3D-RISMを中心とした生体機能解析：理論と京への実装そして応用～京スパコンで可能になること～

吉田紀生
九州大学大学院理学研究院化学部門

平田文男（分子研・立命館大），丸山豊（大阪大），
Saree Phongphanphanee(分子研), Daniel J. Sindhikara(立命館大),
清田泰臣（北里大）

SS研HPCフォーラム2012

次世代ナノ統合シミュレーション ソフトウェアの研究開発

・目的

開発中の次世代スーパーコンピュータを分子・物質の計算科学のシミュレーションで最大限に活用できるよう、計算科学理論・方法論の確立と、それらに基づくソフトウェアの開発を行う。電子・原子・分子レベルからの精緻な大規模計算に基づいて、ナノスケールの物質で発現する特有の現象・特性を解明し、予測することを可能とする。具体的に以下の3つのグランドチャレンジ課題を設置し、それぞれの分野の解明にブレイクスルーをもたらす計算科学の飛躍知の発見・発明に繋げる。研究開発拠点に産学連携体制を構築し、研究成果の産業界への展開を図る。

（ナノ統合拠点ホームページより）

中核アプリケーション

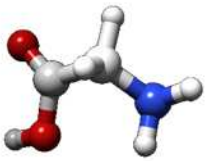
- 6つのアプリケーション
- RISM/3D-RISM (代表：分子科学研究所　平田文男*)

  液体の統計力学理論に基づき蛋白質などナノ分子の水和構造や水和自由エネルギーを計算するプログラム。分布関数（相関関数）に関する積分方程式を解くために、フーリエ変換を多用する。そのため、高速フーリエ変換（FFT）の速度が重要である。

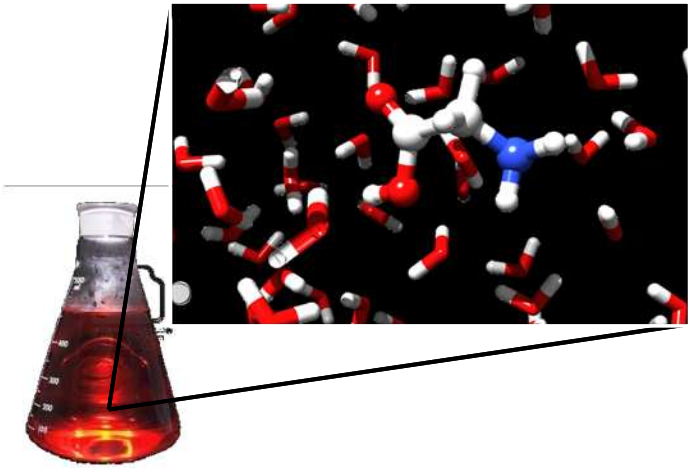
（ナノ統合拠点ホームページより）

*現・立命館大学

実在系の分子理論



理想系
(単一の分子･孤立系)
量子化学, Schrödinger方程式



実在系
(アボガドロ定数個の分子)

10個, 20個の積み上げではなく,
パラダイムシフトが必要

統計力学, 液体論

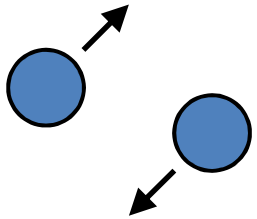
液体の統計力学理論：3D-RISM理論

液体の構造を記述する
統計力学理論

液体の積分方程式理論：
3D-RISM理論

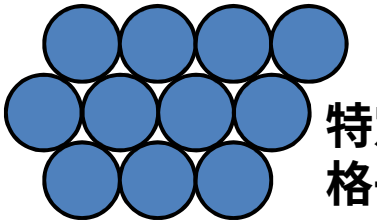
液体の構造

気体の構造



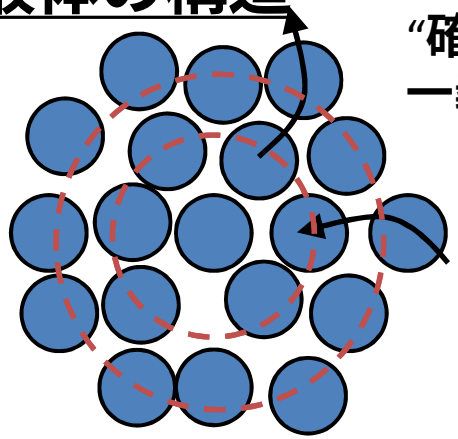
特定の構造を
持たない

固体の構造



特定の構造を持ち、
格子定数で決まる

液体の構造



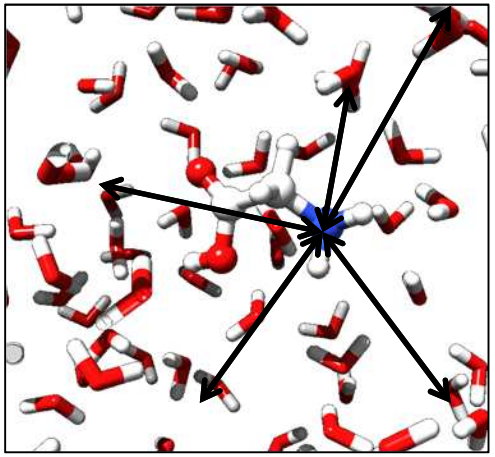
“確率論的構造”を持つが、
一義的には決まらない

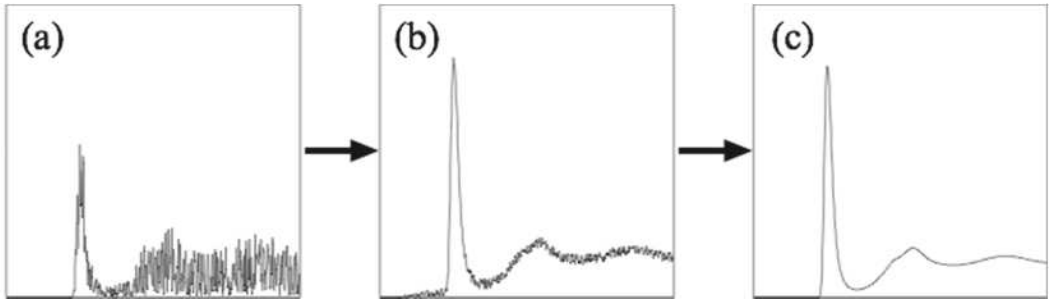
ある粒子からの
“平均的な存在確率”
で記述する

分布関数

液体の構造(分布関数)を求める

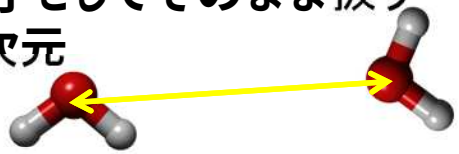
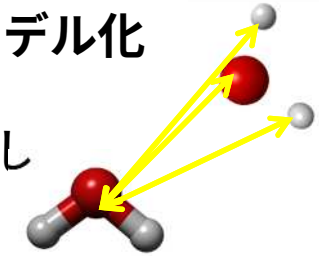
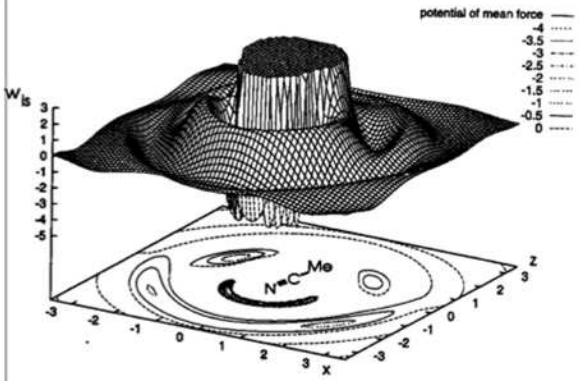
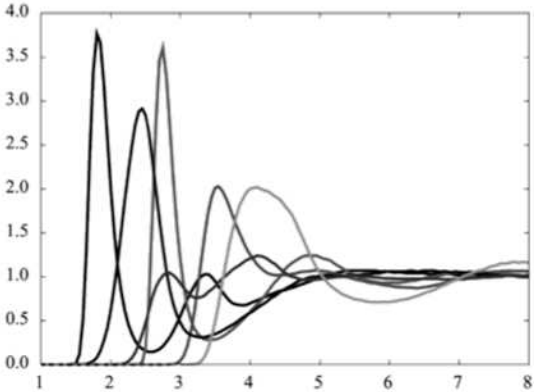
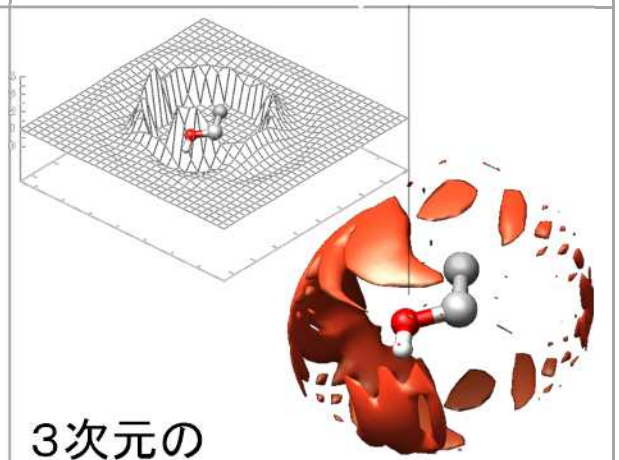
・実験的にはX線や中性子で求められる
・分子シミュレーション
　サンプリングにより求める
(MD: ニュートン方程式, MC:乱数)





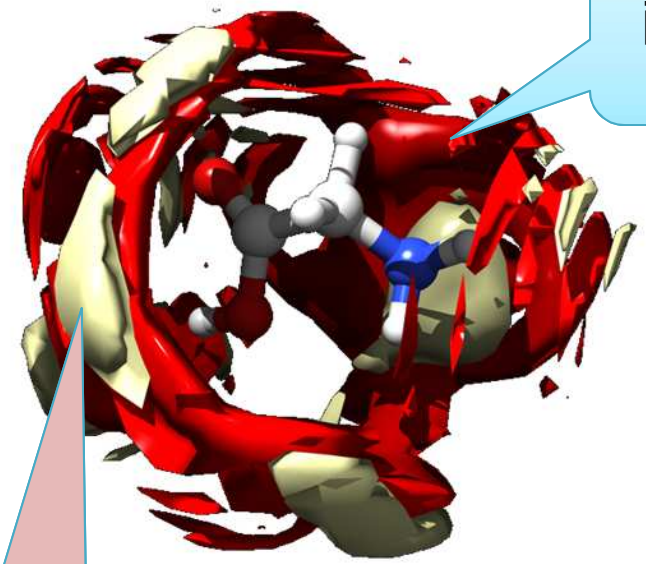
液体の積分方程式理論
“解析的”に“直接”得られる

液体の積分方程式理論

MOZ	RISM	3D-RISM
モデル化なし 分子としてそのまま扱う 6次元 	相互作用点モデル 1次元 	溶媒のみモデル化 溶質は モデル化なし 3次元 
 全自由度を扱える 角度に依存した分布など	 1次元（サイト間距離）の分布関数	 3次元の 空間分布関数
厳密な方程式で理論的精度が高い 計算コストが高く， 大きい分子は難しい	簡便で高速 適用範囲が広い 近似が大きく得られる情報は限定的	生体高分子の溶媒和を計算可能

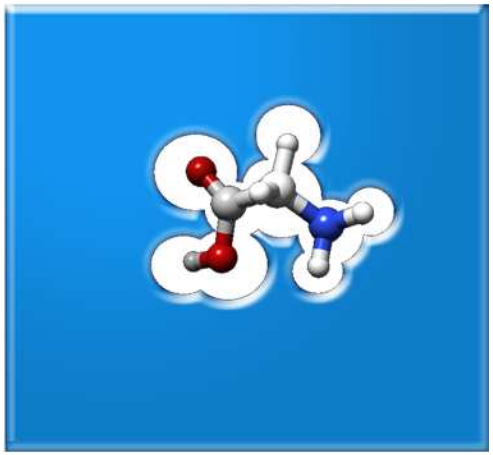
3D-RISM理論の魅力

溶媒を“分子”として扱う

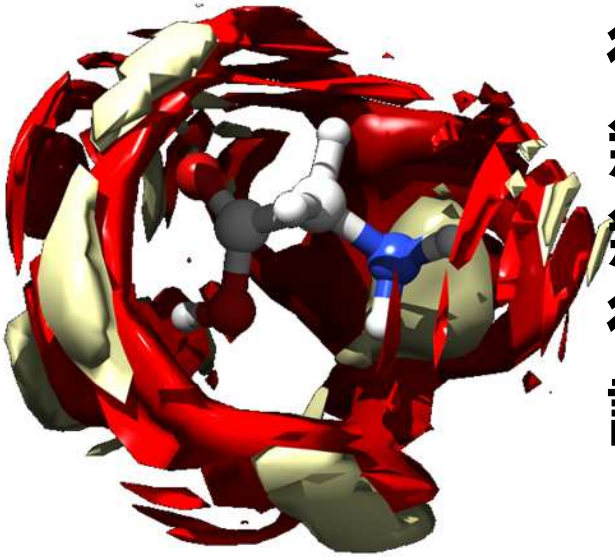


水素結合などの溶媒の“分子としての特徴”を記述できる。

連続体モデルでは溶媒の分子的特徴は記述できない。

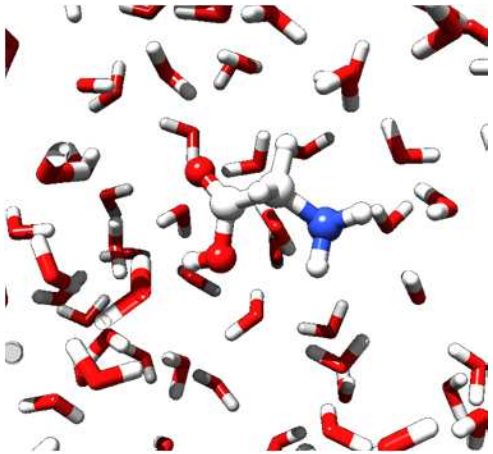


熱力学極限を満たす



得られた分布は
無限個の溶媒分子，
無限時間の統計平均により
得られるものに対応。
計算の初期条件に依存しない

分子シミュレーションでは
溶媒の個数，サンプリングは有限。



熱力学量を(簡単に)計算できる

溶媒和自由エネルギー:

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

内部エネルギー： $U = \frac{3}{2}Nk_BT + \frac{N}{2}\rho\int_0^\infty u(r)g(r)\cdot 4\pi r^2 dr$

圧力： $P = \rho k_BT - \frac{1}{6}\rho^2\int_0^\infty r\frac{du(r)}{dr}g(r)\cdot 4\pi r^2 dr$

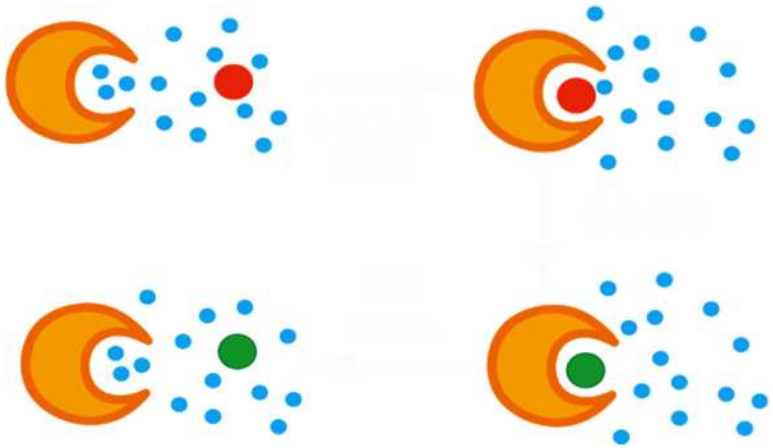
圧縮率： $\chi_T = \frac{1}{\rho k_BT} + \frac{1}{k_BT}\int_0^\infty \left(g(r)-1\right)\cdot 4\pi r^2 dr$

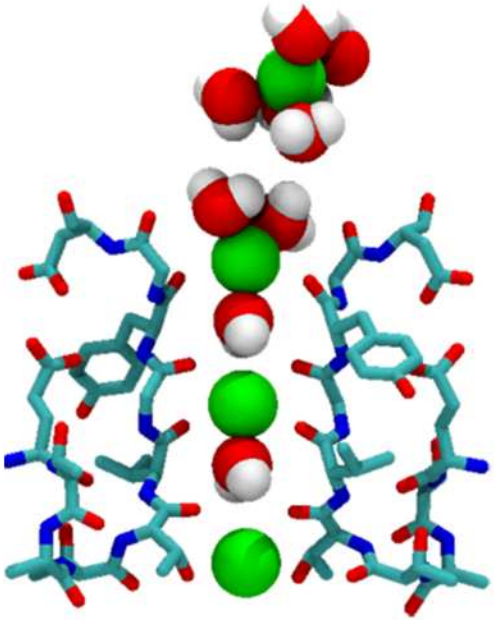
溶媒和エネルギー： $E = \rho\int_0^\infty u_{uv}(r)g_{uv}(r)\cdot 4\pi r^2 dr$

