

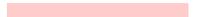
# 計算生命科学の基礎



## 分子動力学計算における 拡張アンサンブルシミュレーション

---

神戸大学大学院工学研究科 応用化学専攻  
中津井雅彦



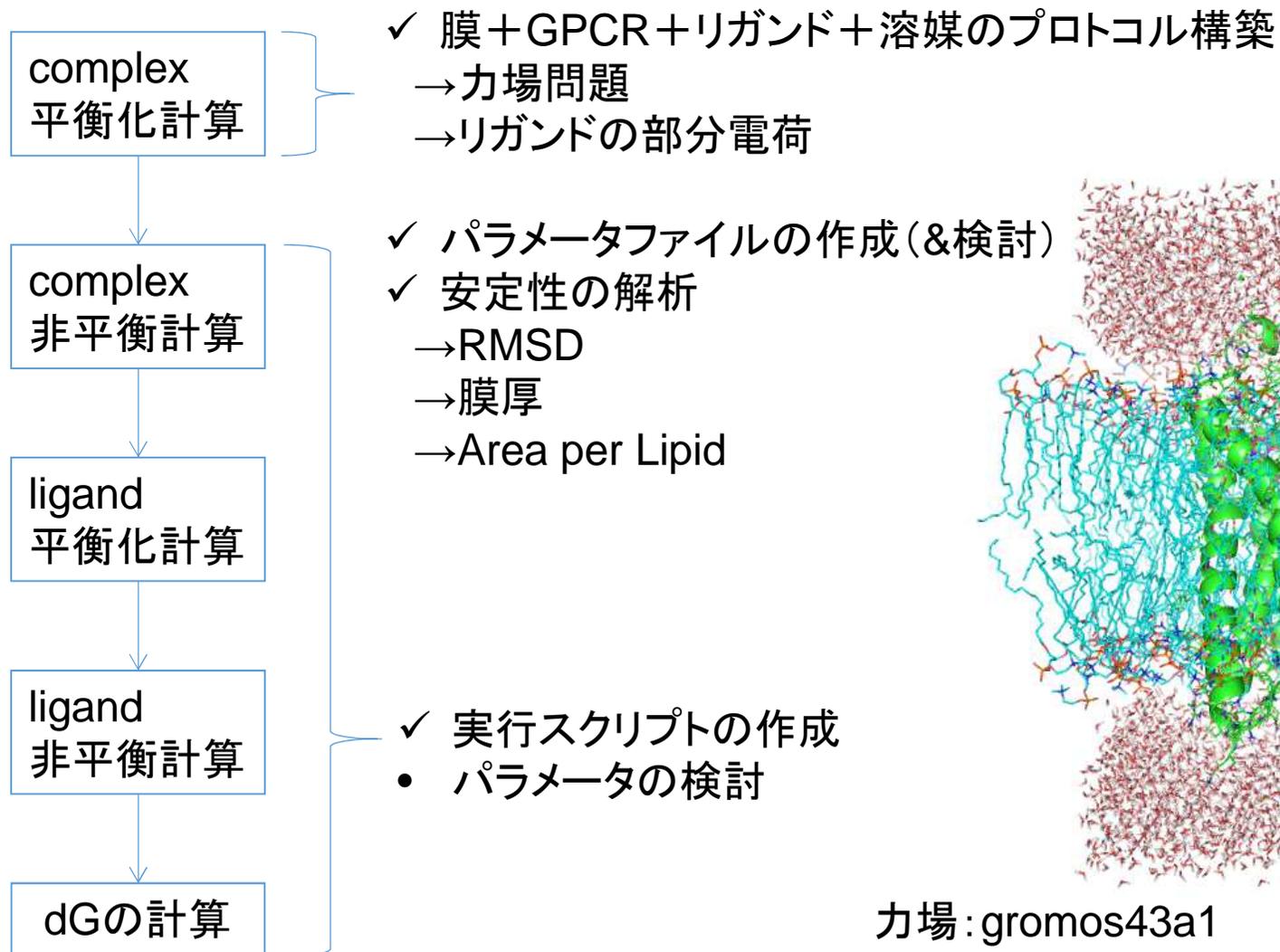


- 先週の補足
  - KBDD: GPCR
- 統計アンサンブル
  - 運動方程式の一般化
  - NVTアンサンブル(カノニカルアンサンブル)
  - NPTアンサンブル(T-Pアンサンブル)
- 拡張アンサンブル
  - 必要性
  - マルチカノニカル法
  - レプリカ交換法
  - 実例



# MP-CAFEE法のGPCRへの適用

## GPCRのMP-CAFEE計算手順構築



力場:gromos43a1

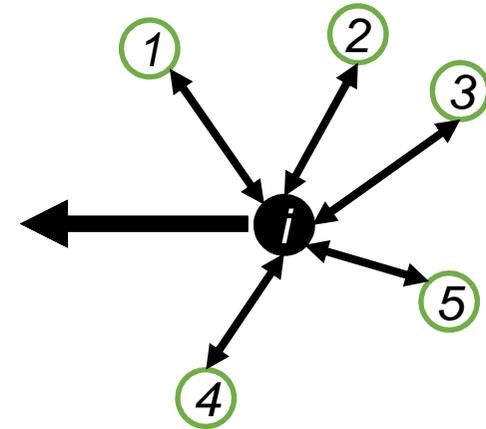
リガンド:PRODRG+GAMESS+RESP

# 運動方程式の一般化



ニュートンの第二法則

$$\mathbf{F}_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = m_i \mathbf{a}_i$$



ラグランジアン  $L$

$$L = K - U$$

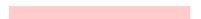
ポテンシャル  
運動エネルギー

ラグランジュの運動方程式

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} \right) - \frac{\partial L}{\partial q} = 0$$

一般化座標

$$\mathbf{q} = (q_1, q_2, \dots, q_n)$$



# ハミルトニアン



ハミルトニアン( $H$ ) : 系の全エネルギー

$$H(\mathbf{q}, \mathbf{p}) = \mathbf{p}\dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}})$$

$\mathbf{p}$ : 一般化座標

$\mathbf{q}$ : 一般化運動量

ハミルトニアンの正準方程式

$$\dot{q} = \frac{\partial H}{\partial p}, \dot{p} = -\frac{\partial H}{\partial q}$$

全エネルギーの保存

$$\frac{dH}{dt} = \sum_{i=1}^n \left\{ \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} - \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} \right\} = \sum_{i=1}^n \left\{ \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} \right\} = 0$$

全エネルギーの時間変化が0なので、エネルギー保存則が成立する



# 代表的なアンサンブル



## アンサンブル

個々の座標や運動量は異なるが、熱力学的には同じ状態にある座標・運動量の集団  
分子動力学計算では、運動方程式を解いて得られるタンパク質構造のスナップショットの集合(トラジェクトリ)

- ミクロカノニカルアンサンブル (NEVアンサンブル)
  - 粒子数( $N$ ), エネルギー( $E$ ), 体積( $V$ )一定

通常分子動力学計算はNEVアンサンブル  
ニュートンの運動方程式, ハミルトニアン  
の正準方程式はエネルギー保存則が成り立つため

- カノニカルアンサンブル (NVTアンサンブル)
  - 粒子数( $N$ ), 体積( $V$ ), 温度( $T$ )一定
- T-Pアンサンブル (NPTアンサンブル)
  - 粒子数( $N$ ), 圧力( $P$ ), 温度( $T$ )一定



# 温度一定のMD (NVTアンサンブル)



## カノニカルアンサンブル

運動エネルギーと温度の関係

$$\sum_{i=1}^N \frac{m_i}{2} \mathbf{v}_i^2 = \frac{3}{2} (N - 1) k_B T$$

## 速度スケーリング法

温度を制御するためには、速度を制御すればいい

$$\mathbf{v}_i = s \mathbf{v}'_i$$

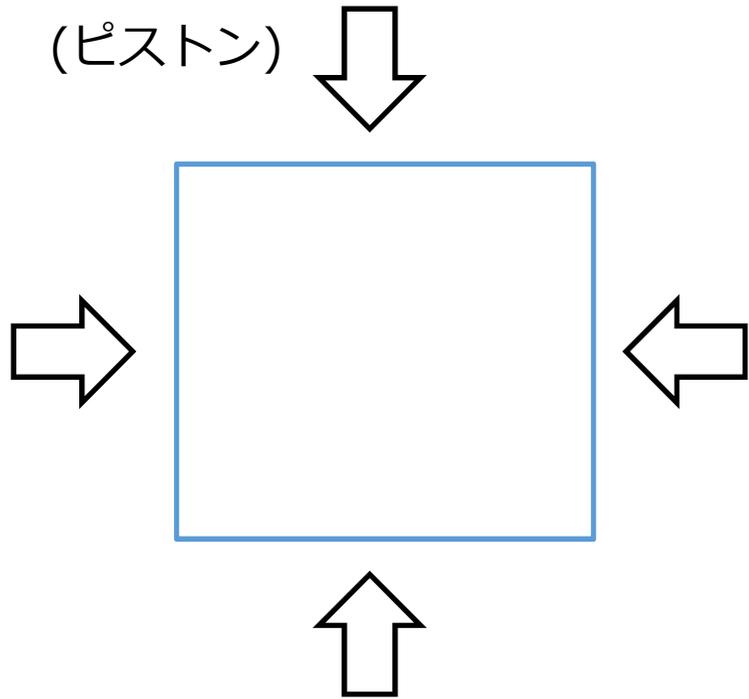
$$s = \left\{ \frac{(3N - 1) k_B T_0}{\sum_i m_i \mathbf{v}_i^2} \right\}$$

# 温度・圧力一定のMD (NPTアンサンブル)



能勢・アンダーソンの方法  
(温度制御) (圧力制御)

圧力を制御する方法  
(ピストン)



外界からの圧力 $P_0$ を考える

ポテンシャルエネルギー

$$V_V = P_0 V$$

全ての座標を、 $V^{\frac{1}{3}}$ でスケーリング

座標

$$\mathbf{r}_i = V^{\frac{1}{3}} \mathbf{q}_i$$

速度

$$\dot{\mathbf{r}}_i = V^{\frac{1}{3}} \dot{\mathbf{q}}_i$$

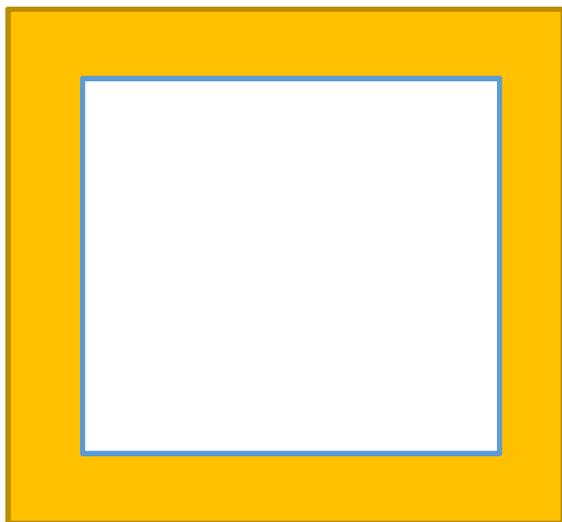


# 温度・圧力一定のMD (NPTアンサンブル)

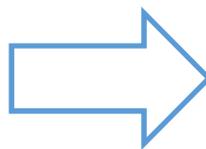


## 温度を制御する方法

熱浴



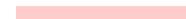
時間の進み方が異なる( $t \rightarrow t'$ )と考える



$$dt = \frac{dt'}{s}$$

運動量の関係

$$p_i = \frac{p'_i}{sV^{1/3}}$$



# 温度・圧力一定のMD (NPTアンサンブル)



仮想系のハミルトニアン  $H'$

$$H' = \sum_{(i)} \frac{\mathbf{p}'_i{}^2}{2m_i V^{\frac{2}{3}} s^2} + U \left( V^{\frac{1}{3}} \mathbf{q}^N \right) + \frac{P_s^2}{2Q} + 3Nk_B T_0 \ln s + P_V V + P_O V$$

