

ドッキングソフトの原理と実際

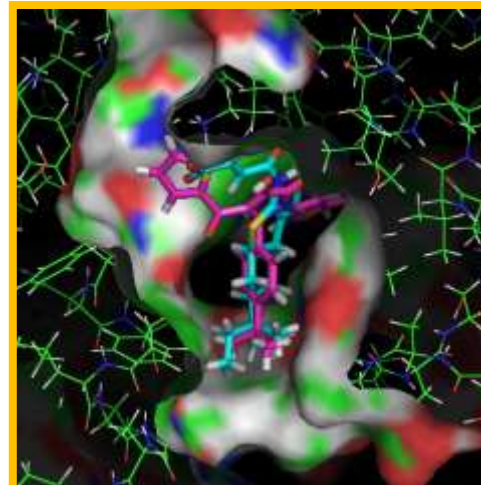
3D分子設計チーム

研究チーム長： 福西 快文

神戸大学計算科学教育センター
セミナー室208

12月6日

17:00-18:30



myPresto

[Documents](#) | [Download](#) | [Download\(Osaka Univ.\)](#) | [myPresto\(JBiC\)](#) | [Contact](#)

myPresto

"myPresto (Medicinally Yielding PRotein Engineering SimulaTOr)" is a free software for Structure Guided Drug Development (SGDD). You can download "myPresto" from this web site.

[myPrestoは、医薬品開発支援用分子シミュレーション計算ソフトおよびデータベースの集合体です。このウェブサイトより、無償プログラムmyPrestoのダウンロードができます。]

>> [DOWNLOAD](#)



独立行政法人
産業技術総合研究所

創薬分子プロファイリング研究センター

National Institute of Advanced Industrial Science and Technology

Molecular Profiling
Research Center
for Drug Discovery

分子設計に必要なこと

(0) スクリーニングなどにより、活性化合物を得る

(1) タンパク質と活性化合物の正確な複合体構造を知る。
X線構造解析、NMR実験、分子シミュレーション計算

(2) タンパク質と活性化合物の正確な結合活性を知る。

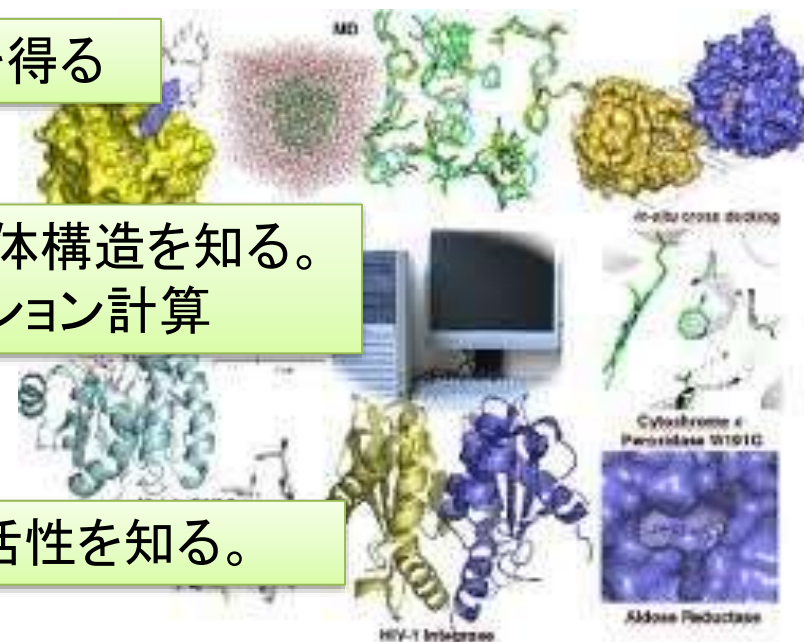
(3) 化合物を改変する(人間の知恵によるアートの世界)

3-① 「合成できる化合物」で、

3-② 水に溶け、吸収できる化合物でなければならない。

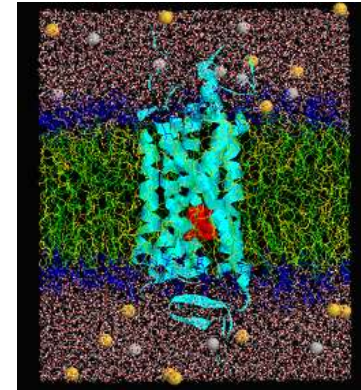
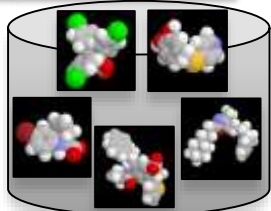
3-③ Kinase/GPCRの場合は、Off-targetに作用しないこと。

(4) 合成、アッセイ実験



医薬品候補物質の探索から、分子設計までを、
一通り、**ラフ**に実行可能な環境”myPresto”を整備している

myPrestoの概要

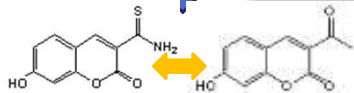


①4000万化合物データベース
天然物を含む世界最大級・購入可能

②分子シミュレーション
スパコンで使われるGPU利用
で30倍高速・世界最高速

④薬物ドッキング
世界最高速 世界標準の10-100倍

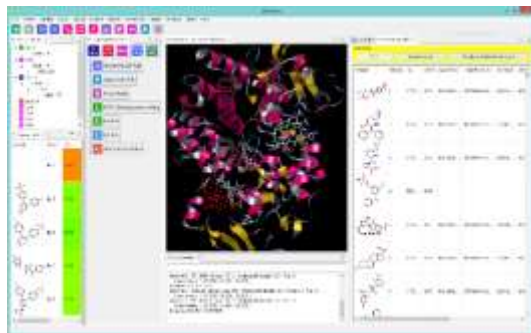
③薬物結合部位予測
世界最高精度レベル



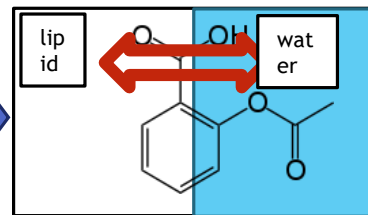
⑥類似化合物探索
世界最高レベル・
プレスリリース

⑤薬物スクリーニング
：世界標準の10倍正確

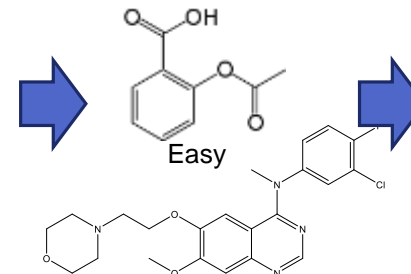
① 薬物候補物質の探索



⑦分子設計



⑧溶解度など物性予測

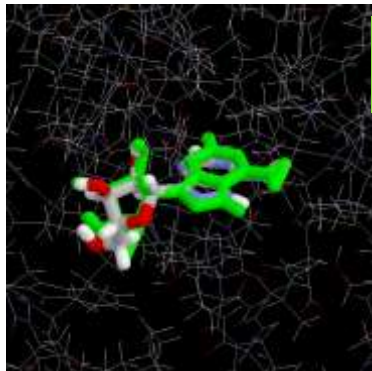


⑨合成可能判定
世界で2つのみ

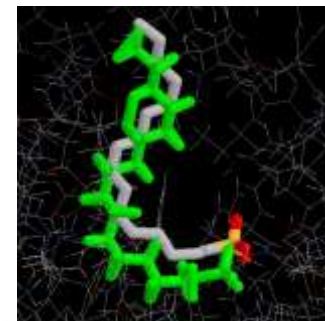
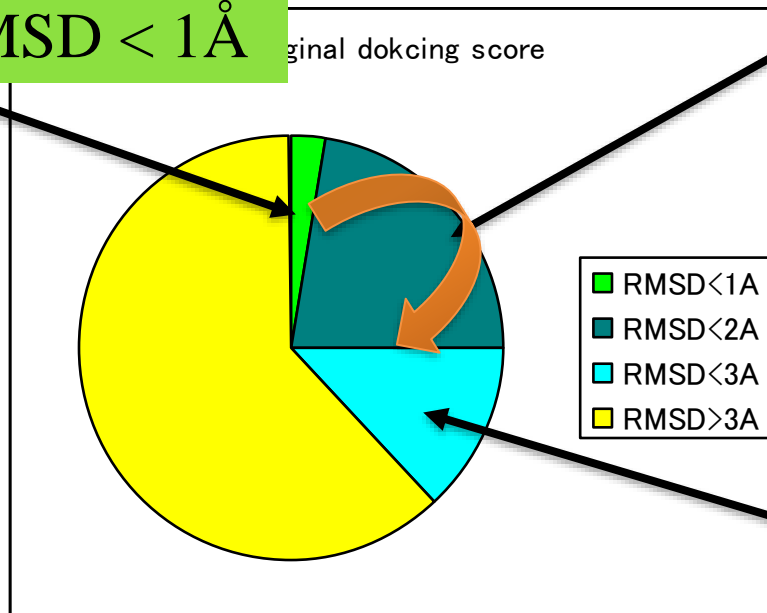
②薬物の最適化

1.ドッキングソフトの精度

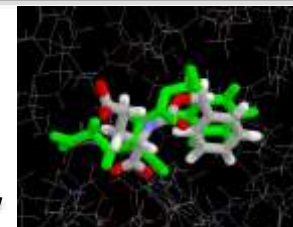
構造の予測精度: 25%



RMSD < 1Å



RMSD < 2Å



RMSD < 3Å

2.活性予測の精度

誤差 1~3 kcal/mol

手法と情報の精度に依存

3.薬物スクリーニングソフトの精度

ヒット率 3~5%程度

硬いタンパク質に限る

4.タンパク質薬物結合ポケット予測ソフトの精度

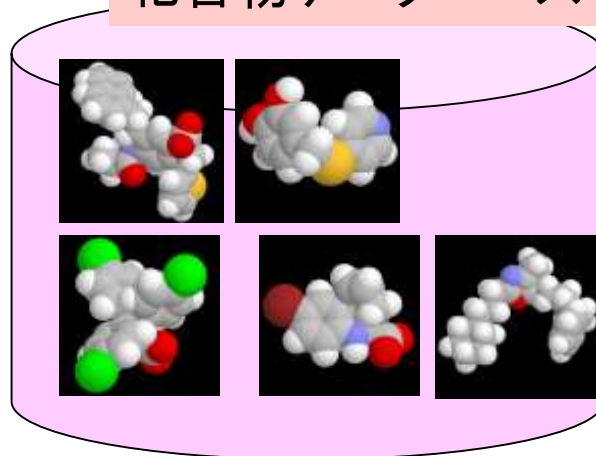
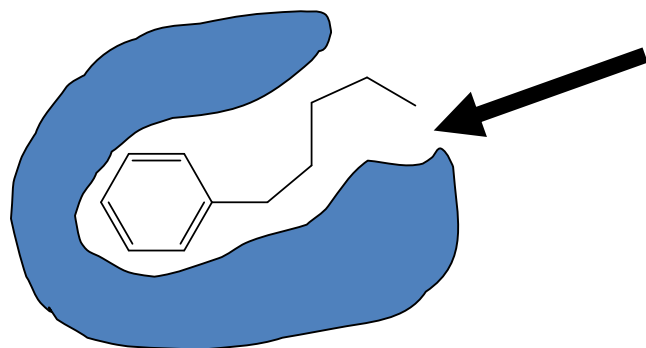
80%程度

天然の基質が存在する場合

薬物スクリーニング手法は主に2つ

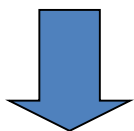
化合物データベース

標的蛋白質の立体構造がある

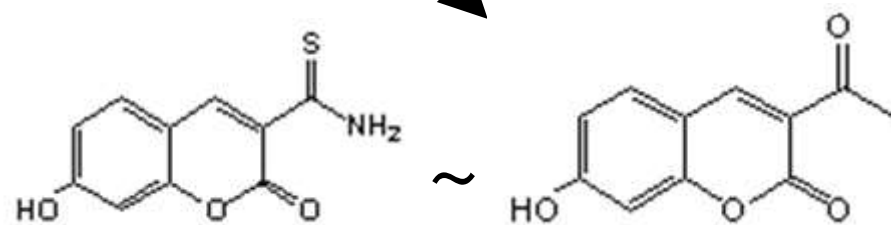


既知の活性化合物がある

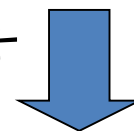
標的蛋白質に結合する分子を探す



Structure-based drug screening

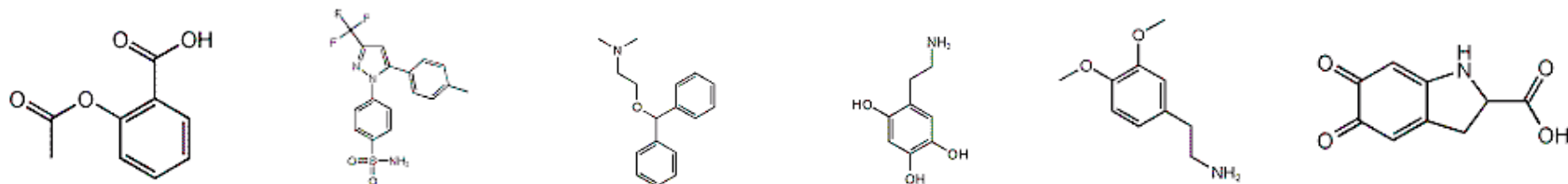


似た分子を探す



Ligand-based drug screening

化合物3次元構造データベース *LigandBox* : HPで誰でも利用可能

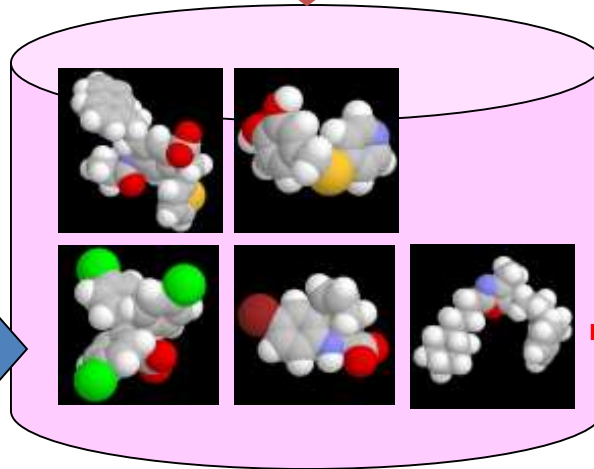


市販化合物は全世界で700万件。数百社がカタログ(化合物の絵)とともに販売している。

化合物カタログ情報

myPrestoで3D化し、
電荷も付与。
(水素、異性体も
発生させる)

*In-silico*スクリーニング
ソフトウェア

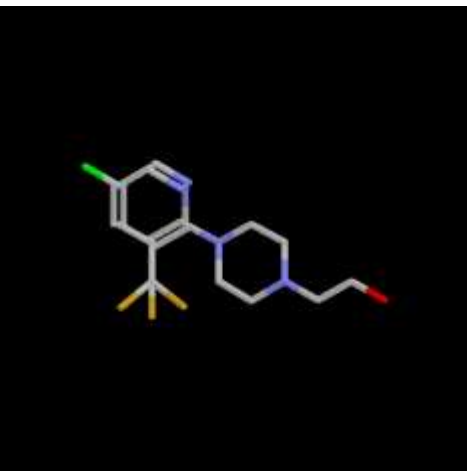


molprof/ナミキ商事
から配布

インターネットでも。

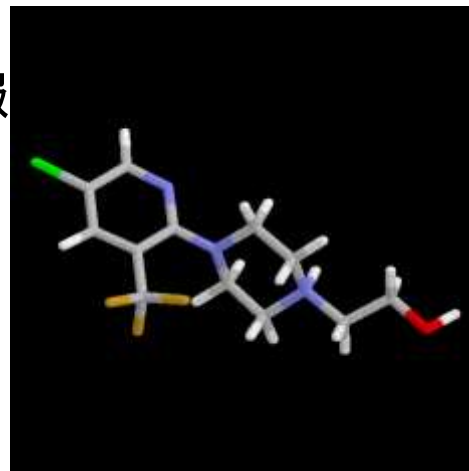
Googleで、
「Ligandbox」で検索

2次元電子カタログからの化合物3次元データベースの生成

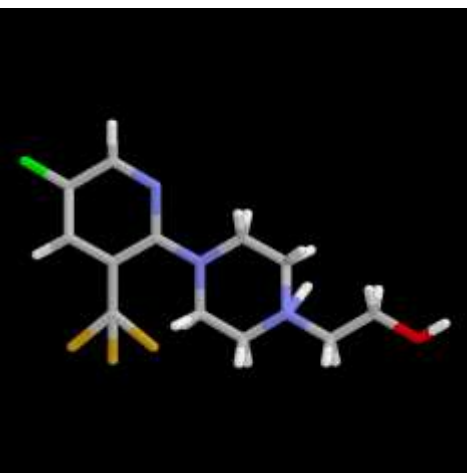
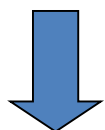


電子カタログの情報

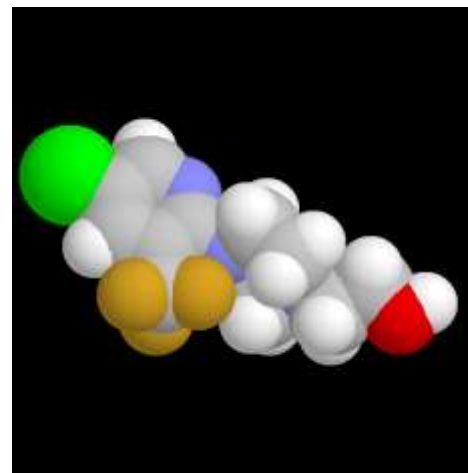
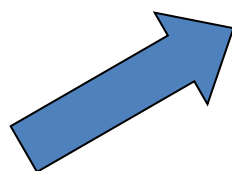
- ・2次元平面構造
- ・水素原子なし
- ・電荷情報なし
- ・光学異性なし



- ・エネルギー的に安定な立体構造の発生
- ・光学異性体の発生



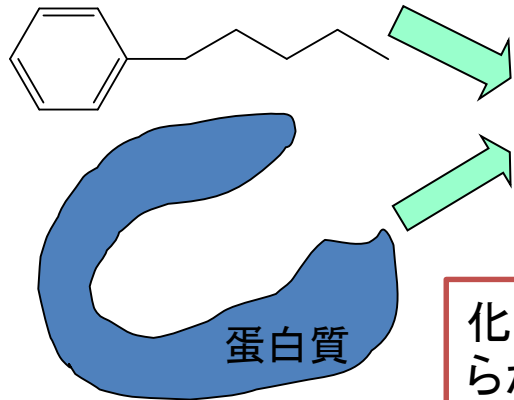
- ・水素原子の付加
- ・芳香族環の判定
- ・原子タイプの判定
(同じ炭素でも結合が2, 3, 4本あるものは性質、大きさが異なる)



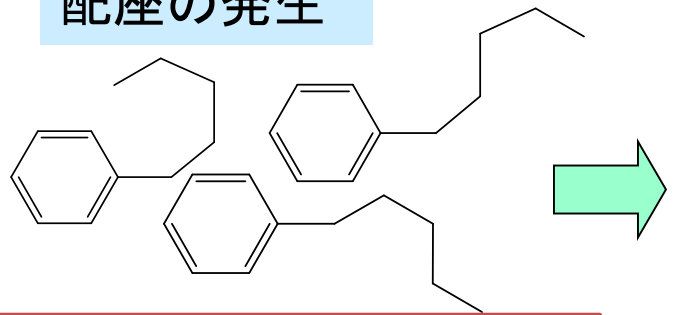
- ・量子化学計算による原子電荷発生
- ・mol2ファイル形式へ変換
- ・データベース化

化合物ドッキングソフト myPresto/Sievgeneの概要

入力分子

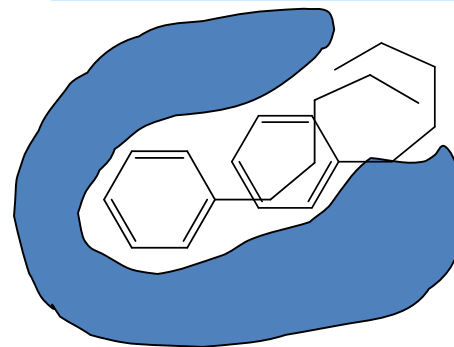


配座の発生



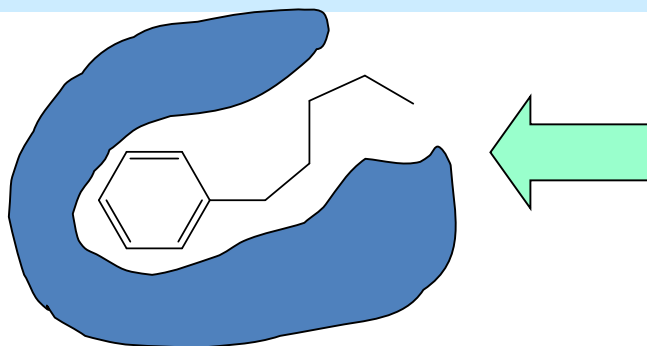
化合物の柔軟な構造変化はあらかじめ計算して準備する。

蛋白質への結合

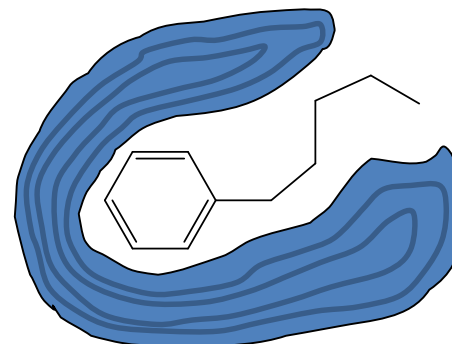


1秒間に10万回、化合物を蛋白質の表面に配置。

最適な複合体モデルの選択



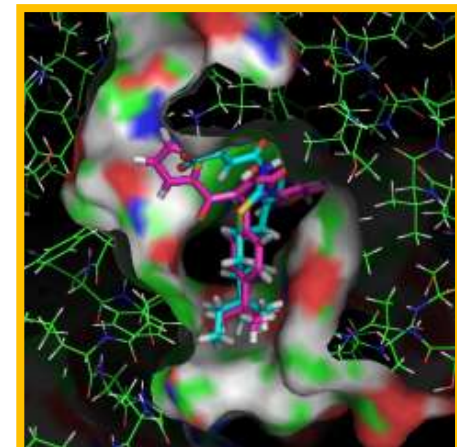
相互作用の見積もり



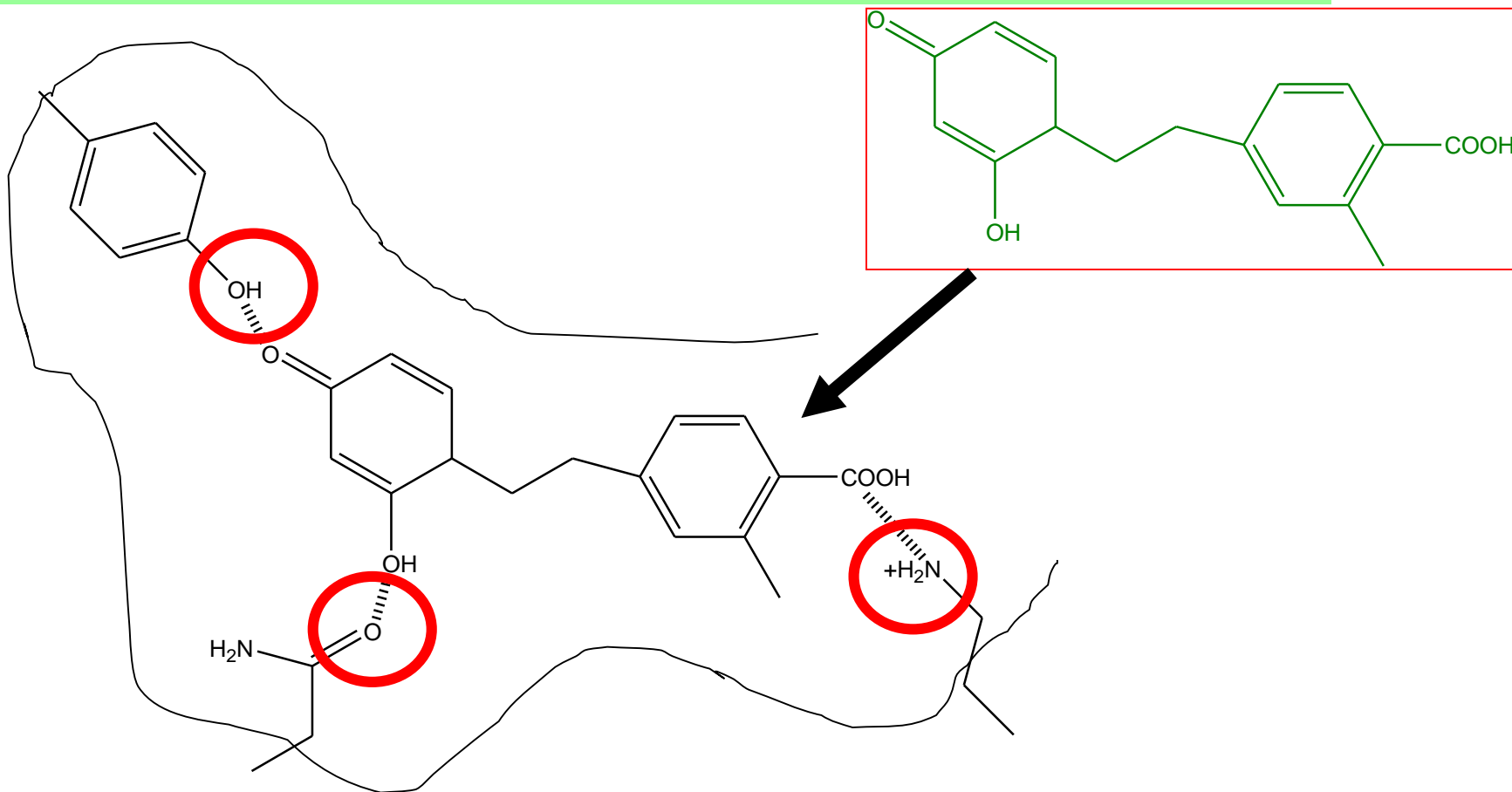
10万回配置された化合物と、弾性体としての蛋白質の相互作用を見積もり、結合エネルギーを推定する。

最適な蛋白質-化合物の結合構造とそのエネルギーを出力する。

結合自由エネルギー = 活性 = ドッキングスコア



蛋白質への化合物の結合 : Global search : ADAM/FlexX

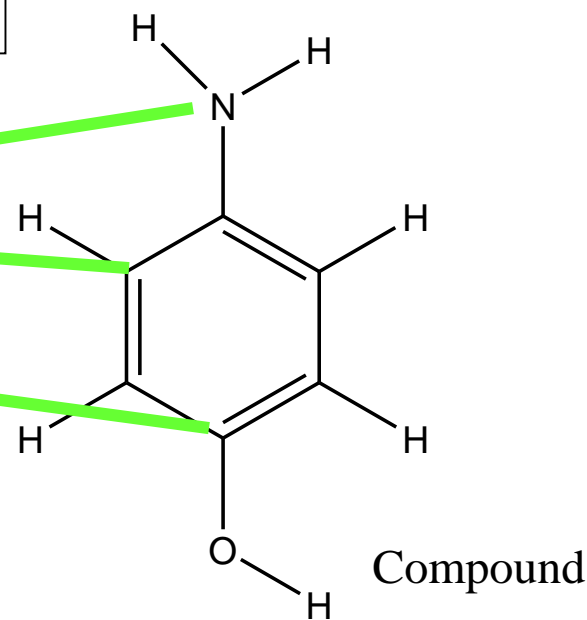
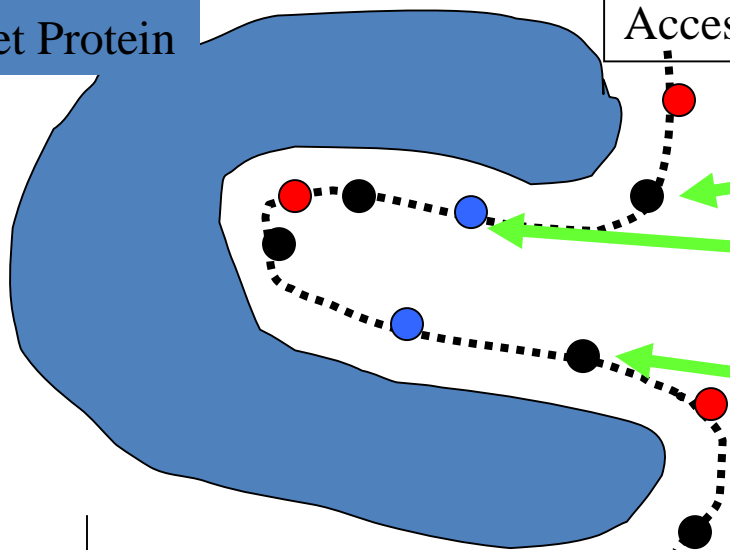


蛋白質表面の水素結合可能な残基をチェックして、水素結合可能なサイトへ、化合物の水素結合可能な官能基を重ね合わせるように、化合物を配置する。

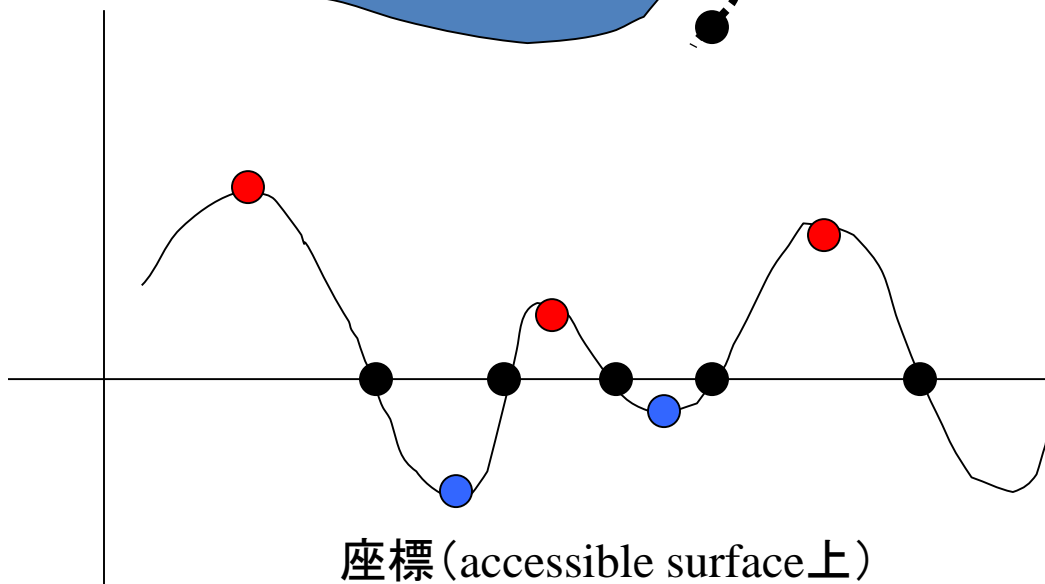
蛋白質への化合物の結合 Global search: myPresto sievgene