部分モル体積： $\bar{V} = \chi_T k_BT - \int_0^\infty \left(g_{uv}(r)-1\right)\cdot 4\pi r^2 dr$

生体分子の分子認識への展開

分子認識
～最も重要な生体内素過程





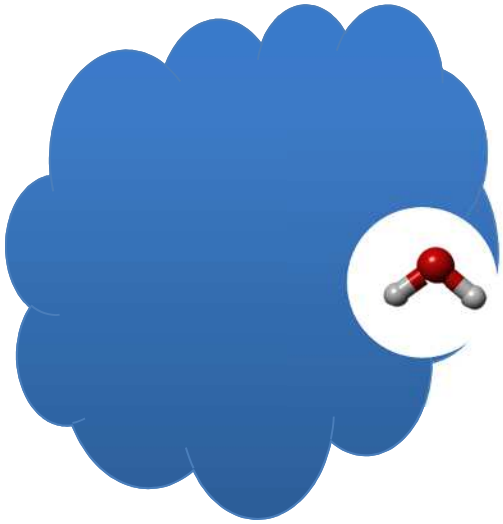
- 酵素反応の初期過程
- チャネルタンパクの選択性
- 抗原･抗体反応　　など

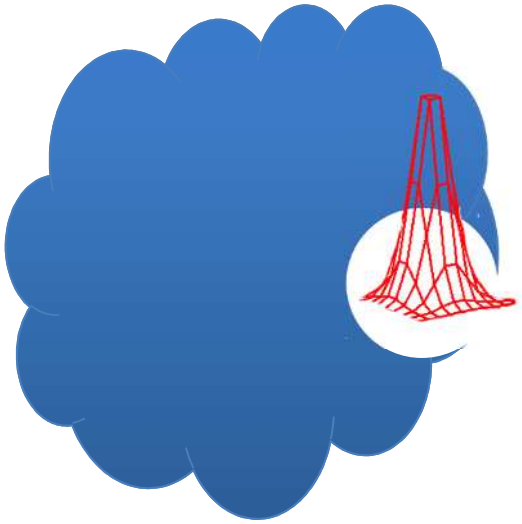
系全体におよぶエントロピー変化を含む自由エネルギー変化を駆動力とする

“統計的”,“確率論的”過程

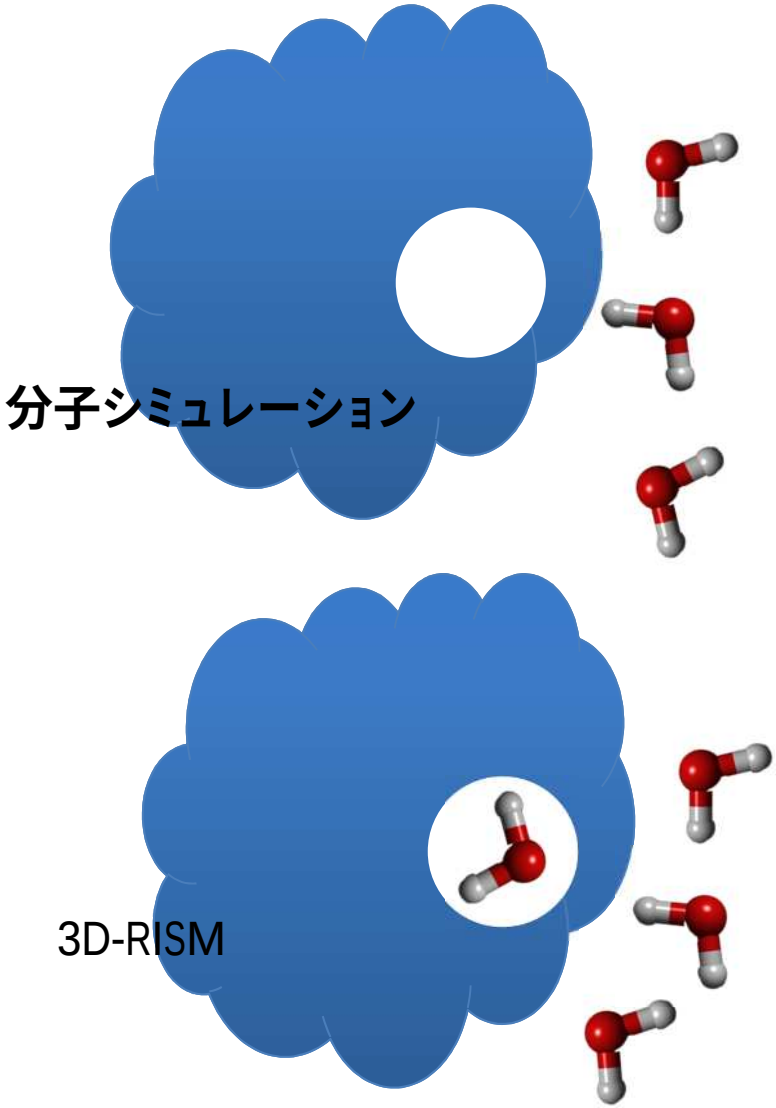
液体の理論で分子認識？？

分子認識 → 高い親和性 すなわち 高い分布確率 → 溶媒・リガンドの分布



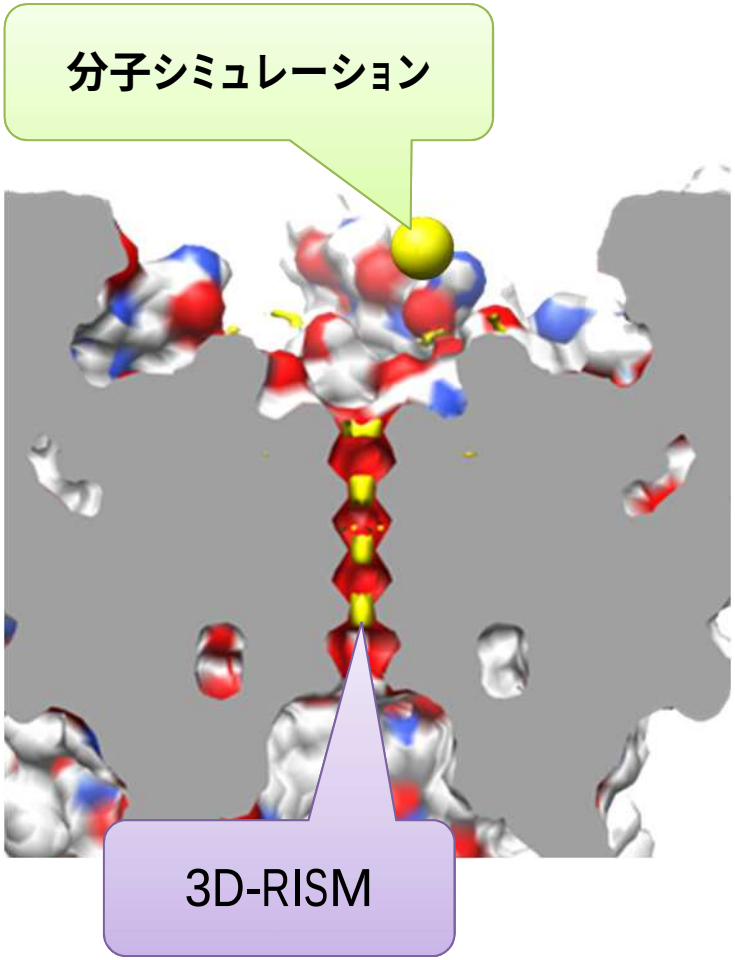


従来の分子認識解析の難点(1)
内部に閉じ込められた分子



- 分子シミュレーションでは分布は初期条件に依存
  - あとからは，入れない
  - 事前に内部の分子分布の予備情報が必要(水やイオンの有無，個数)
- 3D-RISM理論は初期条件に依存しない
  - 完全な統計平均量としての分布が得られる
  - タンパク質の構造以外の予備情報は不要

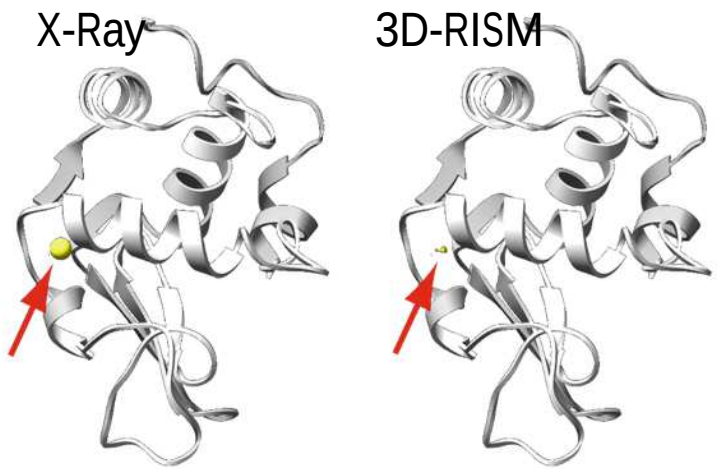
従来の分子認識解析の難点(2)
“レアイベント”の解析



- 分子シミュレーションの時間スケールでは十分な統計量が得られない
  - “良く起こること”は容易だが，“起きにくいこと”の解析は困難
  - 反応座標を設定しサンプルを限定
    J. S. Hub, B. L. de Groot, *Proc. Natl. Acad. Sci. U.S.A. 2008, 105, 1198.*
  - 人工的に圧力や電場を加えるなどして，加速して解析する
    M. Ø. Jensen, E. Tajkhorshid, K. Schulten, *Biophys. J. 2003, 85, 2884.*
- 3D-RISMでは，完全な統計平均量としての分布が得られる
  - 外場や反応座標の設定は不要

例：リゾチームの選択的イオン認識

N. Yoshida, et al., *J. Am. Chem. Soc. Communication*, 2006, 128, 12042



リゾチームによる
Ca^{2+}認識部位を再現

Kuroki, R.; Yutani, K.
J. Biol. Chem. 1998, 273, 34310

野生型，ミュータントに対して，
Na^{+}, Ca^{2+}, Li^{+}の親和性を求め，
実験を再現した

分子シミュレーションでは，
結合部位の予備情報無しでは
不可能

タンパク質の構造さえあれば，
イオンの選択性，認識部位を
“非経験的･分子論的に決定できる”
ことを示した

3D-RISM理論に基づく大きなリガンド分子の分子認識解析手法の開発と応用

清田泰臣, 吉田紀生, 平田文男, JCTC, 2011

清田泰臣　助教 (北里大学)

RISM理論による分子認識解析

- RISM理論

  ⇒溶媒分子の分布関数（＝確率密度）が得られる

溶媒としてリガンド分子を含めれば・・・

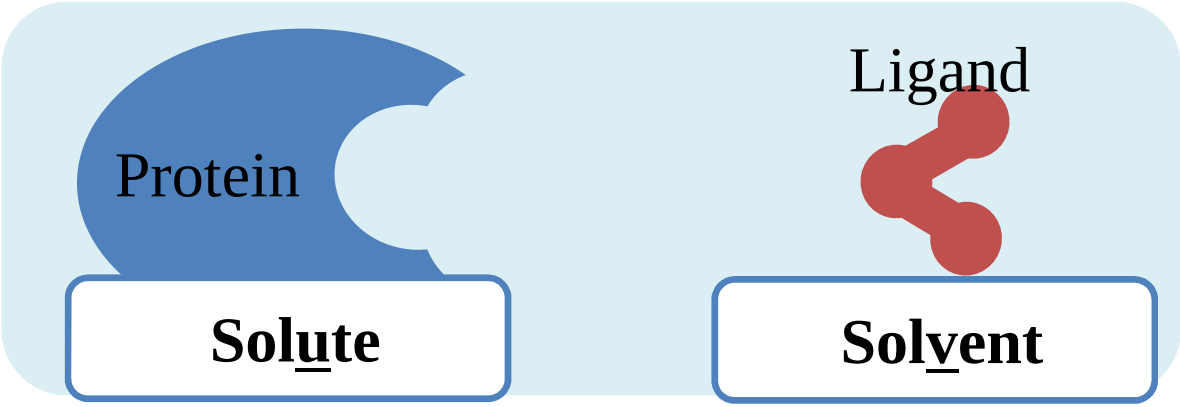
RISM理論によりリガンド分子の分布が求まる

＝分子認識問題について応用できる

- **従来法の問題**
  - リガンドを溶媒として扱う
  - 溶媒（リガンドを含む）のバルク構造に関する計算が難しい

•従来

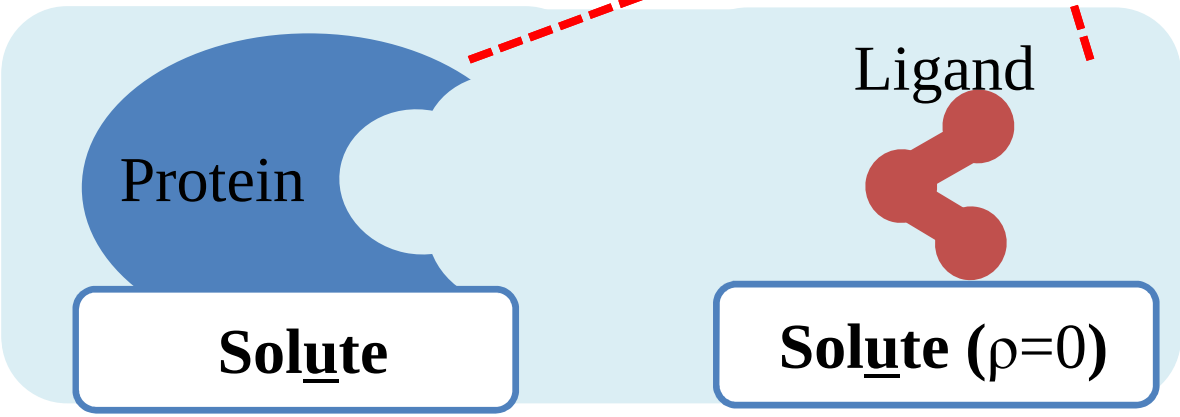
Solute-Solvent



•本研究

Solute-Solute

$$h^{uu}_{\alpha\beta}(1,2) = c^{uu}_{\alpha\beta}(1,2) + \sum_{c}^{N_{solvent}} \rho_c \int c^{uv}_{\alpha c}(1,3) h^{vu}_{c\beta}(3,2) d(3)$$



•Procedure to obtain uu 3D-DFs.

- Calculate correlation function of solvent(water, ion, ...) by vv 1D-RISM.

$$h_{ab}^{vv''}(r) = \sum_{c,d} \varpi_{ac}^{v} * c_{cd}^{vv''}(r) * \varpi_{db}^{v''} + \sum_{v'} \rho^{v'} \sum_{c,d} \varpi_{ac}^{v} * c_{cd}^{vv'}(r_{cd}) * h_{db}^{v'v''}(r_{db})$$

- Calculate correlation funciton of protein by uv 3D-RISM.

$$h_a(\mathbf{r}) = \sum_{v'} \sum_{b} \int c_b(\mathbf{r}') \left(\omega_{ba}^{v'v}(|\mathbf{r}' - \mathbf{r}|) + \rho_{v'} h_{ba}^{v'v}(|\mathbf{r}' - \mathbf{r}|) \right) d\mathbf{r}'$$

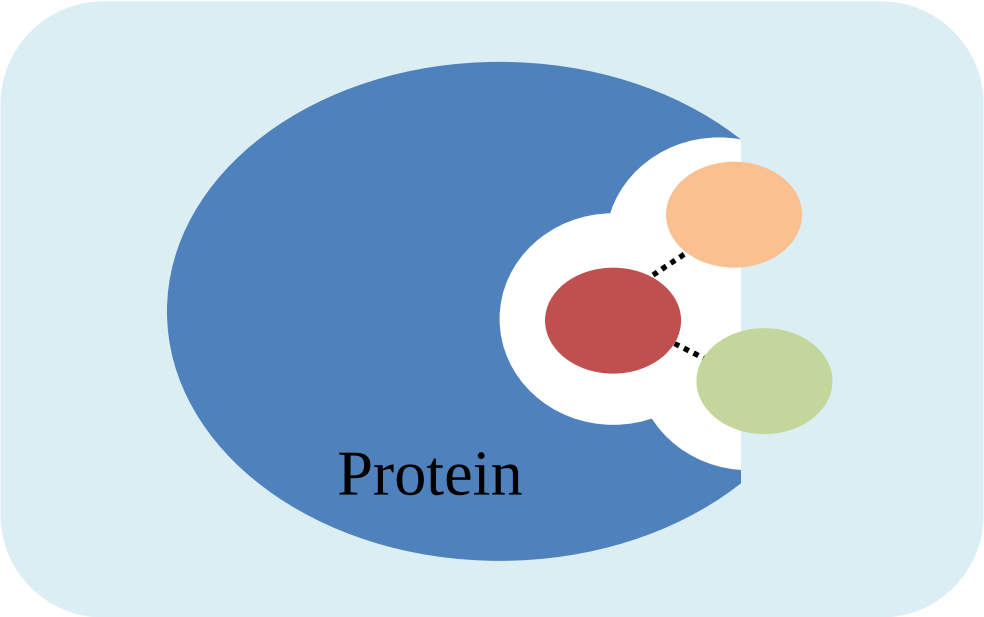
- Calculate correlation function of ligand by uv 1D-RISM.

$$h_{\alpha b}^{uv}(r) = \sum_{\beta,c} \varpi_{\alpha\beta}^{u} * c_{\beta c}^{uv}(r) * \varpi_{cb}^{v} + \sum_{v'} \rho^{v'} \sum_{\beta,c} \varpi_{\alpha\beta}^{u} * c_{\beta c}^{uv}(r) * h_{cb}^{v'v}(r)$$

- Calculate correlation function between protein-ligand by uu 3D-RISM.

$$h_{\alpha}(\mathbf{r}) = \sum_{\beta} \int c_{\beta}(\mathbf{r}') \omega_{\beta\alpha}^{uu}(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r}' + \sum_{v'} \sum_{a} \int c_a(\mathbf{r}') \rho_{v'} h_{a\alpha}^{v'u}(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r}'$$

分布関数からの認識部位決定



Information of ligand:
Distribution function

Where?