現実系の運動方程式

$$\ddot{\mathbf{q}}_i = - \frac{1}{m_i V^{\frac{2}{3}}} \frac{\partial U}{\partial \mathbf{q}_i} - \left( \frac{2\dot{V}}{3V} + \frac{\dot{s}}{s} \right) \dot{\mathbf{q}}_i \quad \ddot{s} = \frac{2s}{Q} \frac{3Nk_B}{2} (T - T_0) + \frac{\dot{s}}{s}$$

$$\ddot{V} = \frac{s^2}{W} (P - P_0) + \frac{\dot{s}}{s} \dot{V}$$

$V, p_v, s, p_s$  に関する項を消去

: ミクロカノニカルアンサンブル (NEV)

$V, p_v$  に関する項を消去

: カノニカルアンサンブル (NVT)

# 評価データセット

共結晶構造が存在し、比較可能な $K_i$ や $K_d$ 値が登録されているものを検証データとして使用

ドッキングスコア開発用データセット



- CHK1 : 6 リガンド
- CDK2 : 12 リガンド
- ERK2 : 6 リガンド
- uPa : 7 リガンド



1セットあたり 約130,000時間(約15年) / コア  
合計 5,330,000時間 (約608年) / コア

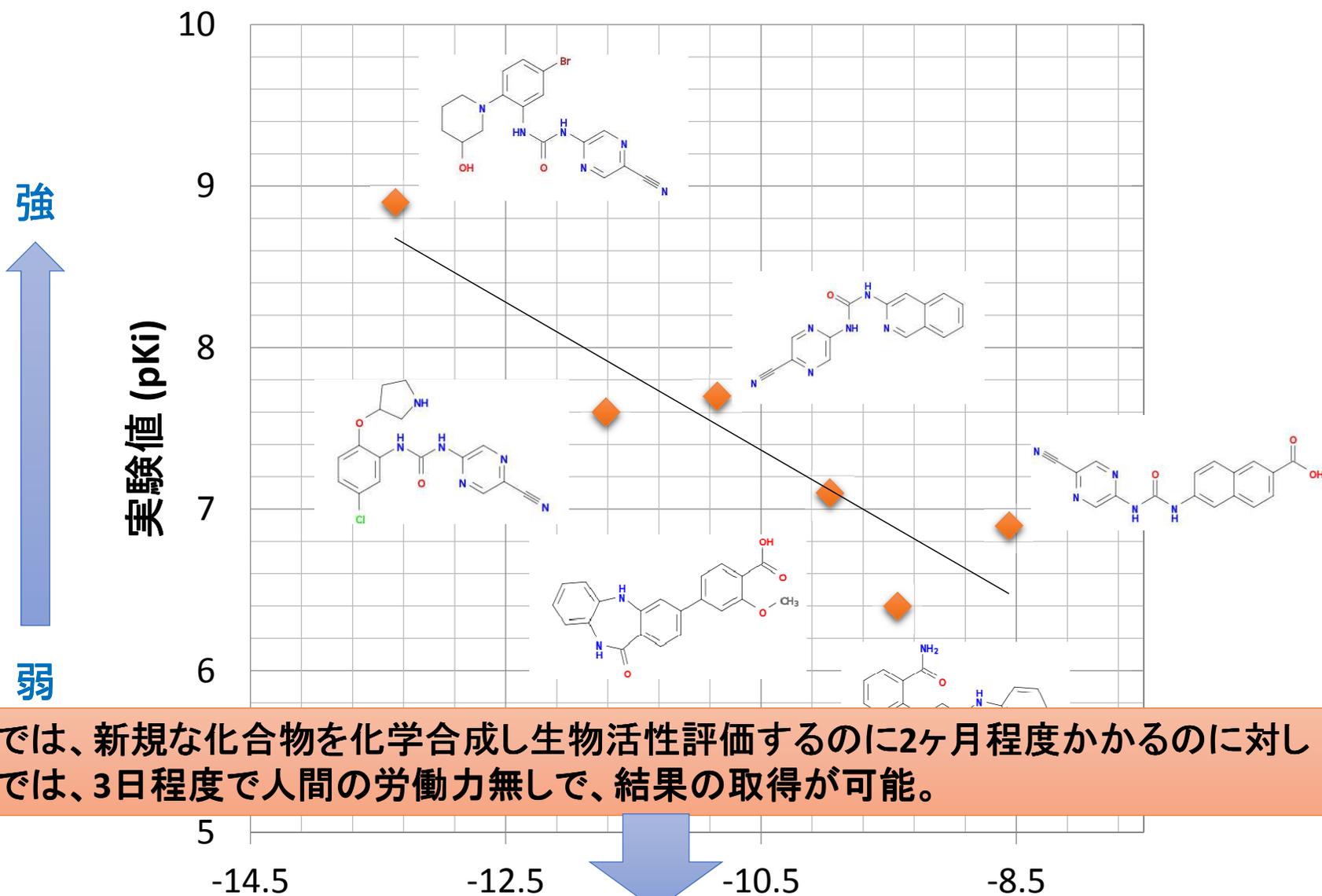


「京」の30%を使用して**約6日**

文献調査により選択したGPCR

- GPCR : 13 リガンド

# CHK1: MP-CAFEE法による予測と実験結果比較

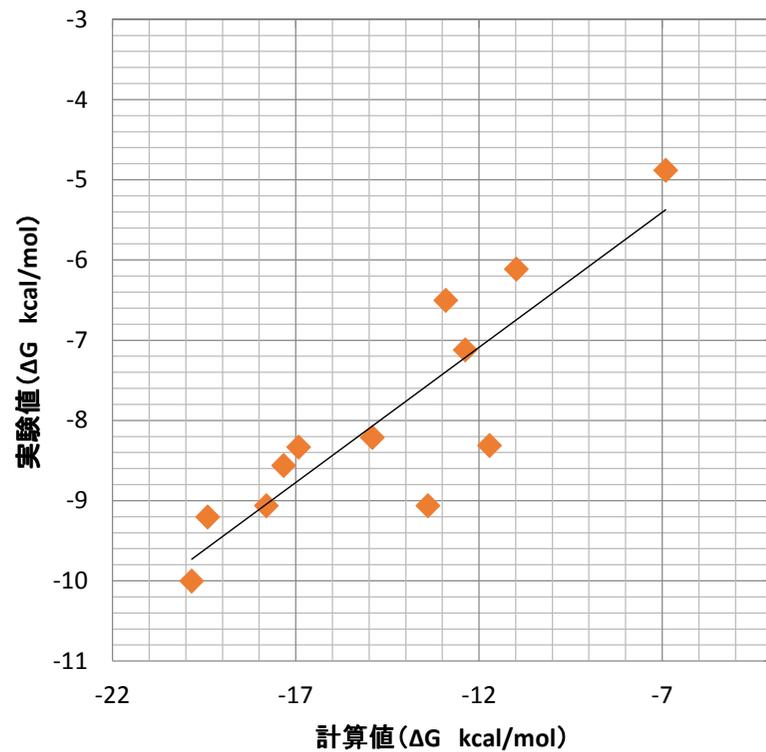


実験では、新規な化合物を化学合成し生物活性評価するのに2ヶ月程度かかるのに対し「京」では、3日程度で人間の労働力無しで、結果の取得が可能。

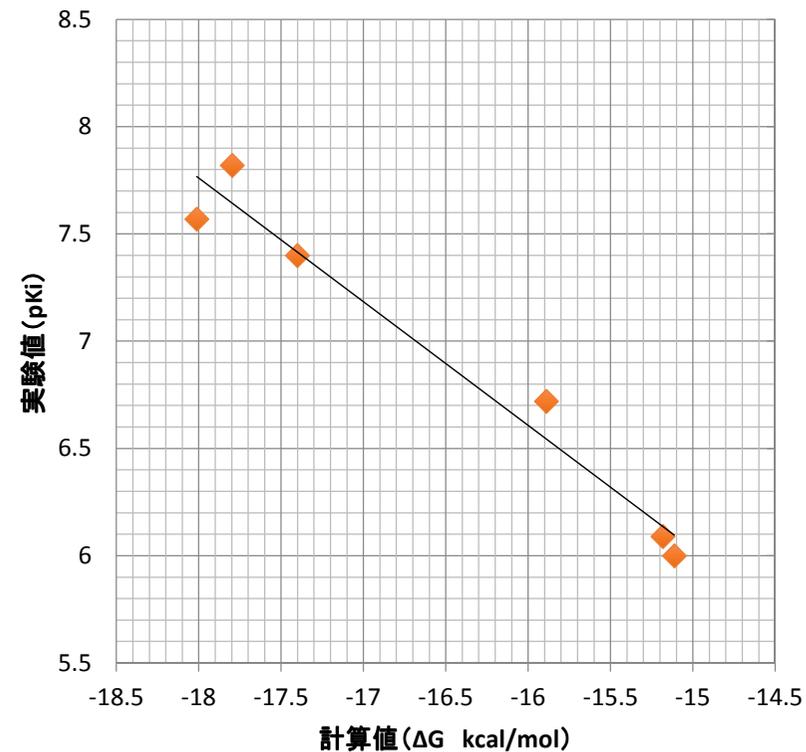
数多くの化合物を計算するためには、もっと高速な計算機が必要  
⇒ ポスト「京」への強い期待！

