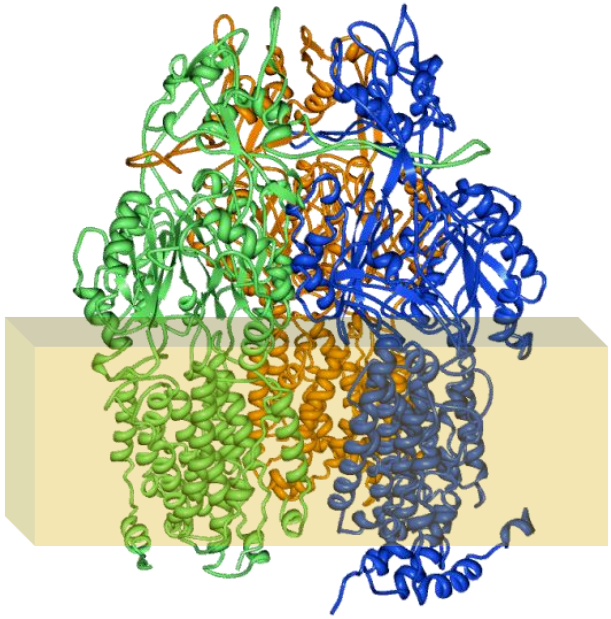
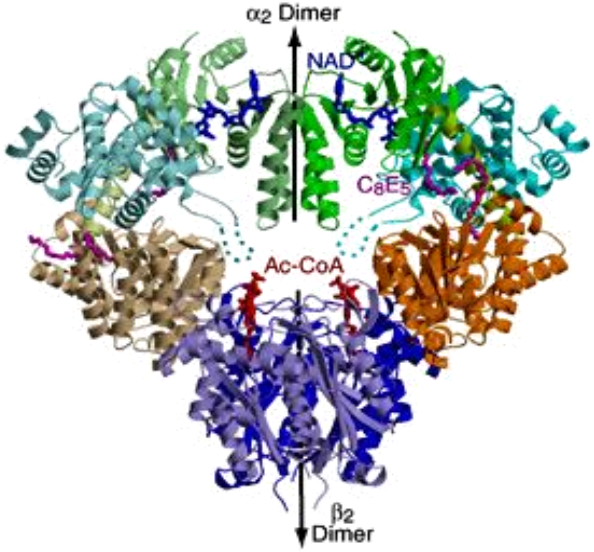
分子シミュレーションの展開
スーパーコンピューティングに向けて

次世代生命体統合シミュレーション研究開発
分子スケールチーム





理化学研究所 次世代計算科学研究開発プログラム
横浜市立大学大学院生命ナノシステム科学研究科
木寺 詔紀

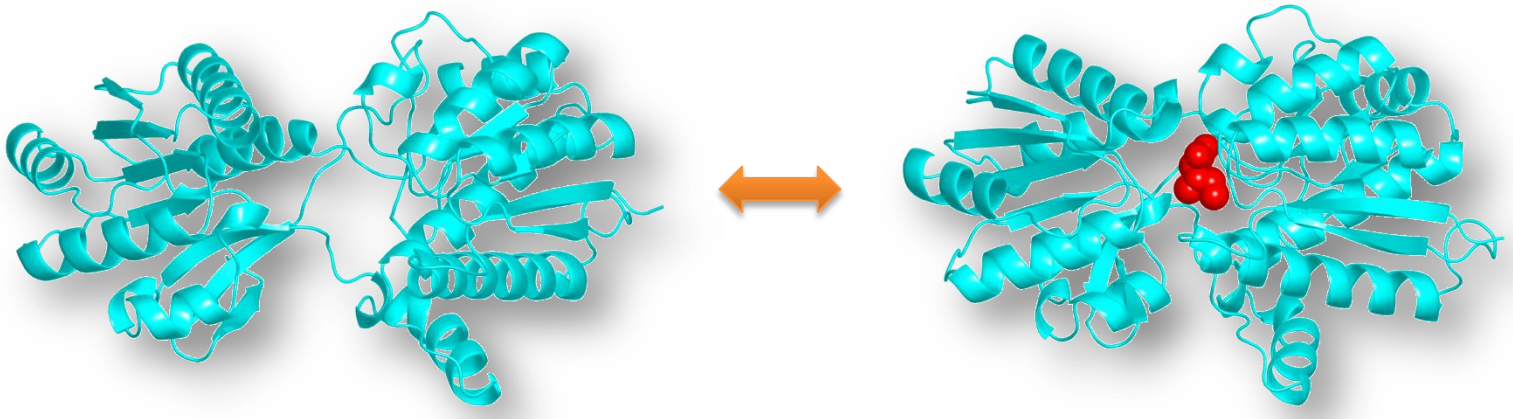
タンパク質は動く

1．このように動く
　1−1．データベースから見えるタンパク質の応答的運動
　1−2．応答的タンパク質機能像
2．そのように動くタンパク質機能の分子シミュレーション
　2−1．もっと速さを！
　2−2．次世代生命体統合シミュレーション 分子スケール
　　(1) 多剤排出トランスポーターを例に
　　分子動力学計算
　　粗視化モデル計算
　　(2) 量子化学計算
　　(3) マルチスケールシミュレーション

