

積分方程式理論の GPUによる高速化と 創薬への応用

大阪大学 蛋白質研究所

丸山 豊

drmaruyama@protein.osaka-u.ac.jp

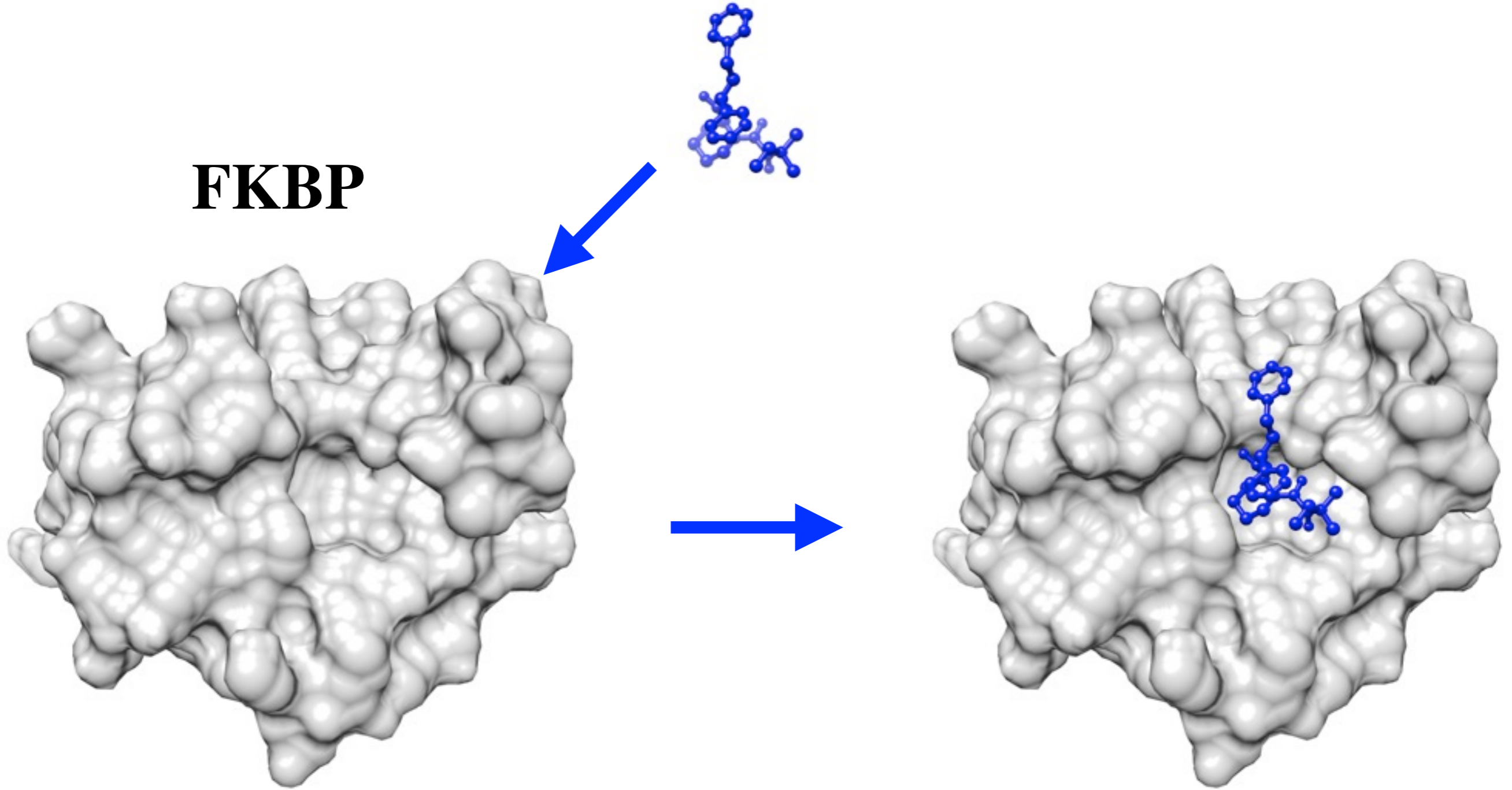
1. 創薬の課題

2. 積分方程式理論

3. 創薬への適用

蛋白質と薬効分子の結合

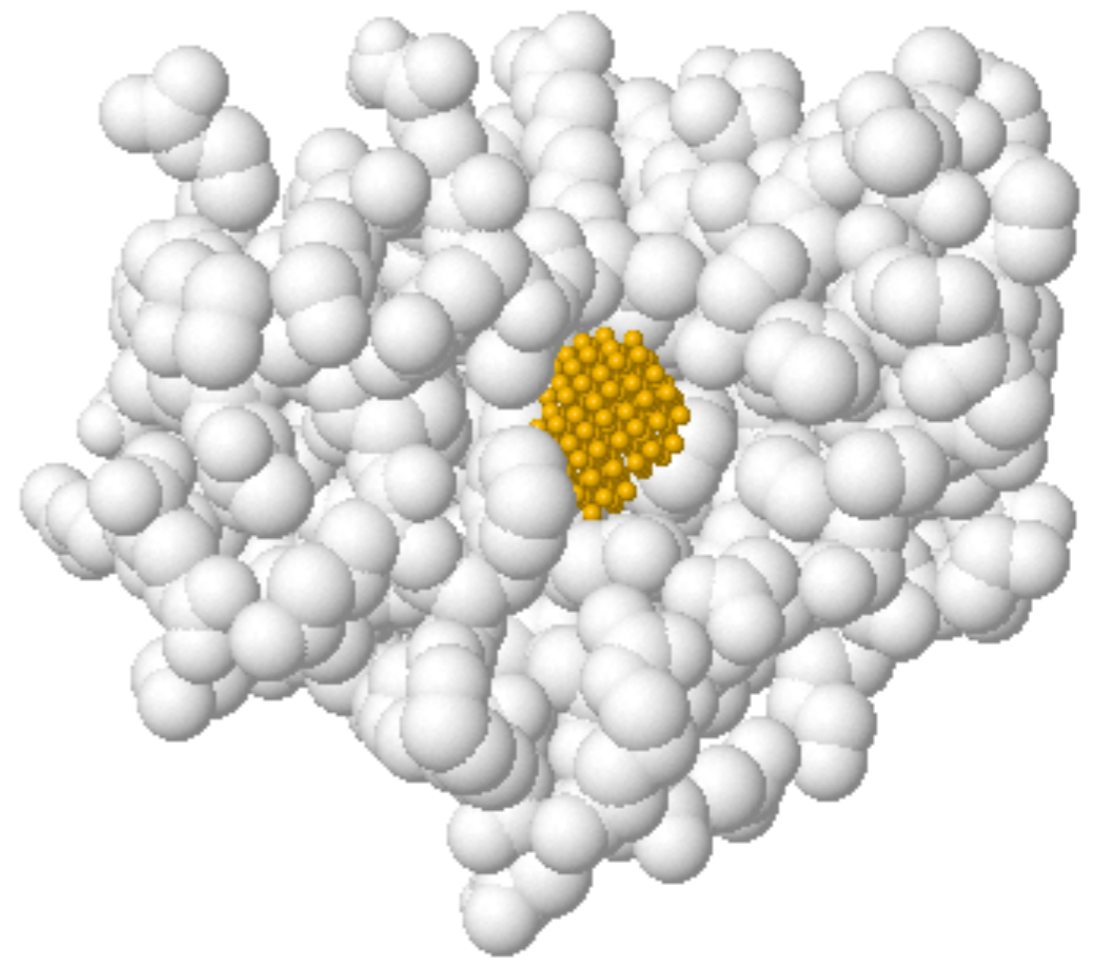
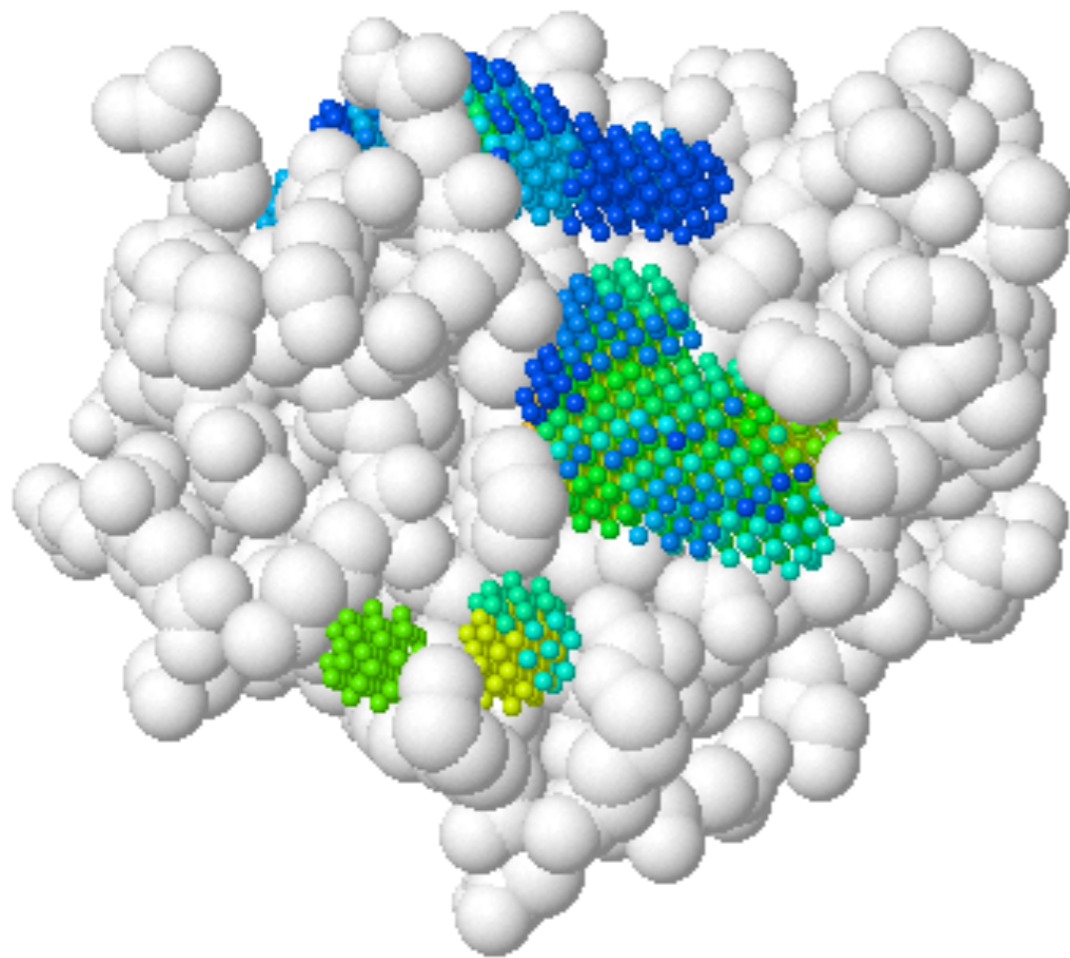
FKBP