Location

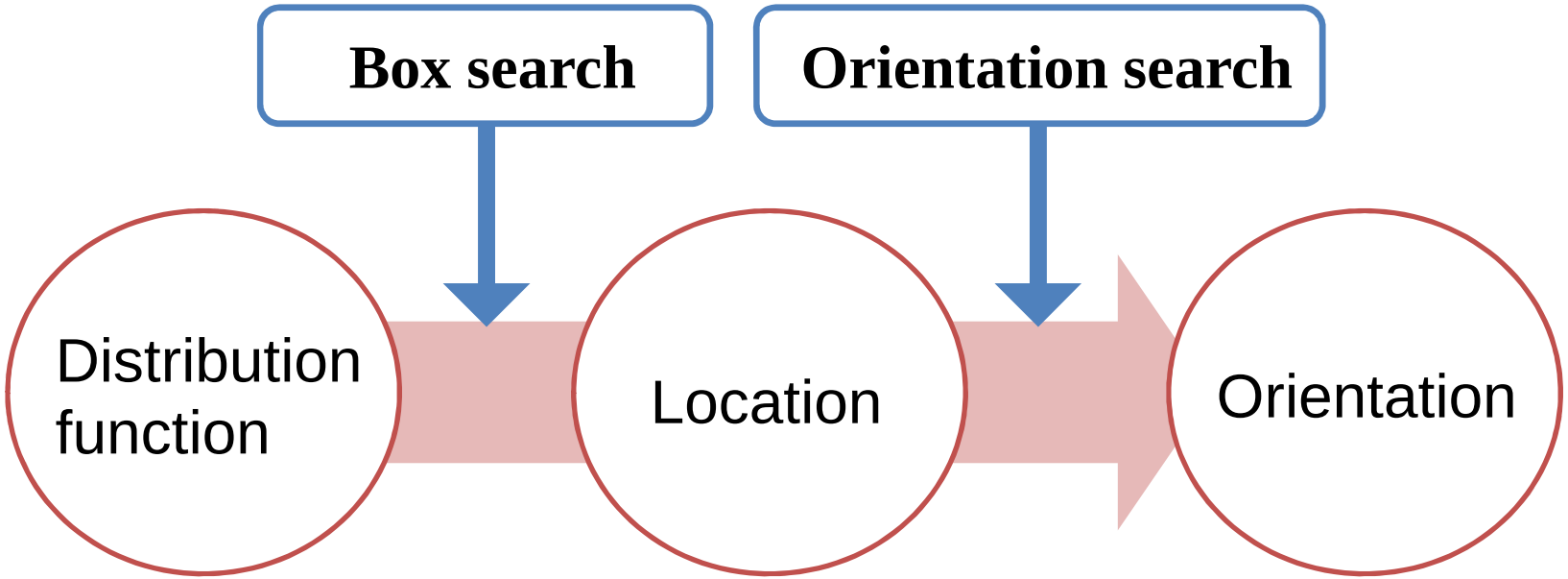
How?

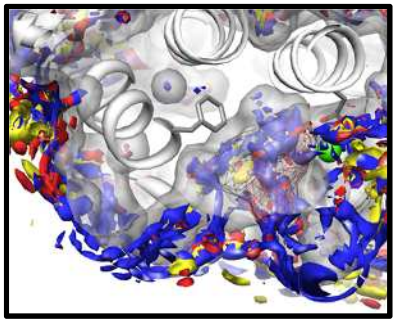
Orientation

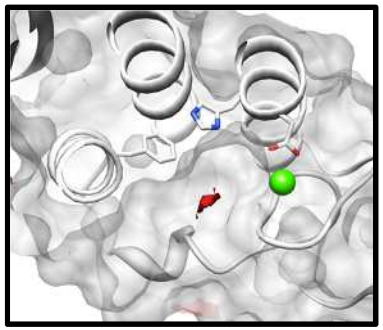
Two different types of score function

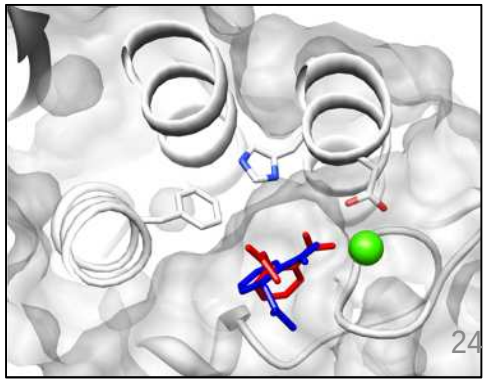
Approach

For determining the location and the orientation





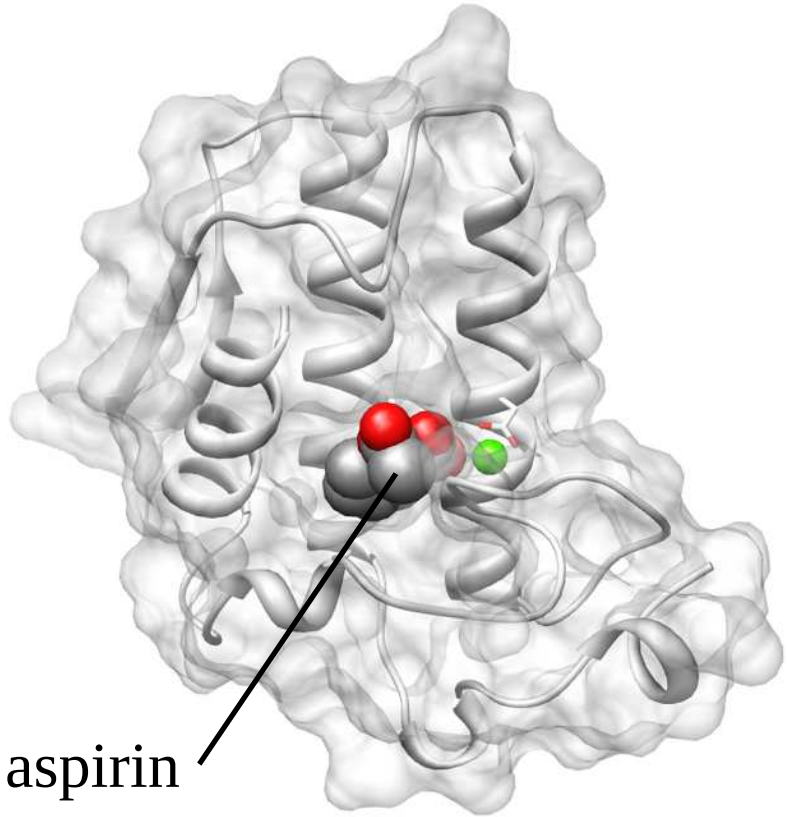




System Phospholipase A2 - Aspirin

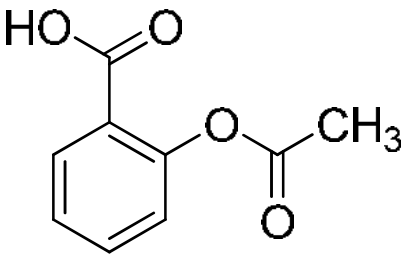
Phospholipase A2 and Aspirin form the complex for inducing Anti-inflammatory effects through the specific binding site. [1]

X-ray structure (PDB: 1OXR)

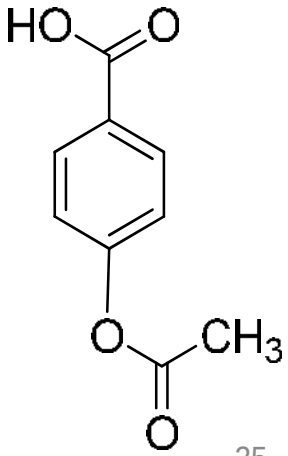


Protein and ions: Amber99
Water: TIP3P
ligand: Antechamber (GAFF)

(a) 2-acetoxy-benzoic acid (aspirin)



(b)4-acetoxy-benzoic acid (p-aspirin)



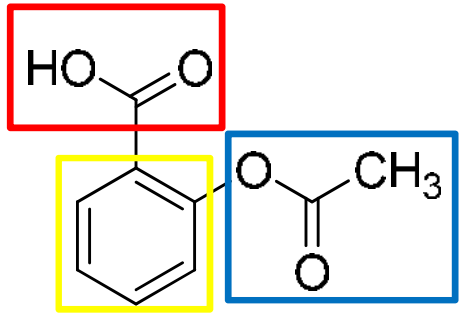
[1] R. K. Singh, A.S. Ethayathulla, T. Jabeen, S. Sharma, P. Kaur, and T. P. Singh., JDT. *13*, **2005**, 133.

• 3D-DFs of Aspirin

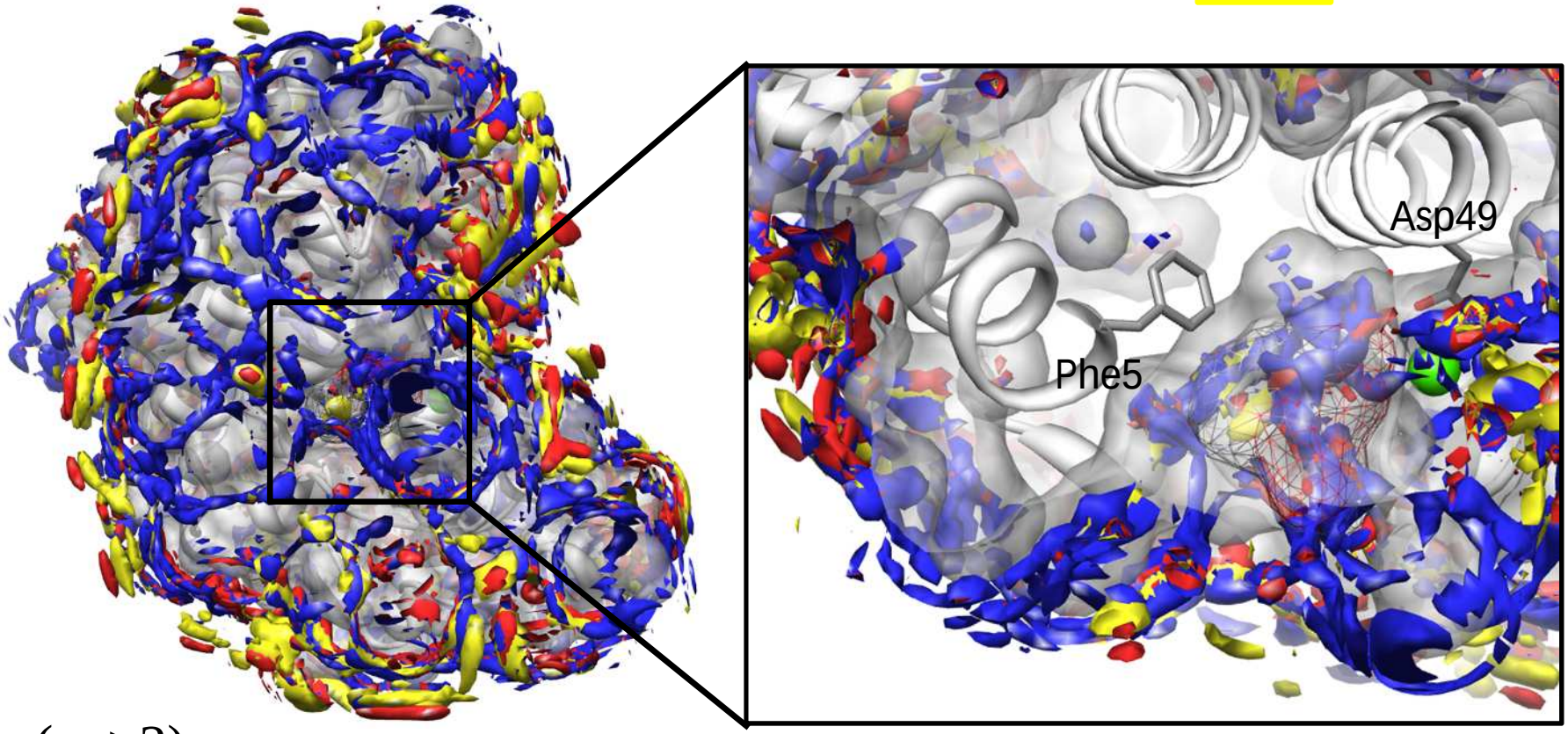
Red: COOH sites

Yellow: Aromatic sites

Blue: $OCOCH_3$ sites



aspirin



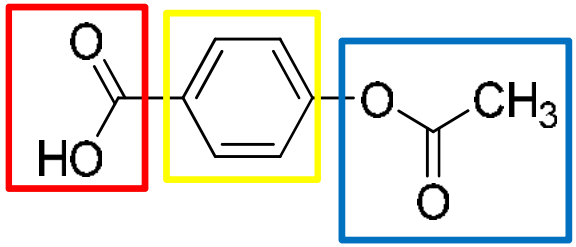
(g_γ >2)