1．このように動く
1−1．データベースから見えるタンパク質の応答的運動

タンパク質は一般的に「動く」と言えるものなのか？
「動く」とは意味のあることなのか？
「動く」、どのように？

Protein Data Bank

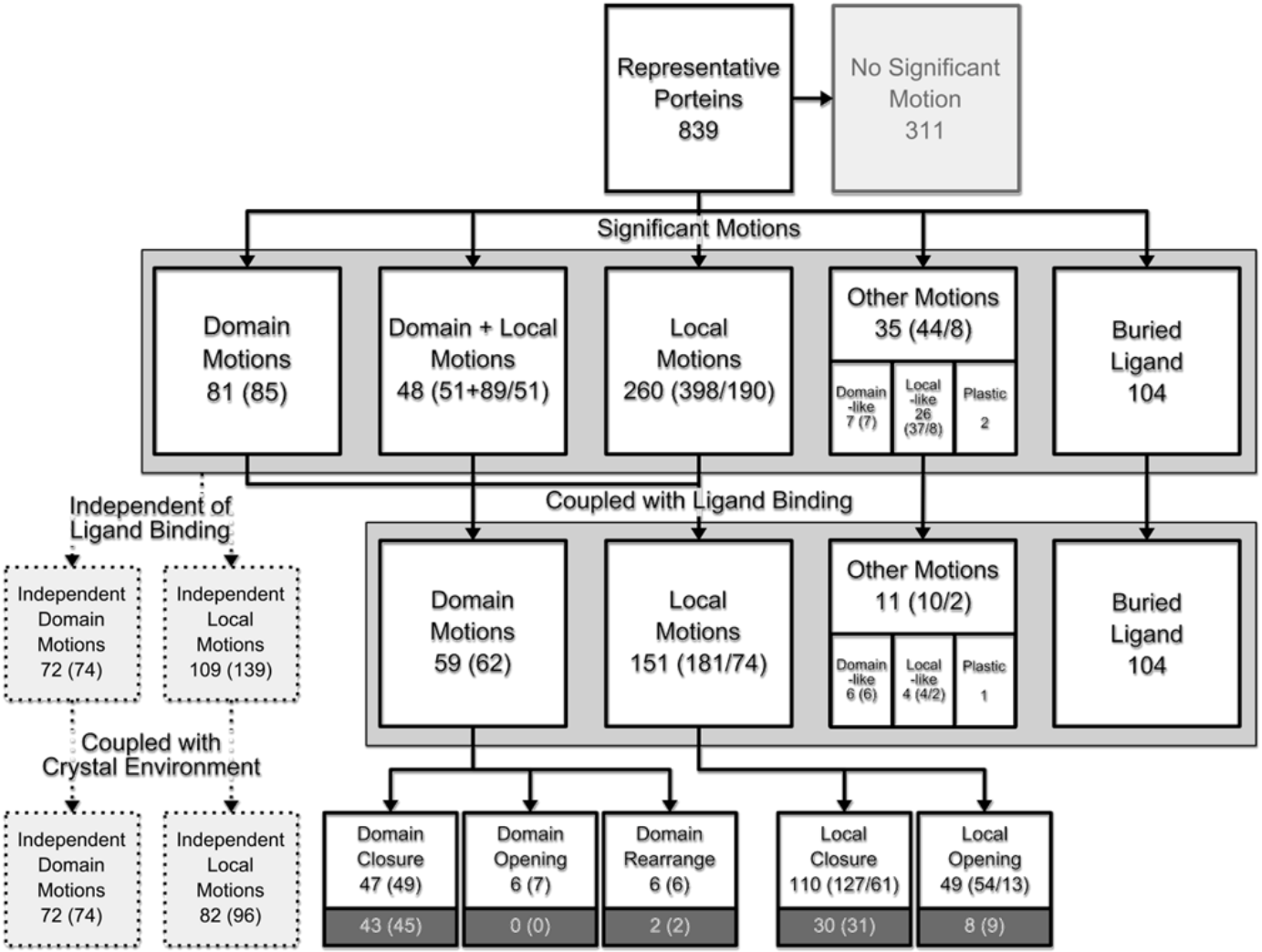


リガンド非結合状態　　リガンド結合状態

因果関係

リガンド結合　→　構造変化　　Single Molecules

Classification

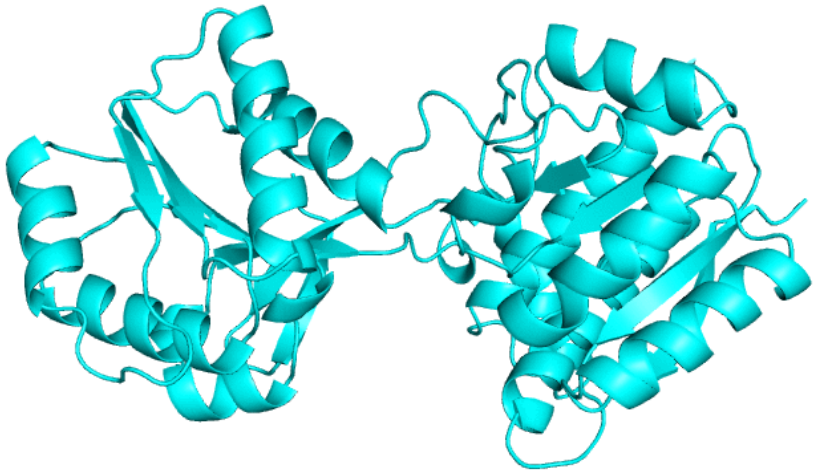


T. Amemiya *et al*. *J.Mol.Biol.* **408**, 568 (2011)

この程度動く

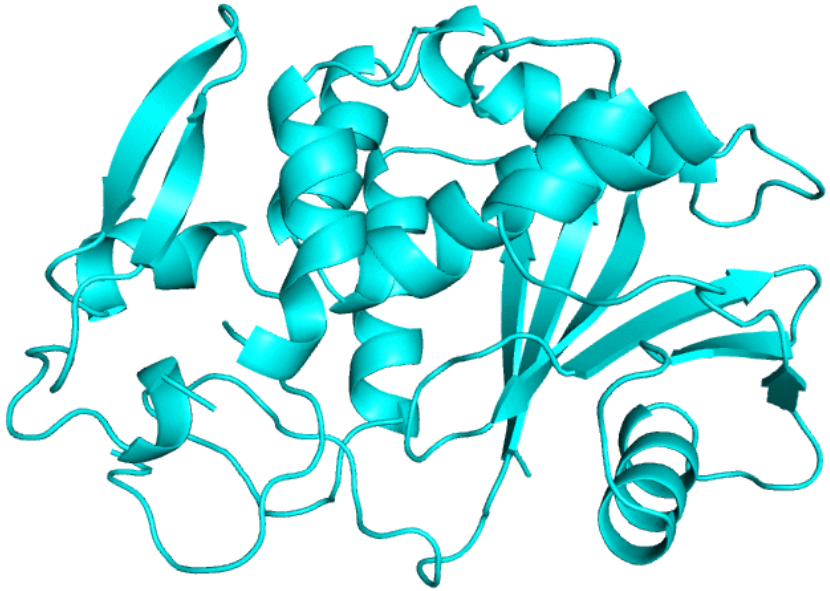
CαRMSD > 1Å

立体構造変化あり



528

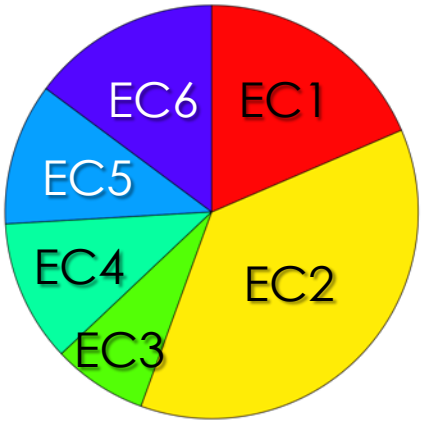
立体構造変化なし



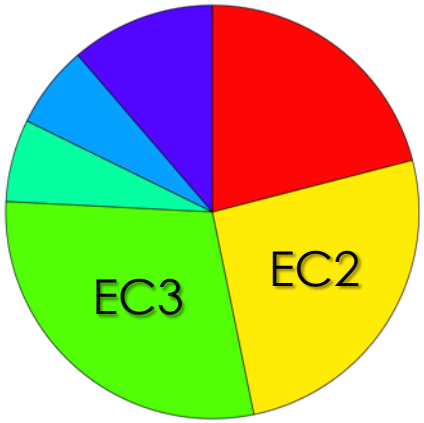
311

動かないことの意味？

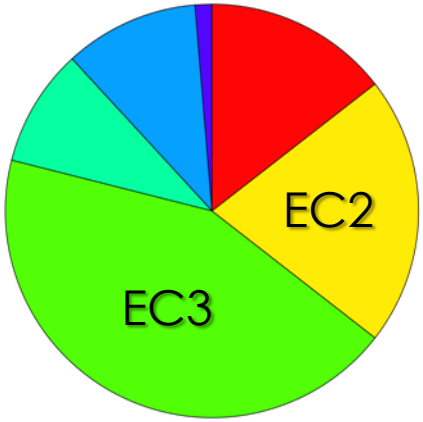
酵素機能の分布



Domain Motion



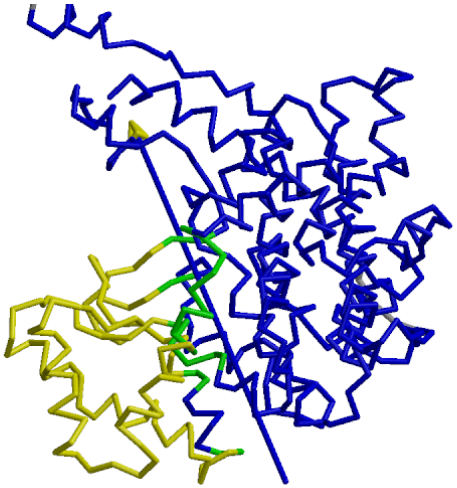
Local Motion



No Significant Motion

R. Koike *et al. J.Mol.Biol.* **379**, 397 (2008)

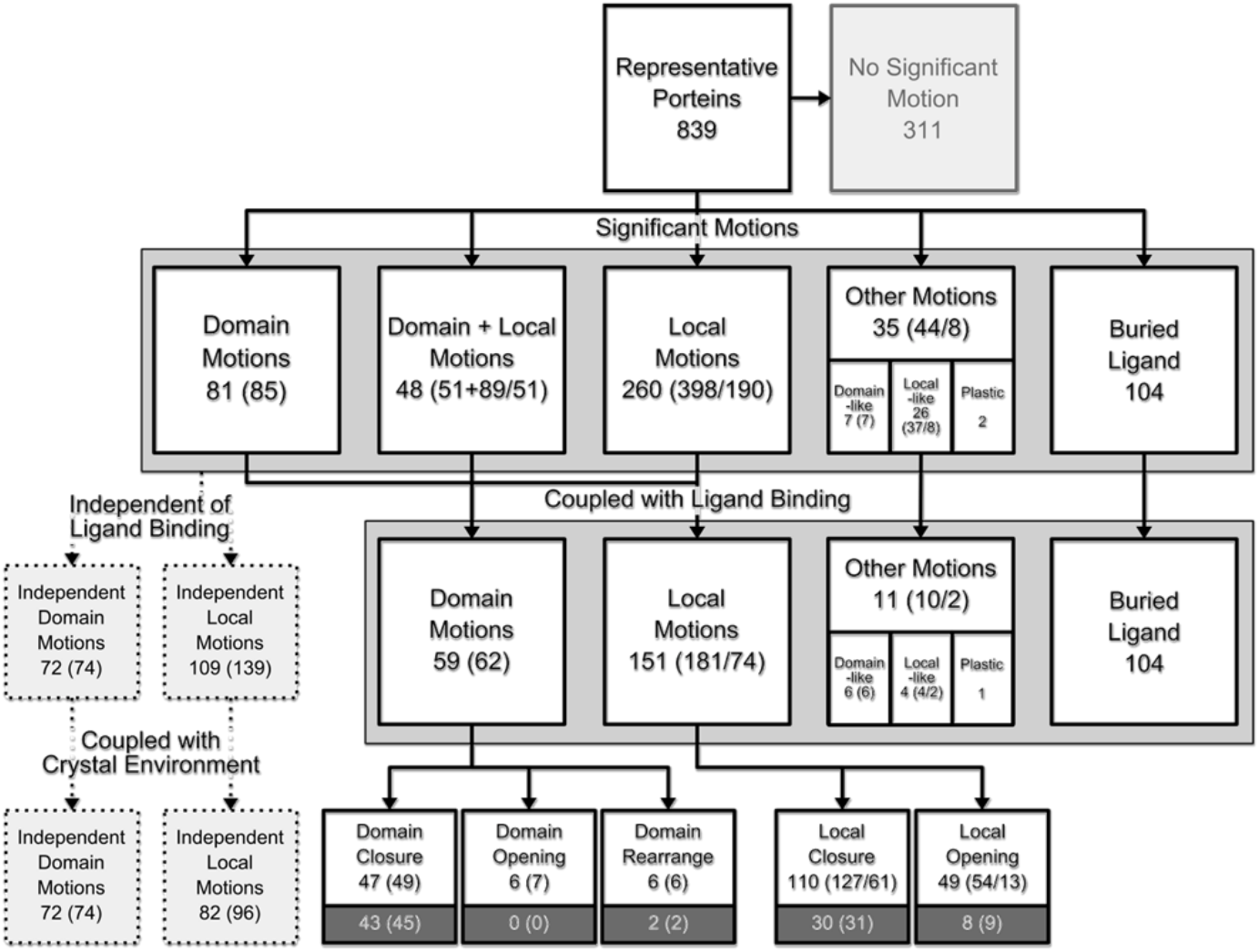
このように動く



Domain Motion

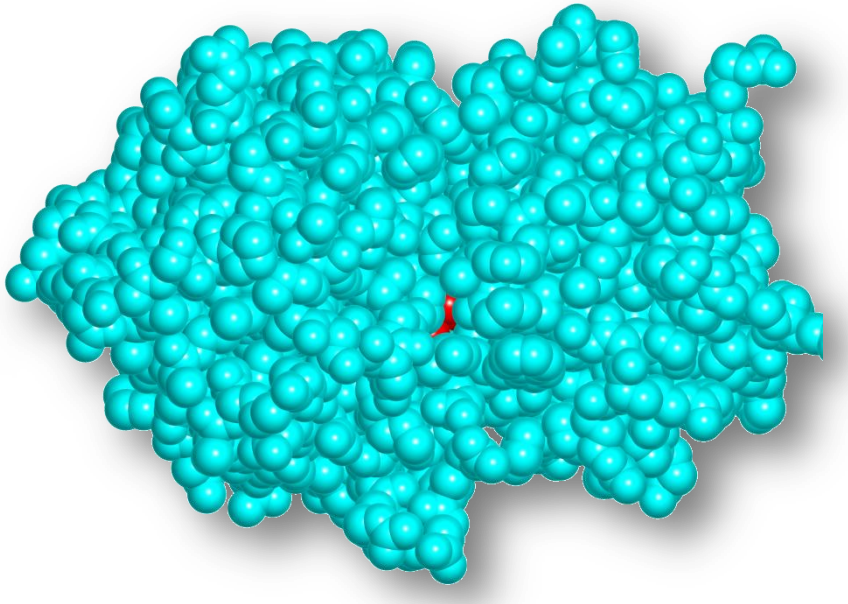
Local Motion

Classification



Buried Ligand

CαRMSD < 1Å



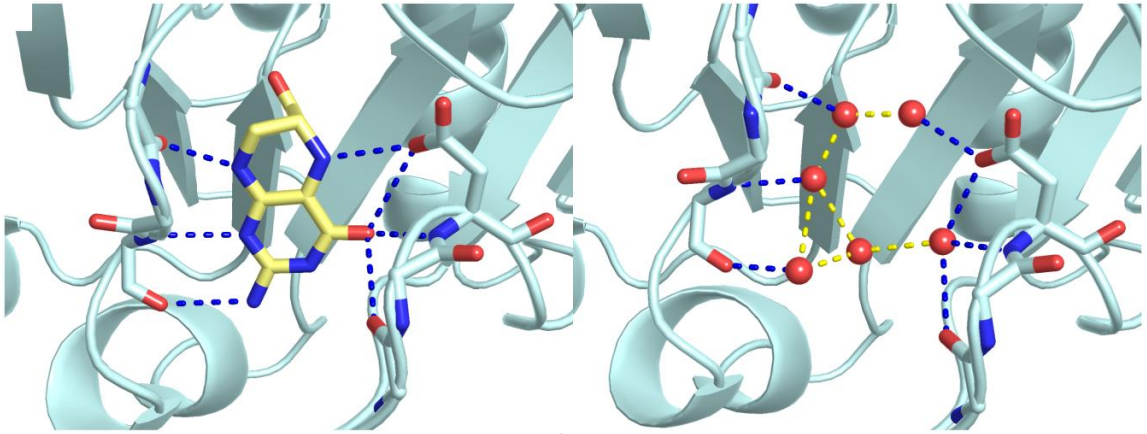
104

動かなければリガンドは入れない！

Buried Ligand Classの特徴

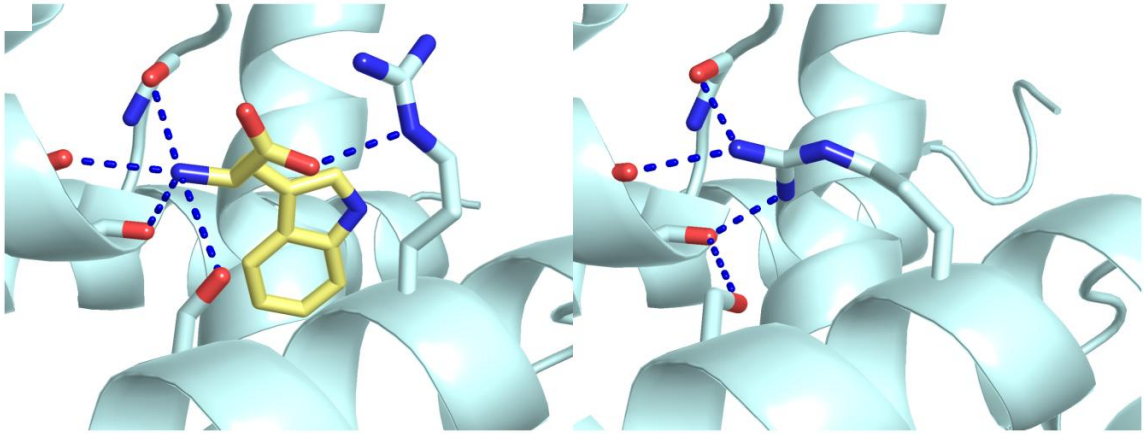
結合状態　　非結合状態

結晶水
284



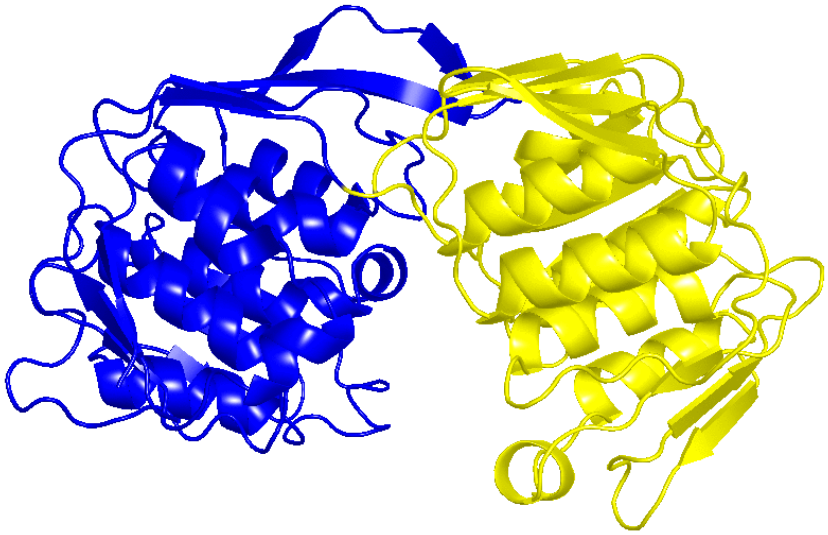
10-Formyltertrahydrofolate dehydrogenase

側鎖
177

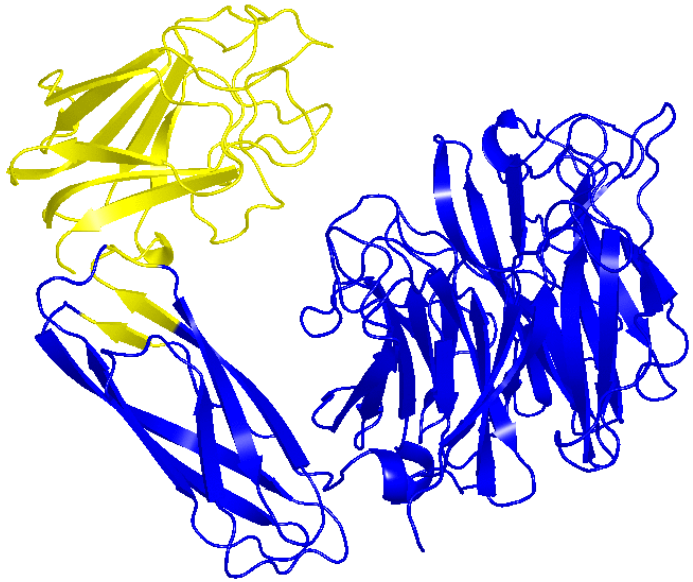


Trp operon repressor

リガンド結合との因果関係



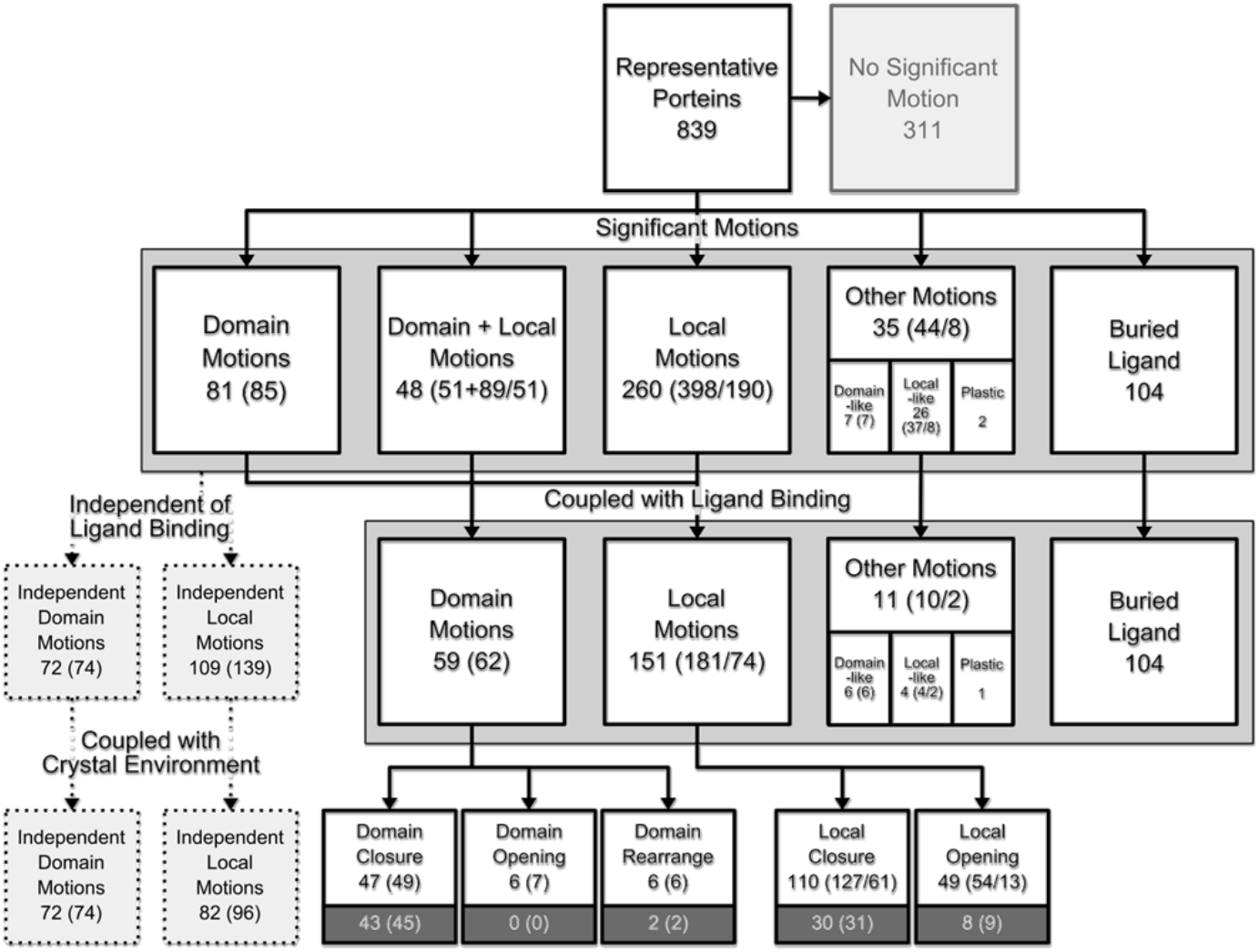
原因がリガンド結合である
221



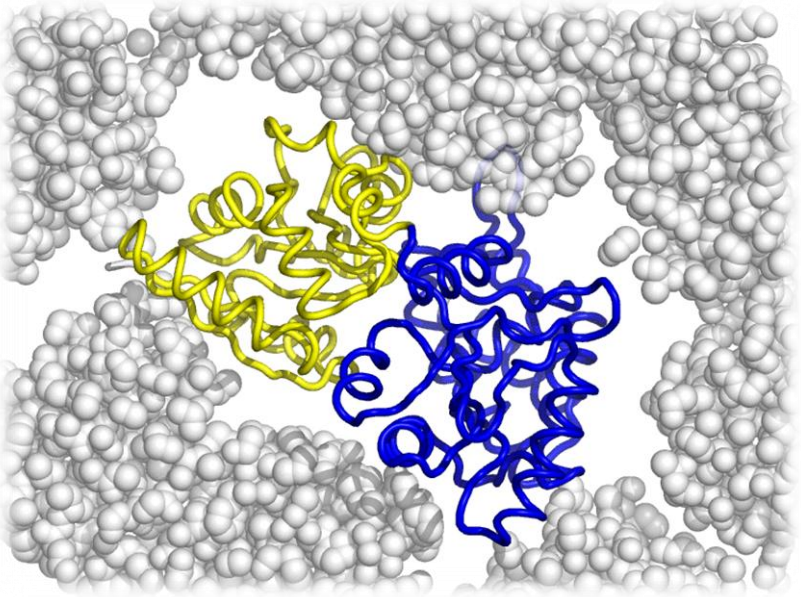
原因がリガンド結合でない
181

other than buried ligand

Classification



結晶場の影響



結晶場によってひずみを受けている

網羅的シミュレーション

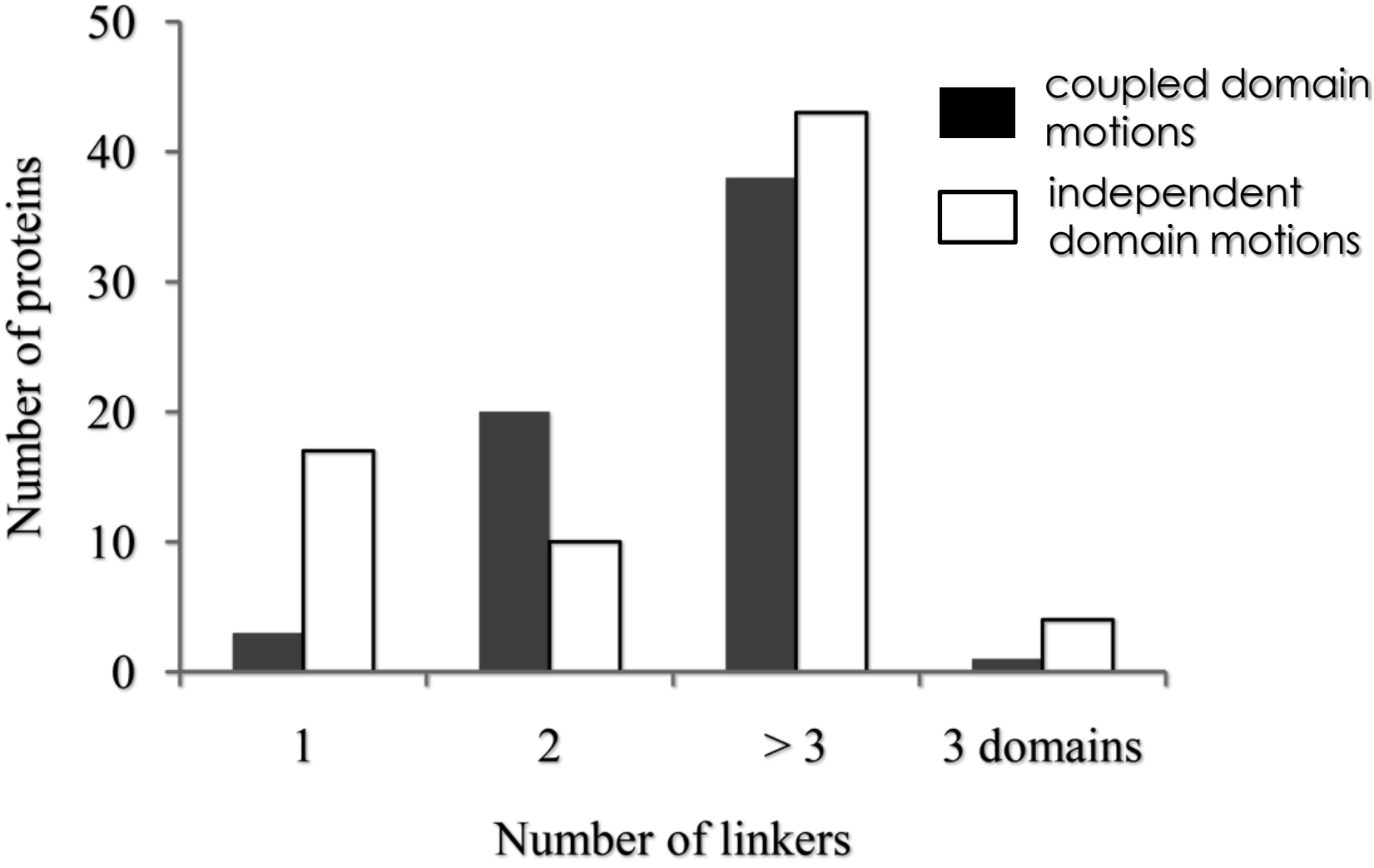
PDB ID1	PDB ID2	RMSDs from initial structures at 10 ns [Å]		Pair-wise RMSD [Å]	
		Protein 1	Protein 2	Initial	Minimum (time [ns])
1n0u_A	1n0v_D	38.6	28.1	34.7	18.6 (5.2/0.4)
2qyu_A	2qza_A	30.6	11.4	6.1	9.0 (0.1/0.3)
1eut_A	1w8n_A	23.1	14.8	2.7	2.0 (0.1/0.1)
2zgv_A	3c3b_A	3.6	2.3	7.0	1.8 (1.1/6.0)
1e4f_T	1e4g_T	1.6	3.3	2.7	1.7 (5.2/5.4)
2v8i_A	2v8j_A	1.9	1.1	2.1	1.0 (7.4/3.3)
2ps0_A	2ps3_A	3.2	2.6	2.0	1.6 (7.9/3.7)
3b8s_B	3b9d_A	5.4	3.0	2.4	1.3 (2.6/1.3)
2ddb_A	2epf_D	3.3	4.9	1.8	1.2 (0.7/0.9)
2o1c_A	2o1c_D	2.0	2.1	1.5	0.8 (2.6/1.1)
1rif_A	1rif_B	4.7	2.9	1.1	1.3 (0.7/0.2)

Single linker

Multiple linker

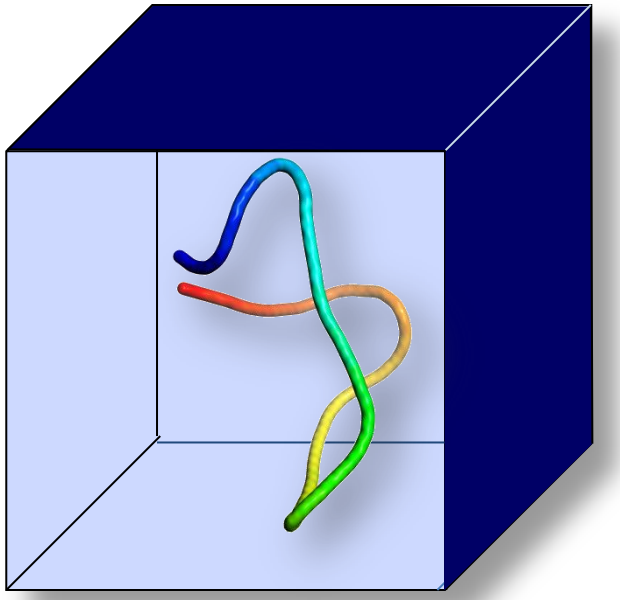
寺田透（東京大学、理研）

Correlation with the number of linkers

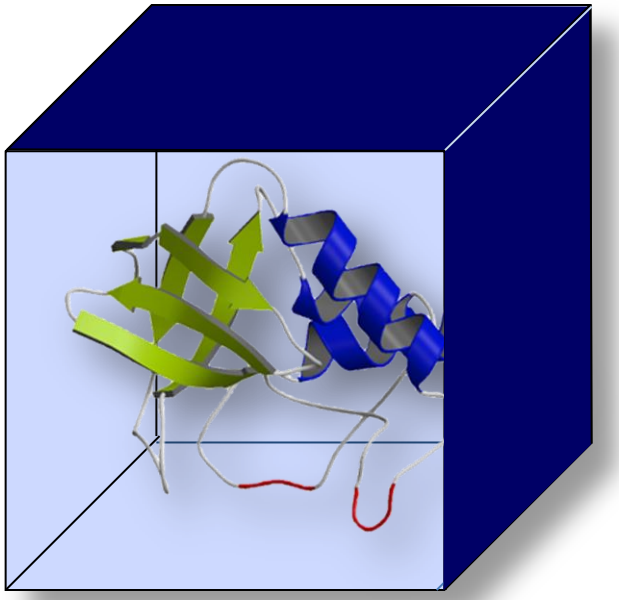


1．このように動く
1－2．応答的タンパク質機能像　予測可能性

Folding

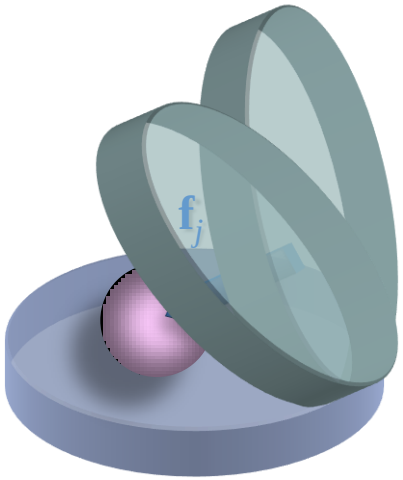


Function



機能シミュレーション
＝　応答的非平衡シミュレーション

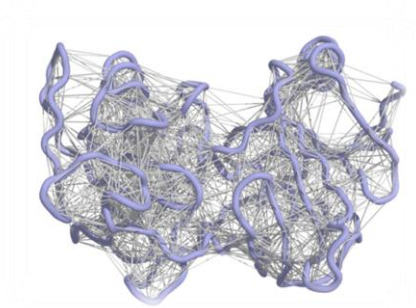
線形応答理論



$$\langle \Delta \mathbf{r}_i \rangle_1 \simeq \frac{1}{k_B T} \langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j \rangle_0 \mathbf{f}_j$$