Target Protein

Accessible surface



静電場ポテンシャル



Accessible surface上の静電場の極大・極小点及び電位=0の点の集合に、化合物の任意のHeavy原子を重ね合わせる。

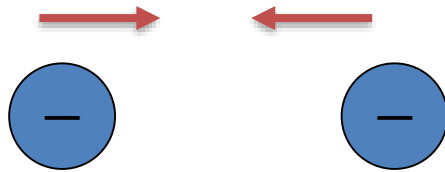
(1) 水中での原子・分子の相互作用の基本の考え方

水の中では、タンパク質と薬物の結合力の半分は、溶媒である水の運動自由度の減少が原因。

残り半分は、タンパク質と薬物の直接の相互作用（静電気的な引力など）である。



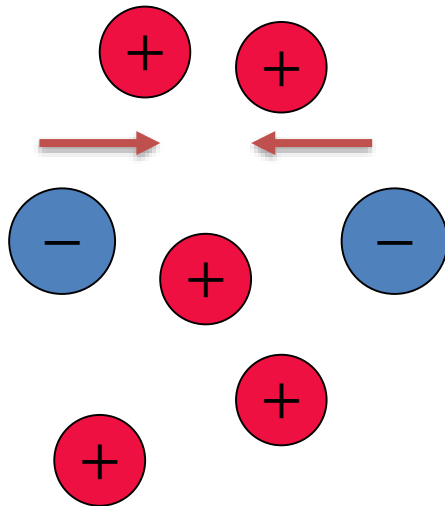
マイナスの電荷同士は
真空中では反発する。



マイナスの電荷同士は
水中では引き合う。
(核酸は、ばらばらにならずに
固まって沈殿したりする)



マイナスの電荷同士は
真空中では反発する。

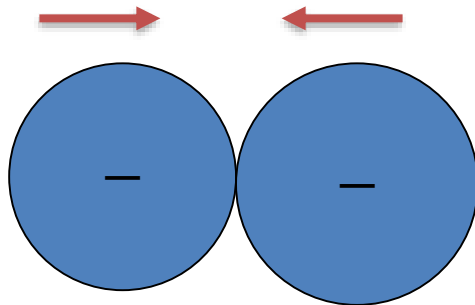


マイナスの電荷同士は
水中では引き合う。
(核酸は、ばらばらにならずに
固まって沈殿したりする)

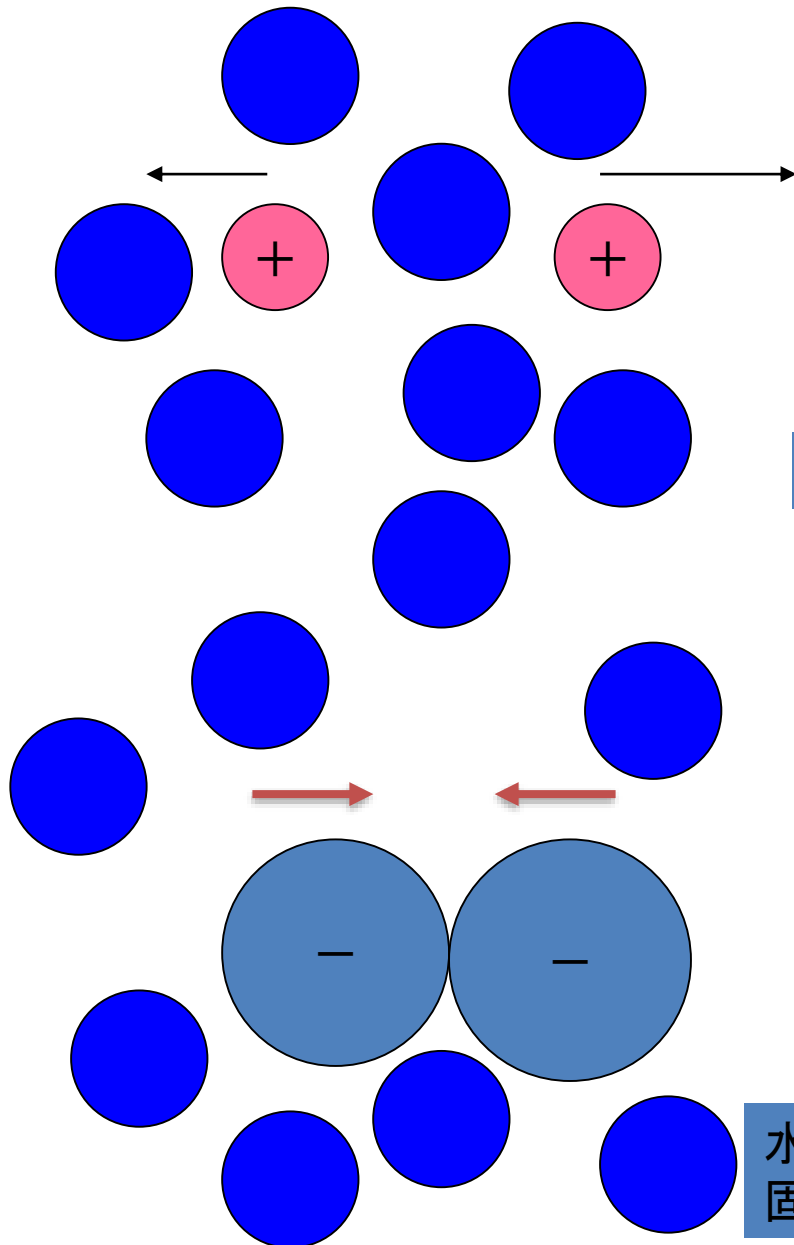
+電荷のイオンが集まって、「のり」の働きをする。



小さなプラスの電荷同士は
水中では反発しやすい。



大きなマイナスの電荷同士は
水中では引き合う。
(核酸は、ばらばらにならずに
固まって沈殿したりする)

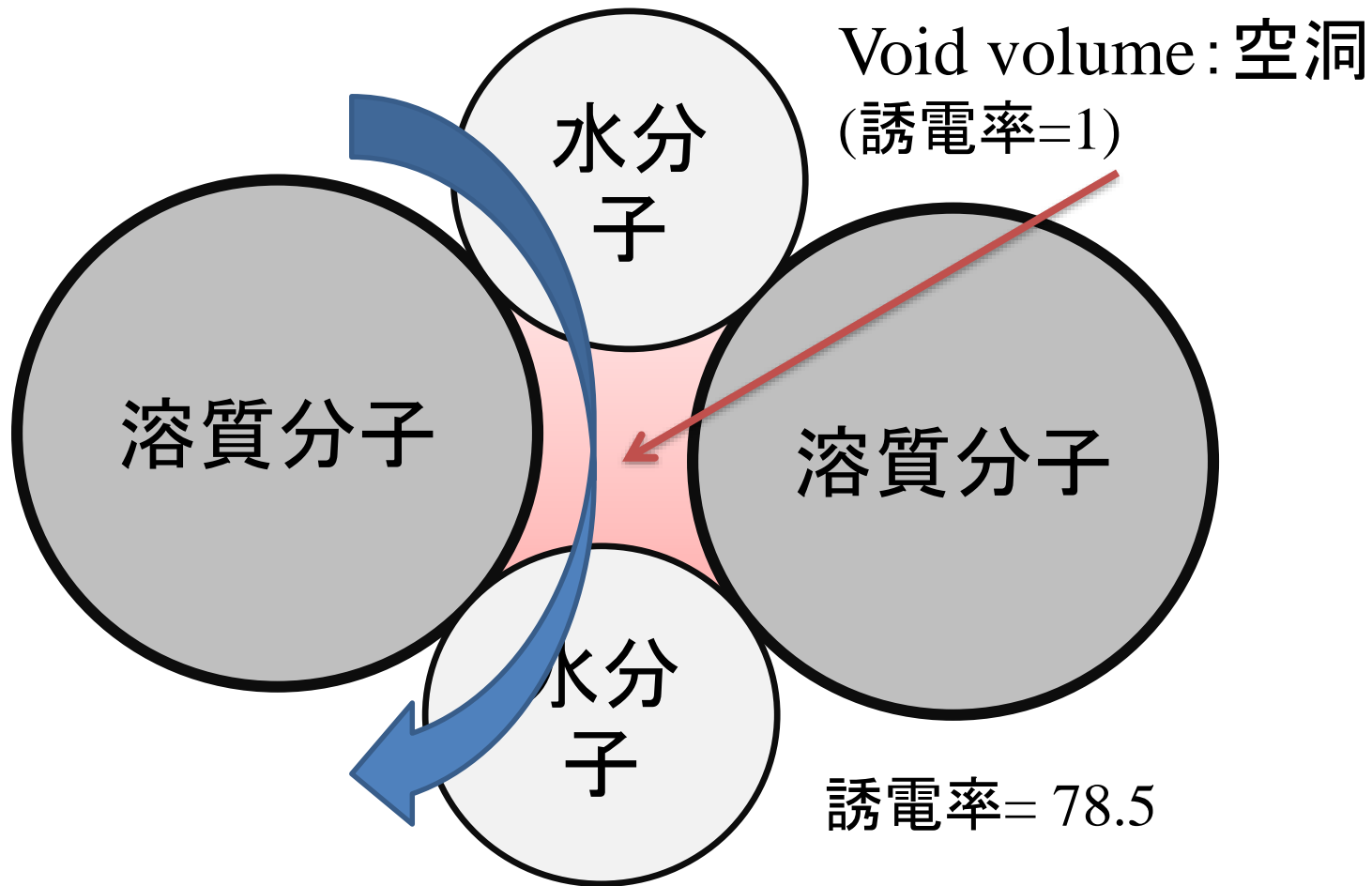


小さなプラスの電荷同士は
水中では反発しやすい。

水の中であまりじゃまにならない。

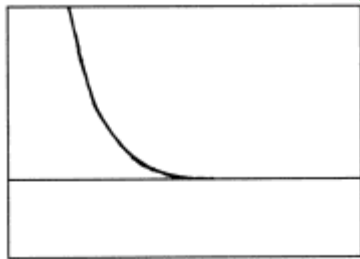
大きなマイナスの電荷同士は
水中では引き合う。
(核酸は、ばらばらにならずに
固まって沈殿したりする)

水分子にとって、おおきなじゃまもの。
固めておいたほうが、水にとってじゃまにならない。



- (1) 水分子が通り抜けられない = エントロピー現象 = 不安定化
- (2) 誘電率が小さい領域の出現 = 系の静電場エネルギーの不安定化

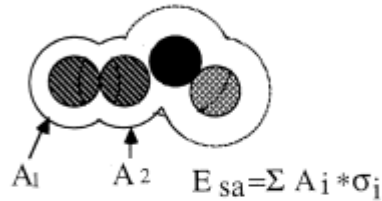
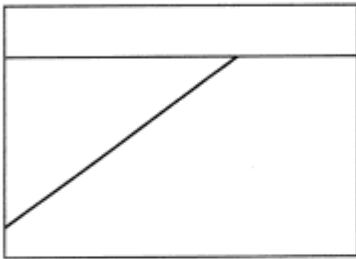
水中での分子間相互作用



van dar Waals repulsion energy

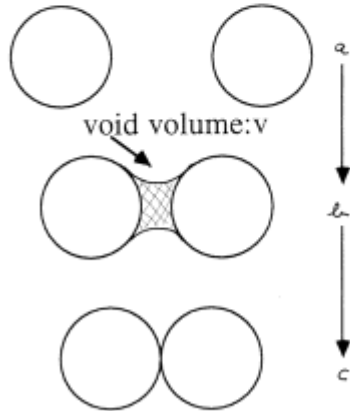
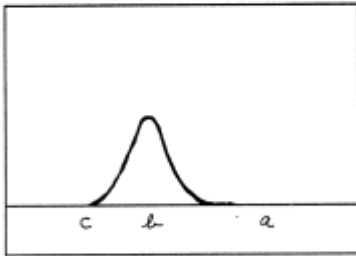


Energy from solvent accessible surface area

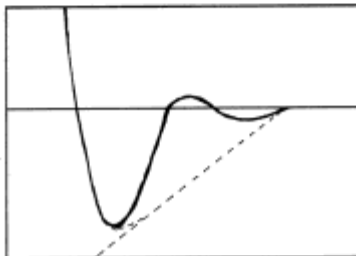


Energy from void volume

$$E_{vo} = \theta * v$$



Total energy



Interatomic distance

① VDW力は、ほとんど反発性。引力は、各原子対で、0. 1kcal/mol程度。

②疎水性相互作用は、溶媒水分子の自由度を低減することによって由来する。だいたい分子の表面積に比例する。

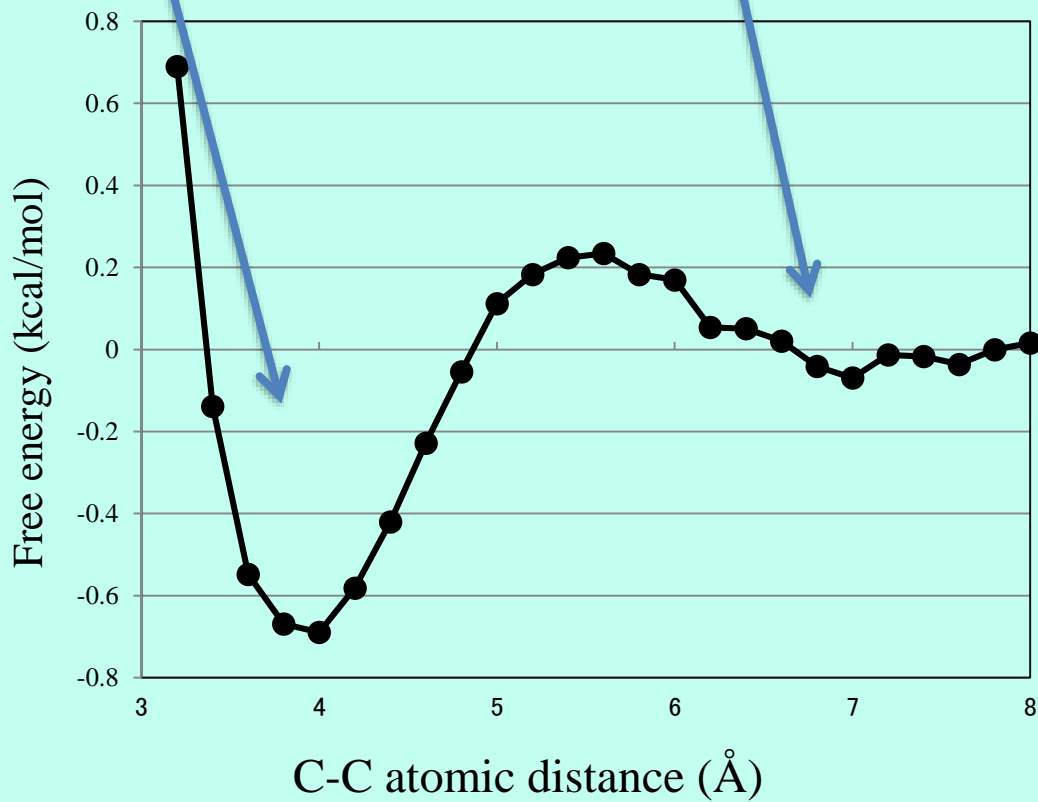
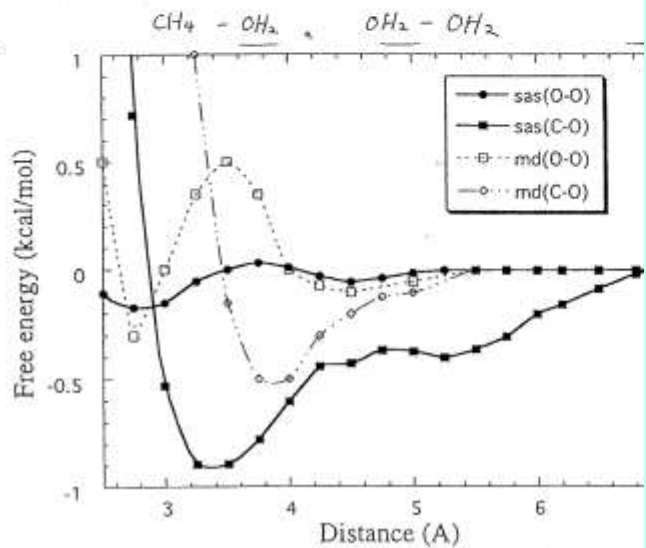
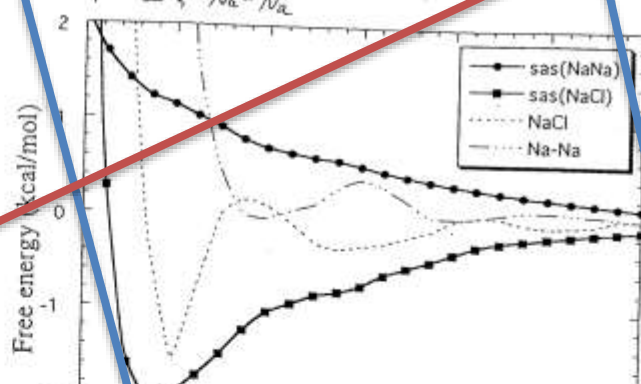
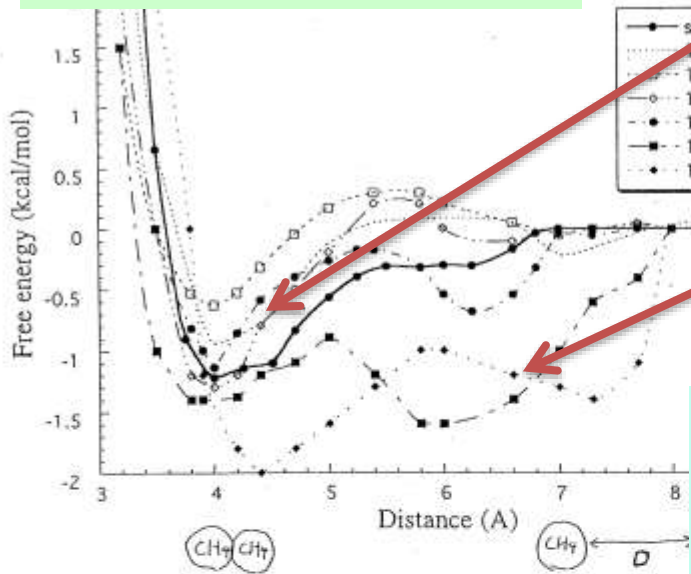
③分子のパッキングが悪い場合は、分子の排除体積効果が現れる。これは、2分子間を水分子が通り抜けられないことから生じる。

●全体を合計すると、図のように、分子間相互作用は波打った形となる。低分子の場合、引力は1kcal/mol程度。

(2) 分子間相互作用の例

精密な分子シミュレーションによる
分子間相互作用の実像

水中での分子間相互作用



CH₄

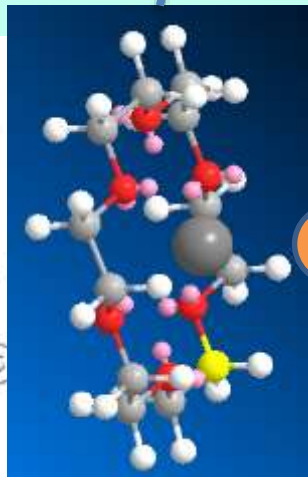
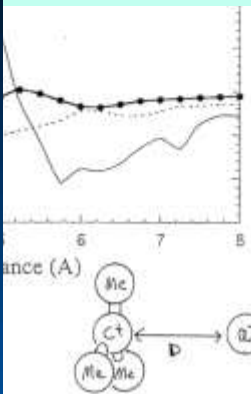
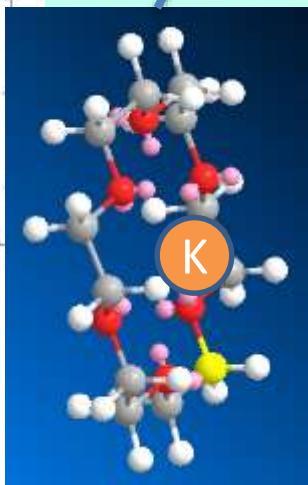
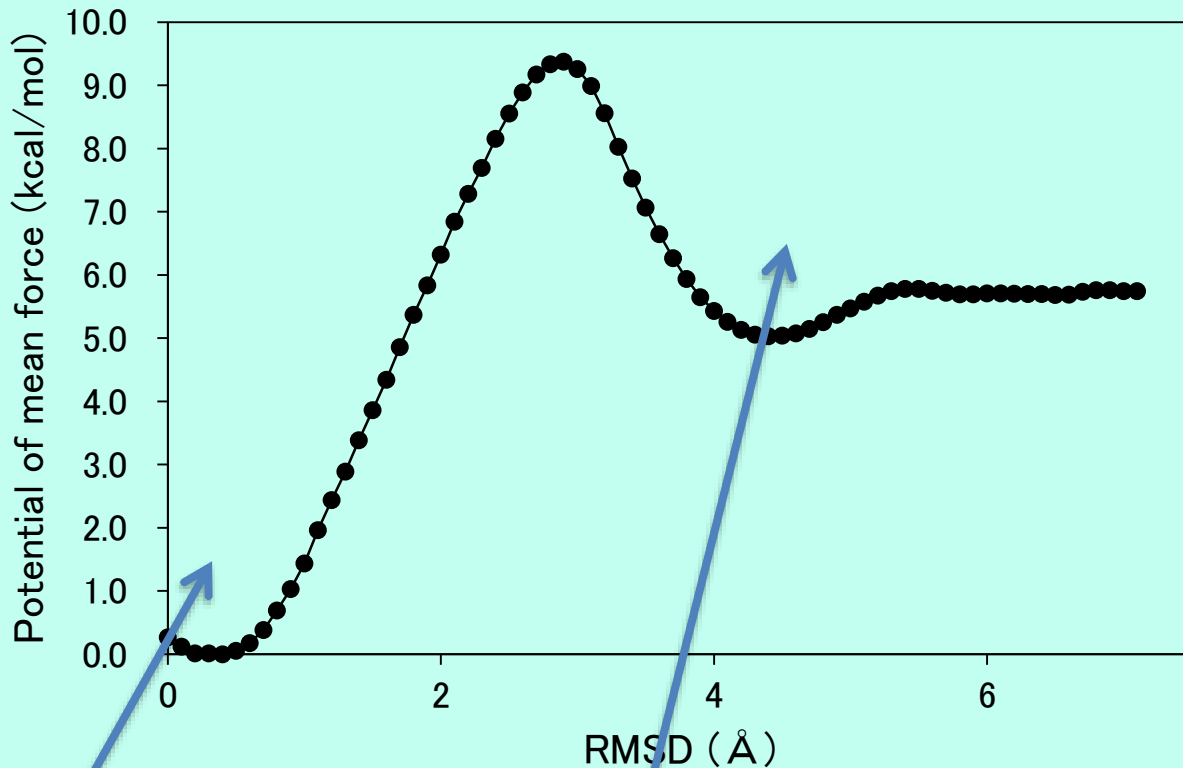
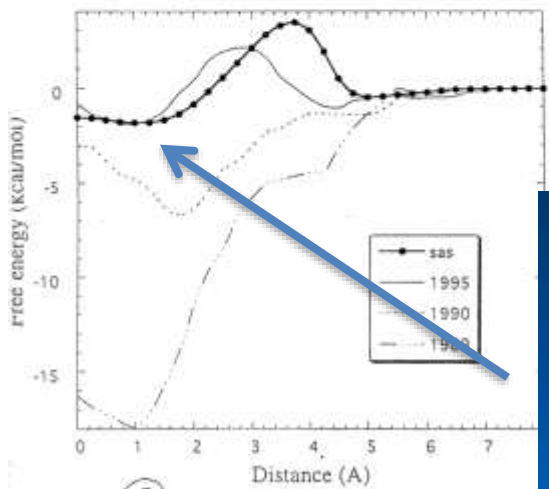
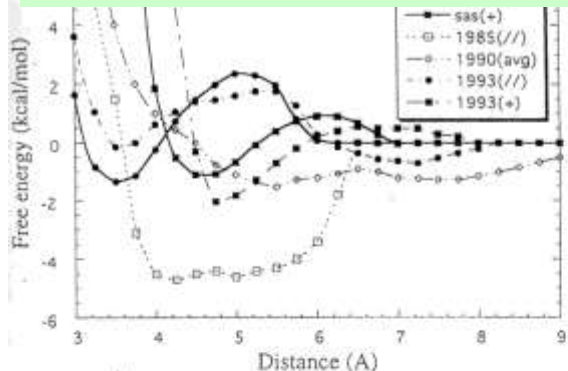
CH₄

CH₄

H₂O

CH₄

水中での分子間相互作用



... 視可物性子におい
て計算困難な課題で

(4) 分子間相互作用の 基礎理論

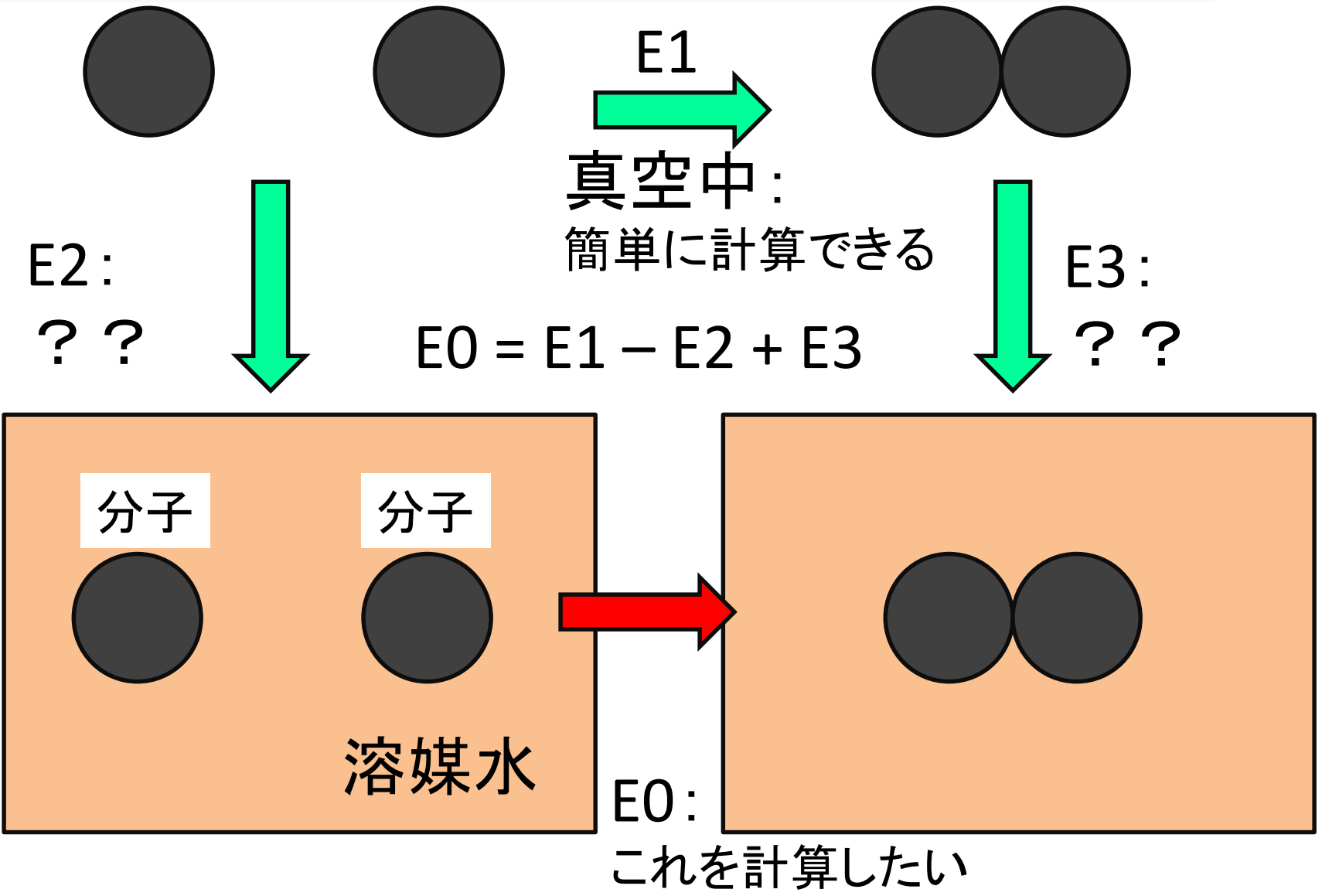
Scaled Particle Theory

タンパク質と化合物の疎水性相互作用は、統計物理学を適用することにより、解析的な近似式で記述できる。

一般に、疎水性相互作用は、その表面積(界面、accessible surface area; ASA)に比例する式で近似できる。

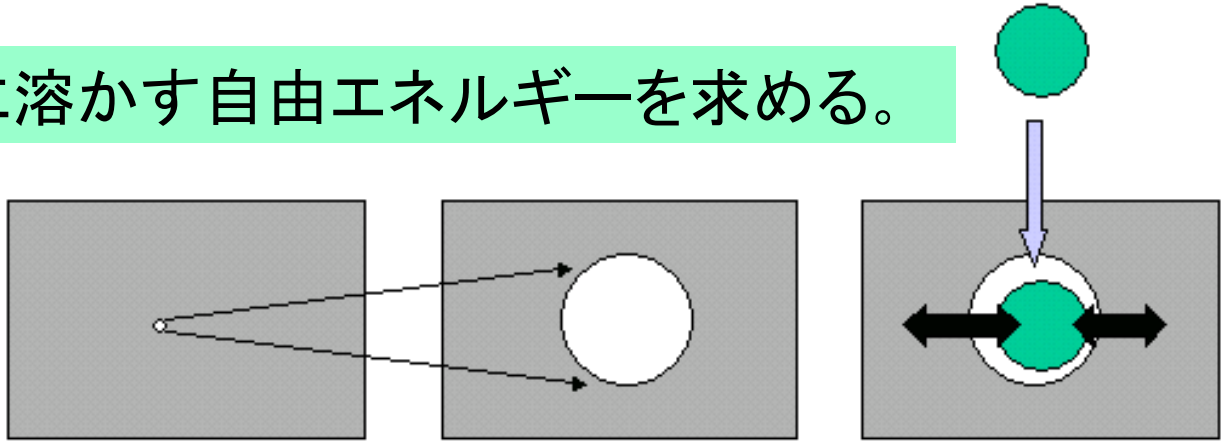
その基礎理論は、Scaled Particle Theoryという。

水中での分子間相互作用の経験的な計算の仕方: ASA法などの基礎



SPT: Scaled particle theoryによって、分子間のPMFは、分子の表面積で見積もることができる

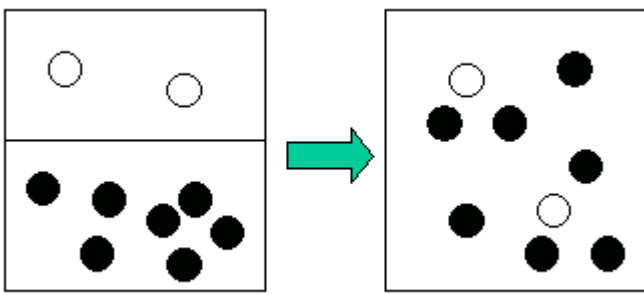
分子を水に溶かす自由エネルギーを求める。



Cavity formation

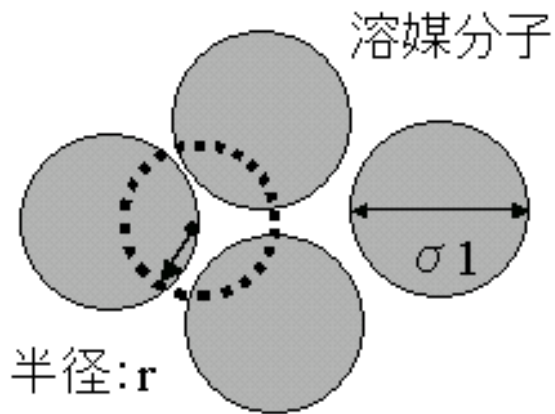
Solute-Solvent interaction

$$\Delta G = E_{cav} + E_{sol} + RT \ln \left(\frac{RT}{V_1} \right)$$



混合のエントロピー。
値は小さいので無視





溶媒分子は直径 σ_1 のhard-core potentialをもつとする。このとき、半径 $r()$ の領域に溶媒分子の核が見出されない確率 P_0 は溶媒の密度を ρ として

$$P_0 = 1 - \frac{4}{3} \pi r^2 \rho$$

である。すなわち、溶媒分子の原子核が半径 $r()$ の中に入れないような状態をつくるのに必要な自由エネルギー(仕事) W_0 は

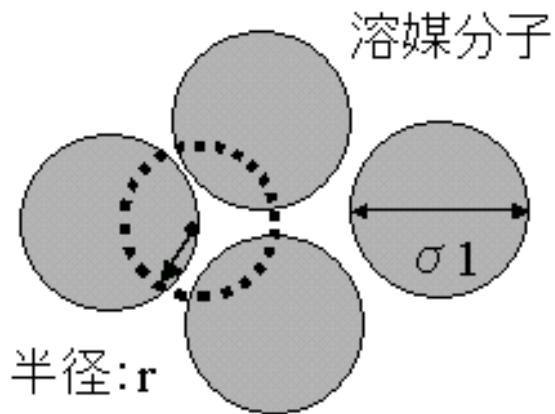
$$W_0 = kT \ln \left(1 - \frac{4}{3} \pi r^2 \rho \right)$$

