LBNL-CCS workshop 2018
March 5th–6th @U Tsukuba

Developments of Computational Methods for Understanding Biological Functions of Proteins

Ryuhei Harada and Yasuteru Shigeta

Center for Computational Sciences
University of Tsukuba

Protein Crowding Simulations for Understanding Biological Functions in Cellular Environments

Computational Biomolecular Medical Science

Main Target
- Medium size drugs
- Covalent bond-type drugs
- Single Nucleotide Polymorphism

Problems
- Diversity in Structures (Vast Conformations)
- Reaction Contorl (First-principles Calculations)
- Fast Gene Analysis

Drug discovery needs long time (10～15 year) and ($1billion), But the success ration is less than 1/20000
We needs a breakthrough for in silico drug discovery,
.e. detection of target, design of drug and proteins & etc.

→ **Fusion of Biology, Medicine, and Physics to solve above problems**

Evolution

Prof. Y. Inagaki
Department of Biology

Computational Biology

Next-generation sequencer

Fast Single Nucleotide Polymorphism detection

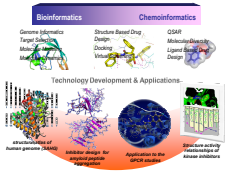
Phylogenetic tree analysis by GPU
Mitocondrial protein detection by supercomputer

Gene analysis
SNP analysis

Drug Simulation

Prof. T. Hirokawa
Department of Medicine & AIST

Bioinformatics,
In Silico Drug discovery

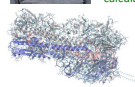


In Silico drug discovery project by AMED
since 2017.

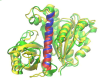
First-principle

Prof. Y. Shigeta
Department of Physics

Biophysics, First-principles calculations



Influenza HA3
First-principles Calculation



Molecular dynamics of FtsZ In cell fission

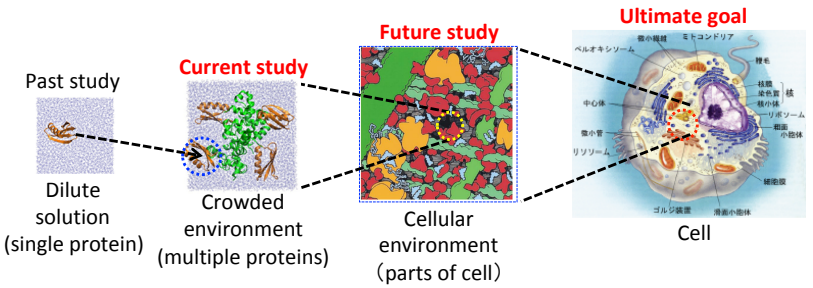
Fragment molecular orbital by GPU architecture
Efficient sampling method by Molecular Dynamics

Post K Computer Project by MEXT
since 2015

Toward molecular cellular-scale simulations

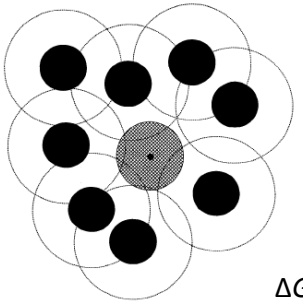
～to understand biological functions in cell～

To better understand cell,
→analyze ***crowding effects***, which control biochemical reactions in cell
①Dynamics：***Protein dynamics*** under cellular environments
②Thermodynamics：***Protein interaction/stability*** under cellular environments
③Function: ***Protein functions*** under cellular environments



The limitations of models

★Hard sphere model (Minton,1993)
(*Entropy-centered* model)

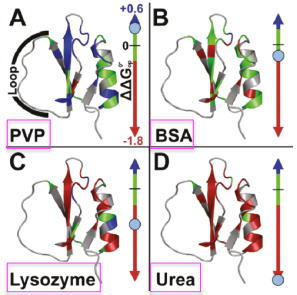


Compact folded structures are to stabilized

A *limitation* of the model

$\Delta G=\Delta H-T\Delta S$

★NMR amide-exchange experiments of CI2 (Pielak, 2011)

