Complex Structure Modeling

Agricultural Bioinformatics Research Unit,
Graduate School of Agricultural and Life Sciences, The University of Tokyo
Tohru Terada

Contents

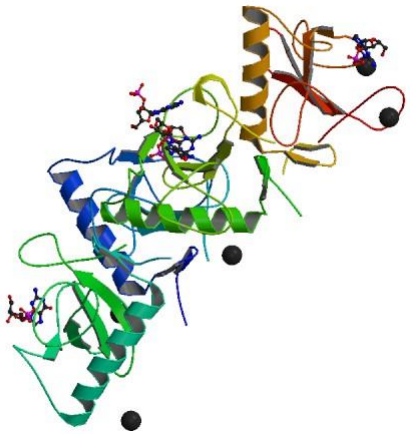
- Preparations for simulation
- Protein-protein docking
- Protein-small molecule docking
  - Exercise
- Perspectives of molecular simulation

Procedures for MD simulation

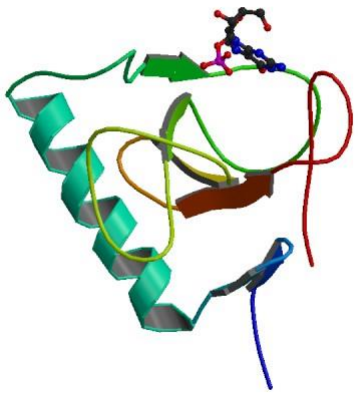
1. Preparation of the initial structure
   - Obtain the structure
   - Add missing atoms and residues
   - Add hydrogen atoms
   - Obtain ligand force field parameters
   - Solvate the system
2. Energy minimization
3. Assignment of the initial velovities
4. Equilibration
5. Production

Preparation of the initial structure (1)

- Obtain the structure
  - Download the experimental structure from PDB (http://www.rcsb.org/pdb/)
  - Usually, simulations are performed for the biological units of the biomacromolecules.
  - Example: Ribonuclease T1 (PDB ID: 1I0X)



Asymmetric unit



Biological Unit

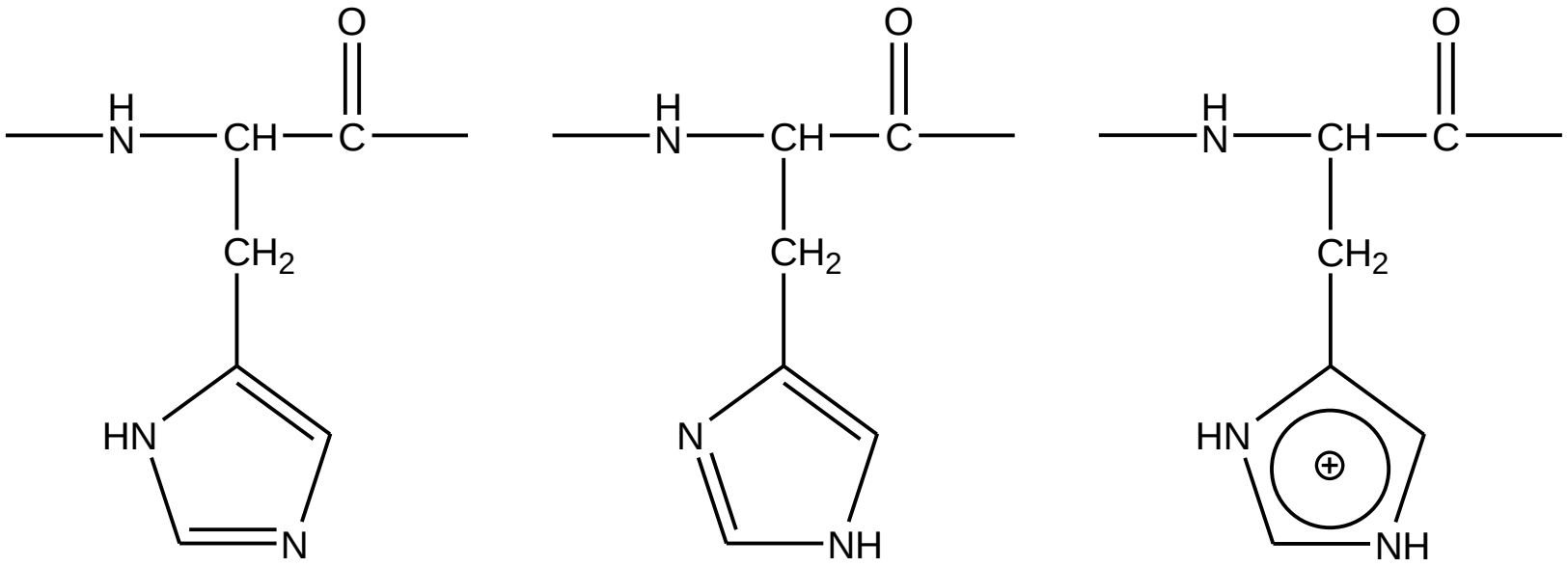
Preparation of the initial structure (2)