結合位置をどうやって調べるか？

従来法による結合位置探索

球の大きさ: 橙 < 黄緑 < 緑 < 青



複数の結合位置を示唆

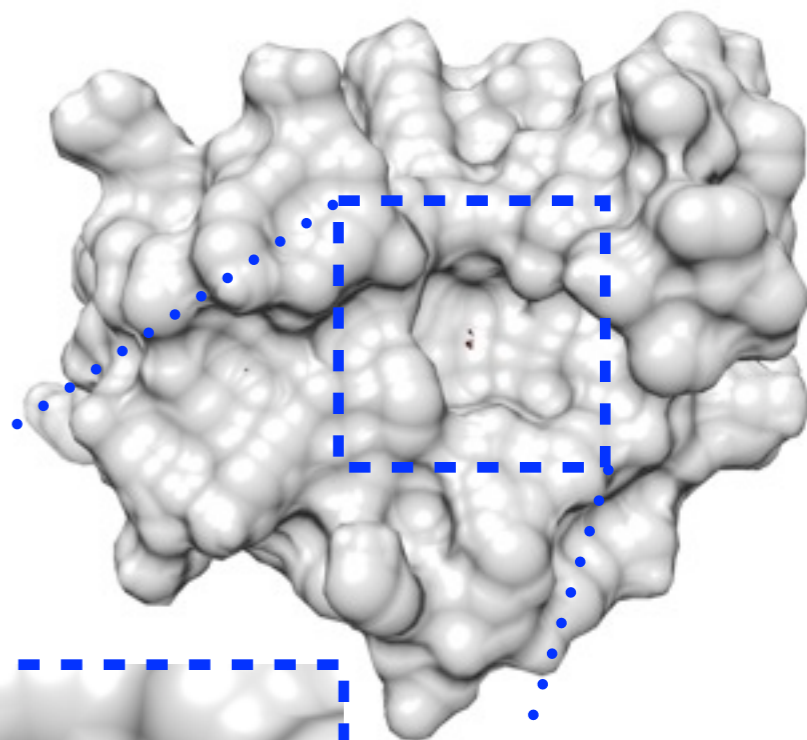
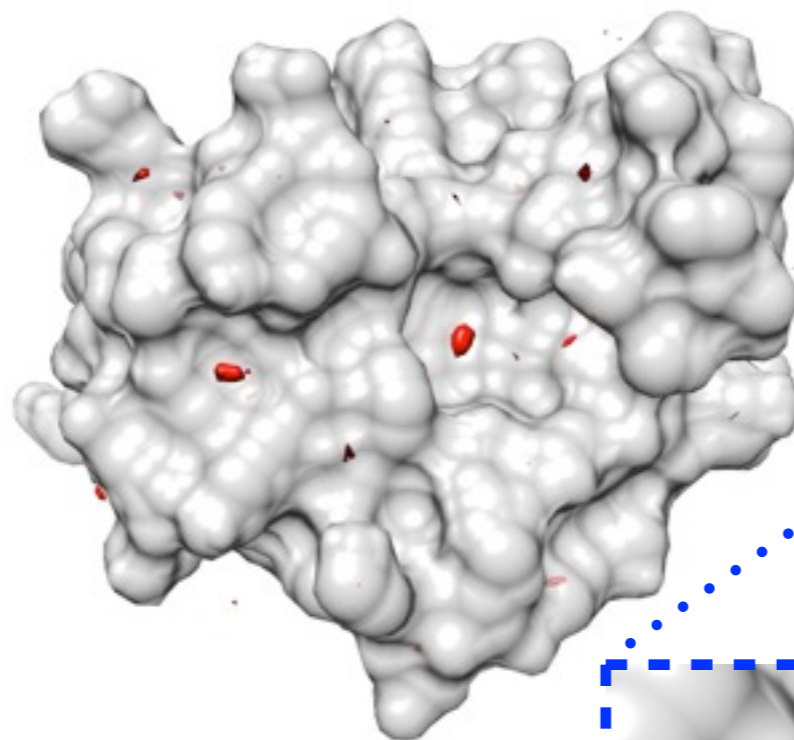
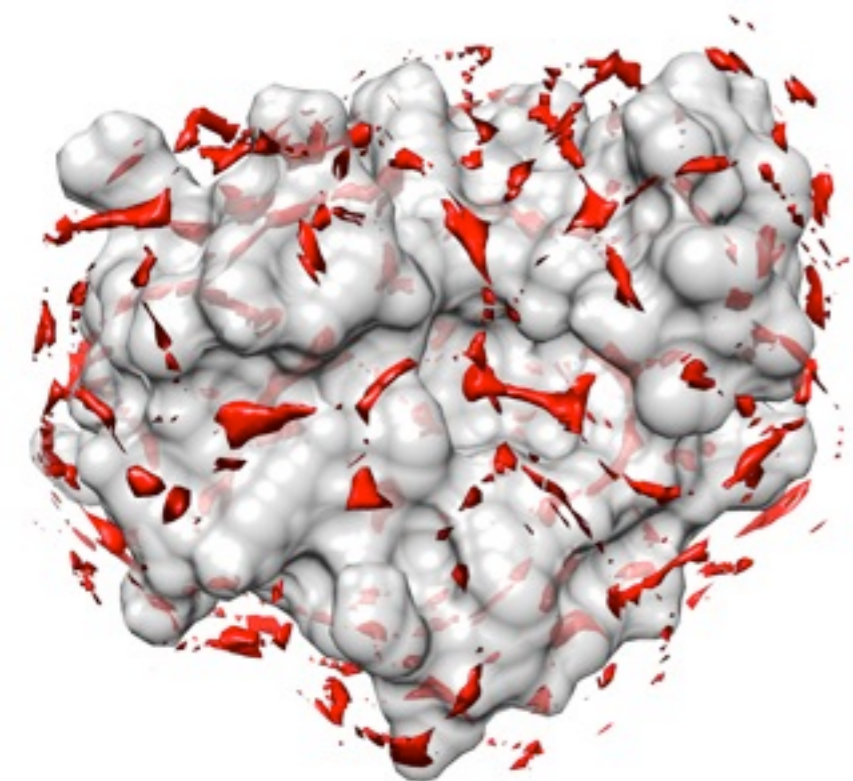
GHECOM 1.0 (川端、阪大蛋白研)

積分方程式理論による結合位置の予測

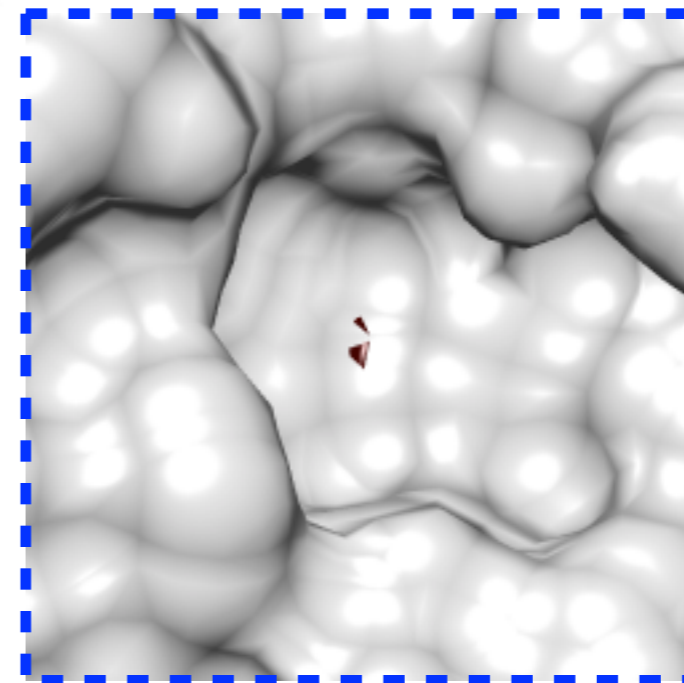
$g(r) > 3$

$g(r) > 5$

$g(r) > 7$



最も可能性の高い位置



1. 創薬の課題

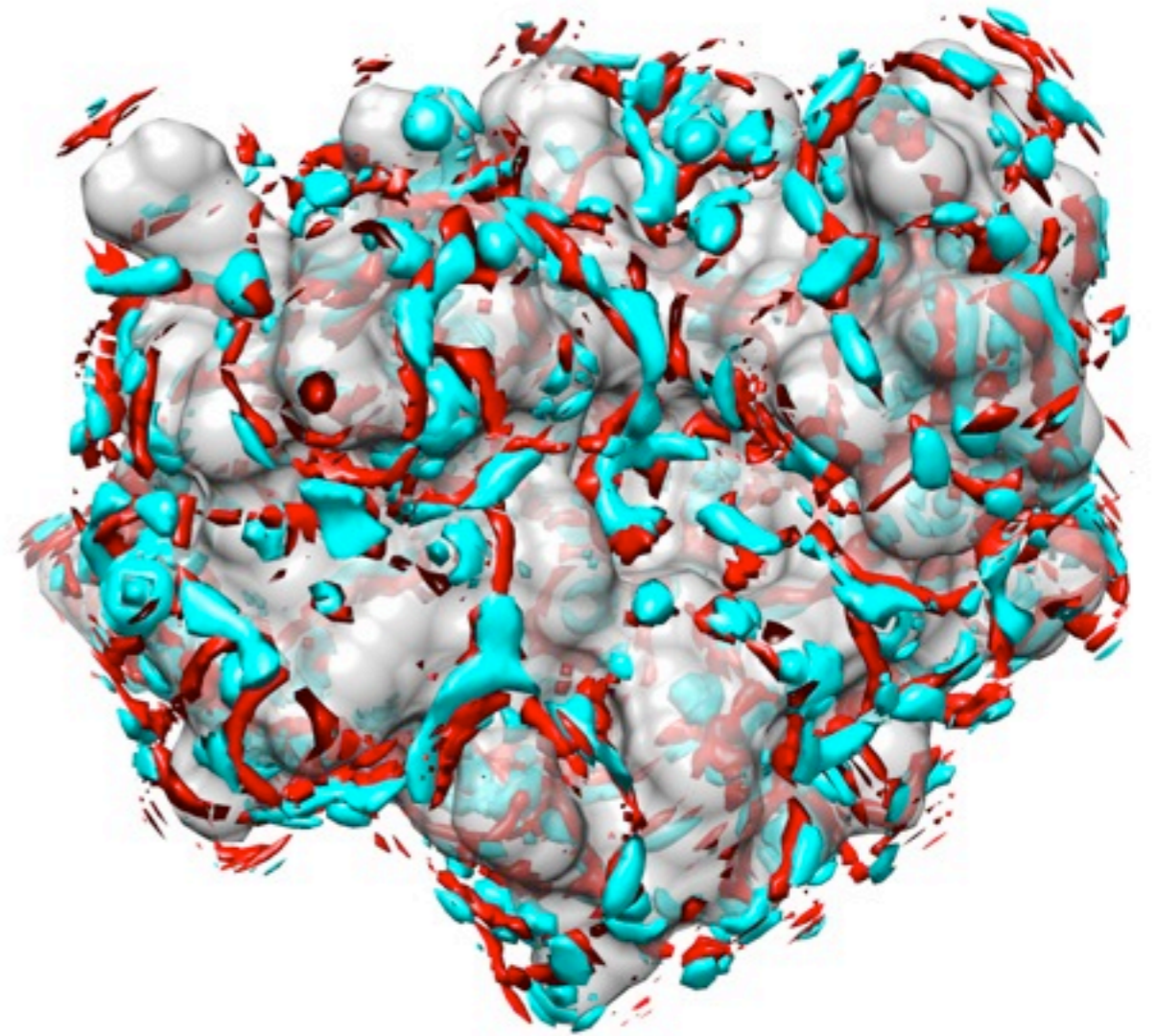
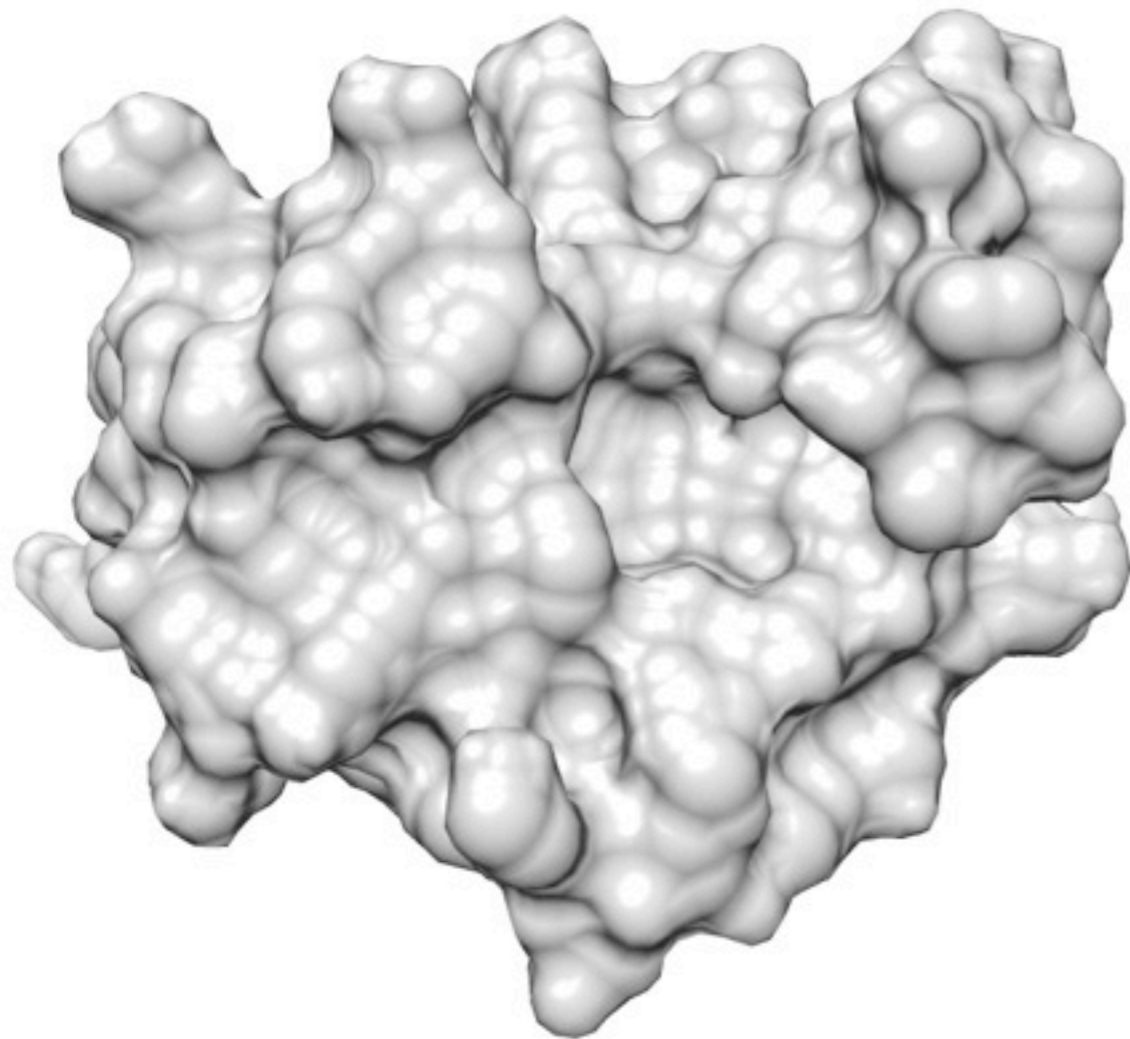
2. 積分方程式理論

3. 創薬への適用

蛋白質の周りの水の分布

水の酸素

水の水素



水素結合などの分子的特徴を記述

統計力学的平均

熱力学量の計算

溶媒和自由エネルギー

$$\Delta\mu = \frac{\rho}{\beta} \int_0^\infty \left[\frac{1}{2} h(r) - c(r) - \frac{1}{2} h(r)c(r) \right] \cdot 4\pi r^2 dr$$

圧縮率

$$\chi_T = \frac{1}{\rho k_B T} + \frac{1}{k_B T} \int_0^\infty (g(r) - 1) \cdot 4\pi r^2 dr$$

部分モル体積

$$\bar{V} = \chi_T k_B T - \int_0^\infty (g_{uv}(r) - 1) \cdot 4\pi r^2 dr$$

実験との対応

3D-RISM理論のプログラム

(旧) 分子研平田グループ

- OpenMP版
- 超並列版 (京スパコン)
- **CUDA版**

商用ソフトウェア

- Amber
- ADF
- MOE

3D-RISM理論のプログラム

(旧) 分子研平田グループ

- OpenMP版
- 超並列版 (京スパコン)
- **CUDA版** ← (丸山、平田 JCTC 2012)

商用ソフトウェア

- Amber
- ADF
- MOE

3D-RISM理論

3D-RISM方程式

$$h_{\alpha}(\mathbf{r}) = \sum_{\gamma} \int d\mathbf{r}' \mathbf{c}_{\gamma}(\mathbf{r} - \mathbf{r}') \chi_{\gamma\alpha}(\mathbf{r}')$$

