

## 2.3 研究開発項目③「高精度in silico スクリーニング等のシミュレーション技術」

集中研究: 社団法人バイオ産業情報化コンソーシアム[バイオメディシナル情報研究センター(BIRC)、大阪大学分室、情報数理研究所分室、塩野義製薬分室、アステラス製薬分室、三井化学アグロ分室]

共同研究先: 独立行政法人産業技術総合研究所バイオメディシナル情報研究センター、国立大学法人大阪大学蛋白研究所

### 【中期目標】

高精度のインシリコスクリーニングを実現するため、以下の技術を開発する。さらに、①、②の技術開発との連携により、産業上有用な化合物を5個以上取得する。

- 1) タンパク質の動的性質を正しく評価し、タンパク質受容体への基質結合能を高い精度で計算できる新しい計算科学手法の開発し、インシリコスクリーニングの効率を従来法に比べ5倍程にあげる。
- 2) タンパク質と化合物とのドッキング計算手法の精度を高め、ターゲット選択性能を従来法に比べ5倍程度上げる
- 3) タンパク質間相互作用を阻害・制御する低分子化合物の選択・設計の技術開発を行う。

### 【達成度】

高精度のインシリコスクリーニング技術を開発し、国際的な業界標準であるソフトウェアに対する優位性を確認した。その手法により、 $\mu$ オピオイド受容体アゴニスト(48ヶ)や農薬のシードとなる化合物(23ヶ)など70ヶを超えるヒット化合物を得た。このうち構造上、前者では少なくとも2ヶ、後者では16ヶについて新規性が高く産業上有用と考えられ、また、横浜市大との共同研究によるインフルエンザウイルスRNAポリメラーゼPA-PB1複合体阻害剤として3ヶの化合物の発見など、将来有用となるもの、有用になりうる化合物を、総計20個以上を得ることができた。

→ 目標を達成できた。

1) タンパク質のドッキング効率を上げる新たな手法や、分子動力学計算に基づく結合自由エネルギー計算の新規な方法を開発した。また、 $\mu$ オピオイド受容体に対する*in-silico*スクリーニング法によって200万化合物から選択された399化合物に対して、ヒット化合物48ヶ(12%)が見出され、また膜タンパク質に対する新規阻害剤が187ヶの評価化合物中23ヶ(12.3%)見出された。これらのヒット率は、通常の0.1%~1%のオーダーのヒット率に比較すると、少なくとも10倍以上のスクリーニング効率となっている。

→ 目標を達成できた。

2) データベースエンリッチメント・カーブにおいて上位1%における選択的なヒット率は、ssDSM法では通常のドッキング手法の5.1倍(45.8/8.9%)、標的タンパク質の動的モデルによるMTS法ではGOLDに比べ20倍(66/3.3)、機械学習法とMTS & DSIを組み合わせた方法ではS社のドッキングソフトに比べて18.0倍(22.1/1.2)を得た。GOLDも後者のソフトも国際的に広く利用されている標準的な

ッキング・ソフトウェアである。一方、不整脈等の原因となるhERGチャネル阻害活性予測を行う新規手法の開発においても、市販プログラムに搭載されている従来法の7.9倍の選択性を得た。

→ 目標を達成できた。

3) 類似化合物探索手法としてMD-MVO法を開発したところ、代表的なソフトウェアROCSより有意に高いヒット率を示し、ペプチド・リガンドからの低分子探索が可能であることを示した。

→ 目標を達成できた。

## (1) in silicoドッキング計算の高精度化

### 【研究内容】

創薬プロセスにおけるin-silico ドッキング計算において、タンパク質の動的性質を正しく評価するため、動的性質を抽出する手法を開発するとともに、ドッキングスコアの精度を高め、タンパク質および低分子リガンドの動的構造、及び各種相互作用を考慮した複合体の構造予測法及び結合エネルギー算出法の開発を行う。

### 【研究成果】

(i) ドッキングスクリーニング法の改良: MTS 法の改良 [BIRC 集中研]

我々は今までに、蛋白質—化合物ドッキングソフト、複数の in silico スクリーニング手法を開発してきた。これらスクリーニング手法は、蛋白質—化合物相互作用行列に基づく手法であり、標的蛋白質の3D 構造をもとにした Multiple target screening method (MTS 法)、標的の構造を用いずに、活性化化合物の情報のみを用いる Docking score index method (DSI 法) である。ドッキングスコアの精度そのものを改善する手法も提案してきた。1つは、蛋白質—化合物相互作用行列に基づく手法である Direct score modification method (DSM 法)、もうひとつは、活性化化合物の情報を用いる Machine-learning score modification method (MSM 法) である。

標的蛋白質の3D 構造が既知であり、かつ活性化化合物の情報もあれば、Machine-learning score modification method と MTS 法を組み合わせることで、安定して高いヒット率を出すことができる (MSM-MTS 法)。しかし、活性化化合物の情報がない場合、Direct score modification method と MTS 法を組み合わせることになるが (DSM-MTS 法)、標的蛋白質の種類によってヒット率にばらつきがあり、またヒット率は、MSM-MTS 法に比べてかなり低いのが問題だった。

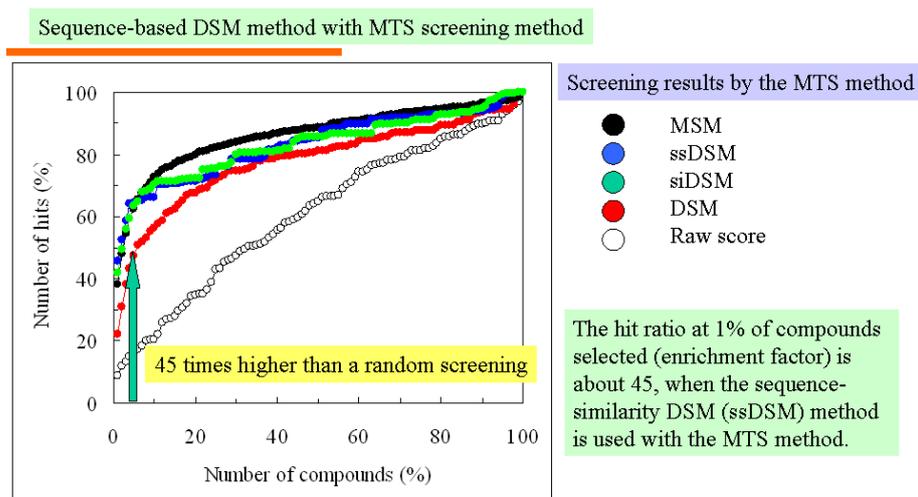
### 【sequence-based DSM 法の開発】

本年度では、DSM-MTS 法の欠点を修正し、標的蛋白質の3D 構造が既知である場合、従来より高いヒット率を、標的蛋白質の種類に関わらずに安定して高いヒット率を出すことを目指す。スコアの修正を下記の式で行う。つまり蛋白質 a、化合物 i のドッキングスコア  $s_a^{\text{mod } i}$  は、蛋白質 b、化合物 i のドッキングスコア  $s_b^i$  の線形結合で近似する。 $M_a^b$  は、パラメーターである。

$$s_a^{\text{mod } i} = \sum_b s_b^i M_a^b \quad \text{式(1)}$$



より不活性化化合物として収集した。180 蛋白質x約1万化合物の相互作用行列を SievGene を用いて作成し、ssDSM/siDSM、オリジナルの MTS,DSM-MTS, MSM-MTS を適用して、データベースエンリッチメント曲線を計算したのが下図である。MSM-MTS 法は、既知活性化化合物が発見される確率を最大化するようにパラメータを調整する手法なので、事実上の上限を与えている。ssDSM/siDSM は、オリジナルの MTS、DSM-MTS 法の結果を大きく上回り、MSM-MTS 法に近い結果を与えた。上位1%の化合物を採択した場合のヒット率は45%に達し、ヒット率を向上させる初期の目的は果たされている。



ただし、ssDSM/siDSM は、通常の MTS 法でヒット率が高かった標的に対し、より高いヒット率を示す反面、通常の MTS 法でヒット率が低かった標的に対しては、より低いヒット率を与えてしまう。平均すれば、ヒット率を完全しているが、標的による結果のばらつきが大きくなるのが欠点であることが分かったので、適用に際しては注意が必要である。

## (ii) 膜蛋白質の分子動力学計算による動的性質の抽出

〔阪大分室〕

創薬プロセスにおけるin-silico ドッキング計算において、タンパク質の動的性質を正しく評価し動的性質を抽出するため、分子動力学計算結果における蛋白質分子の各原子のトラジェクトリーから、主成分解析(PCA: Principal Components Analysis)を行って動的構造を解析する手法を開発・確立した。また、主鎖原子間をバネでつなぐモデルである弾性ネットワーク・モデル(ENM: Elastic Network Model)についても、その解析法を開発し、上記PCA解析と同様に利用できる仕組みとした。

まず、膜蛋白質であるモノアミン酸化酵素-A(MAO-A)の膜分子と水分子をあわらに取り扱った計81,647原子からなる系(水分子:17125, POPC

(1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) リン脂質分子: 163, Na<sup>+</sup>: 76, Cl<sup>-</sup>: 78)に対し、myPresto/cosgeneを用いてNPTアンサンブル(Nosé-Anderson法)による1気圧、310Kでの分子動力学計算を、1 step = 2 fs として、20 nsの長さで3回初期速度を変えて行った(周期境界条件: 72 Å x 94 Å x 124 Å)。同様に、膜貫通ヘリックスを削除し、膜分子のない水溶液系(計

63,605原子, 周期境界条件: 72 Å x 92 Å x 96 Å)も同時に3回の分子動力学計算を行い、トータル6ヶのトラジェクトリーを比較・解析した。

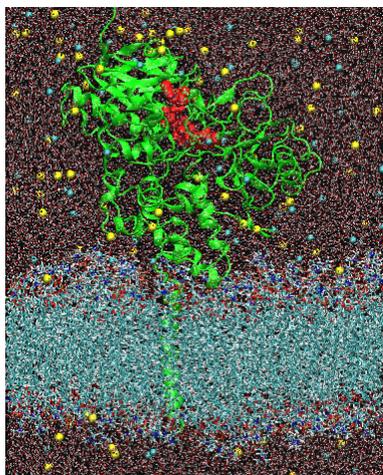


図1: 膜蛋白質 MAO-A に対するシミュレーション計算の系。MAO-A (緑のリボン) と、膜分子 (POPC:水色)、溶媒の水 (赤)、Na<sup>+</sup>イオン (黄色)、Cl<sup>-</sup>イオン (水色の球)。

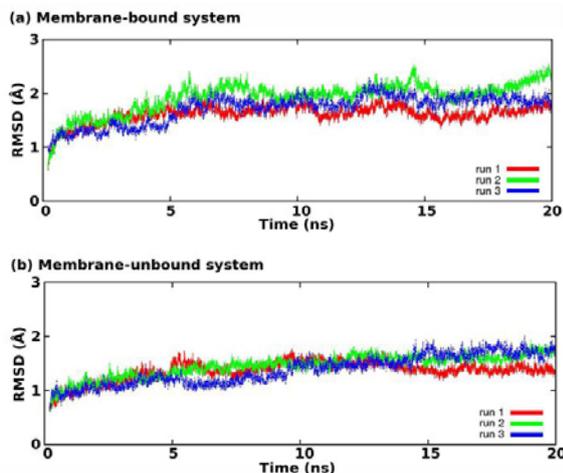


図2: (a)膜貫通ヘリックスがある場合のトラジェクトリにおける初期構造からのズレ。(b)水溶液系のシステムにおけるトラジェクトリ。

この時、特に、MAO-Aの内部のドメイン構造の動的性質を明らかにするため、ゆらぎの相関を

$$Corr_{ij} = \frac{\langle (\vec{r}_i - \langle \vec{r}_i \rangle) \cdot (\vec{r}_j - \langle \vec{r}_j \rangle) \rangle}{\sqrt{\langle (\vec{r}_i - \langle \vec{r}_i \rangle)^2 \rangle \langle (\vec{r}_j - \langle \vec{r}_j \rangle)^2 \rangle}} \quad (\vec{r}_i: i \text{ 番目の残基の } C_{\alpha} \text{ 原子の座標})$$

と定義して、全ての60 ns ずつのトラジェクトリを計算し、膜貫通ヘリックスが膜にアンカーしている状態のもの(左上)と、膜貫通ヘリックスを含まず水溶液中のもの(右下)とを比較した。