# MP-CAFEE法による予測と実験結果比較

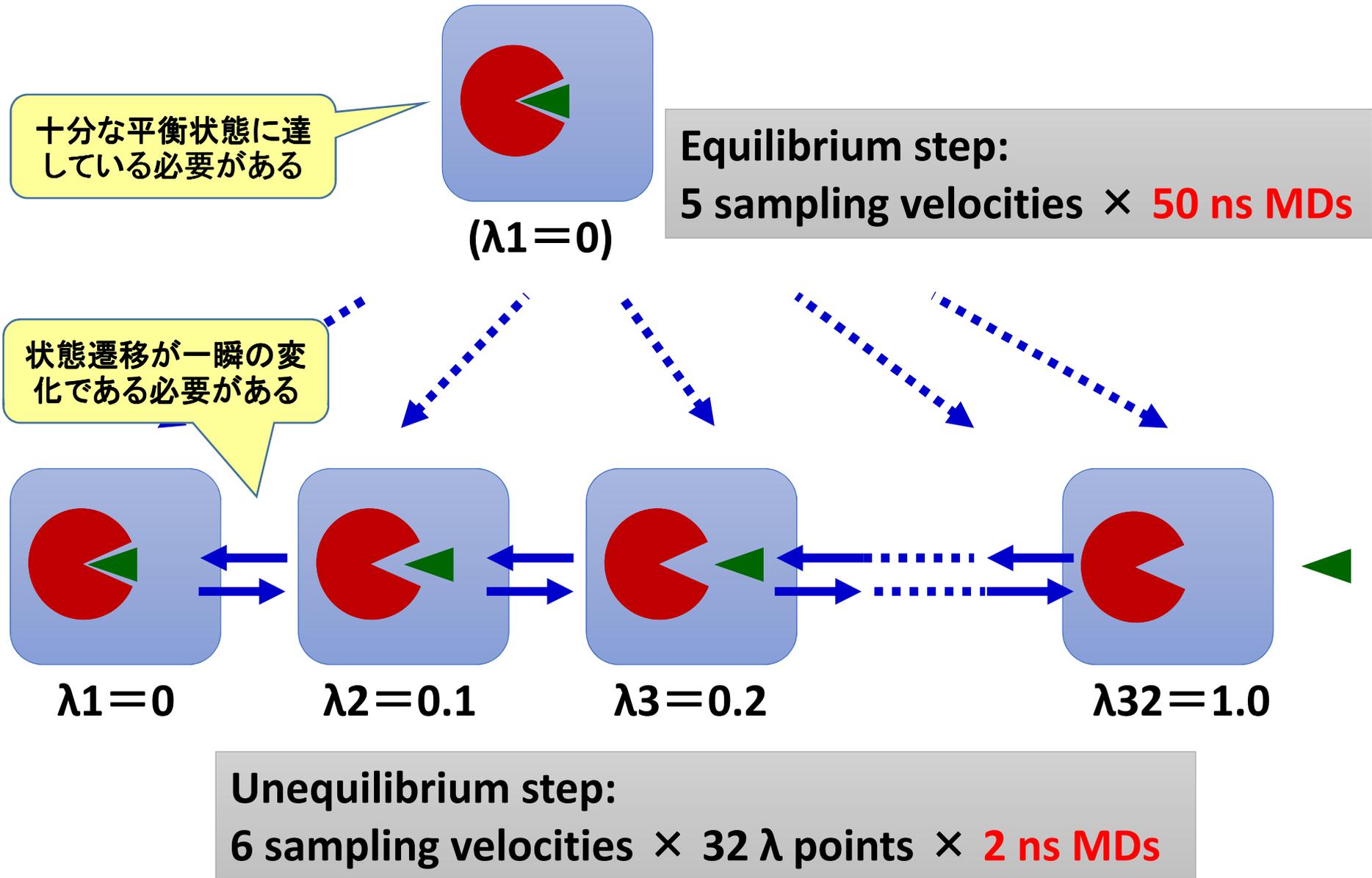
## CDK2



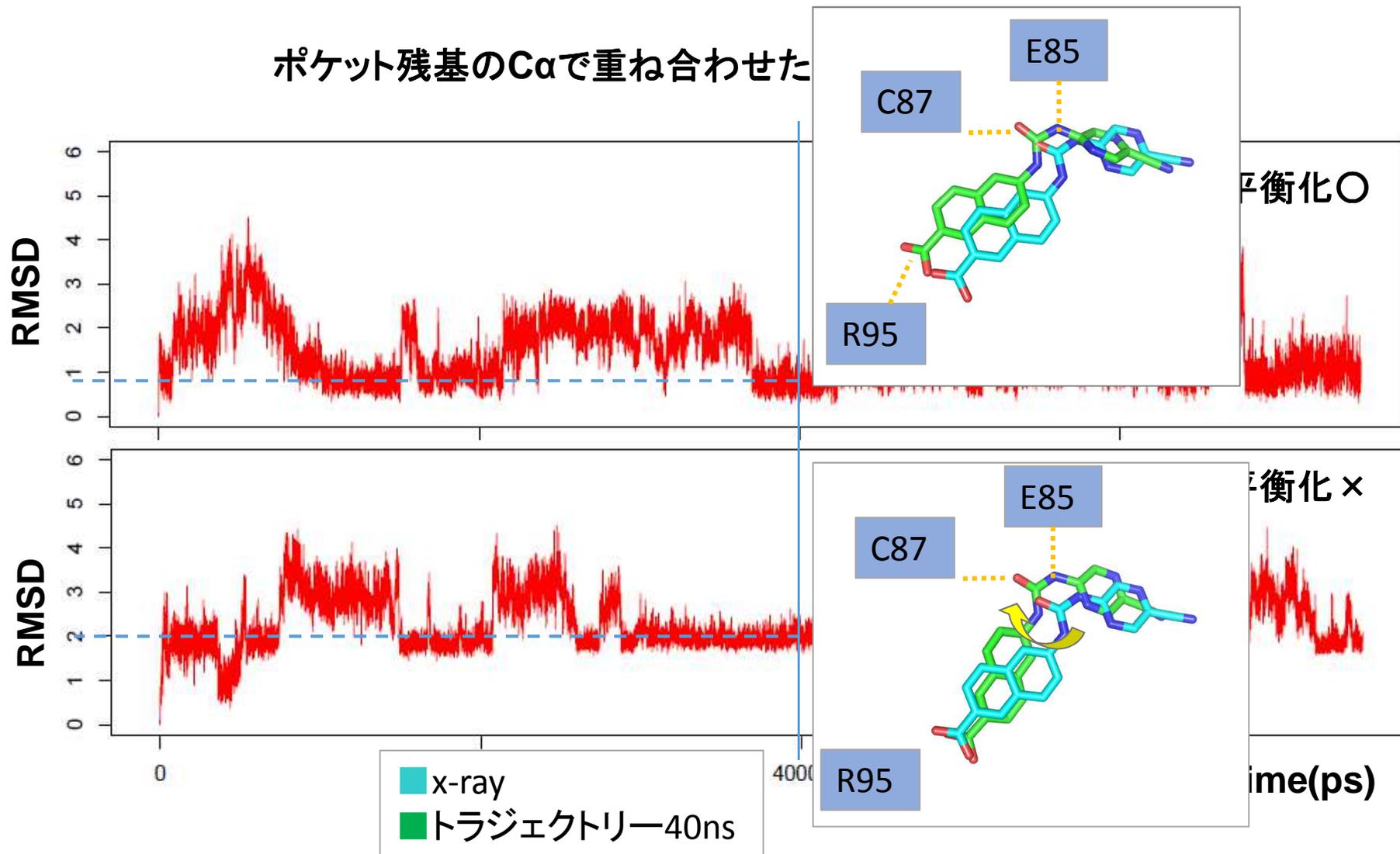
## ERK



# 平衡化したドッキングポーズとは？

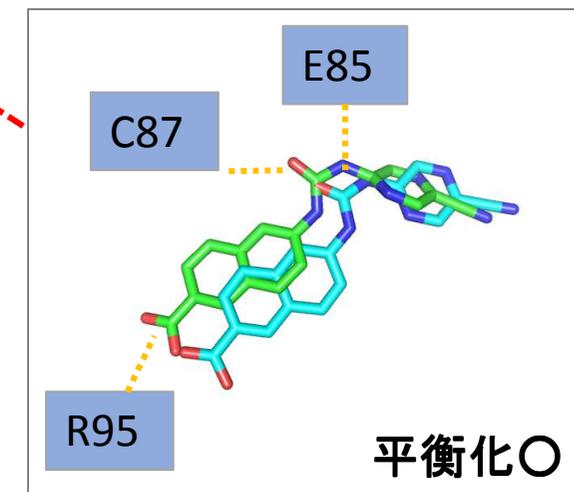
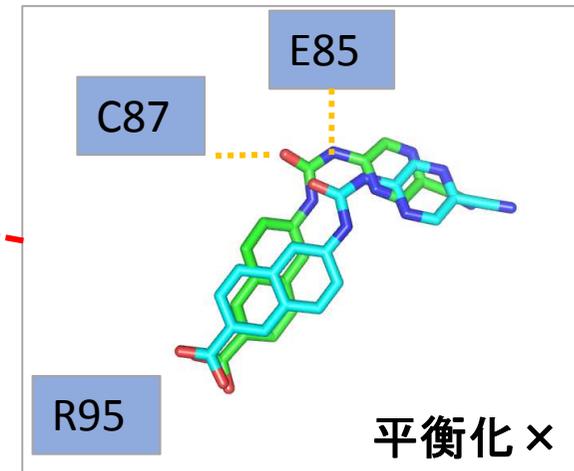
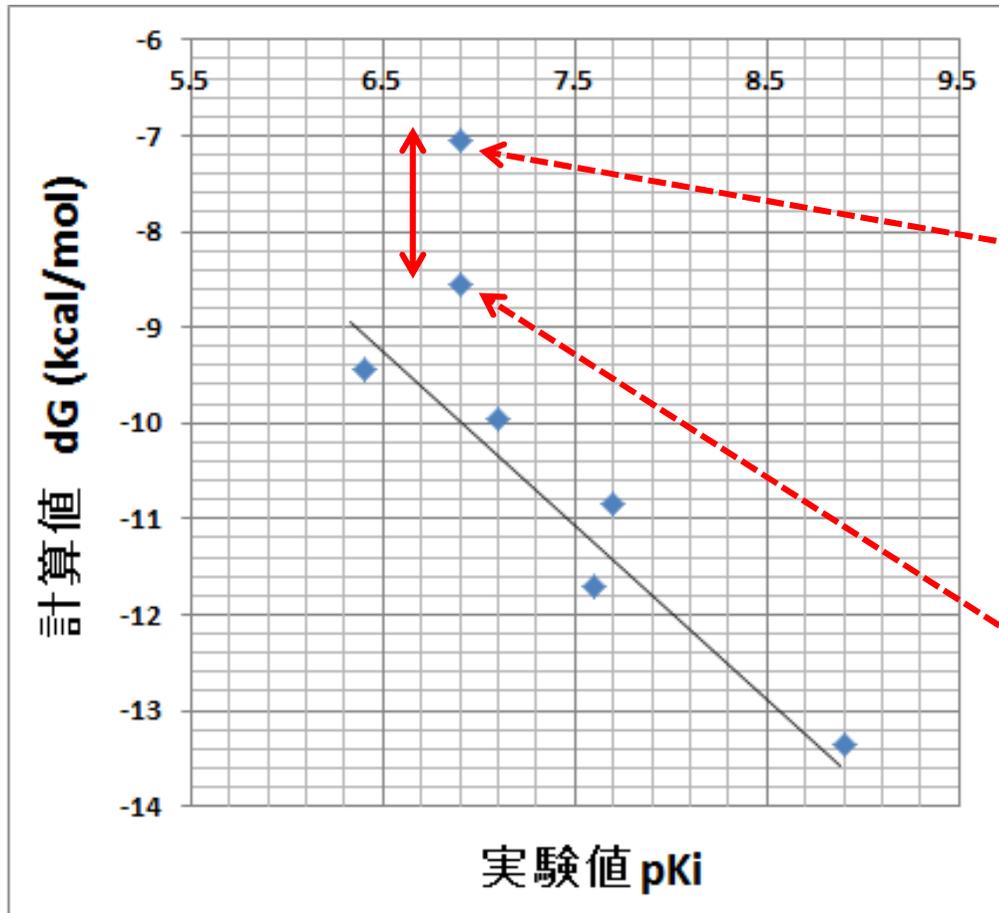