果　　因

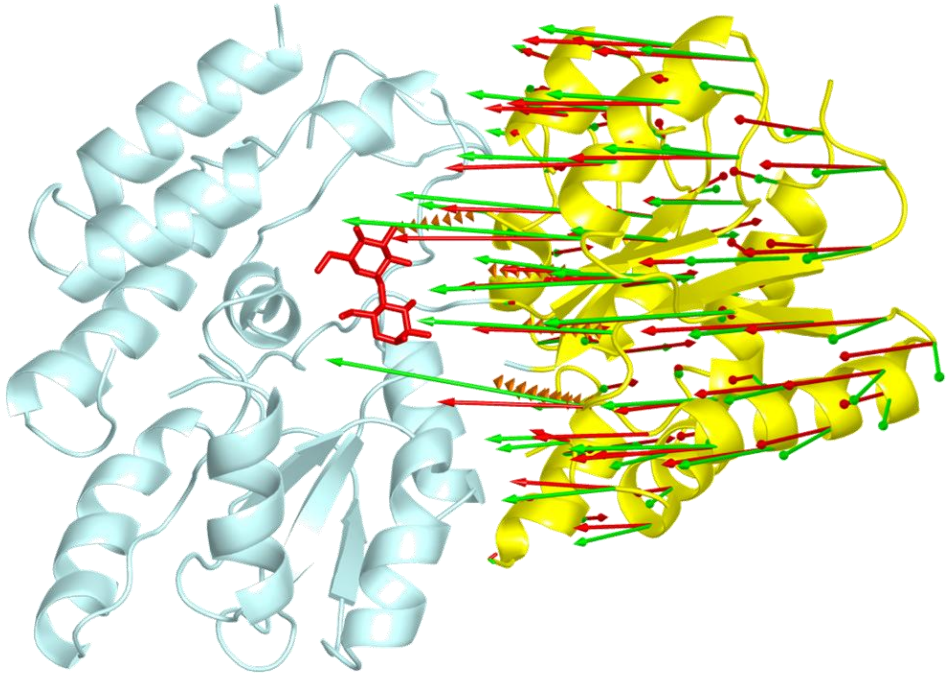
Molecular Dynamics Simulation



Elastic Network Model

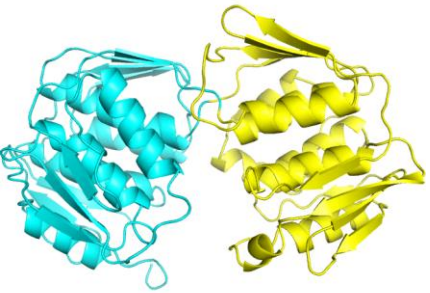
Linear Response Theory of Protein Structural Changes

$$\Delta \mathbf{r} = \beta \boldsymbol{\Sigma} \mathbf{f}$$

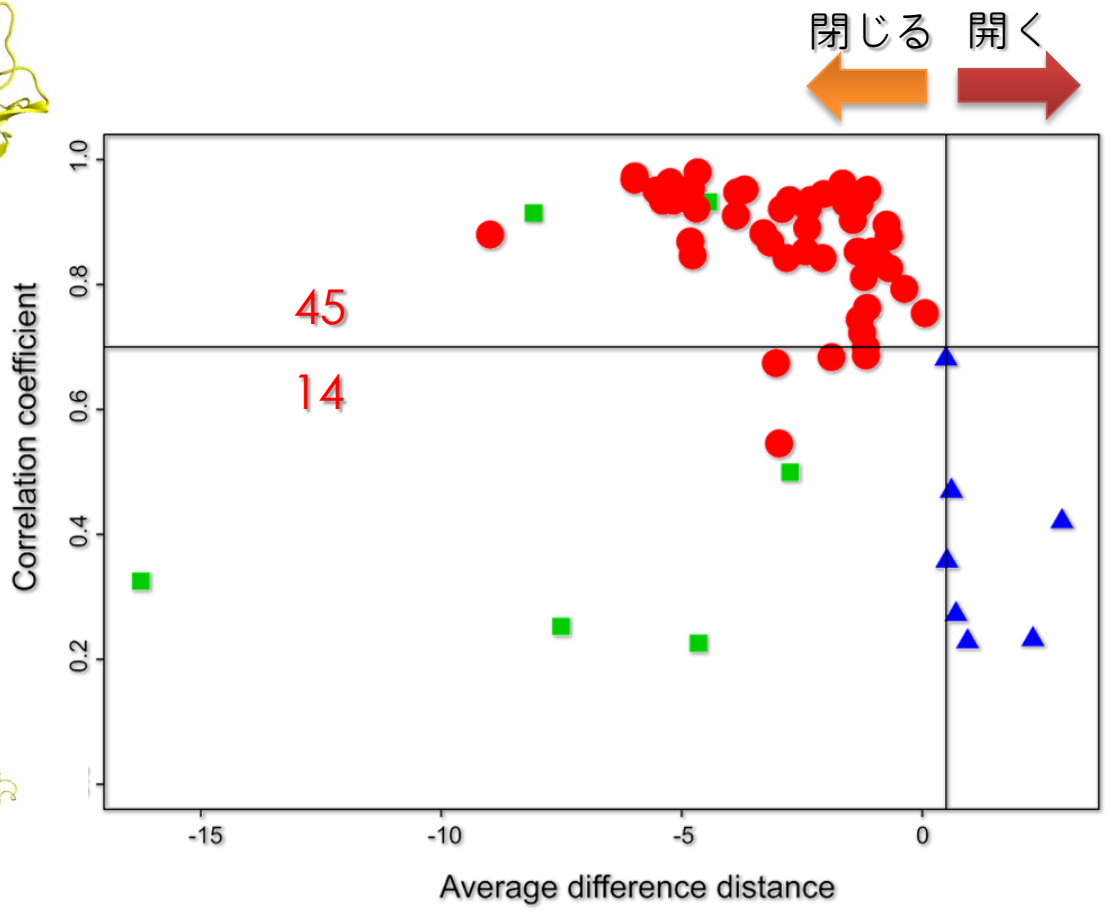


Maltose-binding periplasmic protein

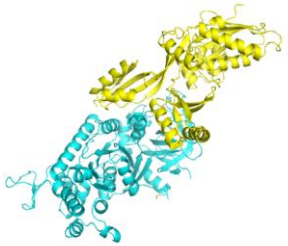
ドメイン運動：線形応答理論の予測



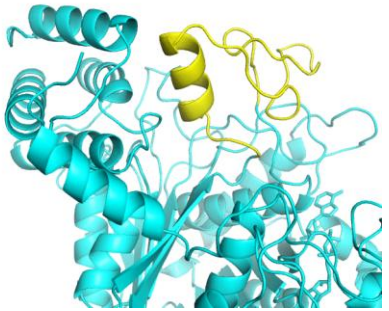
閉運動



閉運動：●
開運動：▲
再配置運動: ■

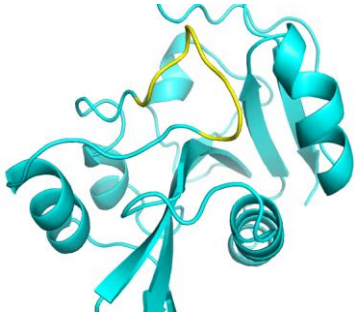


再配置運動

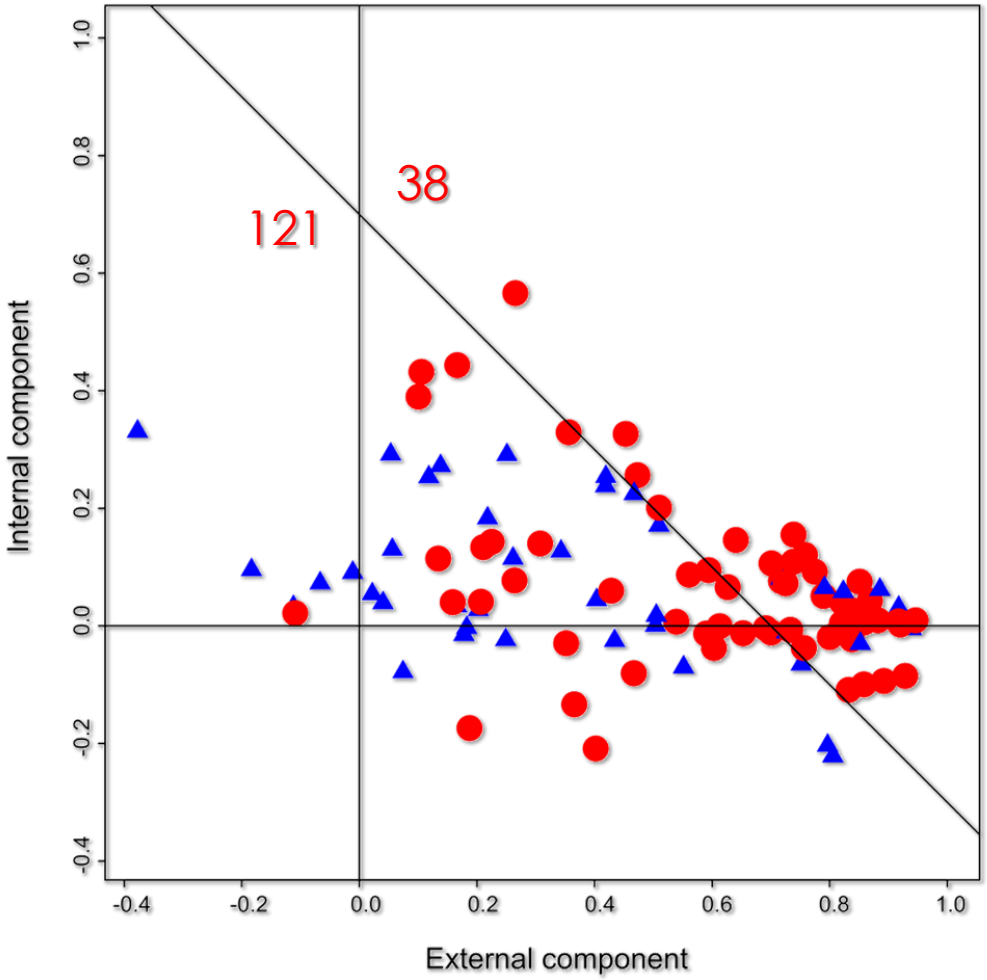


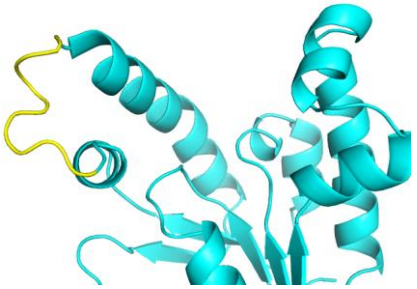
開運動

ローカル運動：線形応答理論の予測



開運動



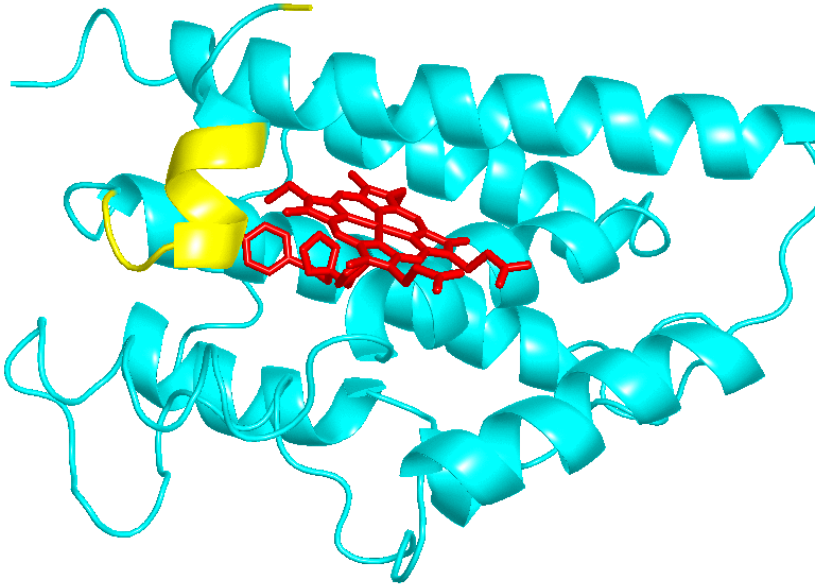


閉運動

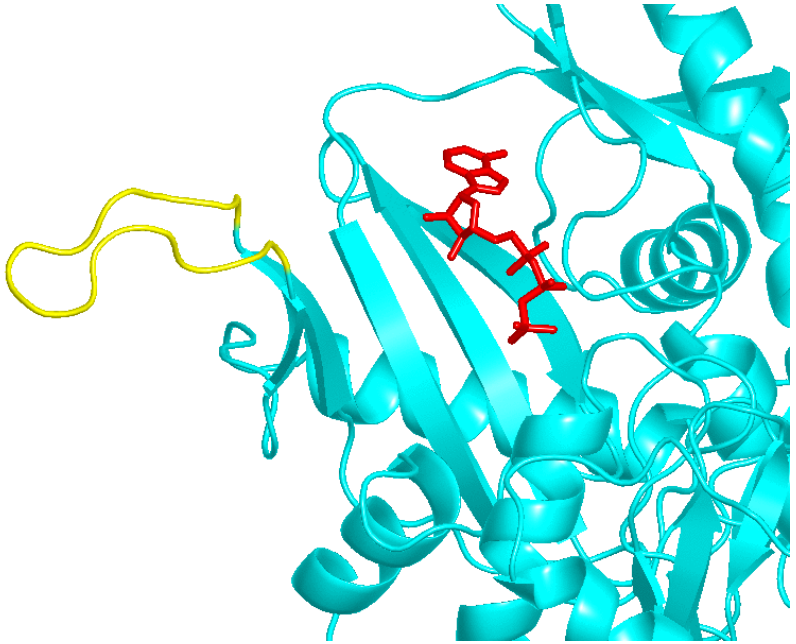
閉運動：●
開運動：▲

ローカル運動の多くは非線形的

Order-disorder transition

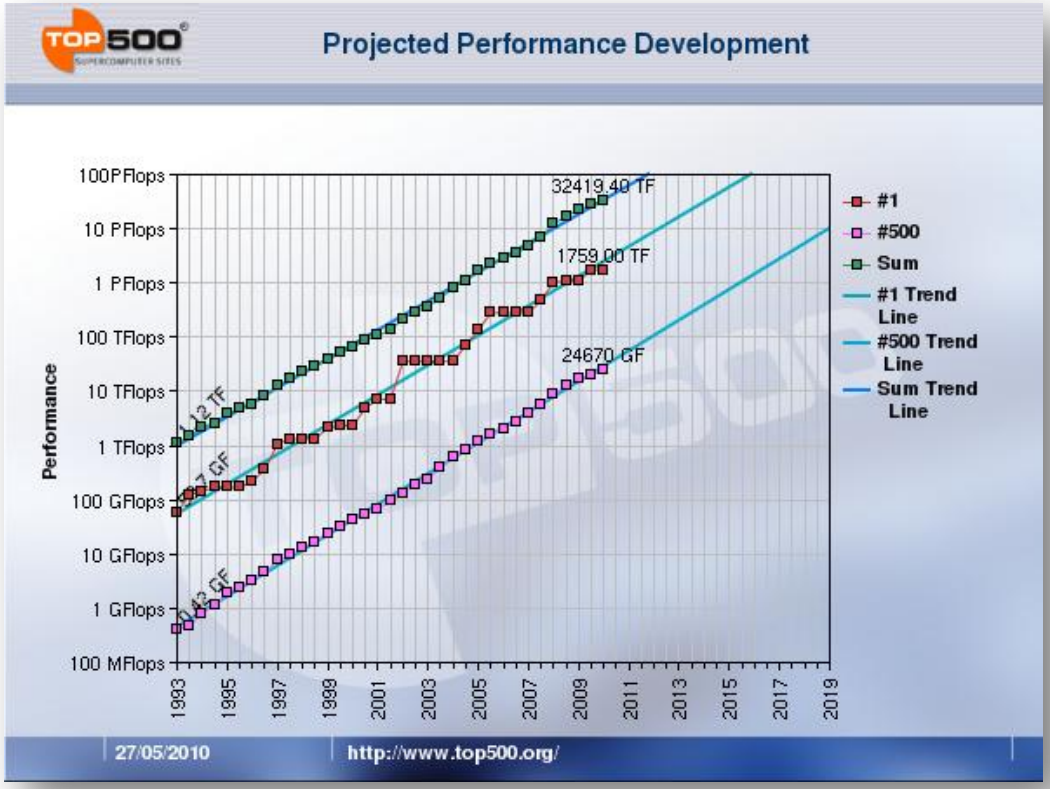


Helix-coil transition



リガンド結合に伴うローカル運動の約60%に見られた

2．タンパク質機能の分子シミュレーション
2−1．もっと速さを！



Molecular dynamics simulations for biomolecules

Nature Vol. 267 16 June 1977

articles

Dynamics of folded proteins

J. Andrew McCammon, Bruce R. Gelin & Martin Karplus

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

The dynamics of a folded globular protein (bovine pancreatic trypsin inhibitor) have been studied by solving the equations of motion for the atoms with an empirical potential energy function. The results provide the magnitude, correlations and decay of fluctuations about the average structure. These suggest that the protein interior is fluid-like in that the local atom motions have a diffusional character.

number of interactions which must be calculated and also permits larger steps in the trajectory calculation since the high frequency hydrogen vibrations have been eliminated. Integration of the equations of motion was performed by means of the Gear algorithm[22] with time steps of 9.78×10^{-16} s. X-ray coordinates[21] were used for the initial positions and the initial velocities were set equal to zero. After 100 equilibration steps, the stresses in the initial structure had partly relaxed and the system had an internal kinetic energy corresponding to a temperature of 140 K. At this point, all velocities were multi-

Bovine pancreatic trypsin inhibitor (BPTI) 454 non-hydrogen atoms In vacuum

~8.8 ps

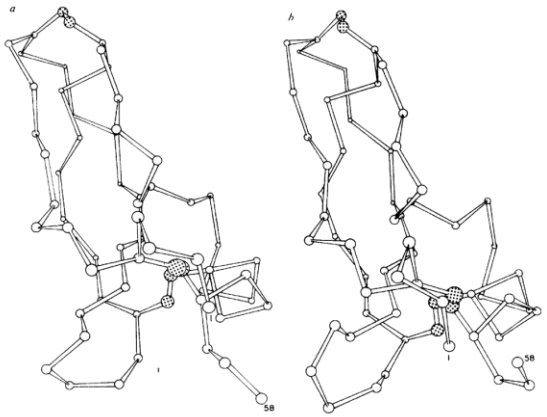
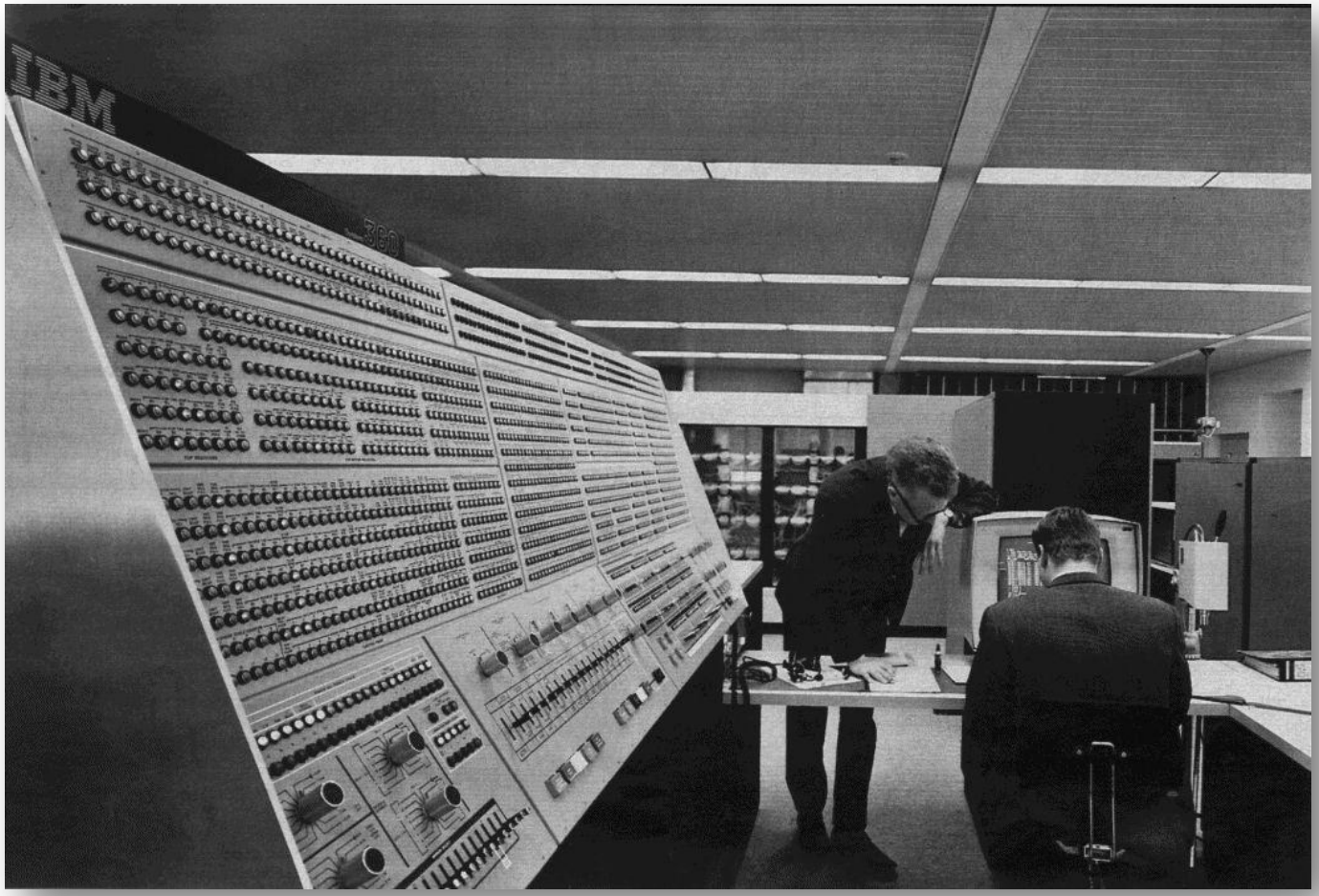
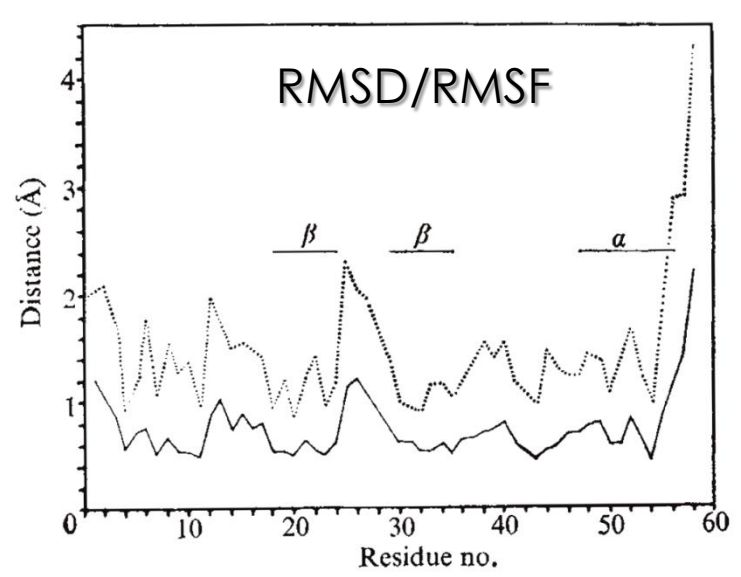