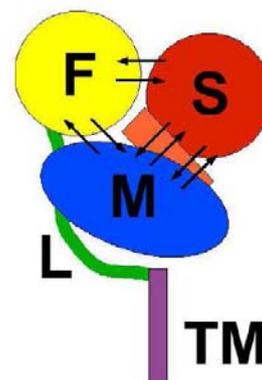
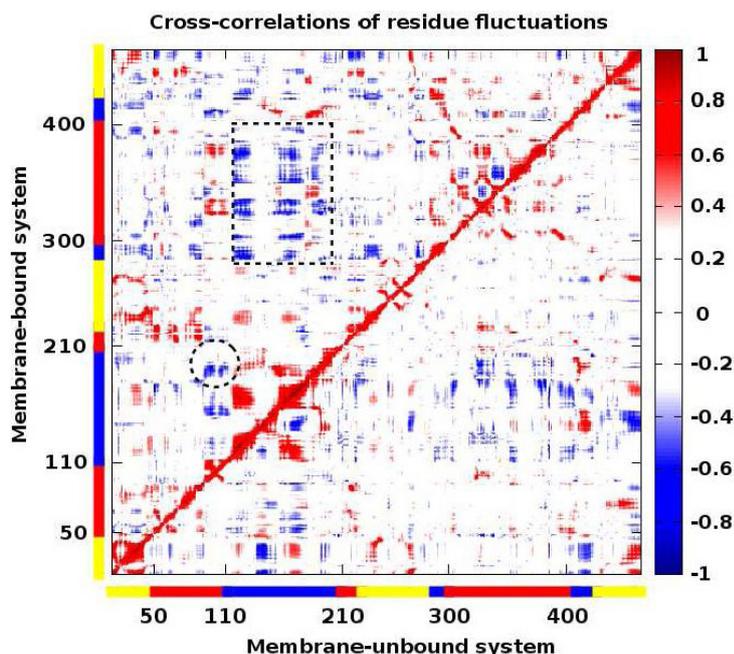


図3:  $i$  番目の残基と  $j$  番目の残基お  $C_{\alpha}$  原子間のゆらぎの相関  $Corr_{ij}$ 。左上は膜貫通ヘリックスを含む系であり、右下は水溶液系。残基部分の色は、右の概略図におけるドメインに対応する。オレンジ色はリガンド結合部位。

明らかに、点線で囲む負の相関(青色で示される領域)が、膜貫通ヘリックスを持つ場合には明瞭であるのに対し、水溶液中のトラジェクトリーには見られないのがわかる。これらは、図3で示すSDメインとMDメイン間の負の相関であり、リガンド結合部位の開閉が、膜貫通ヘリックスがある場合には拡大されることがわかる。

さらに、どのような運動かを詳細にみるため、主成分解析とENM解析とを比較した(図4)。

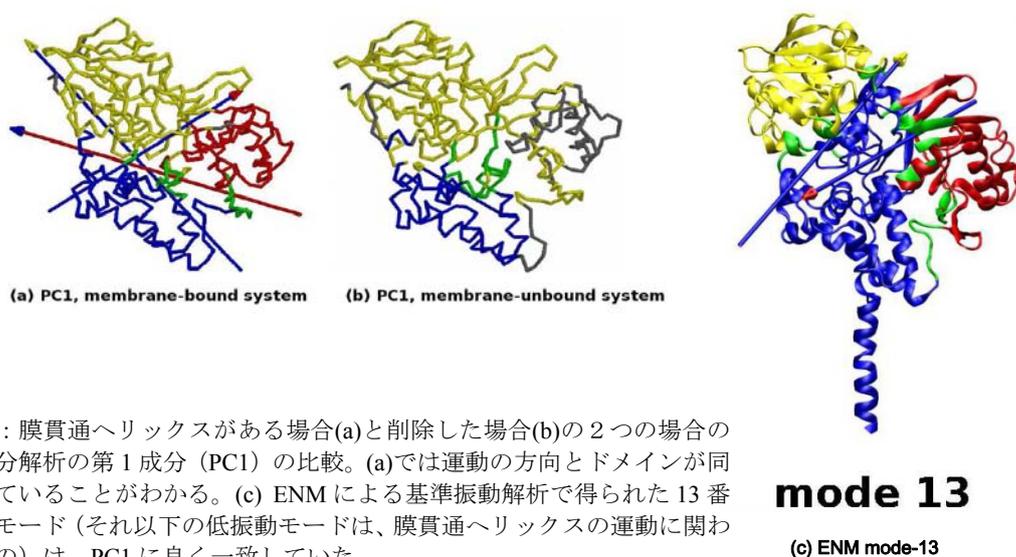


図4：膜貫通ヘリックスがある場合(a)と削除した場合(b)の2つの場合の主成分解析の第1成分(PC1)の比較。(a)では運動の方向とドメインが同期していることがわかる。(c) ENMによる基準振動解析で得られた13番目のモード(それ以下の低振動モードは、膜貫通ヘリックスの運動に関わるもの)は、PC1に良く一致していた。

ENM解析で得られる基準振動で、主成分解析で得られる主な運動が再現されていることがわかり、分子動力学計算と主成分解析によって得られる振動モードは、この蛋白質分子がもともと持っている運動であることがわかる。すなわち、リガンド結合に関係すると思われるドメインの開閉運動は、もともとのMAO-A分子が持っている振動モードであったが、膜貫通ヘリックスが膜にアンカーリングすることによって、テコのように働き、このドメイン間の開閉運動がさらに強調されたものと考えられる。これから、図5のようなリガンドの入出に伴う道筋(トンネル)が推測される。特に右側のピンク色のトンネルについては、リガンド選択性をつかさどるアミノ酸がそのトンネル中に位置することから、この推測の正当性が示唆されている。

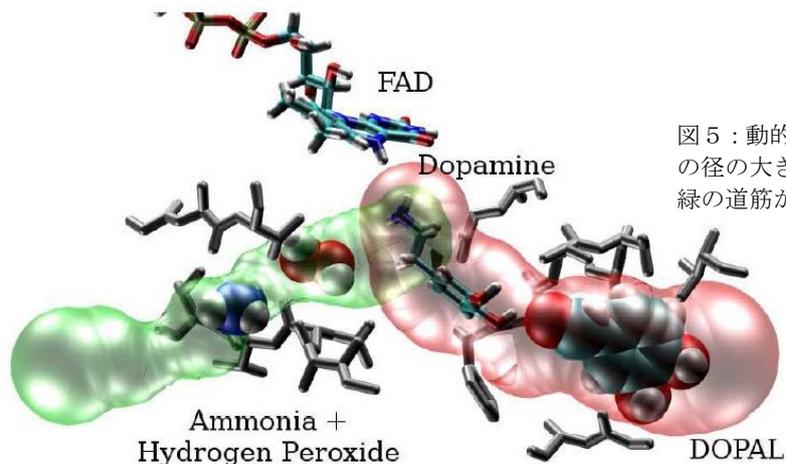


図5：動的なリガンド結合の道筋。トンネルの径の大きなピンクの道筋と、若干小さい緑の道筋が可能なものとして考えられる。

以上、MAO-Aの静止状態の構造からはリガンド結合の道筋が不明であったのが、このようなシミ

ュレーション計算によって、ダイナミックなリガンド結合の仕組みを明らかにすることができた。  
(Apostolov et al. (2009) *Biochemistry*, 48, 5864–5873)

### (iii) 新しいリガンド結合エネルギー計算手法 (Smooth Reaction Path Generation 法) の開発

[BIRC 集中研]

医薬分子の標的への結合活性は、結合自由エネルギーで測ることができる。医薬品の最高活性は、結合自由エネルギーで最大 15 kcal/mol 程度であるとされている。通常、医薬品の平均活性は  $IC_{50} = 30$  nM であり、70% の化合物が  $IC_{50}$  が 50 nM より強い活性をしめす。 $IC_{50}$  が nano-, pico-M である場合、結合自由エネルギーは 12.6 – 16.8 kcal/mol となる。

結合自由エネルギー ( $\Delta G$ ) の計算は、計算化学の究極の目的の一つだが、計算は難しい。蛋白質—化合物ドッキングソフトによる見積もりでは、様々なソフトが開発されてきたものの、実測  $\Delta G$  と計算値の相関は、相関係数 0.6~0.7、平均誤差 2.5 kcal/mol となる。この誤差は、ヒット化合物 (シードレベル) の  $\Delta G$  に相当するエネルギーで、その結果、計算によるヒット化合物探索の精度は低い。MMPBSA、MMGBSA、LIE (Linear Interaction Energy)、COMBINE 法などいろいろな半経験的  $\Delta G$  推算方法が開発されてきたが、精度がドッキングソフトに比べて飛躍的に高いわけでもなく、既知活性化合物の  $\Delta G$  がないと推算ができなかったりする欠点があった。 $\Delta G$  の推算誤差を 1 kcal/mol 程度にすることができれば、活性の推算は、従来の 10 倍程度に改善されたといえるので、この値を最終目標とする。なお、実験での pH、塩濃度によって活性値は 10 倍程度変化するので、誤差 1kcal/mol 以上の精度は、期待できない。

分子動力学シミュレーションで  $\Delta G$  が計算できれば、既知の  $\Delta G$  を必要とせず、理想的であるが、その計算は、極めて計算量が多く、困難である。我々は、2003 年に filling potential (FP) 法を開発した。この手法では、蛋白質—化合物複合体から、人為的なポテンシャルを用いて、化合物を解離させ、その経路に沿った自由エネルギー面を計算することで、少ない計算時間で  $\Delta G$  を求める方法である。この FP 法を用いた全原子シミュレーションによって、酵素—阻害剤の  $\Delta G$  の高精度の計算に成功した。しかし、FP 法は、化合物が水中ではなかなか解離しない、ヒストグラム法という計算方法を使うため、慎重な計算が必要で、計算 5 回~10 回に 1 回しか成功しないという扱いの難しい方法だった。そこで、我々は、より簡便に、高い確率で  $\Delta G$  が計算できる方法を開発することにした。

#### 開発した計算手法: Smooth Reaction Path Generation (SRPG) 法

蛋白質と化合物の複合体構造を最初に準備し、蛋白質から化合物を、真空中で簡単に解離させ、その解離の経路をルジャンドル多項式を用いて、滑らかな経路で近似する。この滑らかな経路に沿って、化合物が様々な位置に存在する状態を作り、タンパク質、水、イオンと合わせて系を作り、各系で MD を行い、リガンドの対象原子に働く平均力を計算する。滑らかな経路に沿って、力を積分することで自由エネルギー面を計算し、結合状態と解離状態の存在確率を求めて、その比率から  $\Delta G$  を計算する。各状態をパラレルに計算することができることで計算時間を大き

く短縮させることができるという利点がある。

SRPG 法による  $\Delta G$  計算では、

$$\Delta G = -k_B T \ln \frac{P_B}{P_U} \quad \text{eq. 1}$$

結合状態の規格化されていない確率  $P_B$  と非結合状態の規格化されていない確率  $P_U$  を用いて  $\Delta G$  を計算する。このとき、

$$P_B = \int_{R_B} \exp(-\beta G(r)) dr \quad \text{eq. 2}$$

$$P_U = \int_{R_U} \exp(-\beta G(r)) dr \quad \text{eq. 3}$$

ここで  $G(r)$ ,  $\cdot$ ,  $R_B$ , and  $R_U$  は、座標  $r$  での自由エネルギー(PMF),  $1/k_B T$ , リガンドの結合状態の領域、リガンドの非結合状態の領域である。

もし、結合状態の確率  $P_B$  が、PMFの最安定座標の周辺( $r_0$ ),に限定され、 $G(r)$  がパラボラポテンシャルで式4のように近似されるなら、

$$G(\vec{r}) = G(\vec{r}_0) + \left( \frac{k_x}{2} \Delta x^2 + \frac{k_y}{2} \Delta y^2 + \frac{k_z}{2} \Delta z^2 \right) \quad \text{eq. 4}$$

ここで

$$\vec{r} = \vec{r}_0 + \vec{\Delta r}, \quad \vec{\Delta r} = (\Delta x, \Delta y, \Delta z) \quad \text{eq. 5}$$

結合状態の確率  $P_B$  は、式6で与えられる。

$$P_B = \int_{V_B} \exp\left(-\beta\left(G(r_0) + \frac{k_x}{2} x^2 + \frac{k_y}{2} y^2 + \frac{k_z}{2} z^2\right)\right) dx dy dz \quad \text{eq. 6}$$

PMF の原点を解離状態に置く( $G(\infty) = 0$ ). もし  $G(r)$  の値が、0 kcal/mol に近いなら、

$$P_U = \int_0^R 4\pi r^2 \exp(-\beta G(r_\infty)) dr = \frac{4\pi}{3} R^3 \exp(-\beta G(r_\infty)) = V_0 \exp(-\beta G(r_\infty)) \quad \text{eq. 7}$$

化合物の密度が 1M density だと、1 個の化合物分子の占める体積  $V_0$  は  $1661 \text{ \AA}^3$  となる。

以上、まとめると、 $\Delta G$  は式8で与えられる。

$$\Delta G = G(r_0) - G(r_\infty) - k_B T \ln \frac{\int_{-\infty}^{\infty} \exp\left(-\beta \frac{k_x}{2} x^2\right) dx \int_{-\infty}^{\infty} \exp\left(-\beta \frac{k_y}{2} y^2\right) dy \int_{-\infty}^{\infty} \exp\left(-\beta \frac{k_z}{2} z^2\right) dz}{V_0} \quad \text{eq. 8}$$

式8の第3項の分子は、ガウス積分なので、解析的に解くことができる。

$$\Delta G = G(r_0) - G(r_\infty) - k_B T \ln \frac{\sqrt{\pi}}{\sqrt{\beta k_x / 2}} \frac{\sqrt{\pi}}{\sqrt{\beta k_y / 2}} \frac{\sqrt{\pi}}{\sqrt{\beta k_z / 2}} \quad \text{eq.9}$$