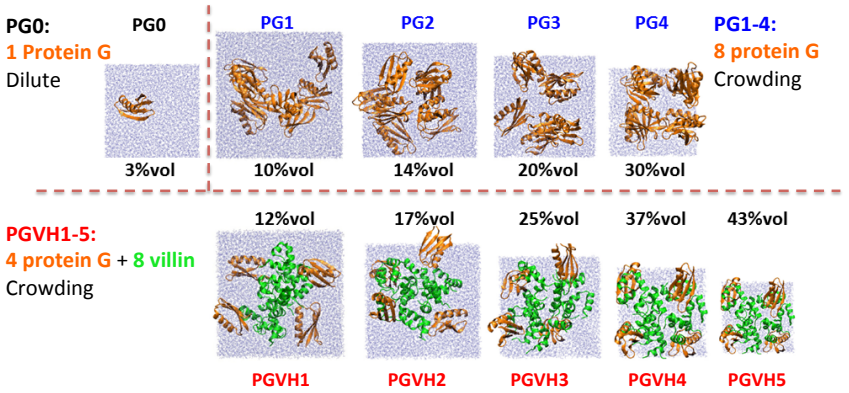


Crowding stabilizes or destabilized proteins

☆***All-atom MD simulations*** were performed to consider *enthalpy-centered* protein-protein or protein-water interactions!

Protein crowding simulations

★We performed a series of
all-atom MD simulations of two different protein systems,
Changing protein volume fraction

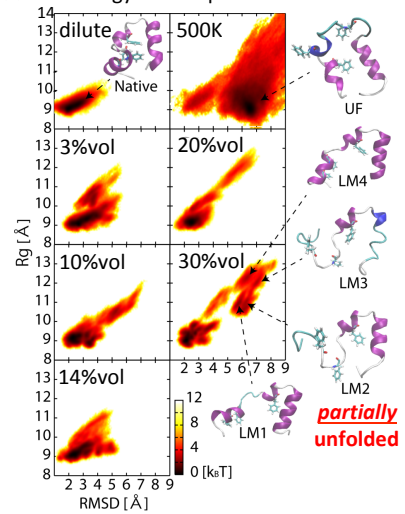


***J. Am. Chem. Soc.* 2012, 134, 4842-4849**

NPT(1bar, 298K 300ns)

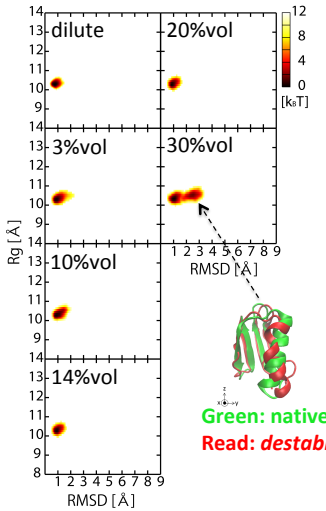
Destabilizations of proteins due to crowding

★Free energy landscapes of ***villin***



***partially* unfolded**

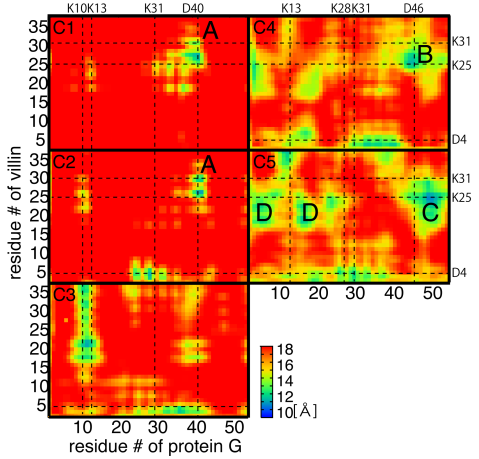
★Free energy landscapes of ***protein G***