畳み込み積分

$$h_{\alpha}(\mathbf{k}) = \sum_{\gamma} \mathbf{c}_{\gamma}(\mathbf{k}) \chi_{\gamma\alpha}(\mathbf{k})$$

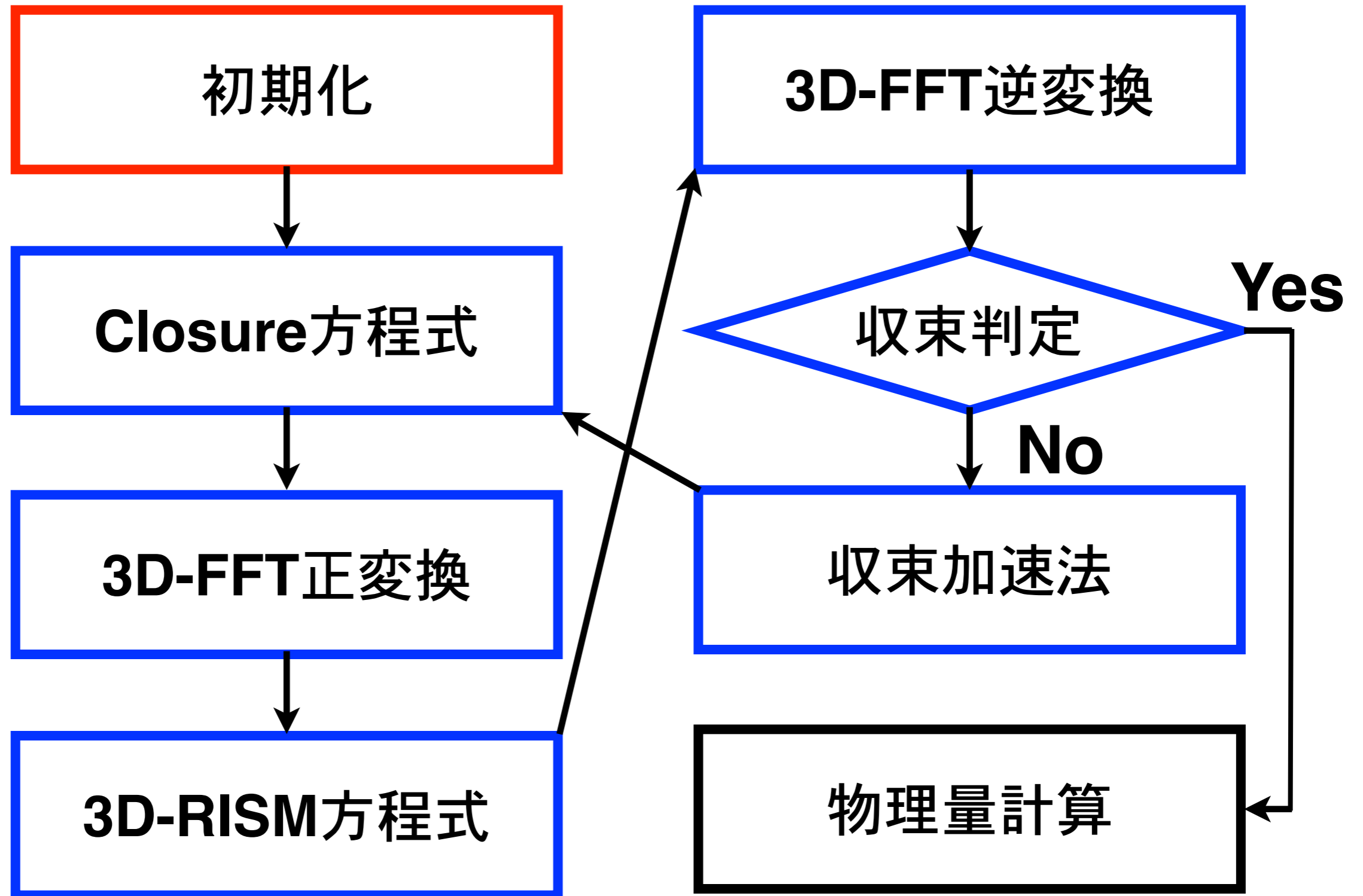
Closure方程式

3D-FFT

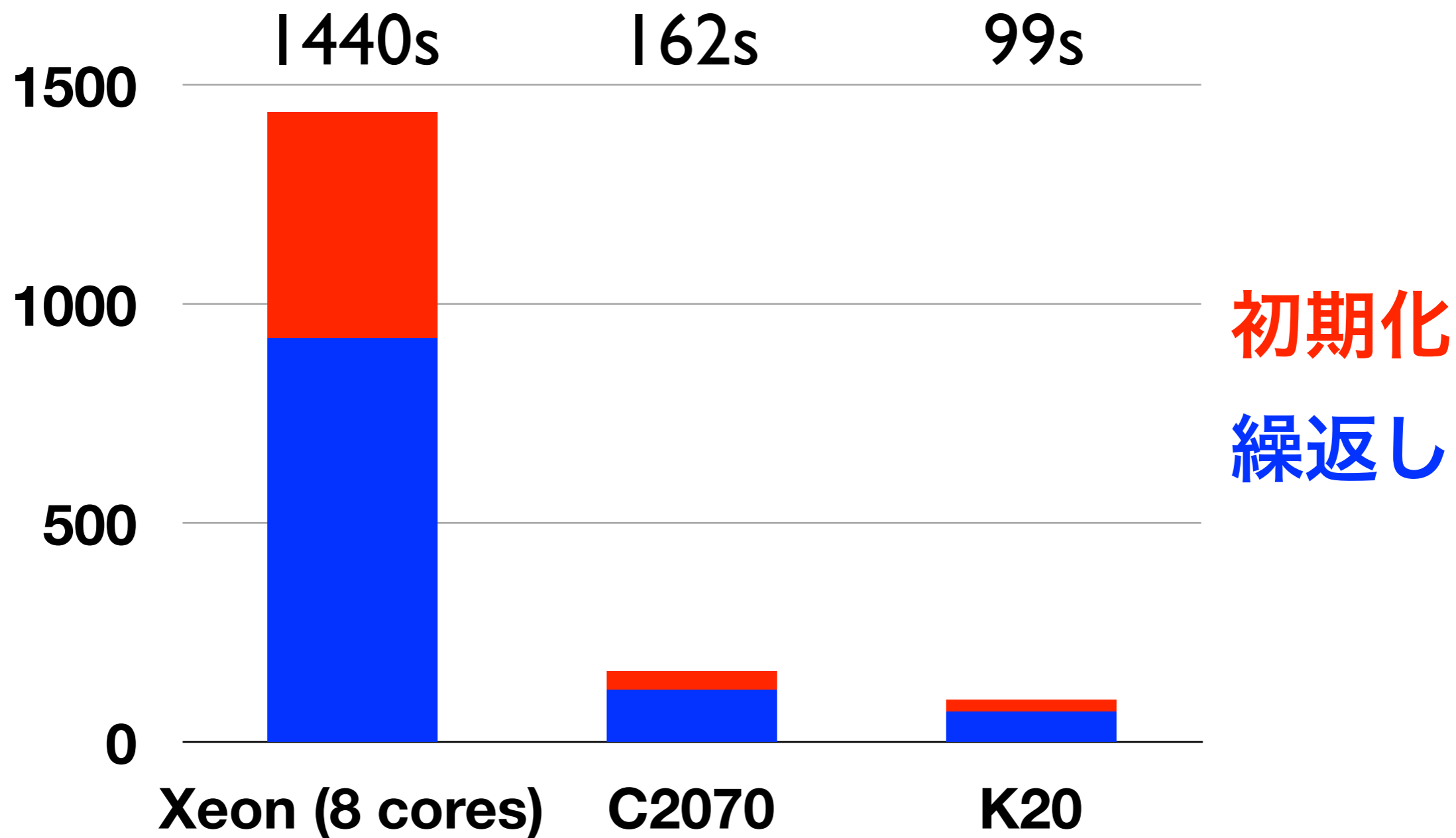
$$h_{\alpha}(\mathbf{r}) = \exp(-\beta \mathbf{u}_{\alpha}(\mathbf{r}) + \mathbf{h}_{\alpha}(\mathbf{r}) - \mathbf{c}_{\alpha}(\mathbf{r})) - 1$$

3D-RISM方程式とClosure方程式を交互に計算

解法の流れ

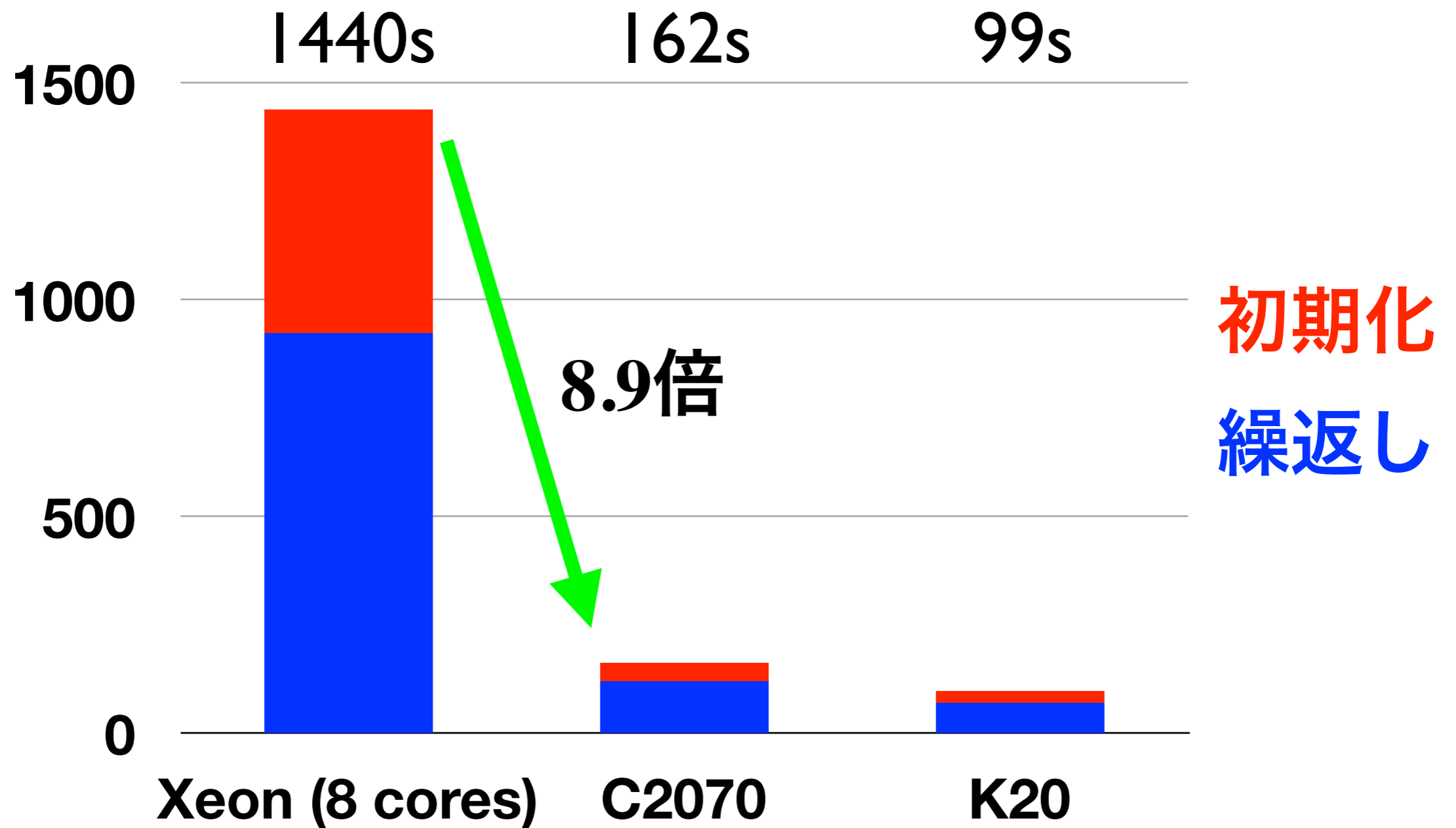


計算時間 (FKBPの周りの水)



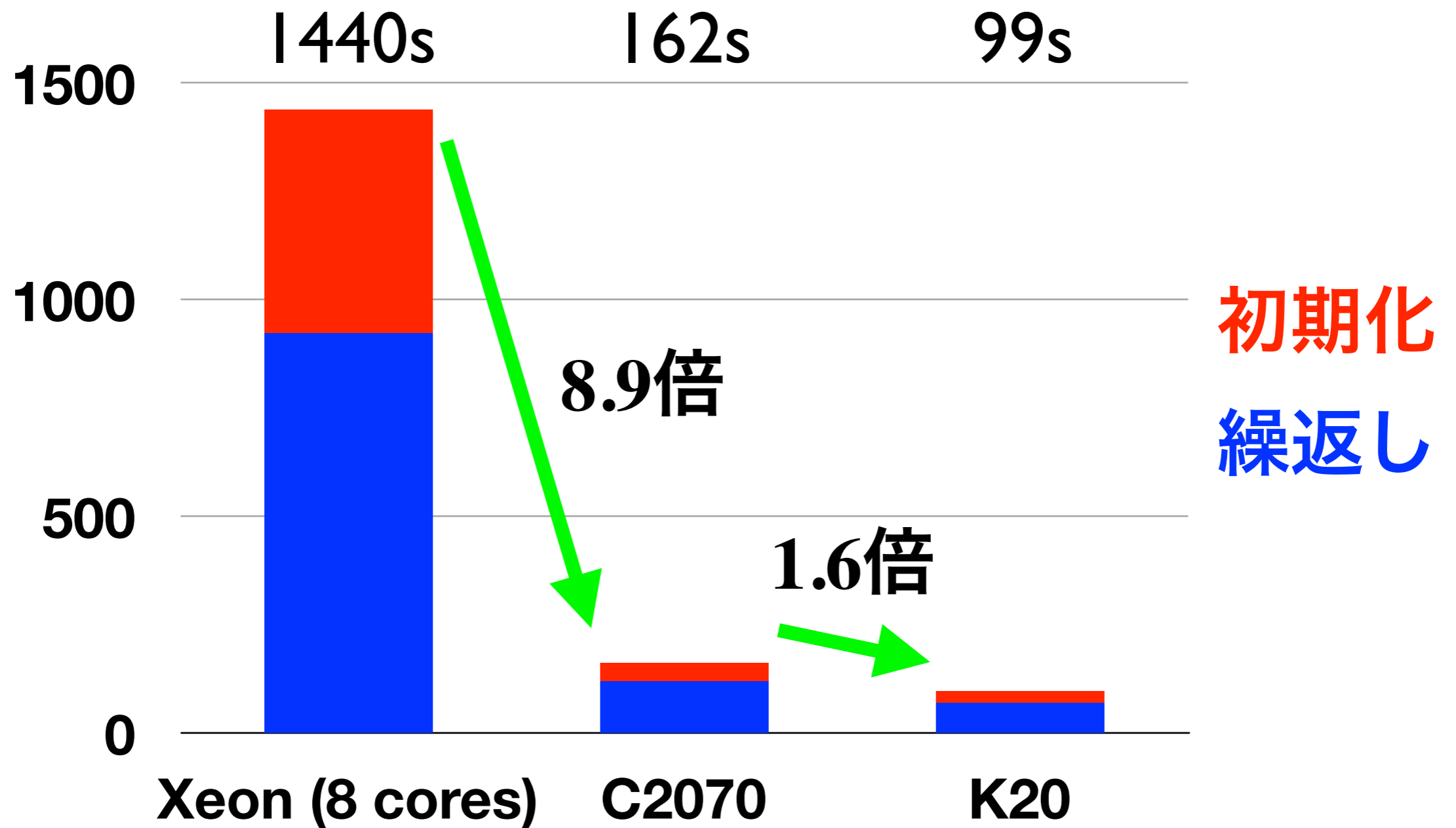
Xeon E5-2650 2.5GHz, gcc 4.6.1, FFTW 3.2.2
CUDA5.0, Thrust