# 平衡化したドッキングポーズとは？



同じ初期構造でも初速度によりドッキングポーズが変化する

# 非平衡計算開始構造の違いによる MP-CAFEE法の結合自由エネルギー値への影響

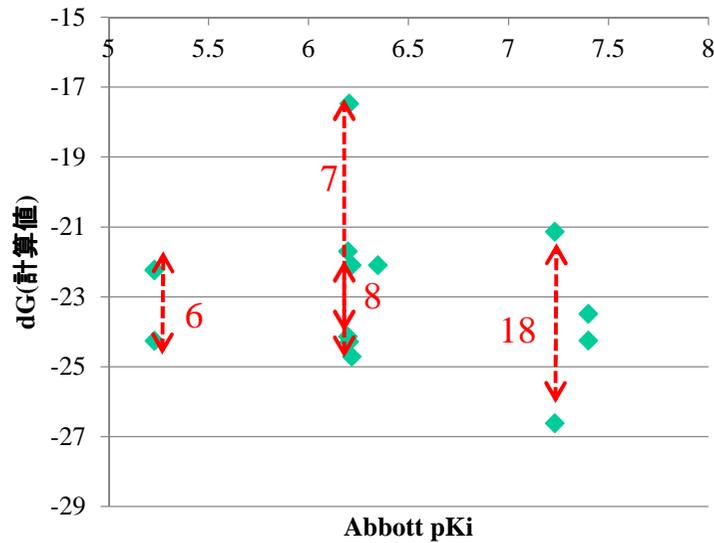


平衡化状態におけるドッキングポーズを  
取得することが重要課題

# 現在の課題①～平衡化した構造とは？～

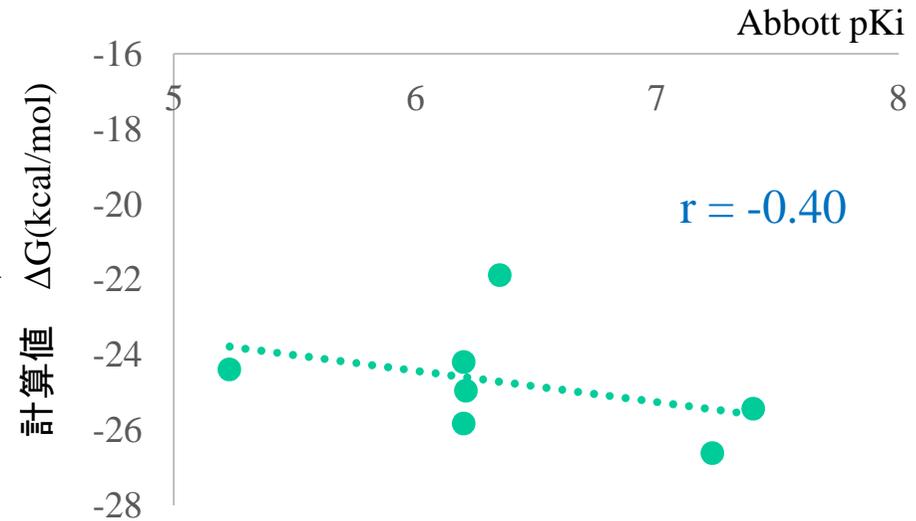
## プロテアーゼ (uPa) 阻害剤の結合自由エネルギー計算結果

各化合物に対して $\Delta G$ を幾つか算出してプロット



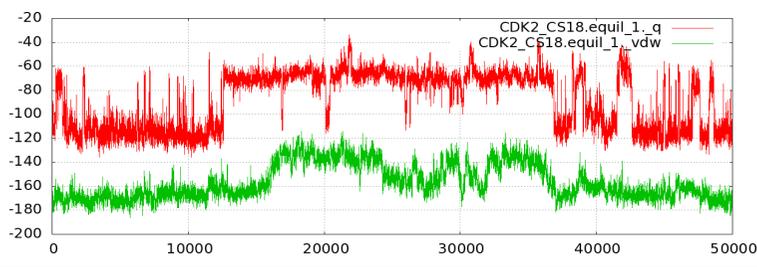
開始構造 (平衡化MDの選択時間) によって $\Delta G$ のバラつきが大きい

各化合物に対して $\Delta G$ の最小値をプロット



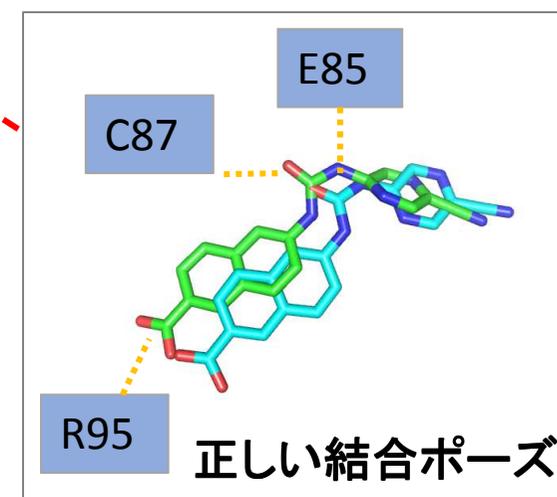
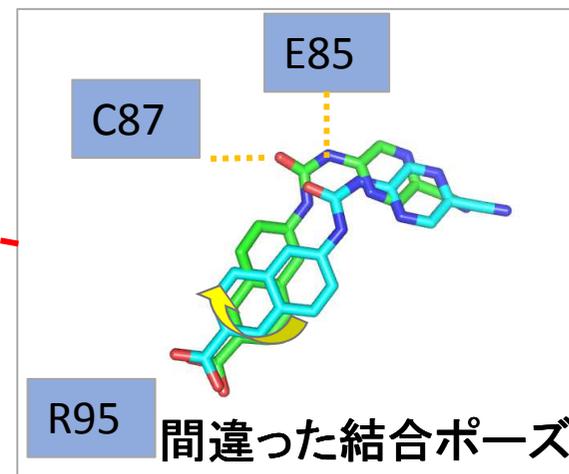
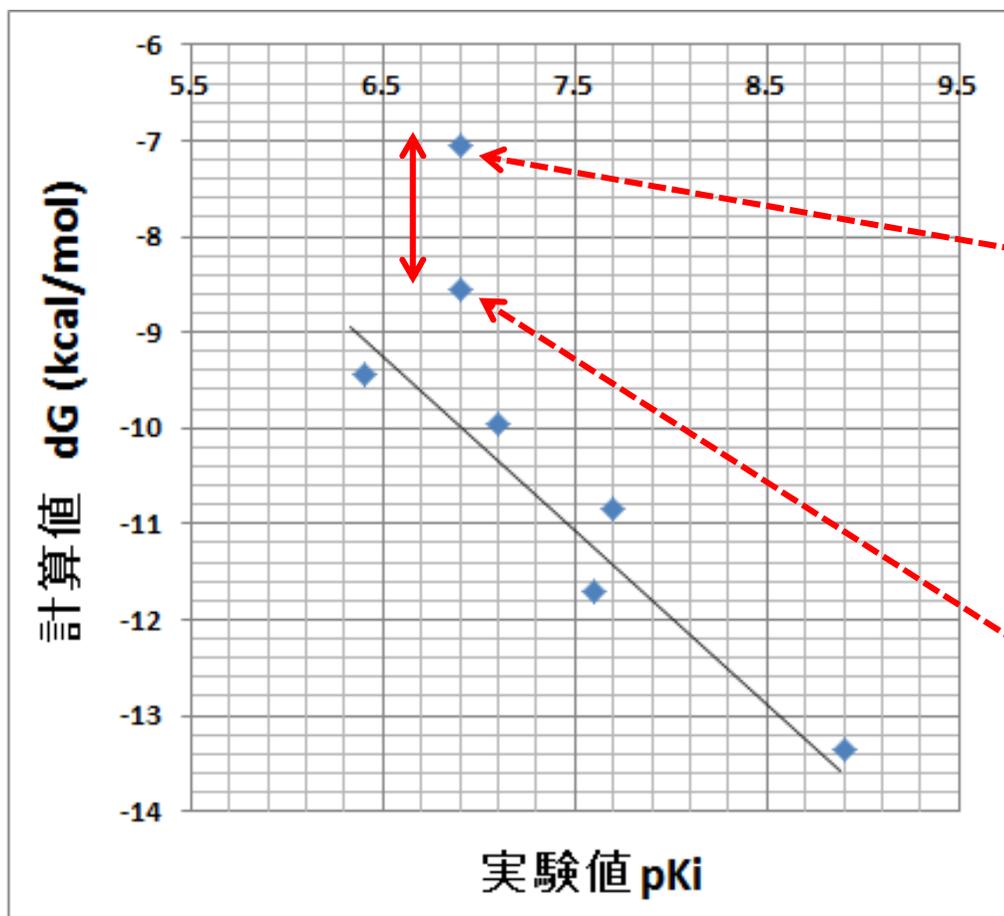
実験値(Ki)を再現する傾向がみられる

### 平衡化MDのエネルギープロファイル



どのようにして、平衡化プロファイルから最安定構造を選択するか？

## 現在の課題②～結合ポーズの正確な予測～



あらかじめ正しいドッキングポーズを取得することが重要課題

理化学研究所 計算科学研究機構 荒木氏  
三井情報(株) 佐藤氏

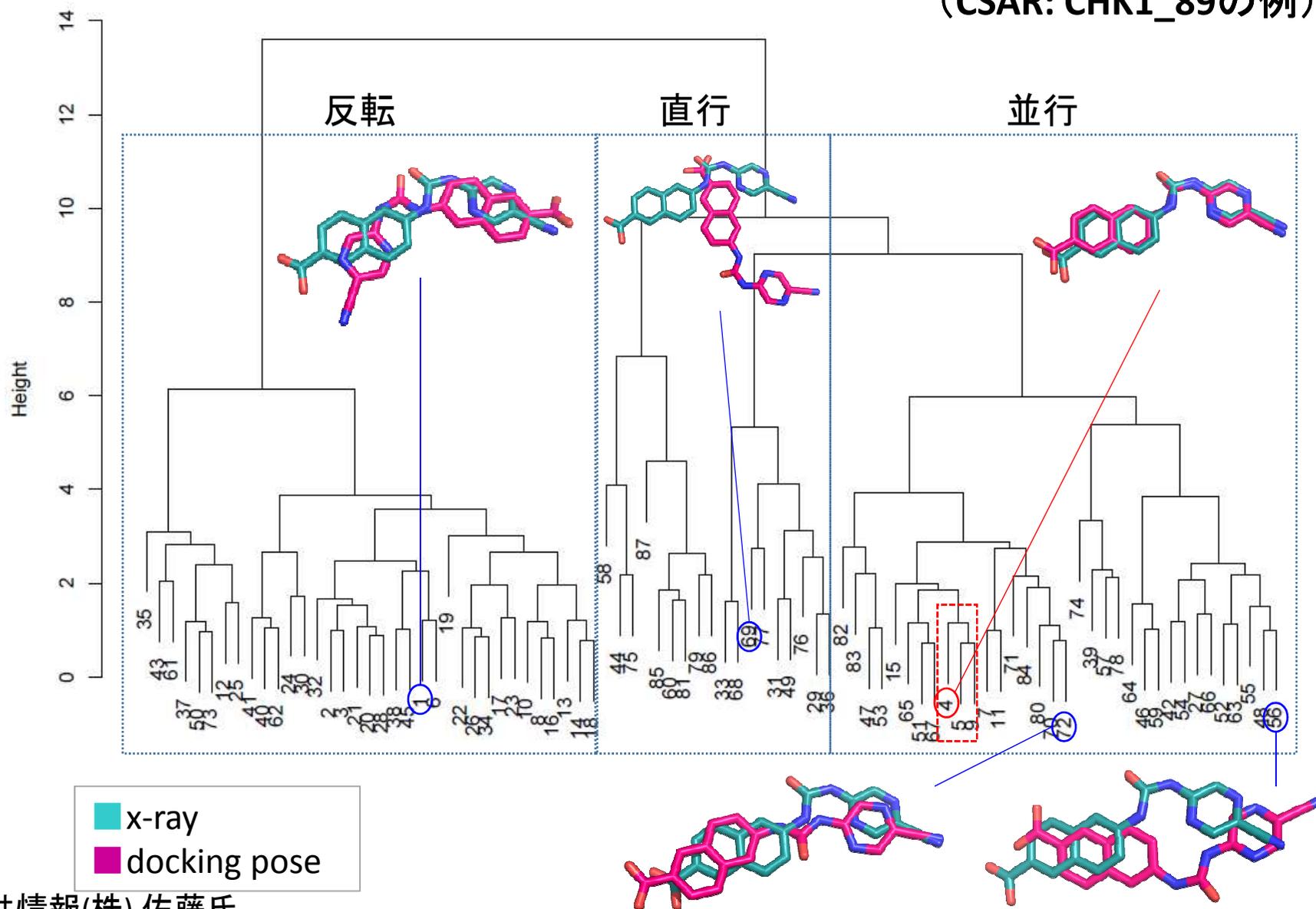
# CHK1のドッキング計算(Redocking)

共結晶構造のRedockingでも結晶構造を当てられない場合もある

CSARID	#Pose	GSCORE TOP DOCKING POSE	RMSD TOP DOCKING POSE	
		RMSD	RANK	RMSD
5	37	0.29	1	0.29
19	90	1.32	11	0.47
40	97	0.81	2	0.31
70	54	0.31	1	0.31
72	6	7.90	6	3.99
89	87	9.26	4	0.51