1661

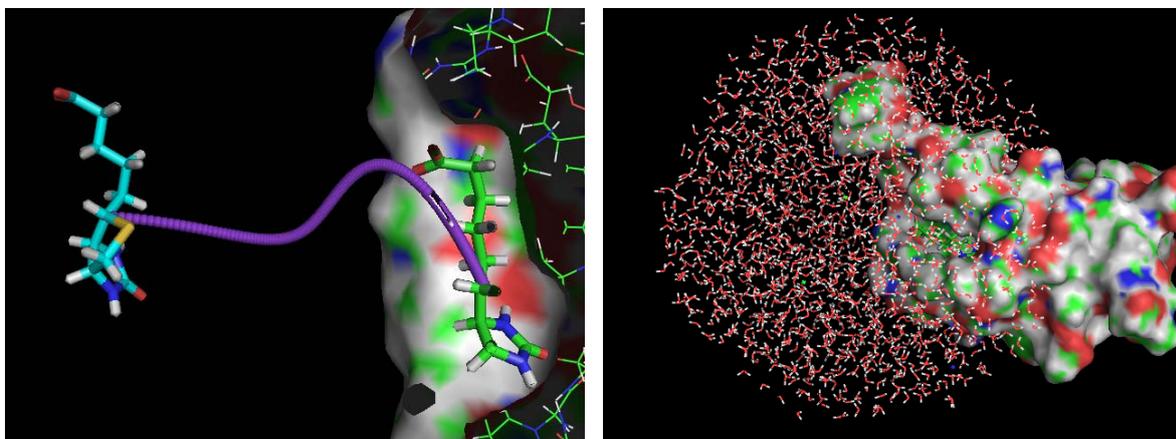
解離の経路  $p(t)$  は、ルジャンドル多項式として、

$$\begin{cases} p_x(t) = \sum_{i=0}^L c_x^i P_i(t) \\ p_y(t) = \sum_{i=0}^L c_y^i P_i(t) \\ p_z(t) = \sum_{i=0}^L c_z^i P_i(t) \end{cases} \quad \text{eq. 10}$$

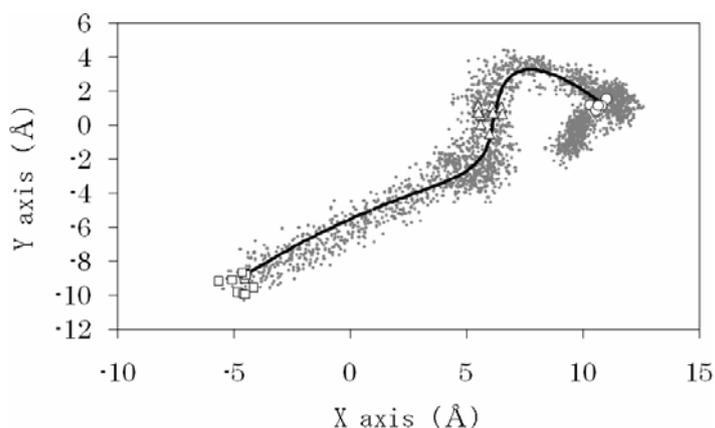
FP法で解離させた経路  $p^0$  に対して、 $S$  を最小化するように、係数  $c$  をモンテカルロ計算で最適化するように決定した。

$$S = \sum_m^M \sum_n^N D(\vec{p}(n/N), \vec{p}^0(m))^2 \quad \text{eq. 11}$$

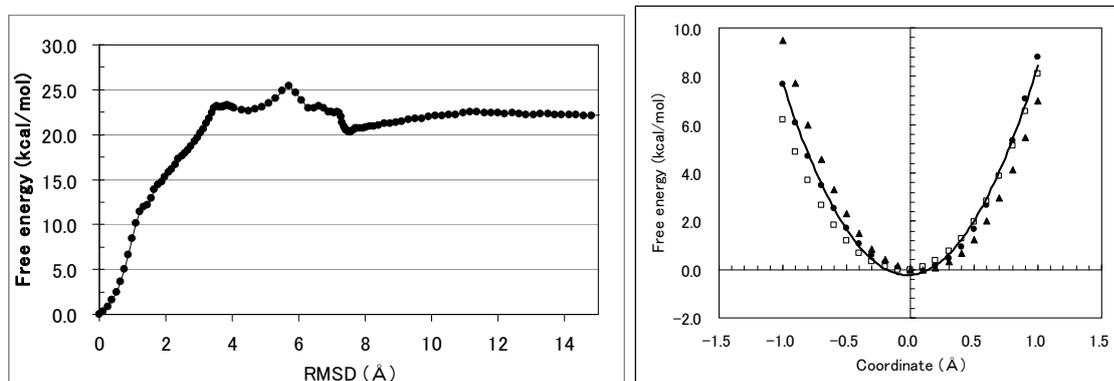
蛋白質—化合物の  $\Delta G$  の計算: ストレプトアビジンとビオチンの系について計算を行った。



図にストレプトアビジンからビオチンが解離していく経路を示す。上図左のように、真空中で解離させ、右のように、後から溶媒水を加えて、経路にそって、150 個の系を作成した。



FP 法によって、生成した解離していくリガンドの座標を点で表し、それに対してルジャンドル多項式で生成した滑らかな経路を実線で示す。



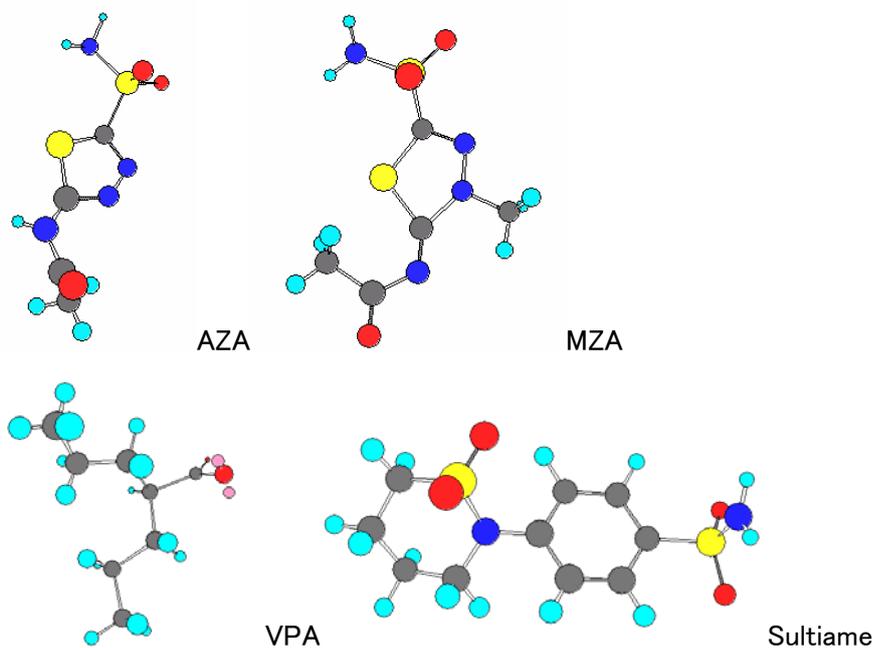
この経路に沿った自由エネルギー面は上左図のように得られた。リガンドの結合状態周辺の自由エネルギー面は、上右図のようであり、2次曲線で近似でき、理論が適用できることが示された。

	SRPG 法	実験値
$\Delta G(\text{kcal/mol})$	-16.5	-18.27

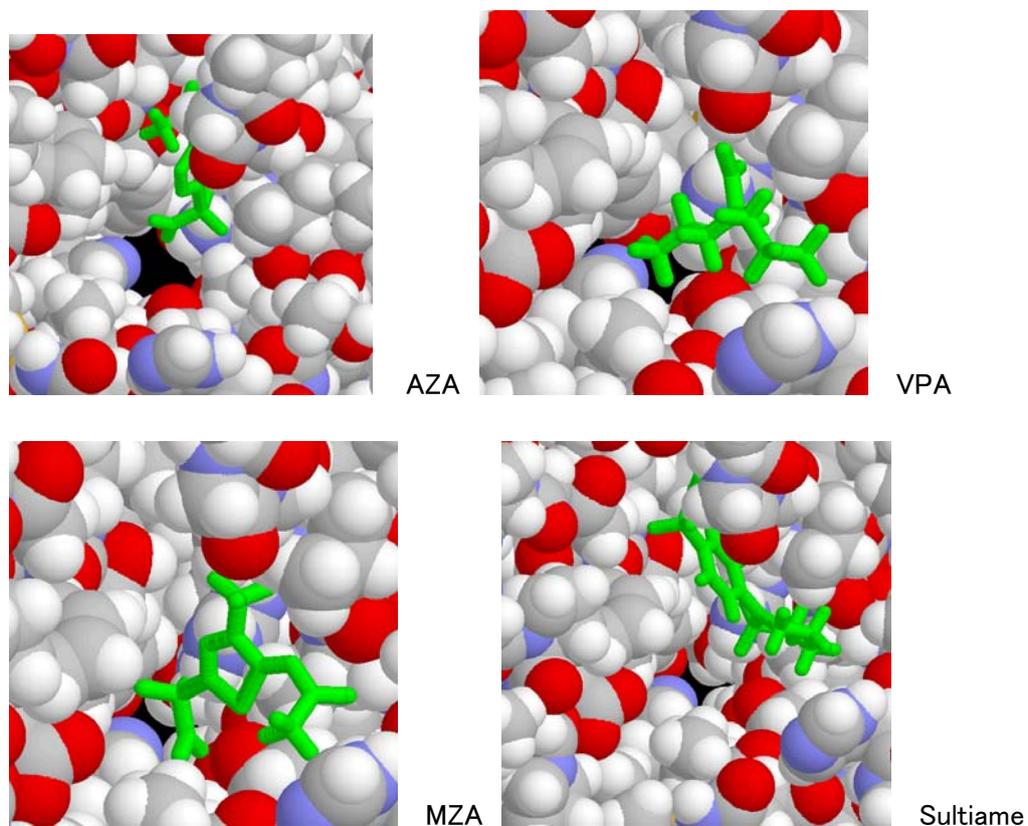
こうしてSRPG法では、誤差1.8kcal/mol(9.8%)という高い精度で、 $\Delta G$ を計算することができた。一例に過ぎないが、ドッキングソフトでの $\Delta G$ 平均誤差を1kcal/mol近く上回り、活性値で10倍の精度に達している、ともいえる。

#### アクアポーリン4(AQP4)への応用

藤吉チームで研究している AQP4 阻害剤探索に対して、SRPG法を適用した。試した化合物候補は、藤吉チームと同じ以下、4化合物である。



Sievgene でこれら4化合物を AQP4 へドッキングした。ドッキングポーズは、以下のものである。

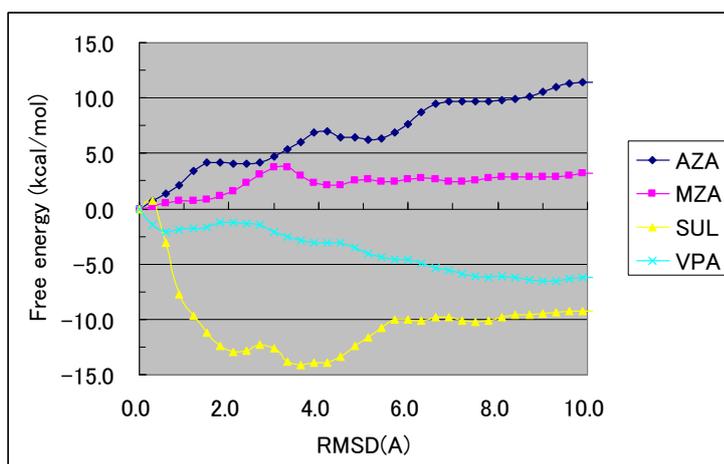


#### 水中でのMD

SRPG 法を適用した。2 nsec の水中でのMDを実施した。膜は作成せず、通常の球状蛋白質と同

じようにCAP水に入れて計算した。各化合物を(2)でドッキングした位置から、10 Å 引き離れた位置まで解離させることにする。0.3 Å 刻みに解離経路上に化合物を配置し、化合物に加わる力を計算することで、自由エネルギー一面を計算する。

#### 計算された自由エネルギー一面



自由エネルギー一面

自由エネルギー一面での結合状態と解離状態の自由エネルギー差から約 5kcal/mol を差し引いた値が  $\Delta G$  に相当する。AZA は、結合しやすいと推測される。AZA と MZA は、分子構造ではメチル基1つの違いしかないが、自由エネルギー一面には大きな違いが現れた。Sultiame, VPA は、結合状態が不安定で、結合しえない可能性が高い。この傾向は、実験と一致することが示された。

#### (iv) McMD (マルチカノニカル分子動力学法) 計算法による蛋白質-阻害剤の結合自由エネルギー計算 [阪大分室]

結合自由エネルギーを高い精度で算出するため、溶媒をあらわな水分子モデルで扱った系における、長時間のマルチカノニカル計算によるドッキング・シミュレーション手法を開発した。この手法により、リゾチーム蛋白質とその阻害分子 (NAG (N-acetyl-D-glucosamin) trimer) との結合についての、ab-initio 的なシミュレーションを行った。643 の TIP3P 水分子を CAP 境界として部分的に付加し、マルチカノニカル分子動力学 (McMD) 計算法によって、700K から 300K までをカバーするアンサンブルを得た (図6)。この時、蛋白質側は距離拘束条件を付加することによって、高い温度でも大きな構造変化がおきないように工夫した (図7)。