計算時間 (FKBPの周りの水)



Xeon E5-2650 2.5GHz, gcc 4.6.1, FFTW 3.2.2
CUDA5.0, Thrust

計算時間 (FKBPの周りの水)



Xeon E5-2650 2.5GHz, gcc 4.6.1, FFTW 3.2.2
CUDA5.0, Thrust

メモリバンド幅

GPU	メモリバンド幅(GB/sec)
-----	-----------------

C2070	144
-------	-----

K20	208
-----	-----

C2070からK20で**1.44**倍

速度向上は**1.6**倍

プロセス毎の計算時間の割合(%)

	C2070	K20
初期化	22.5	23.5
3D-RISM	31.6	36.9
収束加速法	12.0	14.7
FFT(cufft)	32.7	22.9
Data転送	0.5	1.2

C2070からK20

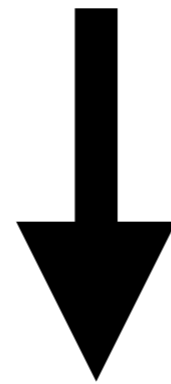
バンド幅1.44倍で速度向上は1.6倍

**一方、ライブラリで提供されるcufftの
計算時間における割合は減少**

C2070からK20

バンド幅1.44倍で速度向上は1.6倍

一方、ライブラリで提供されるcufftの
計算時間における割合は減少

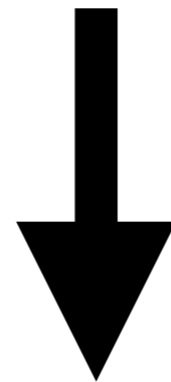


自作コード部分はK20のパフォーマンス
を引き出せていない？

C2070からK20

バンド幅1.44倍で速度向上は1.6倍

一方、ライブラリで提供されるcufftの
計算時間における割合は減少



自作コード部分はK20のパフォーマンス
を引き出せていない？

今後の課題

1. 創薬の課題
2. 積分方程式理論
- 3. 創薬への適用**

1. 薬効分子が不明

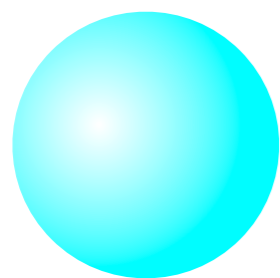
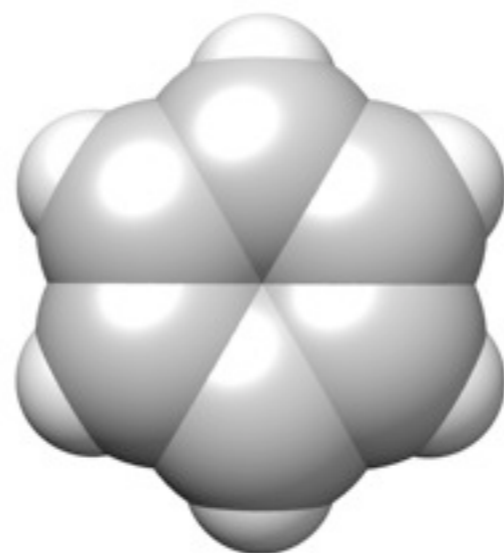
2. 薬効分子が明瞭