• 3D-DFs of p-Aspirin

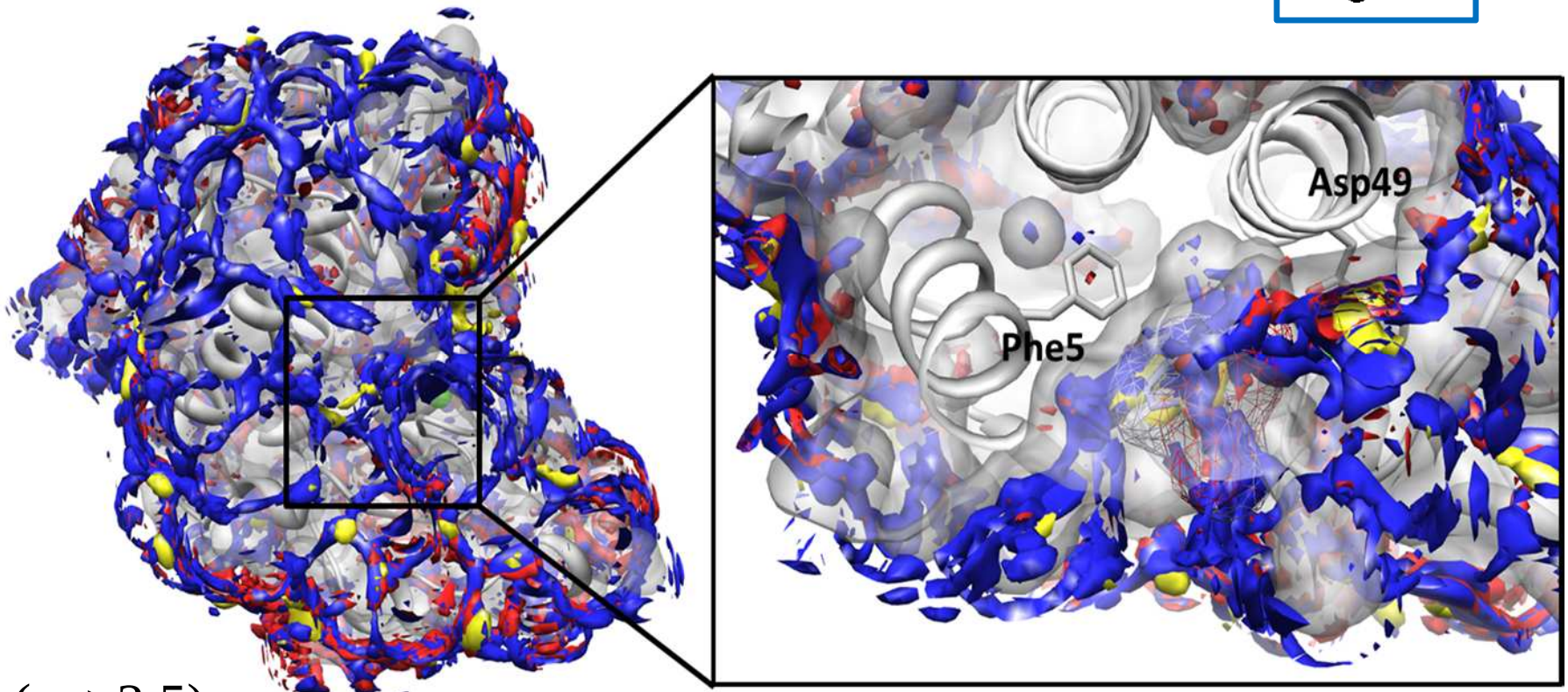
Red: COOH sites

Blue: $OCOCH_3$ sites

Yellow: Aromatic sites



p-aspirin

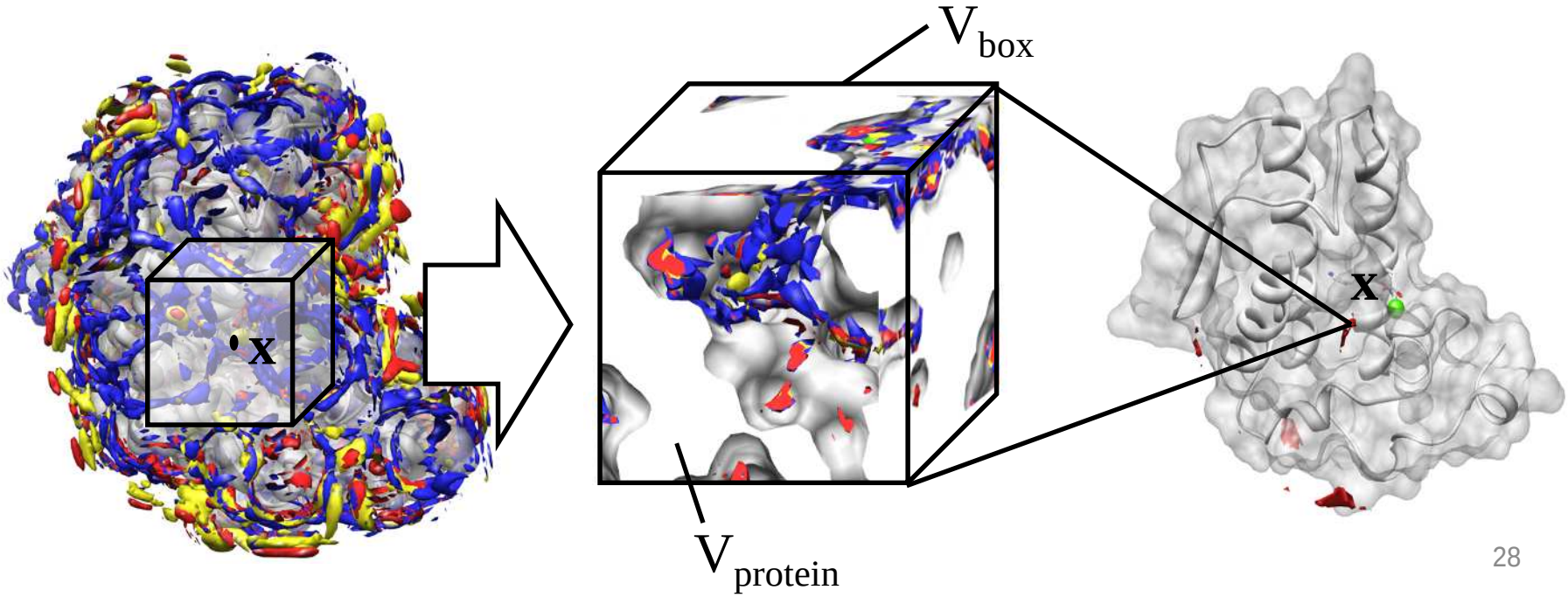


(g_γ >2.5)

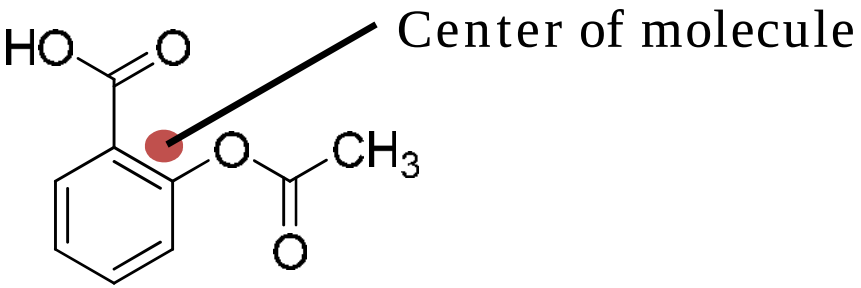
Box Search

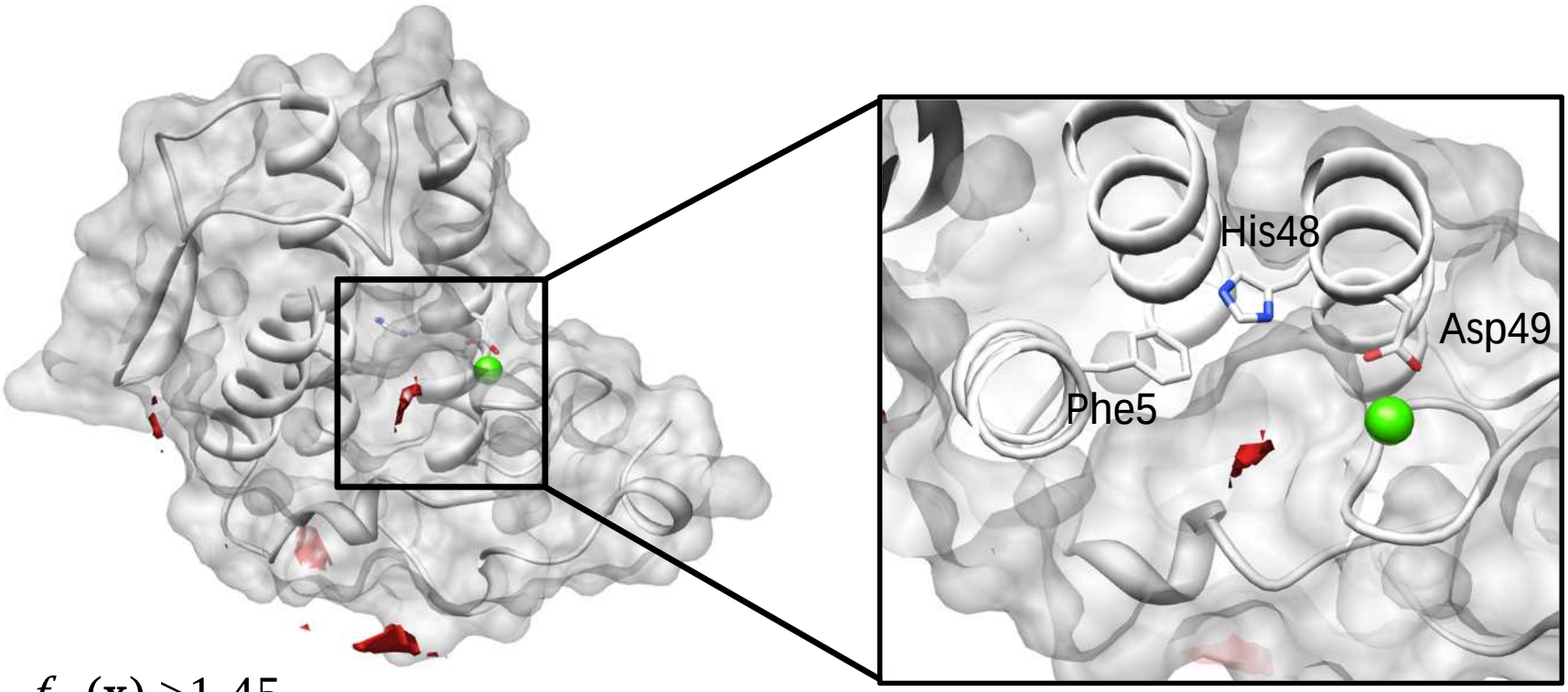
Affinity based on the probability of ligand:

$$f_{loc}(\mathbf{x}) = \begin{cases} \left(\prod_{\gamma}^{N} \frac{1}{V_{box} - V_{protein}(\mathbf{x})} \int_{V_{box}} g_{\gamma}(\mathbf{r}, \mathbf{x}) d\mathbf{r} \right)^{\frac{1}{N}} & for \quad V_{box} - V_{protein}(\mathbf{x}) \geq V_{ligand} \\ 0 & for \quad V_{box} - V_{protein}(\mathbf{x}) < V_{ligand} \end{cases}$$



- **Result of Box Search** aspirin



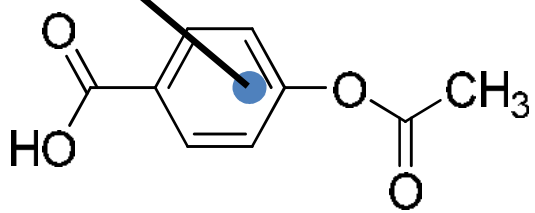


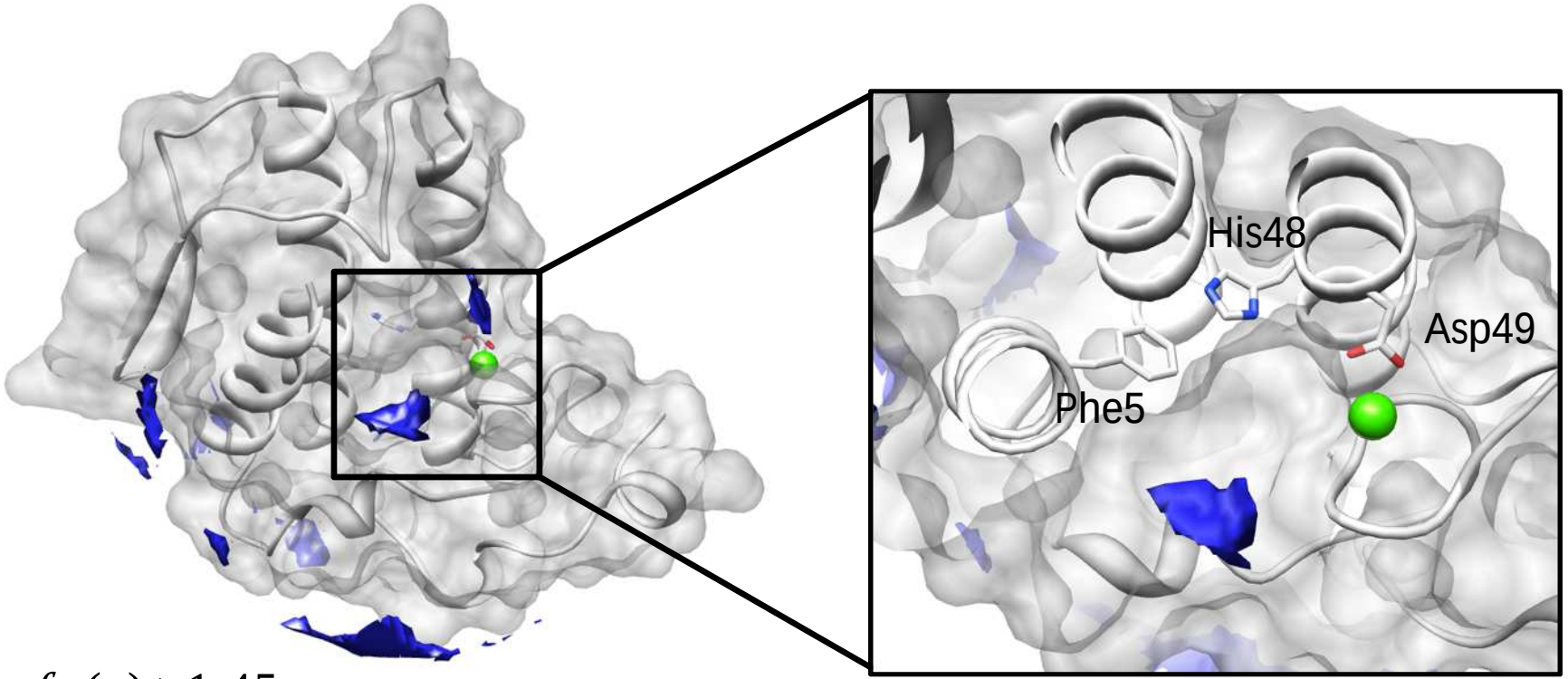
$f_{loc}(\mathbf{x}) > 1.45$

- **Result of Box Search**

p-aspirin

Center of molecule





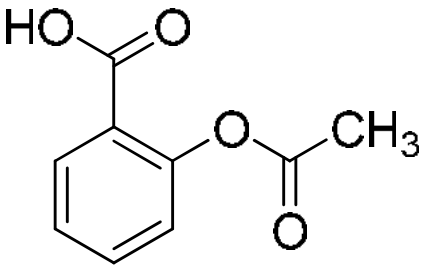
$f_{loc}(\mathbf{x}) > 1.45$

- **Result of Box Search**

Overlap

aspirin

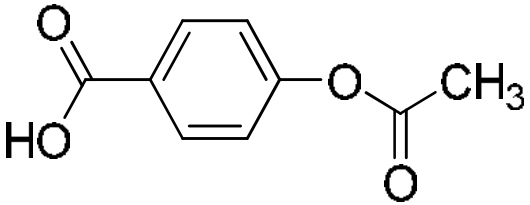
Red:



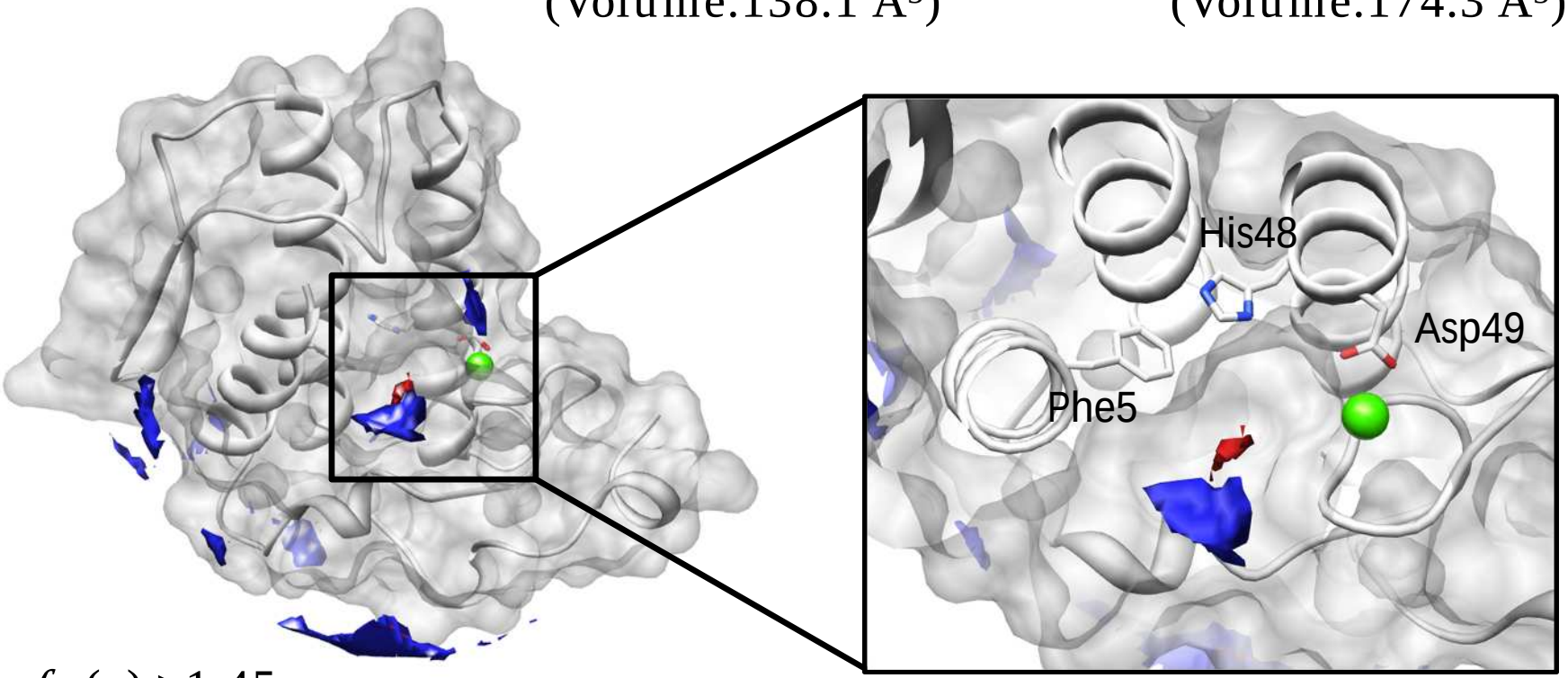
(Volume:138.1 Å^3)

p-aspirin

Blue:



(Volume:174.3 Å^3)