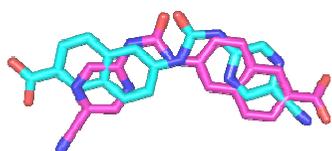
# Docking計算法で予測された結合ポーズとX線構造との比較

(CSAR: CHK1\_89の例)

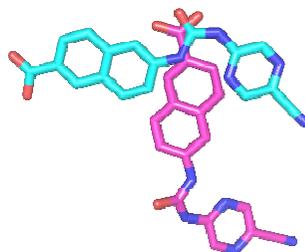


# 平衡化MD計算によるドッキングポーズ変化

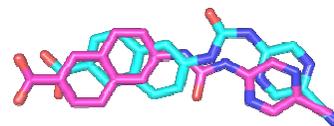
rank1



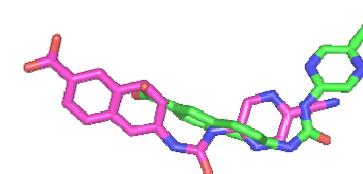
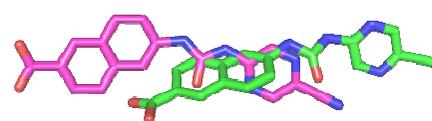
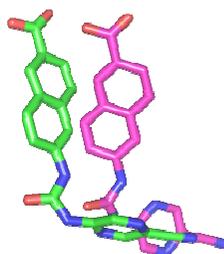
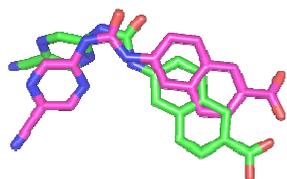
rank69



rank72



rank56



■ x-ray  
■ docking pose  
■ md50ns

50nsMDでは結晶構造と同じ結合ポーズは現れない  
⇒ レプリカ交換MDなどサンプリング手法を検討する必要あり？

# 分子動力学シミュレーションの課題(1)

## シミュレーション時間



1ヶ月程度で到達可能なシミュレーション時間  
(タンパク質、水系)

~100ns: 通常の計算機サーバ

~10 $\mu$ s: GPU搭載型計算機(20日  $\div$  10<sup>6</sup>秒)

~10 $\mu$ s: 京(並列計算の合計値)

~1ms: Anton (データ通信がボトルネック)



生命化学反応の時定数  
(ミリ秒~秒)

### Antonを用いた計算例

ユビキチンのfolding,  
unfolding

リガンドのタンパク質結合過程のシミュレーション

ユビキチンのfolding, unfolding

Piana et al., PNAS 2013, 110, 5915–5920,  
Fig. 1 A

リガンドのタンパク質結合過程のシミュレーション

Src kinase(チロシンキナーゼ)と  
dasatinibの結合

Shan et al., J. Am. Chem. Soc. 2011, 133,  
9181–9183,  
Supplementary Material (1\_si\_001 (AVI))

Piana et al., PNAS 2013, 110, 5915–5920

Shan et al., J. Am. Chem. Soc. 2011, 133, 9181–9183

## リガンドのタンパク質結合過程のシミュレーション

Src kinase(チロシンキナーゼ)と  
dasatinibの結合

Shan et al., J. Am. Chem. Soc. 2011, 133, 9181–9183,  
Supplementary Material (1\_si\_001 (AVI))

# 分子動力学シミュレーションの課題(2)

## 分子数



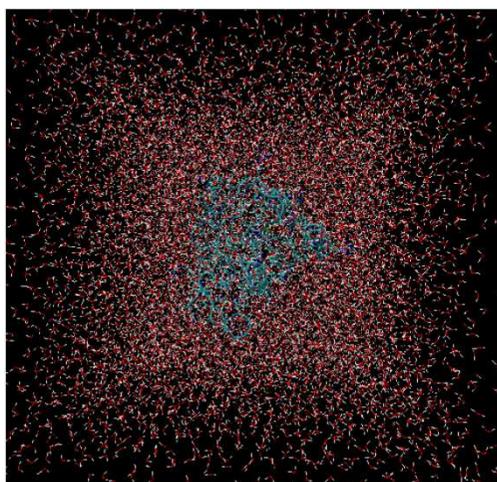
タンパク質のシミュレーションの場合

システム構成:

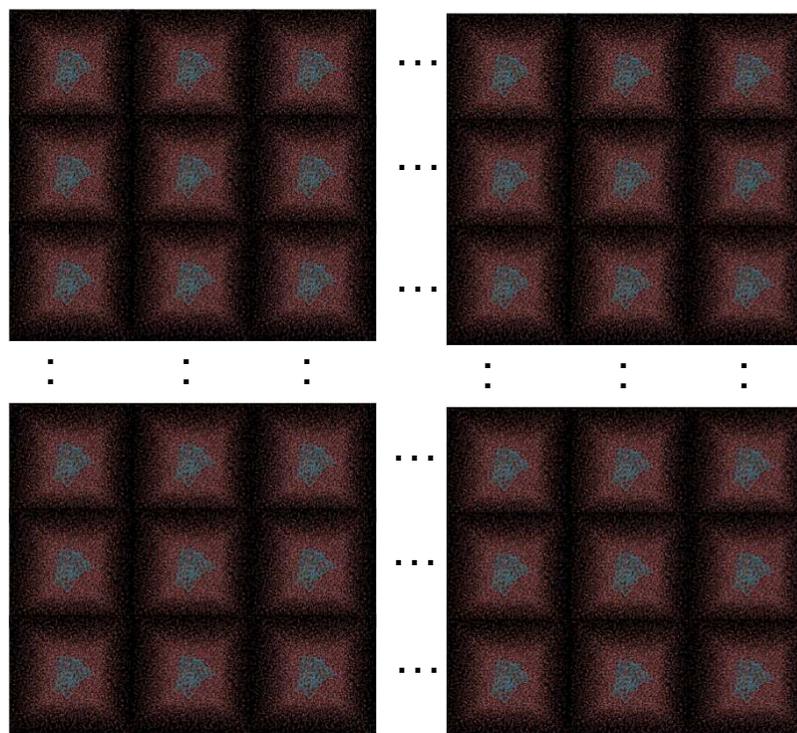
タンパク質: 1個 (水:  $10^4$ - $10^5$ 個)

<<

アボガドロ数 ( $10^{23}$ 個)



~10nm



1分子シミュレーションから算出された物理量で熱力学量を推定しなければならない

京: 88000ノードを使用して、1mMのタンパク質溶液1mLの挙動を1秒間シミュレートしようとする、...

⇒  $1\text{s} / (10\text{ns/day}) \times 10^{17}\text{個} / 10^5\text{個} = 10^{20}\text{年かかる!!}$

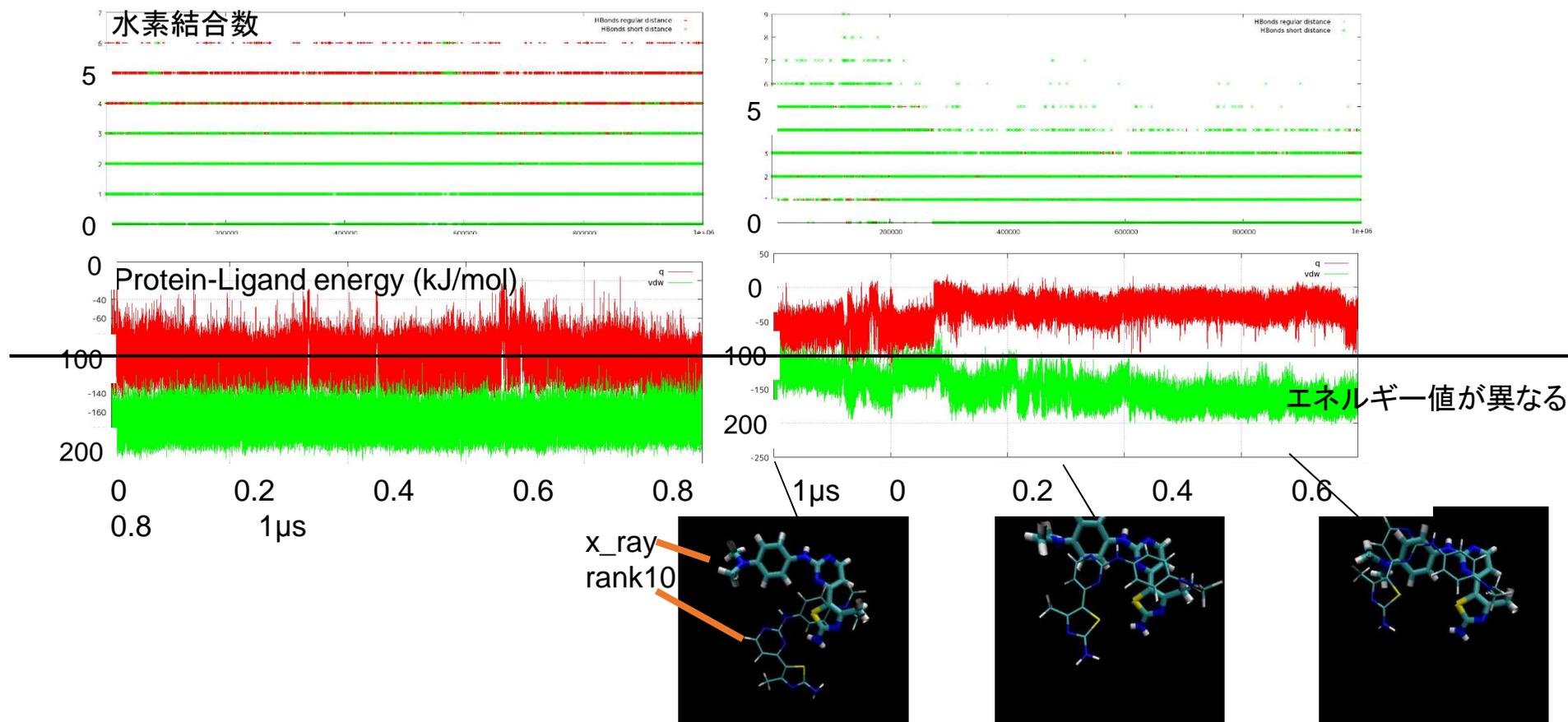
# 準安定構造からの長時間MDで最安定構造にたどり着けるか？



長時間MDによる化合物結合ポーズの追跡 (CDK2\_CS18)

結晶構造 (equil\_2)

docking rank10 (equil\_3)