分子の粗視化

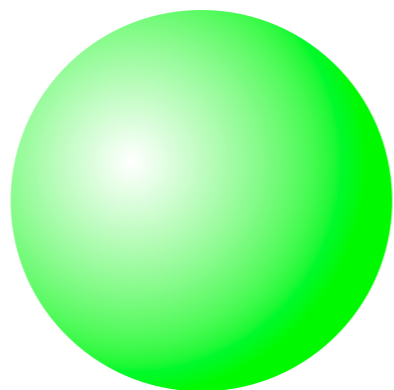
水



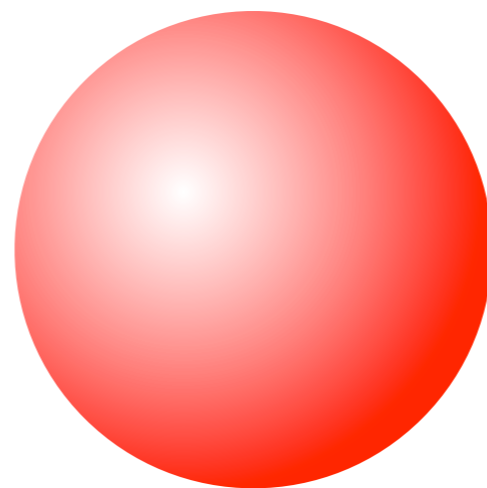
ベンゼン



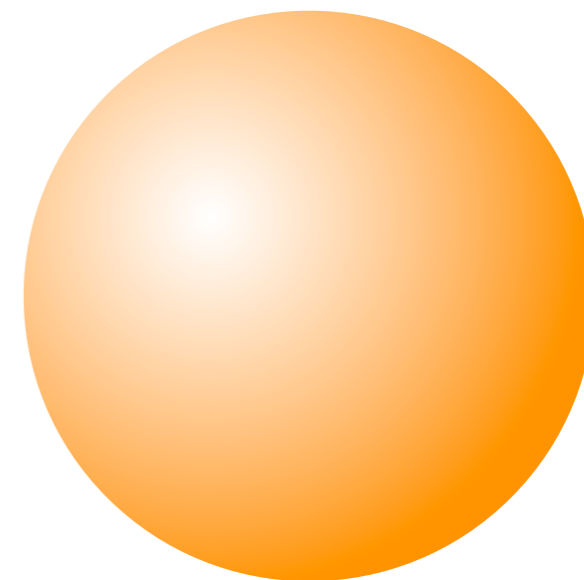
2.8Å



4.0Å



5.0Å



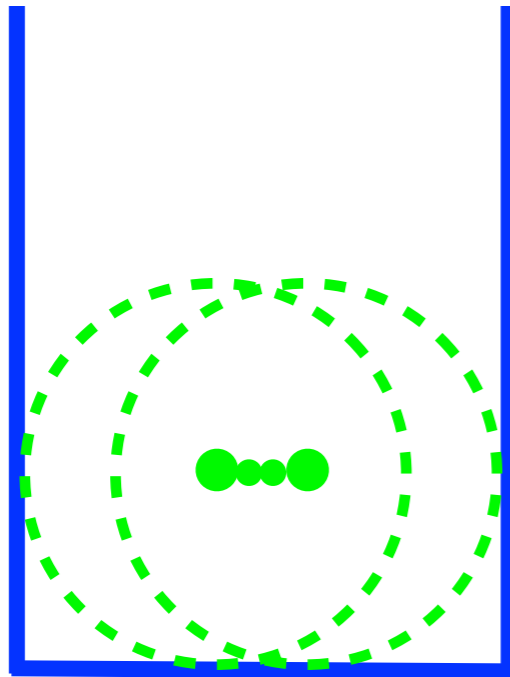
6.0Å

計算条件

- 引力のないソフトコアのポテンシャル
- 各大きさの球を同時に計算
- 2.8Åの球で水の効果

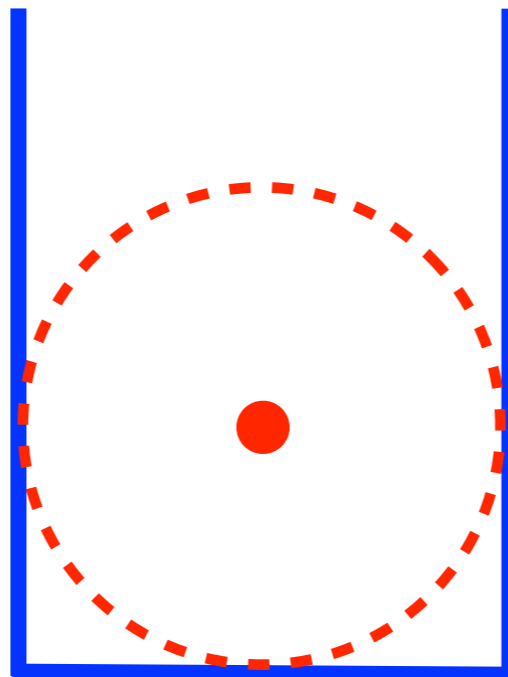
球の大きさとピークの関係

小



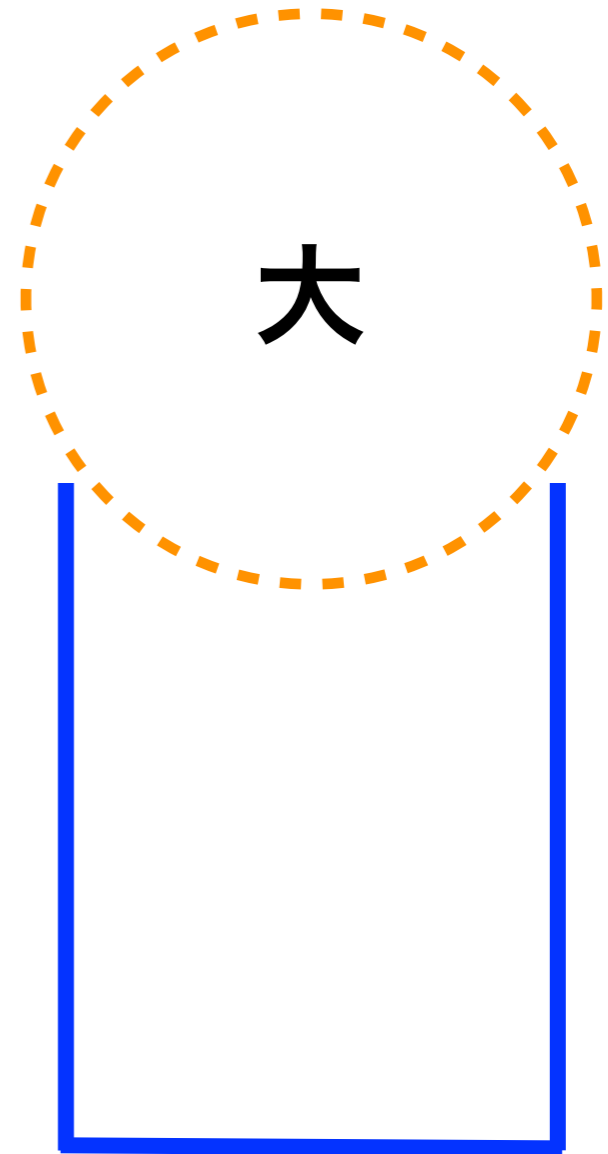
ピークが分散

適



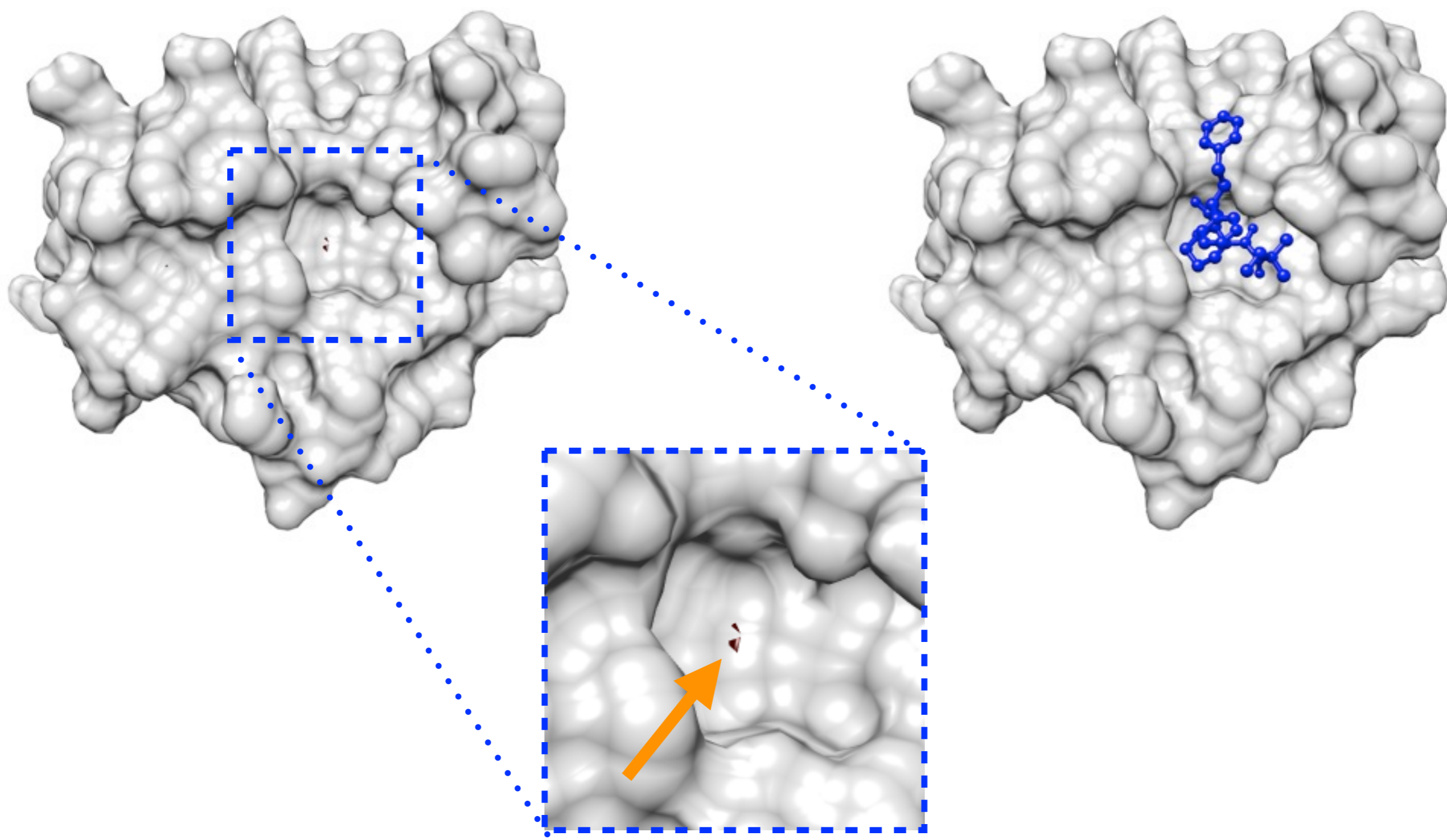
強ピーク

大



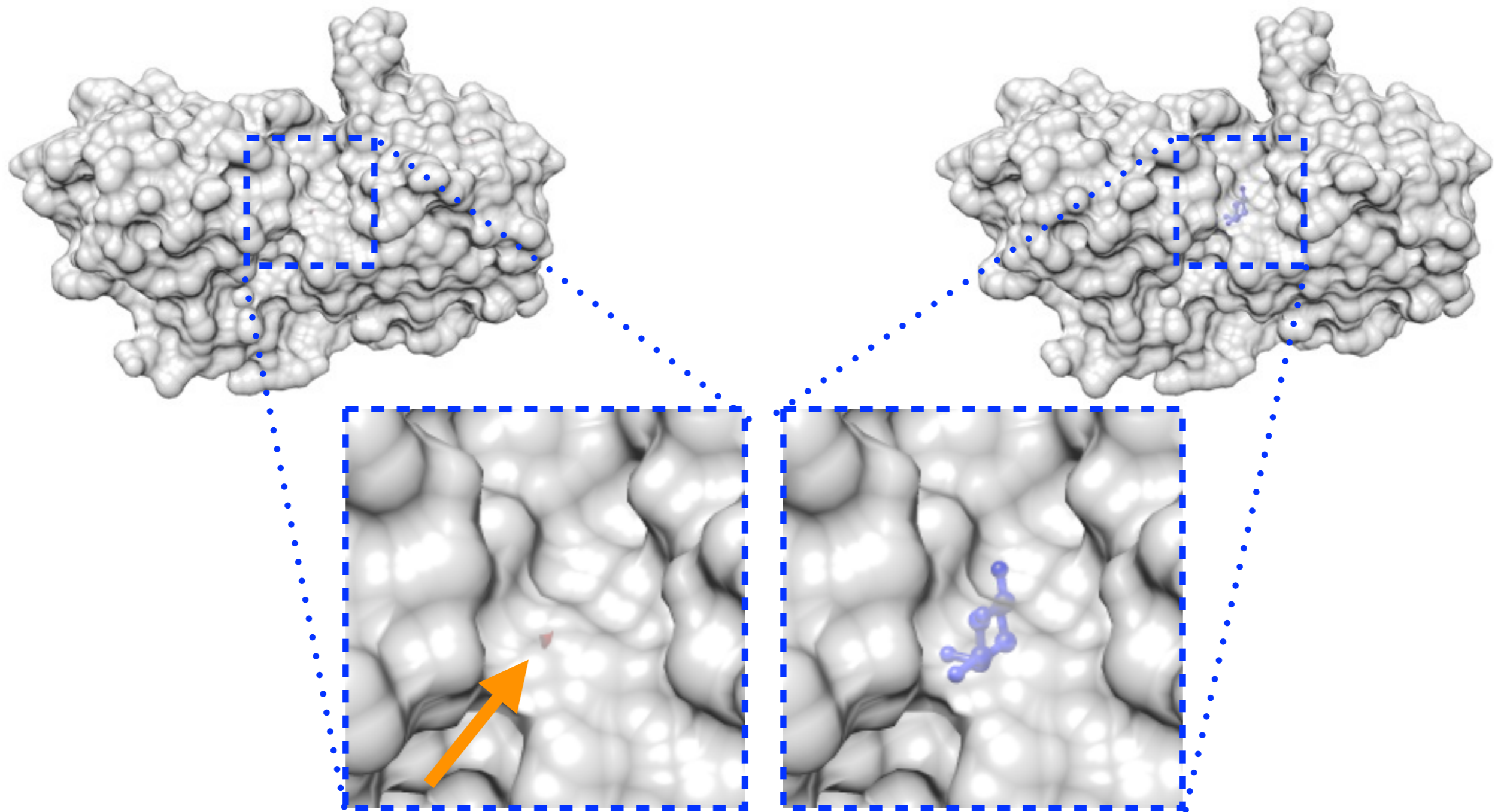
ピーク無し

FKBP



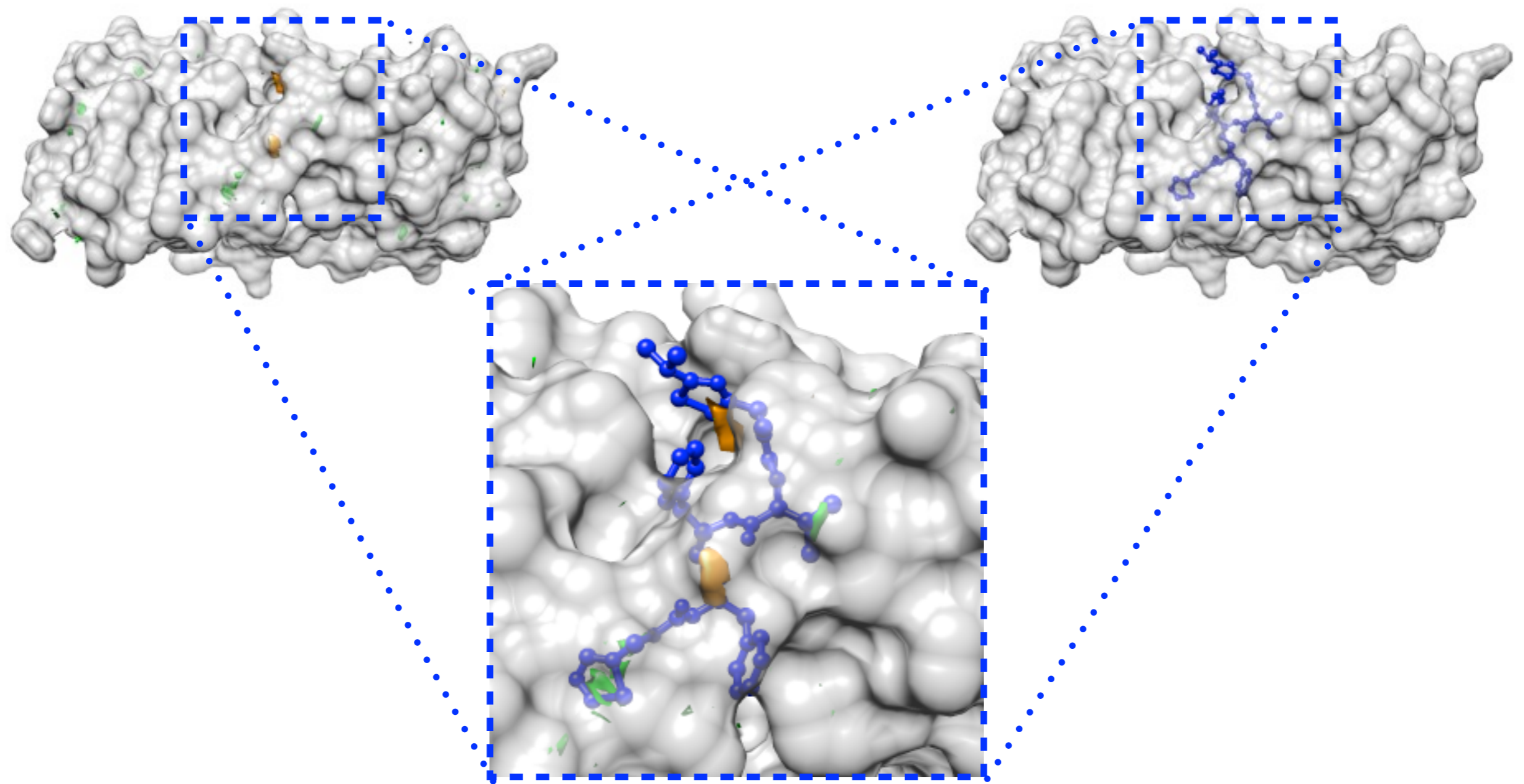
結合位置の予測

アラビノース結合蛋白質



蛋白質内部の薬効分子の位置を予測

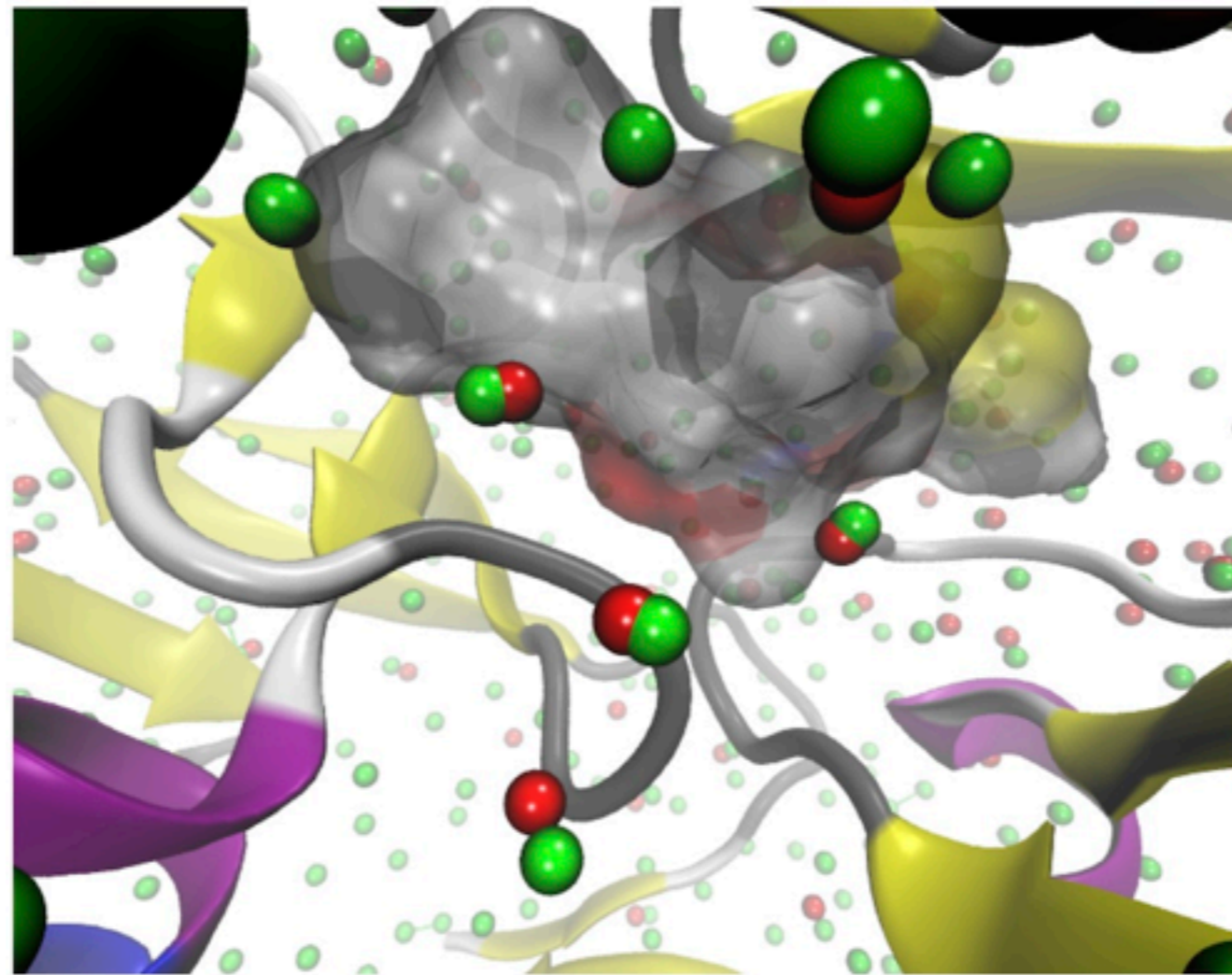
HIVプロテアーゼ



複雑な形状の薬効分子の中心や末端位置を予測

分布から位置を抽出

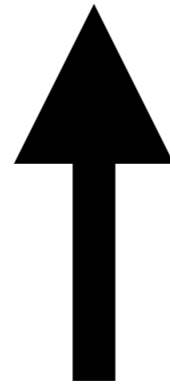
Predicted solvent structure
in HIV-1 Protease



