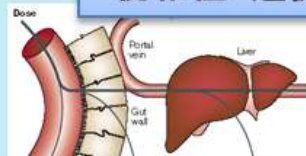
開発候補品



創薬に必要な
項目の予測



ターゲットへの
親和性・選択性



体内動態



安全性
(毒性)

インシリコ創薬からシームレスなAI創薬へ

生体分子シミュレーション・
統計解析

新時代の実験技術・
ビッグデータ



MD
シミュレーション



QM(FMO法)

機械学習

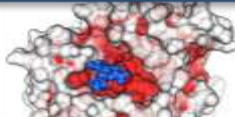
スパコン・
MD専用機



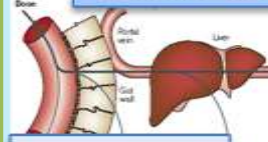
より多くの
創薬現場のニーズを
満たすための課題

AI創薬への
昇華が可能か？

創薬に必要な
項目の予測



ターゲットへの
親和性・選択性



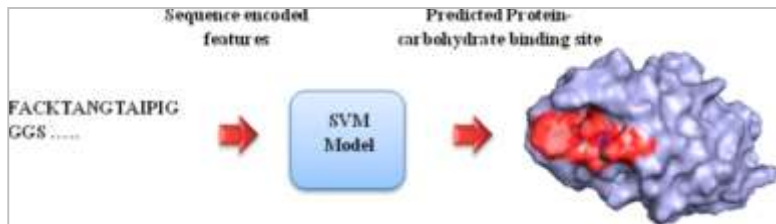
体内動態



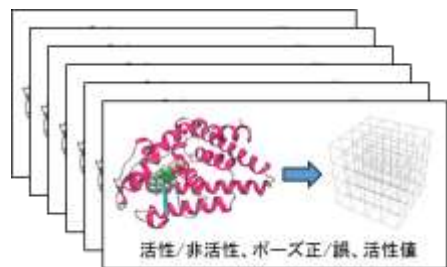
安全性
(毒性)

現状の成功例：
得意なところでは
成功するが、苦
手な対象も多い

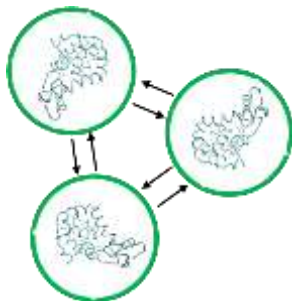
PJ11 ターゲットビリティー



PJ12 ドッキングポーズ



PJ13 MD解析

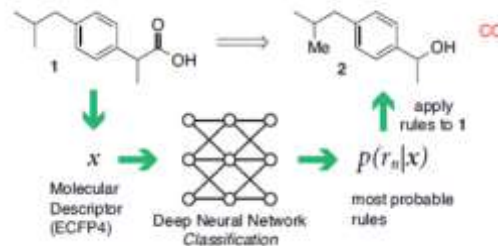


PJ14 QM力場

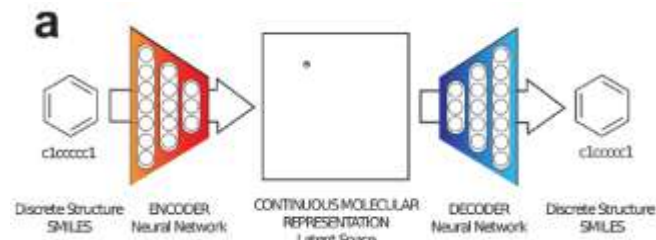


PJ15 合成経路

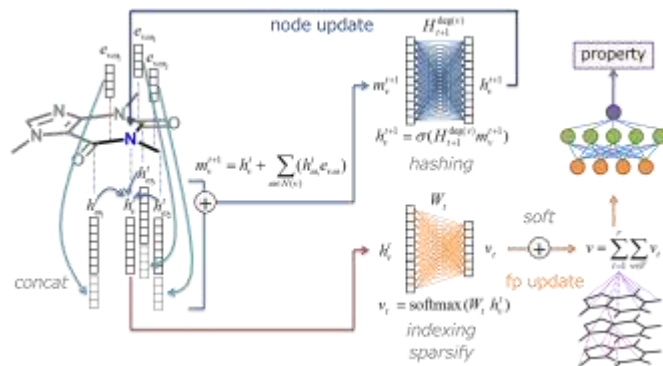
e) Neural Networks determine the Actions