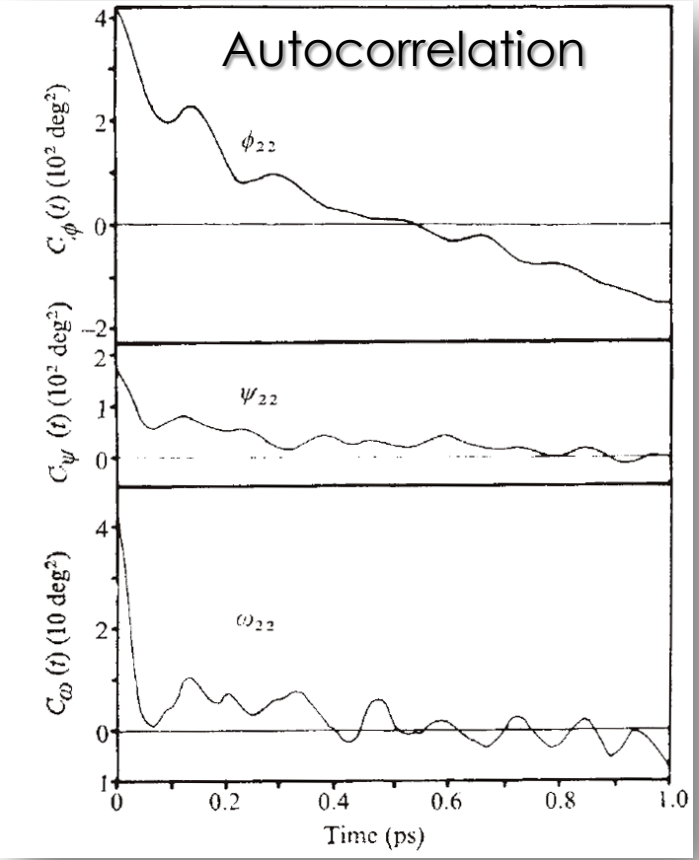
Fig. 1 The peptide backbone (α carbons) and disulphide bonds of PTI. *a*, X-ray structure[21]. *b*, Time evolved structure after 3.2 ps of dynamical simulation.

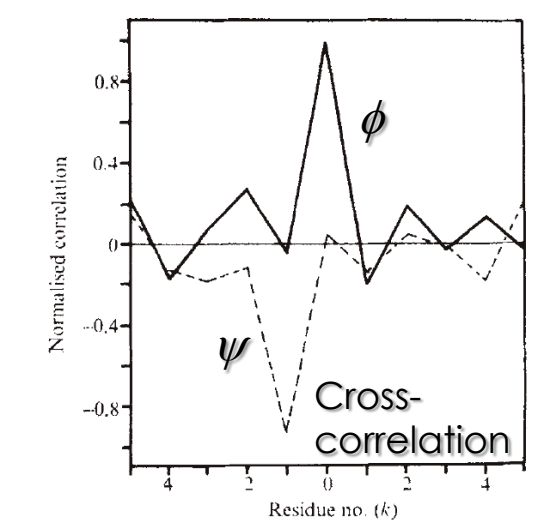


IBM 360/91 : ~1MFLOPS

Analyses of Trajectory Data

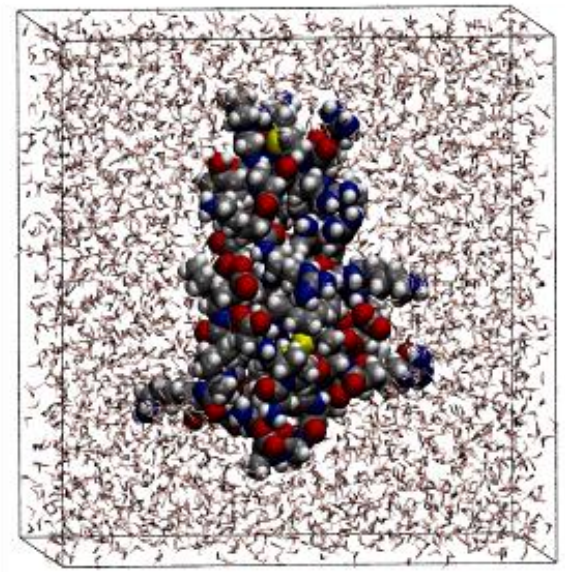






Protein's vibrational dynamics

BPTI



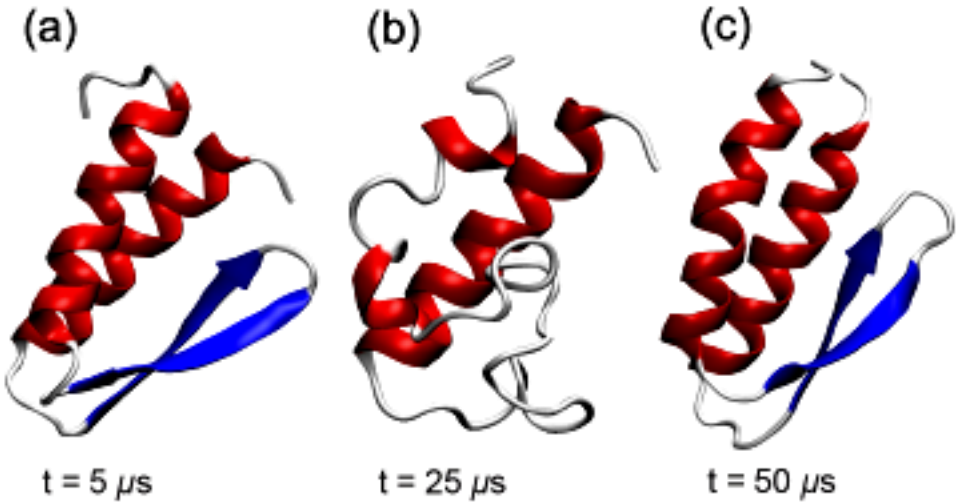
Length (μs)	Protein	Hardware	Software	Citation
1031	**BPTI**	**Anton**	**[native]**	**Here**
236	**gpW**	**Anton**	**[native]**	**Here**
10	WW domain	x86 cluster	NAMD	[10]
2	villin HP-35	x86	GROMACS	[6]
2	rhodopsin	Blue Gene/L	Blue Matter	[25]
2	rhodopsin	Blue Gene/L	Blue Matter	[12]
2	β_2AR	x86 cluster	Desmond	[5]

Table 1: The longest (to our knowledge) published all-atom MD simulations of proteins in explicitly represented water.

<u>> 1ms</u>

~10,000 atoms including explicit water

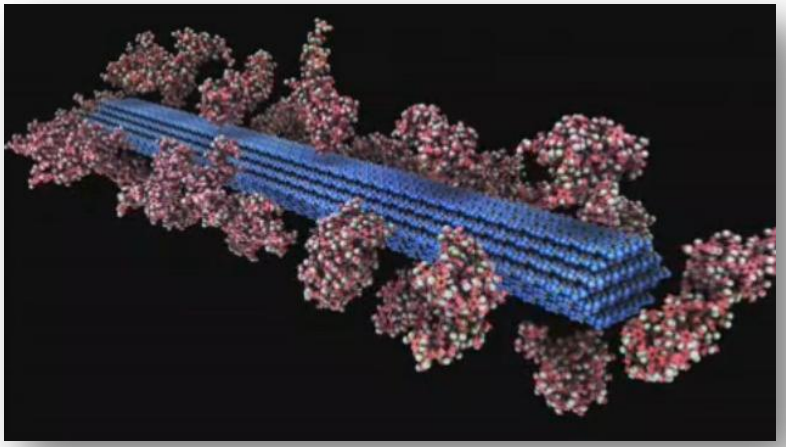
10^8 time longer simulation



Bacteriophage lambda protein W (gpW)
58 residues

D. E. Shaw et al. Science 330, 341 (2010)

Lignocellulosic Biomass



A system was constructed of lignocellulosic biomass containing 52 lignin molecules each with 61 monomers, the same cellulose fibril as described above and 1,037,585 TIP3P water molecules, totaling 3,316,463 (or 3.3 million) atoms.

7,300 times larger system
(3 x 10^7 times more interactions)

JaguarPF, a Cray XT5 massively parallel processing computer with 30,000 Opteron 2.3 GHz cores
200TFLOPS
2x10^8 times larger power

R Schulz, et al. *J. Chem. Theory Comput* 5, 2798 (2009)

33 years: 1977 → 2010

Quantitative Measure

Computer Power 2×10^8

Target Systems

Size (Interaction) 3×10^7

Time Scale 1×10^8

Qualitative Measure

Physics: Vibrational Dynamics

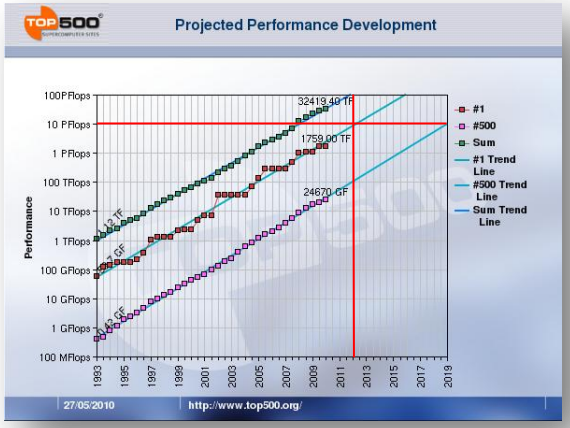
↓

Biology: Biomolecular Functions

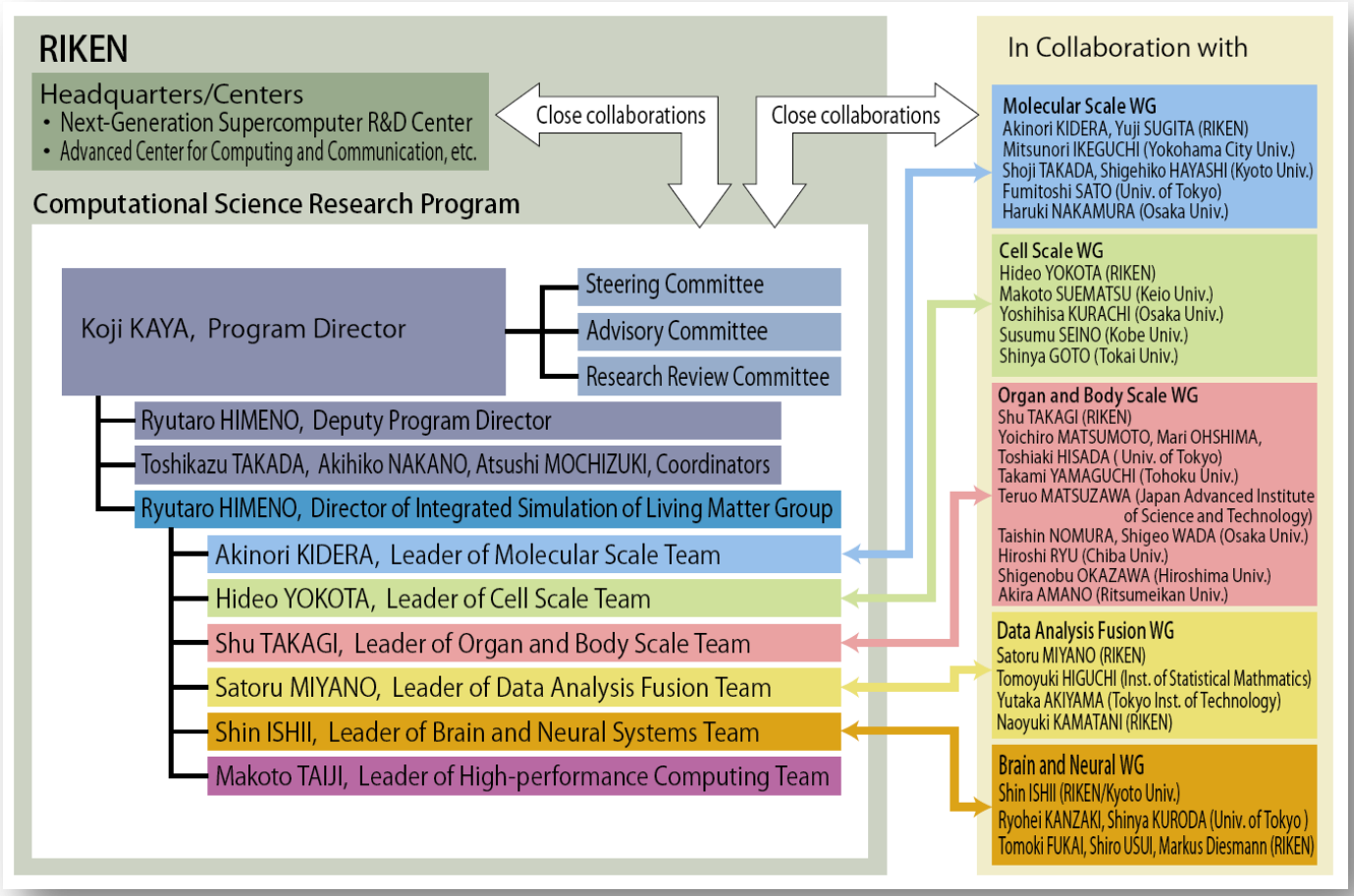
Real Simulations

Next-Generation Supercomputer 「京」

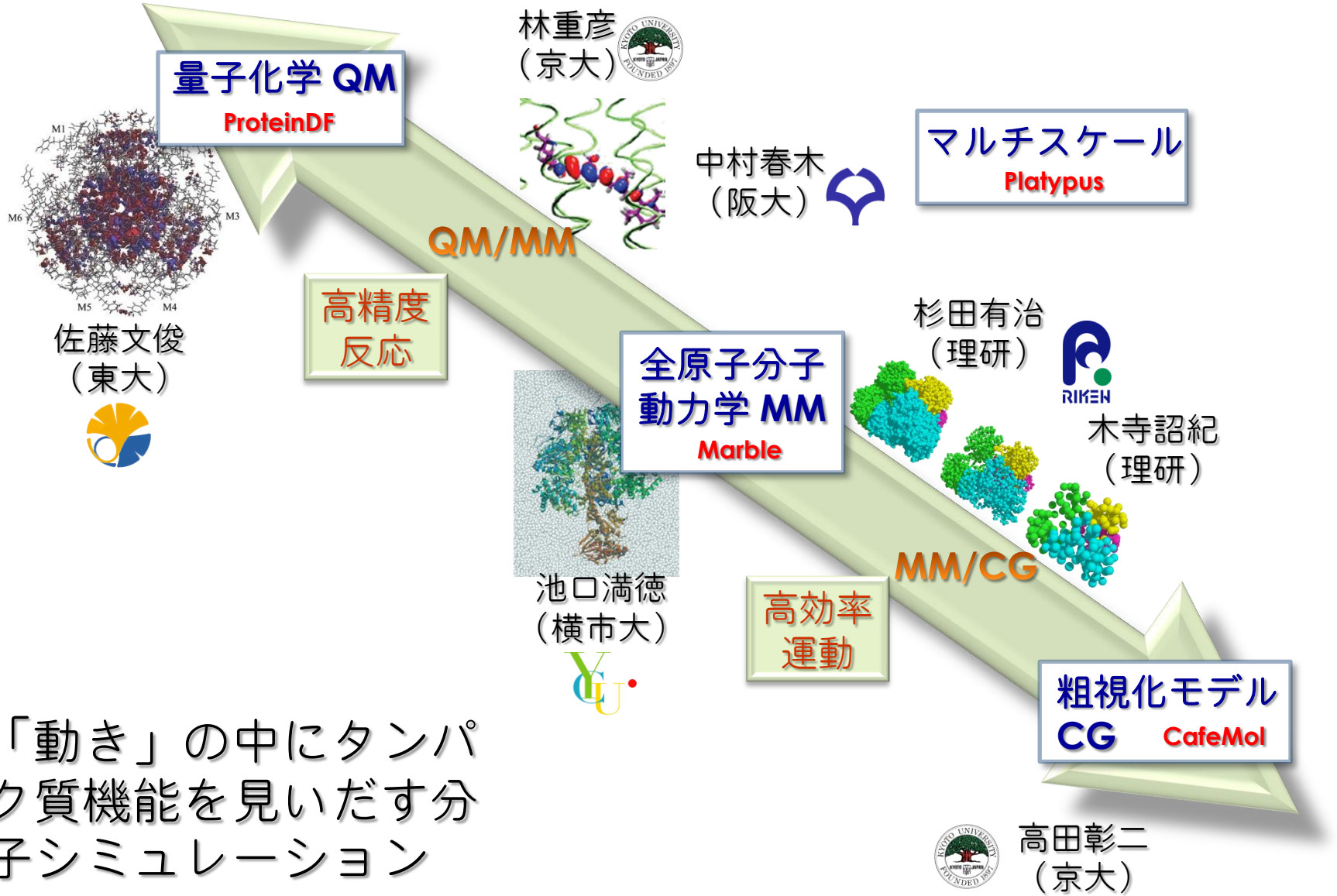
10 PFLOPS
6.4 x 10^5 cores
1PB memory



Next Generation Integrated Simulation of Living Matter



2 タンパク質機能の分子シミュレーション
2−2 次世代生命体統合シミュレーション 分子スケール

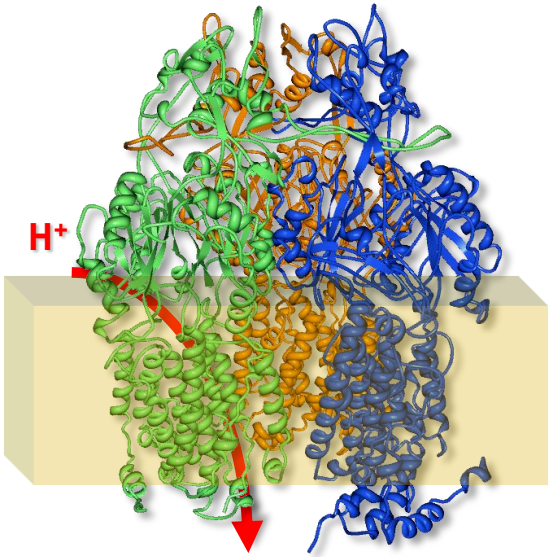


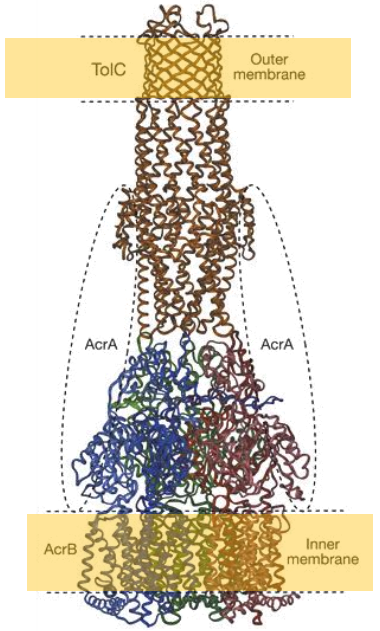
「動き」の中にタンパク質機能を見いだす分子シミュレーション

(1) 多剤排出トランスポーターを例に
Multidrug Efflux Transporter AcrB

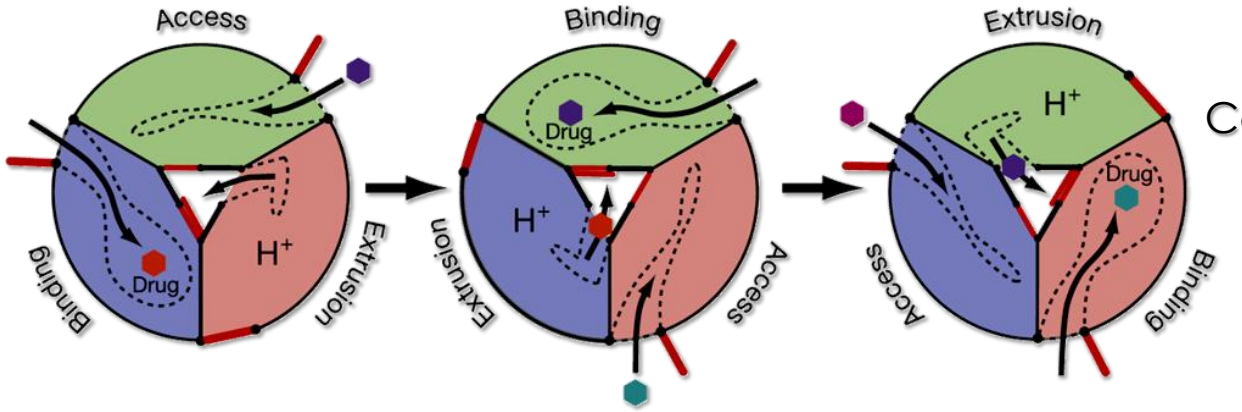
Multidrug resistance:
A serious problem in the chemotherapy and in the antibiotic treatment

AcrB is a principal multidrug exporter.