● Experiment ● This work

球でないと計算できないの？

球でないと計算できないの？

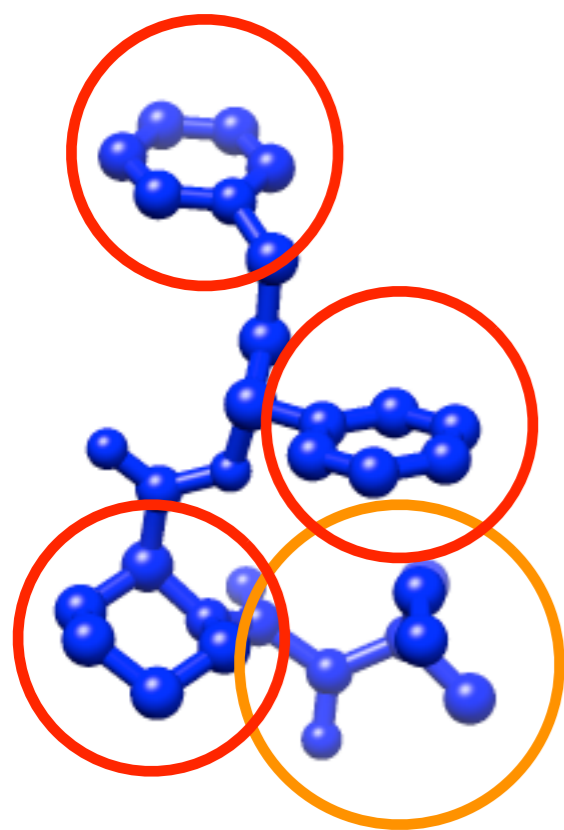


違います！

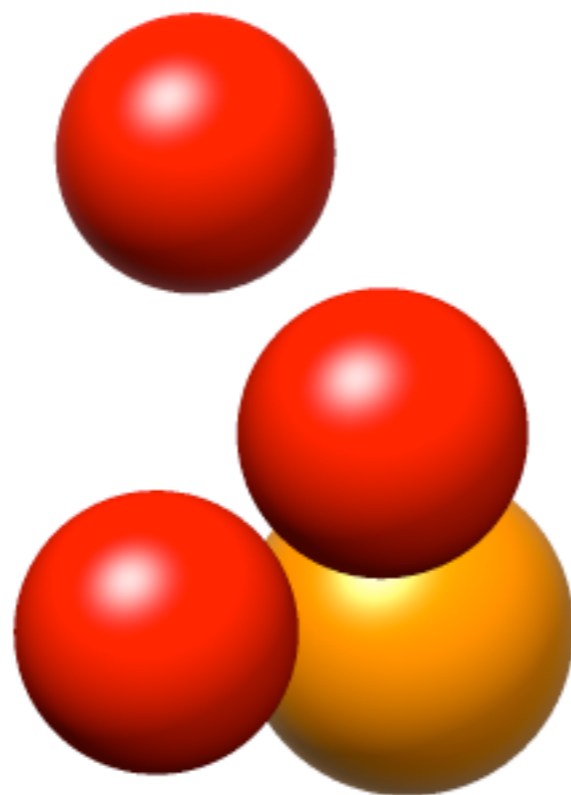
1. 薬効分子が不明

2. 薬効分子が明瞭

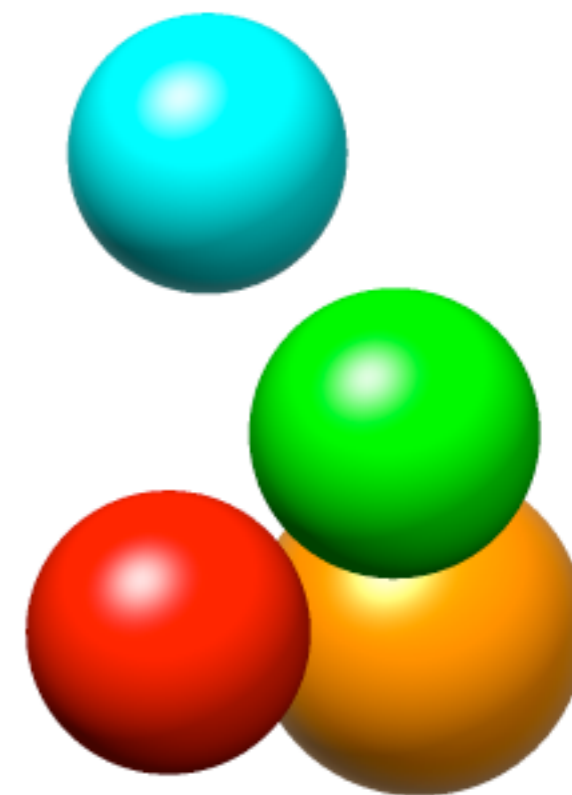
分子の形を維持して粗視化



粗視化位置

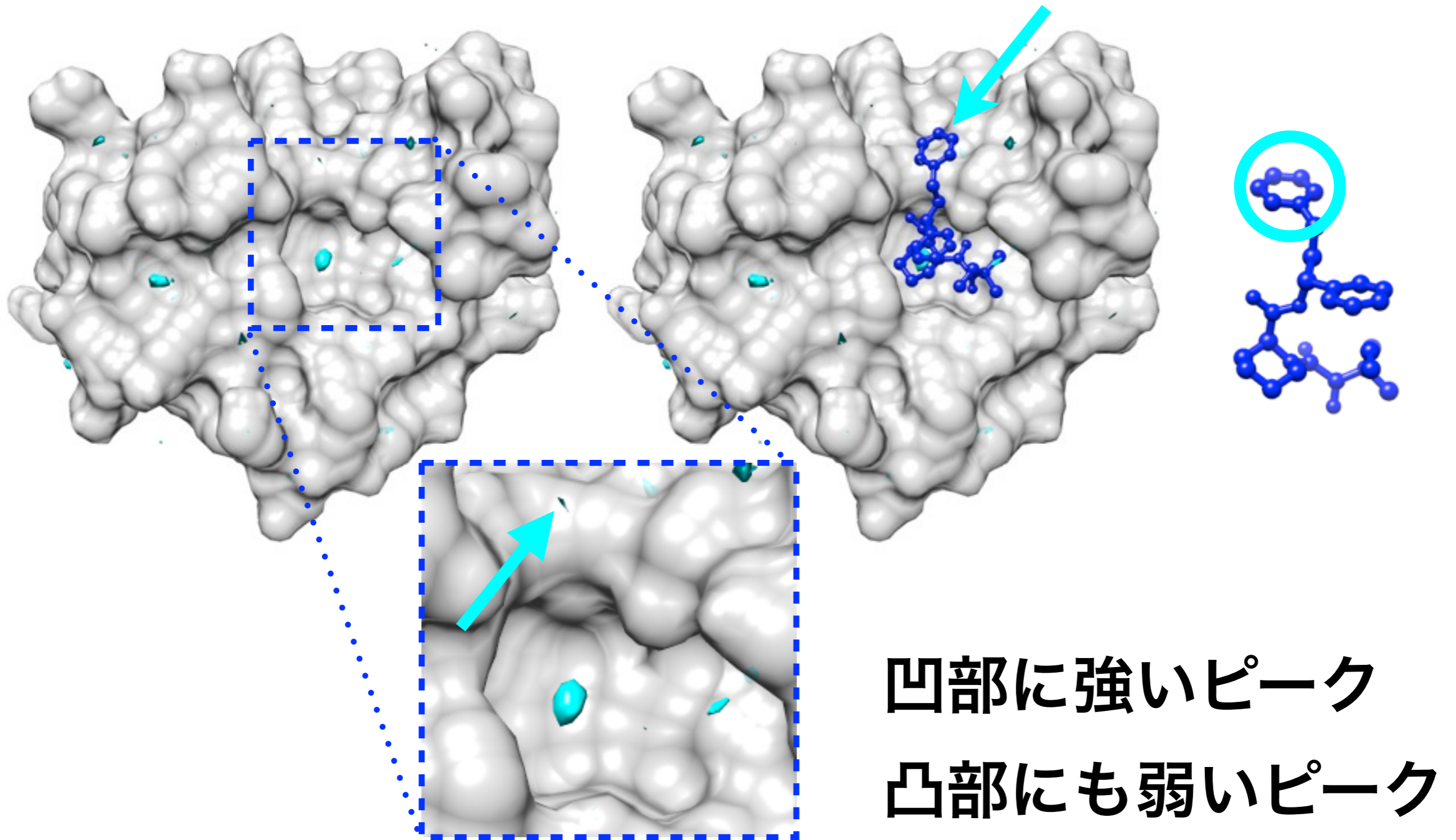


球の大きさ