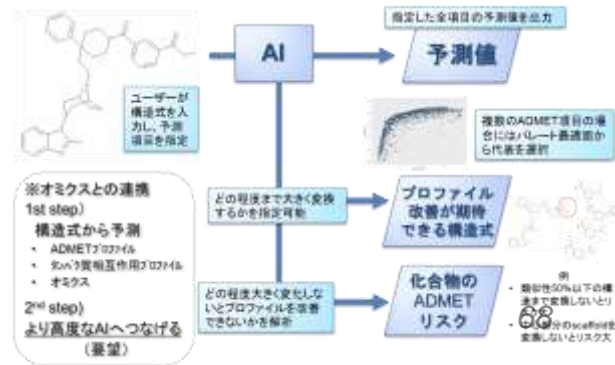
PJ16 新規構造発生

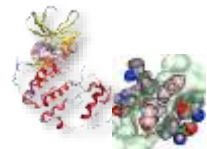


PJ17 記述子



PJ18 ADMET





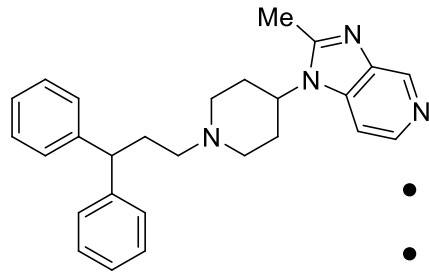
予測の実現・高精度化

- これまで困難・不可能であった予測を可能に
- 従来の予測の高精度化

シームレス化・統合

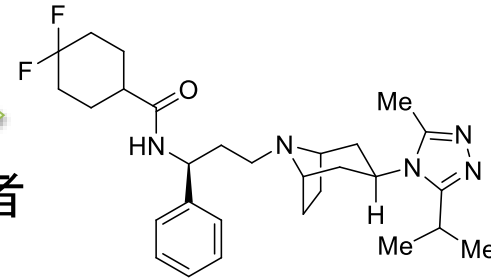
- 創薬は典型的な多項目最適化問題
- 複数の項目の異なる予測モデルを統合し、最適な判断を促す創薬AIが重要

カギは複数のAI・シミュレーションの統合

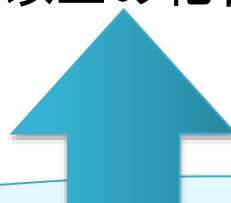


ヒット

- 5-10人の創薬化学者
- 2-4年
- 1000個以上の化合物



開発候補品



PJ11 ターゲットアピリティー



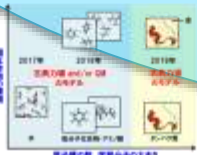
PJ12
ドッキング
ポーズ



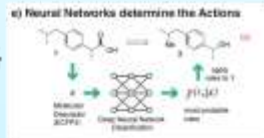
PJ13
MD解析



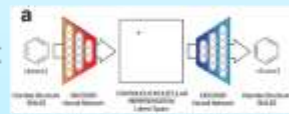
PJ14
QMカ場



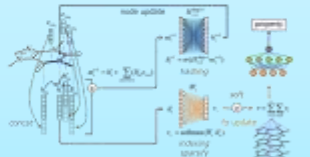
PJ15
合成経路



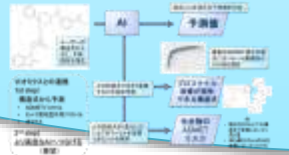
PJ16
新規構造
発生



PJ17
記述子







PJ18
ADMET



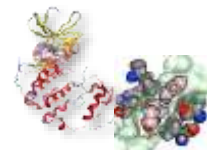
ヒットから開発候補品への最短距離を提案する「AI創薬」は可能か？

- 高精度の予測
- 合成難易度を考慮した構造変換候補のリストアップ
- 体内動態・毒性など多数の項目の同時最適化
これらを高いレベルで統合する必要がある

理研・データプラットフォーム事業(奥野G)の概要

グループ名	ユニット名	研究概要
医薬プロセス最適化プラットフォーム推進グループ 奥野GL 本間副GL	創薬バイオメディカルインテリジェンスユニット 奥野UL 	基礎研究と臨床研究のブリッジング技術開発 大学病院等と連携し、体内動態予測による薬物治療の最適化などを行う。また、システムケミカルバイオロジー(多階層の生体系と分子との相互作用ネットワーク解明)も連携して実施。
	メディシナルケミストリーインテリジェンスユニット 本間UL 	医薬品合成化学者の暗黙知の人工知能化 AIを活用したメディシナルケミストリー(創薬化学)の効率化を行う。特に新たな医薬候補の創出過程を合成反応予測やADMET(薬物動態、毒性)予測等を組み合わせてAI化する。
	分子設計インテリジェンスユニット 池口UL 	AIとシミュレーションを活用した、生体系を制御する低・中・高分子の設計法の開発 今まで薬剤の開発が難しかった動きがあり複雑な生体分子について、動きをシミュレーションで予測、複雑さをAIで解明し、制御する分子を設計する。
	創薬インテリジェンス連携基盤ユニット 奥野UL 	DMPの創薬基盤(創薬化学基盤ユニット、創薬分子設計基盤ユニット)と連携し、創薬化学の研究者の経験知に基づいている化合物の設計プロセスのAI化に向けて、DMPで進める具体的テーマを題材に合成候補化合物の構造提案や薬効予測等の明確なミッションを定めた研究に取り組む。

謝辞に変えて: 研究室のメンバー



幸 瞳
研究員
医薬分子設計全般



渡邊 千鶴
研究員
量子化学計算



保田 真友子
技師
(調査・
アシスタント業務)



佐藤 朋広
研究員
機械学習



佐藤 敦子
客員研究員
製薬企業(共同研究)



永瀬 俊平
テクニカルスタッフ



高谷 大輔
研究員
タンパク質モデリング



神坂 喜久子
派遣職員



小高 明日香
派遣職員



小倉 圭司
委託研究員
製薬企業(共同研究)