Functionally rotating mechanism



Proton Transfer
(QM)

Cooperative Structural Transitions
(MM-MM/CG)

Drug Transport
(MM/CG-CG)

S Murakami S, et al. Nature 443, 173 (2006)

Proton transfer → Structure change

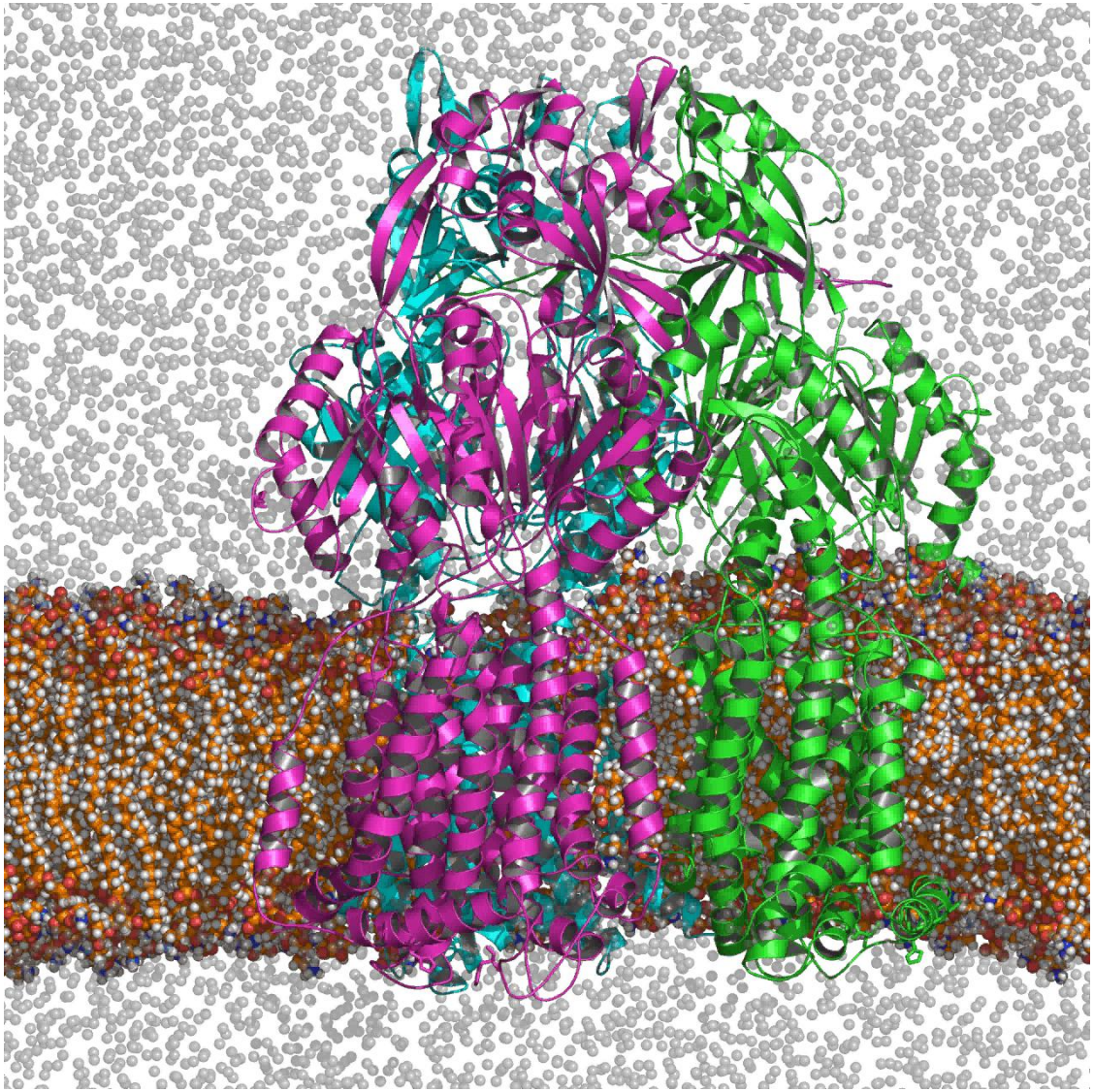
a. MM Protonation states & Gate opening
All atom molecular dynamics simulation
T. Yamane, M. Ikeguchi (YCU)

Structure change → Drug transport

b. MM/CG Ligand binding sites
3D-RISM calculation
T. Imai, A. Kidera (Riken), F. Hirata (IMS)

c. CG Cooperative motions
Triple Go model simulation
X. –Q. Yao, S. Takada (Kyoto U)

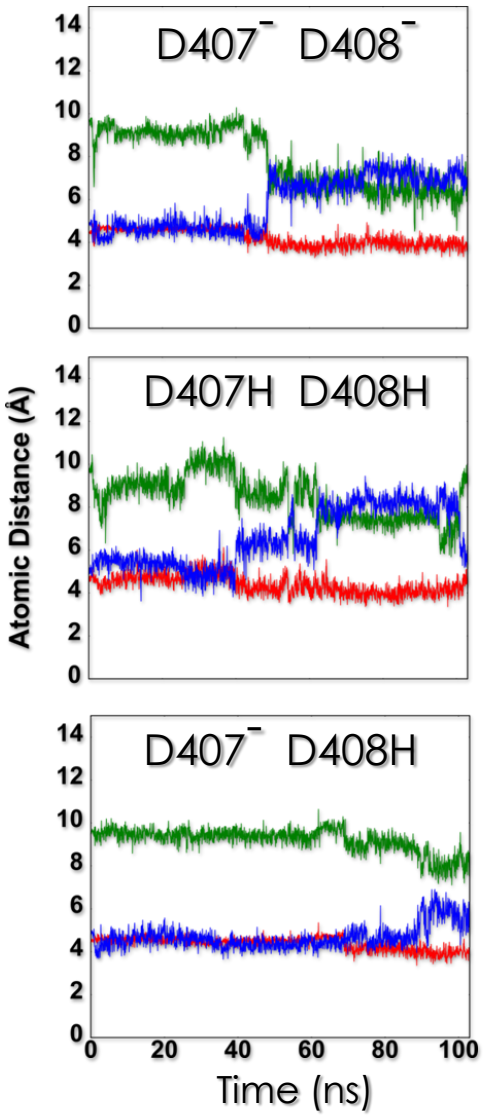
a. MM All atom molecular dynamics simulation

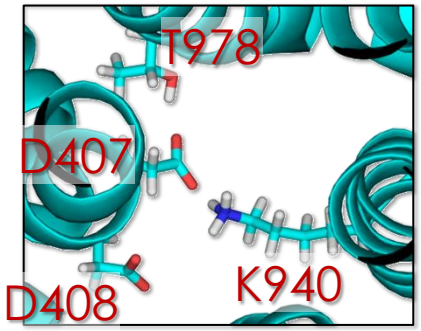


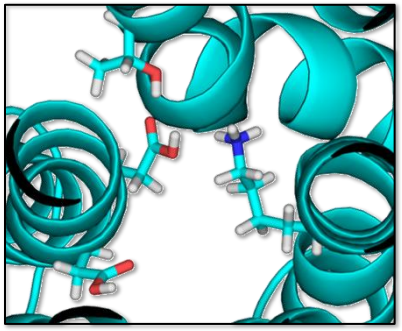
T. Yamane, M. Ikeguchi (YCU)

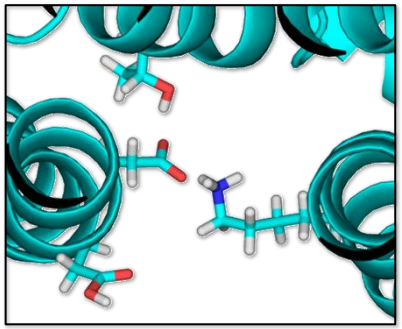
MM Extrusion state

D407Cβ – K940Nζ
D408Cβ – K940Nζ
T978Cβ – K940Nζ









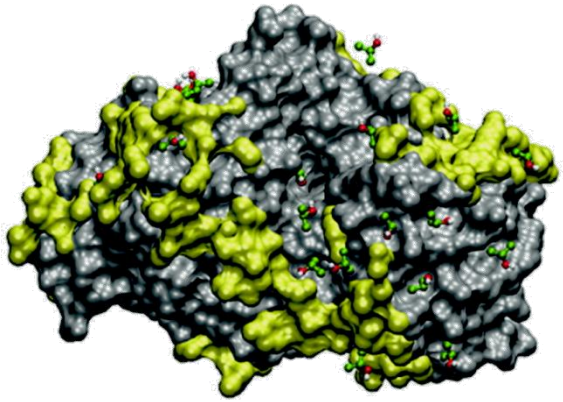
b. MM/CG 3D-RISM Calculation

3D-RISM/KH equations

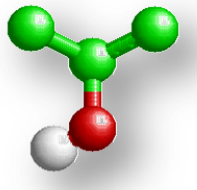
$$h_\gamma(\mathbf{r}) = \sum_\gamma \int c_\gamma(\mathbf{r}')\left(w^{vv}_{\gamma'\gamma}(|\mathbf{r}'-\mathbf{r}|) + \rho_\gamma h^{vv}_{\gamma'\gamma}(|\mathbf{r}'-\mathbf{r}|)\right) d\mathbf{r}'$$

$$h_\gamma(\mathbf{r}) = \begin{cases} \exp\left(d_\gamma(\mathbf{r})\right) - 1 & \text{for } d_\gamma(\mathbf{r}) \le 0 \\ d_\gamma(\mathbf{r}) & \text{for } d_\gamma(\mathbf{r}) > 0 \end{cases}$$

where $d_\gamma(\mathbf{r}) = -\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})$

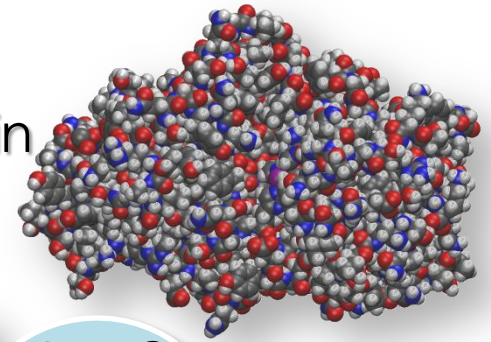


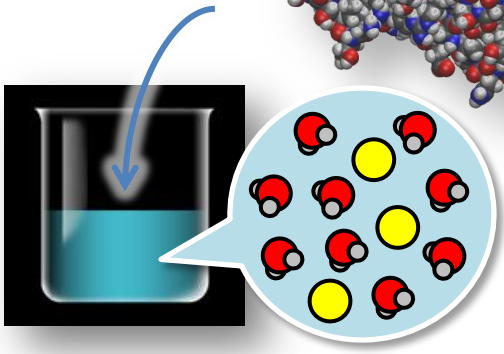
Theoretical prediction



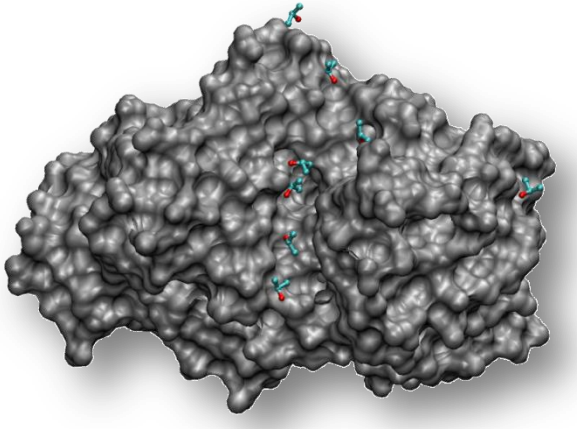
Isopopanol
Acetone
Acetonitrile
Phenol

Thermolysin





Isopopanol
+ water

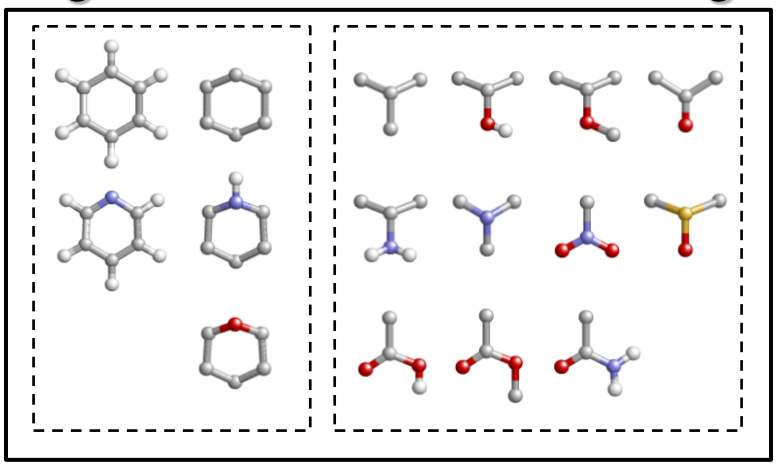


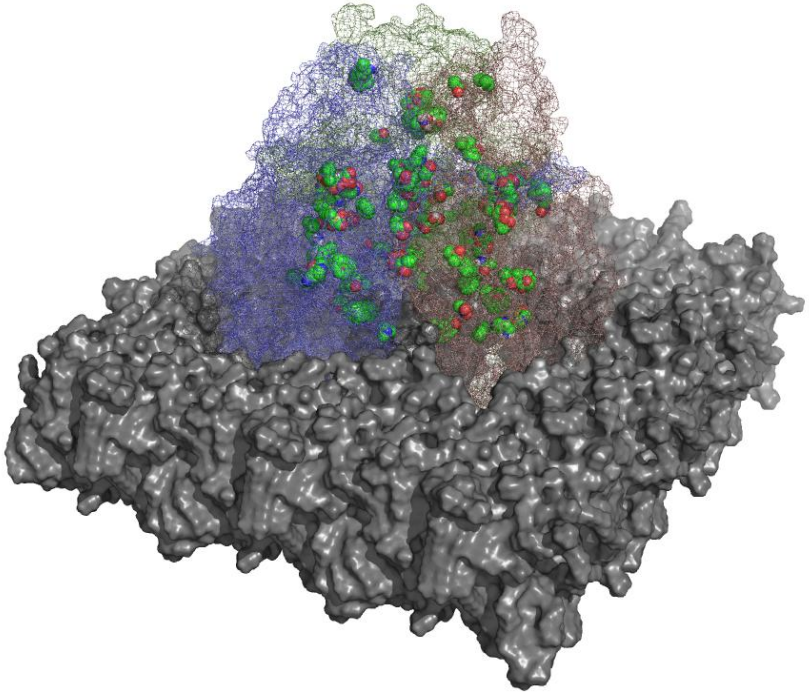
Crystal structure

T. Imai, et al. *J. Am. Chem. Soc.* **131**, 12430 (2009)

Appllication to AcrB

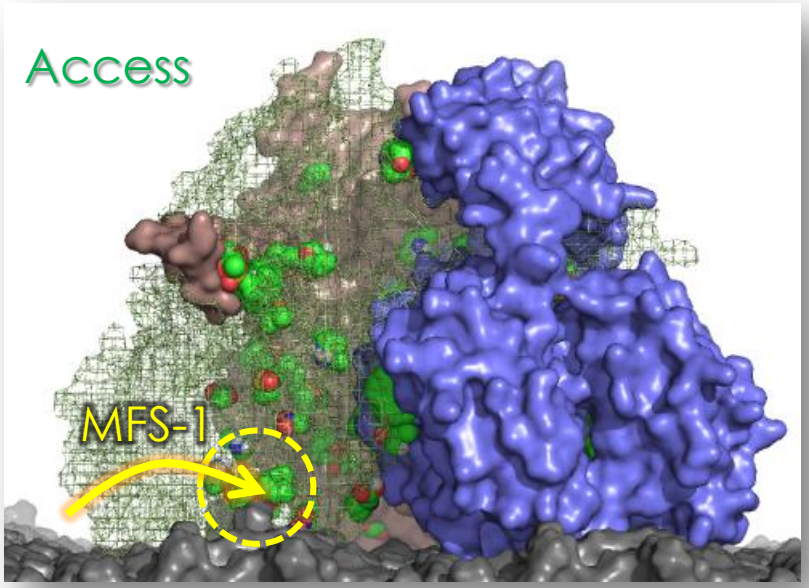
Fragment Molecules Used as Ligands

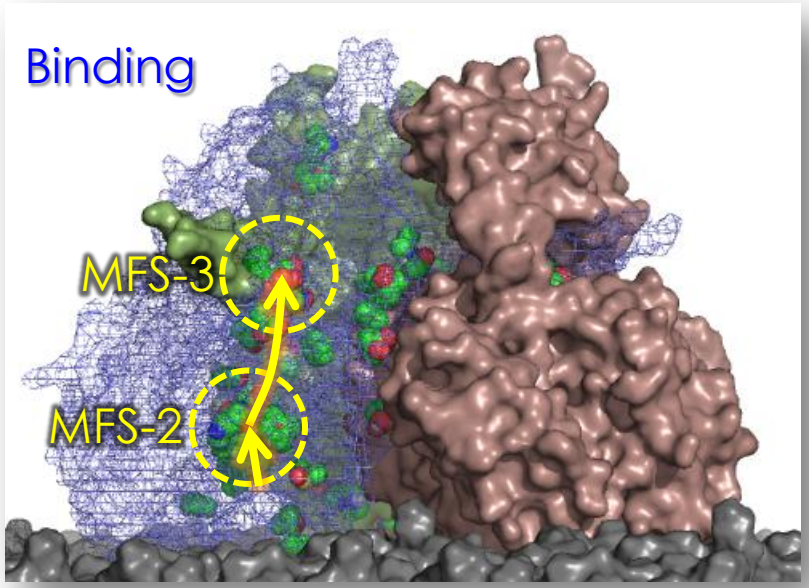


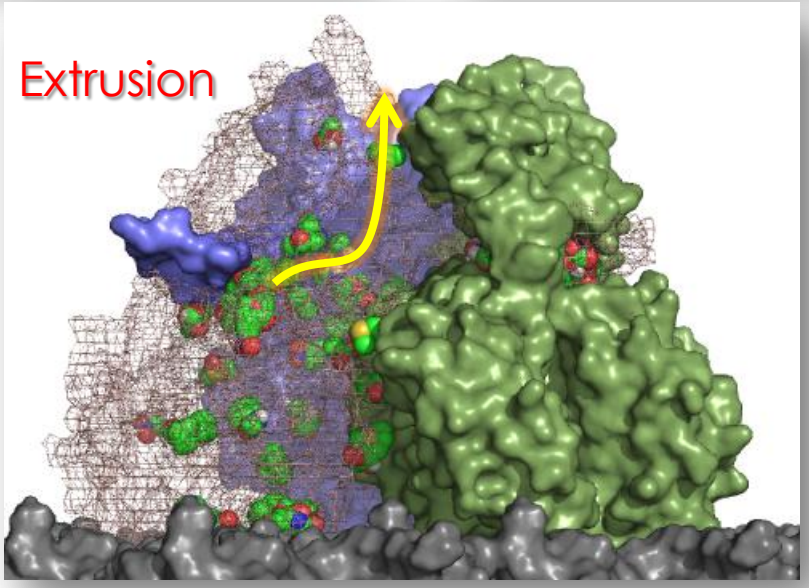


T. Imai, *et al.*, *J.Phys.Chem B* in press

MM/CG "Multi-functional" Sites







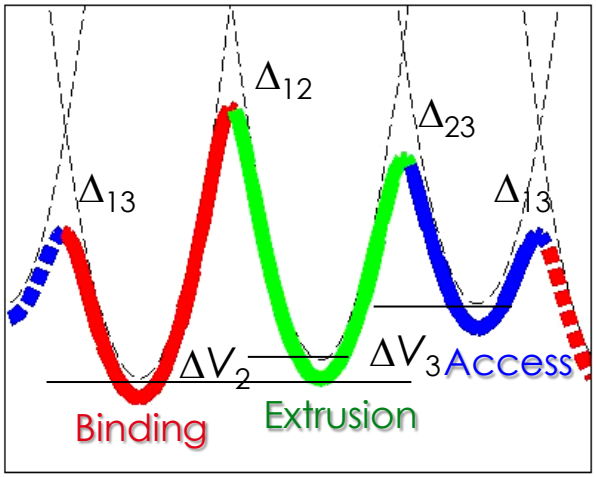
c. CG Triple Go-model Simulation

C_α CG model of Porter domain

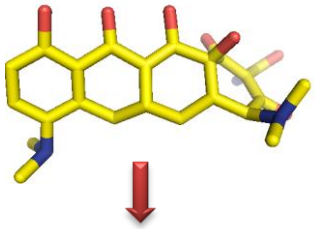
Multiple-basin model

Okazaki, et al. *PNAS* **103,** 11844 (*2006*)