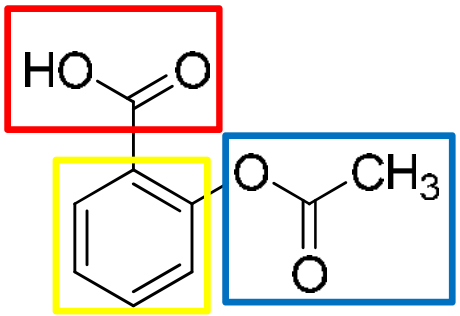


$f_{loc}(\mathbf{x})$ >1.45

Orientation Search

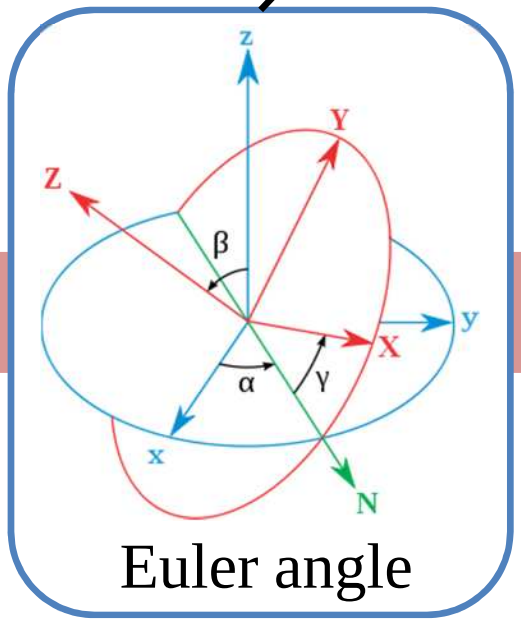
Affinity based on overlap:

$$f_{ori}(\mathbf{x},\mathbf{\Omega}) = \prod_{\gamma}^{N} g_{\gamma}(\mathbf{x} + \mathbf{l}_{\gamma} \cdot \hat{\mathbf{R}}(\mathbf{\Omega}))$$

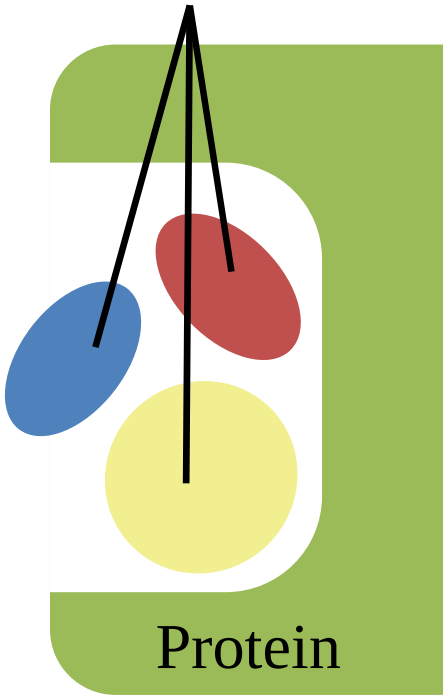


+

Center of molecule



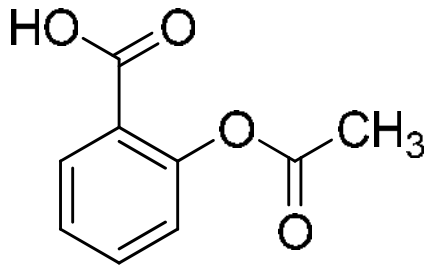
Distributions



• Result of Orientation Search

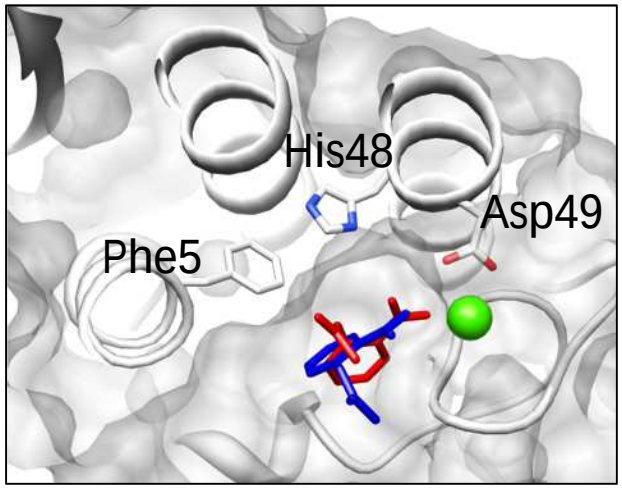
order		α	β	γ	overlap
(a)	1	10	330	50	2.33E+05
(b)	2	40	260	40	2.02E+05
(c)	3	60	220	50	1.83E+05

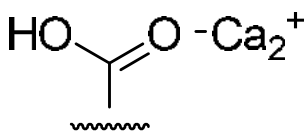
aspirin



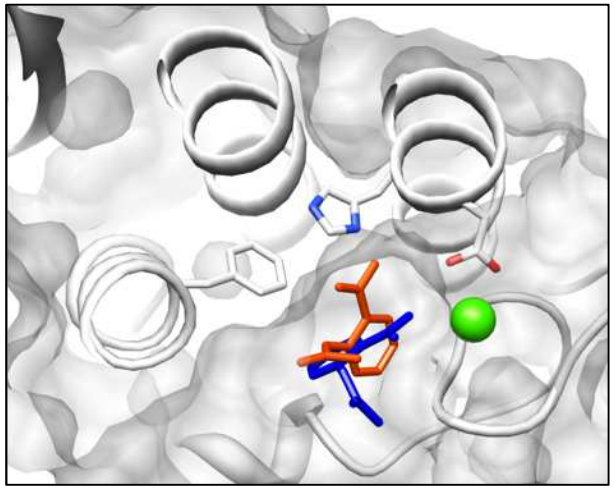
Red: Predicted
Blue: X-ray

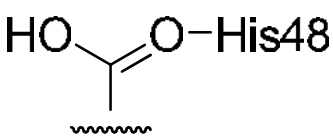
(a)



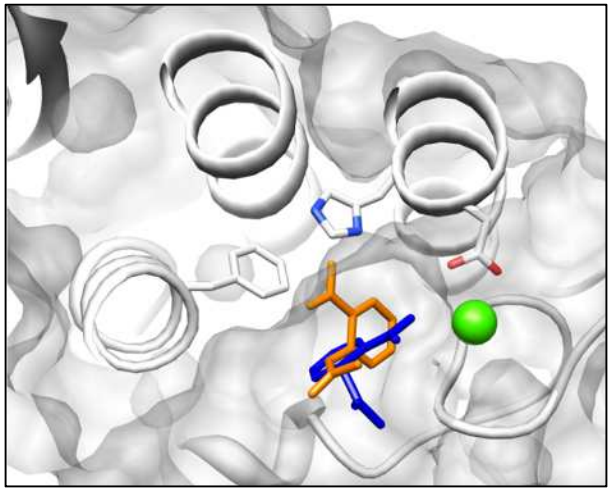


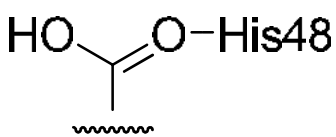
(b)





(c)

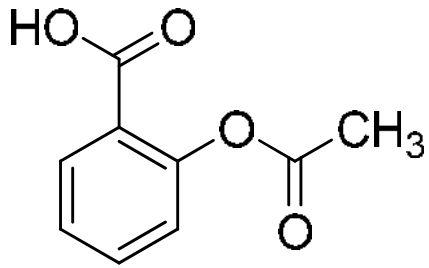




• Result of Orientation Search

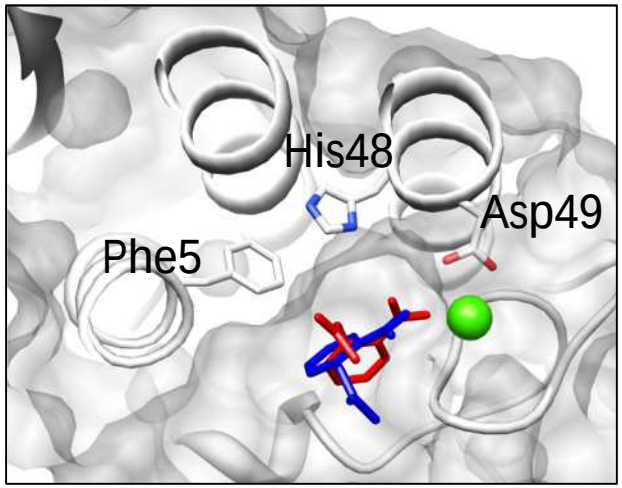
order		α	β	γ	overlap
(a)	1	10	330	50	2.33E+05
(b)	2	40	260	40	2.02E+05
(c)	3	60	220	50	1.83E+05

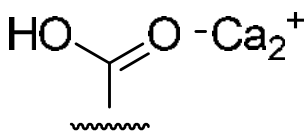
aspirin



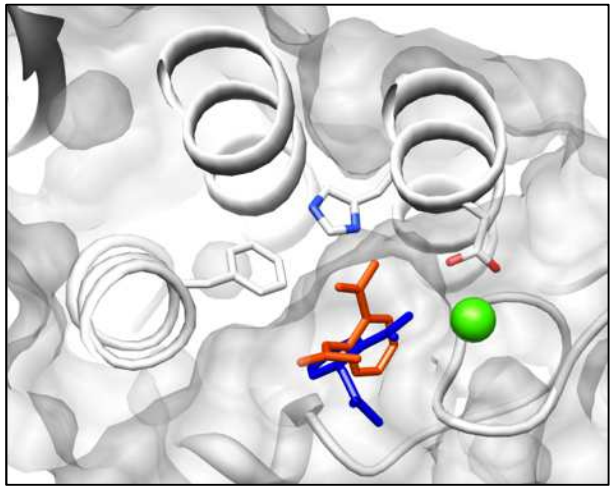
Red: Predicted
Blue: X-ray

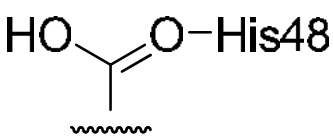
(a)



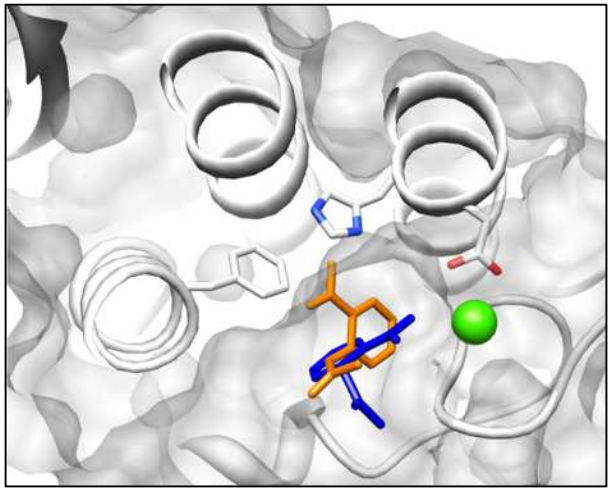


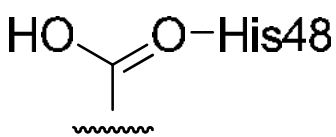
(b)





(c)

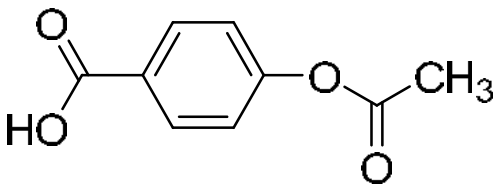




• Result of Orientation Search

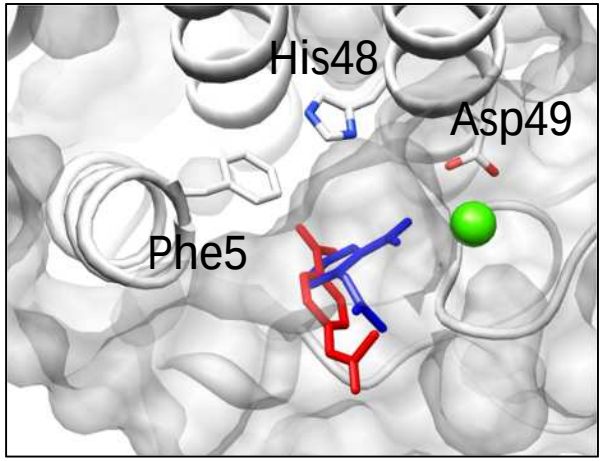
	order	α	β	γ	overlap
(a)	1	250	30	30	9.26E+07
(b)	2	60	300	150	7.72E+07
(c)	3	300	250	110	5.90E+07

p-aspirin

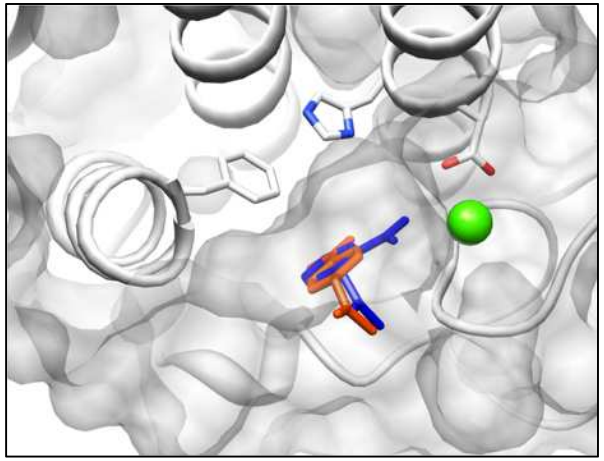


Red: Predicted
Blue: X-ray

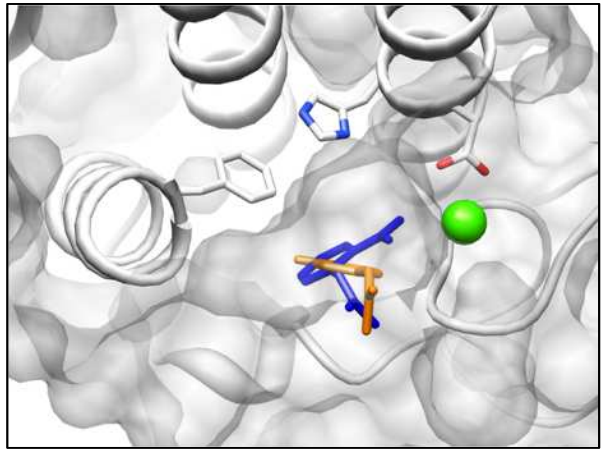
(a)

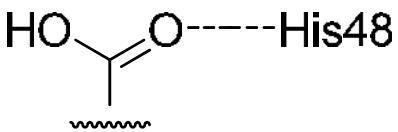


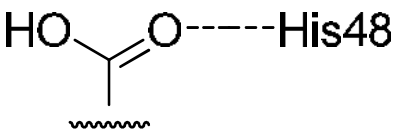
(b)

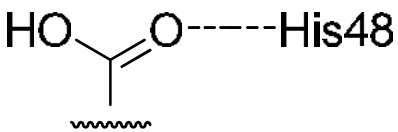


(c)