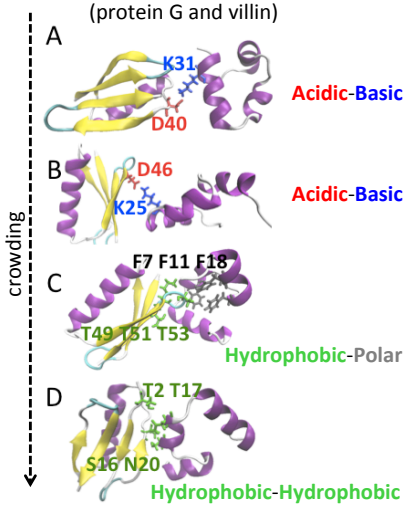
Green: native
Read: *destabilized*

Protein interactions due to crowding

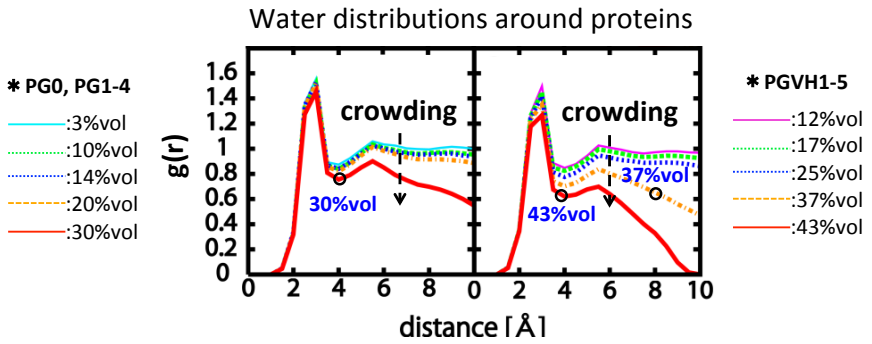
★***Contact maps*** between protein G and villin
(***average minimum distances***)



★Representative dimer structures
(protein G and villin)

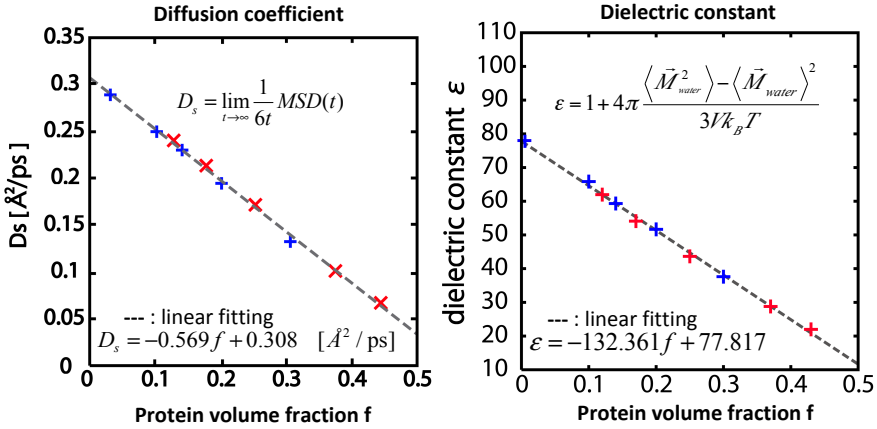


Radial distribution function



★ Distributions of bulk water (> 6Å) *drastically* decreased with protein volume fraction > ***30%*** .