最終的に 72 ns ( $7.2 \times 10^7$  steps) の McMD ランを行い、その後、300K、500K、700K のカノニカル・アンサンブルを再構築し、阻害剤 (NAG)<sub>3</sub> の中心位置の分布をプロットした (図8)。面白いことに、図8の (a) 700K や (b) 500K では、阻害剤の位置は蛋白質から離れているのに対し、(c) 低温 (300K) になると蛋白質に近い所の存在確率が大きくなっている。

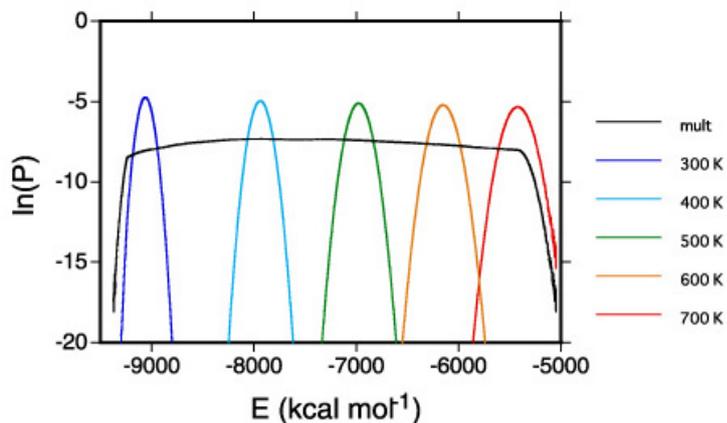
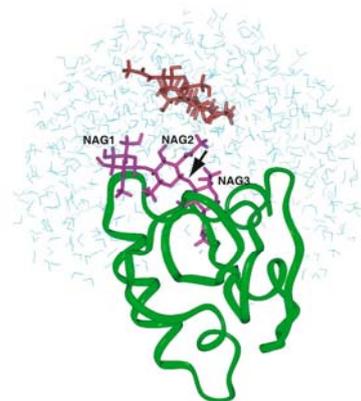


図6：MeMD アンサンブルと再構築したカノニカル・アンサンブルのエネルギー分布。



Kamiya et al. Figure 1

図7：対象としたリゾチームとその阻害剤((NAG)<sub>3</sub>)およびCAPの水分子。阻害剤は、初期の位置を赤で示し、紫は結晶構造。

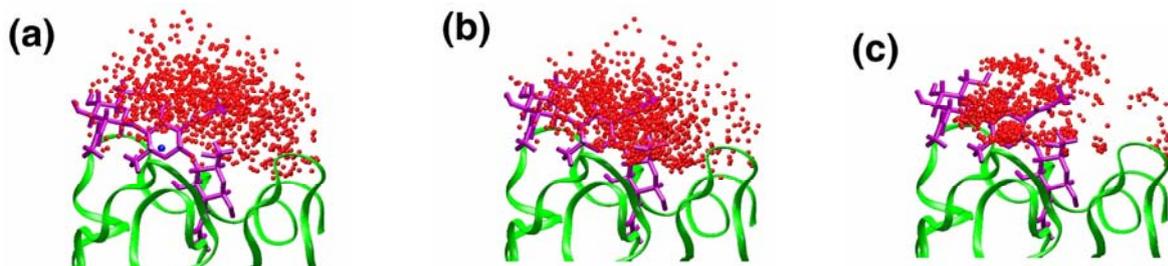


図8：(a) 700K, (b) 500K, (c) 300K として再構築した系のカノニカル・アンサンブル中の、リゾチーム(緑のリボンで示される主鎖構造)に対する阻害剤の重心の位置の分布。紫の棒モデルは、この阻害剤のリゾチームとの複合体のX線結晶構造中の構造を参照構造として示している。

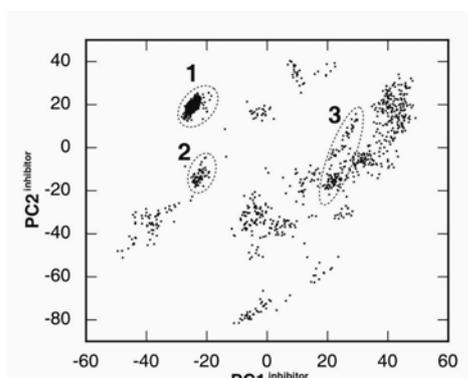
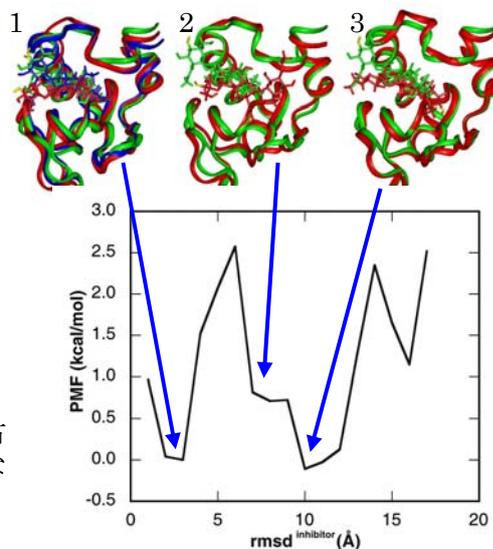


図9：自由エネルギー地形(左上と右下)と、リゾチームと阻害剤(NAG)<sub>3</sub>との複合体構造(右上)(緑：X線結晶構造、青と赤とは典型的なクラスター1, 2, 3中の構造)



このカノニカル・アンサンブルから、阻害剤の座標に対する主成分解析を行って、自由エネルギー地形を描いた(図9)。複数の結合様式の中で、天然に観測される結合状態が安定となることが理解された。(Kamiya et al. (2008) *PROTEINS*, 70, 41-53)

## (2) 構造生理学アプローチによるタンパク質間相互作用解析

### 【研究内容】

タンパク質間相互作用及び超分子複合体の構造情報まで含めた詳細な解析「構造インタラクトーム」に踏み込み、タンパク質間相互作用の阻害等の創薬において有用な機能を有するものの活性の維持等の観点から医薬品化が困難な生理活性ペプチドから、医薬品となりやすい非ペプチド性の低分子化合物等へ展開するため、ペプチドと同様あるいはそれ以上の強い結合性を有する非ペプチド性化合物(低分子化合物等)を探索・設計する新しい手法の開発に着手する。

### 【研究成果】

(i) タンパク質複合体構造予測法の開発 [BIRC集中研、阪大分室]

タンパク質複合体構造予測法の開発を行い、複合体構造予測のブラインドコンテストであるCAPRI(Critical Assessment of Predicted Interactions)に参加してその性能を評価した。

構造プロテオミクスの発展に伴い、多くのタンパク質の立体構造が公開されるようになったが、それらからタンパク質の複合体構造を予測することは、タンパク質の機能を理解し制御するために重要であるが、尚困難な課題の一つである。

タンパク質複合体構造の予測においては、1)タンパク質間相互作用部位の予測、2)候補構造の発生と評価、が重要なポイントである。我々は進化系統樹を利用した機能部位予測法であるET(Evolutionary Trace)法とタンパク質形状の相補性を利用した独自の構造探索エンジン・構造評価関数を組み合わせた手法を開発した。

この手法をCAPRIの問題に適用したところ、いくつかの例で、高精度の複合体構造をサンプルすることができた。

下図は予測結果の一例である、ARF-1とARF-BDの複合体の予測(T25)では、44番目の候補として、結晶構造とのRMSD=2.59 Åの構造が得られた。また、ToIBとPalの複合体の予測(T26)では14番目の候補として、結晶構造とのRMSD=2.55 Åの構造を得ることができた。

今後は相互作用部位予測法と評価関数の改良により、さらに高精度の予測を目指す。

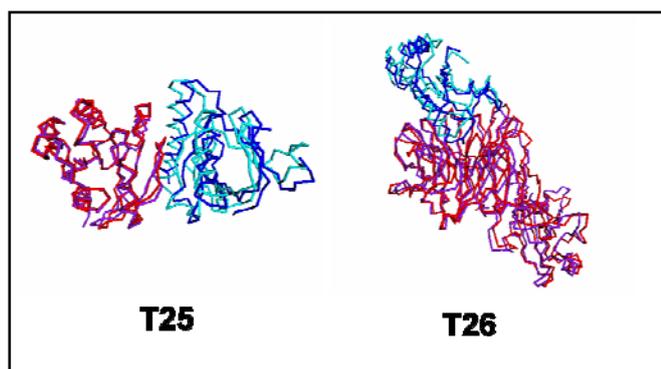


図) 予測結果の例

結晶構造のサブユニットを赤及び青、予測構造を紫及び薄青で示す。

## (ii) アミノ酸選択的交差飽和法の実験データを用いたタンパク質複合体構造の構築

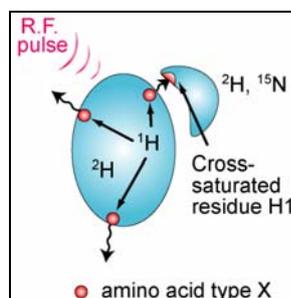
タンパク質-タンパク質間の相互作用は、シグナルの伝達など生体内で重要な役割を果たす。近年、タンパク質複合体の形成を妨げることで効果を発揮する医薬品の研究開発にも重点が置かれるようになってきている。

このような医薬品の開発において、タンパク質の複合体の立体構造を知ることは、薬剤を結合させる部位を決定するために重要であるが、X線結晶構造解析等の手法を用いて実験的に構造を決定することは、単体の状態の立体構造を決定するよりもさらに困難であることが知られている。

そこで今回、嶋田グループが開発した、NMRによる実験手法である、アミノ酸選択的交差飽和法(ASCS法)の実験データと、分子動力学計算とを組み合わせることにより、それぞれのタンパク質の単体の構造が既知の場合に、それらが形成するタンパク質複合体の立体構造を高精度に構築する方法を開発した。

## 方法

ASCS法は、一方のタンパク質の特定のアミノ酸種を標識して(交差飽和ドナー)、他方のタンパク質での交差飽和( $\eta$ )を観測する実験手法で(下図)、強い交差飽和が観測された残基の近傍に、標識したアミノ酸種が存在するという情報が得られる。



複数種のアミノ酸を標識したデータを組み合わせることで、ドナー側の残基とアクセプター側の残基

の位置関係についてある程度特定はできるが、原子レベルでの立体構造を構築することはできない。

そこで、この実験データに分子動力学計算を組み合わせることで複合体構造を構築することを試みた。実験データを再現するほどエネルギーが低下する、次の疑似エネルギー項を加えた分子動力学計算を行う。

$$E_{sat} = \frac{1}{2} W_{sat} \sum_{all\ exp.} \sum_{acceptor\ residues} (\eta_{calc} - \eta_{exp.})^2$$

ここで  $W_{sat}$ : 重み、 $\eta_{calc}$ : 計算値、 $\eta_{exp.}$ : 実験値である。

## 結果

(a) Ubiquitin(Ub)とYUHの複合体構造構築における本手法の検証

計算には嶋田グループで実施した5種類のアミノ酸を標識した実験データを用いた。

当該タンパク質複合体の立体構造は、PDBコード:1cmxとして結晶構造が知られているものであり、UbとYUHを分離した状態から、本方法を用いて結晶構造と同様の構造を得ることを試みた。

### 初期構造

交差飽和が観測された残基と交差飽和源の残基が近接するという条件から、初期構造(2つのタンパク質の相対位置)を作成した。可能性の高い42構造を初期構造として各初期構造につき2例ずつ(初期速度の乱数が異なる)分子動力学計算を実行した。

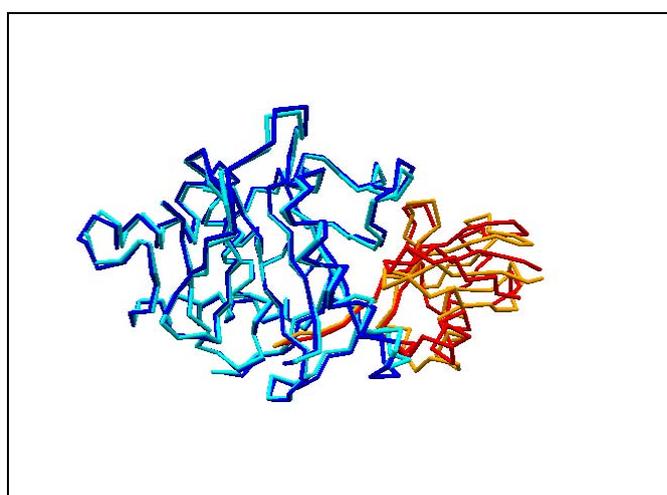
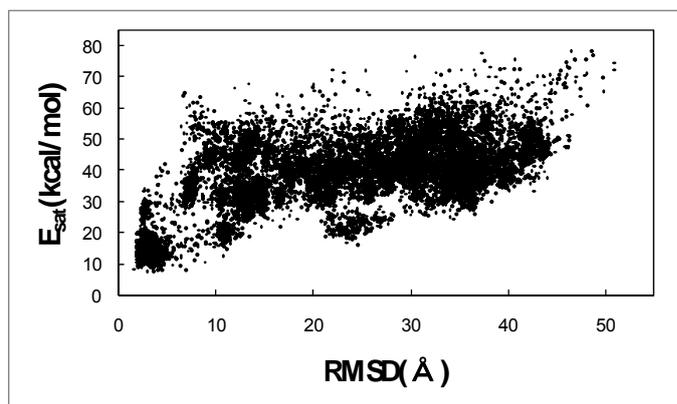
### 計算条件