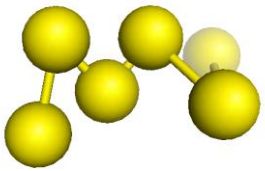
Trimer interface
Three types of multiple LJ type



Ligand

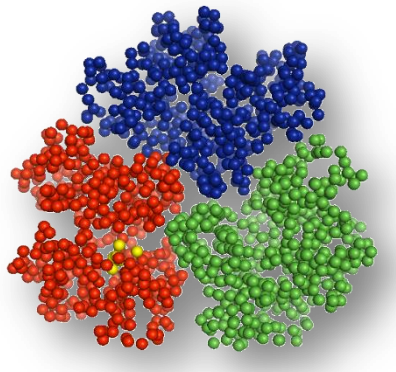


Minocycline

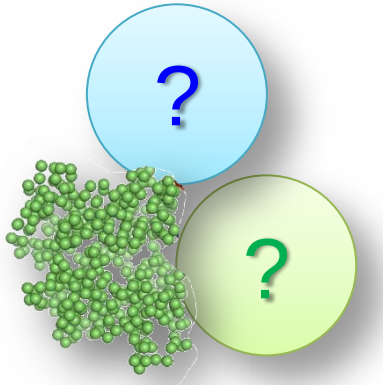


Repulsive potential

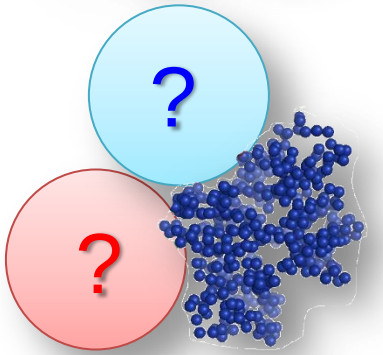
Xin-Qiu Yao, *et al. Nature Commun.* **1**, 117 (2010)



B→E



E→A



A→B

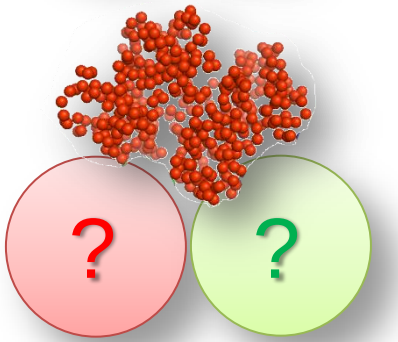
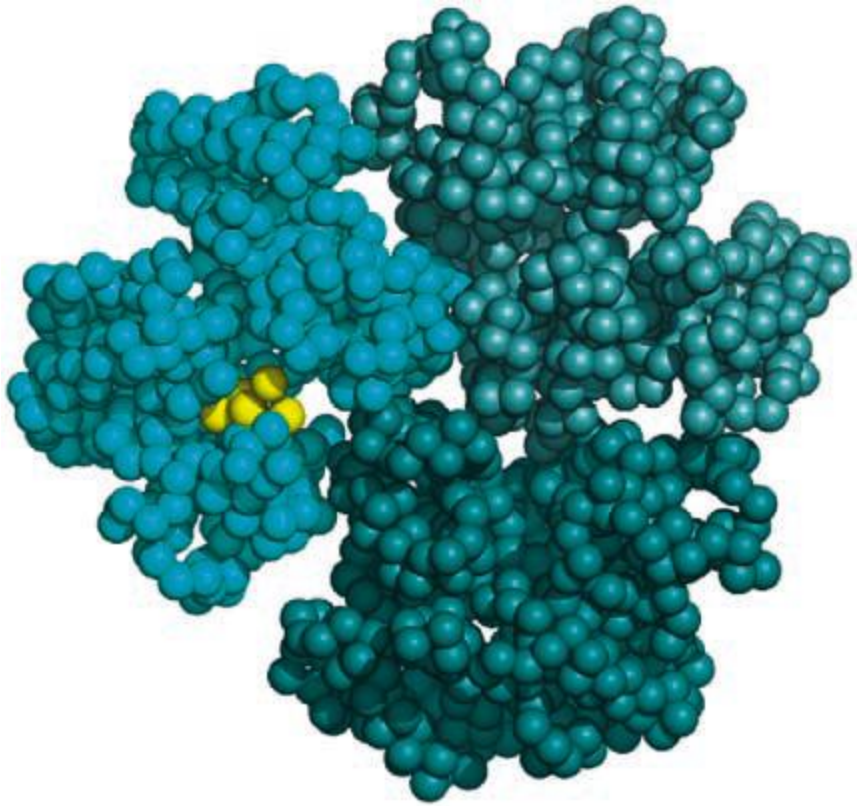


Table 1 | Structural characteristics of the three states.

	Entrance	Cleft	Binding pocket	Exit
A/access	O	O	X	X
B/binding	O	O	O	X
E/extrusion	X	X	X	O

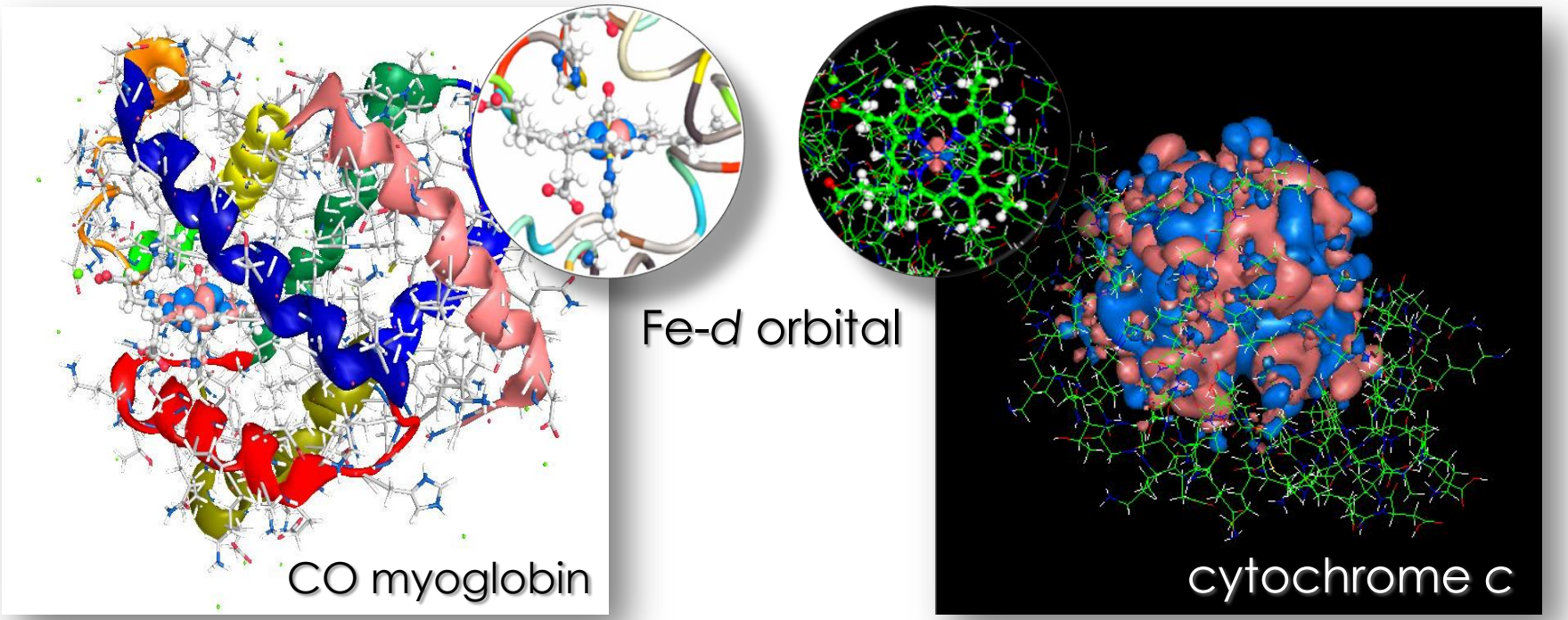
O, open; X, closed.

B → E



(2) 量子化学計算
Have you seen MO of protein?

HOMO



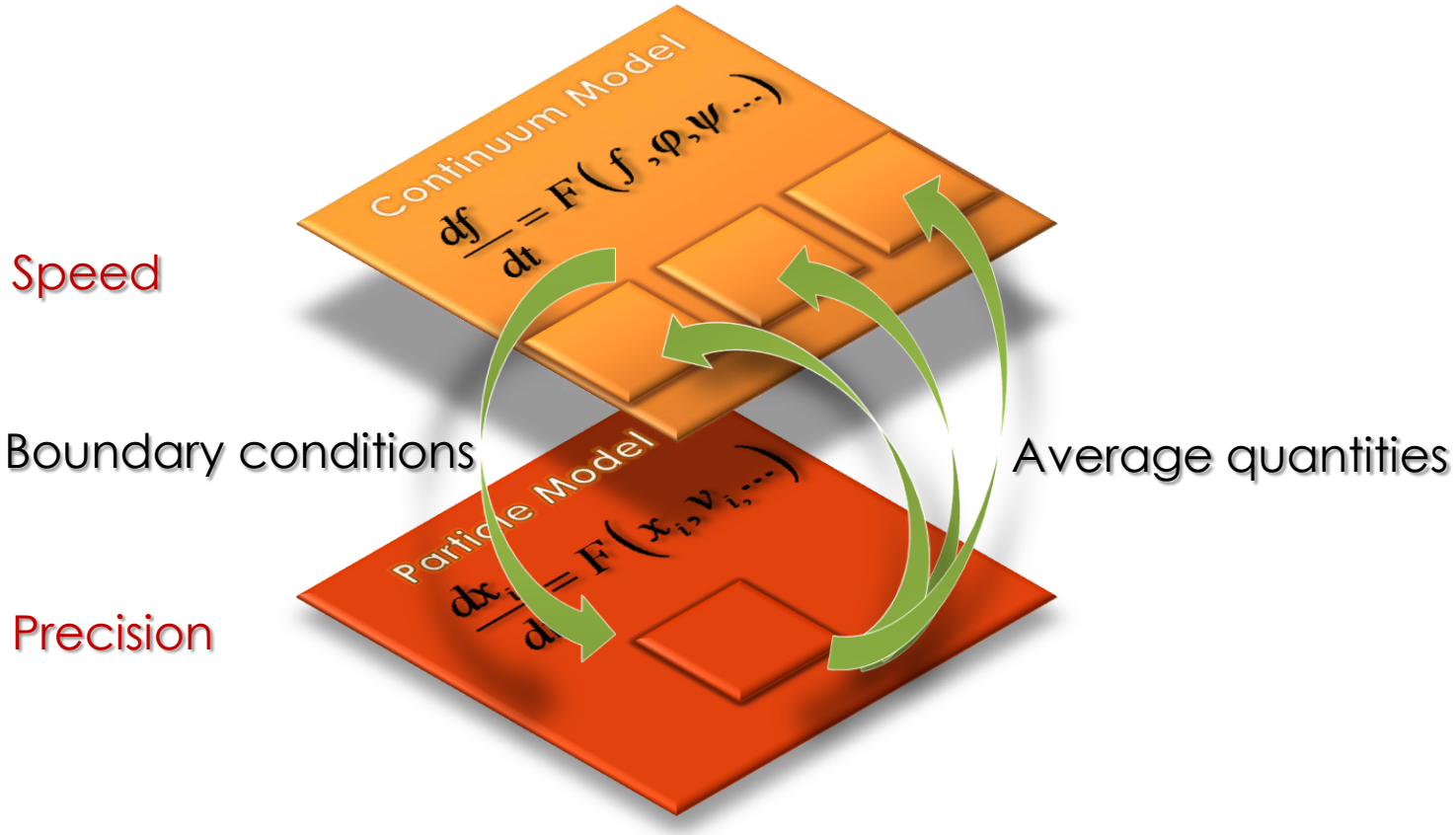
localized　　delocalized

All-electron calculations (not divided) using ProteinDF

佐藤文俊（東大)

(3) マルチスケールシミュレーション

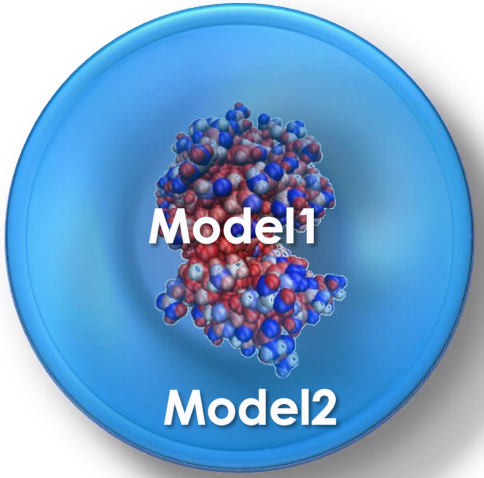
Coupled Simulation



Homogeneous Systems

Coupled Simulation in Biomolecules

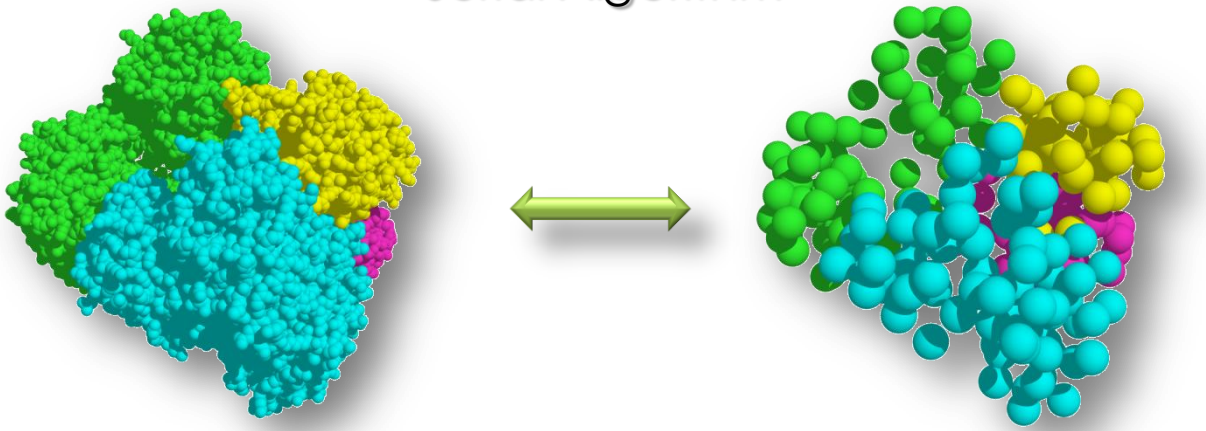
Parallel Algorithm



QM/MM

MM/CG

N Serial Algorithm M



$N > M$

GS Ayton, WG Noid, GA Voth, Current Opinion in Structural Biology 17, 192 (2007)

QM/MM

Coupling between Enzymatic Reaction and Protein Motions

1. On-the-fly QM/MM MD simulation - S. Hayashi (Kyoto U)

$$H_{\mathrm{QM/MM}}\left(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)$$

2. QM/MM umbrella sampling - H. Nakamura (Osaka U)

$$H_{\mathrm{QM/MM}}\left(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)$$

$$\mathbf{r}_{\mathrm{QM}}=\lambda\mathbf{r}_{\mathrm{QM}}^{0}+\left(1-\lambda\right)\mathbf{r}_{\mathrm{QM}}^{1}$$

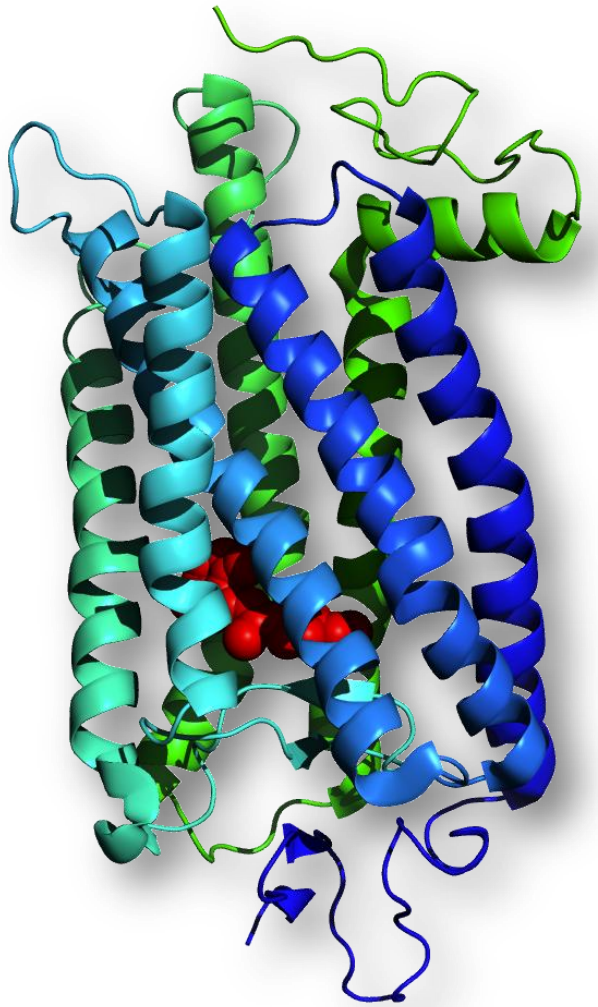
Nonequilibrium MM

3. QM/MM free energy - S. Hayashi (Kyoto U)

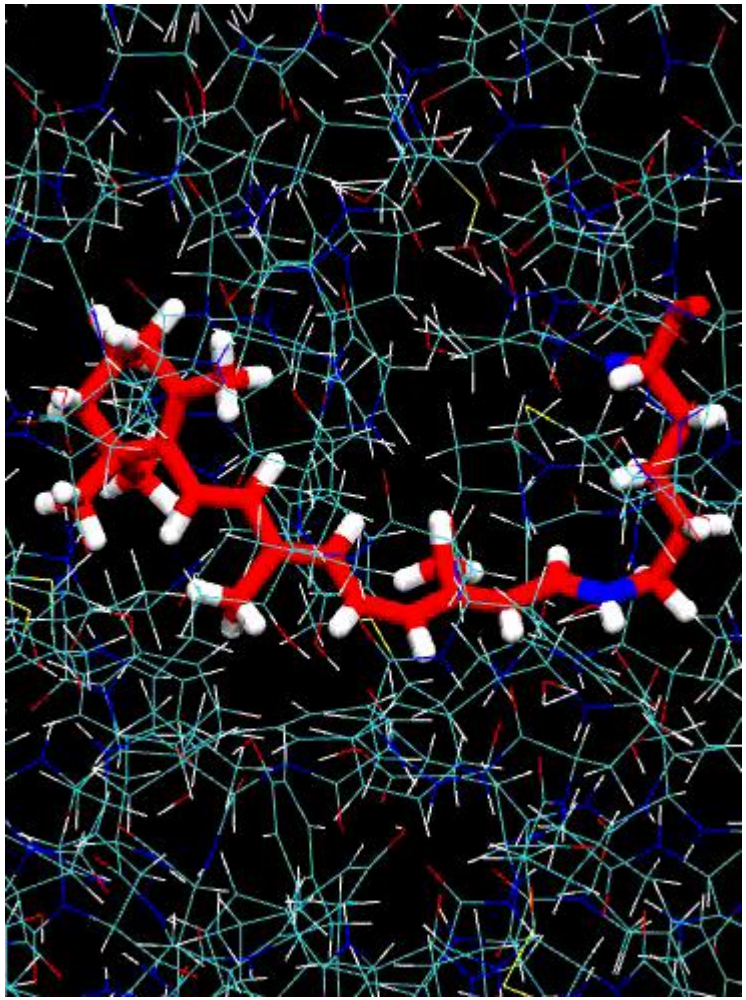
$$F\left(\mathbf{r}_{\mathrm{QM}}\right)=-\frac{1}{\beta}\ln\int d\mathbf{r}_{\mathrm{MM}}\exp\left[-\beta H_{\mathrm{QM/MM}}\left(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)\right]$$