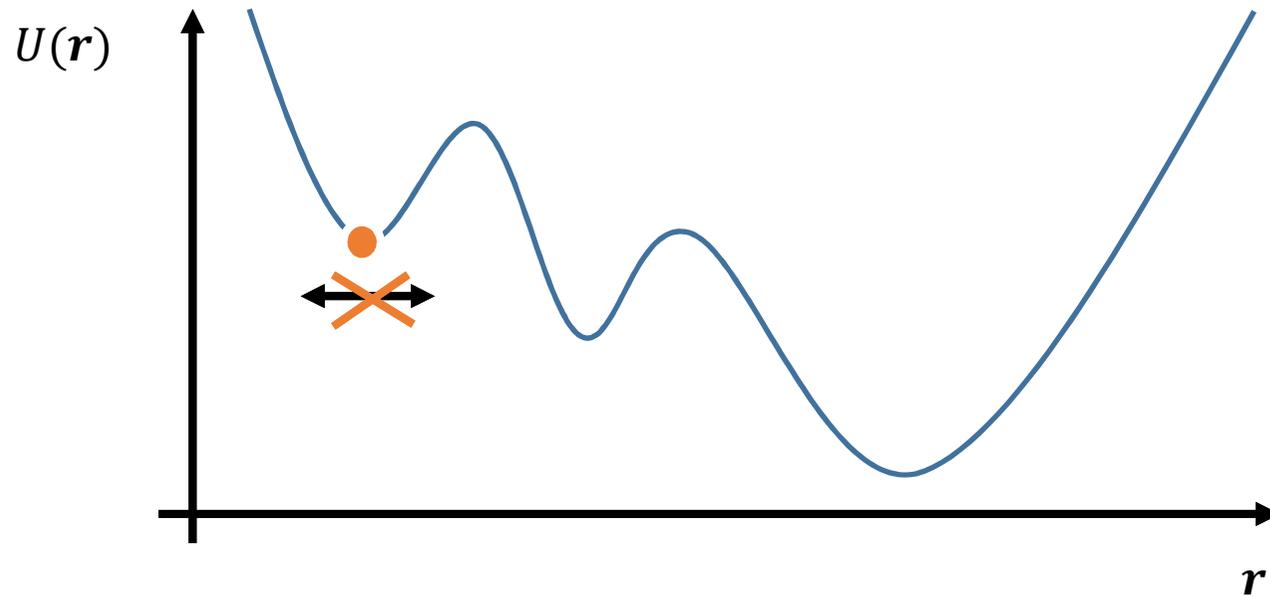
結合ポーズは1μs後も同じにならない

⇒正しい結合ポーズ(最安定構造)を予測するには温度レプリカ等の拡張サンプリングが必要

# トラップからの脱却



通常のMDシミュレーションでは、タンパク質の構造がエネルギー極小の領域にトラップされることがある



エネルギーが上がれば、トラップから抜け出すことができる  
原子の速度を上げる、温度を上げる

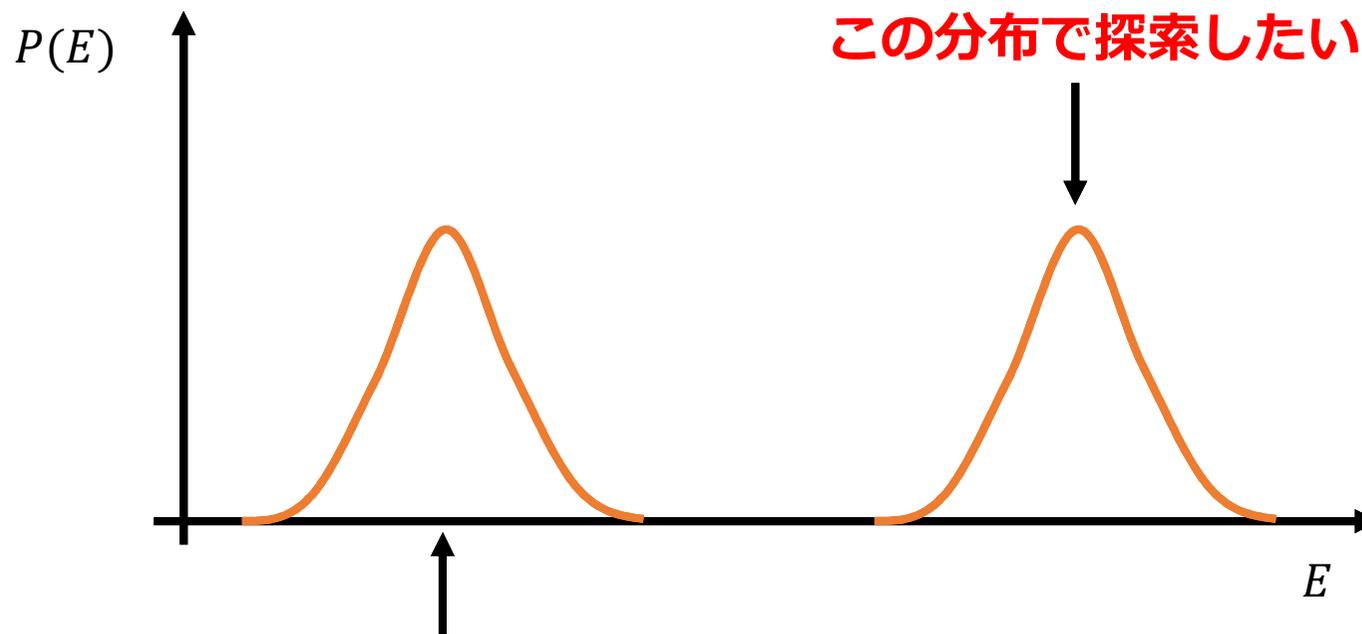


# カノニカルアンサンブルでのエネルギー分布



温度一定のMD

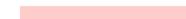
高エネルギーでの探索  
トラップから抜けられるが、  
求めるエネルギー(温度)での構造は  
得られない



**この分布で探索したい**

このエネルギー分布内で座標が移動  
なかなか高エネルギーにはならない

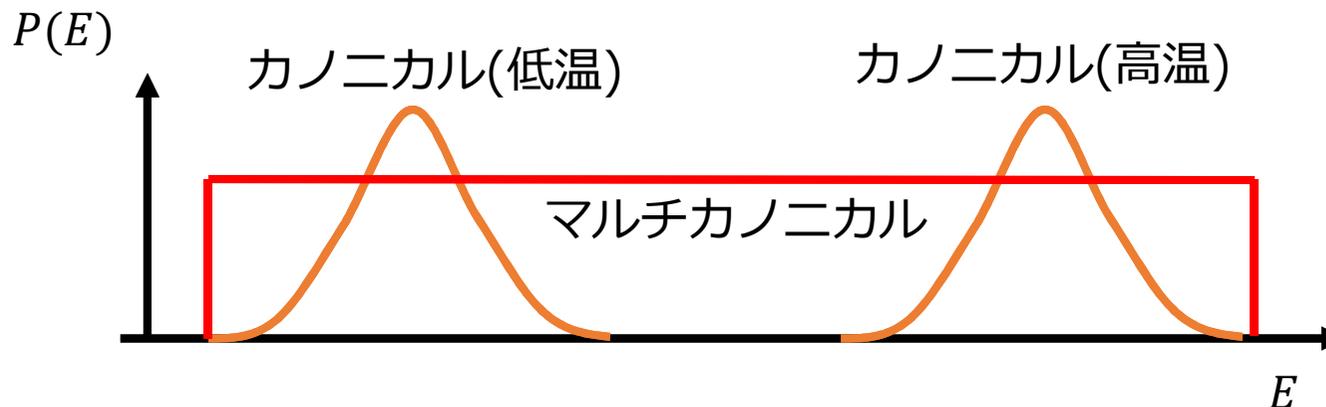
**この分布での構造が欲しい**



# 拡張サンプリング

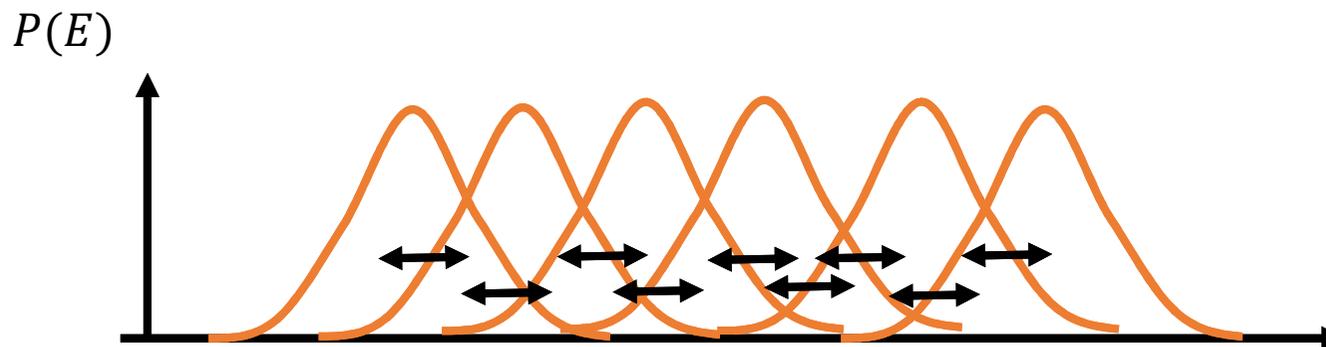


## マルチカノニカル法

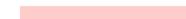


一様のエネルギー分布空間において、構造探索を行う

## レプリカ交換法



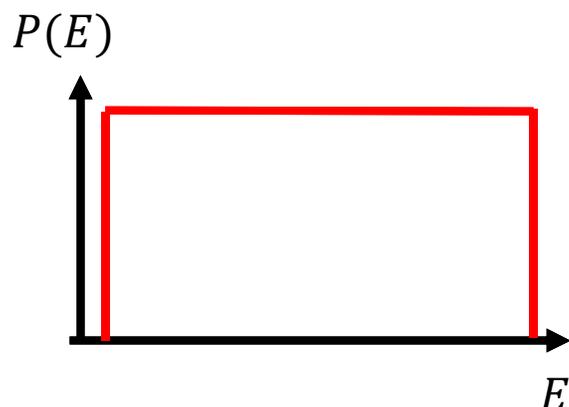
複数のエネルギー分布でMDを実行し、それらの間で温度または構造の交換を行いつつ、構造探索を行う