Universality in crowding

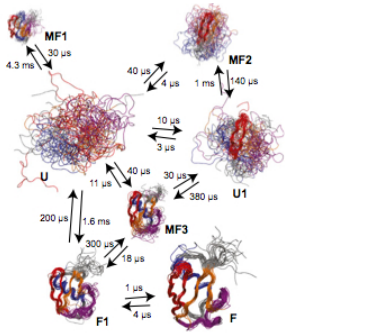


★ **Diffusion coefficient and Dielectric constant *linearly* decreases with protein volume fractions**

Developments of Computational Methods for Promoting Biologically Relevant Rare Events of Proteins

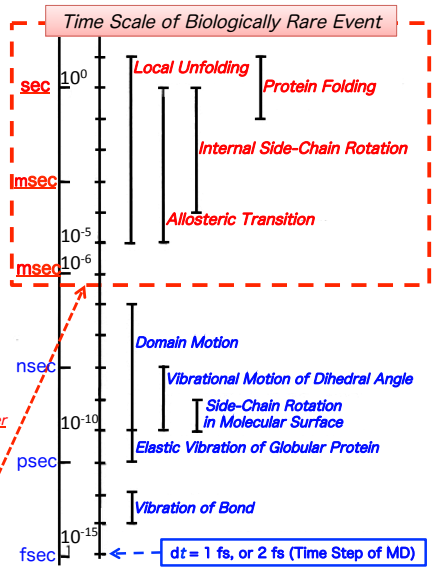
Background: *Biologically Rare Event*

◆ *Brute-Force Molecular Dynamics (MD)* by a Special Purpose Supercomputer (ex. *ANTON from D. E. Shaw Research*)



Millisecond-order Folding Simulation of Ubiquitin in Explicit Water by a Brute-Force MD. (*PNAS*, **110**, 5915, (2013). **Fig. 1**)

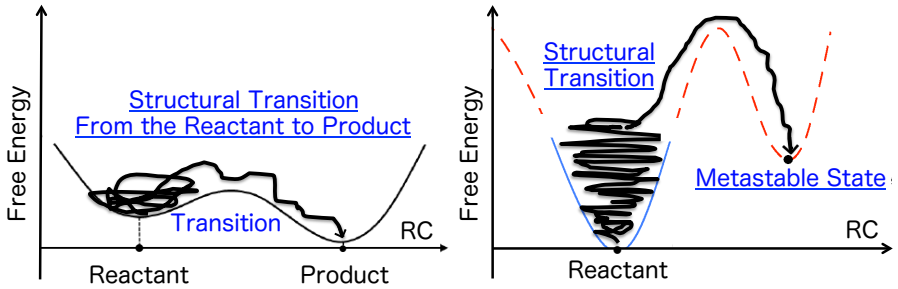
The accessible timescale of normal MD simulation is far from the timescale of rare events.



Purpose: *Developments of Methods for Reproducing Rare Events as Structural Transitions of Proteins*

1. Parallel Cascade Selection MD (PaCS-MD) (*J.Chem.Phys.* ***139****, 035103* (2013))

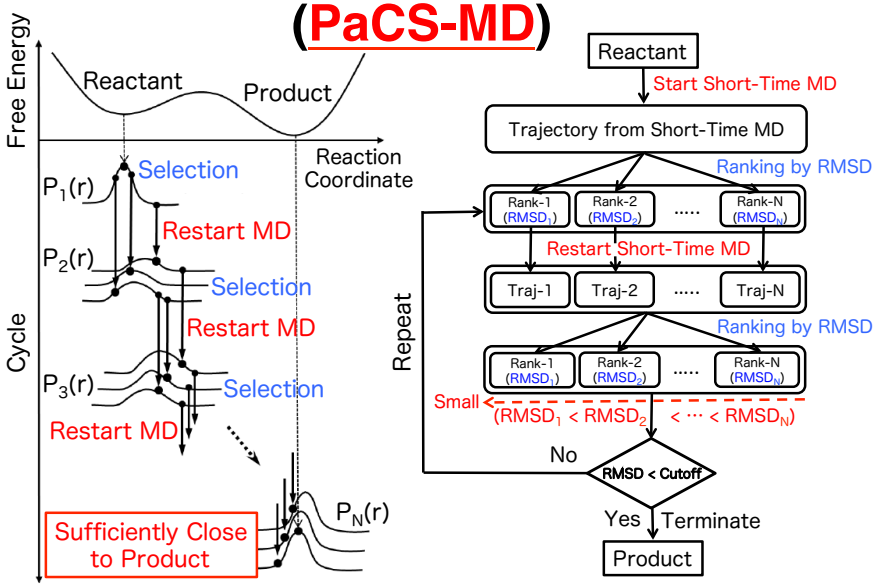
2. Outlier FLOODing (OFLOOD) Method (*J. Comput. Chem.* **36** 97 (2015))