Equilibrium MM

On-the-fly QM/MM MD Simulation



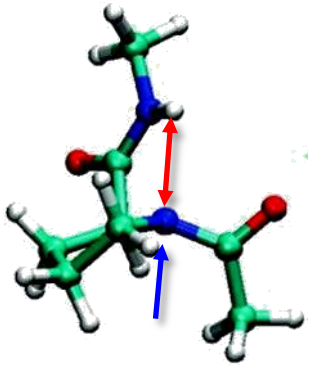
Bovine rhodopsin



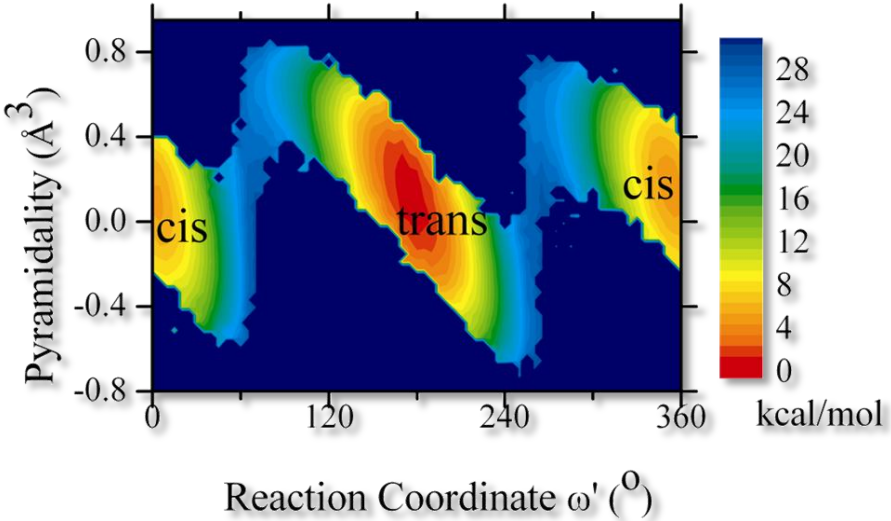
Hayashi S, Tajkhorshid E, Schulten K, *Biophys J.* **96**, 403 (2009)

QM/MM umbrella sampling

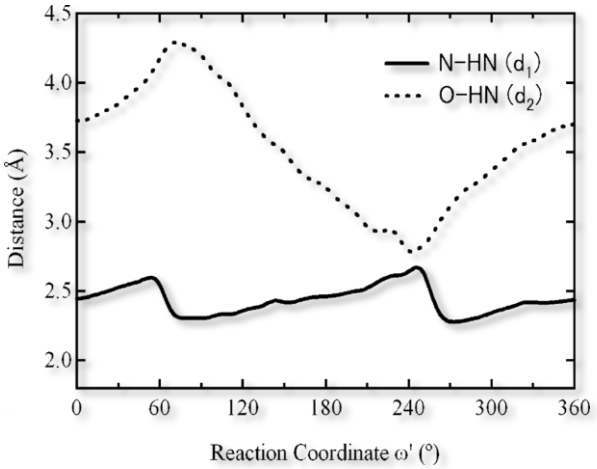
Cis-trans isomerization of a solvated Proline dipeptide



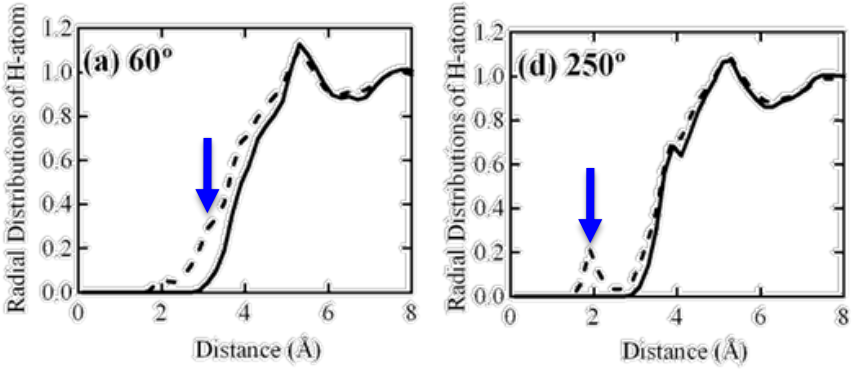
$$\mathbf{r}_{QM} = \lambda \mathbf{r}^0_{QM} + (1-\lambda)\mathbf{r}^1_{QM}$$

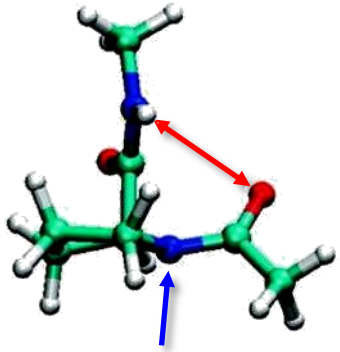


Hydrogen bond



Hydration





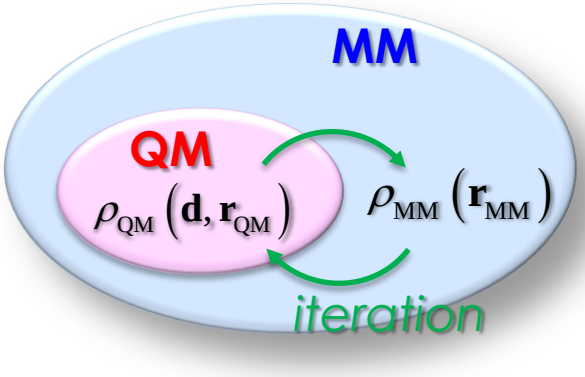
Y. Yonezawa et al., J. Am. Chem. Soc., 131, 4535 (2009)

QM/MM free energy

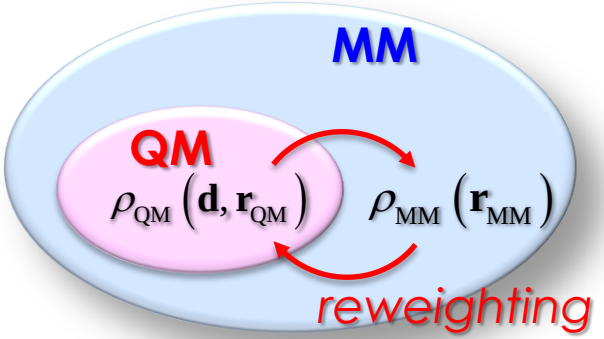
$$F\left(\mathbf{d},\mathbf{r}_{\mathrm{QM}}\right)=-\frac{1}{\beta}\ln\int d\mathbf{r}_{\mathrm{MM}}\exp\left[-\beta H_{\mathrm{QM/MM}}\left(\mathbf{d},\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)\right]$$

Complete separation of QM/MM optimization and MD sampling

Conventional



Present



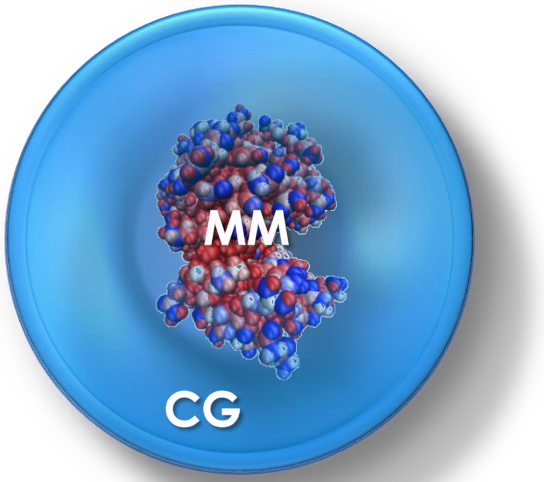
$$\rho_{\mathrm{MM}}\left(d,\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)=\frac{\exp\left(-\beta\left[H_{\mathrm{QM/MM}}\left(d,\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)-H_{\mathrm{QM/MM}}\left(d^{0},\mathbf{r}^{0}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)\right]\right)}{\int d\mathbf{r}_{\mathrm{MM}}\exp\left(-\beta\left[H_{\mathrm{QM/MM}}\left(d,\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)-H_{\mathrm{QM/MM}}\left(d^{0},\mathbf{r}^{0}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)\right]\right)}\rho_{\mathrm{MM}}\left(d^{0},\mathbf{r}^{0}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}\right)$$

S. Hayashi (Kyoto U)

MM/CG

Parallel algorithm

3D-RISM

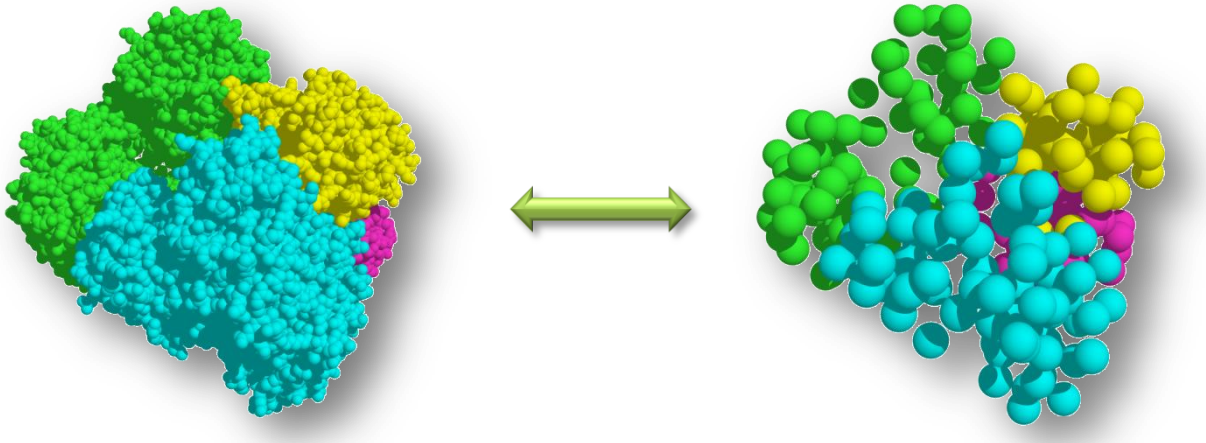


Serial algorithm

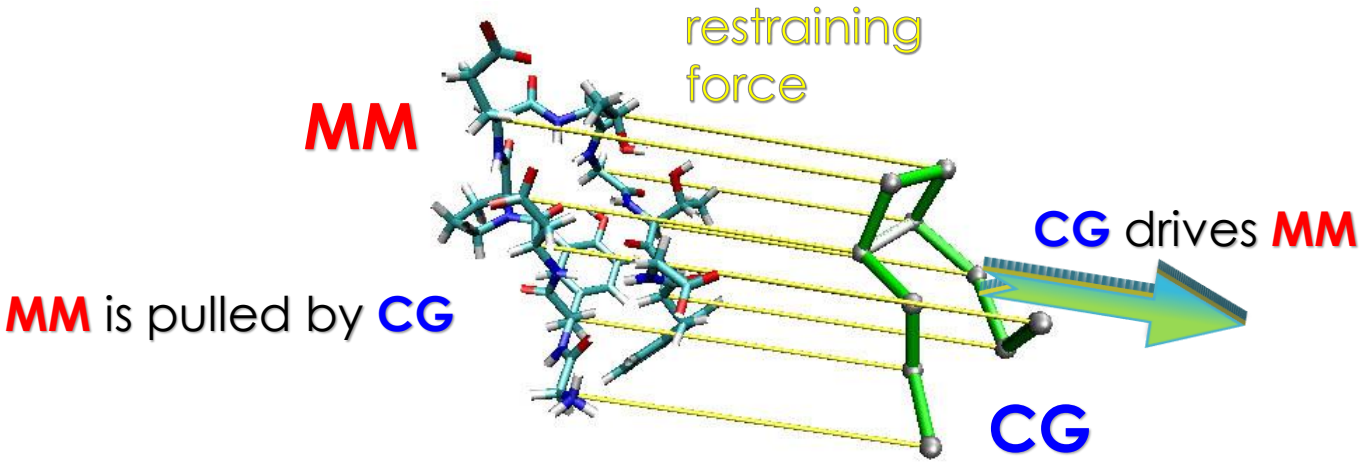
Coupled Hamiltonian

Path Search – Path Ensemble

Sequential Monte Carlo Simulation



MS-MM/CG Simulation



$$H(\mathbf{x};\mathbf{z}) = V(\mathbf{x}) + U(\mathbf{z})$$

MM (x) CG (z)

Removing the influence from the coupling

$$H(\mathbf{x};\mathbf{z}) = V(\mathbf{x}) + U(\mathbf{z}) + k/2(\theta(\mathbf{x}) - \mathbf{z})^2$$

MM $\mathbf{x}$ **CG** $\mathbf{z}$ ***coupling***

Hamiltonian Exchange (multicopy)

$k \to 0$

$$H(\mathbf{x};\mathbf{z}) = V(\mathbf{x})$$

MM

Multiscale Enhanced Sampling (MSES)

$k \to \infty$

$U(\mathbf{z}) = 0$

Path Search (multicopy)

$$-F(\mathbf{z}) \to \ln \int d\mathbf{x} e^{-\beta V(\mathbf{x})} \delta(\theta(\mathbf{x}) - \mathbf{z})$$

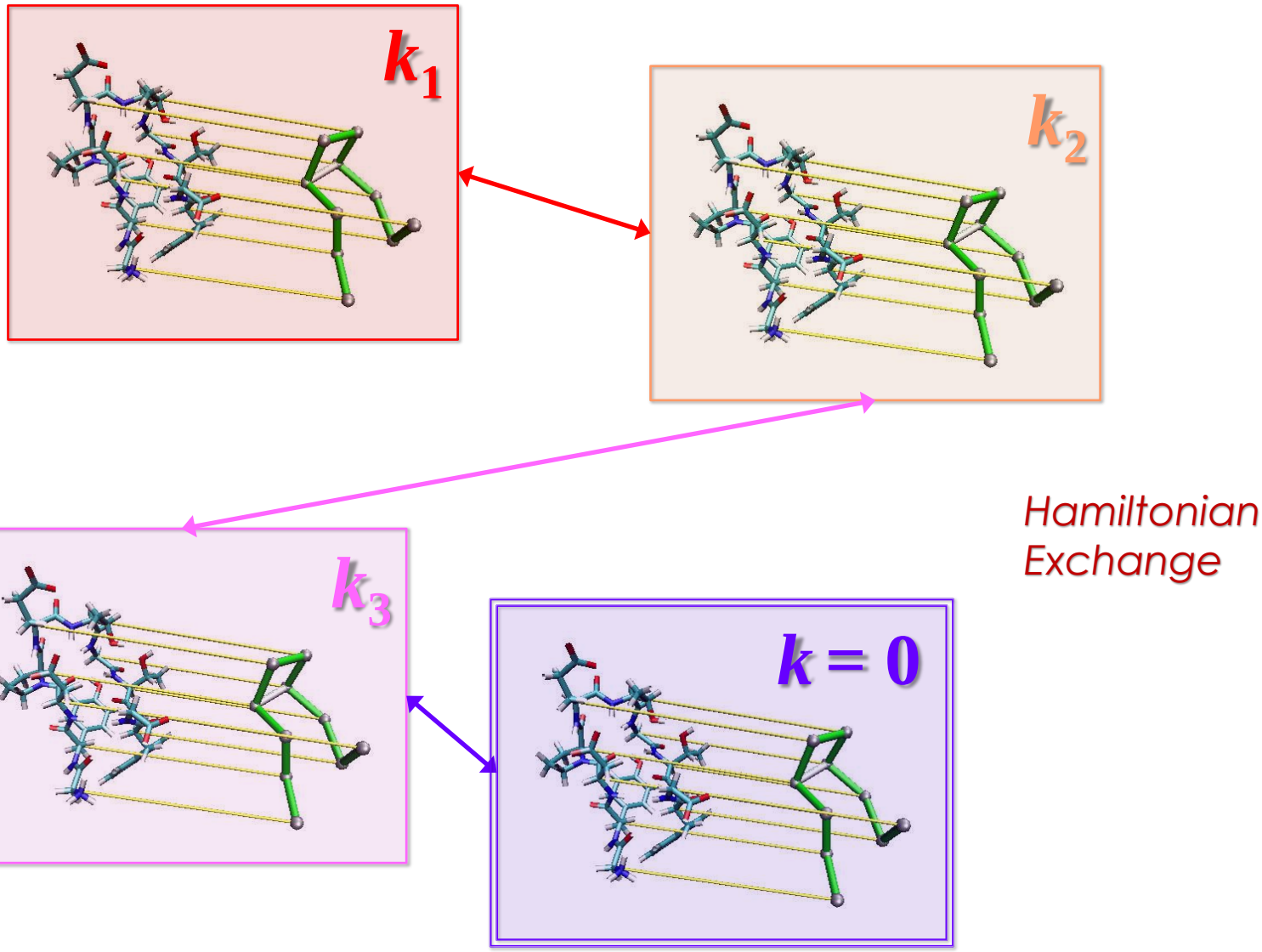
potential of mean force of **MM**
potential on $\mathbf{z}$

On-the-fly String Method

L. Maragliano, E. Vanden-Eijnden, *Chem. Phys. Lett.* **446**, 182 (2007).

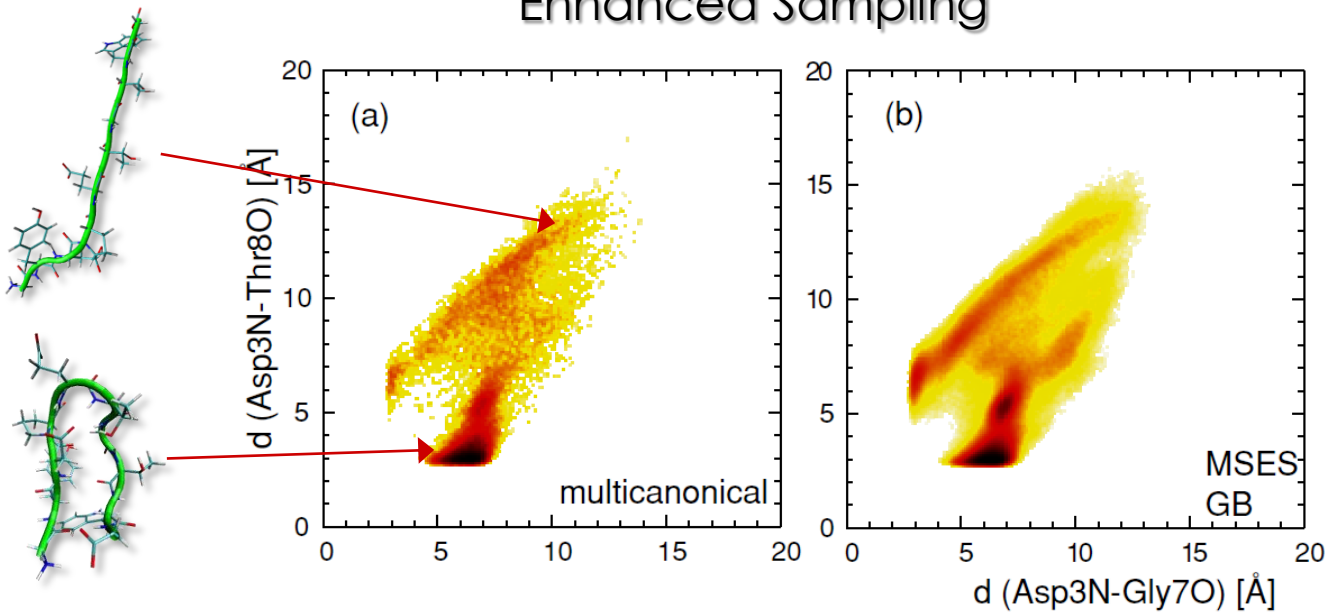
2-2. Multiscale Enhanced Sampling (MSES)

K. Moritsugu (Riken)



K Moritsugu, T. Terada, A. Kidera, J. Chem.Phys. 133, 224105 (2010)

Enhanced Sampling



Metropolis Criterion $\min\left(1, \exp\left(\Delta_{mn}\right)\right)$

Temperature RE

$$\Delta_{mn} = \left(\beta_m - \beta_n\right)\left[V\left(\mathbf{x}_m;\beta_m\right) - V\left(\mathbf{x}_n;\beta_n\right)\right]$$

MSES

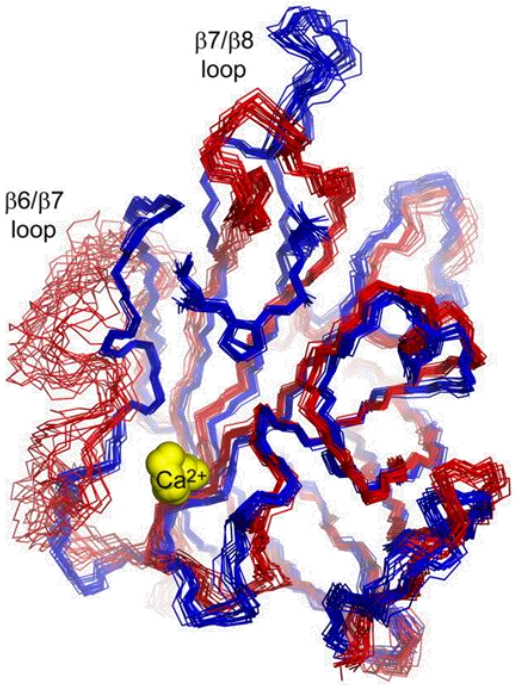
$$\Delta_{mn} = \beta\left(k_m - k_n\right)\left[W\left(\mathbf{x}_m,\mathbf{z}_m\right) - W\left(\mathbf{x}_n,\mathbf{z}_n\right)\right]$$

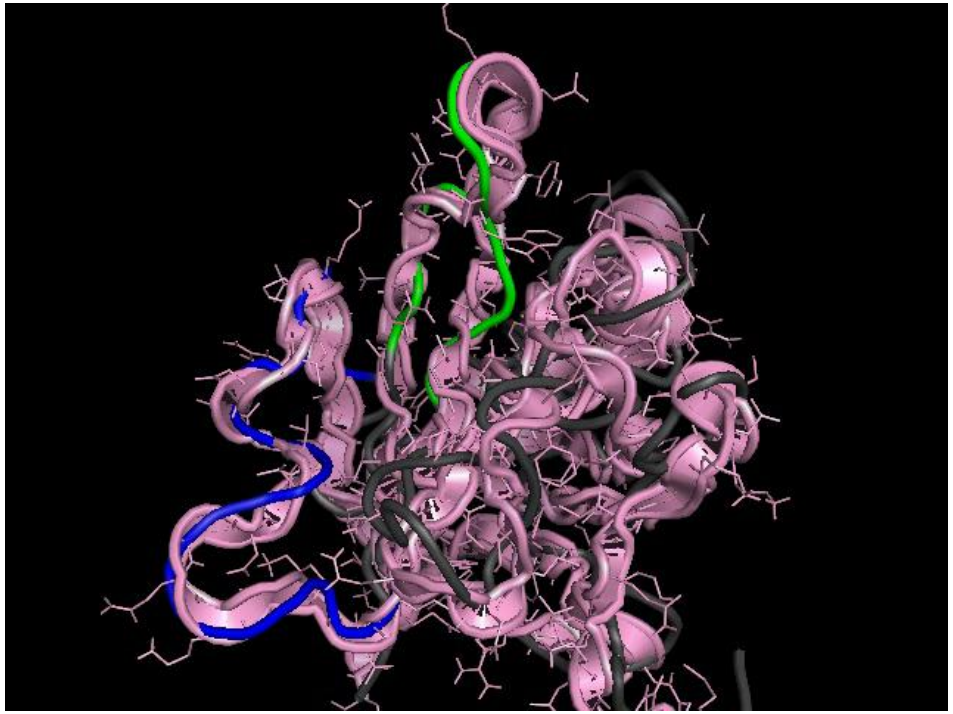
$$W(\mathbf{x},\mathbf{z}) \equiv \frac{1}{2}\left[\theta(\mathbf{x}) - \mathbf{z}\right]^2$$

1. The constraint W can be set independently of $V(\mathbf{x})$
2. The number of degrees of freedom in W can be small
3. The potential is harmonic

Excellent scalability has been attained in MSES.