半径 r のcavity formation energyを

$$W(r) = K_0 + K_1 r + K_2 r^2 + K_3 r^3$$

NPTでは圧力 P 、界面張力 γ の時、

$$W(r) = \frac{4}{3} \pi r^3 P + 4 \pi r^2 \gamma \left(1 - \frac{4\delta}{r} \right)$$



溶媒分子は直径 σ_1 のhard-core potentialをもつとする。このとき、半径 $r()$ の領域に溶媒分子の核が見出されない確率 P_0 は溶媒の密度を ρ として

$$P_0 = 1 - \frac{4}{3} \pi r^2 \rho$$

である。すなわち、溶媒分子の原子核が半径 $r()$ の中に入れないような状態をつくるのに必要な自由エネルギー(仕事) W_0 は

表面張力係数 * 表面積

$$W_0 = kT \ln \left(1 - \frac{4}{3} \pi r^2 \rho \right)$$

圧力 * 体積

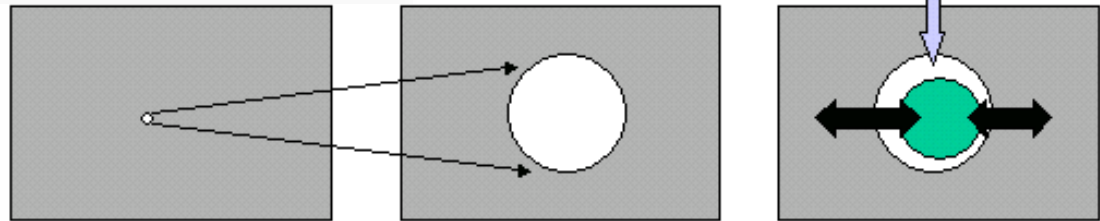
水における、点における隙間の体積
(= 1 - 空間重点率)

半径 r のcavity formation energy

$$W(r) = K_0 + K_1 r + K_2 r^2 + K_3 r^3$$

各種パラメータとメタンの 溶媒和自由エネルギー ΔG

$$\Delta G = E_{vdw} + E_{cav}$$



時代によってMDパラメータは変遷してきた。

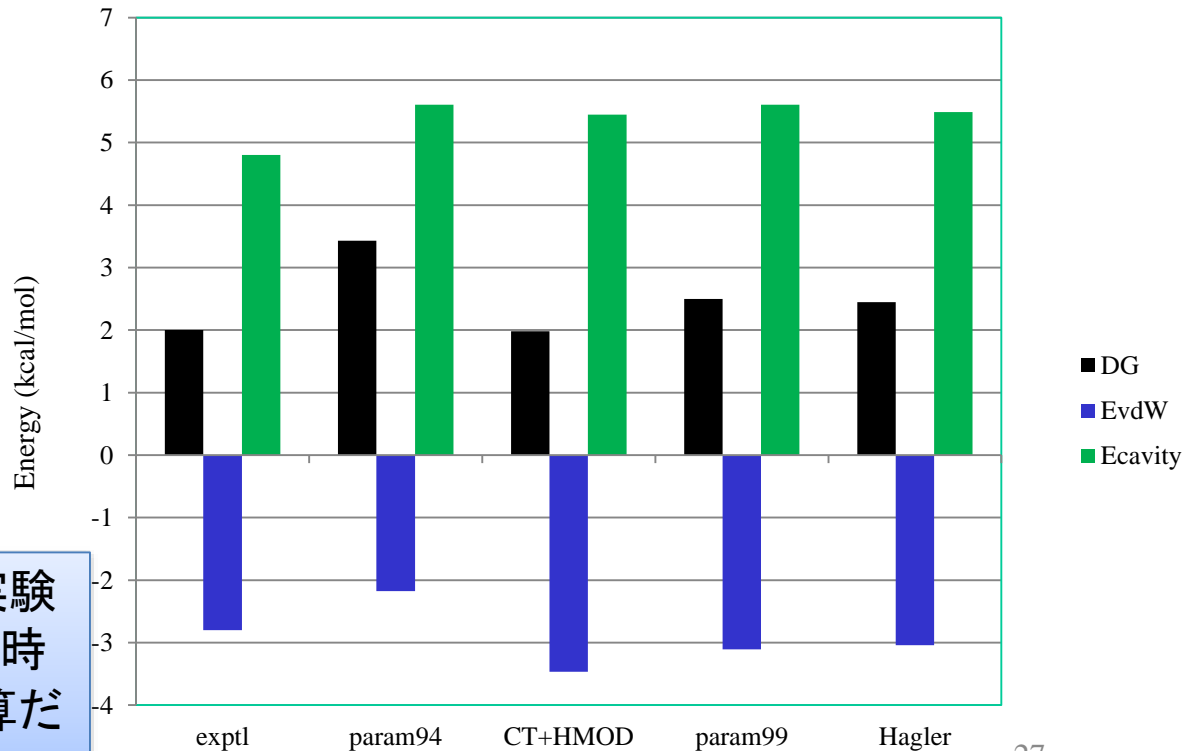
Cavity formation

Solute-Solvent interaction

MD計算による ΔG の値の変化は、SPTによって、分子シミュレーションの結果が再現される。

MD計算による ΔG の値のばらつきは、

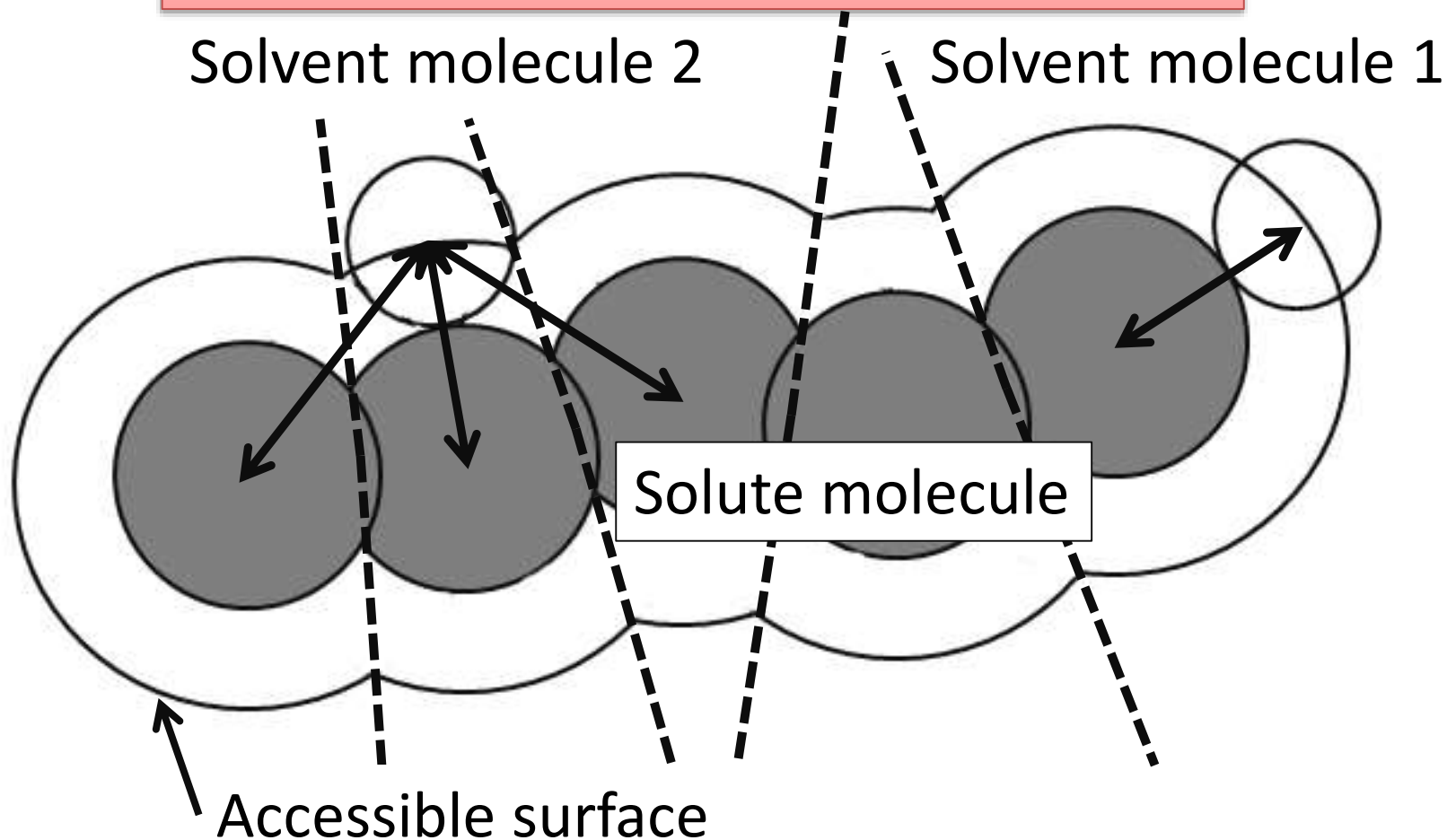
- ①vdW半径による空洞の生成エネルギーと、
- ②溶媒-溶質相互作用のバランスで決定される。



現在(AMBER parm99/gaff)では実験値の再現は可能だが、1992年当時のパラメーターでは不可能な計算だった。

分子の末端は、疎水的である

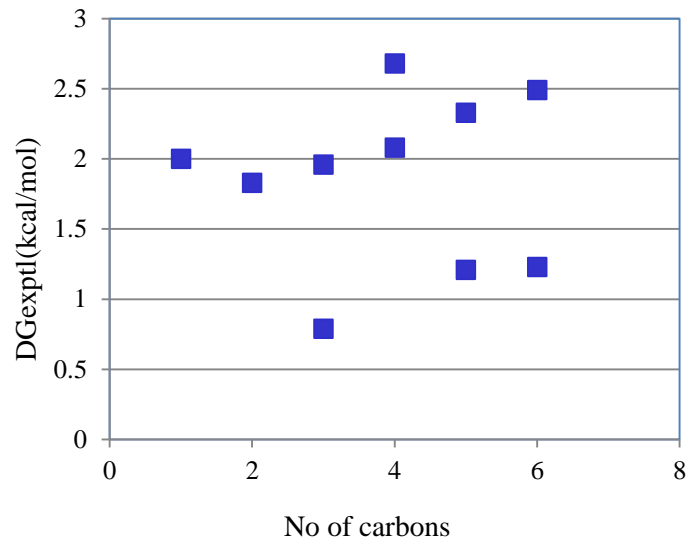
Cavity formationエネルギーがどこでも同じとしても、
分子の末端では、溶媒-溶質相互作用が弱い。
→ 疎水性が強くなる。



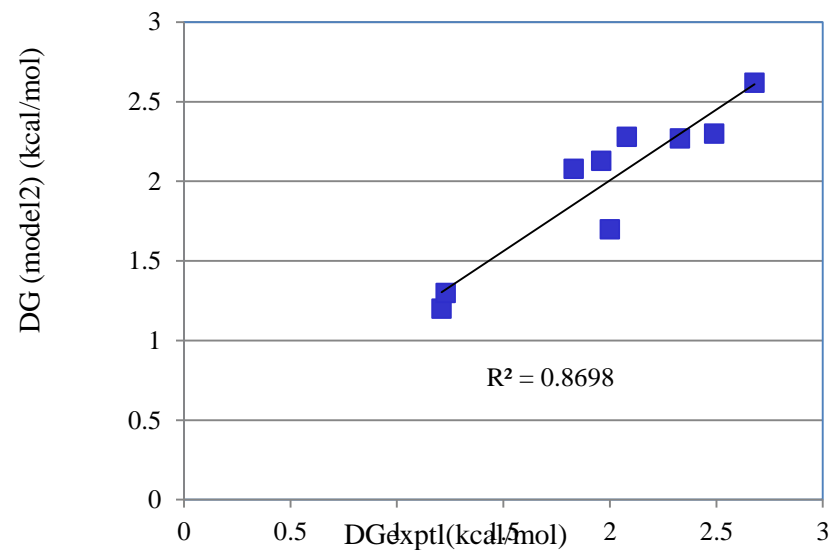
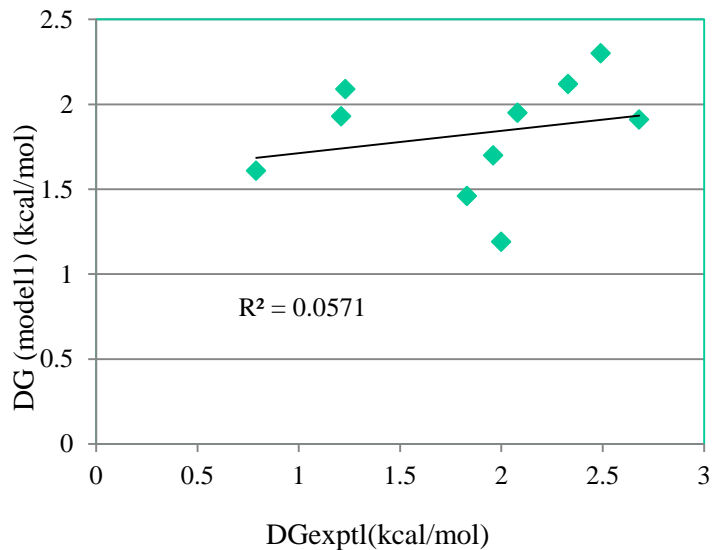
炭化水素の ΔG の例

ΔG と、炭素の数は、ほとんど関係がない

ASAと、炭素の数も、ほとんど関係がない。
($\sigma = 8.2 \text{ cal/mol/\AA}^2$)



CH_3 の $\sigma = 11.7 \text{ cal/mol/\AA}^2$ とし、その他の $\sigma = 5.1 \text{ cal/mol/\AA}^2$ と、末端の疎水性を高く扱うように、パラメーターを2つにすると、実験値が完全に再現できる。



Hofmeister effect

塩分濃度が高いと、疎水性相互作用は、強くなる。

$$\Delta G_{cav} = -k_B T \ln(P_{cav} - \rho_{ion} V_{ion}) = -k_B T \ln\left[P_{cav} \left(1 - \frac{\rho_{ion} V_{ion}}{P_{cav}}\right)\right]$$

$$\cong -k_B T \ln P_{cav} + \frac{\rho_{ion} V_{ion}}{P_{cav}}$$

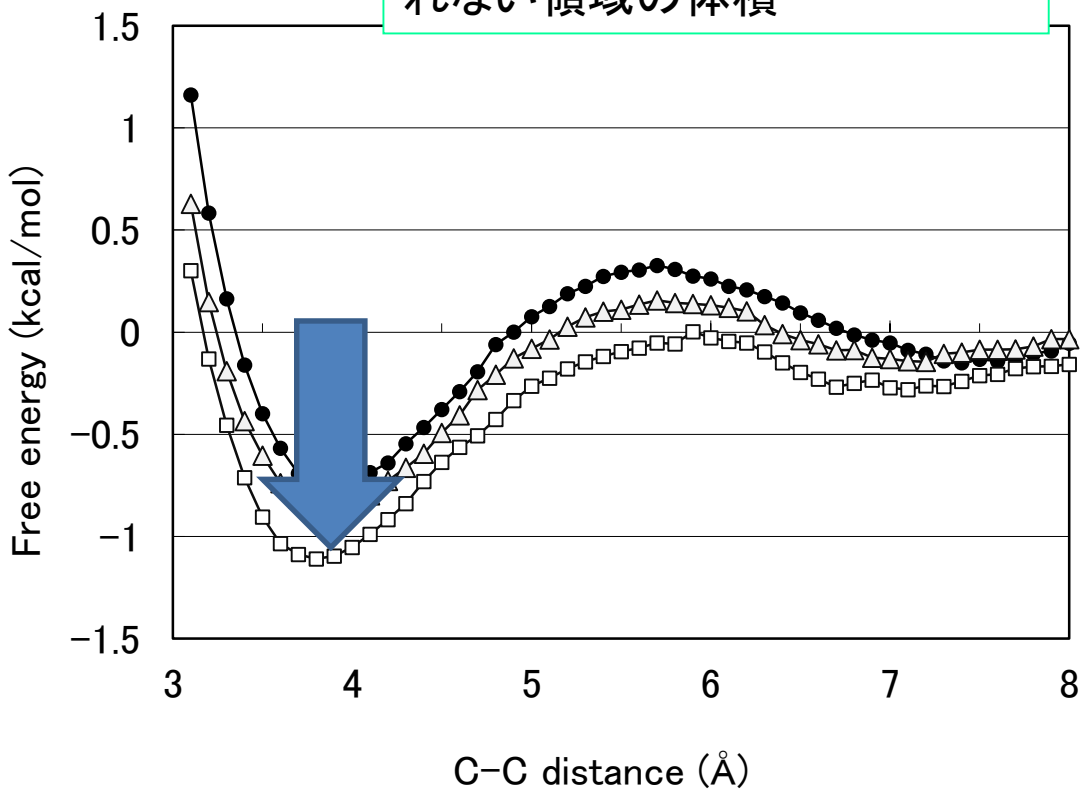
純水での疎水性相互作用

塩によって、相互作用が、強められる効果

塩の濃度

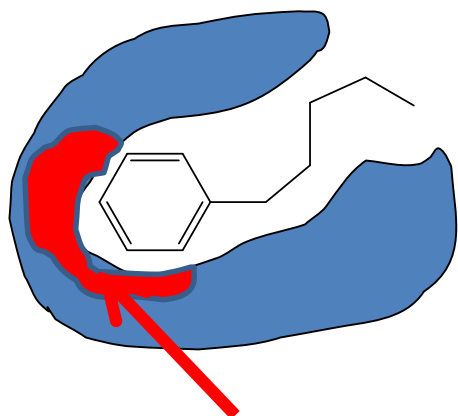
塩の周囲の、疎水性分子の入れない領域の体積

自由な水分子の少ない環境では、疎水性相互作用は強くなる。



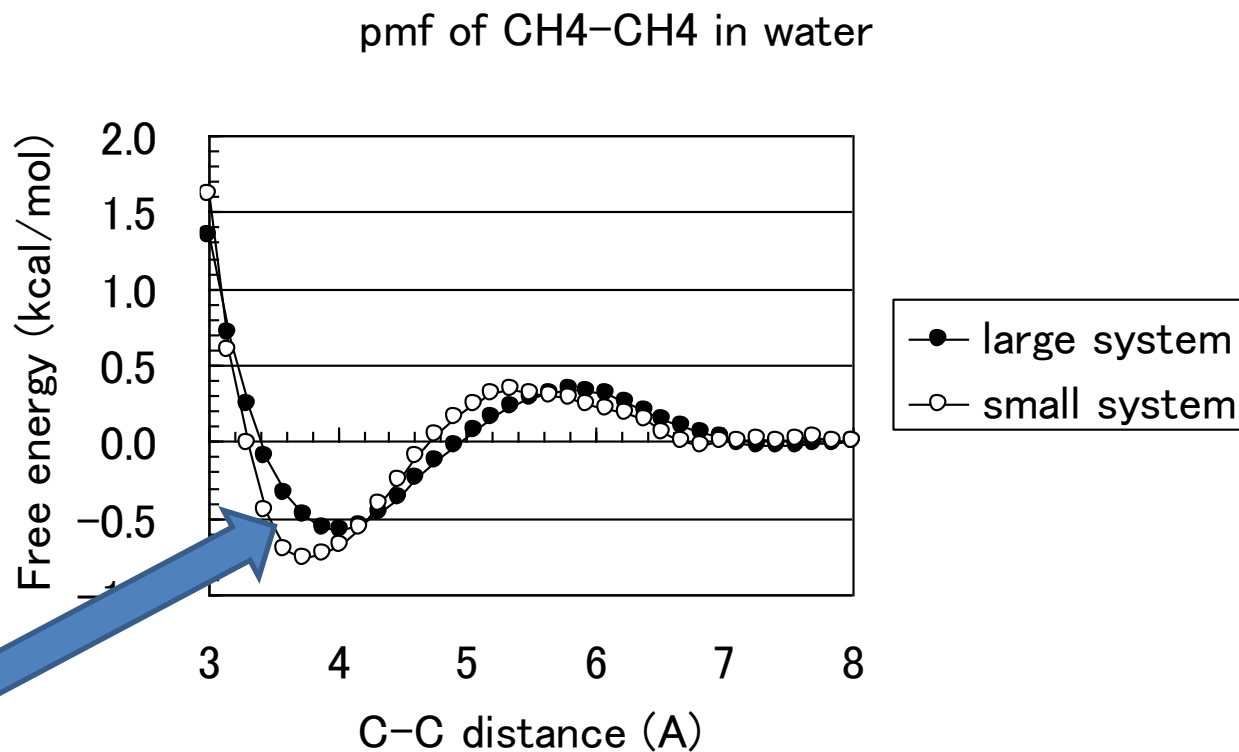
蛋白質のポケット内部では、自由な水が少なく、疎水性相互作用は強まる

自由な水が多いほど、疎水性相互作用は弱まる。
水の多い大きなセルと、水の少ない小さなセルで、メタンの相互作用を計算。



疎水性が強い

ドッキングでASA項
に考慮するとドッキ
ング結果改善



一辺18.64775Å, 214 水分子 + 2 メタン分子 (small system)
一辺21.7557Å, 341 水分子 + 2メタン分子 (large system)

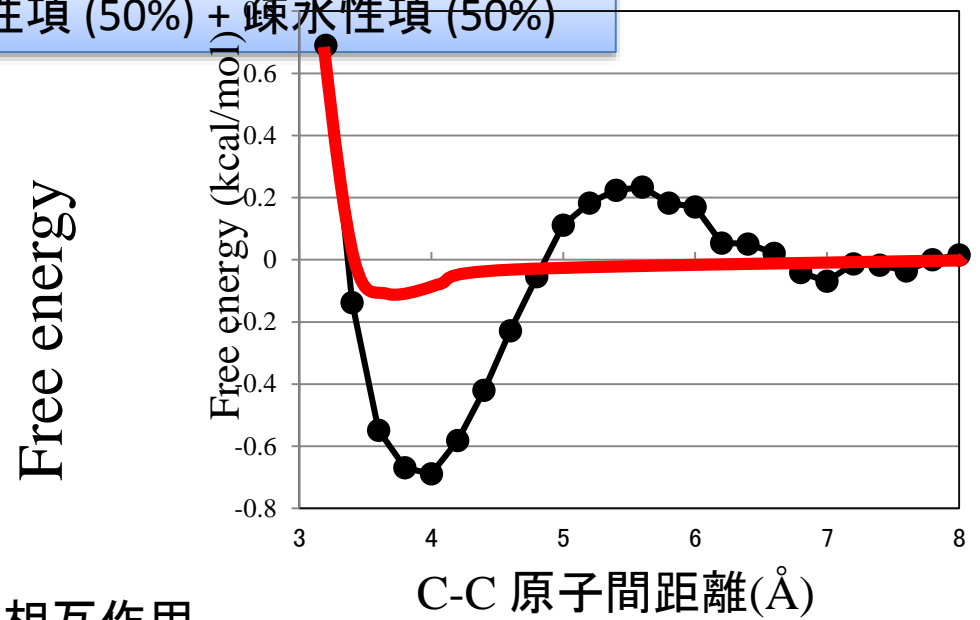
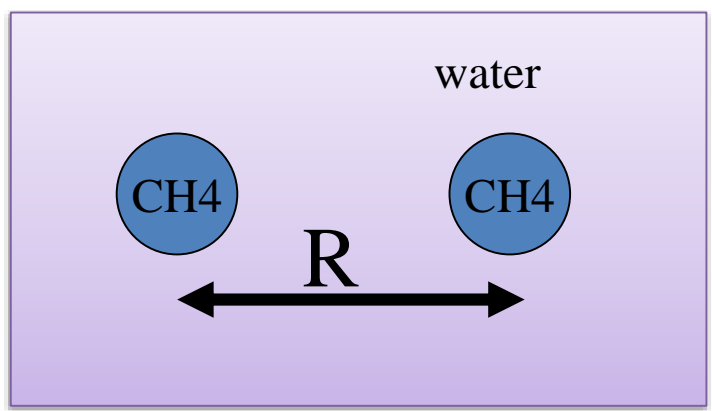
(3) 蛋白質-化合物相互作用での基本的な考え方

計算機スクリーニング・
ドッキング計算などでの扱い

タンパク質-化合物間の代表的な相互作用(ドッキング計算用)

- **van der Waals potential (ファンデアワールス力)**
- **Hydrophobic interaction (疎水相互作用)**
- Coulomb interaction
- Hydrogen bonding energy

タンパク質-薬物相互作用 = 親水性項 (50%) + 疎水性項 (50%)



Van der Waals 力の引力項 << 疎水性相互作用

疎水性相互作用 ~ **BSSE** of RHF/6-31G (真空中)