# マルチカノニカル法



一様分布になるように、人工的なエネルギー分布 $P_{mu}(E)$ を定義



$$P_{mu}(E) = n(E)W_{mu}(E) = \text{定数}$$

$$W_{mu}(E) = n(E)^{-1}$$

$n(E)$ : 状態密度

系の状態密度に反比例するように、  
マルチカノニカル重み因子 $W_{mu}(E)$ を定義

任意の温度 $T_0$ におけるマルチカノニカルポテンシャルエネルギー

$$E_{mu}(E; T_0) = k_B T_0 \ln n(E) = T_0 S(E)$$

マルチカノニカル法の運動方程式

$$\dot{\mathbf{p}}_k = \frac{\partial E_{mu}(E; T_0)}{\partial E} \mathbf{f}_k$$

ニュートンの運動方程式

$$\dot{\mathbf{p}}_k = \mathbf{f}_k$$



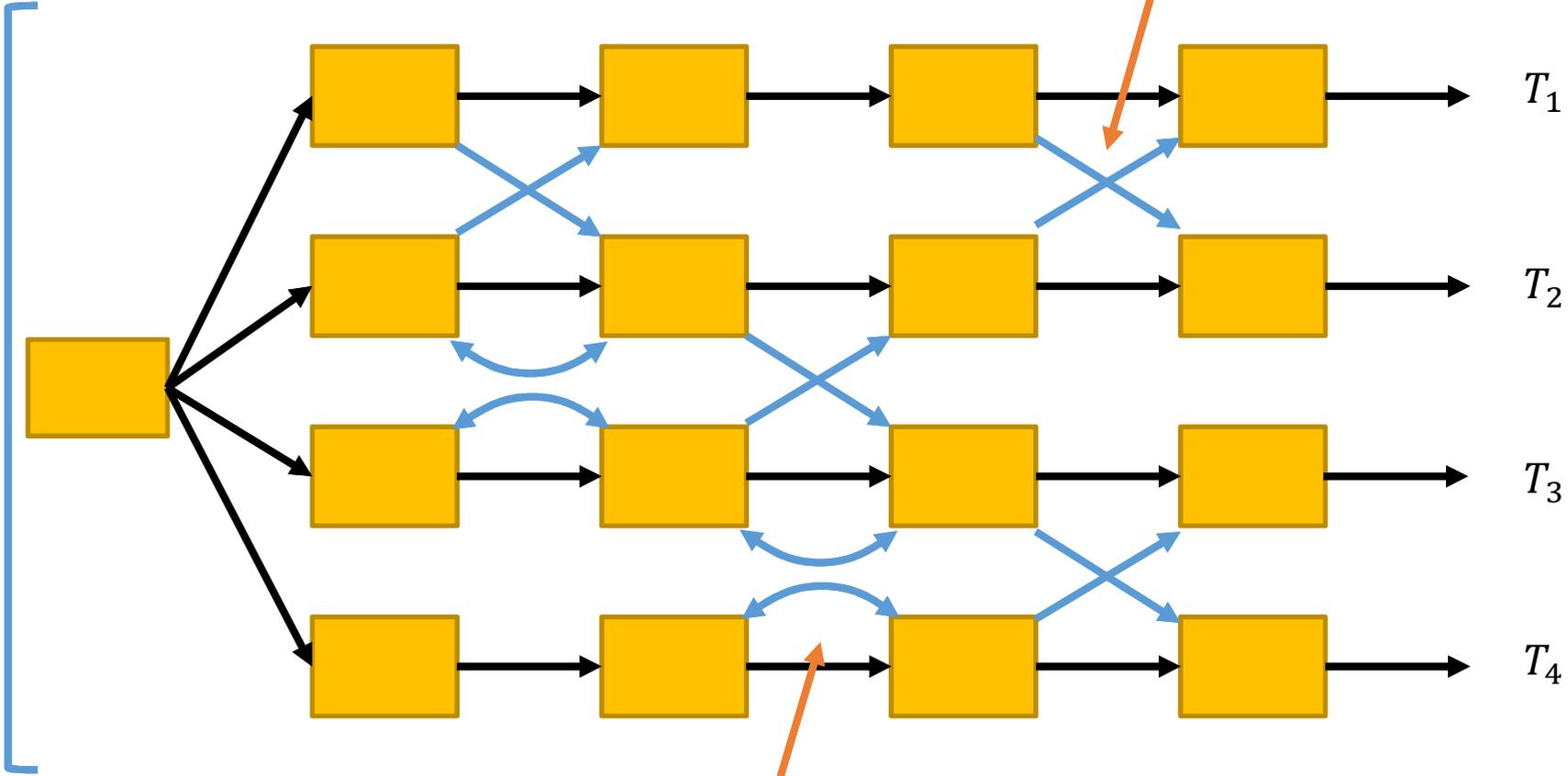
※事前に、短いMDでマルチカノニカル重み因子 $W_{mu}(E)$ を決定しておく必要がある





# レプリカ交換法

複数の温度で、MD計算を並列して実行する  
温度を交換する



同じ系のレプリカ  
温度が異なるシミュレーション

交換不成立

シミュレーション時間



# レプリカ交換法の手順



1. 互いに相互作用しない系のコピー(レプリカ)を複数用意する
2. 各レプリカで、異なる温度において独立にカノニカルシミュレーションを実行する
3. 温度値が隣接した二つのレプリカを、詳細つり合いの条件に基づいて交換する

$$\Delta = (\beta_m - \beta_n) \{E(q^{[j]}) - E(q^{[i]})\}$$

$$\text{遷移確率 } w(X \rightarrow X') = \begin{cases} 1 & \Delta \leq 0 \\ \exp(-\Delta) & \Delta > 0 \end{cases}$$

4. 2. 3.を繰り返す。

事前に重み因子を決定する必要がない。

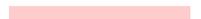
独立したMDシミュレーションを複数実行するため、並列コンピュータに適する。





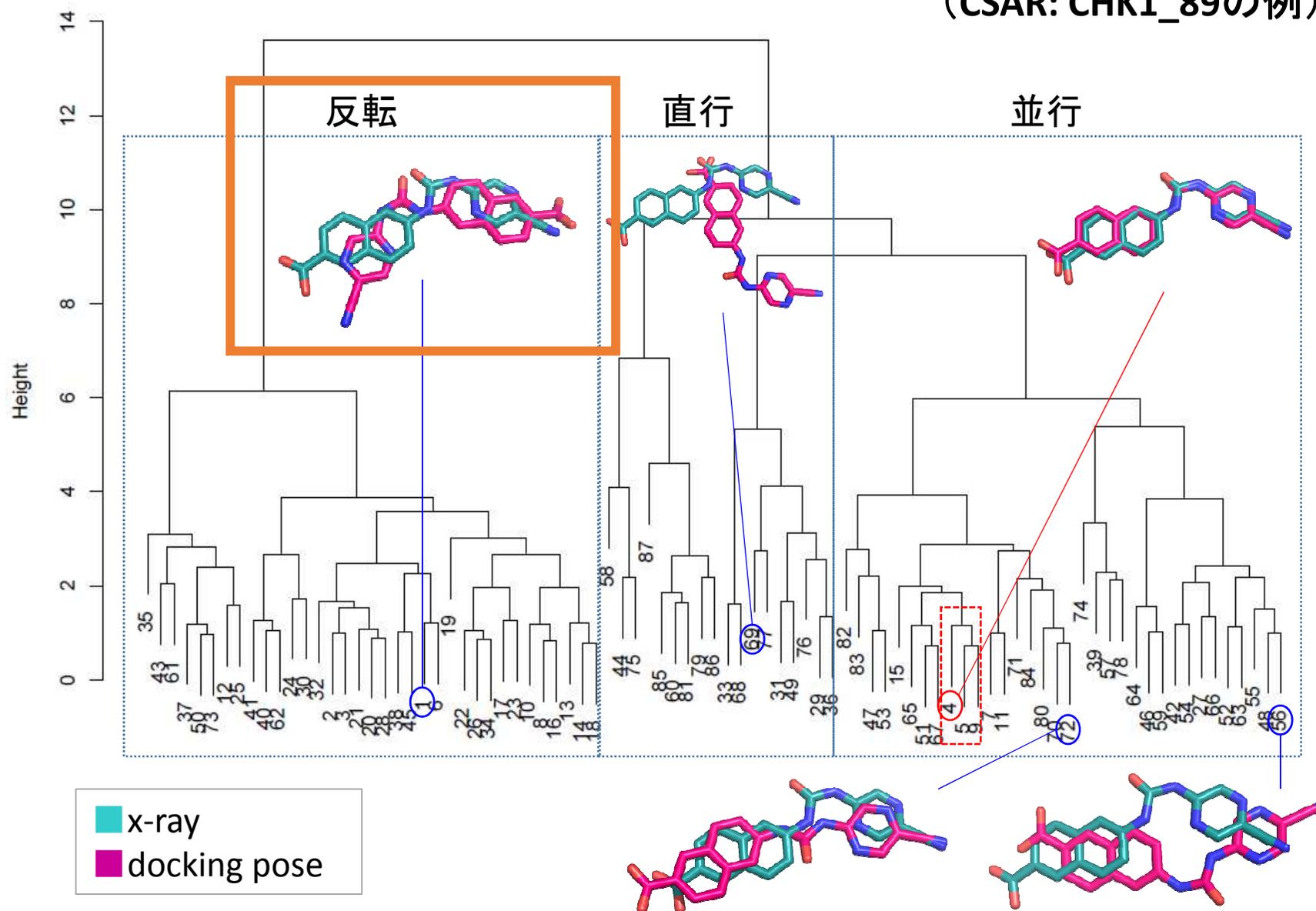
- KBDDプロジェクト

- CSAR CHK1\_89 re-docking #1からの、X線結晶構造の探索

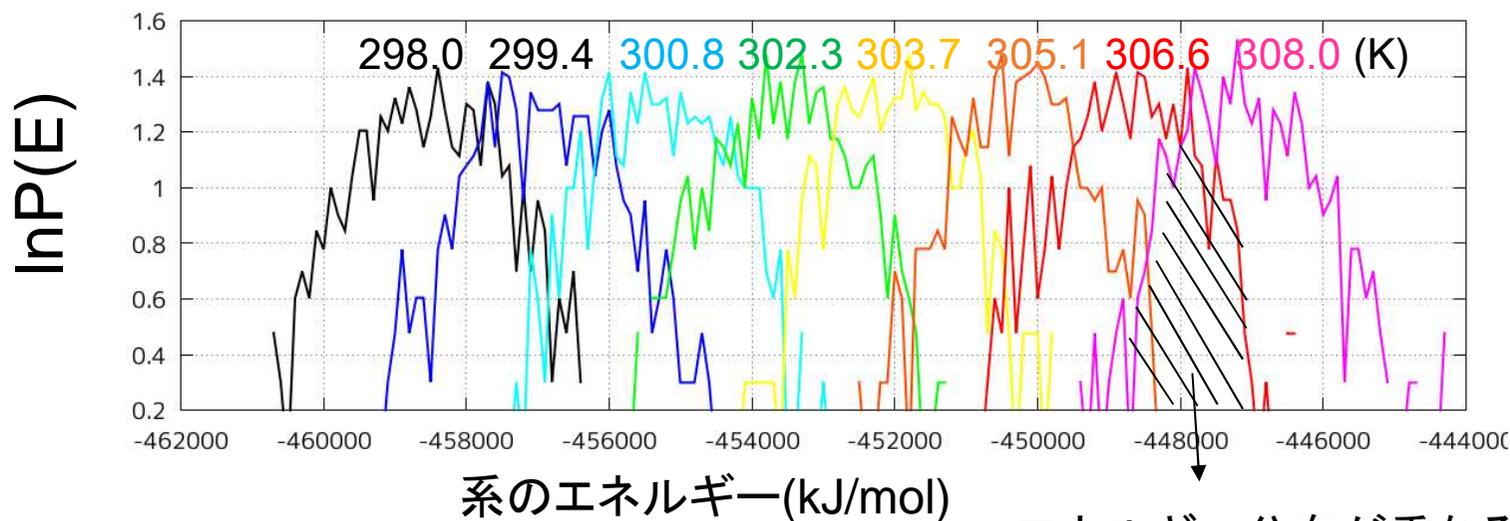


# Docking計算法で予測された結合ポーズとX線構造との比較

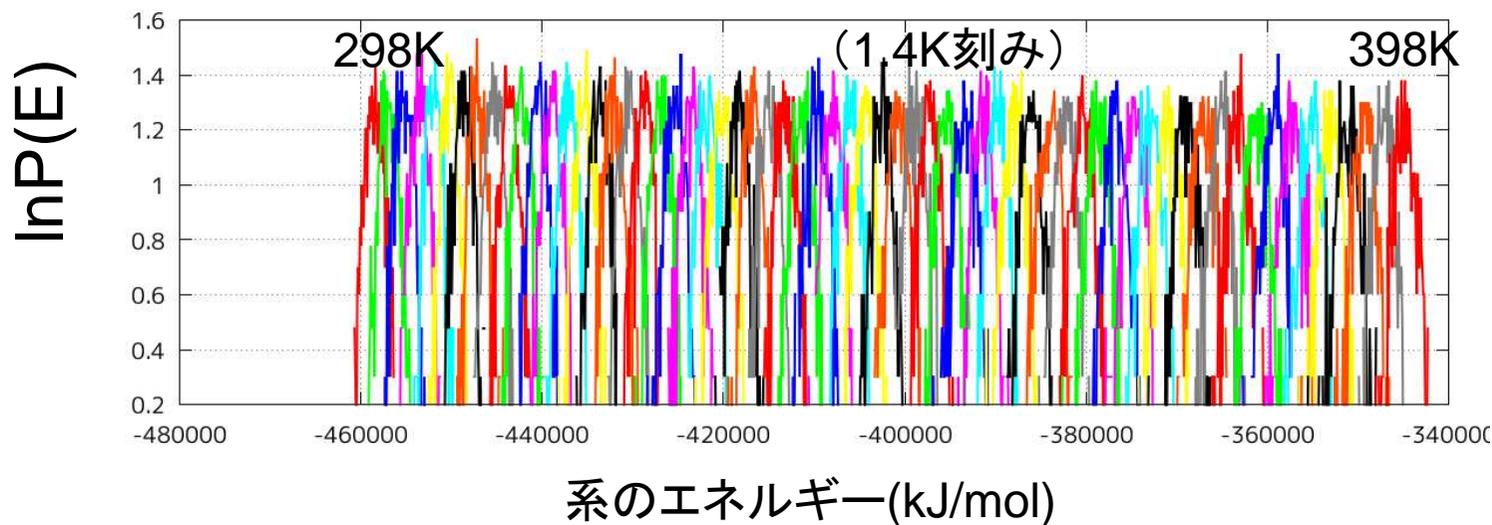
(CSAR: CHK1\_89の例)