Intrinsically Disordered Protein: Sortase A

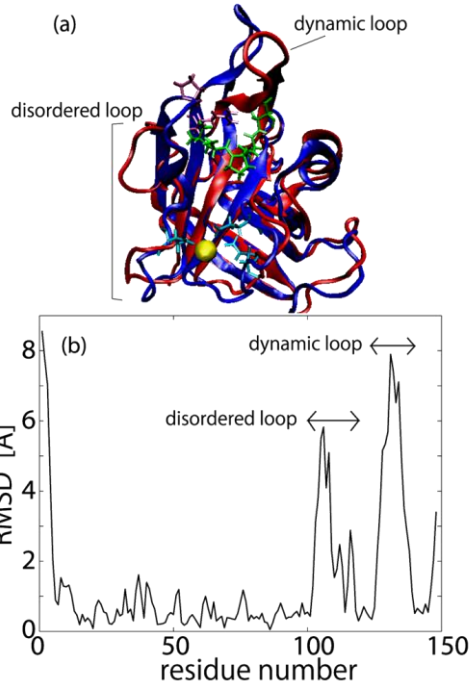


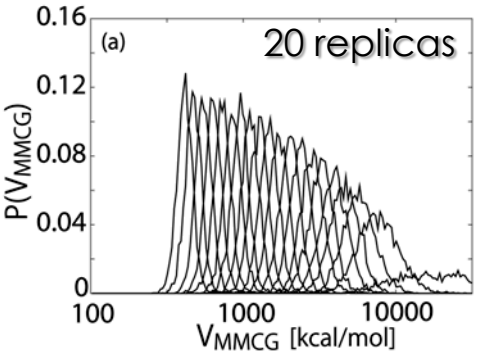


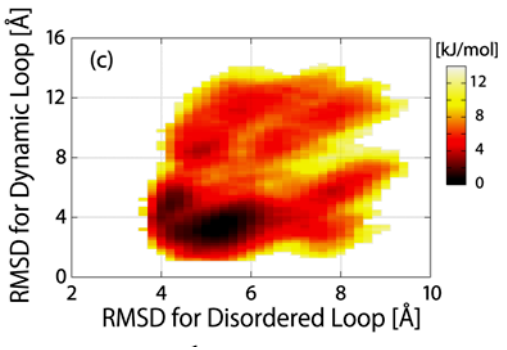
Enhance the disordered regions

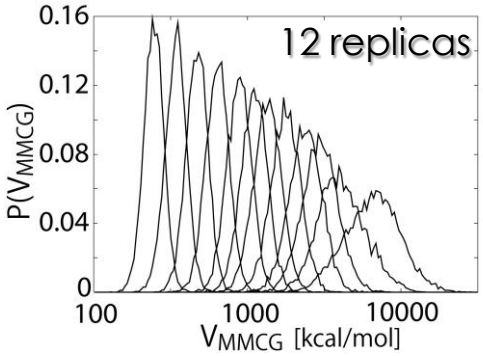
N Suree, *et al*. *J. Biol. Chem*. **284**, 24465 (2009)

Dynamic and Disordered Loops

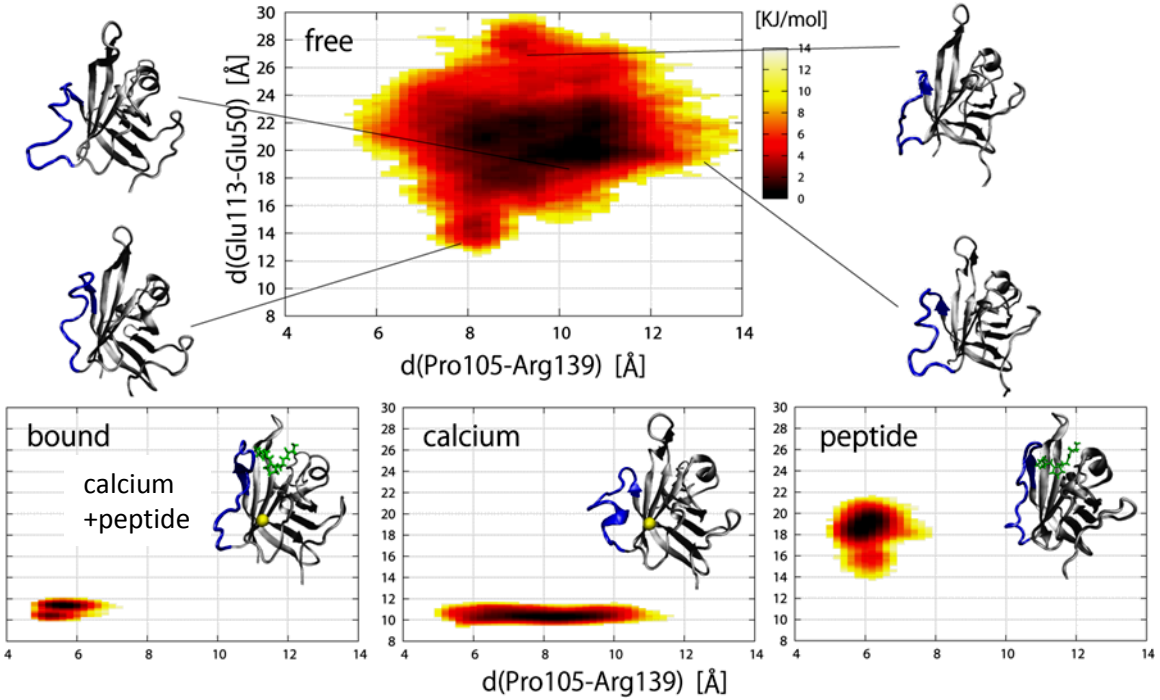








Disordered Loop only

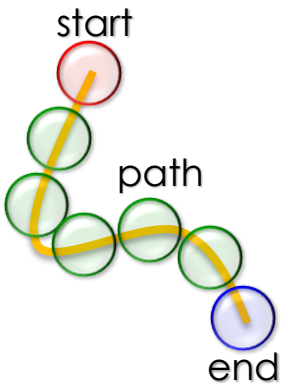


2-3. Path Search: On-the-fly String Method

$$H(\mathbf{x};\mathbf{z}) = V(\mathbf{x}) + U(\mathbf{z}) + k/2(\theta(\mathbf{x}) - \mathbf{z})^2$$

MM CG coupling

Y. Matsunaga (Riken)



$k \to \infty$

$U(\mathbf{z}) = 0$

Path Search (multicopy)

potential of mean force on $\mathbf{z}$

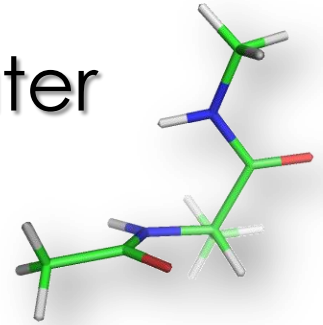
No guiding potential → Entropy driven
Require some constraints → Path search

Path Search Methods

Minimum energy paths
- Intrinsic reaction coordinate (K Fukui, 1981)
- Self-penalty walk (R. Elber, 1987)
- Conjugate-peak refinement (M. Karplus, 1992)
- Nudged-elastic band (H. Jonsson, 1994)
- String method (E. Vanden-Ejinden, 2002)

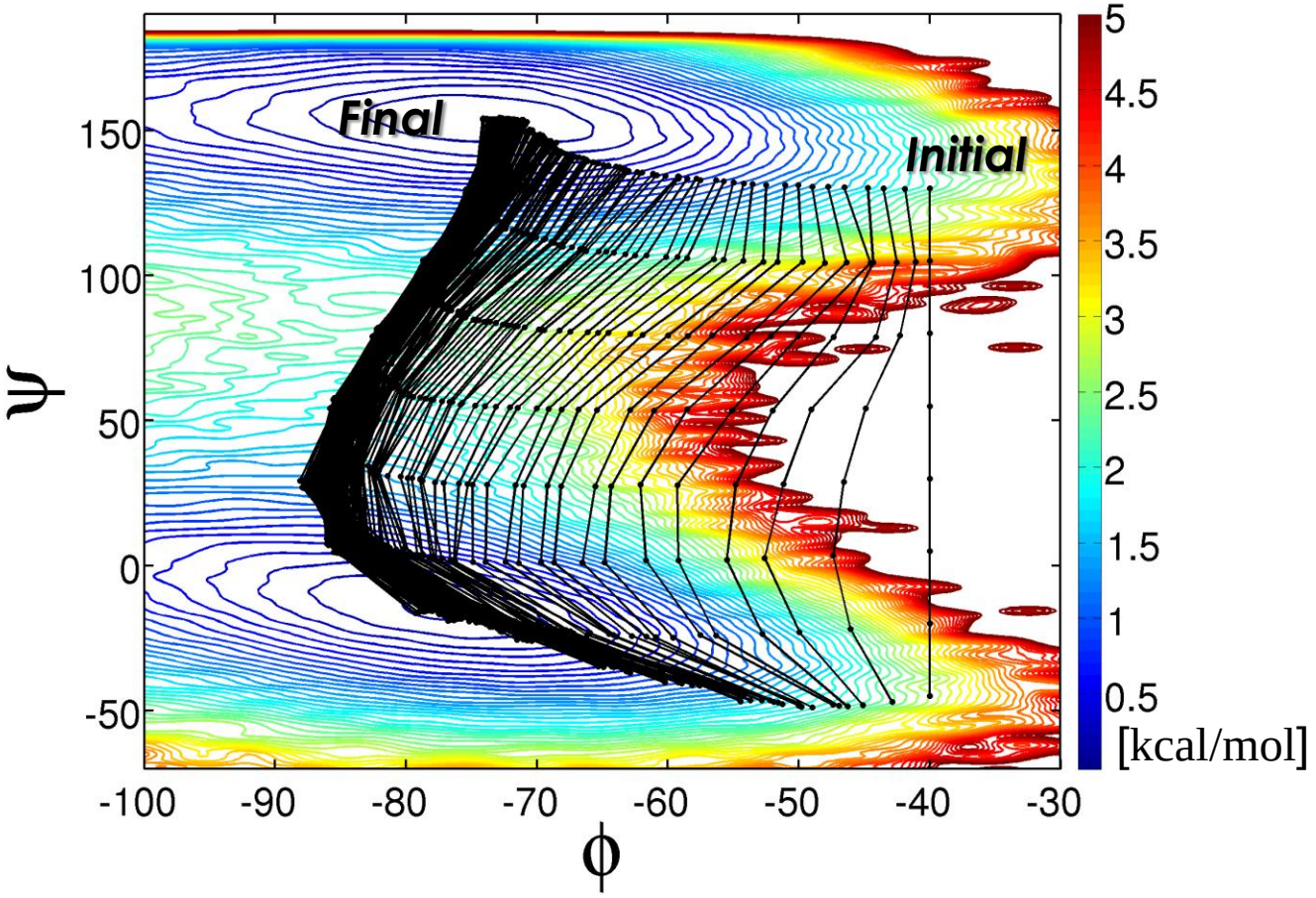
Finite tempearture paths
- Max flux method (J.E. Straub, 1997)
- Finite tempearture string method (E. Vanden-Ejinden, 2005)
- On-the-fly string method (E. Vanden-Ejinden, 2007)

Alanine-dipeptide in GB water

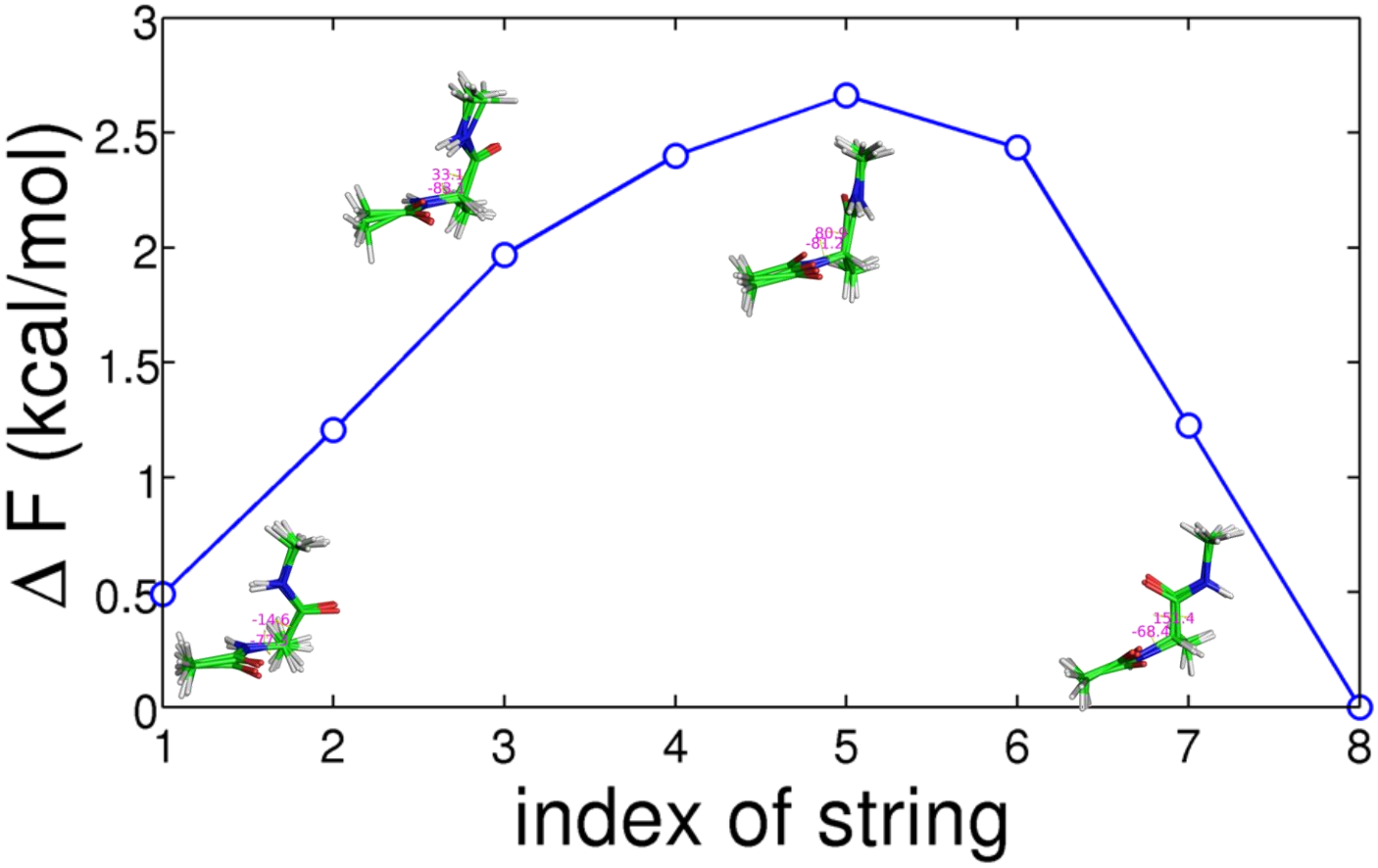


$$\mathbf{z} = (\phi, \psi)$$

8 replicas



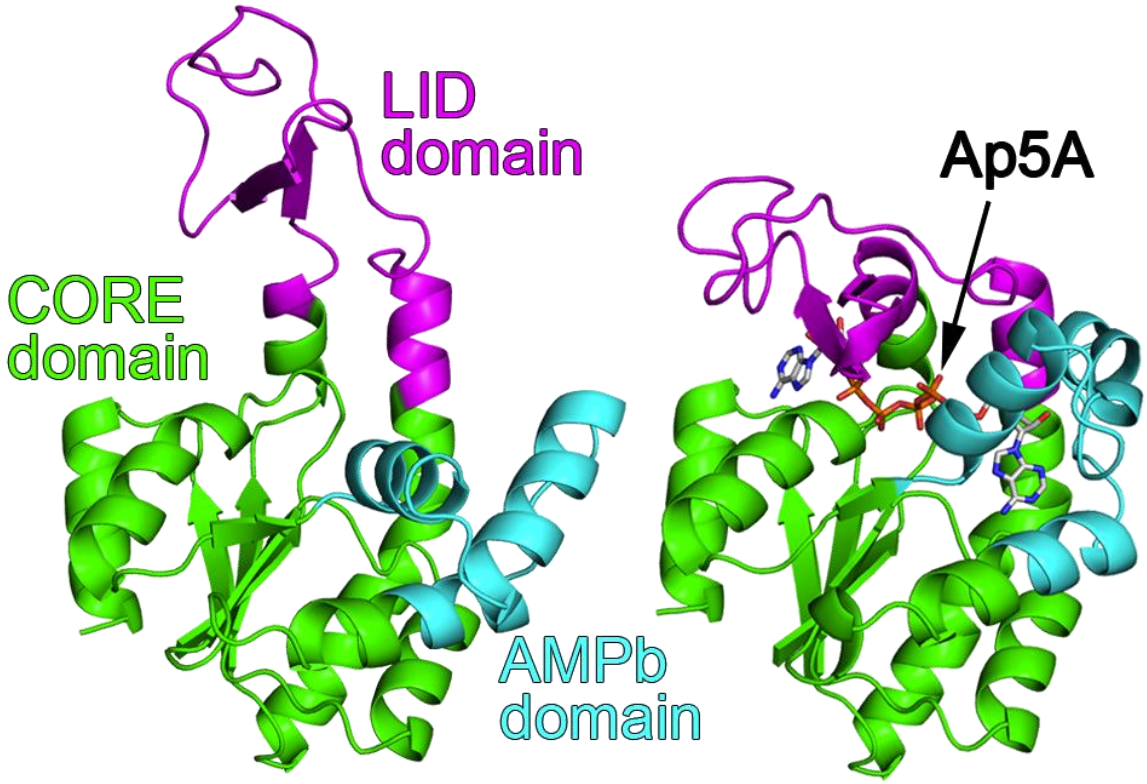
Minimum Free Energy Path



Adenylate kinase in water (65,000 atom system)

z: 20 principal components

65 replicas



タンパク質は動く

1．このように動く
　1−1．データベースから見えるタンパク質の応答的運動
　1−2．応答的タンパク質機能像
2．そのように動くタンパク質機能の分子シミュレーション
　2−1．もっと速さを！
　2−2．次世代生命体統合シミュレーション 分子スケール
　　(1) 多剤排出トランスポーターを例に
　　分子動力学計算
　　粗視化モデル計算
　　(2) 量子化学計算
　　(3) マルチスケールシミュレーション