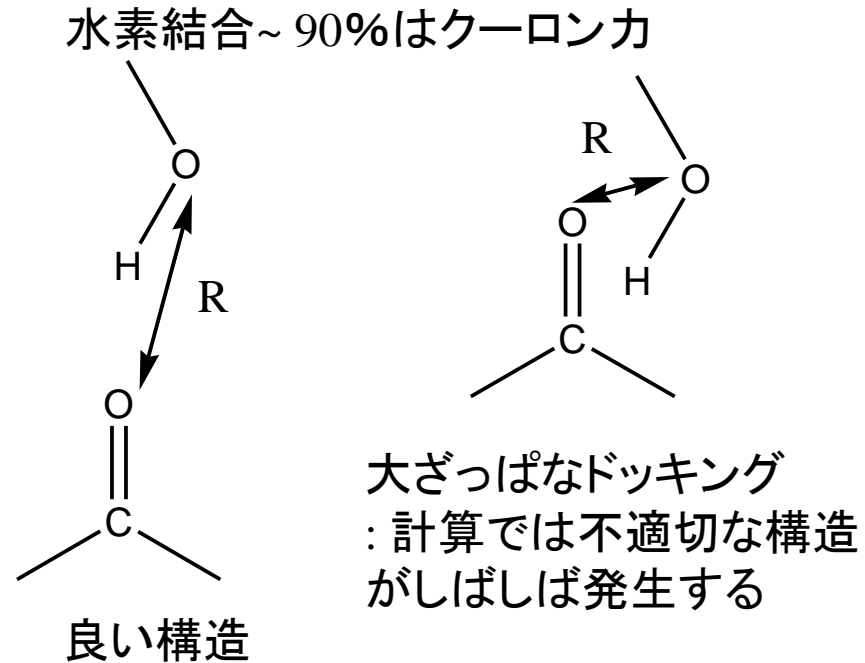
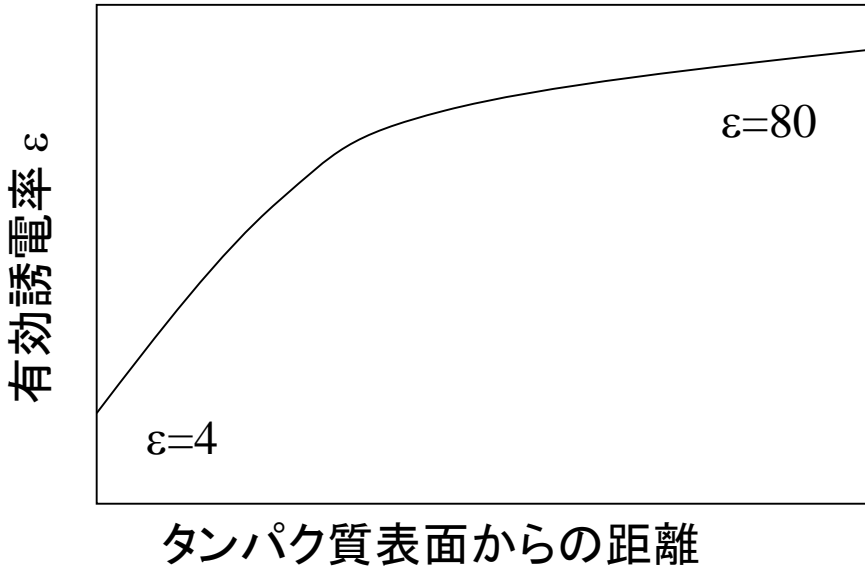
疎水性相互作用は、水のエンタルピー変化からきている

page 33

$$E(\text{疎水性相互作用}) = \text{定数} * (\text{Accessible Surface Area: 分子表面積})$$

タンパク質-化合物間の代表的な相互作用(ドッキング計算用)

- van der Waals potential
- Hydrophobic interaction(疎水性相互作用: 溶媒のエントロピー)
- **Coulomb interaction (静電相互作用、クーロン相互作用)**
- **Hydrogen bonding energy (水素結合)**
- 化合物内部のエントロピー
- 化合物内の相互作用: 無視する



$$E_c = 332 q_i * q_j / 4 R^2$$

有効誘電率 $\epsilon = 4 R$

水素結合の見積もりは大ざっぱに行う
 $E_h = \exp(-aR^2)$

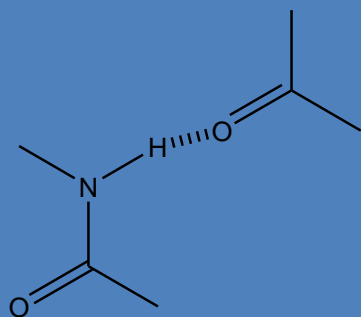
水中では、誘電率が大きい＝静電的相互作用は弱い
タンパク質に結合するとき、
水がタンパク質で排除されるので、誘電率は小さい＝静電的相互作用は大きくなる。

この効果は、誘電率を「原子間の距離に比例する」して簡単に近似する。

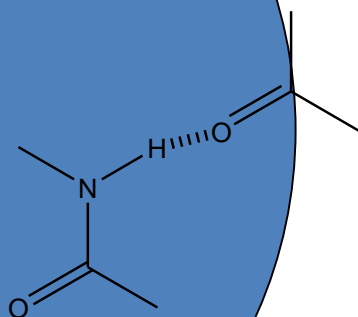
$$\text{静電気力} = \text{電荷1} * \text{電荷2} / (\text{距離} * \text{誘電率})$$

タンパク質
誘電率 = 2 ~ 4

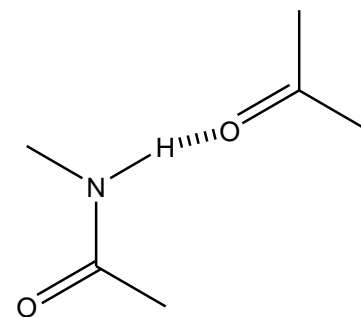
水
誘電率 = 78.5



水素結合は強力



水素結合はやや強い

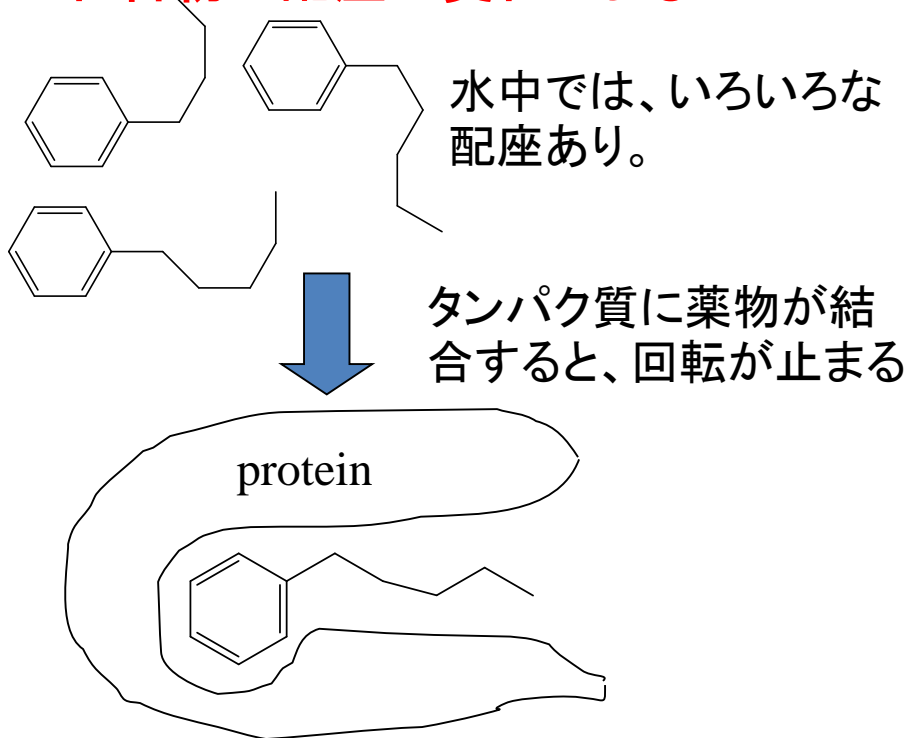


水素結合は弱い

タンパク質-化合物間の代表的な相互作用(ドッキング計算用)

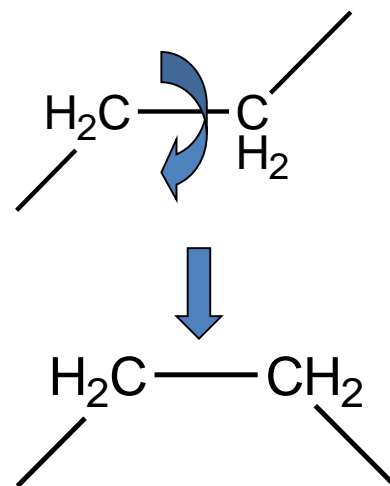
- van der Waals potential
- Hydrophobic interaction
- Coulomb interaction
- Hydrogen bonding energy

・化合物の配座の変化によるエントロピーの変化



エントロピー変化による自由エネルギー増大は: $E_r = -kT \log N_{\text{conf}}$

120度ずつ回転可能。
1結合あたり3つの可能性あり。



1重結合は、回転できるのでエントロピーあり。
タンパク質に薬物が結合すると、回転が止まるのでエントロピーが減少し、タンパク質との相互作用を弱める。

スコア関数（＝タンパク質－化合物間の相互作用エネルギー）

- 物理化学的スコア
クーロン力、VDW力、溶媒和自由エネルギーなど、物理化学的計算
- 統計的スコア
X線結晶構造(PDBやCCDC)をもとに、
どの各官能基同士が隣接しやすいか空間分布を計算する。

→ 最近はやりのAI (Convolution Neural Network)が使われる例
古典的なベイズ統計法を、AIでの繰り返し学習に置き換えたもの。
- 経験的スコア
水素結合や、疎水性原子の接触数を数え上げる。

DOCK: 物理化学的スコア

DOCKは最も古い.

DOCKでのリガンド構造探索は、グラフトポロジカルな手法が用いられている.

タンパク質と薬物の相互作用は、静電相互作用と、原子の接触で見積もる.

$$E = \sum_{i=1}^{ligand} \sum_{j=1}^{receptor} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332 \frac{q_j q_i}{D r_{ij}} \right)$$

- vdW力はLJ12-6ではなくLJ9-6など、反発力の小さなものを用いる.
- DOCK ver4よりリガンドをflexibleにし、accessible surface area(ASA)の計算を導入している. タンパク質の構造は変化しない.

AutoDock

GRID用：探索用のスコア関数。蛋白質-リガンド間（vdW力,水素結合,静電相互作用）とリガンド分子内のエネルギーを考慮する。蛋白質内のエネルギー変化はない。

$$\Delta E = \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i,j} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}}$$

$$+ [\sum_{\text{torsions}} K_{\phi} (1 + \cos(n\phi) - \delta) + \Delta H_{vdW}^{\text{ligand}} + \Delta H_{elec}^{\text{ligand}} + \Delta H_{hbond}^{\text{ligand}}]$$

評価用のスコア関数：蛋白質-リガンド間（vdW力,水素結合,静電相互作用）とリガンド分子内のエネルギーを考慮する。水素結合はLJ12-10型に加えて、水素結合の数を数えて補正する。リガンド分子内部のエネルギーは考慮しない。ASAに相当する項を近似的に見積もって加算する（末尾fsol）

$$\Delta G = f_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + f_{hbond} \left(\sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + E_{hbond} \right)$$

$$+ f_{elec} \sum_{i,j} \left(\frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) + \Delta G_{tor} N_{tor} + f_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2 / 2\sigma^2)}$$

$$E_{hbond} = (n_A + n_D - n_{OH}) E_{hb}$$

PRO-LEAD

タンパク質と薬物間の水素結合や、疎水性原子の接触の数を数え上げる。

経験的スコア関数を用いている。

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} \sum_{il} g_1(\Delta r) g_2(\Delta \alpha) +$$

$$\Delta G_{metal} \sum_{aM} f(r_{aM}) + \Delta G_{lipo} \sum_{iL} f(r_{iL}) + \Delta G_{rot} H_{rot}$$

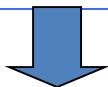
$$g_1(\Delta r) = \begin{cases} 1 & \dots & \Delta r \leq 0.25 \text{ \AA} \\ 1 - \frac{\Delta r - 0.25}{Du - 0.25} & \dots & 0.25 \text{ \AA} < \Delta r \leq Du \\ 0 & \dots & Du < \Delta r \end{cases}$$

$$g_2(\Delta \alpha) = \begin{cases} 1 & \dots & \Delta \alpha \leq 30^\circ \\ 1 - \frac{\Delta \alpha - 30}{Au - 30} & \dots & 30^\circ < \Delta \alpha \leq Au \\ 0 & \dots & Au < \Delta \alpha \end{cases}$$

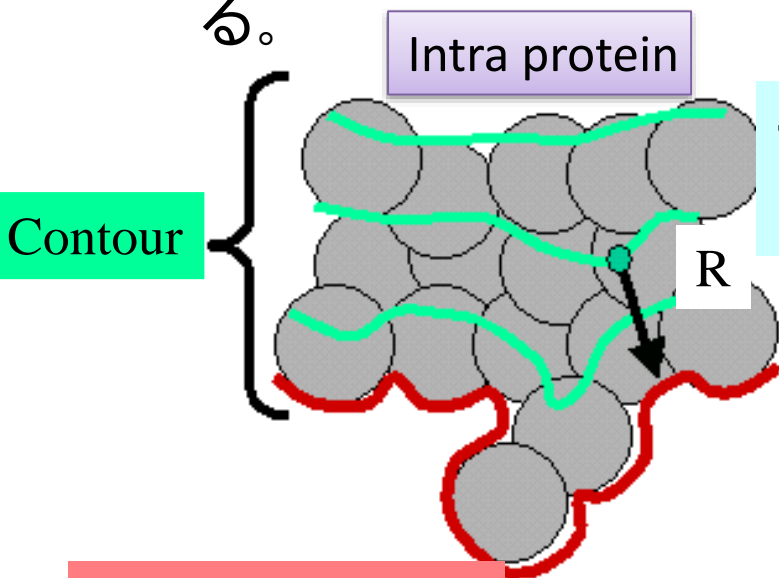
水素結合は、全てのリガンド-蛋白質間の水素結合をなしうる対に対し計算する。Duは、水素結合とみなされる最大の原子間距離（3 - 4 Å）。Δαは、O-H.....O/N-H.....Oなどの角度。

Sievgene:なめらかなポテンシャル面の形成

リガンドをタンパク質表面にラフに配置する。



タンパク質の原子とリガンド原子が重なる。



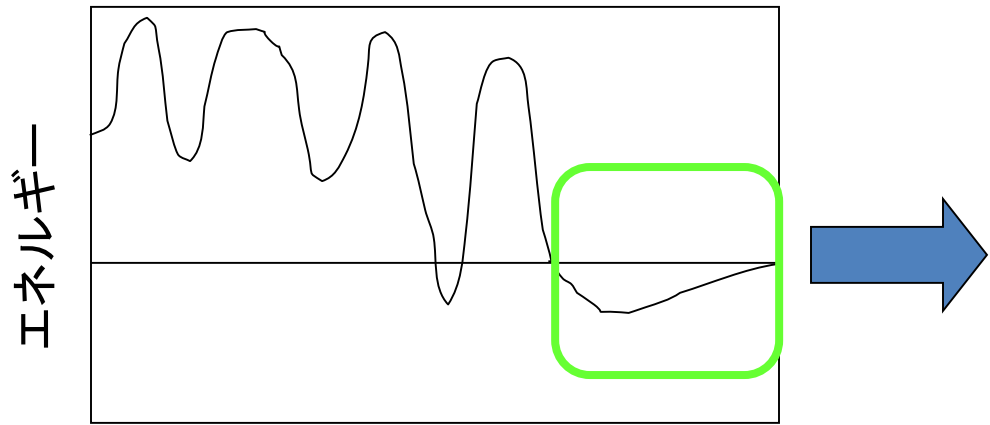
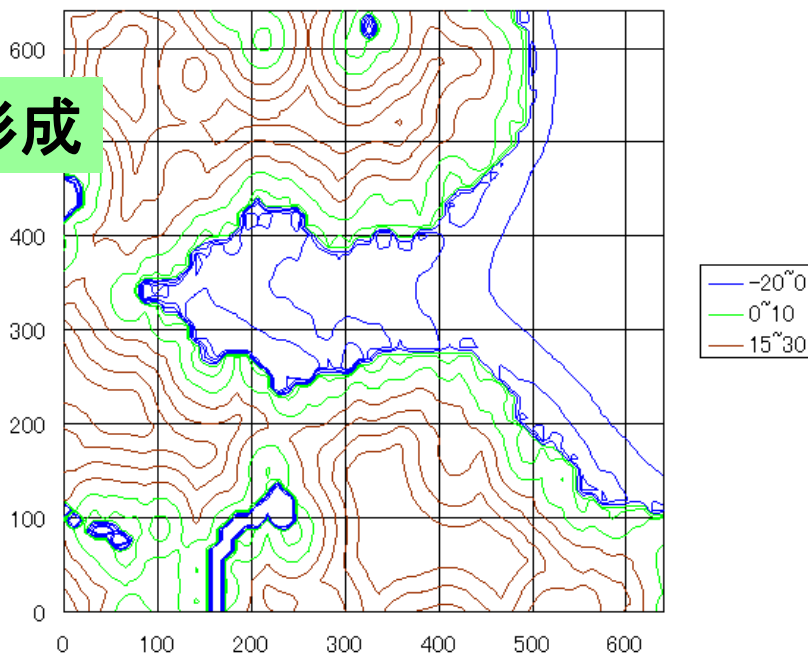
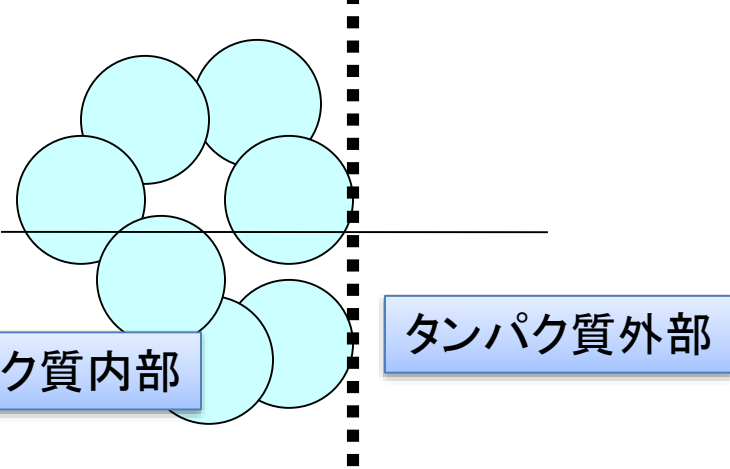
ラフな原子の配置が可能なノイズに強いポテンシャルの作成が必要



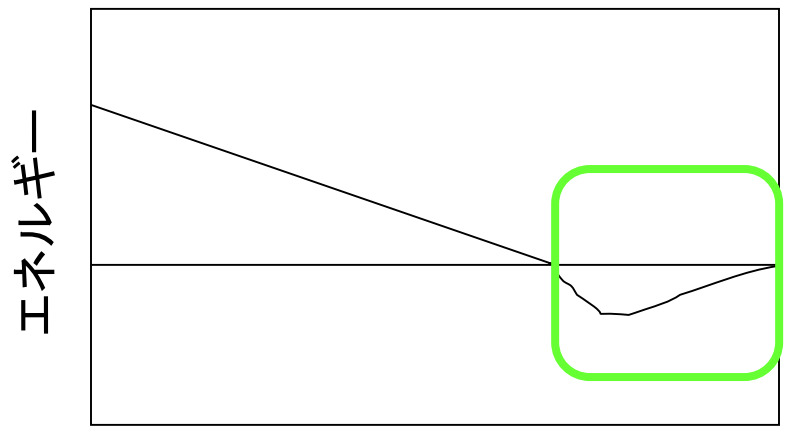
タンパク質領域の外には、普通の力場を適用する。
タンパク質内部は弾性体で近似

R : accessible surfaceまでの距離

Sievgene:なめらかなポテンシャル面の形成



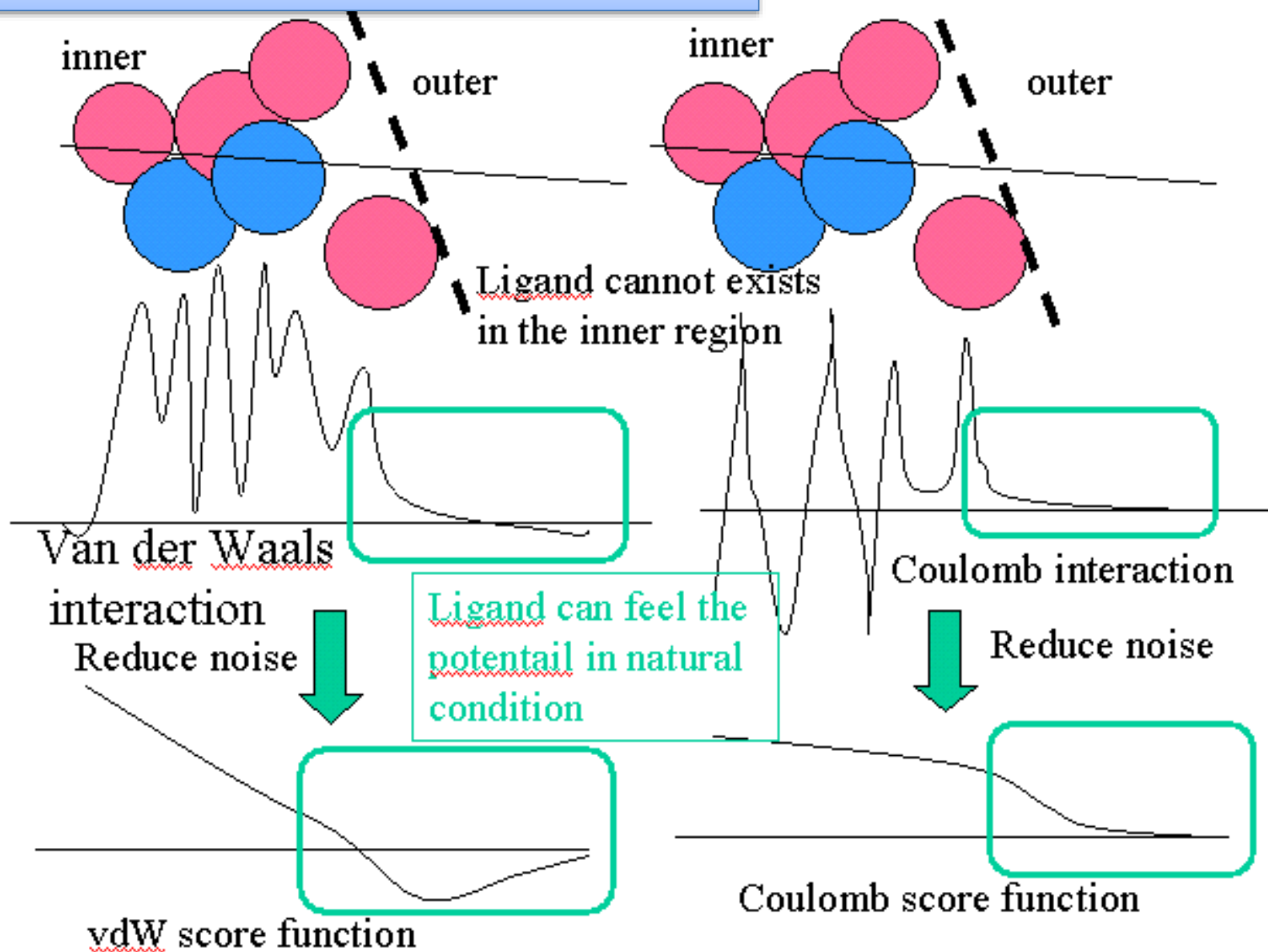
通常のポテンシャル



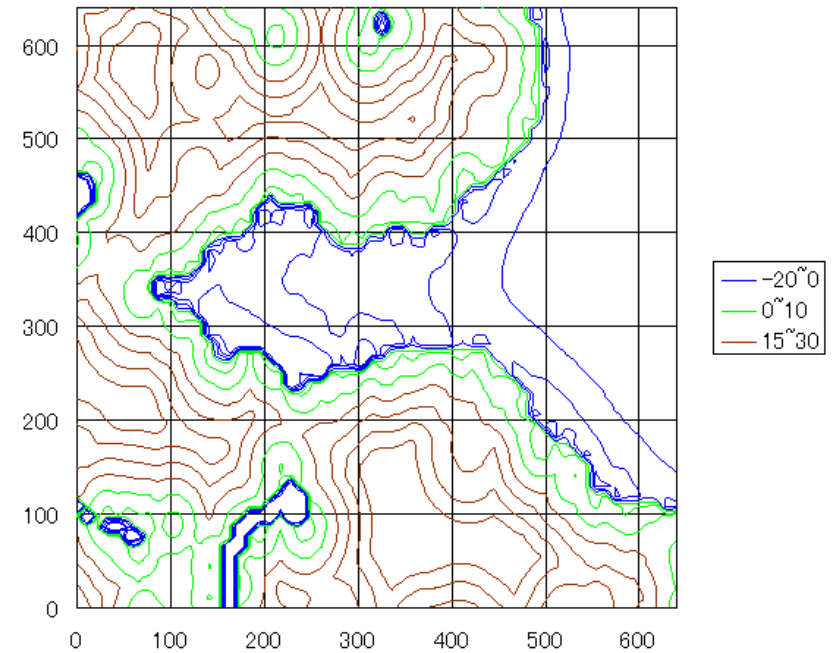
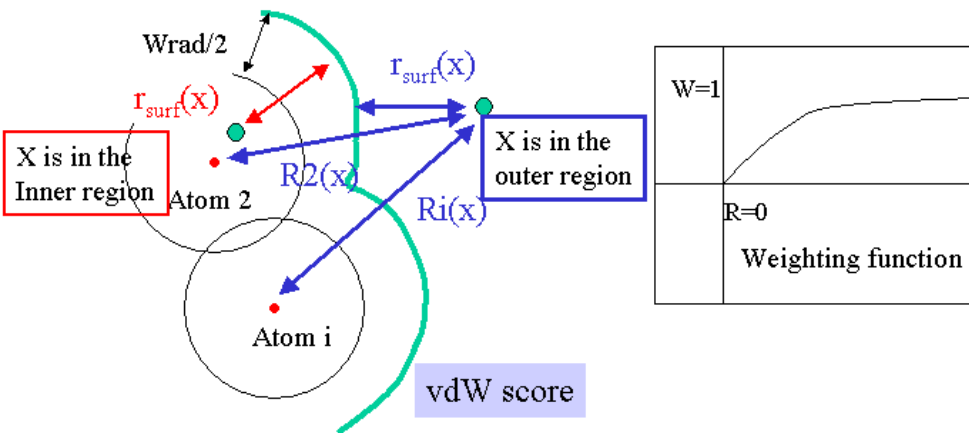
平滑化されたポテンシャル