In these methods, *multiple and short-time MD simulations are used instead of a long-time MD simulation.*

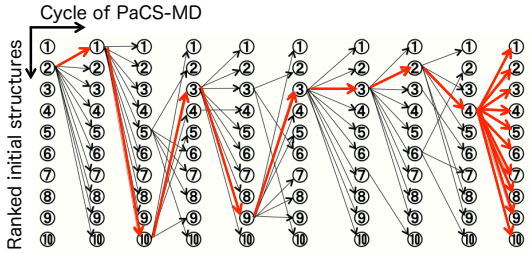
For a review of other methods, see *Phys.Chem.Chem.Phys.*, R. Harada *et. al.*, **17**, 6155-6173 (2015)

Parallel Cascade Selection MD (PaCS-MD)



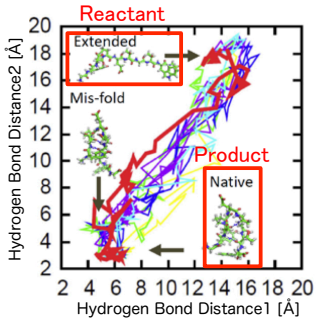
Selections of Initial Structures in PaCS-MD: *Folding of Chignolin*

Selections of Initial Structures at Each Cycle



∗Computational cost: 100 ps X 10 initial structures / cycle (1ns / cycle)

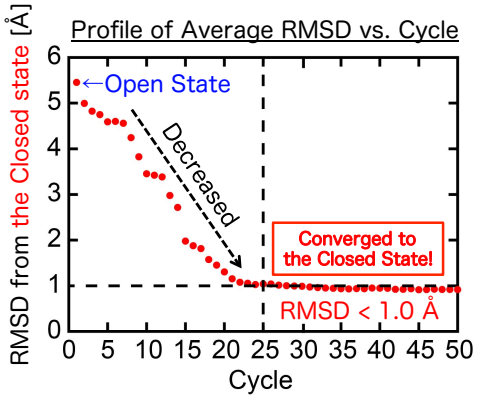
Folding Pathways Reproduced by PaCS-MD



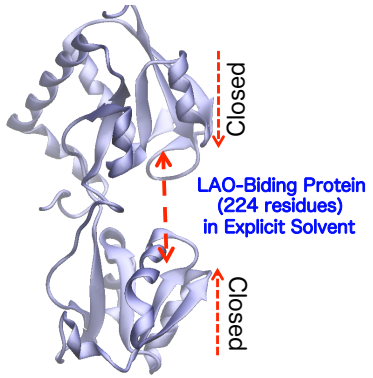
∗ *Snapshots are Ranked with RMSD from Product at Each Cycle.*
∗ *Highly ranked Top Ten Snapshots were Selected by low RMSD values.*

J.Chem.Phys., R. Harada *et. al.*,**139**, 035103 (2013)