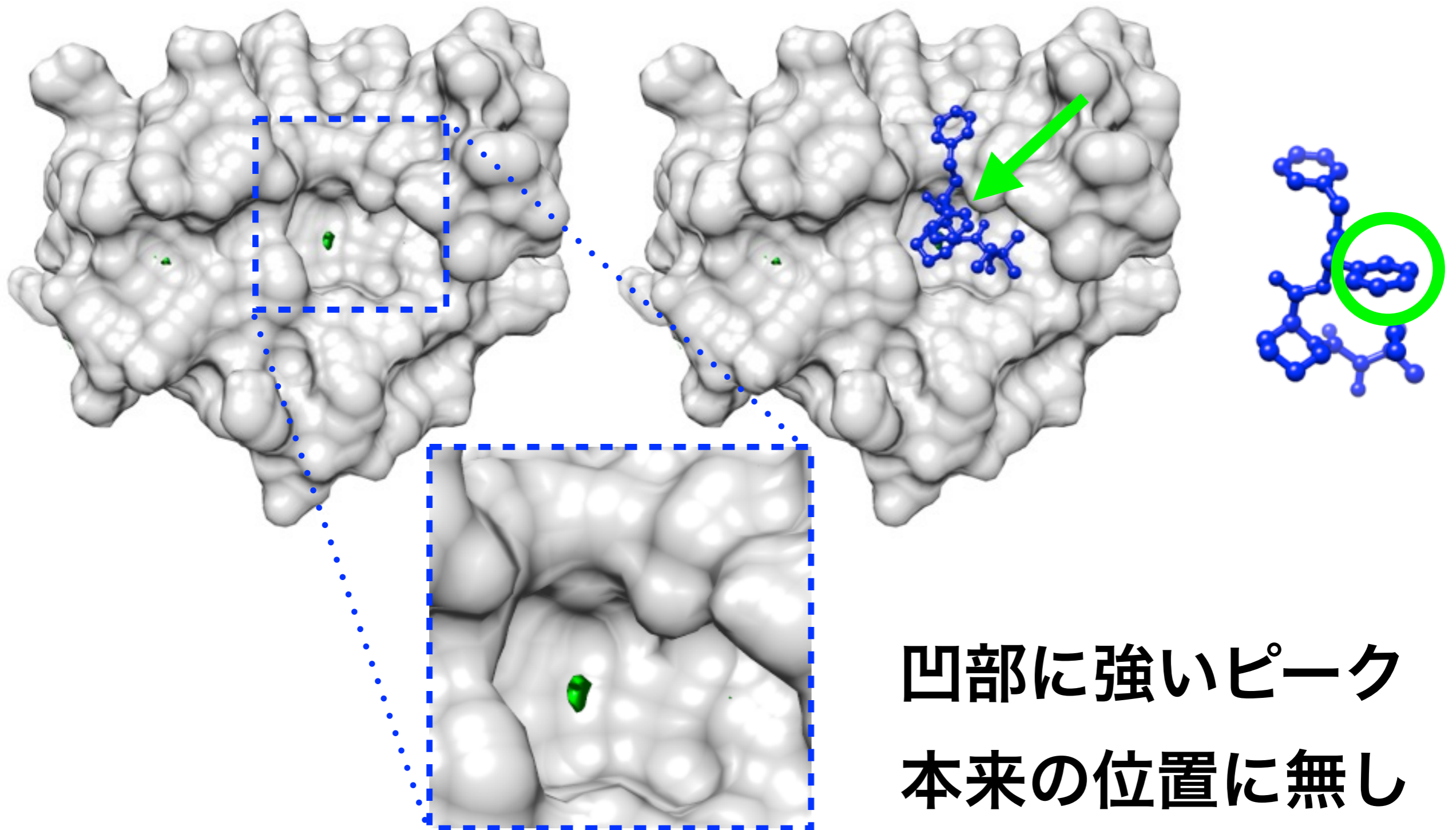


表示の識別

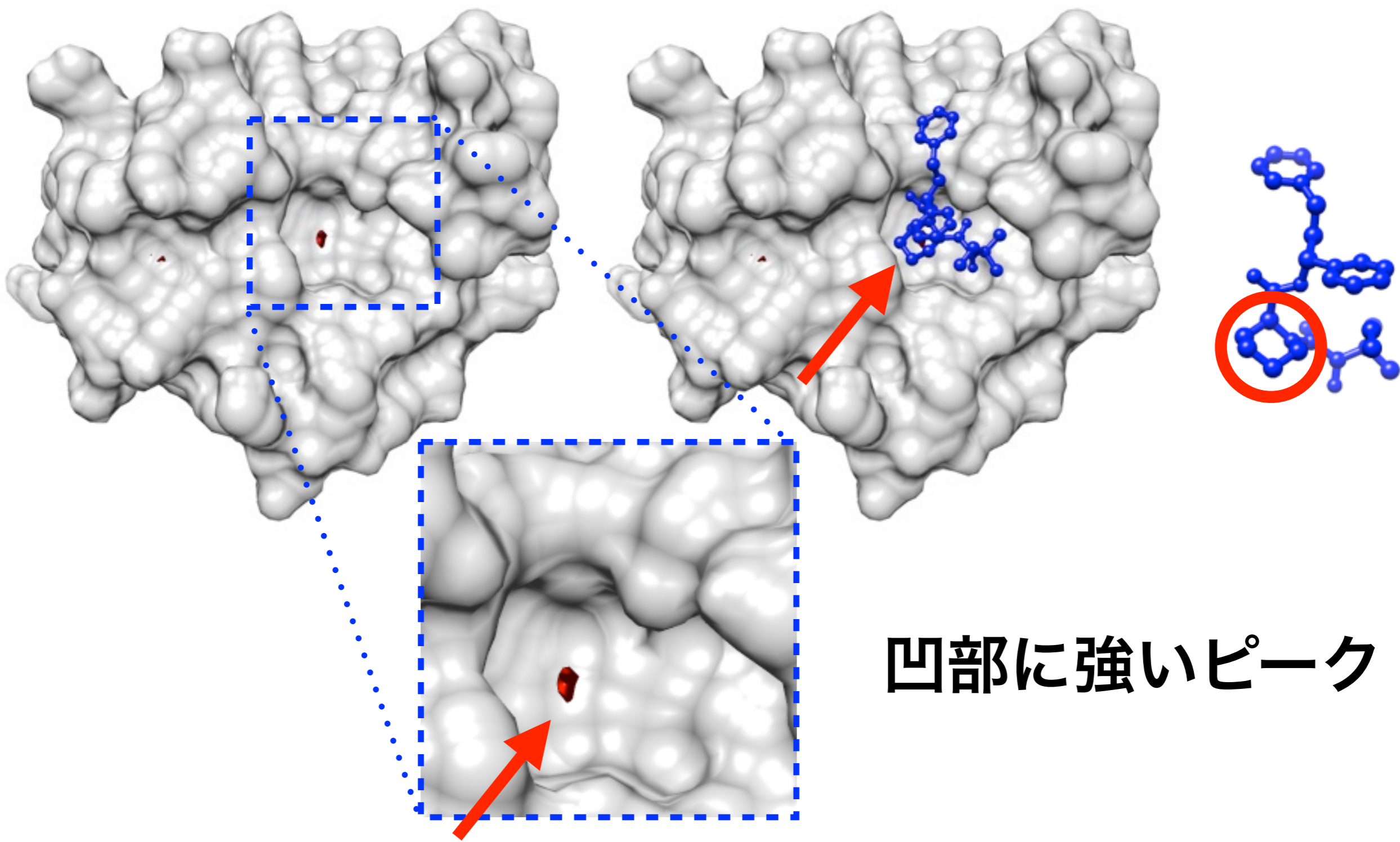
分子の形を取り入れた計算(1)



分子の形を取り入れた計算(2)

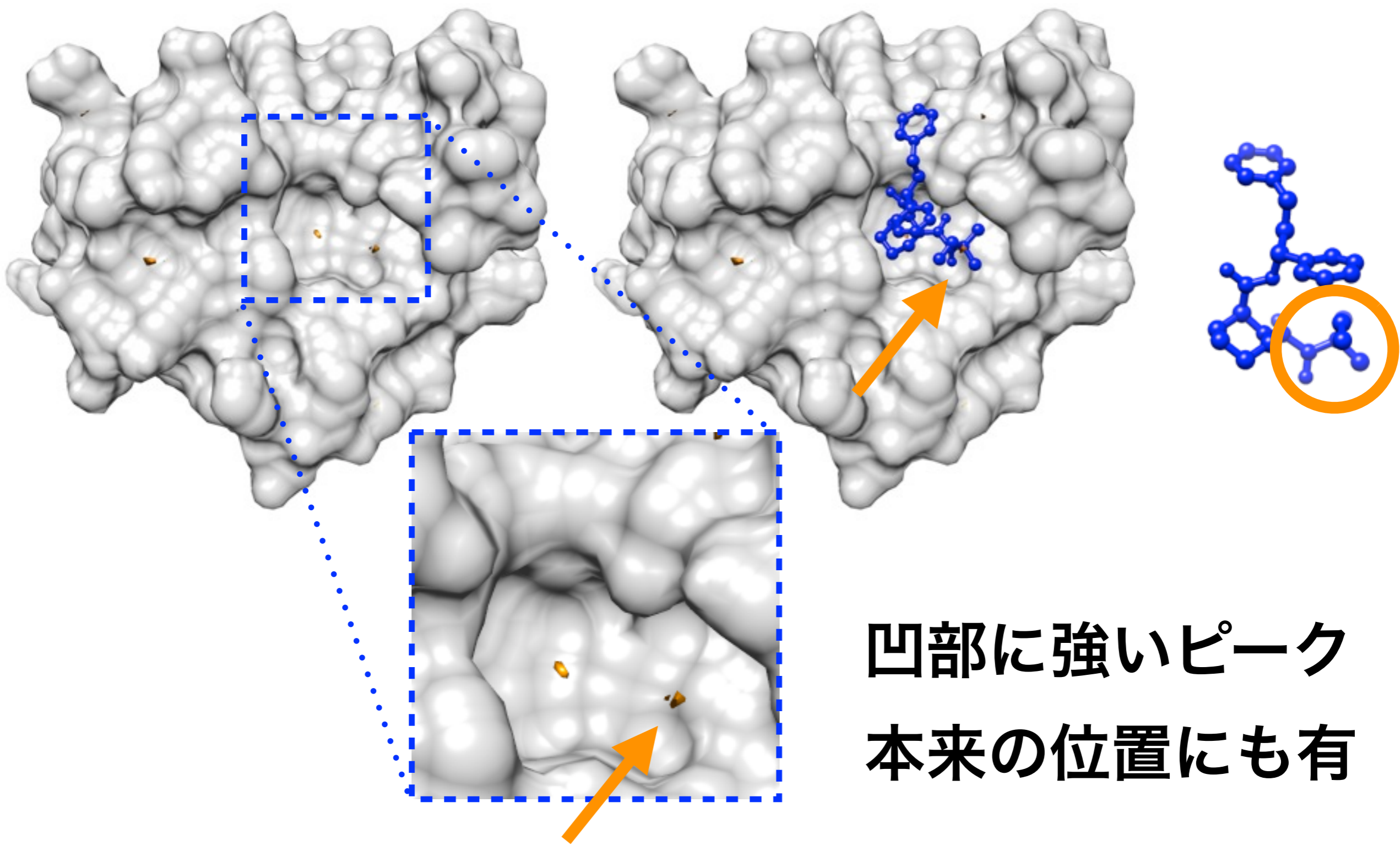


分子の形を取り入れた計算(3)

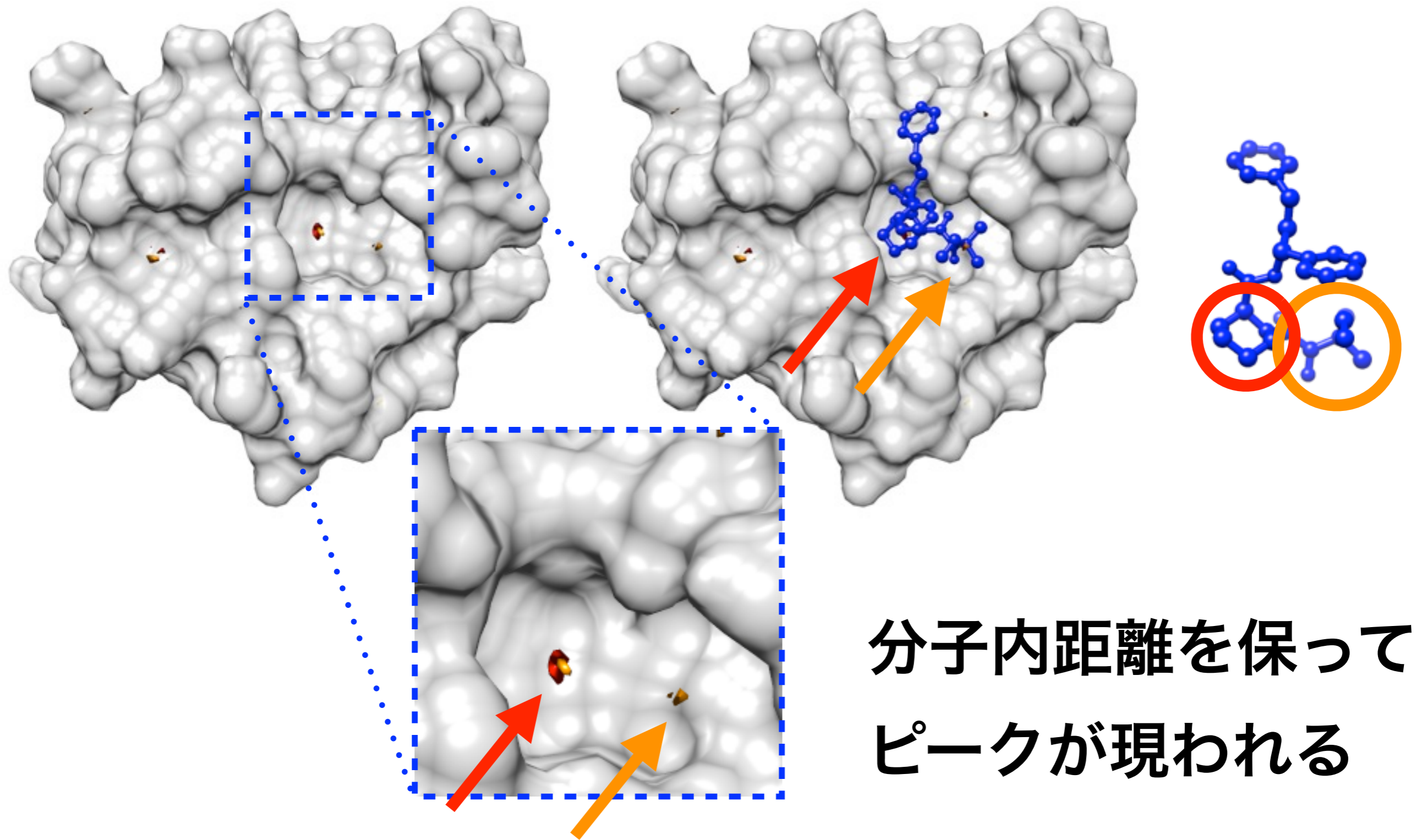


凹部に強いピーク

分子の形を取り入れた計算(4)

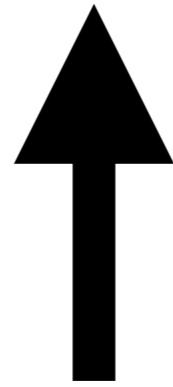


分子の形を取り入れた計算(5)