タンパク質領域内部には、vdW原子間反発を弱めるように、タンパク質表面までの距離に依存した新しい力場を適用する。

Sievgene力場 (score) 関数の作成



Sievegene : Van der Waals相互作用

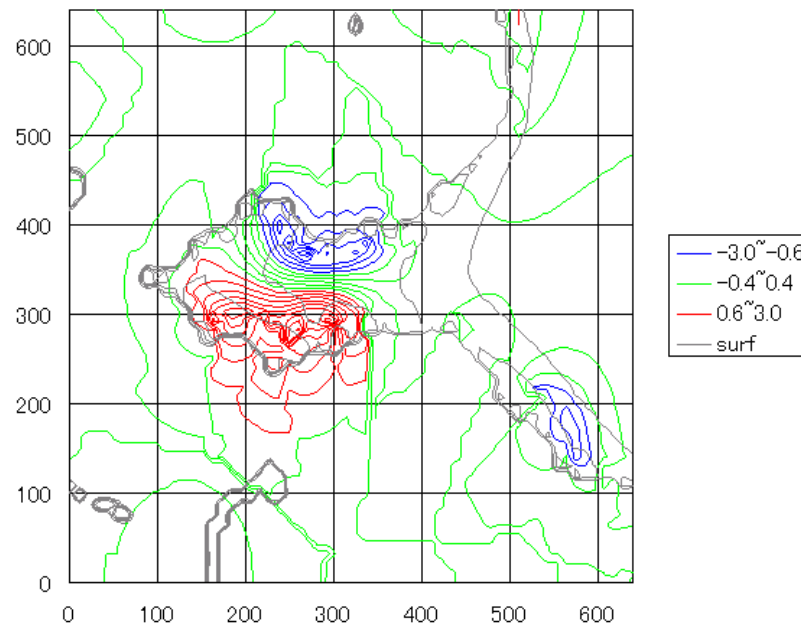
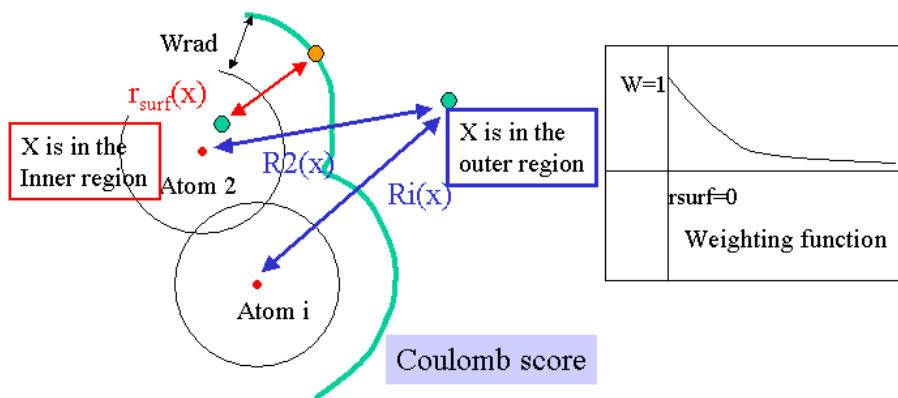


$$score(x) = w(r_{surf}(x)) \sum_i \left\{ \frac{A}{R_i(x)^{12}} - \frac{B}{R_i(x)^6} \right\} : x \in outer$$

$$score(x) = 10.0 r_{surf}(x) : x \in inner$$

$$w(r) = 1 - e^{-Cr}, C = 2.0$$

Sievgene: 静電相互作用



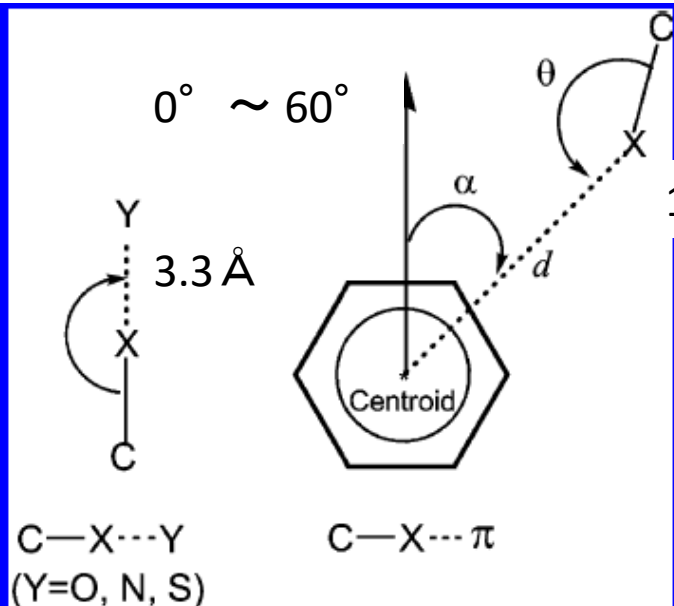
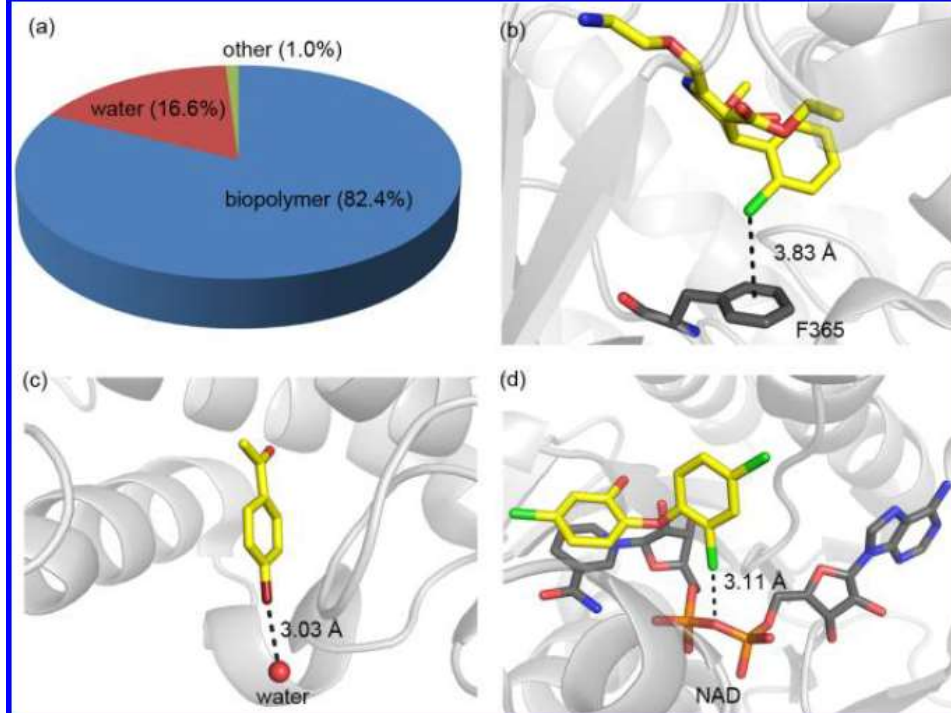
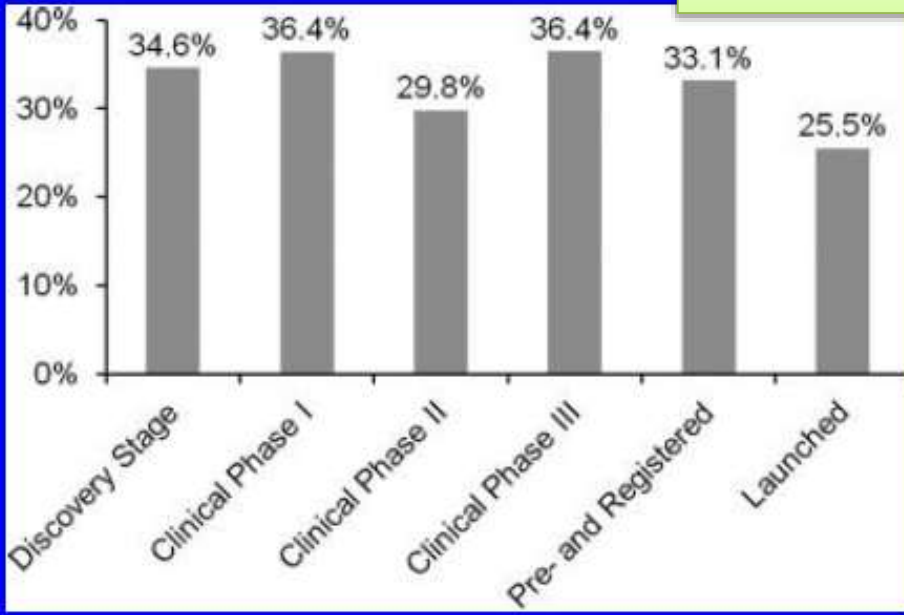
$$score(x) = \sum_i \frac{332q_i}{4R(x)_i^2} : x \in outer$$

$$score(x) = w(r_{ASA}(x)) \sum_i \frac{332q_i}{4R(ASA)_i^2} : x \in inner$$

$$w(R) = \frac{1}{1+R}$$

where q_i is charge of i -th atom of protein

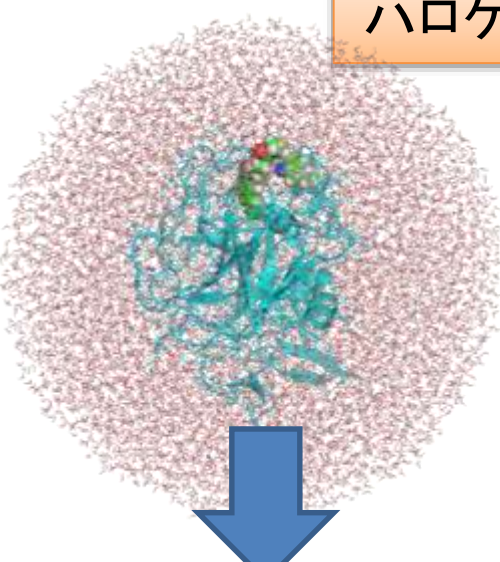
ハロゲン原子は、医薬品の3割が含有している



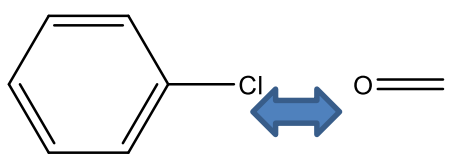
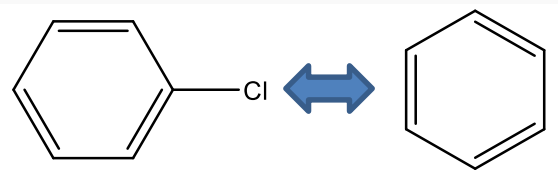
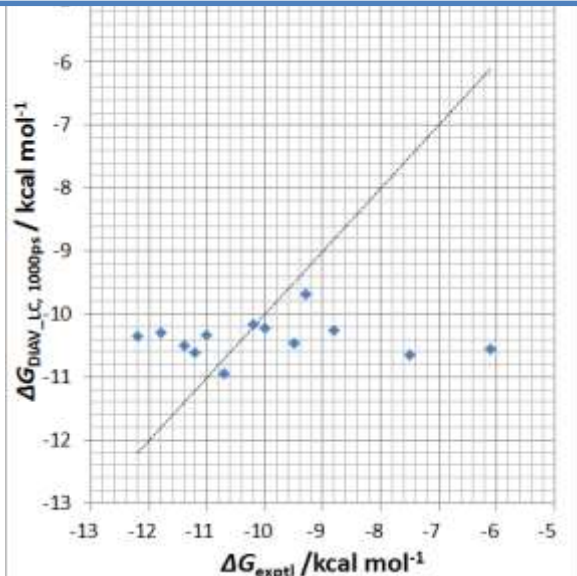
ハロゲン (Cl, Br, I) は、マイナス電荷の O/N と結合する

ハロゲン (Cl, Br, I) は、 π 電子と結合する

ハロゲン原子モデルの改良: Extra-charge point model (OPLS3 など)



通常の古典力場では結合エネルギーをまったく再現しない



ハロゲン (負電荷) 酸素 (負電荷)

マイナス電荷を帯びているハロゲンは、電氣的に反発するはずのマイナス電荷の原子と結合しやすい。

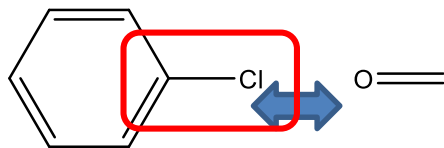
QM:
B3LYP以上 + 基底関数をtriple以上



ハロゲンを2つの点電荷(+と-)で表現する。

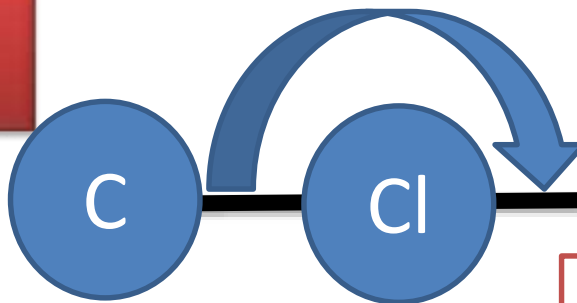


ハロゲン原子モデル: 一見、うまく
いったかに見えるが



ハロゲン
(負電荷)

酸素
(負電荷)



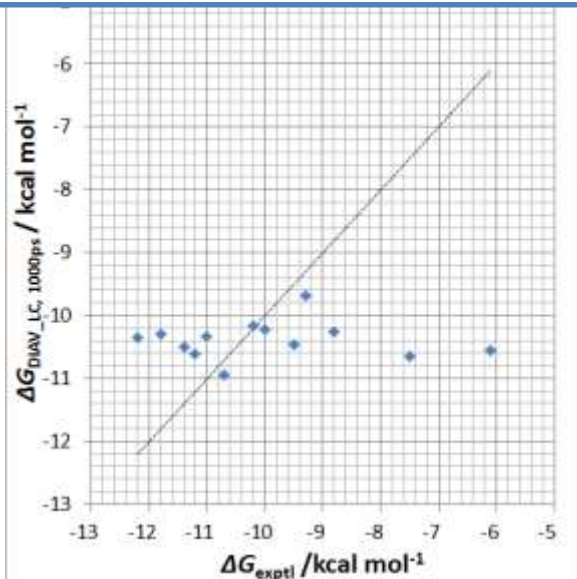
180° 直線状に3原子

3原子のなす面が
定義できない

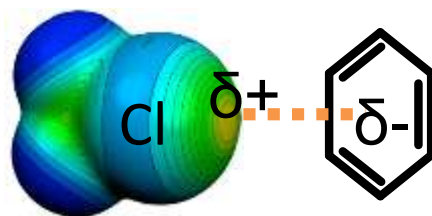
計算技術上、ゼロ割発生

3点が直線に乗る分子計算は計算でき
ない(手法はあるが時間かかる)

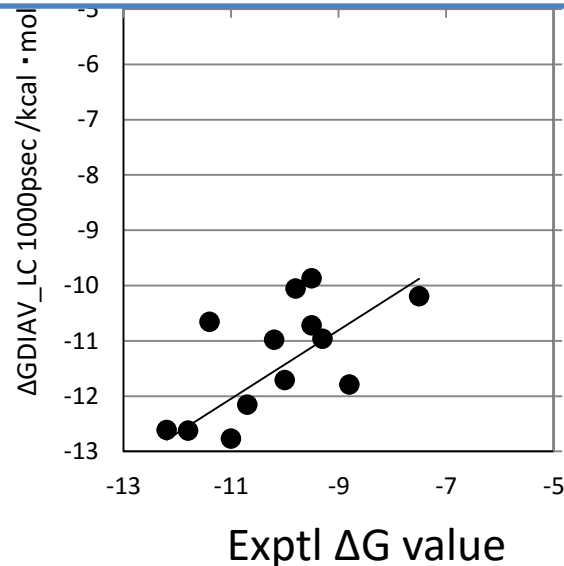
通常の古典力場では結合エ
ネルギーをまったく再現しない



ハロゲンを2つの点電荷(+
と-)で表現する。



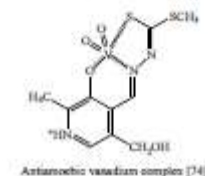
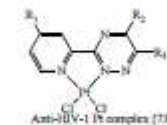
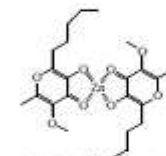
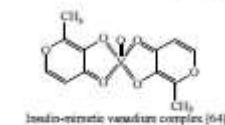
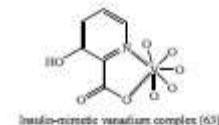
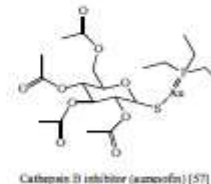
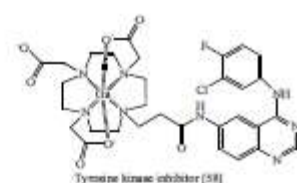
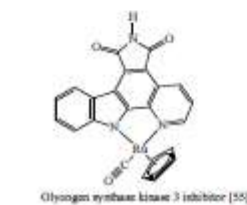
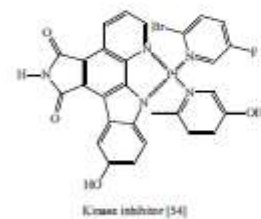
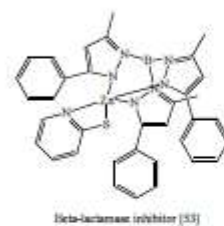
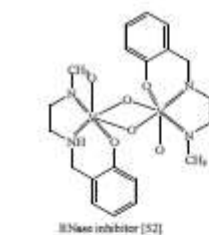
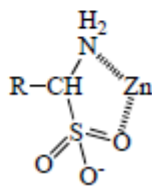
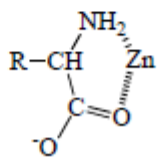
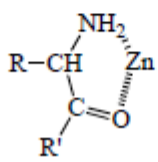
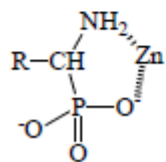
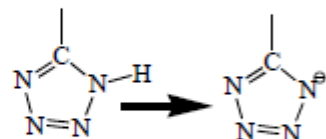
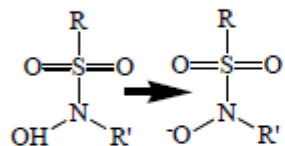
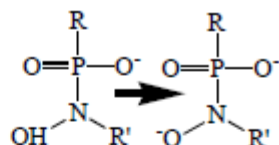
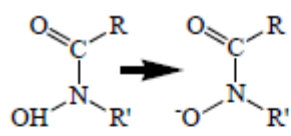
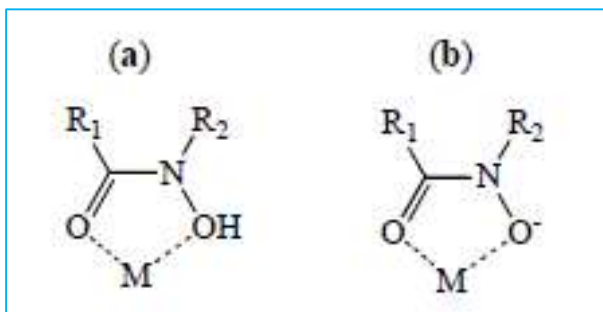
新しい2点電荷モデルでは結
合エネルギーを再現できる



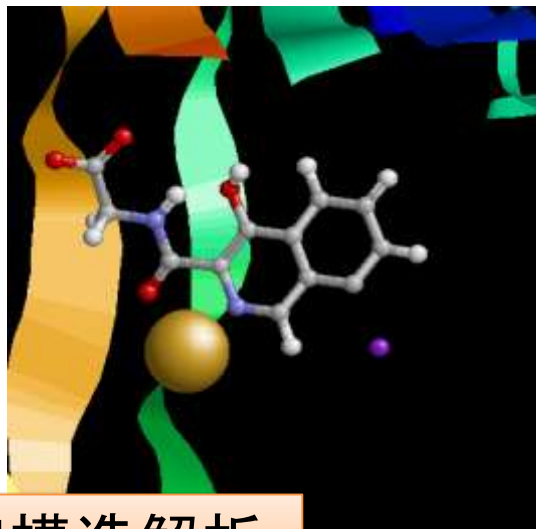
化合物原子のローンペア電子モデル化による金属酵素のモデリング改善

金属に結合する分子のパターン(1) 金属酵素への医薬品の配位

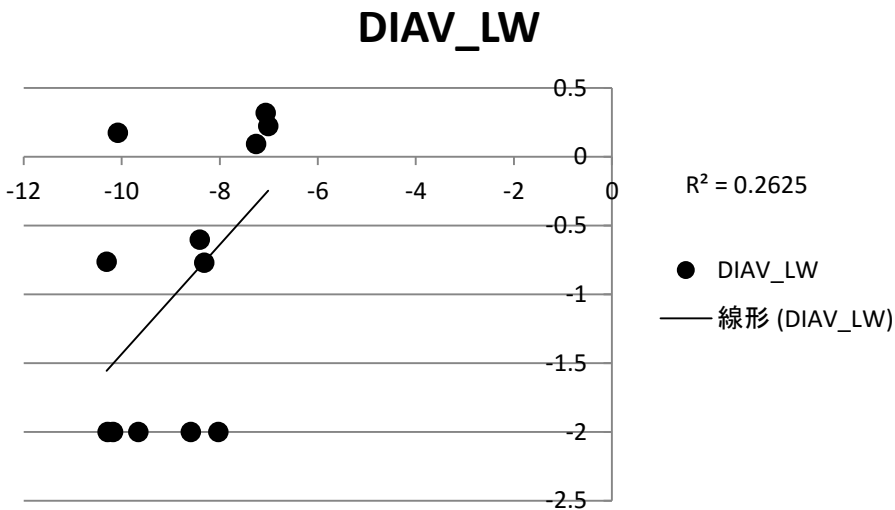
金属に結合する分子のパターン(2) 医薬品に見られる金属錯体



金属の扱いで本質的に困難なこと： 酵素の30%に見られる



X線構造解析



分子シミュレーション

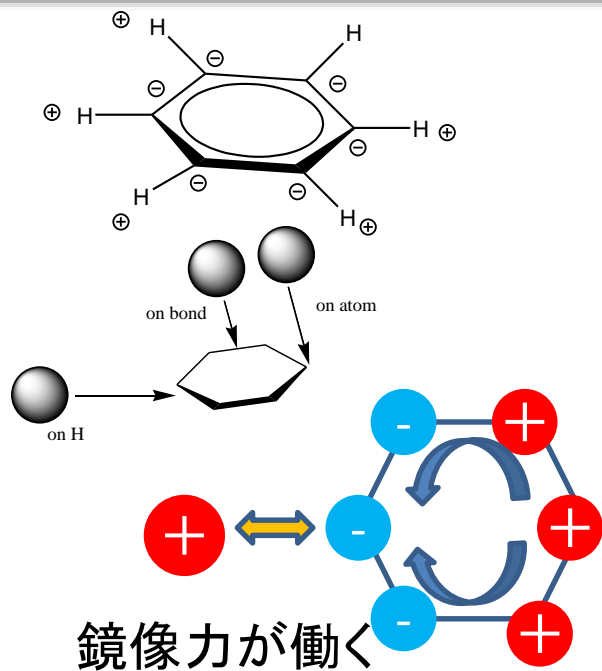
人体の酵素(創薬標的)の30%は金属酵素。しかし通常
の分子シミュレーションで再現不能(スパコン京でも無理)

芳香族のリングカレント

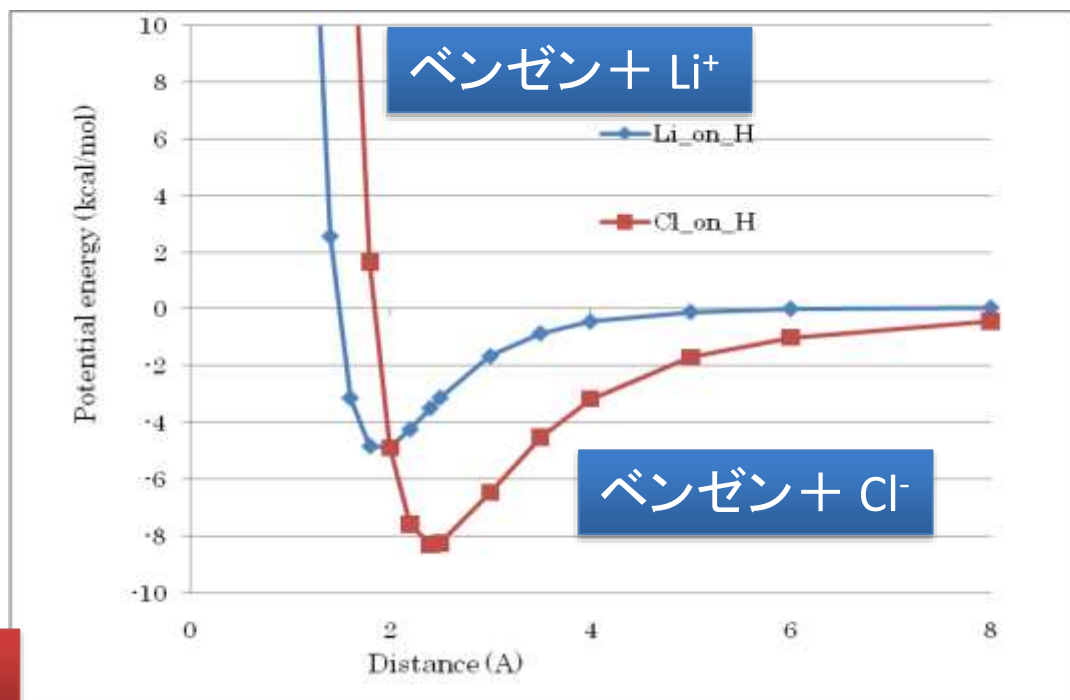
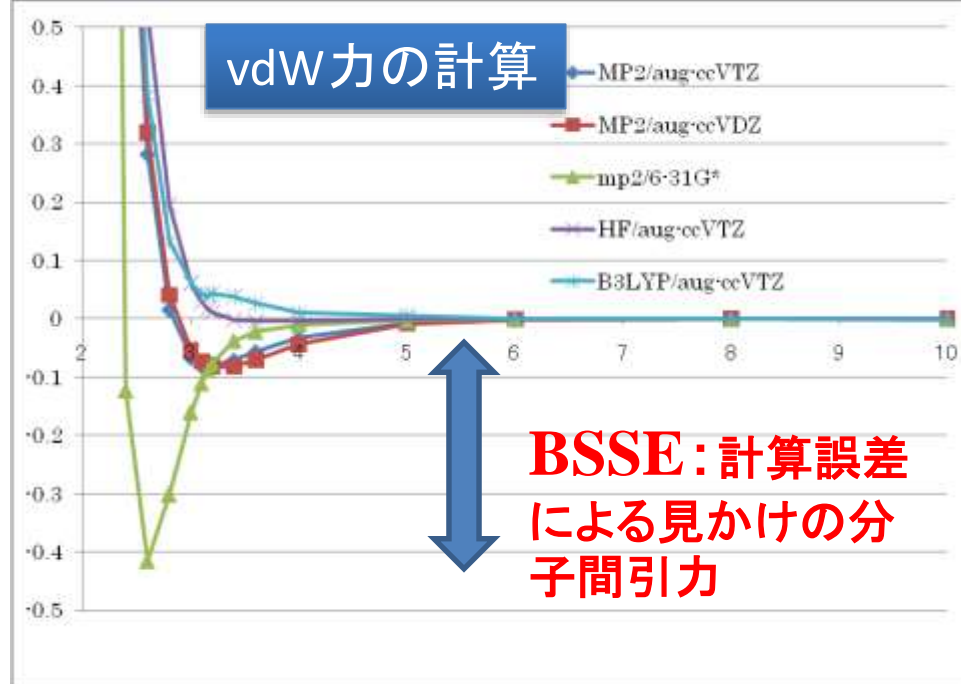
(1) 量子化学計算での条件検討

MP2/aug-cc-pVDZ (6-311++G**) 以上

(2) ベンゼンは、疎水性分子も、 +電荷も、-電荷も引き付ける



電荷平衡法やinduced-dipole法



ハロゲン結合、Lone pair電子の考慮は、X線構造での ΔG 予測を改善
しかし、スクリーニングヒット率は低下

確度の高いモデルが、良い結果を与えるわけではない。
データの準備の良さに応じた、ロバストな計算手法を使う必要がある

高精度・良

Induced fittingなどの無視
(原子間衝突の発生)
水分子の位置の見積もり違いなどが、
精密なモデルほど大きな誤差になる

生物学的な環境、
アンサンブルの解釈の違い、
古典と量子の違いなどが、
精密なモデルほど大きな誤差になる

エネルギーなどの精度

低精度・悪

データの準備が悪い
(スクリーニングレベル)

データの準備が良い
(X線結晶レベル・丁寧に準備した
MD計算結果を使う場合)

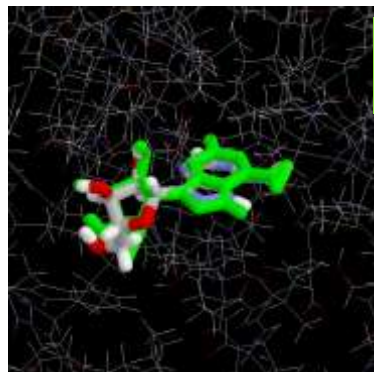
おおざっぱなモデル

精密モデル

モデルの確度(力場、スコア関数、計算手法などの精密さ)

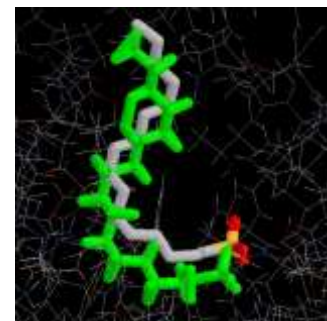
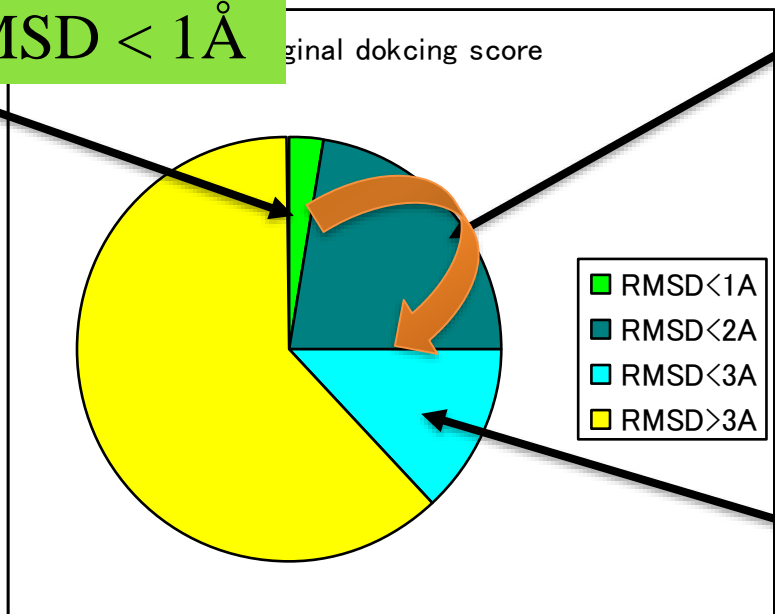
1.ドッキングソフトの精度

構造の予測精度: 25%

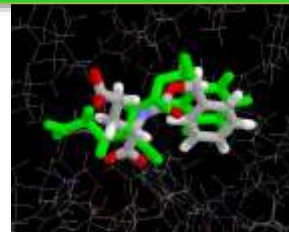


RMSD < 1Å

original docking score



RMSD < 2Å



RMSD < 3Å

2.活性予測の精度

誤差 1~3 kcal/mol

手法と情報の精度に依存

3.薬物スクリーニングソフトの精度

ヒット率 3~5%程度

硬いタンパク質に限る

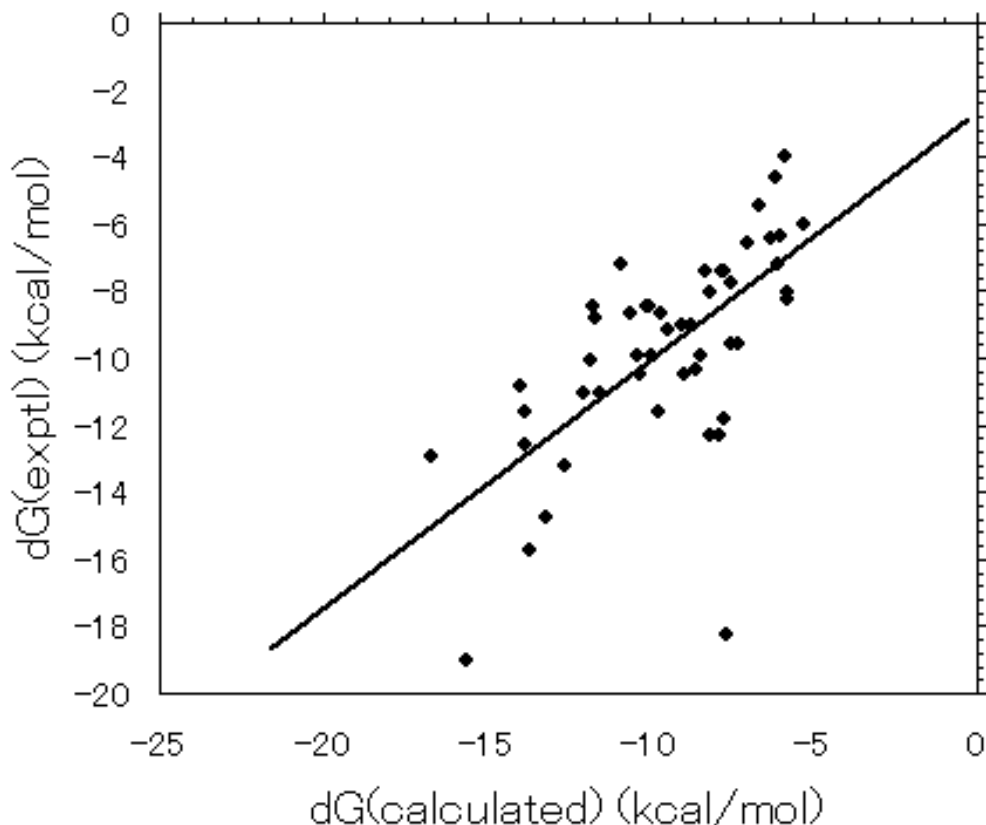
4.タンパク質薬物結合ポケット予測ソフトの精度

80%程度

天然の基質が存在する場合

ドッキング計算によるタンパク質-化合物結合自由エネルギーの再現性

平均誤差は 2.5kcal/mol. 実験と計算の相関係数は: 0.66.