Open-Closed Transitions of Protein Reproduced by PaCS-MD

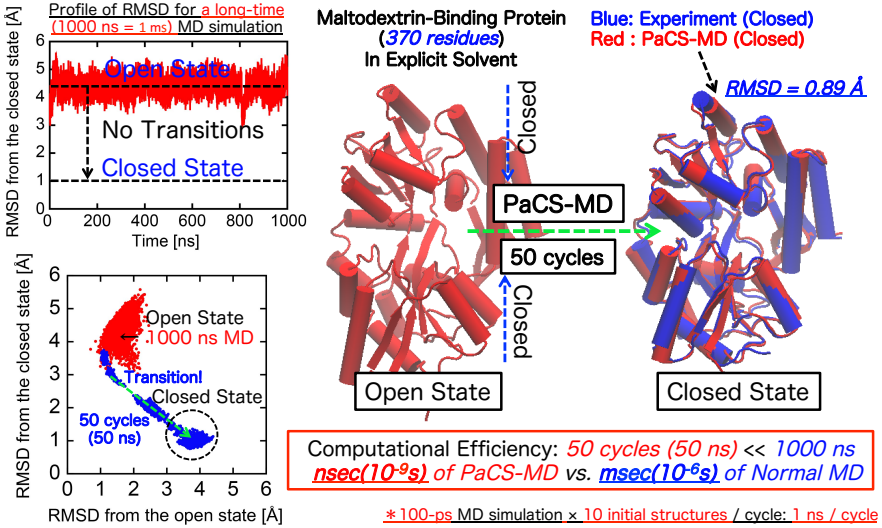


Open-Closed Transition Reproduced by PaCS-MD



∗ Short-time (100 ps) MD simulation × 10 initial structures / cycle: 1 ns / cycle
Totally simulation time after the 50th cycle: 50 ns
∗ $RMSD_{closed}$ < 1.0 Å → Terminate PaCS-MD.

Computational Efficiency of PaCS-MD

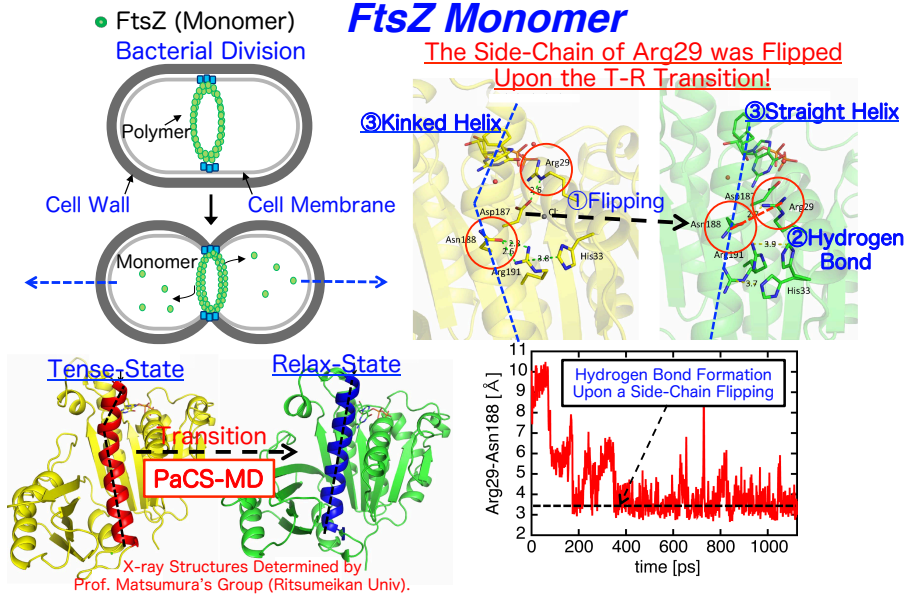


Computational Efficiency: *50 cycles (50 ns)* << *1000 ns*
nsec(10⁻⁹s) *of PaCS-MD vs.* ***msec(10⁻⁶s)*** *of Normal MD*