租視化しないと計算できないの？

租視化しないと計算できないの？

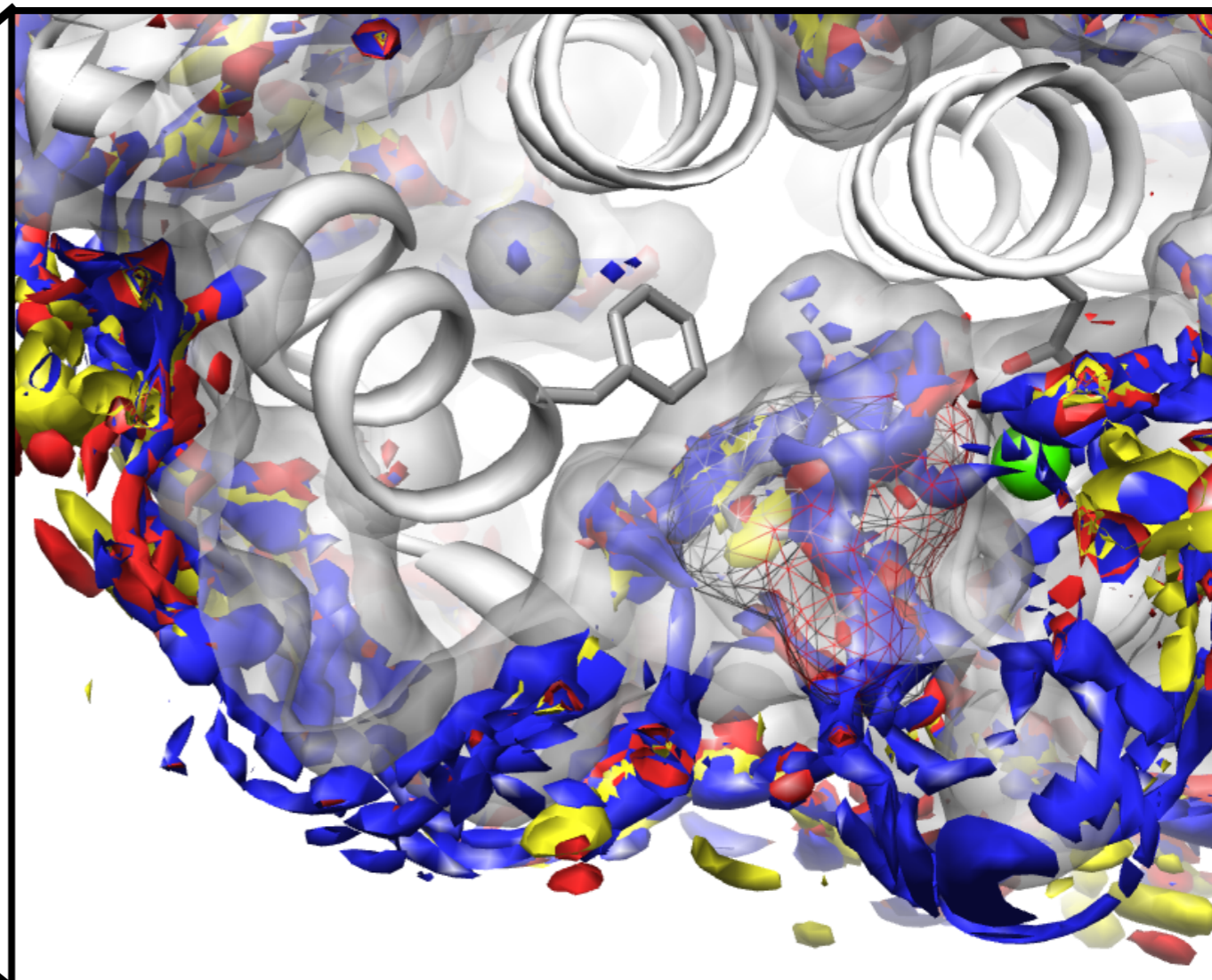
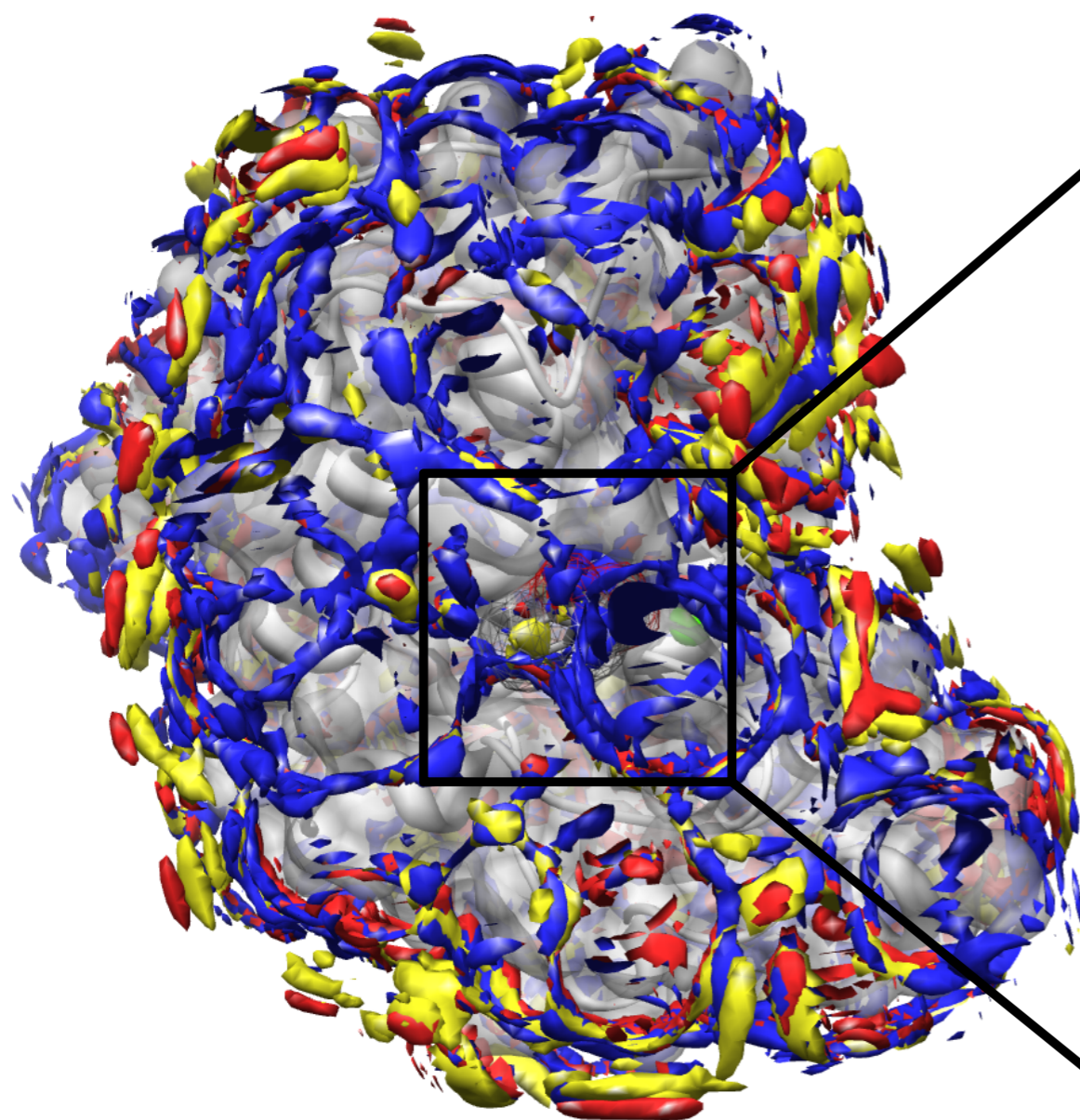
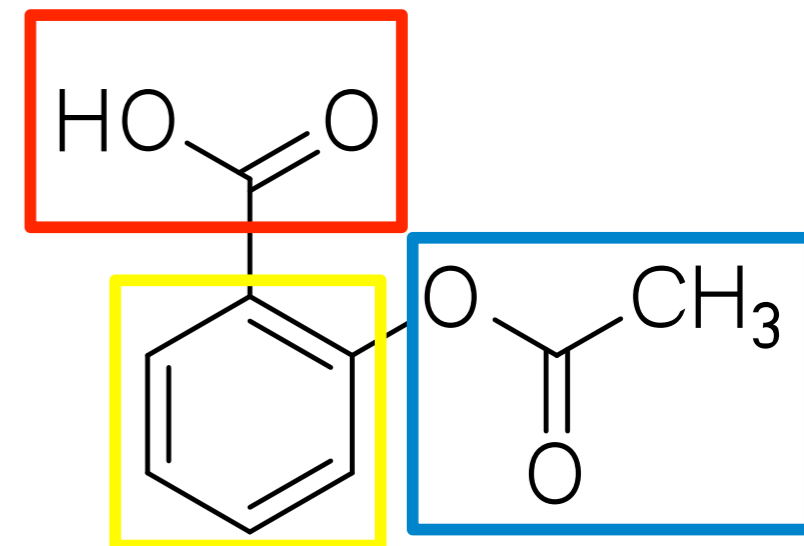


もちろん違います！

アスピリンの分布

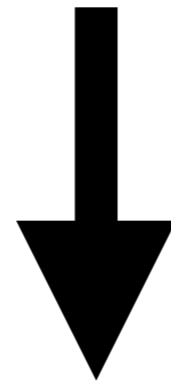
原子表現で計算

(表示はグループごと)



まとめ

- 球形での探索では、最も強いピーク位置とPDBでの位置が一致
- 分子の形状を取り入れた探索では、蛋白質と接触する部位の位置は良く表現



積分方程式理論による薬効分子の結合位置探索は有効

謝辭

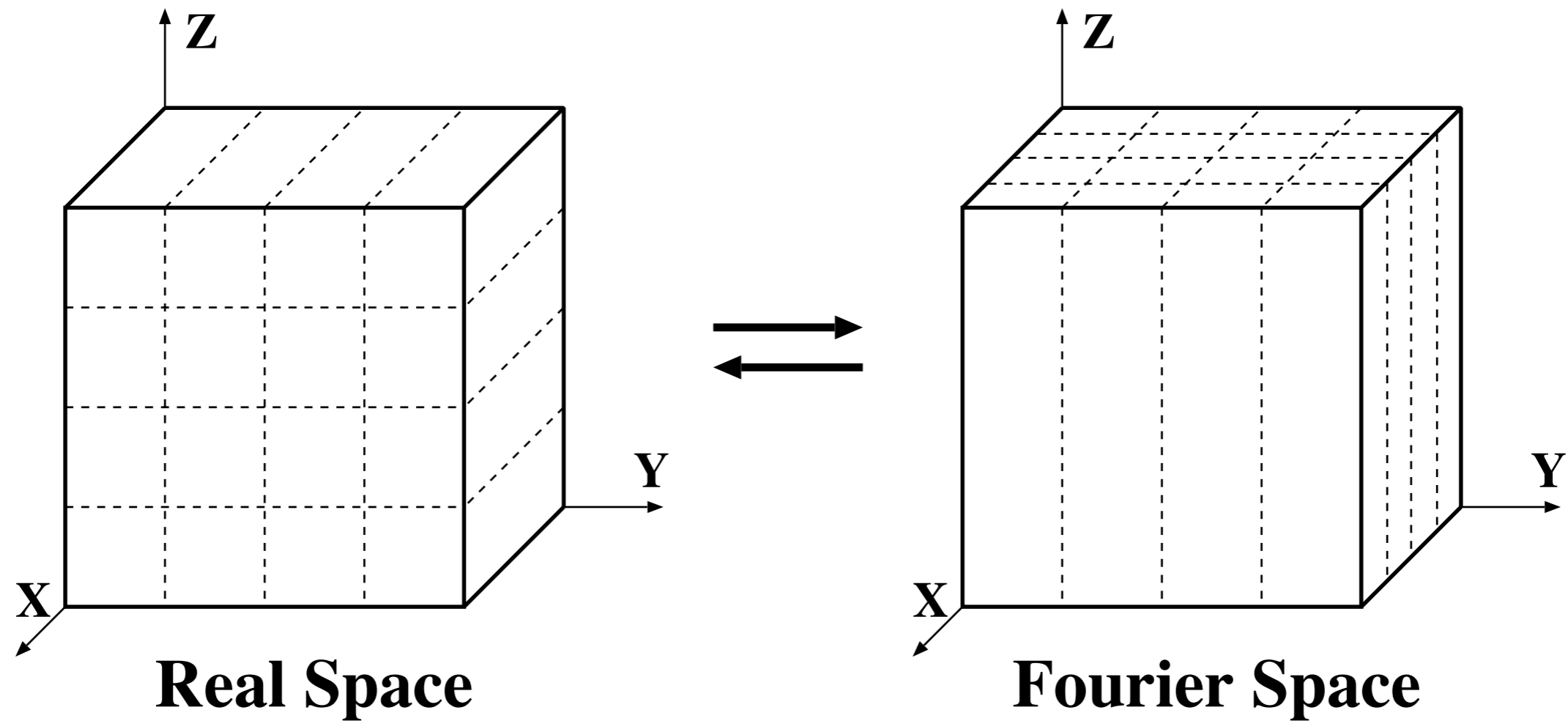
共同研究者

原野 雄一 准教授 (大阪大学)

吉田 紀生 准教授 (九州大学)

額田 彰 博士 (東京工業大学)

3D-RISMプログラムの京対応(1)



実空間でYZ分割、フーリエ空間でXY分割にして3D-FFTの通信量を半減

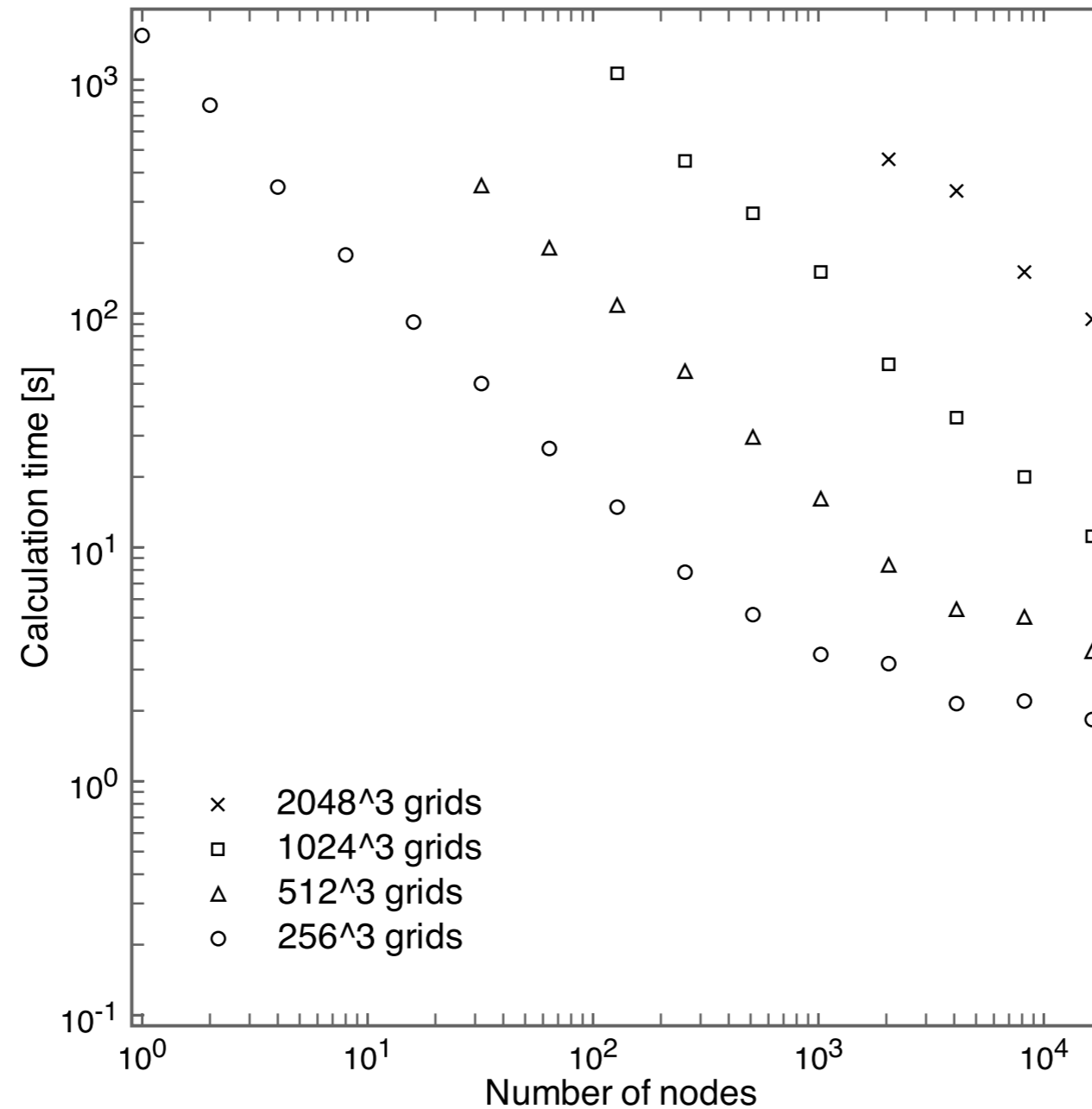
3D-RISMプログラムの京対応(2)

YZ分割の3D-FFTの手順

1. X軸のFFTを実行
2. Y軸のデータを揃える通信
3. Y軸のFFTを実行
4. Z軸のデータを揃える通信
5. Z軸のFFTを実行
6. Z軸のデータを戻す通信
7. Y軸のデータを戻す通信

← 実行しない

3D-RISMプログラムの京対応(3)



計算セルサイズ毎の計算時間のノード数依存性