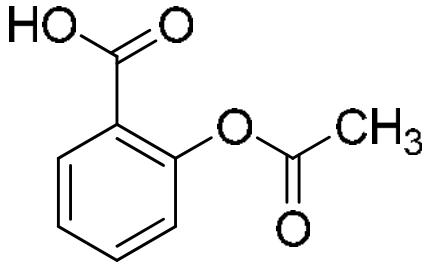


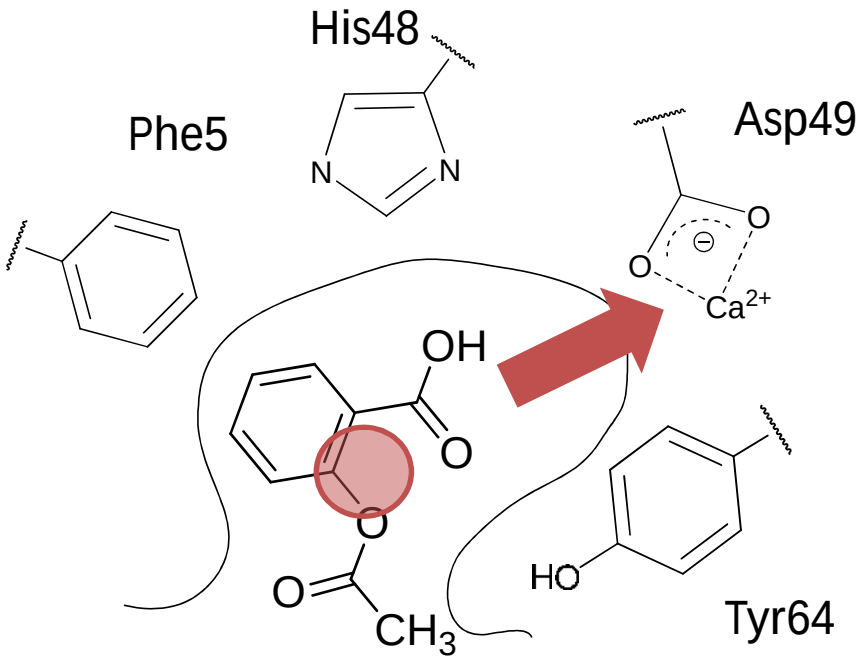




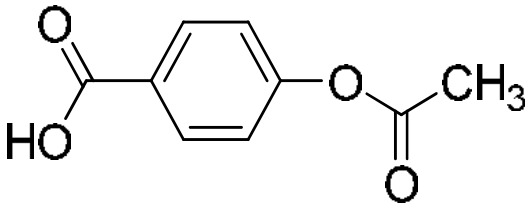
- # Binding mode

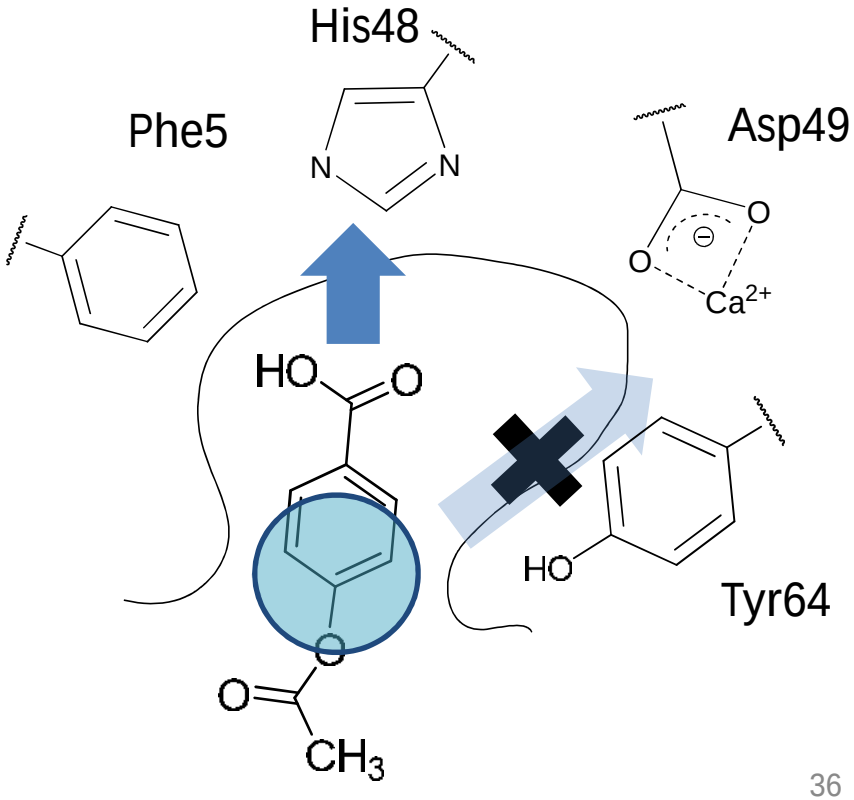
aspirin





p-aspirin





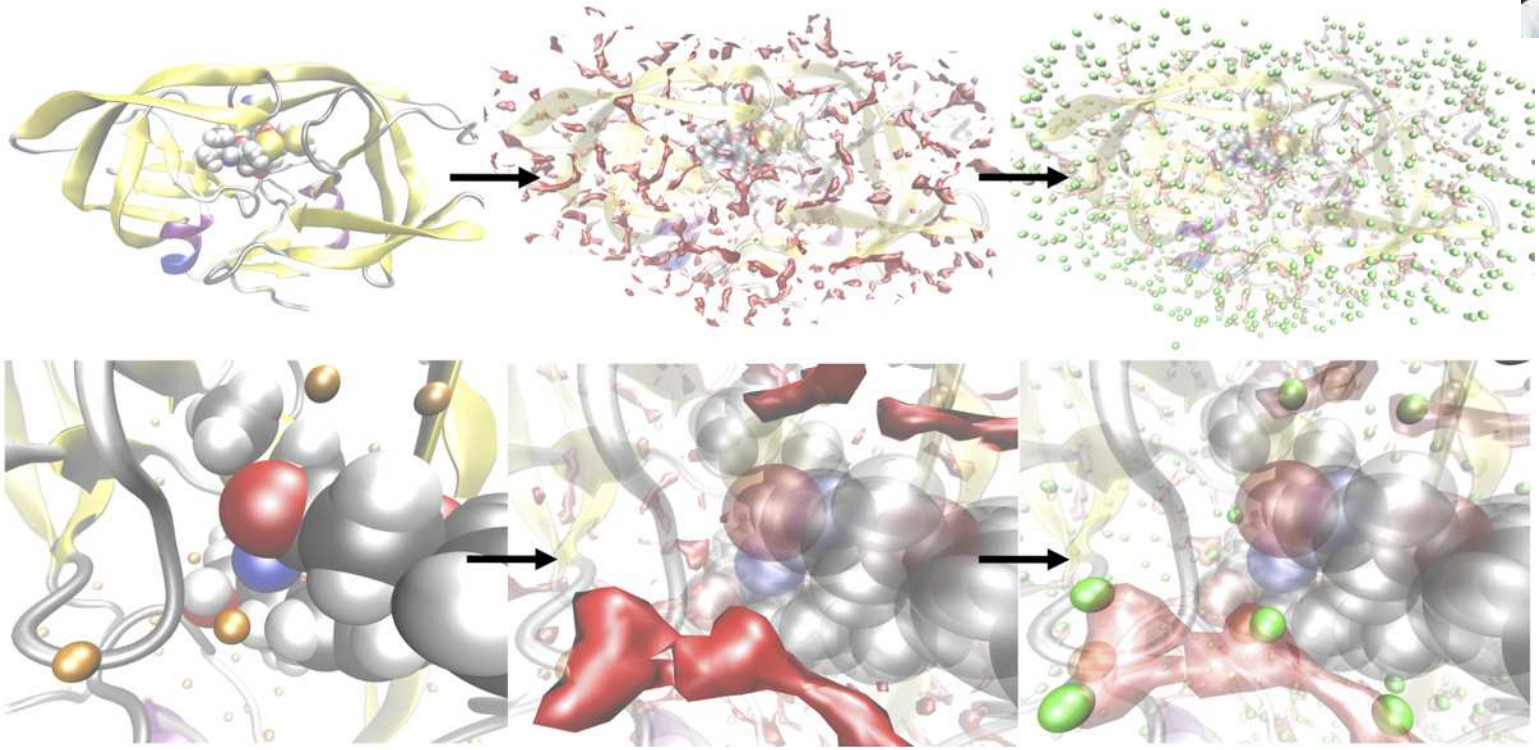
Summary

- これまで困難だったサイズのリガンドの分子認識解析が可能になった
- Phospholipase A2によるアスピリンの認識へ応用

- 課題
  - リガンドの構造揺らぎ
  - タンパク質の構造揺らぎ

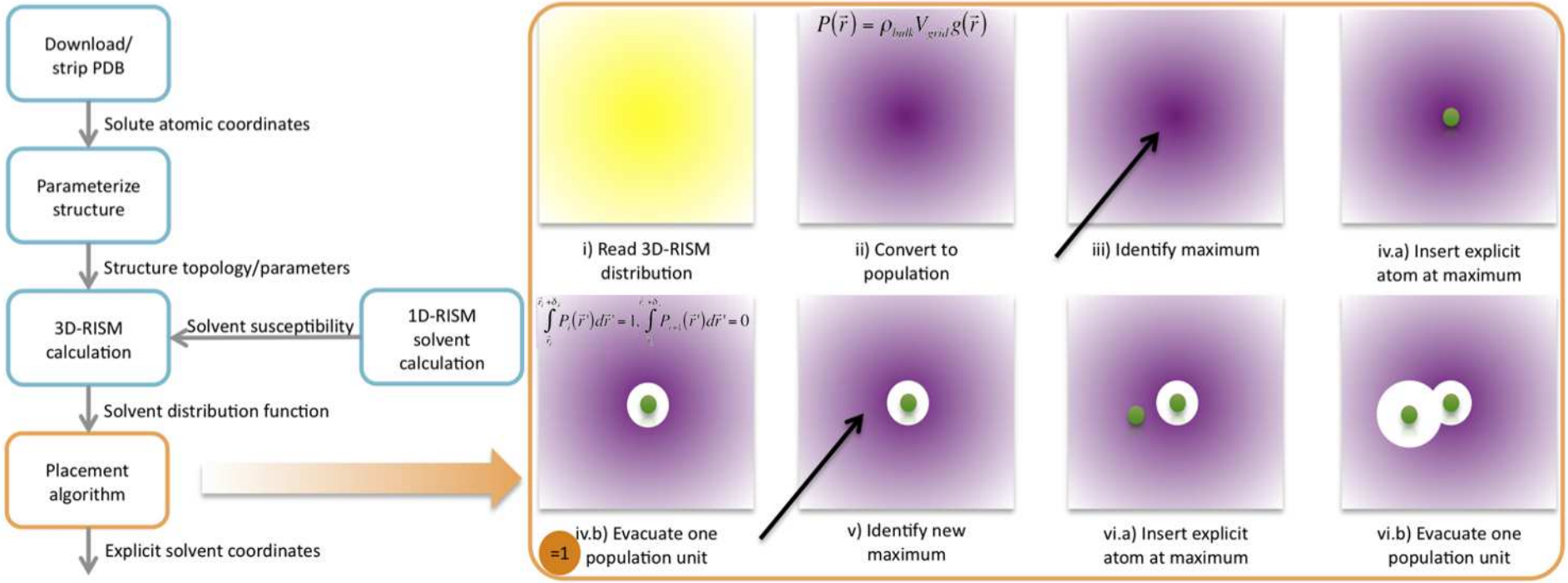
3D-RISMを用いた溶媒配置予測アルゴリズムの開発

Daniel J. Sindhikara, Norio Yoshida, Fumio Hirata, J. Comput. Chem., 2012



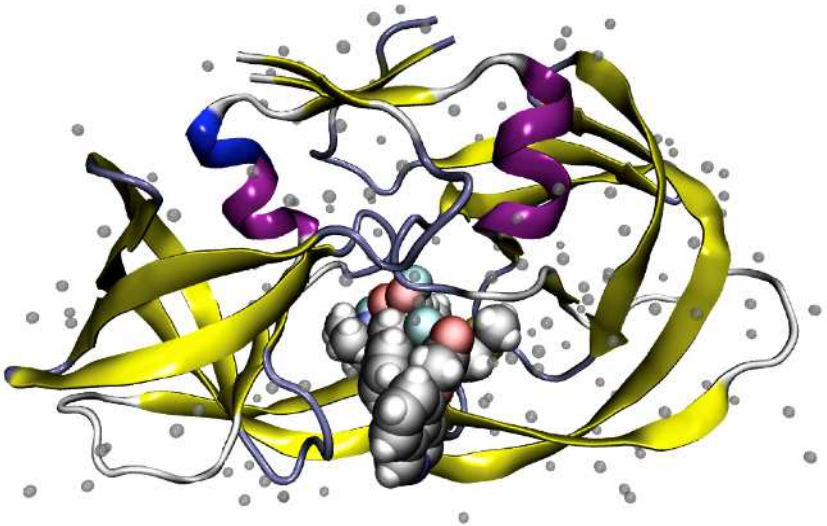
Placement algorithm

$$P(\vec{r}) = \rho_{bulk} V_{grid} g(\vec{r}) \qquad \int_V P_0(\vec{r}) = N_i + \int_V P_i(\vec{r}) \qquad \int_{\vec{r}_i}^{\vec{r}_i+\delta_i} P_i(\vec{r}\,')d\vec{r}\,' = 1, \int_{\vec{r}_i}^{\vec{r}_i+\delta_i} P_{i+1}(\vec{r}\,')d\vec{r}\,' = 0$$

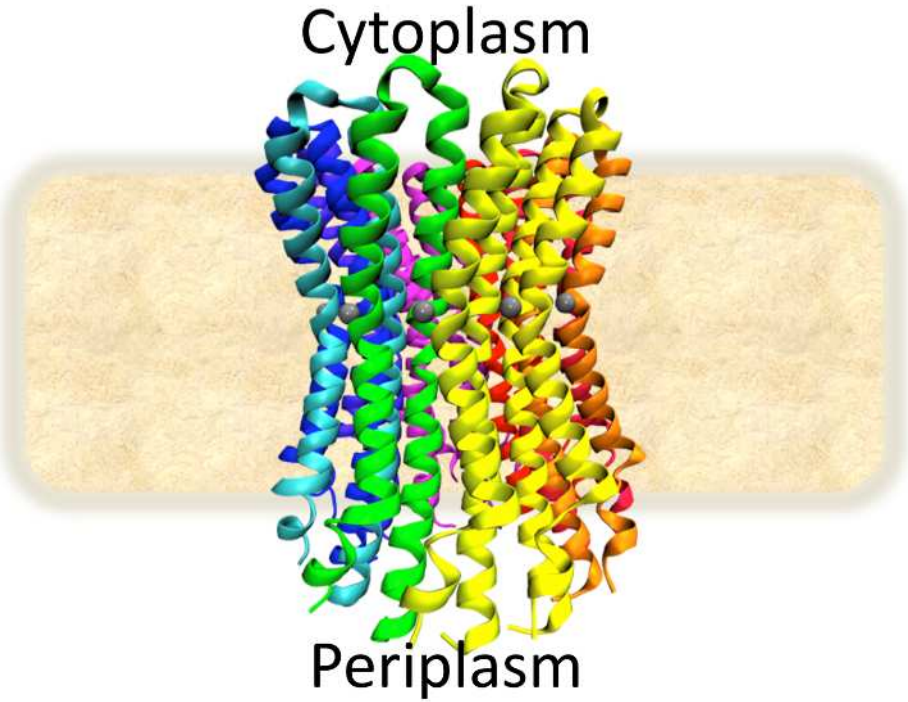


Examples

HIV-1 Protease

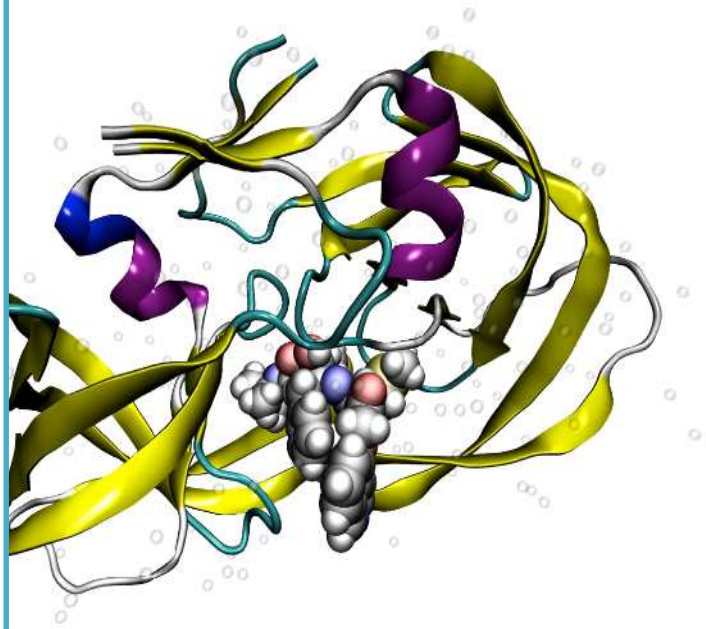


F-ATP Synthase



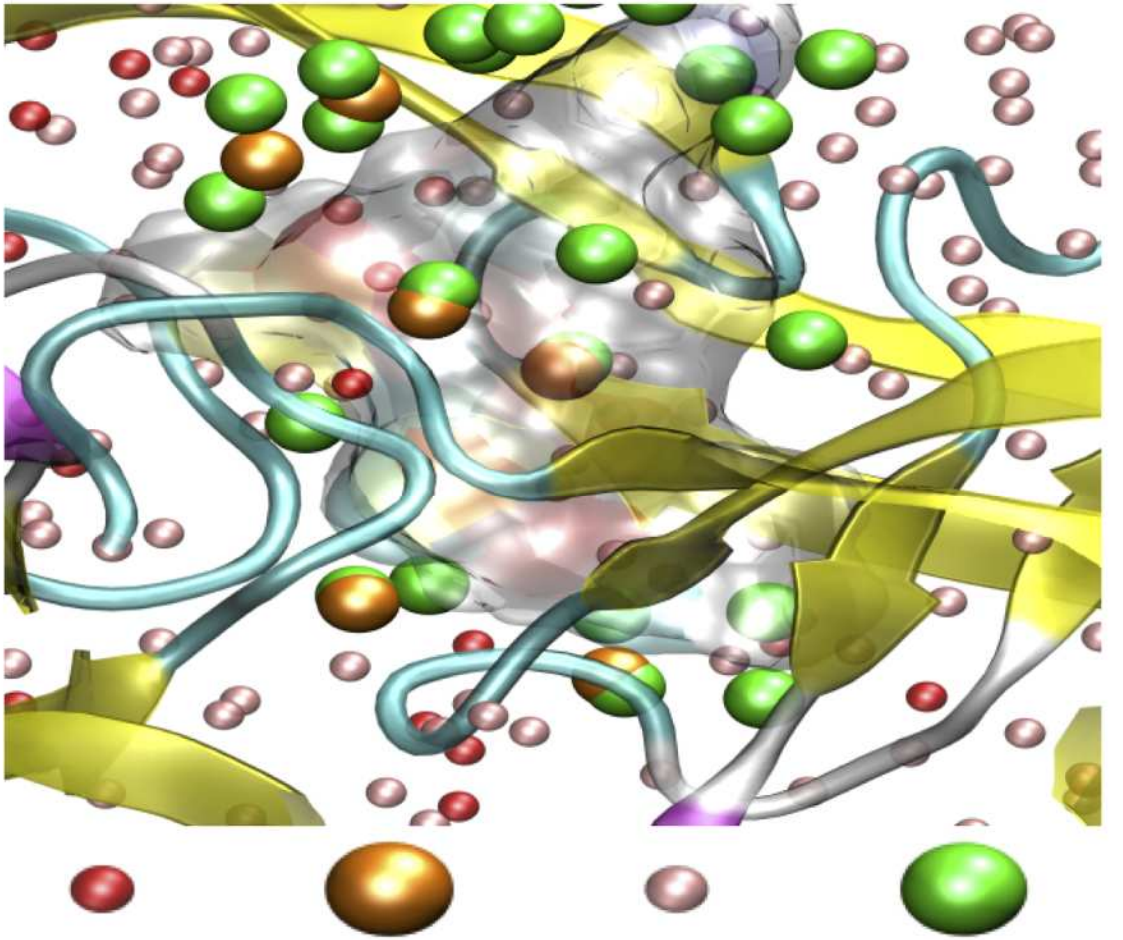
KNI-275 bound form of HIV-1 Protease

- HIV-1 protease
  - Key step in replication
- KNI-275 is a TS-analog with pM activity
- Combined x-ray, neutron crystallography reveal six key water molecules (PDB ID: 2ZYE)



Adachi, M.; Ohhara, T.; Kurihara, K.; Tamada, T.; Honjo, E.; Okazaki, N.; Arai, S.; Shoyama,Kimura, K.; Matsumura, H.; Sugiyama, S.; Adachi, H.; Takano, K.; Mori, Y.; Hidaka, K.; Kimura, Hayashi, Y.; Kiso, Y.; Kuroki, R. P Natl Acad Sci USA 2009, 106(12), 4641-4646.