分子動力学の計算条件を下表のとおりとし、真空中で構造を維持するためにUbとYUHの双方に、原子間の距離拘束を適用した。

time step	1.0 fsec
shake	H only
time	150 psec
温度コントロール	500K→100K (150psecで一定の割合で冷却)
誘電率	4.0R
カットオフ長	20 Å(テーブル更新 10step)

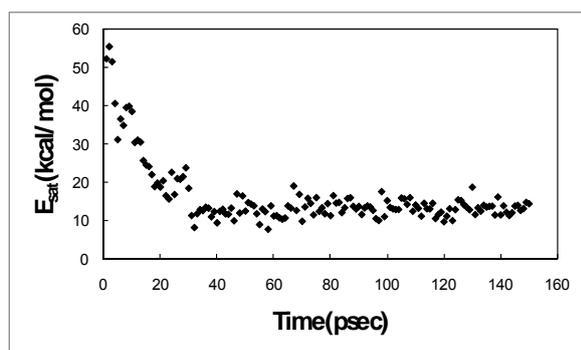
42 x 2個の計算のうち最終構造のRMSD(結晶構造とYUH側で重ね合わせた際のUb側の主鎖のRMSD。以下同じ)が5 Å以下の構造となったものは9個であった。

42x2個の分子動力学計算中に、サンプルした構造のRMSDとEsatをプロットすると、下図のようになり、Esatが小さい構造(実験データをよく説明する構造)を選択するとRMSDが小さい構造を選択できることがわかった。

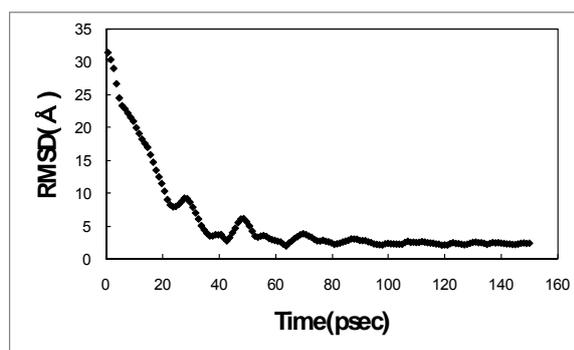


Esat の最小構造（青：YUH1、赤：Ub）（ $E_{\text{sat}}=7.88\text{kcal/mol}$ 、 $\text{RMSD}=3.05\text{Å}$ ）、参照構造（シアン：YUH1、オレンジ：Ub）

また本構造をサンプルした際の  $E_{\text{sat}}$  と RMSD の時間変化は以下の通りであった。

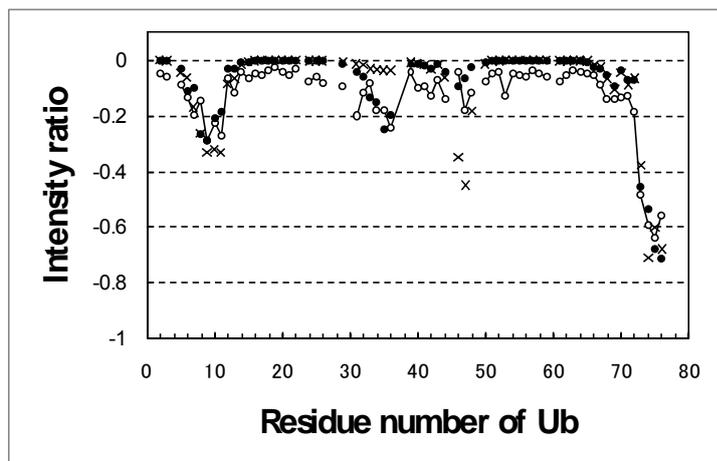


Esat の時間変化



RMSD の時間変化

Esat 最小構造での  $\eta$  の計算値と実験値を比較すると下図 (LEU 標識データ) のようになり、 $\eta_{\text{exp}}$  と  $\eta_{\text{calc}}$  (Esat 最小構造) が十分一致していることがわかる。



○・実線： $\eta_{\text{exp}}$ 、●： $\eta_{\text{calc}}$  (Esat 最小構造)、×： $\eta_{\text{calc}}$  (結晶構造)

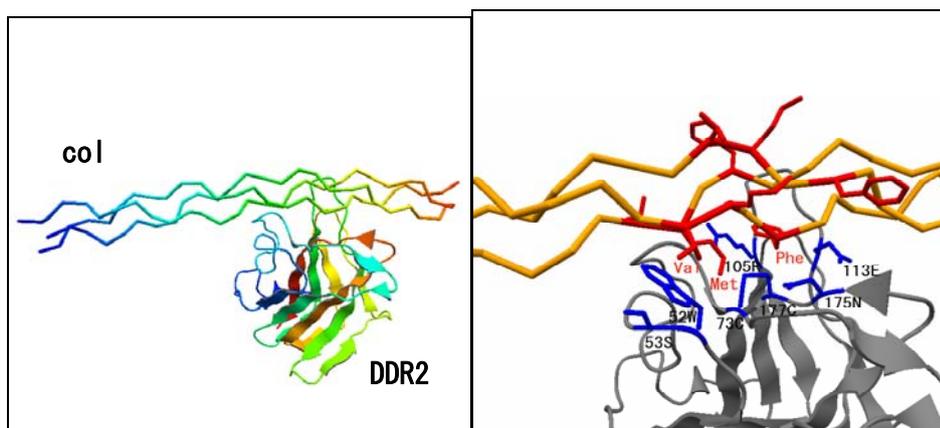
さらに、本構造を CAP 水中で Esat 項を外して再度 simulated annealing を行うと、RMSD は、0.67 Å まで改善される。これは、真空中で Esat 項を用いて計算したことにより生じた構造のゆがみが解消されたためと考えられる。

#### (b) DDR2(Disocoidin domain receptor2)-コラーゲンペプチドの複合体モデルの構築

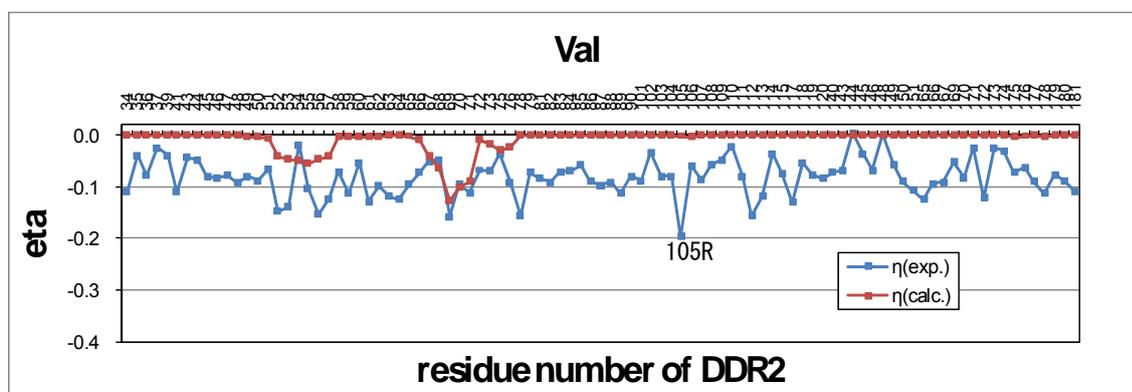
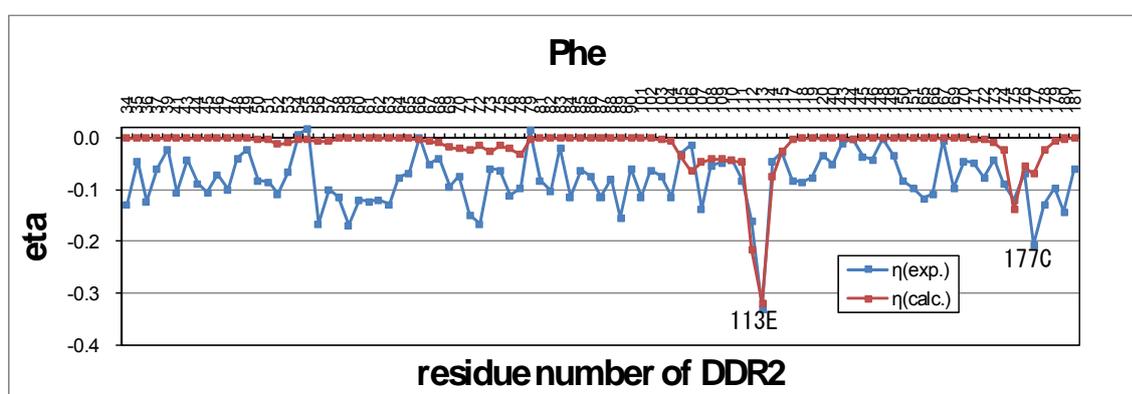
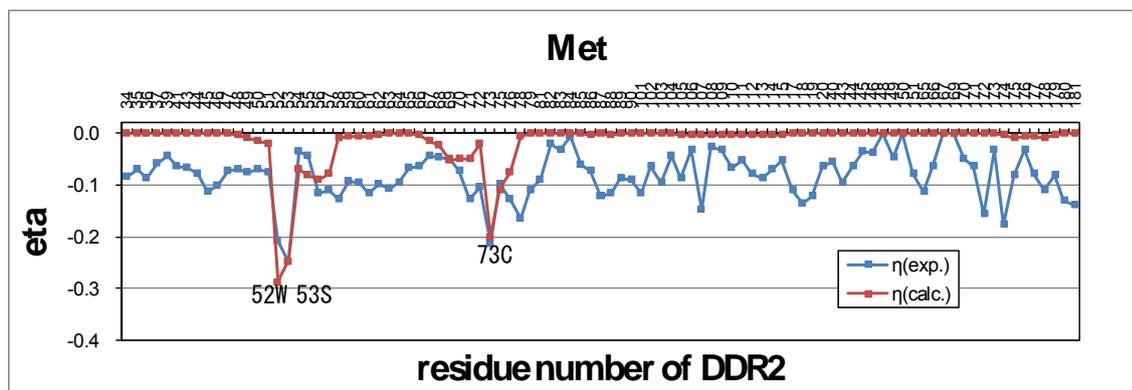
それぞれの単体構造として、DDR2 については嶋田チームが NMR で決定した (pdb code:2z4f) 20 構造を使用し、コラーゲンペプチドについてはテンプレートとして、PDB データベースに登録されているコラーゲンの X 線結晶構造 (pdb code:1k6f) に基づくホモロジー・モデルを用いた。

DDR2 の 20 構造に対して、我々が開発したドッキングソフト(上記「①タンパク質複合体構造予測法の開発」参照)を用いてコラーゲンペプチドをドッキングさせ、各 300 の複合体構造を作成した。嶋田チームがこの系に対して得た ASCS 実験の観測値を用い、この 20x300 構造から ASCS 実験結果と定性的に合致する 119 構造を選び、疑似エネルギー項 Esat を加えた分子動力学計算を行った。

119 個の計算でサンプルされた構造のうち、Esat が小さく磁化変化  $\eta$  が ASCS 実験値とよく一致する構造の 1 つが下図の構造であった。



本構造の磁化変化  $\eta$  の計算値は下図のようになり、主要なピークのほとんどについては実験値とよい整合性を示した。実験値と一致しない点については、側鎖の精密なモデリングの必要性や、複数の複合体ポーズの存在などが考えられる。



## 結論

以上の結果から、今回開発した ASCS の実験データを活用した分子動力学計算は、それぞれの単体の立体構造が既知の場合に、それらが形成するタンパク質複合体構造を高精度に構築する手法として、極めて有用であることが示唆された。これは、界面という2次元の情報を与える ASCS 実験データから、複合体という3次元の情報を得ることができることを示している。

本手法は、NMRの実験データを利用する方法であるため、複合体の結晶が得られない場合や、相互作用が弱い場合にも適用可能な手法である。また、ASCS 実験では、同位体ラベルした残基の界面情報しか得られないが、本手法では、複合体という3次元の情報を得ることで、ラベルされてい

ない残基も含めた蛋白質界面の情報を得ることができる。これにより、蛋白質複合体化を阻害する物質を設計する部位を探索することができる。

今回の研究は、嶋田チームの発表前の NMR 実験値を利用することができ、手法の検証に加えて新たな構造解析にも適用し、その有効性を示すことができた。チームを越えた協力研究としての成果を挙げる事ができたと考えている。