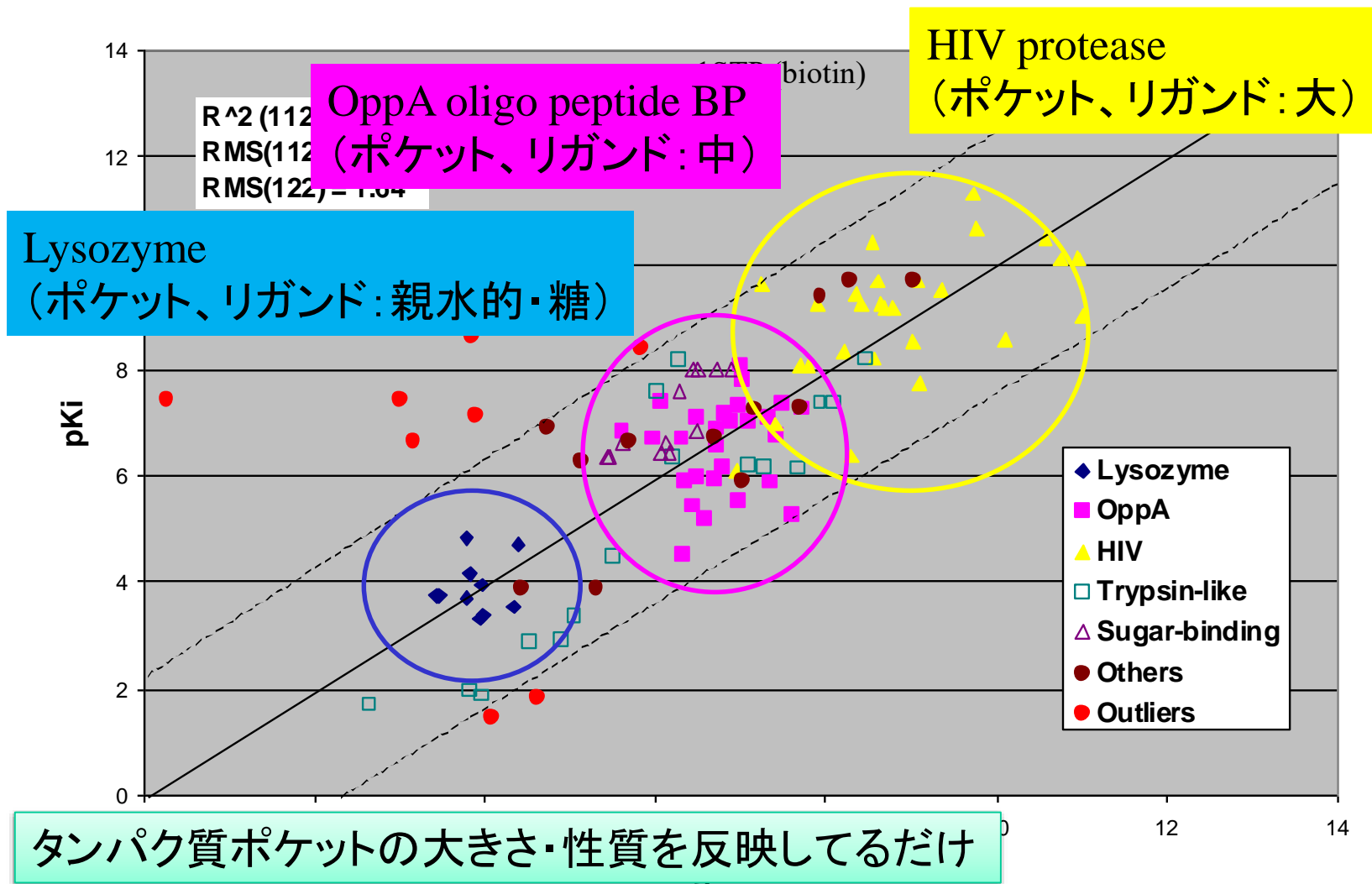
ヒット化合物の結合エネルギーは、
3~8kcal/mol程度。
リード化合物では、結合エネルギーは、
~12 kcal/molに高める。

【開発目標】
ヒット化合物の IC_{50} は、 $1\mu M$ 前後。
市販薬の IC_{50} は、平均 30nM。
70%の薬は50nMより活性が強い。

平均誤差で、2.5 kcal/molとは、実験で薬物の相互作用が数百倍異ならないと、見分けがつかないということ。

活性値(結合自由エネルギー)は再現するか(アクセラリスから拝借:LigandFit)

LigScore2 Exact CFF



構造は再現できるが、活性値は再現が困難=>実験と組み合わせることを前提

化合物の準備

Sievgeneでは、物理化学的スコア関数系を用いる。
＝化合物の3D座標と原子電荷が必要。

また電荷は、水中での支配的イオン形が必要。
金属に配位する場合は、配位形が必要。

通常、ISIS mol (SDF) 又は Sybyl mol2 file formatで準備する。

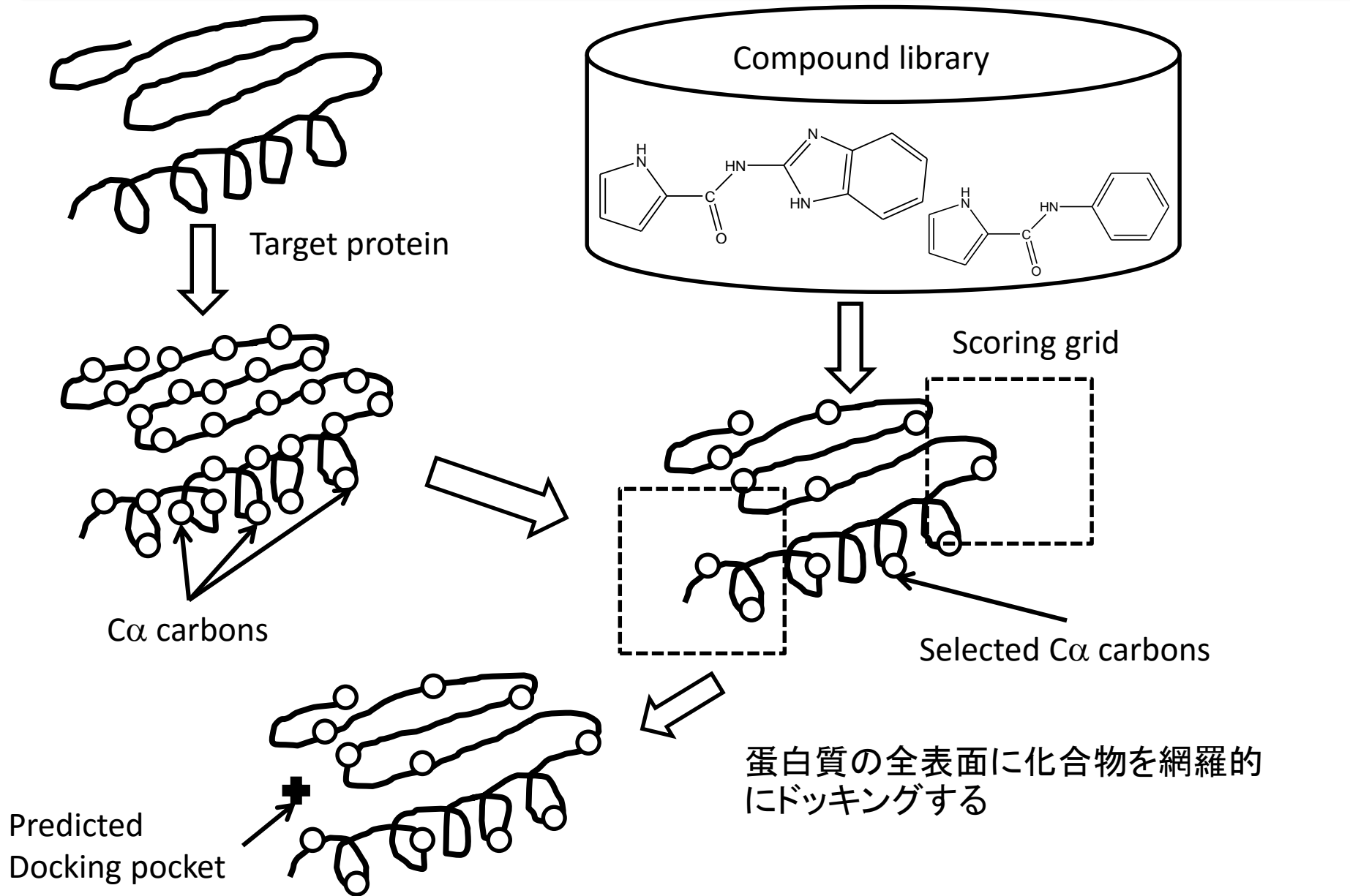
原子電荷の準備方法

- ① Gasteiger電荷
- ② MOPAC AM1電荷 = LigandBoxでの標準
- ③ RHF/6-31G* RESP charge = 正確なドッキングを行いたい場合。

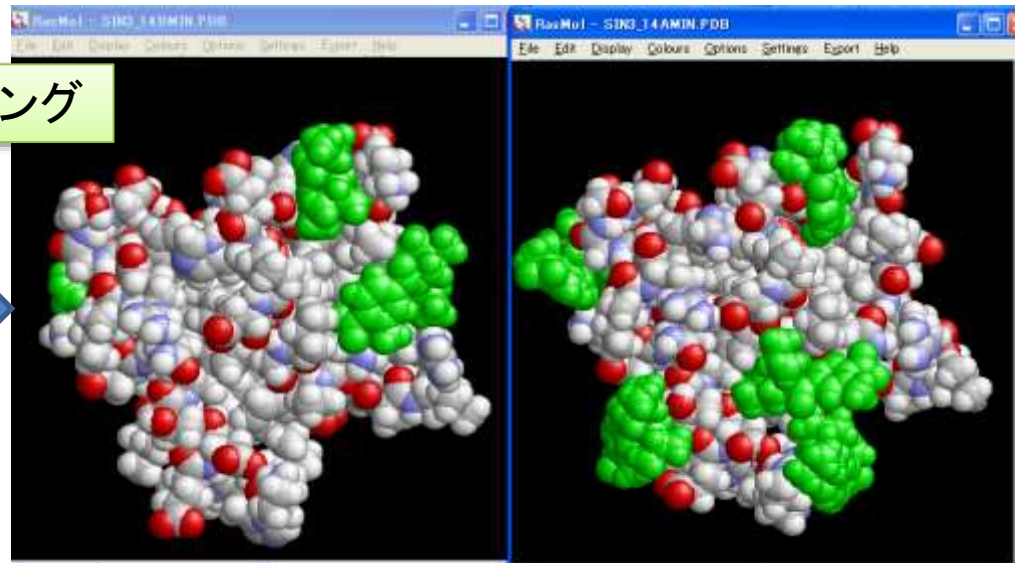
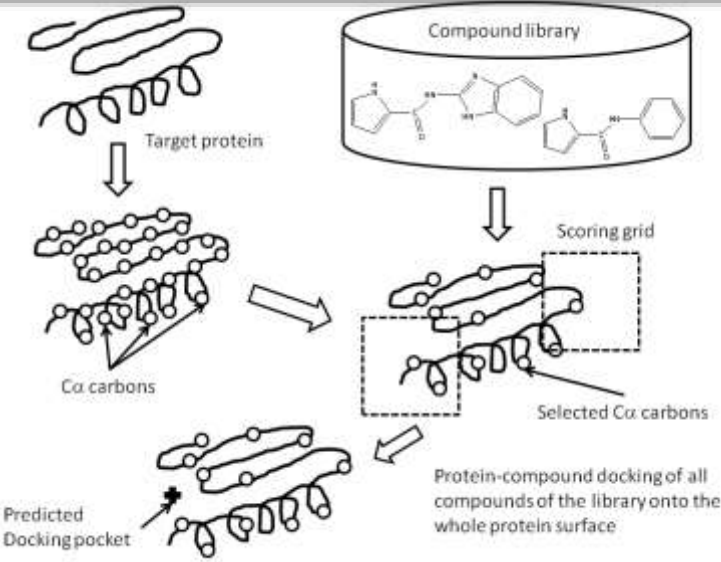
ドッキング結果は、① < ② < ③で、正確なほど良い。

ただし、ドッキング結果には、①－③で数%の違いしか見られない。

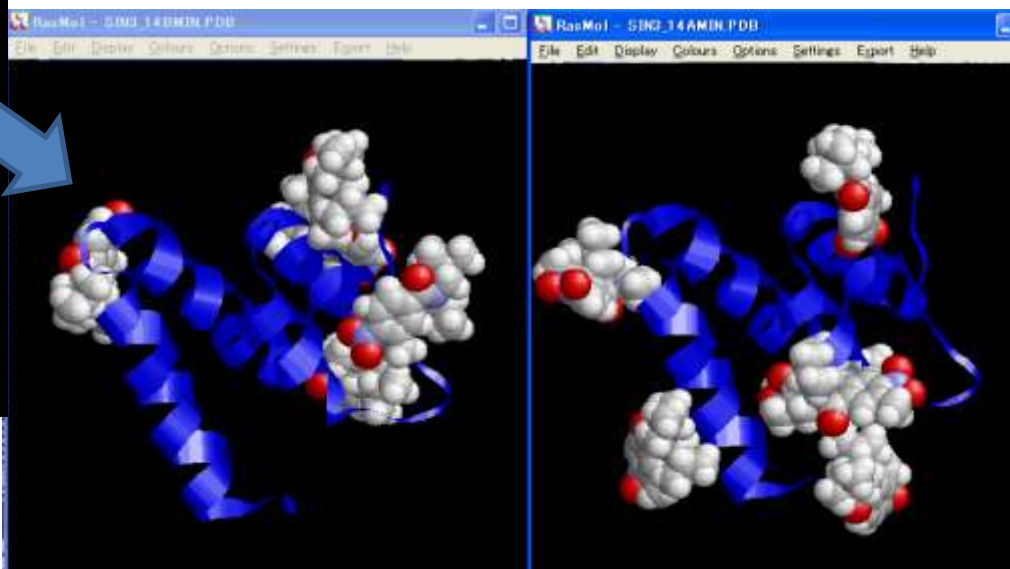
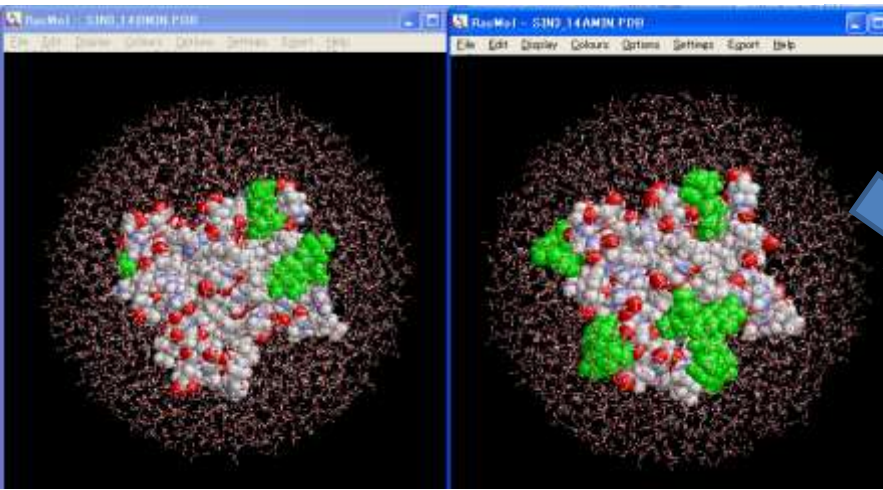
ポケットの分からない標的タンパク質 : MolSite法でのポケット探索と薬物探索



活性化化合物をタンパク質の全表面にドッキング



生理食塩水につける



水中でのMD計算を開始

分子設計に必要なこと

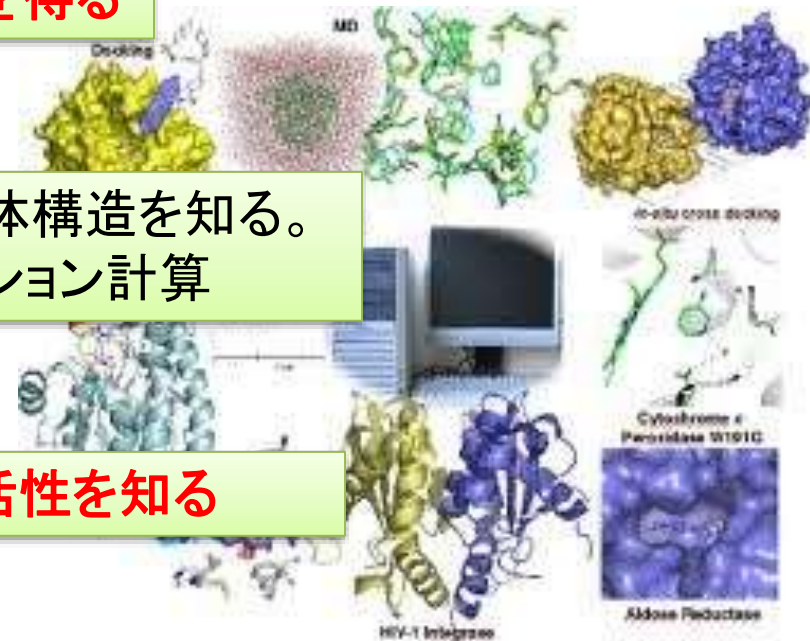
(1) スクリーニングなどにより、活性化合物を得る

(2) タンパク質と活性化合物の正確な複合体構造を知る。
X線構造解析、NMR実験、分子シミュレーション計算

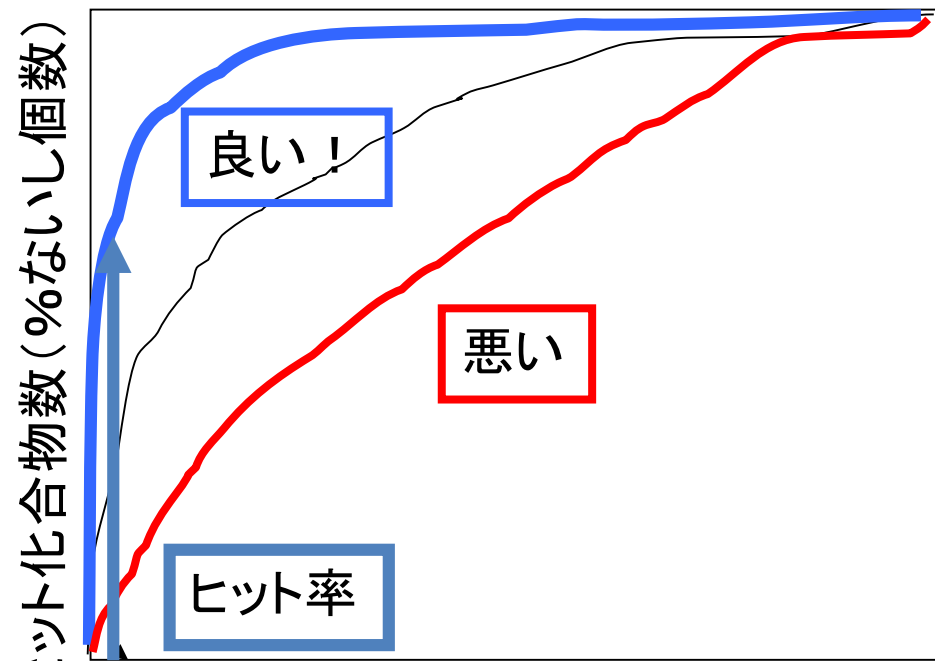
(3) タンパク質と活性化合物の正確な結合活性を知る

(4) 化合物を改変する(人間の知恵によるアートの世界)
4-① 「合成できる化合物」で、
4-② 水に溶け、吸収できる化合物でなければならない
4-③ **Off-target**に作用しないこと

(4) 合成、アッセイ実験



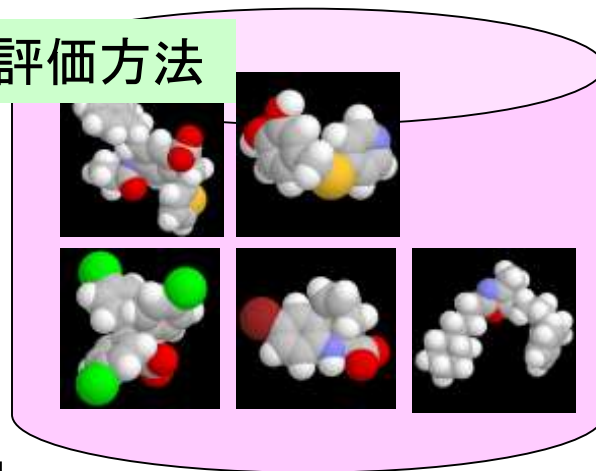
データベースエンリッチメントカーブ:スクリーニングの評価方法



購入化合物数(%ないし個数)

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

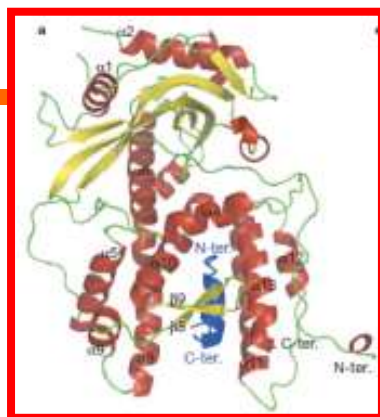
$$\text{ヒット率} = \text{ヒット数} / \text{購入化合物数} \times 100(\%)$$



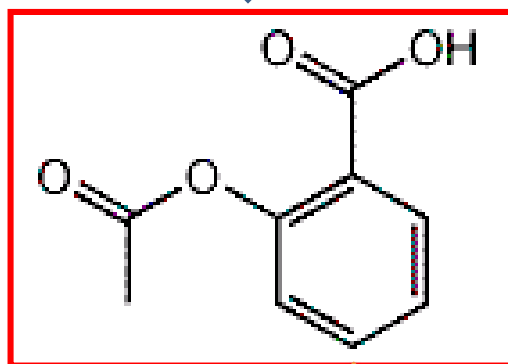
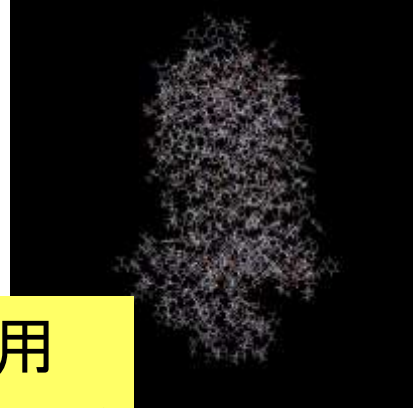
薬物スクリーニング

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

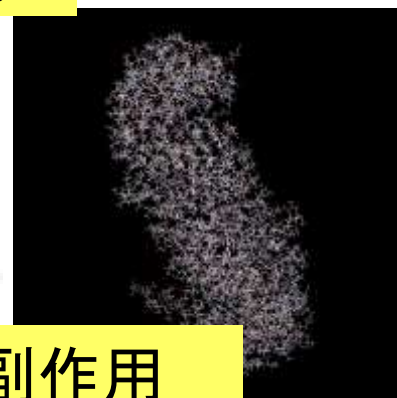
標的蛋白質
Target
protein



副作用
Off target

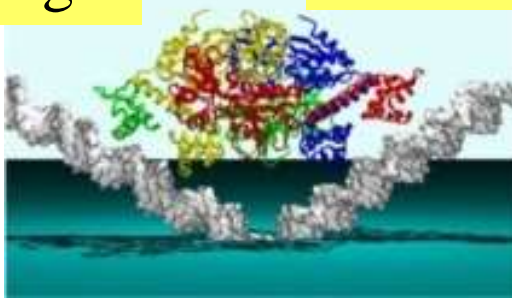


副作用
Off target



副作用
Off target

副作用
Off target



化合物:
compound

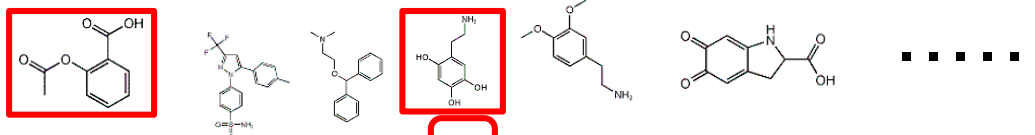
化合物:
compound

In silico screening手法開発 (Multiple target screening)

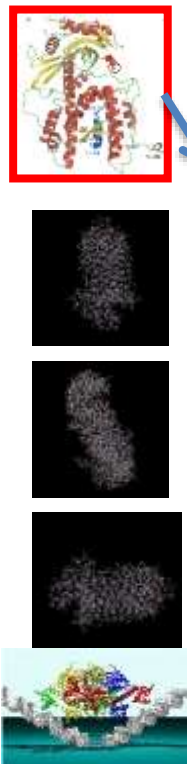
1つの化合物を、
標的蛋白質蛋白質+それ以外の200種類の蛋白質にドッキングし、
結合力を見積もる。

複数の蛋白質の中で、標的に一番強く結合する化合物を選ぶ

化合物ライブラリー



蛋白質ライブラリー



	1	2	3	4	5	6	7	8	9	10	11
1	1	12	8	3	1	6	4	2	4	4	6
2	9	11	9	2	2	1	5	3	3	1	5
3	10	10	10	12	3	5	6	5	2	5	4
4	11	9	5	1	6	4	2	1	1	3	3
5	4	8	7	11	5	3	7	6	11	2	2
6	3	1	6	4	4	12	1	4	10	6	1
7	8	2	2	5	9	2	8	7	9	8	11
8	12	5	11	6	8	11	9	8	8	8	11
9	2	7	3	10	7	10	10	12	12	9	10
10	7	3	12	7	12	9	11	9	7	10	9
11	6	4	4	8	10	8	12	10	6	11	8

標的蛋白質

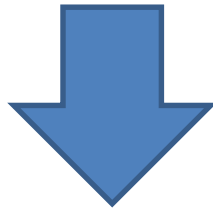
ヒット化合物

In silico screeningの開発 (Multiple target screening)

複数の蛋白質の中で、標的に一番強く結合する化合物を選ぶ。



標的に効いて、副作用の少ない化合物を予測できる。



真のヒット率 (活性化化合物が予測によって見つかる確率) が向上

簡単な機械学習で大量のデータを整理する

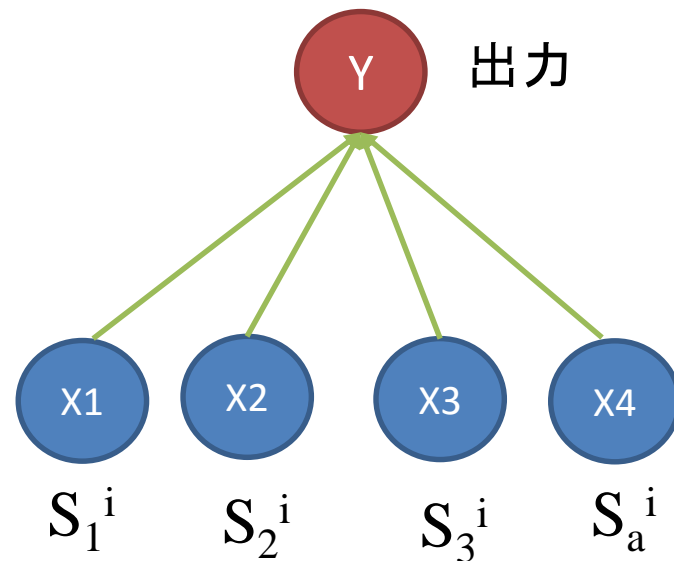
ドッキングスコア～タンパク質－化合物結合エネルギー

似たタンパク質は、似た化合物を結合するだろう。

タンパク質aと化合物iのドッキングスコアは、蛋白質aに類似した蛋白質bと化合物iのドッキングスコアに近い値をとる。
ドッキングスコアは、他のタンパク質へのドッキングスコアの重み付平均で精度を向上できる。

$$S_a^{new\ i} = \frac{\sum_b S_b^i M_a^b}{\sum_b M_a^b}$$

S_a^i : 蛋白質 aと
化合物 iの結合エネルギー推定値



薬物探索での機械学習: 単純なAI利用

機械学習法により、もともと与えた「既存の活性化合物」が予測上位に発見できるように、学習する。

データベースエンリッチメント = 正解発見率が最適化されるまで反復する

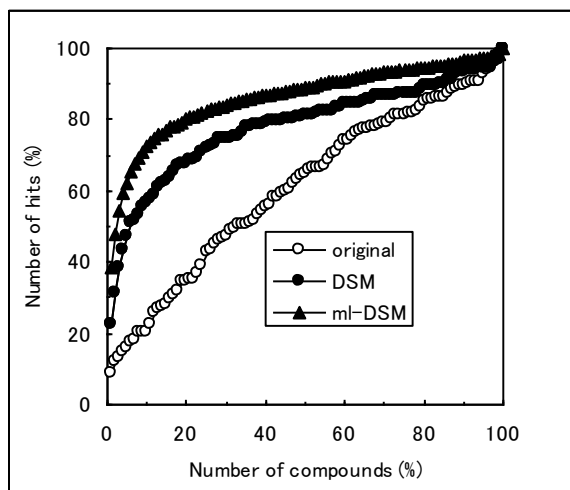
$$S_a^{new^i} = \sum_b S_b^i M_a^b$$

線形結合係数の調整



	1	2	3	4	5	6	7	8	9	10	11	12
1												
2		○	○		○		○					
3								○				
4	○				○							
5					○							
6		○	○		○		○			○	○	○
7												
8		○			○		○					
9					○							
10												
11					○				○			
12					○							