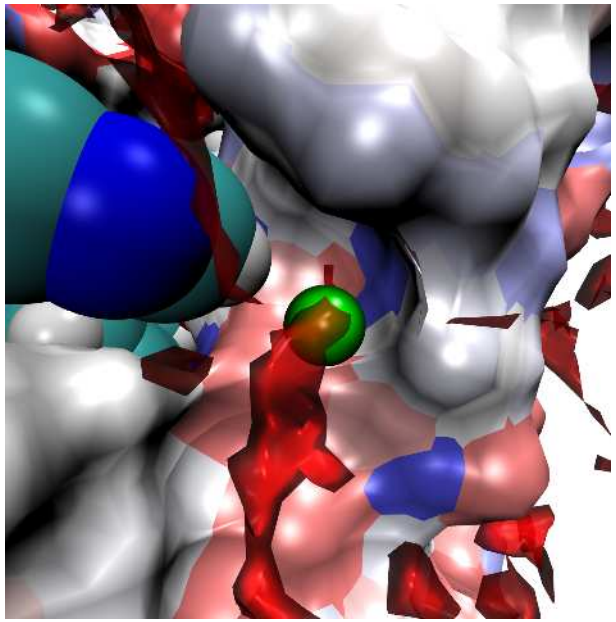
NKI275近くに見つかった水分子



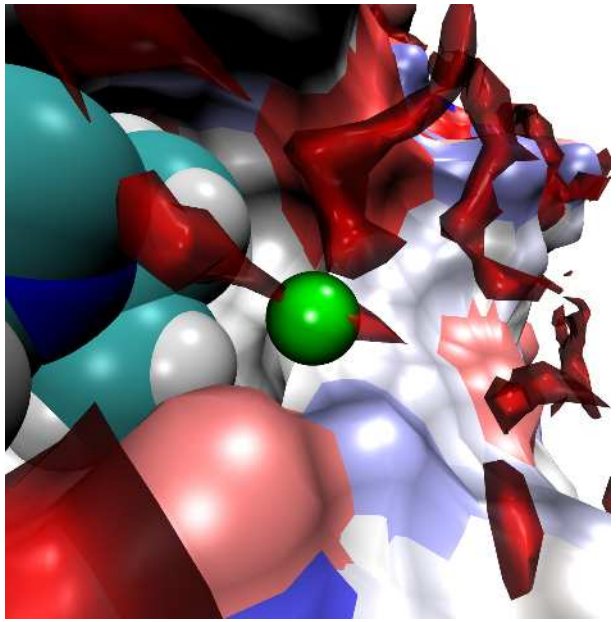
$g(r)_0$	$g(r)_i$	i_{placed}	i_{exp}	d_{exp} (Å)
17.72	17.72	2	301	0.43
10.28	10.28	39	566	0.46
10.2	10.2	43	607	0.26
10.14	10.14	44	354	1.38
9.38	9.38	52	608	0.32
8.38	8.38	89		
7.76	7.76	110		
7.62	7.62	119	322	0.79
7.39	7.39	130		
7.33	7.33	134		
7.25	7.25	142		
7.05	7.05	149		
7.03	7.03	152		
6.56	6.56	182		
4.58	4.58	372		
4.51	4.51	383		
4.23	4.23	425		
3.91	3.91	481		
3.77	3.77	497		
3.5	3.5	548		
3.49	3.49	552		
2.66	2.66	730		
6.68	2.53	763		
4.68	2.46	782		
10.25	2.35	807		
8.01	1.84	885		

http://goo.gl/uLohs

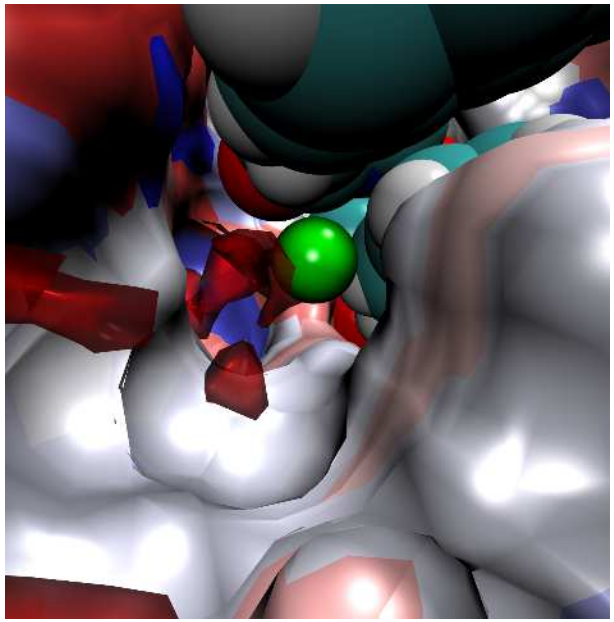
What didn’t crystallography find?



89



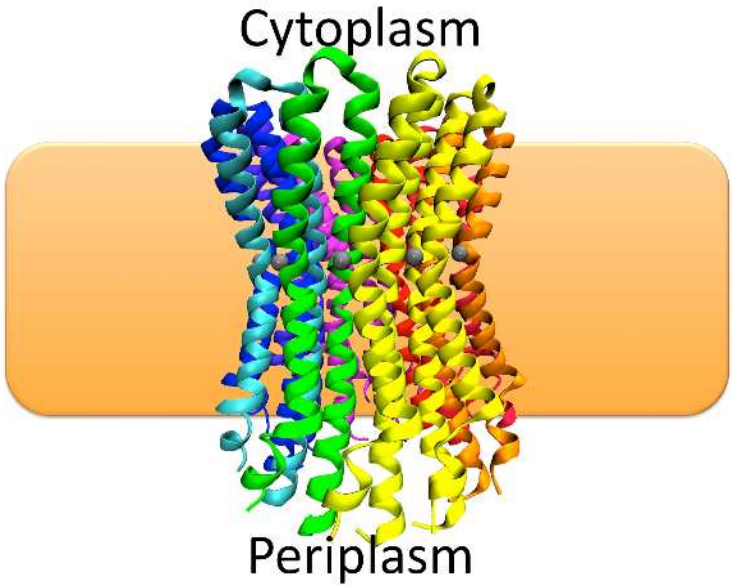
110



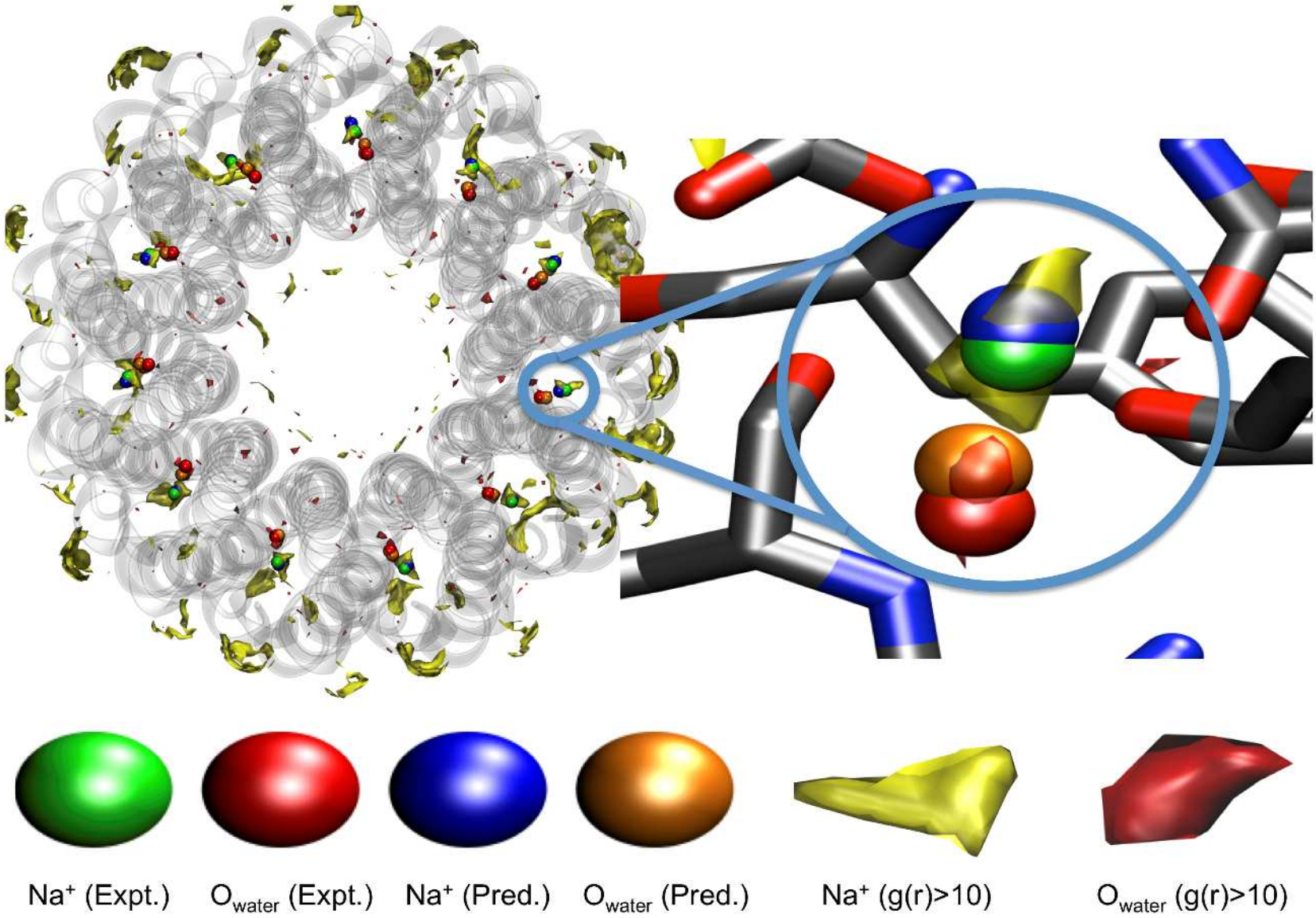
130

Rotor Ring of F-Type Na+-ATPase

- Part of F-ATPase machinery (animation)
- Rotor driven by electrochemical gradient
- Crystallography reveals sodium and (after refinement) coordinating water atomic structure
- 11-fold symmetry



Meier, T.; Polzer, P.; Diederichs, K.; Welte, W.; Dimroth, P. Science 2005, 308(5722), 659-662.
43.
Meier, T.; Krah, A.; Bond, P. J.; Pogoryelov, D.; Diederichs, K.; Faraldo-Gomez, J. D. J Mol Biol 2009, 391(2), 498-507. (PDB:2wgm)


Na^+ (Expt.)
O_{water} (Expt.)
Na^+ (Pred.)
O_{water} (Pred.)
Na^+ (g(r)>10)
O_{water} (g(r)>10)

Sodium				water			
$g(r)_0$	$g(r)_i$	i_{placed}	d_{exp} (Å)	$g(r)_0$	$g(r)_i$	i_{placed}	d_{exp} (Å)
29.19	29.19	1	0.29	15.31	15.31	11	0.88
28.77	28.77	2	0.44	14.7	14.7	17	0.80
27.48	27.48	3	0.59	14.47	14.47	24	0.58
27.34	27.34	4	0.35	14.24	14.24	27	0.44
26.88	26.88	5	0.27	14.03	14.03	30	1.05
26.65	26.65	6	0.32	13.8	13.8	36	0.93
25.86	25.86	7	0.29	13.62	13.62	44	0.78
25.6	25.6	8	0.48	13.31	13.31	51	0.48
25.28	25.28	9	0.52	13.26	13.26	53	0.41
24.89	24.89	10	0.84	13.02	13.02	59	0.60
17.19	17.19	19	0.53	12.94	12.94	63	0.23

Summary

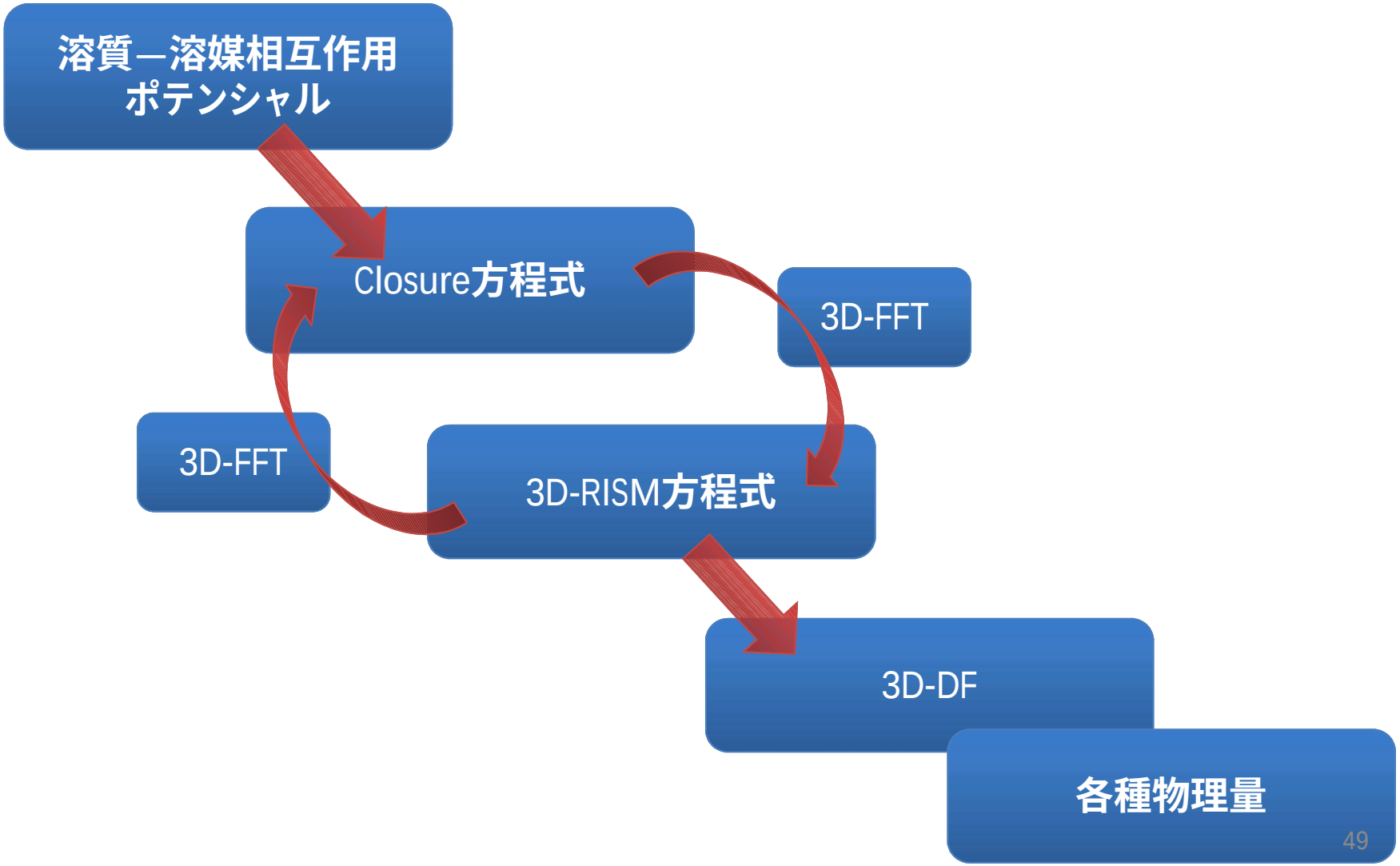
- Discrete distributions can be coerced out of 3D-RISM distributions for both water and ions
- This technique matches quantitatively (about 0.5 Å error) with experiment while providing more information
- Ion concentration required manual adjustment
- Can be used to add perspective to 3D-RISM, get insight into specific water and/or ion interactions, provide an initial snapshot for MD simulations, give insights into drug design
- Info and tutorial on website, work done using free software (Ambertools and scipy) http://goo.gl/uLohs