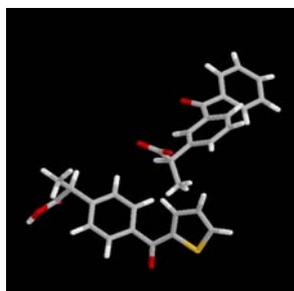
### (iii) ペプチドと同様の結合性を有する非ペプチド性化合物を探索・設計する新しい手法の開発

#### (a) 新規で高精度な類似化合物探索ソフトウェアの開発 [BIRC集中研]

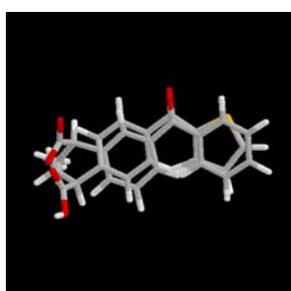
##### (a-1) molecular dynamics maximum volume overlap (MD-MVO)法—の開発と実施

類似化合物探索ソフト molecular dynamics maximum volume overlap (MD-MVO)法を、開発し薬物スクリーニングを開始した。

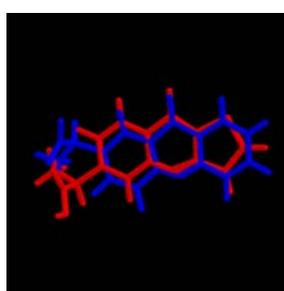
MD-MVO法とは、既知活性化合物に対して、データベースから選んだ化合物を、原子電荷も考慮しながら2化合物の体積重なりが最大になるようにMD計算によって、分子の配座探索と同時に分子重ね合わせをする手法である。この分子重ね合わせでは、周辺の環境である標的蛋白質を考慮した計算を行うこともできる。また、既知活性化合物を、標的蛋白質に結合していると期待されるもっともらしい構造に固定して、データベースから選んだ化合物のみを自由に運動させて2分子の重ね合わせを行うこともできる。



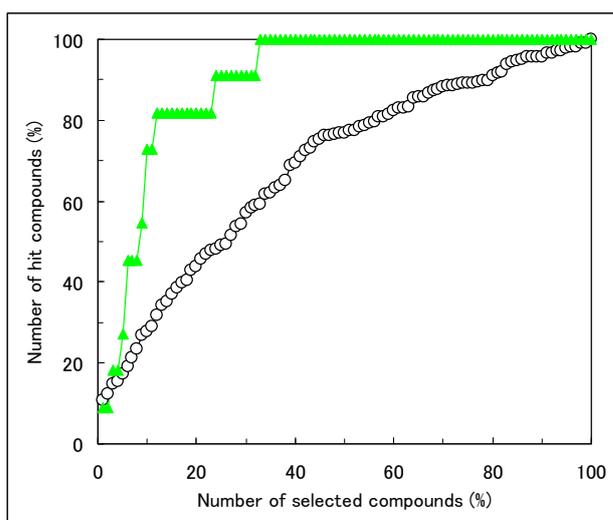
2分子の初期座標



MD-MVO法による重ね合わせの結果



$\mu$ オピオイド受容体のペプチド性リガンドであるエンドモルフィンを既知活性化合物として選択した。 $\mu$ オピオイド受容体について、既知の低分子リガンド(モルヒネ類など)を11種類文献より収集した。そして、エンドモルフィンを既知活性化合物として、1万化合物を含む化合物データベースに既知の低分子リガンド11種類を混ぜ、これらの既知の低分子リガンドを発見できるかテストしてみた。その結果が、下のデータベースエンリッチメントカーブである。我々が以前に開発した、ドッキングソフトを用いたligand-based drug screening法であるDocking score index (DSI)法と、今回のMD-MVO法の結果を比較した。



緑▲がMD-MVO法の結果、白○がDSI法の結果である。ペプチドを活性化化合物とするスクリーニングは困難であり、DSI法の結果がランダムスクリーニングに近くなっていることから、その困難さが伺える。明らかに、MD-MVO法はペプチドを活性化化合物とする場合も、優れたスクリーニング結果を示している。また、世界でもっとも広く使われている類似化合物探索ソフトROCSやChemFinderとの比較テストを複数の標的で行ったが、いずれもMD-MVO法がヒット率において、より優れていることが示された。

		MD-MVO 法		ROCS	
		AUC	ヒット率	AUC	ヒット率
COX2	1cx2	85.8	28.6	52.8	23.1
	1pxx	98.5	50.0	64.5	30.8
	3pgh	85.3	33.3	64.8	30.8
	4cox	74.0	33.3	68.5	15.4
HIV	1aid	59.8	5.0	29.2	4.5
	1bv7	93.3	40.0	75.8	36.4
	1hte	70.3	5.0	59.3	13.6
	1mes	88.2	35.0	77.0	36.4
	Average	81.9	28.8	61.5	23.9
	上位	MD-MVO	ChemFinder		
COX2	9%	38.20%	9.80%		
HIV	22%	51.40%	47.50%		

MD-MVO法とROCSによるデータベースエンリッチメントカーブの下の面積(AUC)と、上位1%化合物を採択したときのヒット率を表に示す。AUCが80%以上が緑色、ヒット率で30倍以上を空色で示

したが、MD-MVO法がROCSに対し優位であることが分かる。また、ChemFinderとの比較では、COX2阻害剤について上位9%、HIVプロテアーゼ1阻害剤に対し上位22%の化合物を採取したときのヒット率を表にしたが、やはりMD-MVO法はChemFinderに対しても優位性を示した。

MD-MVO法は、既知活性化合物とクエリー化合物の重ね合わせにおいて、周辺環境である標的蛋白質を考慮した計算を行うこともできることは、従来の構造重ね合わせ手法と全く異なる優れた点である。シクロオキシゲナーゼ2(COX-2)に、MD-MVO法を適用して、その効果を見てみた。まず、PDB ID:4coxのリガンドである活性化合物compound aと、非活性化合物であるcompound bを用意した。COX-2蛋白質としてPDB ID:1cx2の構造と、既知活性化合物として1cx2のリガンドを選んだ。この結果を表にまとめた。

1cx2に対してcompound a/bを我々の蛋白質—化合物ドッキングソフトsievgeneでドッキングし、スコア上位の構造5つを選んだ。これらのスコア上位5構造に対して、1cx2の構造を考慮したMD-MVO法を適用した。Compound aのMD-MVO法のスコア(MVOscore)は、compound bのスコア0.46より大きく、活性化合物aが、非活性化合物bより、予測活性が高いという合理的な結果になった。

また、化合物aのドッキング構造を正解の構造である4coxの構造に重ね合わせてみると、ドッキング予測構造(蛋白質—リガンド複合体構造)が、極めて正しく予測されることがわかった。Sievgeneによるドッキングでは、compound aの予測複合体構造の座標の誤差は、2~6 Åと大きかったが、MD-MVO法での最終座標では、正解の4cox結晶構造に0.2~1.5 Åとほぼ一致した。このように、標的蛋白質構造を環境として考慮することは、ドッキングソフトによる予測複合体構造の精度向上に貢献することが示された。

ドッキングスコアでの順位	compound a				compound b	
	ドッキングスコア	RMSD (ドッキング計算) Å	RMSD (MD-MVO) Å	MVOscore	ドッキングスコア	MVOscore
1	-3.57	2.25	1.06	0.47	-3.53	0.45
2	-3.50	2.40	0.26	0.45	-3.48	0.45
3	-3.43	1.97	0.99	0.45	-3.46	0.46
4	-3.38	2.59	1.13	0.44	-3.40	0.45
5	-3.34	6.78	1.56	0.40	-3.38	0.37

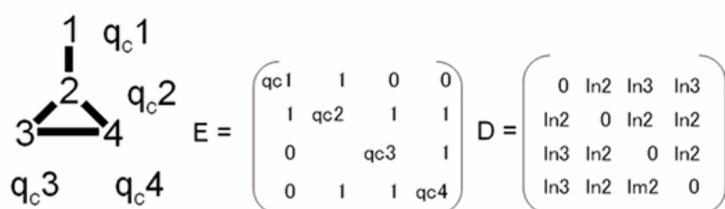
MD-MVO法は、MD計算の一種であるため、計算時間が長い。蛋白質を考慮しない状態でも1 CPUで、1日に約300化合物しかスクリーニングすることができない。従って、蛋白質を考慮しない状態で運用するしかない。現在保有する300コアを使用すると、2週間で100万化合物をスクリーニングすることができるため、 $\mu$ オピオイドに対して、エンドモルフィンを既知化合物としたスクリーニング計算を行い、塩野義製薬でアッセイを行っている。

## (a-2) 分子グラフを用いた高速な類似化合物探索法

〔BIRC集中研〕

MD-MVO法では、1CPUにて1日当たり300化合物しかスクリーニングできない。このままでは、応用範囲が極めて限られてしまうため、1時間当たり1000万分子以上をスクリーニングできる高速なスクリーニング手法を開発した。

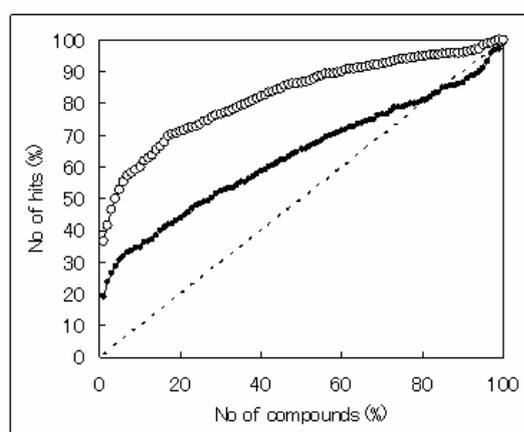
合成する代わりに類似化合物を取得する: 分子グラフを用いた類似化合物探索



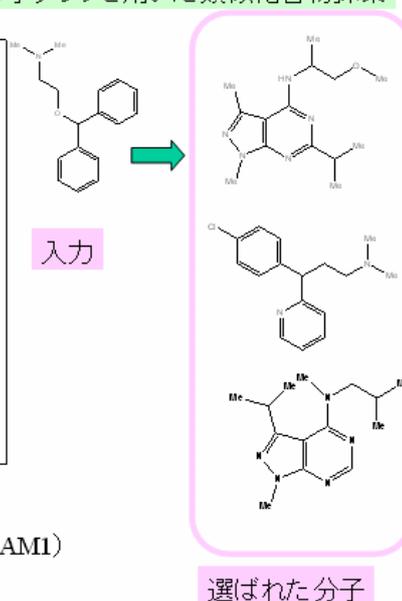
分子グラフのグラフ不変量(行、列の入れ替えに対して不変な量)を用いる。結合行列、距離行列の行列固有値を分子記述子とする。

この手法では、分子をグラフとして捕らえ、その結合行列と距離行列を作成し、この行列固有値をグラフ不変量とする。この固有値を幅を持たせたヒストグラムに変換し、2分子の比較の場合、2つのヒストグラムの重なりによって、分子の類似性を評価する。

合成する代わりに類似化合物を取得する: 分子グラフを用いた類似化合物探索



データベースエンリッチメントカーブ  
○: グラフ類似性探索、■: EEVA(MOPAC AMI)



この手法を実際に適用したのが、上の図である。グラフ類似性探索は、1時間1300万分子を探索するという簡便さのわりには良好なデータベースエンリッチメントを示し、実際、上位に選ばれた分子は、既知活性化合物と構造が似ていることが確かめられた。

(b) 蛋白質の動的構造を反映した薬物スクリーニングー GPCR を標的としたドッキングシミュレーション 〔BIRC 集中研〕

Gタンパク質共役受容体(GPCR)は重要な創薬標的であるが、その立体構造はほとんど解明されていない。そのため、GPCRを標的としてin-silicoスクリーニングを行うにはホモロジーモデリングにより標的タンパク質の立体構造を拘置する必要がある。また、induced-fitやタンパク質の動的挙動を考慮するには分子動力学(MD)計算が有効かもしれない。モデリング構造の信頼性やMDによる構造サンプリングの有用性を検証するために、まず、リガンド分子の構造活性相関が最もよく研究されているGPCRの一つであるヒト $\mu$ オピオイド受容体(MOR)のドッキングシミュレーションを行った。



$\mu$ オピオイド受容体(MOR)と $\beta_2$ アドレナリン受容体(ADRB2)の配列アライメント

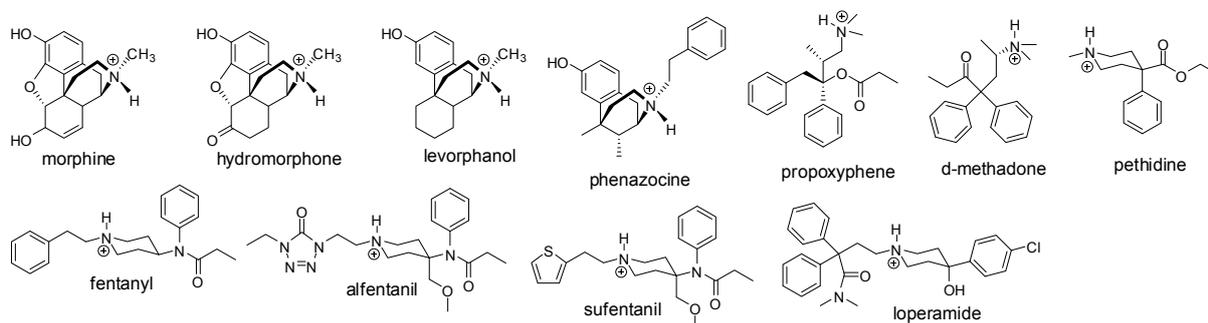
ヒト由来GPCRの $\beta_2$ アドレナリン受容体(PDB entry: 2rh1)をテンプレートとして、3種類のモデリングソフト(MOE, Modeller, Prime)を用いて、MORモデルを21構造作成した。