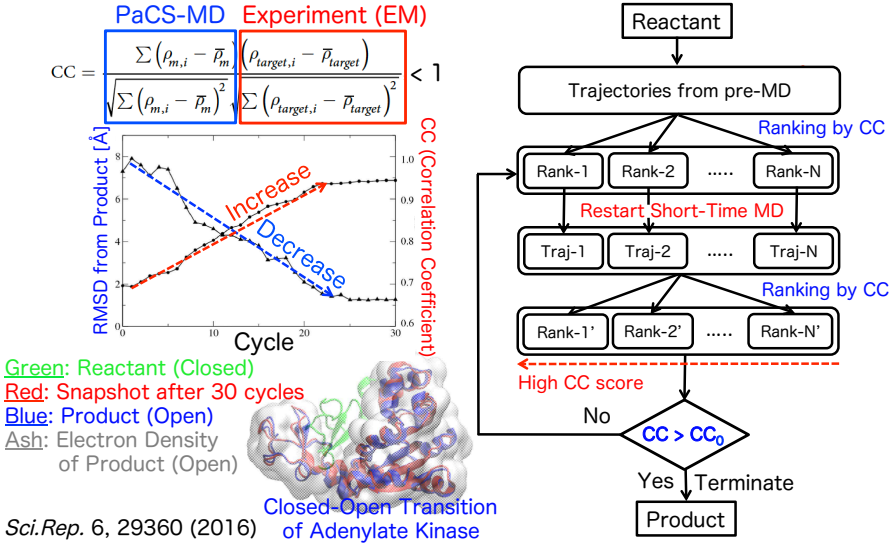
＊100-ps MD simulation × 10 initial structures / cycle: 1 ns / cycle

J. Struc. Biol., **198**, 65 (2017)

Application: *Structural Transition of FtsZ Monomer*



PaCS-MD Combined with *Experimental Data (PaCS-fit)*



Sci.Rep. 6, 29360 (2016)

How to define the scoring function

From PaCS-fit

From experiments

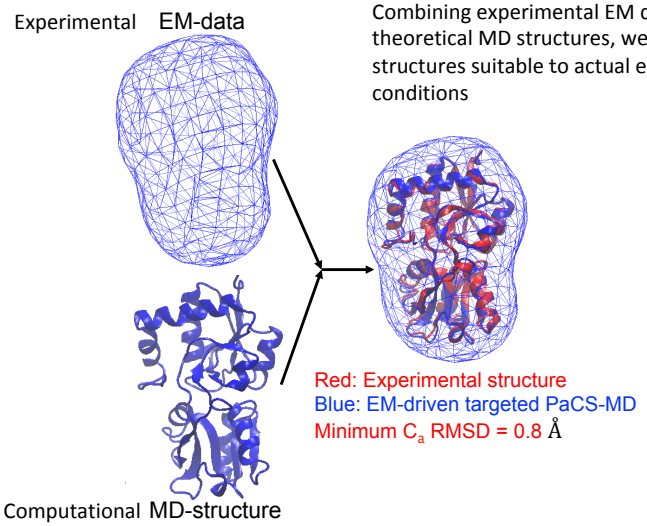
EM. The cross-correlation coefficient (CC) between the simulated map of any sampled conformation and the target EM map is computed as

$$CC = \frac{\sum(\rho_{m,i} - \bar{\rho}_m)(\rho_{target,i} - \bar{\rho}_{target})}{\sqrt{\sum(\rho_{m,i} - \bar{\rho}_m)^2}\sqrt{\sum(\rho_{target,i} - \bar{\rho}_{target})^2}} \quad (2)$$

where $\rho_{m,i}$ is the electron density of simulated map at the i^{th} voxel and $\rho_{target,i}$ is the corresponding density in the target map. During PaCS-Fit, all the CC values were computed by the collage program[47] in the Situs package[36].

In PaCS-fit, initial structures would be selected so that values of CC gradually increase. To reproduce data from EM, PaCS-fit is performed so that values of CC between theoretical and experimental data would be correlated, i.e. CC → 1.

EM-driven PaCS-MD for MBP



Combining experimental EM data with theoretical MD structures, we can predict structures suitable to actual experimental conditions

Red: Experimental structure
Blue: EM-driven targeted PaCS-MD
Minimum C_a RMSD = 0.8 Å

Summary of this talk

1. Molecular Dynamics Simulation in Cellular Condition

Crowding effects on protein folding

2. Parallel Cascade Selection Molecular Dynamics

Transition pathways among metastable structures
Prediction of Structures on the basis of experimental results
Protein-Protein interaction (PPT)