3D-RISMの京への実装

丸山豊博士（大阪大）

筑波大学　佐藤グループ
多田野寛人助教，高橋大介准教授，佐藤三久教授

3D-RISMのアルゴリズム



(従来型) 3D-FFTの並列化の問題点

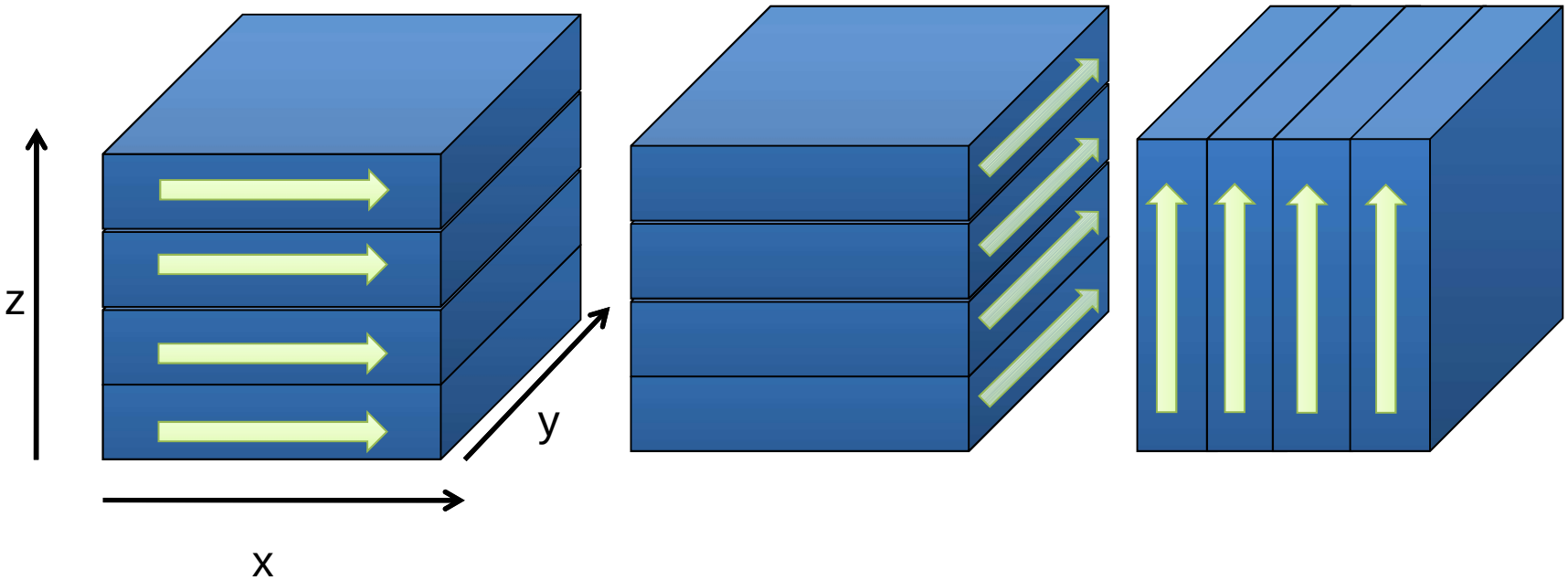
- 三次元空間の1方向 (例えばz軸) を分散して並列化
- 最大で1軸分の離散点数までの並列化しか出来ない
  - 3D-RISMでは次世代スパコンで2048^3のグリッド点を想定
  - 従来型では最大で2048並列しかできない
- 軸交換時に全対全通信を必要とする
  - 通信律速

Volumetric並列三次元ＦＦＴ

- 近年、筑波大学高橋大介准教授らによって開発された並列化手法
- 三次元空間の2軸を分散して並列化
  - 2048^3なら2048^2並列まで可能
  - 軸交換時の全対全通信の通信量を削減可能
- 3D-RISMに実装することで高効率並列計算が可能になる

従来の3D-FFTの並列化

- Z軸方向へ一次元ブロック分割



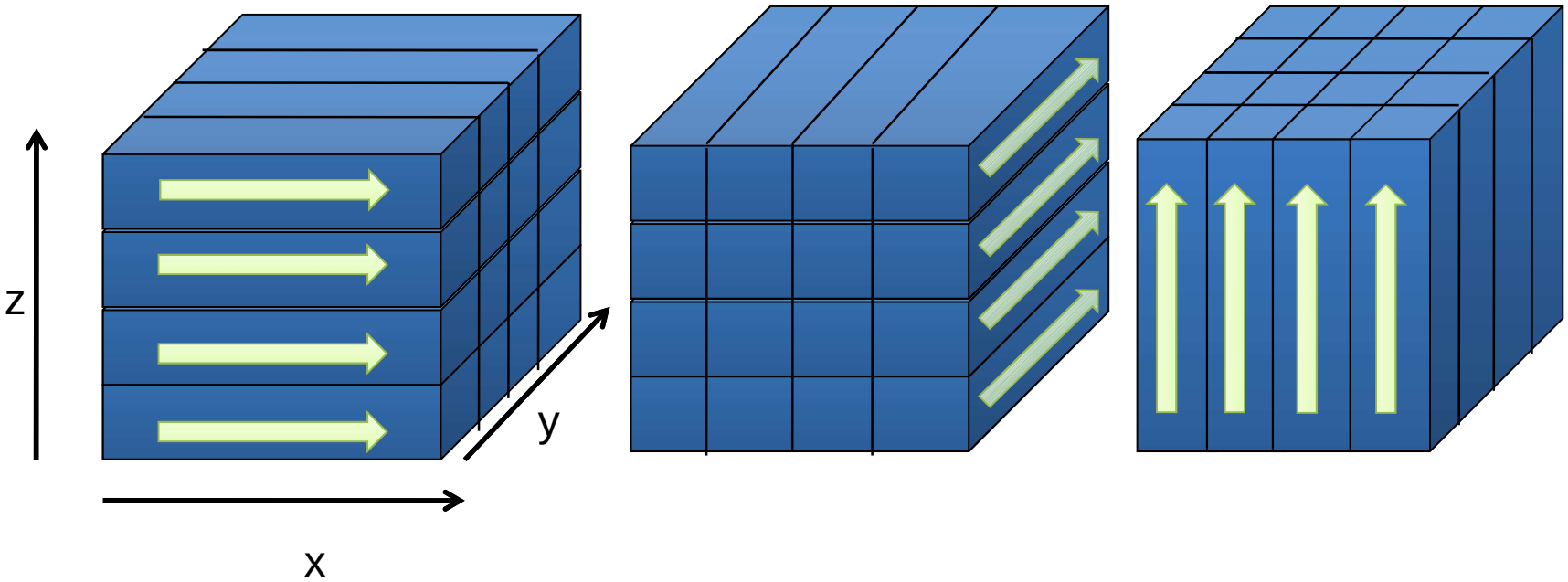
x軸方向にFFT → y軸方向にFFT → z軸方向にFFT

全対全通信による軸交換

全対全通信による軸交換

Volumetric並列3D-FFT

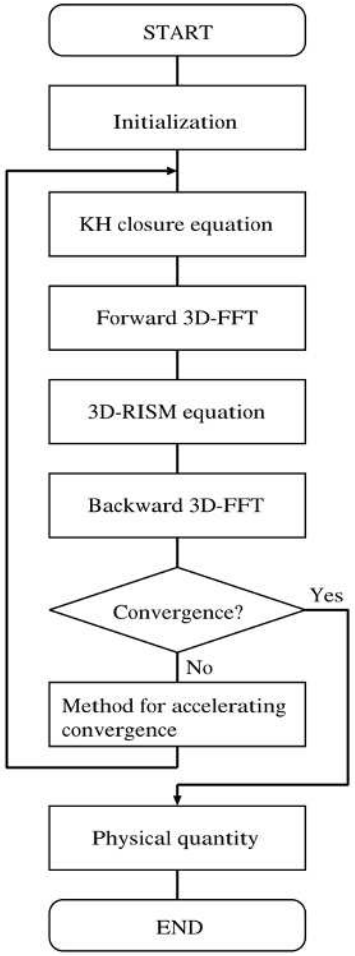
- y,z軸方向へ二次元ブロック分割



x軸方向にFFT → y軸方向にFFT → z軸方向にFFT →

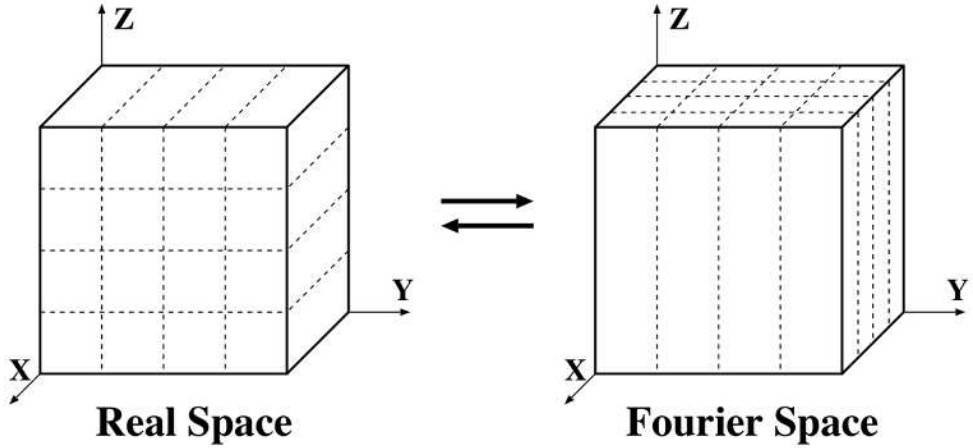
xy面での軸交換　yz面での軸交換　xz面での軸交換

通信の半減



←通信

←通信



実空間とフーリエ空間で分割のレイアウトが変わる事にプログラム側で対応する事によりFFT部分の通信量を半減させた。

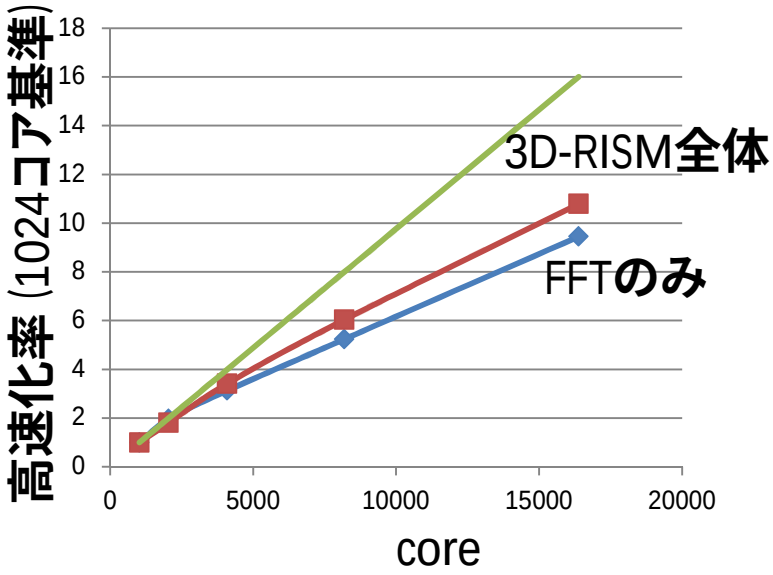
3D-RISMの京への実装

ナノ統合中核アプリとしての目標

- 1万ノード程度でスケール

結果

- 1024^3グリッド
  2048ノードまでテスト済み
  (右図)
- 2048^3グリッド
  8192ノードでスケール
  4096ノード比で88%の効率

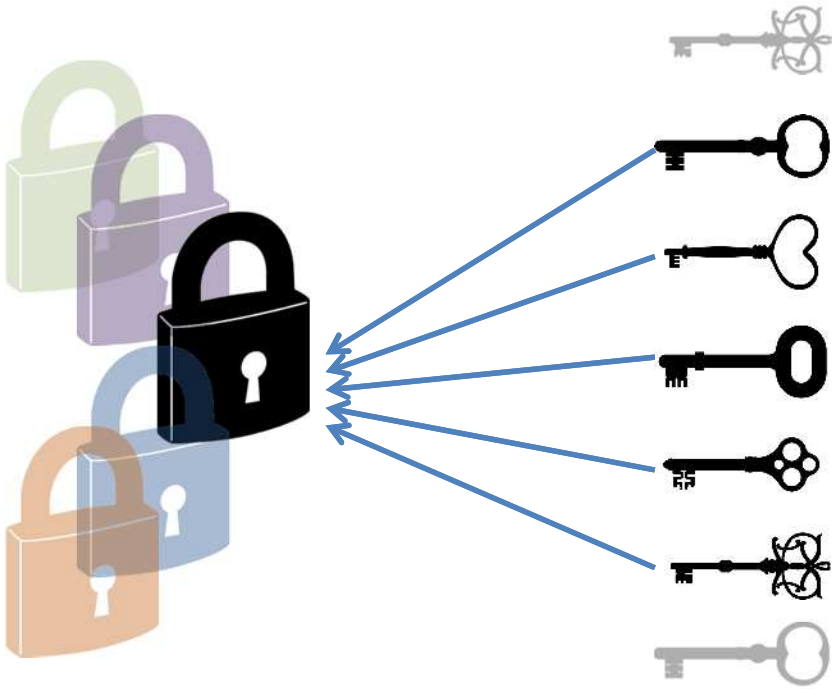


京で可能になること･やりたいこと

- ドラッグスクリーニング

- 各種連成計算
  - 構造サンプリング
  - 電子状態計算

ドラッグスクリーニング



標的たんぱく質とリガンドの認識を予測・解析

- **どこに？　認識部位探索**
- **なにが？　リガンドスクリーニング**

3D-RISMのアドバンテージ

- **たんぱく質を溶質として，リガンドを溶媒（あるいは第2の溶質）として計算することで上記2つを同時に解析可能**

京での展開

- **1つのたんぱく質（構造）・1つのリガンドに対して1回の3D-RISM収束計算が必要**
- **3D-RISMを高速に解くことで多くの構造・リガンドへの解析が可能**

課題

- **たんぱく質の構造揺らぎ**
- **リガンドの構造揺らぎ**

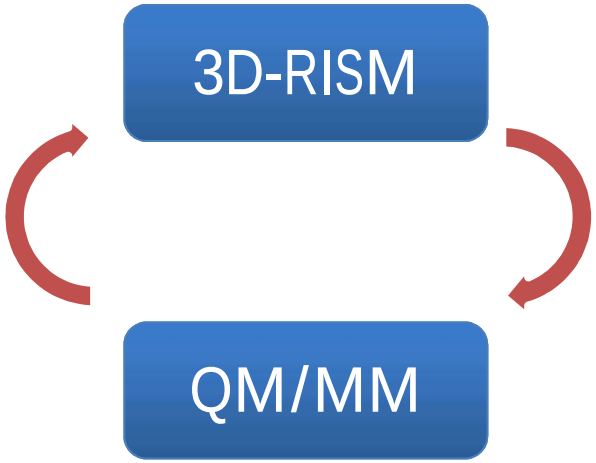
連成計算

各種連成計算でできること
- たんぱく質構造サンプリング
  - 3D-RISMとMD, MCとの連成計算
- 化学反応解析
  - 3D-RISMとQM/MMの連成計算

3D-RISMのアドバンテージ
- 溶媒の温度依存性，密度依存性などが簡単に扱える
- 共溶媒効果や希薄電解質を扱える

京での展開
- セルロース分解酵素反応解析
- リグニン構造の温度依存性，溶媒依存性

課題
- 高効率連成計算



今後の展開

- 理論的課題
  - 溶質分子の構造揺らぎ
  - 溶媒分子の内部構造揺らぎ

- プログラムの課題
  - 連成計算の高度化
  - ポスト/プリプロセスの充実
    - インプット作成支援
    - アウトプット解析支援

エクサスケールコンピューティングに向けて

- **エクサスケールコンピュータ**
  - Many-cores?
  - GPU?
- 3D-RISM　for GPGPU
  - **丸山豊博士 (大阪大)**
    Modified Anderson Method for Accelerating 3D-RISM Calculations Using GPU, Y. Maruyama and F. Hirata, JCTC, 2012
- **大並列化には**multi-GPU**用**3D-FFT**が必要**
  - **額田彰博士 (東工大)**
    High Performance 3-D FFT using multiple CUDA GPUs, A. Nukada, Y. Maruyama, S. Matsuoka, GPGPU-5 proceedings, 2012
- **エクサに対応するなら・・・**
  - 3D-RISM**の理論と計算機科学の境界領域を押さえられる人材**
  - **中～長期的なプロジェクト推進が必要**

以上です