作成したホモロジーベースモデルそれぞれに対して周囲に水分子と電荷中和のためのイオンを配置し、エネルギー最小化と平衡化計算の後、周期境界条件で1n秒のNVT計算を実行し合計462種類のMOR構造をサンプリングした。



【図: MD計算により生成されたMORの構造】

ドッキングシミュレーション評価用化合物ライブラリは、Coelacanth社化合物ライブラリから抽出した18190化合物(デコイセットA)と3次元化合物ライブラリLigandBoxから抽出した11489化合物(デコイセットB)の2種類を準備した。また、MORの既知リガンド(アゴニスト)としてモルヒネやフェンタニルを含む11化合物を用意した。



μ オピオイド受容体の既知アゴニスト

462の標的MOR構造それぞれに対して、2つのデコイセット及び既知化合物のドッキング計算をSievGene/myPrestoで実行し、Multiple Target Screening(MTS)法による統計的スクリーニングを適用した。Area Under the Curve(AUC)として算出したヒット率の値を以下に示す。AUCとは、ヒット率(スクリーニングの精度)を定量的に表す指標で、0%~100%の範囲の値をとり、ランダムスクリーニングでは50%、理想的な高精度のスクリーニングならば100%に近い値を取り、逆に50%を下回る数字は、ランダムスクリーニングより悪い(すなわち計算しないほうがよい)結果を表す。

	MODELLER										MOE										PRIME
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	
initial	33.6	50.9	51.3	51.3	71.8	45.2	47.4	46.8	43.7	28.8	48.7	31.0	32.6	45.2	49.7	34.4	52.6	34.8	53.5	36.6	58.5
minimized	28.5	41.5	55.8	54.2	75.9	57.6	56.8	46.9	74.2	60.5	66.2	79.5	73.5	30.5	32.7	60.1	61.7	64.6	68.7	68.3	69.1
50 ps	68.3	60.5	47.4	52.9	70.6	51.6	67.5	71.8	69.5	44.7	66.3	33.6	65.8	36.9	18.6	50.4	62.2	74.9	75.0	35.7	69.4
100 ps	53.9	47.2	45.2	49.3	71.5	64.0	50.0	55.7	78.1	61.6	61.5	47.0	66.9	65.5	43.3	48.6	69.6	57.6	84.0	38.4	56.8
150 ps	50.9	47.8	45.0	73.8	63.4	43.5	60.0	69.7	86.5	67.4	70.5	59.4	70.6	70.8	28.6	44.3	86.5	60.1	72.6	42.4	57.8
200 ps	41.7	43.1	54.5	77.6	53.4	72.2	47.0	29.1	76.5	38.2	54.5	47.2	63.9	76.9	58.8	43.5	84.7	80.7	87.0	55.1	43.1
250 ps	37.2	39.1	54.0	76.0	57.6	42.5	60.1	43.7	70.9	49.3	64.9	48.8	63.8	51.7	33.9	38.9	80.3	58.0	59.5	64.5	52.8
300 ps	39.7	46.9	54.4	81.0	59.3	67.8	43.9	51.5	82.1	32.5	54.1	39.0	42.4	45.4	47.9	55.2	65.6	72.4	86.7	58.6	73.8
350 ps	53.1	59.9	52.5	70.2	63.8	51.8	45.2	57.5	70.0	41.4	75.4	42.1	77.0	55.2	64.9	62.4	83.0	60.0	61.5	53.8	69.2
400 ps	50.8	56.7	69.2	67.9	67.3	33.5	57.5	57.0	59.2	63.0	48.6	32.4	62.5	43.6	48.7	57.3	77.2	69.9	78.2	48.0	76.6
450 ps	50.5	50.4	52.4	63.4	56.8	55.6	52.8	32.9	80.9	55.8	56.6	38.2	51.2	47.4	75.0	46.7	40.1	67.4	77.1	30.8	56.6
500 ps	37.4	49.3	56.5	79.2	53.5	48.3	45.3	59.5	70.3	48.5	60.3	37.8	59.5	55.8	61.4	43.0	73.9	90.3	75.6	64.0	51.4
550 ps	50.0	67.7	46.3	67.1	62.0	53.6	43.2	47.5	65.5	57.9	53.1	51.8	51.1	46.6	62.3	46.9	74.3	82.6	74.3	64.0	57.1
600 ps	51.1	44.0	54.0	73.7	48.9	73.2	31.3	31.5	52.5	71.5	66.3	45.5	52.0	31.3	65.8	59.6	40.9	70.0	81.6	48.5	64.9
650 ps	44.7	42.4	34.9	66.8	41.4	65.7	34.7	56.5	66.5	66.0	62.8	44.5	53.2	38.8	23.6	65.7	92.5	67.3	70.7	53.3	66.9
700 ps	49.3	62.8	30.6	88.6	70.4	56.5	47.9	48.9	70.4	59.8	48.9	48.4	53.9	55.2	59.6	46.4	85.6	67.4	40.9	63.3	69.2
750 ps	40.2	46.2	61.4	76.8	58.5	39.3	37.6	71.6	39.6	59.3	62.7	38.6	60.1	45.0	62.6	58.6	81.9	49.1	89.0	58.6	54.4
800 ps	45.5	41.3	54.5	73.0	56.9	57.7	54.2	67.3	64.6	62.5	57.9	48.2	42.1	58.2	44.3	48.5	88.9	83.5	70.1	34.4	55.0
850 ps	59.0	54.5	20.5	70.5	50.3	59.7	52.6	54.1	61.0	72.1	76.2	39.9	63.5	59.1	64.7	44.9	60.4	59.5	83.3	44.4	76.0
900 ps	66.9	31.5	60.3	78.8	41.5	50.3	55.4	57.6	59.2	37.2	44.3	50.5	63.4	57.5	41.2	49.3	63.5	73.3	56.1	54.1	80.0
950 ps	31.5	45.0	32.0	70.9	51.5	68.5	47.4	58.9	72.8	46.3	44.8	55.0	48.1	41.2	53.0	62.6	85.1	64.5	60.6	55.9	54.7
1000 ps	41.5	39.0	61.4	82.6	51.4	66.6	65.5	36.4	56.7	35.5	40.3	52.2	61.1	53.0	42.5	54.1	57.5	66.6	47.6	63.9	53.1

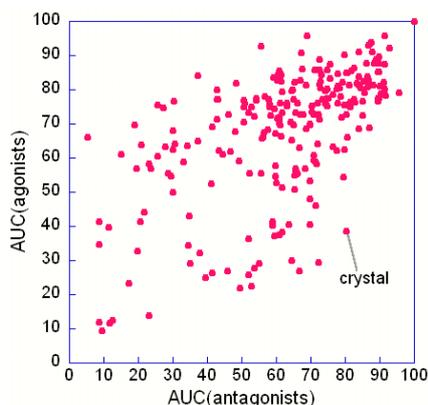
デコイセットAの462の標的構造に対するAUCの値。

	MODELLER										MOE										PRIME
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	
initial	67.4	61.5	74.9	71.2	82.4	65.8	64.5	58.5	54.3	56.9	65.6	38.9	33.8	66.0	44.5	44.7	67.1	48.7	40.5	60.6	59.4
minimized	58.2	64.3	45.5	67.3	85.7	69.1	77.8	62.2	71.8	77.7	65.5	56.2	56.1	42.5	33.7	50.2	64.2	69.9	77.4	63.6	59.9
50 ps	79.6	75.3	29.5	41.2	63.8	32.8	78.4	65.9	72.4	68.3	76.5	44.7	36.4	44.3	31.0	47.9	46.5	67.7	58.7	53.6	49.4
100 ps	65.6	68.0	54.5	46.0	76.4	42.6	56.8	68.3	72.0	82.5	76.4	66.7	46.9	78.2	60.9	56.4	34.9	63.8	77.2	41.0	38.1
150 ps	68.4	52.1	23.5	64.8	69.7	40.5	53.9	74.5	80.9	86.0	70.6	61.6	54.6	76.2	45.4	44.2	66.3	53.6	70.3	42.8	30.7
200 ps	70.8	71.2	52.0	81.4	63.0	50.5	46.3	45.2	80.9	67.9	63.7	59.0	56.0	79.7	46.5	60.2	64.4	66.9	78.5	67.8	17.0
250 ps	74.8	58.5	57.1	61.7	57.0	49.2	57.7	58.5	72.4	71.1	76.5	62.0	67.3	44.9	26.9	56.7	60.8	40.7	53.6	66.9	26.6
300 ps	62.6	55.2	59.5	65.1	63.6	53.5	34.1	47.9	77.3	48.5	59.1	61.5	40.6	41.2	71.5	60.5	49.6	57.5	71.4	59.5	42.5
350 ps	74.7	65.3	66.5	69.1	59.5	45.5	34.8	73.5	65.8	64.3	76.4	57.3	65.5	46.0	66.6	76.6	72.0	41.5	49.3	49.0	37.9
400 ps	79.6	68.5	72.6	67.2	59.6	33.4	48.5	71.8	52.8	84.4	55.8	54.7	60.5	40.4	31.1	74.0	48.6	63.5	75.0	45.0	53.5
450 ps	64.1	70.0	51.7	52.7	55.9	36.6	54.7	58.5	87.0	69.9	58.4	44.5	58.2	46.8	69.5	58.5	24.3	69.0	66.5	29.0	42.9
500 ps	68.2	63.2	61.2	79.9	61.6	34.3	35.8	68.3	77.9	69.3	61.5	60.5	37.8	43.9	58.6	40.2	59.8	85.0	58.8	62.0	27.2
550 ps	70.8	68.5	60.9	63.1	68.1	45.0	27.3	66.2	74.5	69.6	66.0	38.2	47.8	51.2	51.4	52.0	52.2	79.6	72.3	70.8	28.5
600 ps	64.5	58.0	57.9	55.0	49.4	69.6	32.6	59.0	42.5	78.5	75.6	58.3	51.2	53.5	57.8	72.6	33.6	68.5	80.4	56.4	37.4
650 ps	62.2	56.6	43.5	60.5	51.0	63.7	21.5	62.2	69.3	67.1	80.4	49.2	63.7	42.9	45.1	78.6	78.1	72.2	47.7	63.8	48.8
700 ps	61.7	65.1	39.7	75.4	79.4	51.1	50.7	62.3	76.7	76.2	70.8	57.6	63.7	65.8	57.6	53.9	78.4	65.4	28.1	64.5	44.5
750 ps	59.5	51.9	51.5	71.2	59.5	19.6	26.2	70.5	41.2	57.0	72.6	62.8	58.4	52.3	52.7	81.7	59.5	55.1	85.5	64.7	23.3
800 ps	63.5	49.5	35.3	61.5	63.0	42.5	44.8	78.2	67.1	73.4	66.4	61.6	21.7	60.8	44.2	68.2	72.5	65.7	74.5	34.3	25.9
850 ps	77.4	53.8	18.4	59.8	62.3	35.8	49.9	73.8	70.4	80.0	75.0	56.3	47.5	71.5	39.3	58.5	45.4	70.6	84.6	48.1	40.4
900 ps	85.3	45.2	32.2	49.5	68.9	29.3	56.1	66.4	71.0	45.3	50.9	57.2	48.4	61.3	22.5	70.3	45.2	71.3	62.4	48.4	44.5
950 ps	34.4	48.1	24.4	41.5	48.4	61.8	39.5	64.5	70.5	67.9	63.6	54.0	45.9	56.5	28.8	71.9	71.5	62.5	61.2	63.8	28.2
1000 ps	55.5	55.1	28.8	50.8	59.4	66.8	57.1	56.1	57.2	49.1	66.7	61.2	46.7	65.3	35.0	69.7	52.4	65.3	35.7	62.7	26.7

デコイセットBの462の標的構造に対するAUCの値。

MD計算における構造サンプリングした時間によってAUC値はさまざまで、ランダムスクリーニングよりも悪い結果を与えるもの(図の青色部分)もあるものの、80以上の高い値を示す標的構造(赤色部分)も数多く見つかった。膜を頭で考慮していないにもかかわらず、1n秒程度のMD計算で高いヒット率を示す標的構造をサンプリングすることができた。

同様のスクリーニング計算を、b<sub>2</sub>アドレナリン受容体(ADRB2)、A<sub>2A</sub>アデノシン受容体、ロイコトリエン受容体(CysLT1)およびb<sub>3</sub>アドレナリン受容体の4種類のGPCRにも適用した。ADRB2とA<sub>2A</sub>の結晶構造は既知であるが、Modellerにより側鎖配座の異なる初期構造をいくつか用意してMDによる構造サンプリングを実行した。いずれのGPCRの場合も、AUC値が80~90を超える標的構造が多く見ついている。また、アゴニストとアンタゴニストの両方のスクリーニング計算を実行してみると、それぞれに選択性のある標的構造も見出すことができることもわかった。