# 様々な温度でのエネルギー分布 ～通常MD(NPT)でのtotal energyを使用～



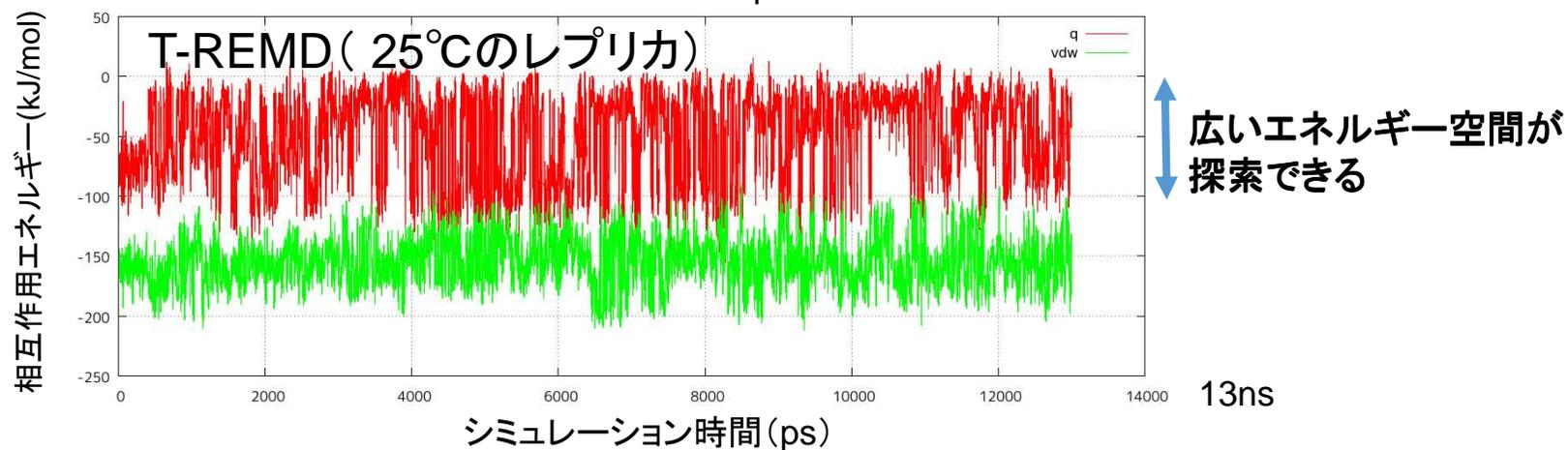
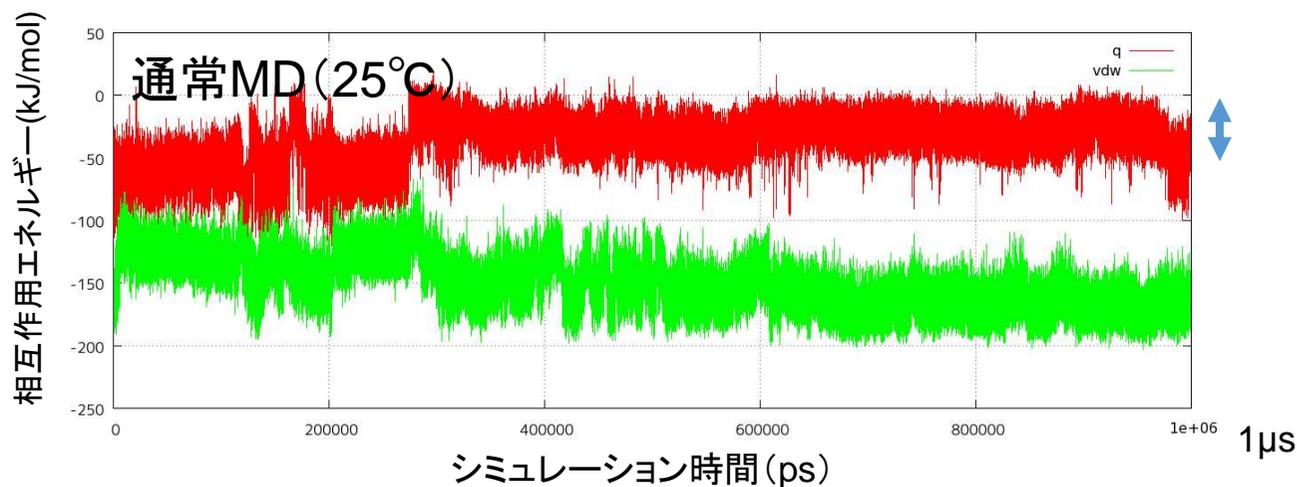
エネルギー分布が重なるように  
レプリカの温度を設定する必要がある



# レプリカ交換法による、 タンパク質-化合物相互作用エネルギーの経時変化



## タンパク質-化合物相互作用エネルギーの経時変化

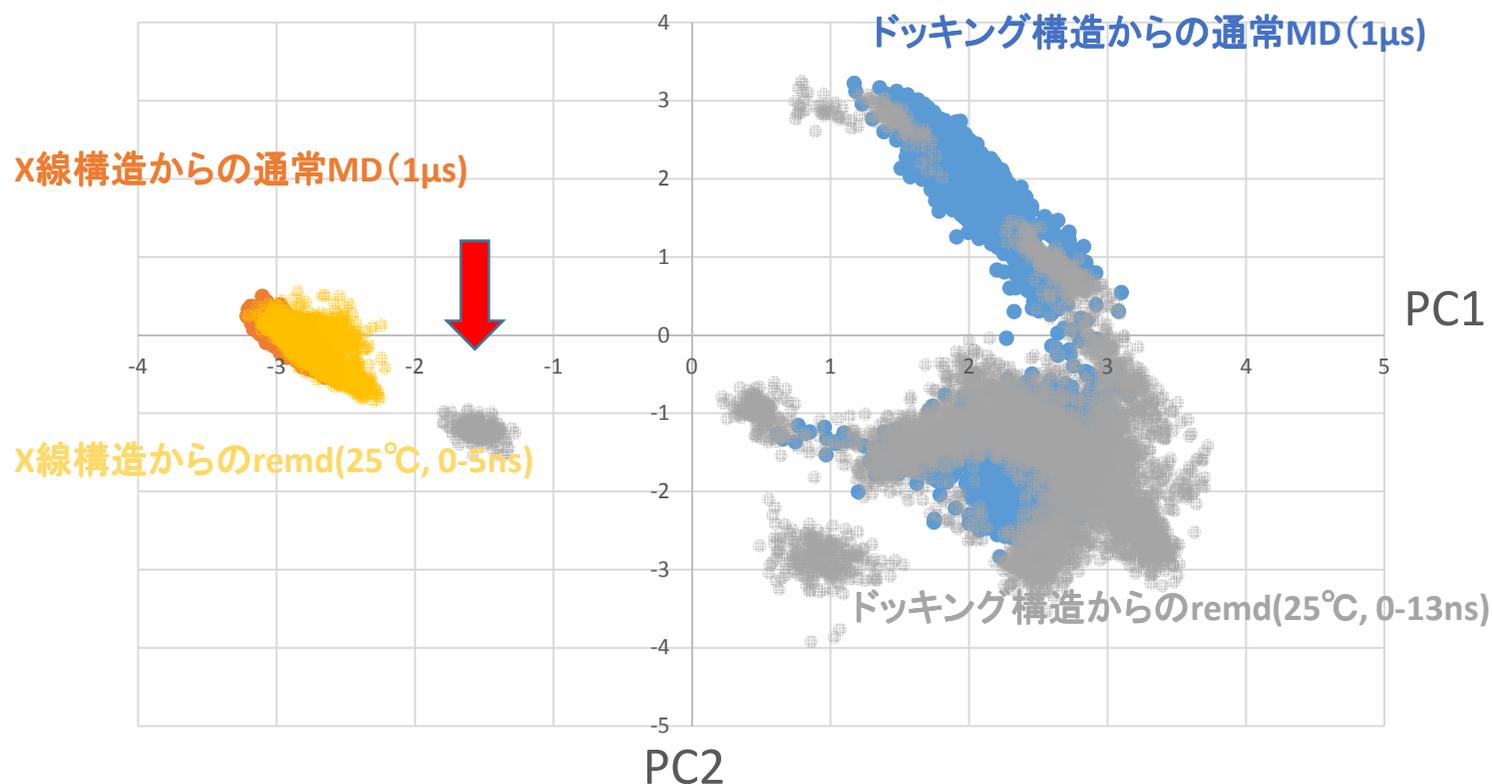


# 準安定構造からの温度レプリカによる構造サンプリング

T-REMD 91レプリカ(25°C~175°C)

for CDK2\_CS18

主成分分析結果(25°C, replica#0)



通常MDよりは広い構造空間を探索できており、X線構造のリガンド配座近傍の探索は行われているものの、たどり着いてはいない(現在温度幅を拡大して検証中)

# バイオグリッドHPCIプロジェクト「新薬開発を加速する「京」インシリコ創薬基盤の構築」 KBDD (K supercomputer-based drug discovery project by Biogrid pharma consortium)

「京」産業利用枠：新薬開発を加速する「京」インシリコ創薬基盤の構築



申請主体: NPO法人バイオグリッドセンター関西,

研究代表者 京都大学医学研究科 奥野恭史

製薬企業(23社): アステラス製薬(株)、アスピオファーマ(株)、エーザイ(株)、小野薬品工業(株)、科研製薬(株)、  
(株)カネカ、キッセイ薬品工業(株)、杏林製薬(株)、協和発酵キリン(株)、参天製薬(株)、塩野義製薬(株)、  
千寿製薬(株)、大正製薬(株)、大日本住友製薬(株)、田辺三菱製薬(株)、帝人ファーマ(株)、東レ(株)、  
日産化学工業(株)、日本新薬(株)、日本たばこ産業(株)、マルホ(株)、MeijiSeikaファルマ(株)、持田製薬(株)

IT企業(2社): (株)京都コンステラ・テクノロジーズ, 三井情報(株)

大学等: 京都大学大学院医学研究科, (独)産業技術総合研究所, 理研HPCI企画調整グループ

# 謝辞



## KBDDコンソーシアムメンバー

NPO法人 バイオグリッドセンター

奥野恭史 教授, 荒木望嗣 様 (理化学研究所計算科学研究機構), J. B. Brown 先生 (京都大学医学研究科), 白石慧 様, 佐藤美和 様 (三井情報(株)), 広川貴次 研究チーム長 ((独)産業技術総合研究所), 金井千里 主任研究員((株)京都コンステラテクノロジーズ)

## 製薬企業 (23社):

アステラス製薬(株)、アスピオファーマ(株)、エーザイ(株)、小野薬品工業(株)、  
科研製薬(株)、(株)カネカ、キッセイ薬品工業(株)、杏林製薬(株)、協和発酵キリン(株)、参天製薬(株)、  
塩野義製薬(株)、千寿製薬(株)、大正製薬(株)、大日本住友製薬(株)、田辺三菱製薬(株)、  
帝人ファーマ(株)、東レ(株)、日産化学工業(株)、日本新薬(株)、日本たばこ産業(株)、マルホ(株)、  
MeijiSeikaファルマ(株)、持田製薬(株)

## IT企業 (2社):

(株)京都コンステラ・テクノロジーズ, 三井情報(株)

(独) 産業技術総合研究所

理研・計算科学研究機構

高度情報科学技術研究機構



<http://sciencechannel.jst.go.jp/M120001/detail/M120001027.html>

JST サイエンスチャンネルにて放送中