ドッキングスコアの修正



データベースエンリッチメント = 発見率の計算

MTS法でスクリーニング

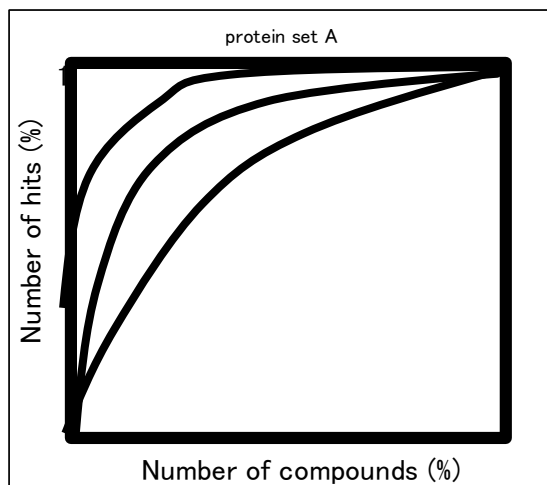
薬物探索での機械学習: 単純なAI利用

機械学習法により、もともと与えた「既存の活性化化合物」が予測上位に発見できるように、学習する。

データベースエンリッチメント = 正解発見率が最適化されるまで反復する

$$S_a^{new^i} = \sum_b S_b^i M_a^b$$

線形結合係数の調整



データベースエンリッチメントの計算

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2		○	○		○		○					
3								○				
4	○				○							
5					○							
6		○	○		○		○			○	○	○
7												
8		○			○		○					
9					○							
10												
11									○			
12					○							

ドッキングスコアの修正

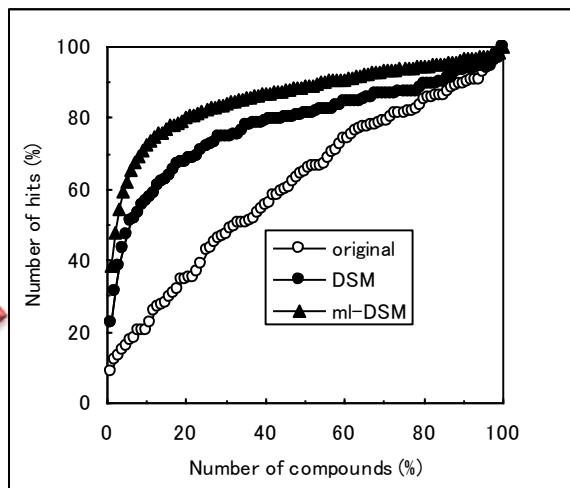
ITS法でスクリーニング

薬物探索での機械学習: 単純なAI利用

機械学習法により、もともと与えた「既存の活性化化合物」が予測上位に発見できるように、学習する。

$$S_a^{new^i} = \sum_b S_b^i M_a^b$$

線形結合係数の調整



データベースエンリッチメントの計算

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2		○	○		○		○					
3								○				
4	○				○							
5					○							
6		○	○		○		○			○	○	○
7												
8		○			○		○					
9					○							
10												
11					○				○			
12					○							

ドッキングスコアの修正

ITS法でスクリーニング

医薬品探索のヒット率(発見率)は改善 (~3% → ~10%)
1層だけでできたAIでも、かなり有効

分子設計に必要なこと

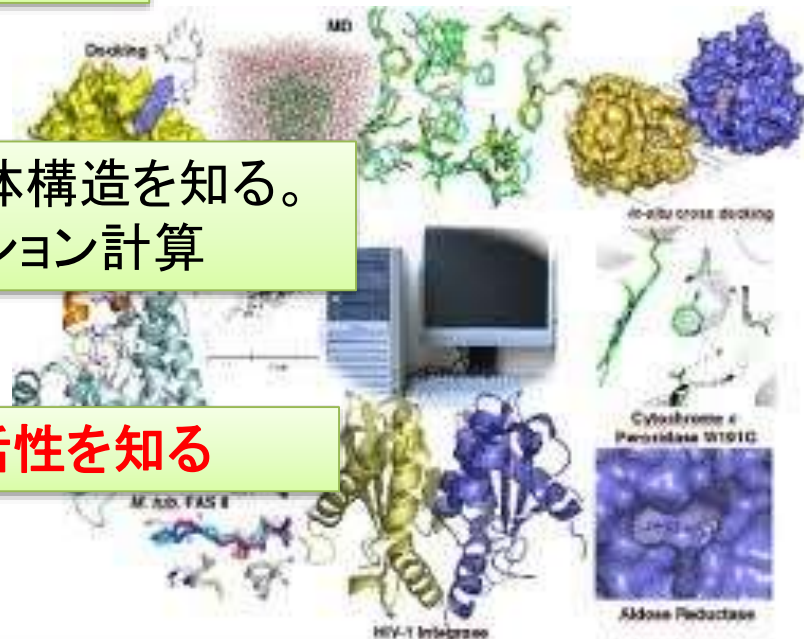
(1) スクリーニングなどにより、活性化合物を得る

(2) タンパク質と活性化合物の正確な複合体構造を知る。
X線構造解析、NMR実験、分子シミュレーション計算

(3) タンパク質と活性化合物の正確な結合活性を知る

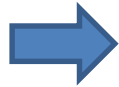
(4) 化合物を改変する(人間の知恵によるアートの世界)
4-① 「合成できる化合物」で、
4-② 水に溶け、吸収できる化合物でなければならない
4-③ **Off-target**に作用しないこと

(4) 合成、アッセイ実験

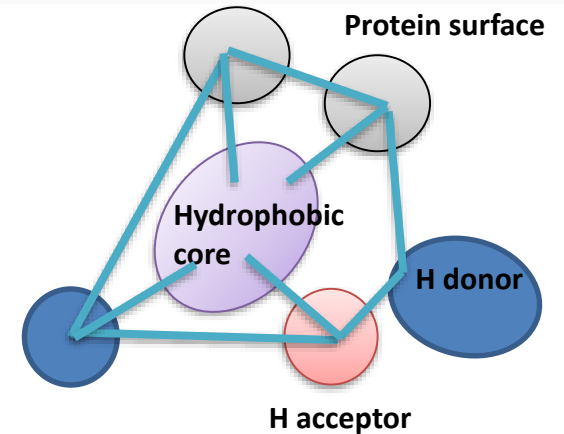


包括的なタンパク質-化合物相互作用エネルギーの解析

タンパク質にも化合物にも通用する記述子



ファーマコフォアを使えば
 (1) 蛋白質ポケットの類似性
 (2) 化合物の類似性
 両方、計算できる



全てのファーマコフォア
 Pharmacophore = { ph1, ph2, ph3, ..., phL } (無限の数)



ファーマコフォア: タンパク質-薬物相互作用の場の空間分布

Cf: ファーマコフォアで化合物を記述

Sc: 化合物同士の類似性

	ph1	ph2	ph3	ph4	ph_L
Compound1	0.1	0.3	0	0.2	:	0.1
Compound2	0.2	0.1	0.9	0.8	:	0.1
Compound3	0.3	0.4	0.1	0.2	:	0.4
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
compoundN	0.1	0.4	0.6	0.3	0.5

	compound1	compound2	compound3	compoundN
compound1	1	0	0	0.4
Compound2	0	1	0.5	0.2
Compound3	0	0.5	1	0.1
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
compoundN	0.4	0.2	0.1	1

Dimension = infinite

$$S_c = C_f * C_f^T \quad : T \text{ 転置行列}$$

各化合物がファーマコフォア ϕ で記述されれば

$$C_f = \{ \phi_1, \phi_2, \phi_3, \phi_4, \dots, \phi_N, \}$$

ファーマコフォアを、化合物とタンパク質ポケットの共通の記述子
 : 結合の本質とするとタンパク質-化合物相互作用が計算できる

T

	ph1	ph2	ph3	ph4	ph_L
化合物1	0.1	0.3	0	0.2	:	0.1
化合物2	0.2	0.1			:	0.1
化合物3	0.3	0.4			:	0.4
:	:	:			:	:
:	:	:			:	:
化合物N	0.1	0.4	0.6	0.3	0.5

Cf

	ph1	ph2	ph3	ph_L
ph1	8.2	1.2	2.1	:	0.1
ph2	1.2		2.4	:	0.1
ph3	2.1		9.5	:	0.4
:	:	:		:	:
:	:	:		:	:
ph_L	-3.1	-2.1	4.4	5.5

G

	ph1	ph2	ph3	ph4	ph_L
蛋白質1	0.2	0.1	0	0.5	:	0.1
蛋白質2	0.2	0		0.3	:	0.1
蛋白質3	0.3	0		0.2	:	0.3
:	:	:			:	:
:	:	:			:	:
蛋白質M	0.1	0.3	0.6	0.5	0.1

Pf

=

	化合物1	化合物2	化合物3	新規化合物N
蛋白質1	1	0.2	0	0.4
蛋白質2	0.1	0.2		0.99
蛋白質3	0	0.5		0.1
:	:	:		:
:	:	:		:
蛋白質M	0.4	0.9	0.1	0.4

M

$M = C_f * G * P_f^T$: Tは転置行列
 Gは、各ファーマコフォアの類似度と、結合エネルギーを表す対称行列
 (単位は加法性のためエネルギー:kcal/mol)

Mreal

	化合物1	化合物2	化合物3	新規化合物N
蛋白質1	1		0	?
蛋白質2			0.5	?
蛋白質3	0	0.5		?
:	:	:	:	:
:	:	:	:	:
蛋白質M	0.4		0.1	?

予測

予測

Gを、現実のデータに合うように、モンテカルロ法で調整すればよい。

(1) ドッキングスコア = 結合エネルギー
 (2) ドッキングスコア = タンパク質と化合物の類似性記述子

異なる化合物でも同じタンパク質に結合するなら、「類似」活性

化合物集団

タンパク質集団 : 200-600

proteins

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2		○	○		○		○					
3								○				
4	○				○							
5					○							
6		○	○		○		○		○	○	○	
7												
8		○			○		○					
9					○							
10												
11					○			○				
12					○							

類似タンパク質

類似化合物

ファーマコフォアはタンパク質と化合物の類似性尺度

ファーマコフォアの代わりに、多数のタンパク質に対するドッキングスコアで代用

包括的なタンパク質-化合物相互作用エネルギーの解析

タンパク質 a への
化合物 i の結合活性を

$$\Delta G_a^i = \sum_b S_b^i M_{(PCA)_a}^b$$

$S_{b,i}$ = ドッキングスコア

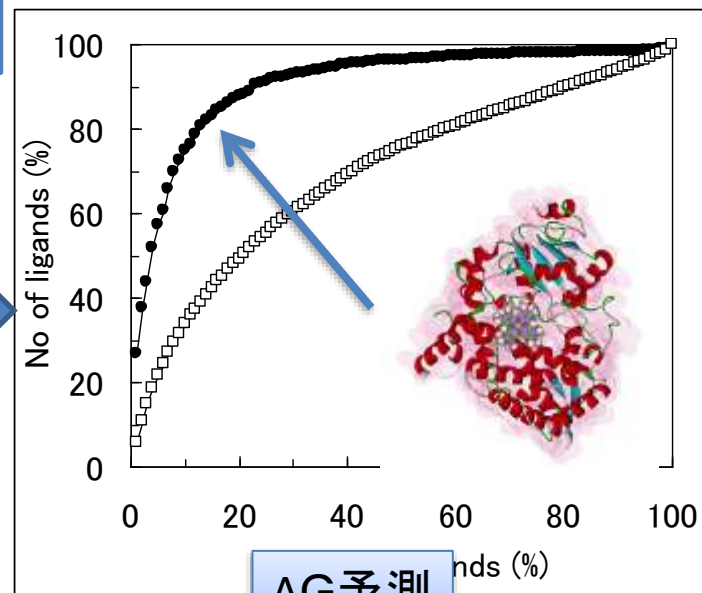
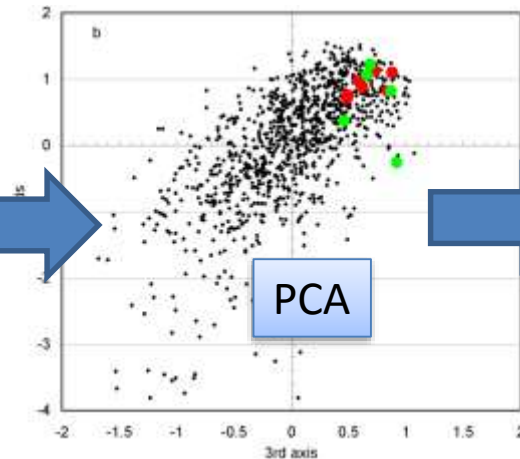
M_{ab} : パラメータ

公的DBをダウンロードし、自動処理

proteins

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2		○	○		○		○					
3									○			
4	○				○							
5												
6	○									○		
7												
8		○			○		○					
9					○							
10												
11					○					○		
12					○							

Docking scores



ΔG 予測

- Y. Fukunishi, S. Yamasaki, I. Yasumatsu, K. Takeuchi, T. Kurosawa, and H. Nakamura. Molecular Informatics (2016).
- Y. Fukunishi, S. Hojo, and H. Nakamura. Journal of chemical information and modeling. 46 (6), 2610-2622 (2006).

包括的なタンパク質-化合物相互作用エネルギーの解析

ChEMBL database

Targets: 10,579 proteins

Compound records: 1,638,394 compounds

Activities: 12,843,338 interactions **Ki, Kd, IC50, %-inhibition, Km, and etc**

Publications: 57,156 papers

いろいろな実験値を ΔG に換算して利用する

$$\Delta G = k_B \cdot T \cdot \ln(K_d)$$

IC50

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_s}}$$

Ki

$$K_d = K_i$$

(assumption)

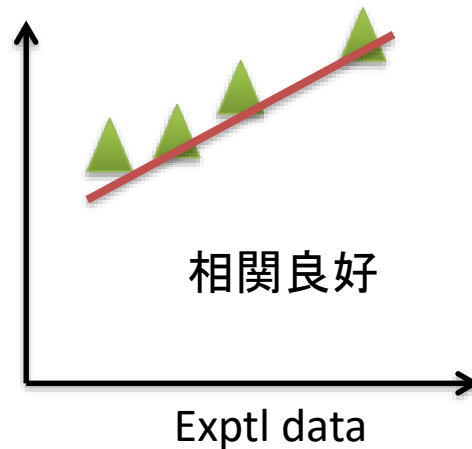
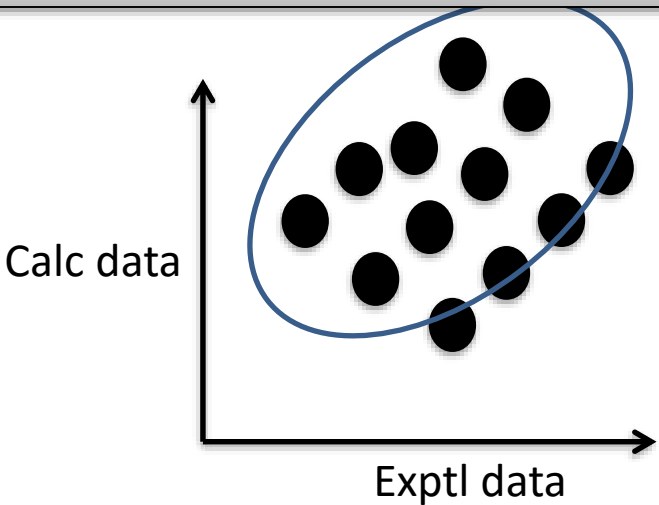
%-residual (r)

%-inhibition = 1 - r

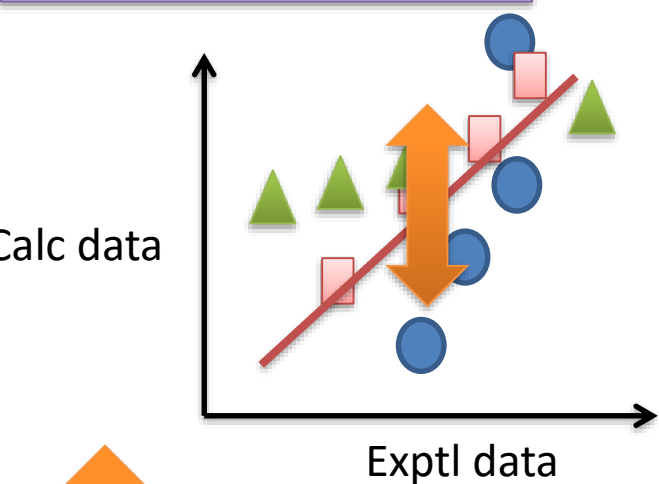
$$K_i = \frac{[I]}{\frac{[S]/r - [S]}{K_s} - 1}$$

実験条件(K_s , $[S]$, S: natural ligand like ATP. $[I]$)がDBに記載なし。
標準値で代用するしかない。

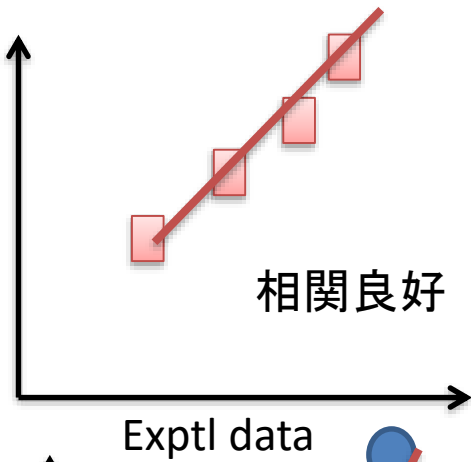
うまく回帰しにくい。実験値が「混ざりもの」なので



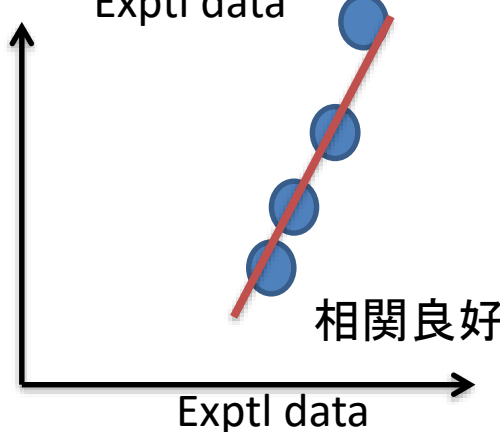
データを元データに分割



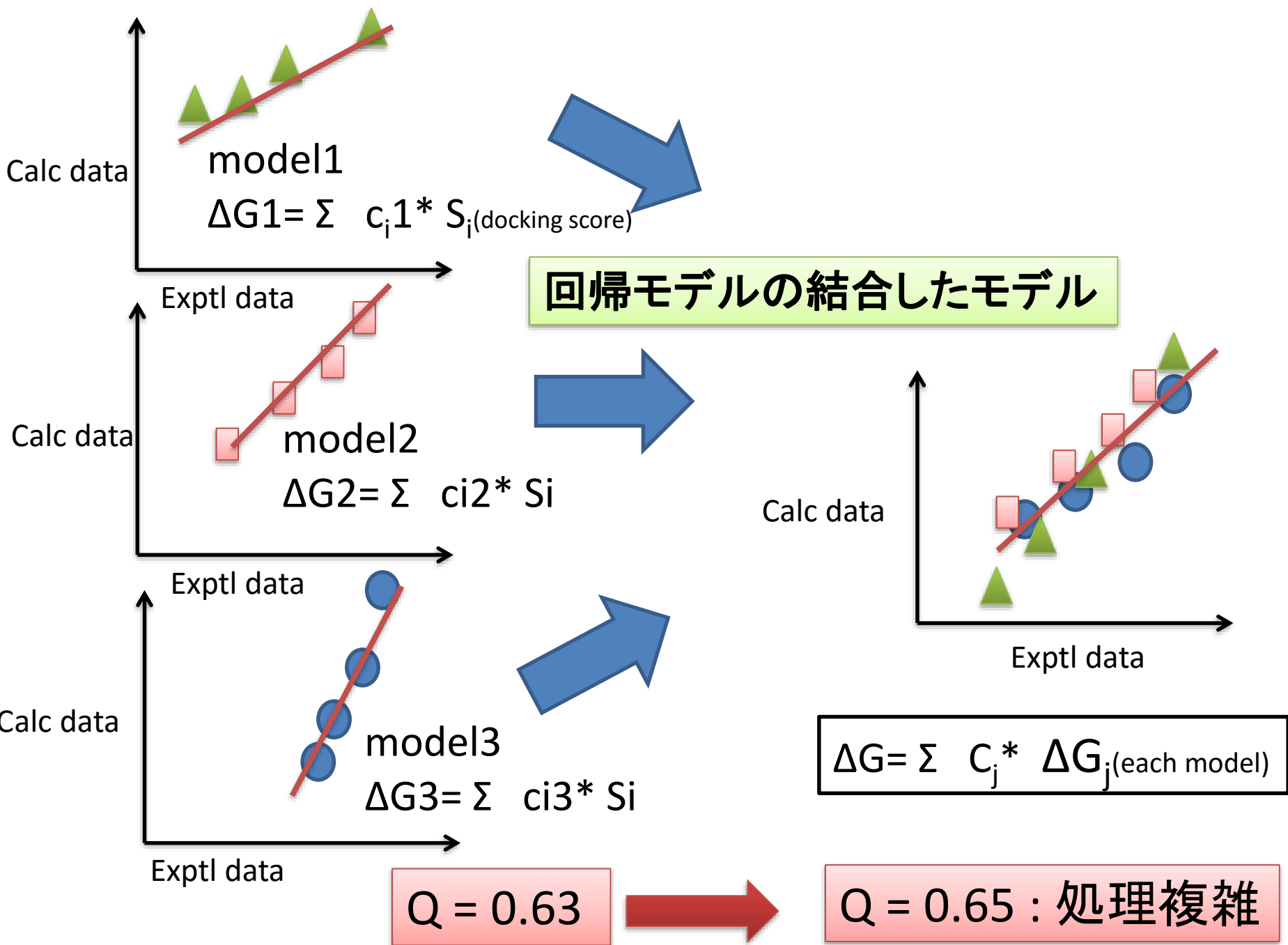
Calc data



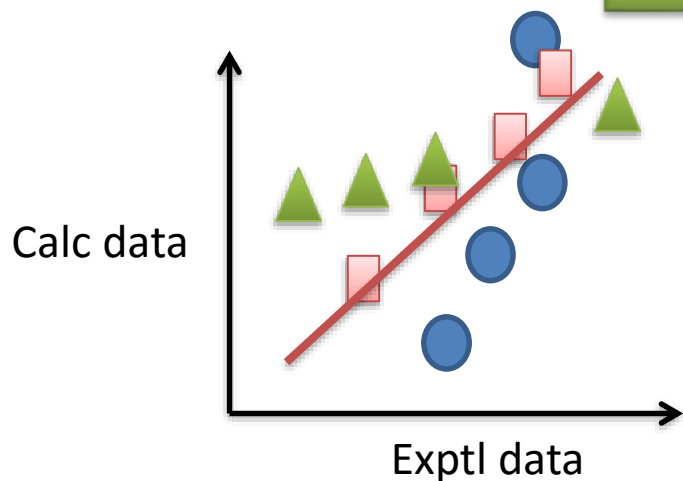
Calc data



Experimental deviation
= 1 - 1.5 kcal/mol



重み付き学習法: Weighted learning model

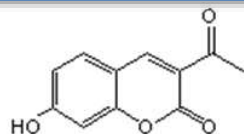


Q = 0.56

重み付き学習

入力(ΔG を知りたい分子)に、分子構造の近い実験値を重点的に利用する

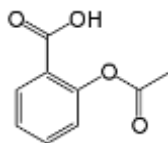
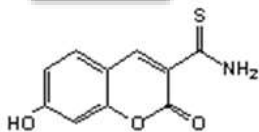
Data 1: 入力に似る



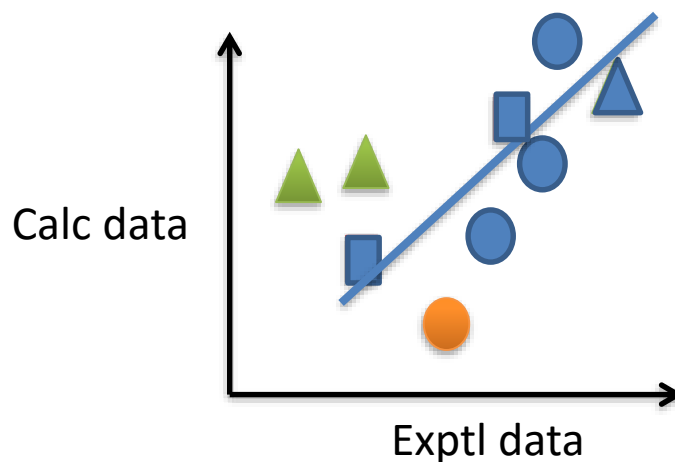
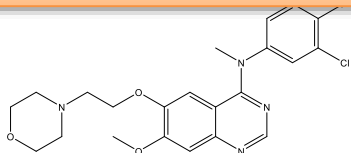
Similar compounds (●) for input are weighted

入力

Data 2: 入力に似てない



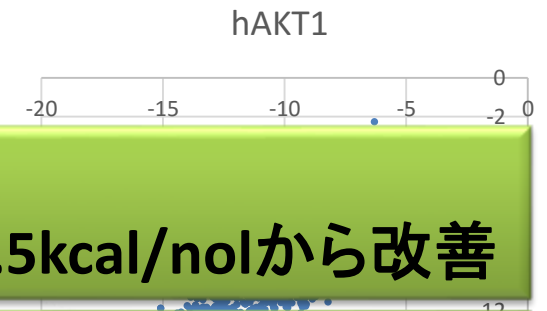
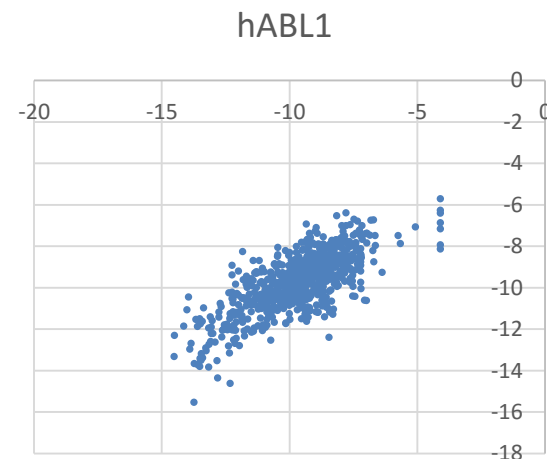
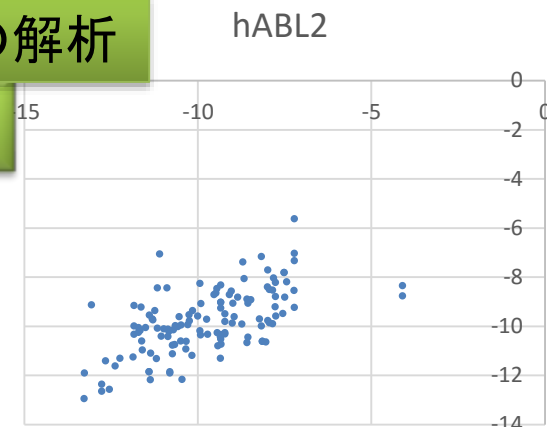
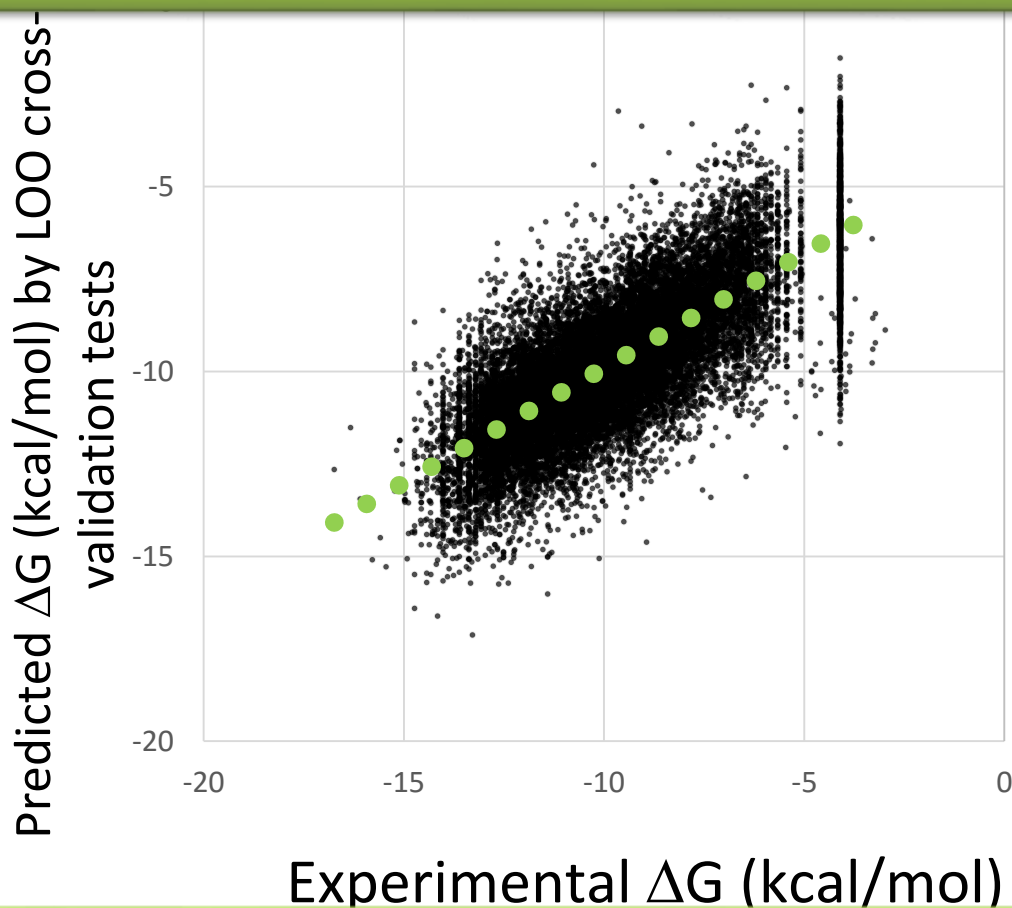
Data 3: 入力に似てない