ADRB2 サンプリング構造を標的として、アゴニストとアンタゴニストのスクリーニング計算を実行したときのAUC値の相関。

以上の結果から、GPCRIに対して、標的タンパク質構造のホモロジー・モデリングとMDシミュレーション計算、およびMTS法による統計的スクリーニングによって、既知リガンドがある場合には、ヒット率が高いモデルをあらかじめ選択できることが示唆される。すなわち、シミュレーション計算に用いる出発構造は、たとえX線結晶構造でないホモロジー・モデルによるものでも、ヒット率としては高いものが得られることがわかった。ここで選択された構造を基に、実証研究を試みることにした。

さらに、agonist/antagonistの結合の違いについても、差別化した探索の可能性も示唆された。すなわち、GPCRIはagonist結合型とantagonist結合型の間の変化する構造を行うが、それら構造変化のうちリガンド結合サイトについては、シミュレーション計算によってサンプリングされる構造が類似の構造をある程度再現できているとも考えられる。複数のagonist/antagonistが既知であれば、この手法によって、未知の化合物が結合するモデルのタイプに依存して、その化合物がagonist型かantagonist型からの推定ができよう。さらに、リガンド結合サイトについて、agonists/antagonist結合型の特徴に基づく分類がなされれば、McMD法やSRPG法との組み合わせによる高効率の構造サンプリング手法と合わせて、より効果的なドッキング計算も期待される。

#### (c)生理活性ペプチドから非ペプチド性化合物(低分子化合物等)を探索・設計する手法の開発

[塩野義製薬分室]

##### 【事前調査】

PDBより、蛋白質と複合体を形成しているペプチドと、同じドッキングポケットに結合している他の低分子のデータを約20対収集した。このデータを用い生理活性ペプチドから非ペプチド化合物が探索できるか、ドッキング計算を基にした類似化合物探索手法(DSI法)で探索する実験を試みたが、ヒット率はランダムスクリーニングと大差なかった。理由としては、ペプチドは自由度が大きくdocking計算に向かないのでDSI法が使いにくいと考えられる。そこで新しい類似化合物探索手法を開発することで、生理活性ペプチドから非ペプチド低分子化合物が探索できるかどうか検討を進めた。

##### 【研究内容】

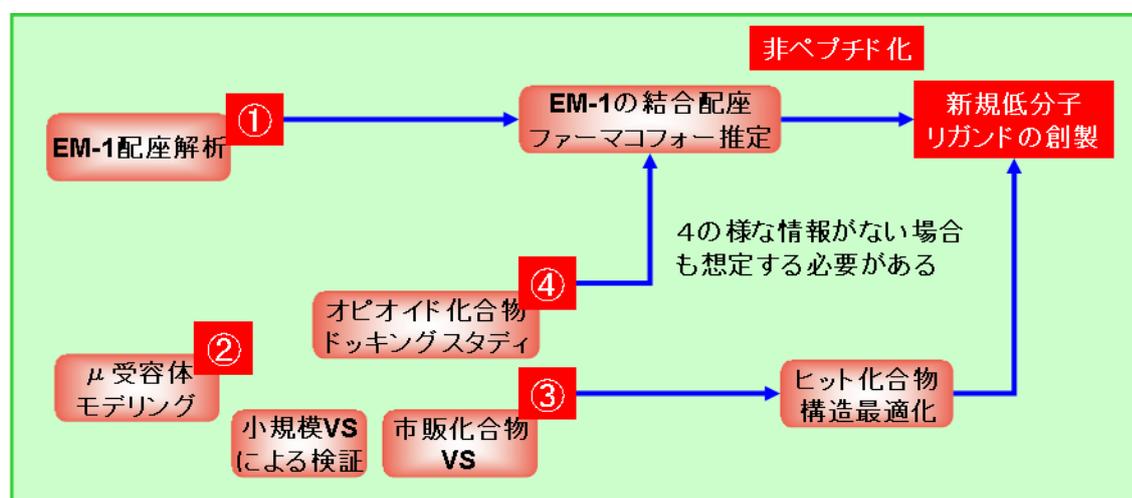
生理活性ペプチドから非ペプチド性化合物(低分子化合物等)を探索・設計する手法の開発の一環として、内在性の $\mu$ -オピオイドの一種であるEndomorphine-1 (Tyr-Pro-Trp-PheNH<sub>2</sub>: 以下、EM-1と略す)から非ペプチド性の新規 $\mu$ 受容体リガンドを探索するための検討を行った。具体的な研究方針を以下に示す。

EM-1の非ペプチド化を行う上では、その結合配座およびファーマコフォアが有効な情報になると考えられる。そこで、これらの情報を得るための解析に必要となるEM-1の立体配座集団および $\mu$ 受容体モデルを得るべく、EM-1の立体配座解析(後述の取り組み1)および $\mu$ 受容体のホモロジーモデリング(取り組み2)を行った。

次に、取り組み2で得られたモデルの検証を行うため、取り組み3において、小規模バーチャルスクリーニング(VS)による検証を行った。また、実証実験として市販化合物を対象としたVSについても検討を行い、新規 $\mu$ 受容体リガンドの探索を試みた。

また、 $\mu$ 受容体については既知のアゴニストリガンドが複数報告されており、これらのファーマコフォー情報についてもEM-1の結合配座およびファーマコフォーを推定する上で有効と考え、取り組み4において、オピオイドリガンドのドッキングスタディについて検討を行った。ただし、既知リガンドがなく生理活性ペプチドの情報しかない場合も想定されるため、このようなケースに対応する方法についても今後検討する必要があると考えている。以下に全体の研究の流れを示す。

図1. 研究の流れ



## 【研究成果】

以下にこれまでの取り組みおよび得られた成果を挙げる。

### 1. EM-1の立体配座解析

ターゲットである $\mu$ 受容体への結合配座の推定やファーマコフォー抽出の解析に活用することを目的に、立体配座の網羅的な探索を行った。

この検討には、myPrestoのcosgeneによる溶液中でのForce-biased Multicanonical Molecular Dynamics (F.B.McMD)シミュレーションにより解析を行った。解析の結果、2318個の立体配座集団が得られた。これら得られた配座集団については、今後のEM-1のドッキングスタディの取り組みにおいて活用したいと考えている。

### 2. $\mu$ 受容体のホモロジーモデリング

上記1の取り組みと合わせたEM-1の結合配座の推定や、バーチャルスクリーニングによる新規 $\mu$ 受容体リガンドの探索に活用することを目的に、 $\mu$ 受容体の立体構造を予測することとした。

human  $\beta$ 2 adrenergic G-protein-coupled receptor ( $\beta$ 2AR)を鋳型構造として、MOE (CCG社)

とPrime (SCHRODINGER社)を用いて計11個の $\mu$ 受容体モデルを構築した。また、これらはモデル構造であることから、構造の精密化およびタンパク質の動的挙動を考慮するため、cosgeneを用いたエネルギー極小化およびMDシミュレーションによる構造サンプリングを行った。解析の結果、最終的に132個の $\mu$ 受容体のサンプリング構造を得た。

### 3. MTS法スクリーニングによる $\mu$ 受容体構造の検証および新規 $\mu$ 受容体リガンドの探索

はじめに、上記2の取り組みにより得られた $\mu$ 受容体のサンプリング構造集団を用いて、MTS法(Multiple target screening法)による小規模スクリーニングの検証を行った。検証セットとしては、モルヒネやフェンタニル等の既知のオピオイドリガンド11個とデコイリガンド11,479個を用いた。この検証の結果、最も高いヒット率で既知リガンドを抽出できる構造を選抜し、この構造を用いて市販化合物に対するMTS法およびDSI法(Docking Score Index法)スクリーニングを行った(この計算についてはJBICで実施)。スクリーニング結果および化合物の在庫状況の確認の後、最終的に399個の候補化合物が選抜された。そしてこれらについて、 $\mu$ 受容体に対する結合親和性試験を実施したところ、複数個の活性化合物が確認された。

### 4. オピオイドリガンドのドッキングスタディ

EM-1の結合配座の推定やファーマコフォー抽出の解析に活用することを目的に、既知のオピオイドリガンドのドッキングスタディを行った。この際、既知リガンドとしては、モルヒネやフェンタニル等10個のリガンドを用いた。また $\mu$ 受容体モデルとしては、Mosberg等により報告されている活性型の $\mu$ 受容体モデルを用いた。cosgeneを用いたMDシミュレーションにより200個の構造をサンプリングした後、Sievene, GOLD(CCDC社), S社のドッキングソフトを用いてアンサンブルドッキングを行い、多数のドッキングポーズを得た。そして、得られたドッキングポーズをエネルギー極小化により精密化した後、MM-GBSA手法により簡易的な結合自由エネルギー( $\Delta G_{bind}$ )を算出した(エネルギー極小化, MM-GBSA計算については、AMBER 8を使用)。最後に、各リガンド毎で $\Delta G_{bind}$ が良好なドッキングモデルを比較し、推定の共通ファーマコフォー情報を得た。

以下、各項目の詳細な検討内容について記載する。

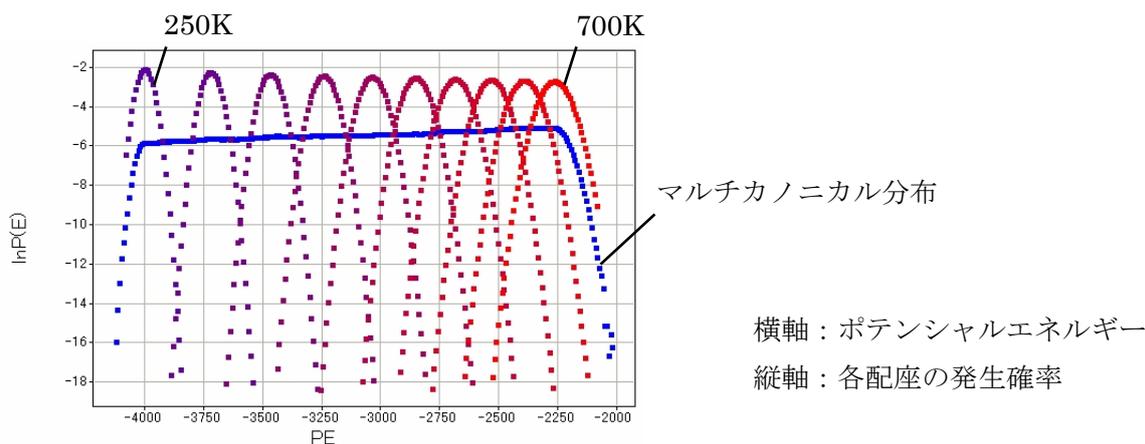
#### 1. EM-1の立体配座解析

EM-1分子が単独で取り得る立体配座の算出を試みた。EM-1は4残基と比較的短いペプチドであること、および2番目のアミノ酸として自由度の少ないProを含んでいるため、その立体配座の数は通常の生理活性ペプチドよりも限定されるものと期待される。

立体配座解析の手法としては、先の生体高分子プロジェクトの成果物であるmyPrestoのcosgeneによる溶液中でのF.B.McMDシミュレーションによる解析を行った。通常、配座解析手法としては分子動力学法や二面角の系統的探索法があるが、ペプチド分子のように回転可能結合が多い分子では、全配座空間を網羅的に探索することは難しく、一般的な分子動力学法による配座解析では局所空間に留まってしまう恐れがある。また、二面角の系統的探索法では、多くの計算時間を要し、また実際には優位に取り得ない配座をサンプリングしてしまう恐れもある。そこで、F.B.McMDシミュレーションを用いることで、網羅的かつ効率的な配座探索を試みた。

F.B.McMDシミュレーションの検討の結果、下図で示すような高いエネルギーから低いエネルギーを持つ配座までのマルチカノニカル分布を示す配座集団が得られた。そして、これらの集団について、myPrestoのreweightFBツールを用いて解析を行ったところ、下記で示すような各温度に対応するカノニカル分布を示す配座集団を得ることができた(50K毎に表示)。そして、これらの結果の内、300Kに該当する配座集団を抽出したところ、2318個のEM-1の立体配座が得られた。

図2. F.B.McMDシミュレーション結果



## 2. $\mu$ 受容体のホモロジーモデリング

$\mu$  受容体は、7回膜貫通型のGPCRであり、その立体構造は明らかにされていない。本検討では、ホモロジーモデリング手法により、その立体構造の予測を試みた。

実際のホモロジーモデリングの作業としては、はじめに、clustalwを用いて、アミノ酸配列の multiple alignmentを行った。用いたアミノ酸配列としては、GPCRの中でRhodopsin-likeファミリーの内、peptideサブファミリーに属するOpioidやAngiotensin、Somatostatin受容体等の38配列、そして結晶構造が明らかにされているRhodopsin受容体(Bovine)および $\beta$  2ARの2配列の計40配列である。以下に、multiple alignmentの結果の中から、 $\mu$  受容体と $\beta$  2AR(結晶構造PDB ID:2rh1.pdbのアミノ酸配列)のアミノ酸アライメントを抜粋したものを示す。