Q = 0.59 : 簡単

包括的なタンパク質-化合物相互作用エネルギーの解析

100タンパク質 (4万化合物)の自動QSAR

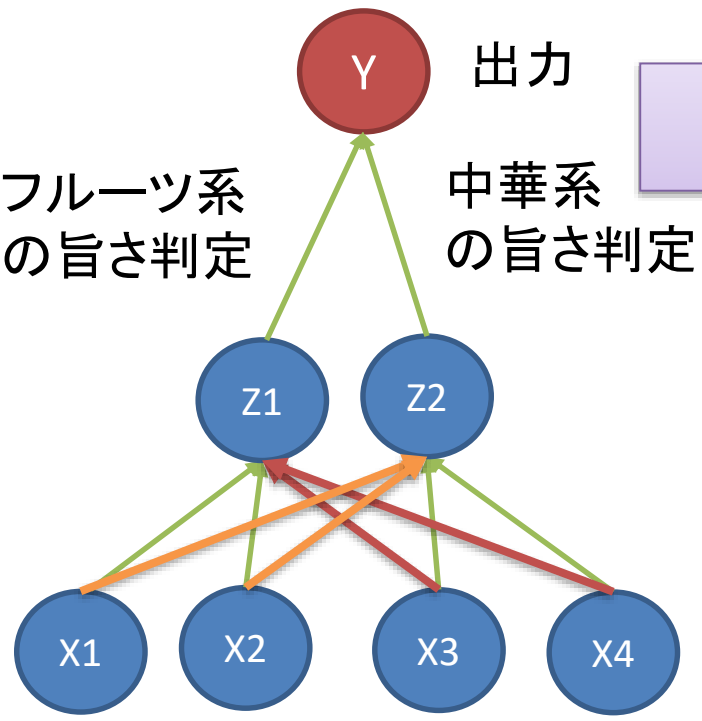


平均誤差 (ΔG)

= 1.5kcal/mol → ドッキングスコアの誤差2.5kcal/molから改善

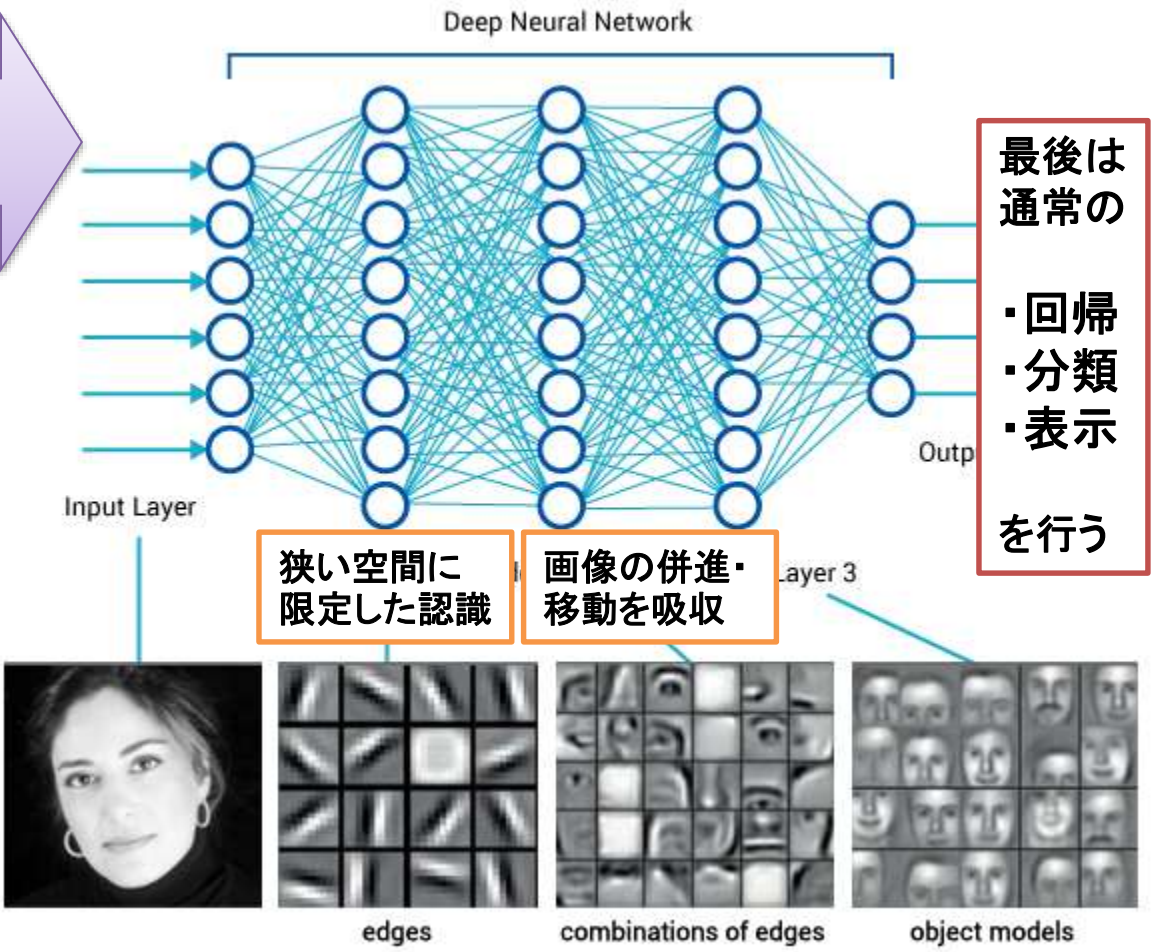
包括的な ΔG 計算手法は1万タンパク質に適用可能

多段階(多層)にすると、複雑な処理ができる



$Y = a * X + b$ の場合と同じく、「答え:実測値」のデータが必要

※: 中間ノードの数を、回帰に本質的なパラメータの数にすることが大事
 =PLS回帰とかでのPRESSとか、R以外にQを最適化するパラメータ調整法と類似



最後は通常の

- ・回帰
- ・分類
- ・表示

を行う

「もの」
「化合物」

「もの」を
記述子に
変換

記述子と
記述子の
位置関係

「記述子位
置関係」の
関係

AIで、できること

回帰分析

分類

の、3つのみ

低次元化

自然言語処理: Googleなどのインデクサー + tf-idf 理論

DNNによる自然言語処理

Xiang Zhang, Yann LeCun." Text Understanding from Scratch" <https://arxiv.org/pdf/1502.01710v5.pdf>

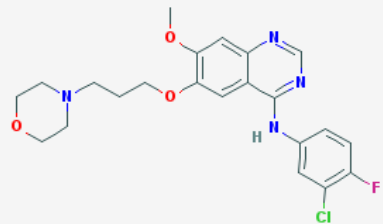
安易な応用例の案なので、自分ではやる気しないが...

化学でできる例1: 分子構造からの物性(LogS/LogP)・活性(ΔG)予測+化合物空間表示など

タンパク質や低分子

文字列情報: SMILES, InChKey, など

COC1=C(C=C2C(=C1)N=CN=C2NC3=CC(=C(C=C3)F)Cl)OCCCN4CCOCC4



CGVBSではアミノ酸2文字等

Convolution NNの局所視覚野が、「分子フラグメント」に相当する

N記述子

	C	F	I	N	O	()	1	2	=
C	1	0	0	0	0	0	0	0	0	0
O	0	0	0	0	1	0	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0
=	0	0	0	0	0	0	0	0	0	1
C	1	0	0	0	0	0	0	0	0	0
(0	0	0	0	0	1	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0	0	1
C	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	1	0	0
C	1	0	0	0	0	0	0	0	0	0

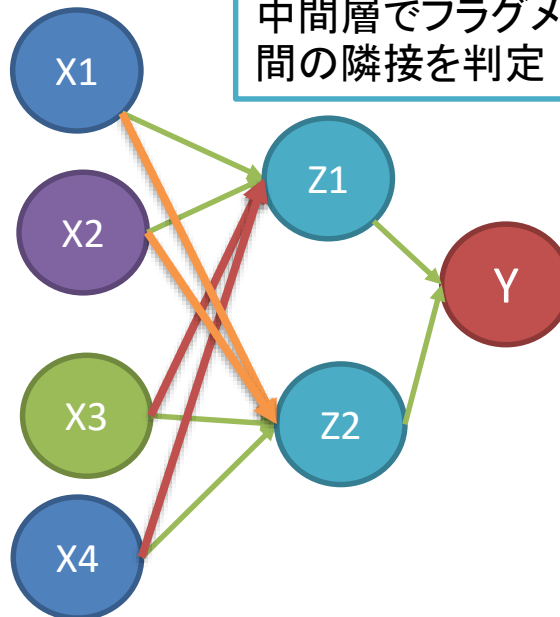
Convolution

2xN convolution

3xN convolution

4xN convolution

中間層でフラグメント間の隣接を判定

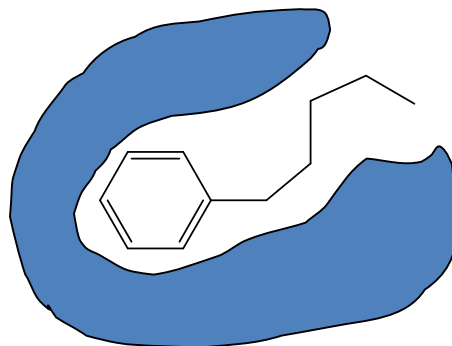


出力
LogS/LogP
 ΔG など

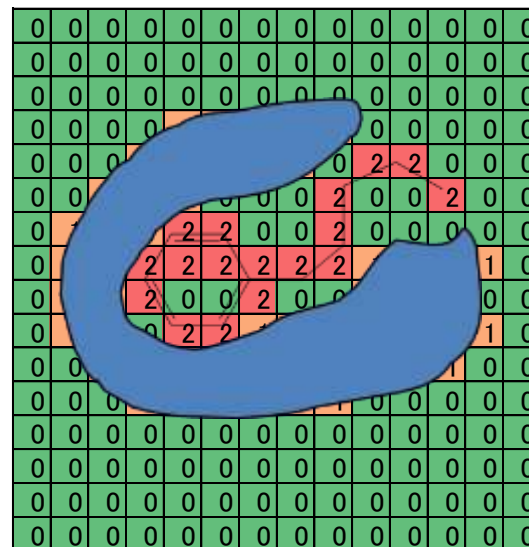
大きさのバラバラな分子も、convolutionで一定次元に変換

化学でできる例2:ドッキング・ドッキングによるスクリーニング計算

タンパク質構造や、タンパク質-化合物複合体

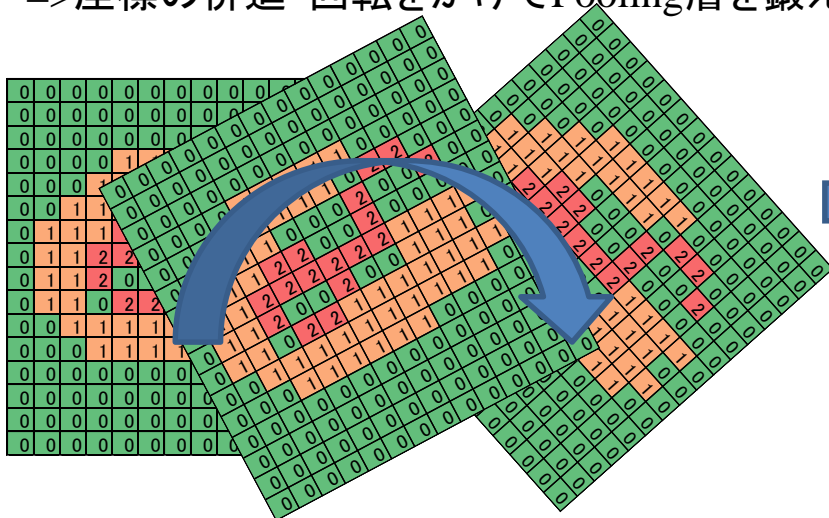


3次元ピクセルで「画像化」



分子間相互作用を捉えるConvolution

Pooling : ユークリッド距離空間は、併進回転群に対し不変
=>座標の併進・回転をかけてPooling層を鍛える



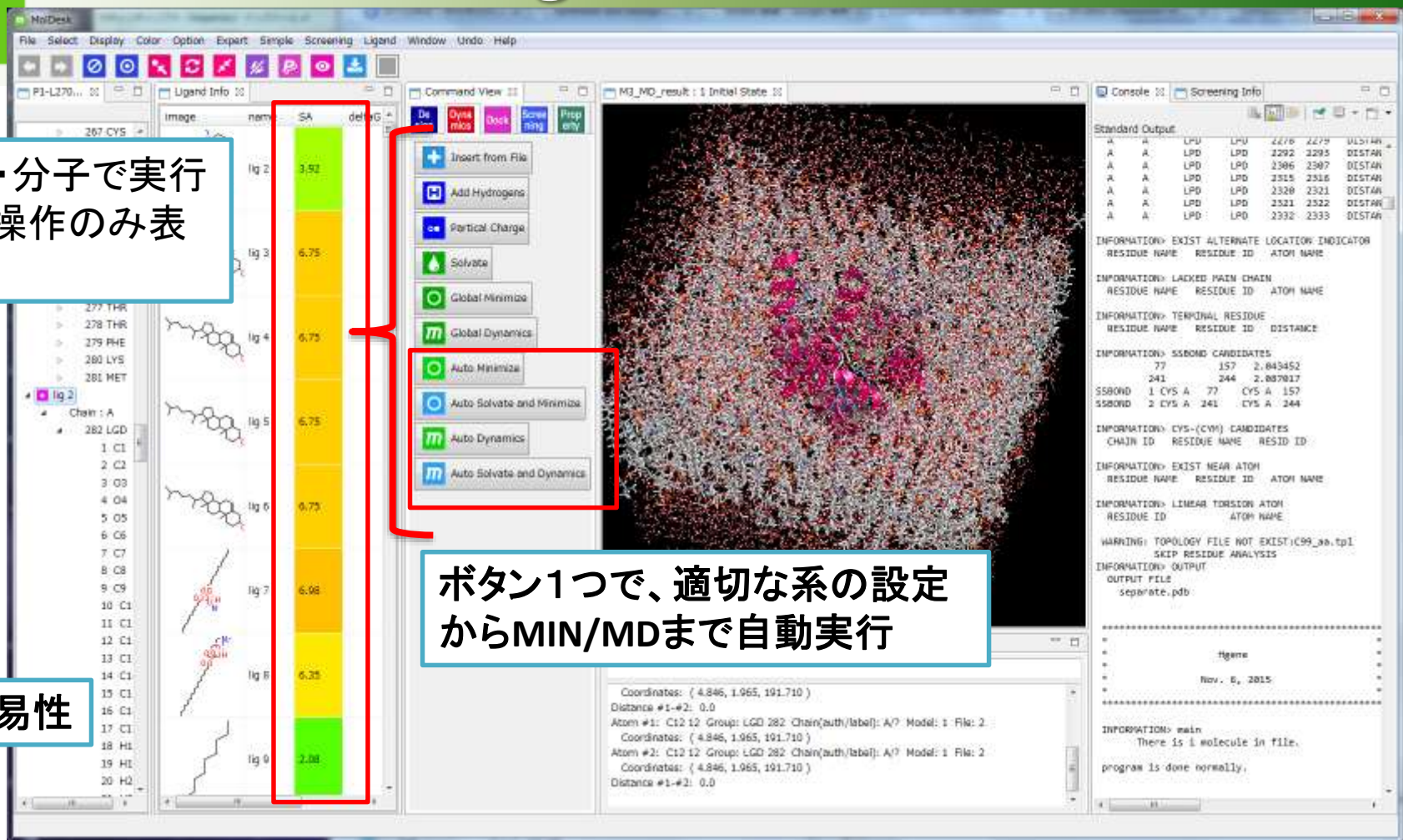
ΔG や、「結合・非結合」
タンパク質のカテゴリなど
回帰・分類したい値を与えて、
バックプロパゲーションでトレーニング

実施例: Thomas Unterthiner, Andreas Mayr, G unter Klambauer,
Marvin Steijaert, J org K. Wegner, Hugo Ceulemans, Sepp Hochreiter.
Deep Learning as an Opportunity in Virtual
Screening. <http://www.bioinf.jku.at/publications/2014/NIPS2014a.pdf>

各状態・分子で実行可能な操作のみ表示

合成容易性

マニュアル不要



The screenshot displays the MolDeskScreening software interface. On the left, a list of molecules is shown with their respective scores: lig 2 (3.92), lig 3 (6.75), lig 4 (6.75), lig 5 (6.75), lig 6 (6.75), lig 7 (6.98), lig 8 (6.35), and lig 9 (2.08). A red box highlights the 'Command View' menu, which includes options like 'Insert from File', 'Add Hydrogens', 'Partial Charge', 'Solvate', 'Global Minimize', 'Global Dynamics', 'Auto Minimize', 'Auto Solvate and Minimize', 'Auto Dynamics', and 'Auto Solvate and Dynamics'. A red arrow points from the 'Auto Minimize' button to a 3D molecular model of a protein-ligand complex. The console window on the right shows the output of the simulation, including standard output, information about alternate locations, and warnings about topology files.

ボタン1つで、適切な系の設定からMIN/MDまで自動実行

ポケット探索
(MolSite法: 世界最高精度)

NMRのデータも考慮できる
高精度ドッキング(予測精度2倍)

ボタン1つで、系の準備
からドッキングまで自動実行

マニュアル不要

Command View

- Insert from File
- Add Hydrogens
- Partial Charge
- Delete
- Make Pocket
- Find Pocket
- Select Receptor Molecule
- Docking
- Docking NMR
- Auto Docking

Screening Info

CLUSTER	SCORE	STRUCTURES
85	-279.3969	
86	-279.3968	
9	-279.3452	
97	-275.0439	
98	-275.0458	
99	-275.0457	
10	-293.345	4

Console

```
Atom #1: O 52 Group: PNT 1 Chain(auth/label): C/? File: 3  
Coordinates: (-1.983, -7.810, 120.064 )  
Atom #2: O 48 Group: PNT 1 Chain(auth/label): C/? File: 3  
Coordinates: ( 0.459, -6.318, 119.656 )  
Distance #1-#2: 2.8906548  
Atom #1: O 52 Group: PNT 1 Chain(auth/label): C/? File: 3  
Coordinates: (-4.622, -5.478, 118.911 )
```



EGFR-エルロチニブのドッキング計算

The screenshot shows the MolDesk software interface. On the left, a project tree shows a receptor and ligands. The central panel displays a 3D ribbon model of a protein with a ligand docked. On the right, a 'Screening Info' window shows a table of results with columns for Image, rank, ID, Score, and Source ID. A red box highlights the 'Export as csv' button in the table's toolbar.

Image	rank	ID	Score	Source ID
		H-04326943-02	4.71	NS-10278135
		H-04320832-03	4.62	NS-10271650
		H-003221131-01	4.53	NS-00458945
		H-00349605-02	4.49	NS-00488718
		H-04534297-02	4.43	NS-10563086
	6	HTS1404-00724014-02	4.41	NS-00949811
	7	HTS1404-03702991-01	4.4	NS-08473205

SBDD

LBDD

部分構造探索
Substructure
search

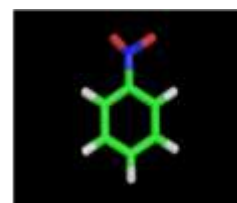
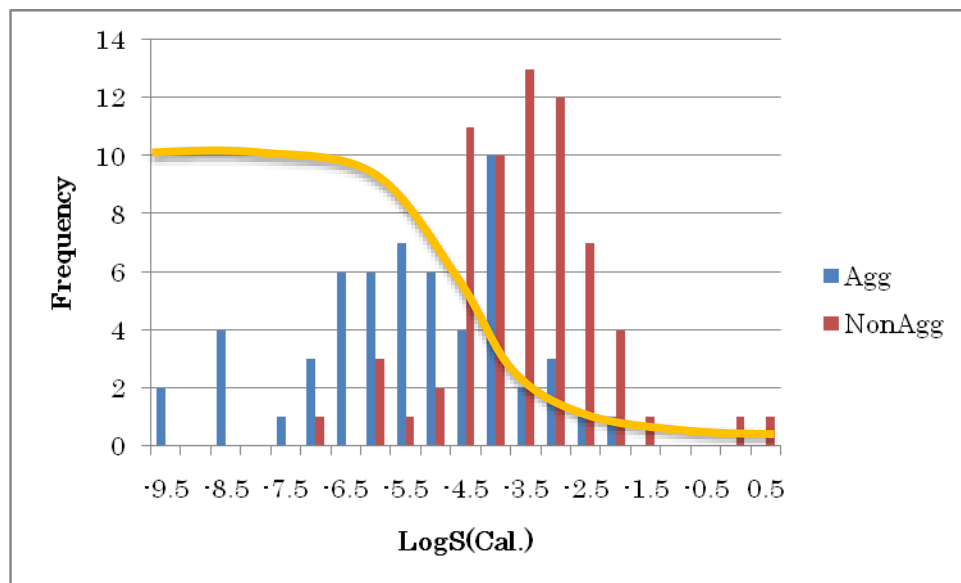
実験データの
再利用でヒット率が上がる
機械学習法
Machine-learning method in
virtual screening

化合物購買リスト作成
メールで試薬会社へ送れば購入できる。
(アッセイも外注可能)
1化合物～1万円
1アッセイ～10万円 (IC50)

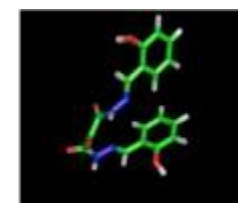
**7時間 (WS) で200万化合物SBDD:
アマゾンクラウドで2 時間 (3,500 円).**

アグリゲータ予測

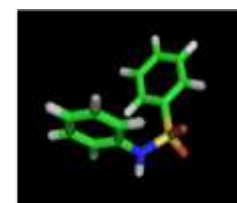
アグリゲータ分子及び非アグリゲータ分子の溶解度推算



(1) -1.66

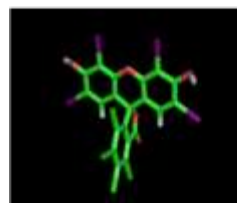


(2) -2.28

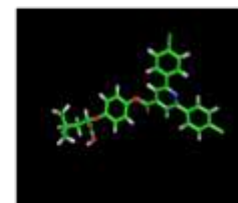


(3) -2.50

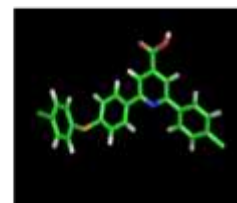
aggregators of high solubility



(1) -9.20



(2) -7.80

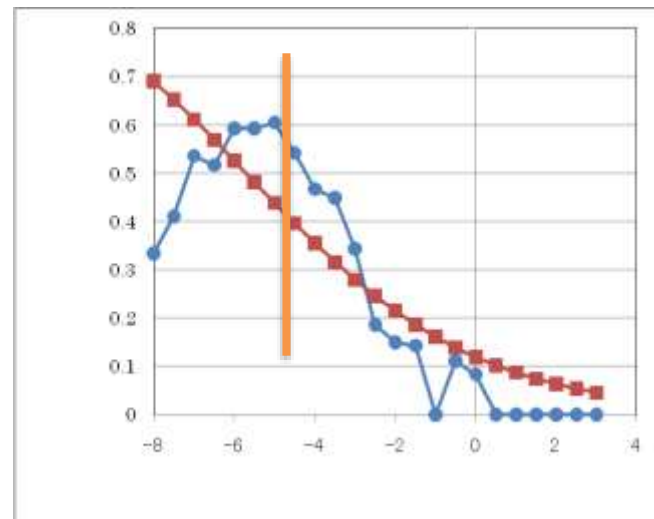
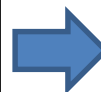
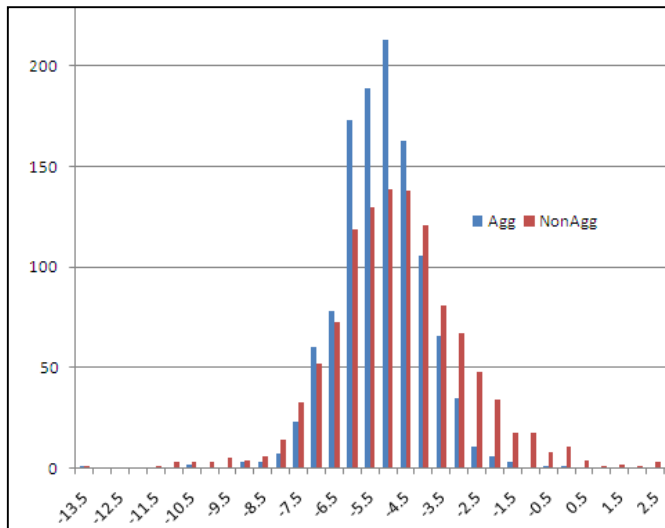


(3) -7.65

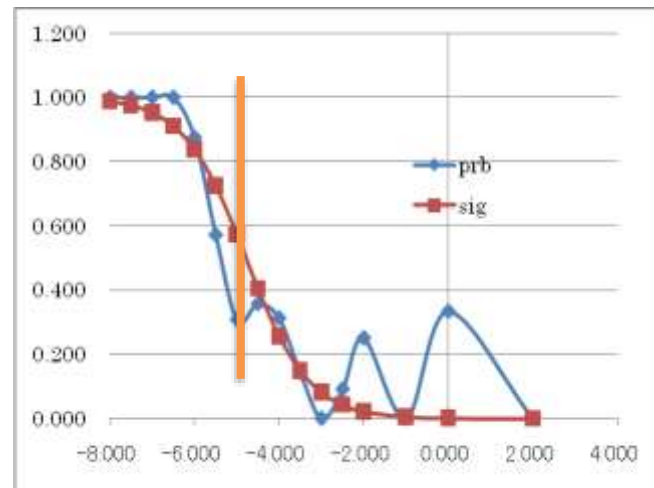
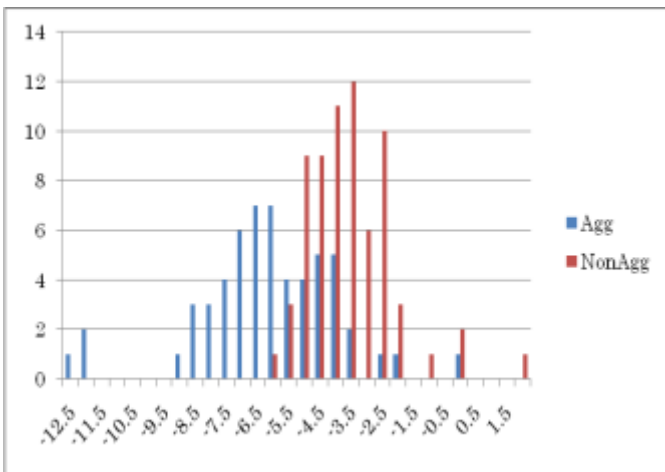
aggregators of low solubility

アグリゲータと非アグリゲータの閾値は $\log S = -5.0$ 程度

HTS データ



文献値 (既知医薬品)



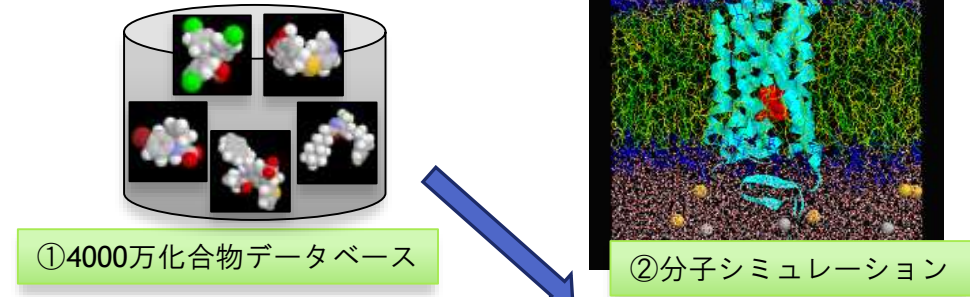
分子量と、ヘテロ原子(N, O, S)の含有原子数と、アグリゲーターになる確率(%)

90%のヒットは、非特異的な「カス」

		MW		
		200-300	300-400	400-500
no. of N,O	0-5	23.022	62.559	91.983
	5-10	9.877	45.094	69.827
	10-	3.172	18.927	36.560

myPresto連携企業群

myPresto



出力ファイルを
メールして試薬購入

メールで合成発注

外注試薬倉庫

メールでアッセイ発注

動物実験発注

メールで結果が届く

実験室・倉庫不要

① 薬物候補物質の探索

② 薬物の最適化

実空間

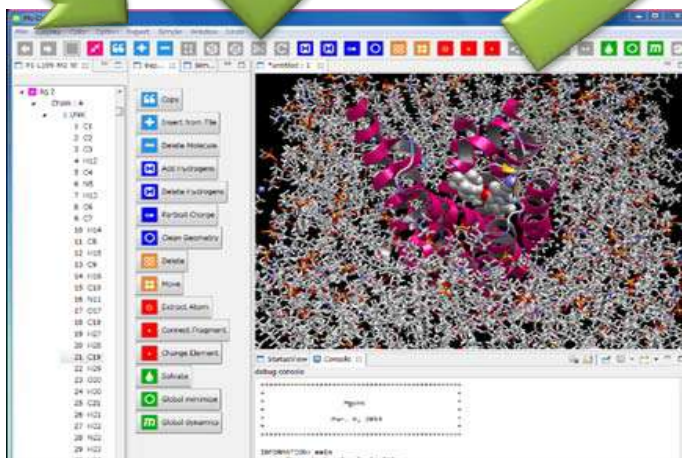


誰もが、実験室・実験経験なしに、創薬研究を開始できる世界

クラウド利用

受託解析・
コンサルティング

買ったその日に創薬開始。
ハード一体型製品



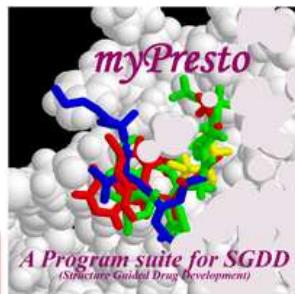
試薬購入

試薬合成

アッセイ実験

誰でも30分で覚えられる簡単操作

謝辞



myPrestoは、経済産業省・NEDO・AMEDの支援を受けています。

大阪大学: 中村春木 教授

JBIC・次世代天然物化合物組合: 榎原 信雄

情報数理バイオ: 真下 忠彰

富士通: 和田 光人

第一三共ノバーレ: 安松 勲

日立ソリューションズ東日本: 三上 義明・黒澤 隆

ユーザー(企業、研究者など)

おわり