- Add missing atoms and residues
  - They can be added by using modeling software.
  - When N- or C-terminal residues are missing, you can block the terminus with an acetyl or N-methyl group.
- Add hydrogen atoms
  - Most of them are added automatically.
  - Pay special attention to SS bonds and protonation states of His.

Operations in Discovery Studio (1)

1. Choose “File”→“Open URL” from the menu, enter “1I0X” for ID, and click “Open.”
2. Change Display Style to Line.
3. Select B, C, and D chains in Hierarchy Window and delete them.
4. Click “Macromolecules” button and expand “Protein Report” in the Tools tab.
5. Click “Protein Report.”
   →Check Incomplete or Invalid Residues. (Lys41, Asp49, Glu102 are colored purple.)
6. Expand “Prepare Protein” in the Tools tab, and click “Clean Protein” in the Manual Preparation section.
   →Missing atoms are added.

Protonation states of His

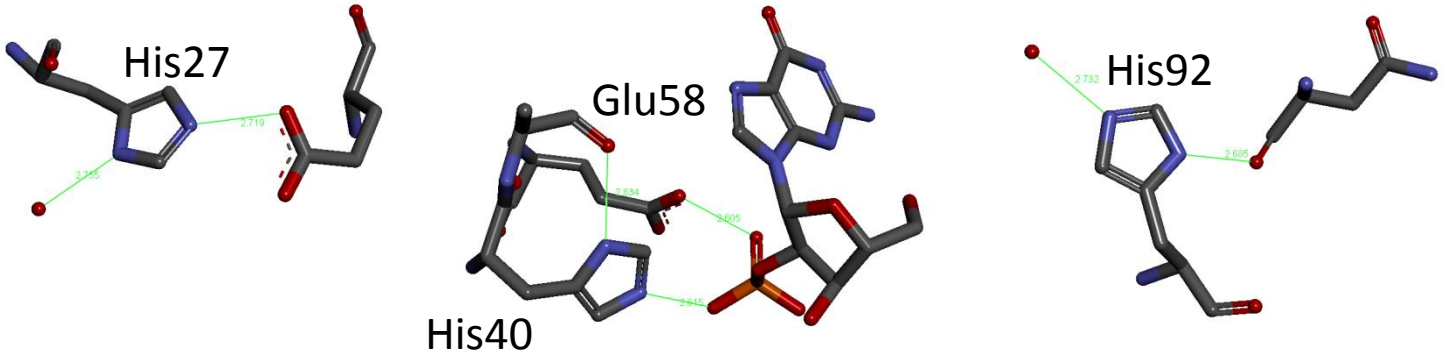


Protonation at δ | Protonation at ε | Protonation at δ and ε

- pK_a of His side chain is close to neutral (~ 6.5).
- You can find the protonation state from the hydrogen bond network where His is involved.

Operations in Discovery Studio (2)

7. Check the interactions of His27, His40, Glu58, and His92 with their surroundings.



8. Apply CHARMm force field, click “Calculate Protein Ionization and Residue pKa” in Protonate Protein section of Prepare Protein, and click “Run.”
→Check the protonation states of the residues.

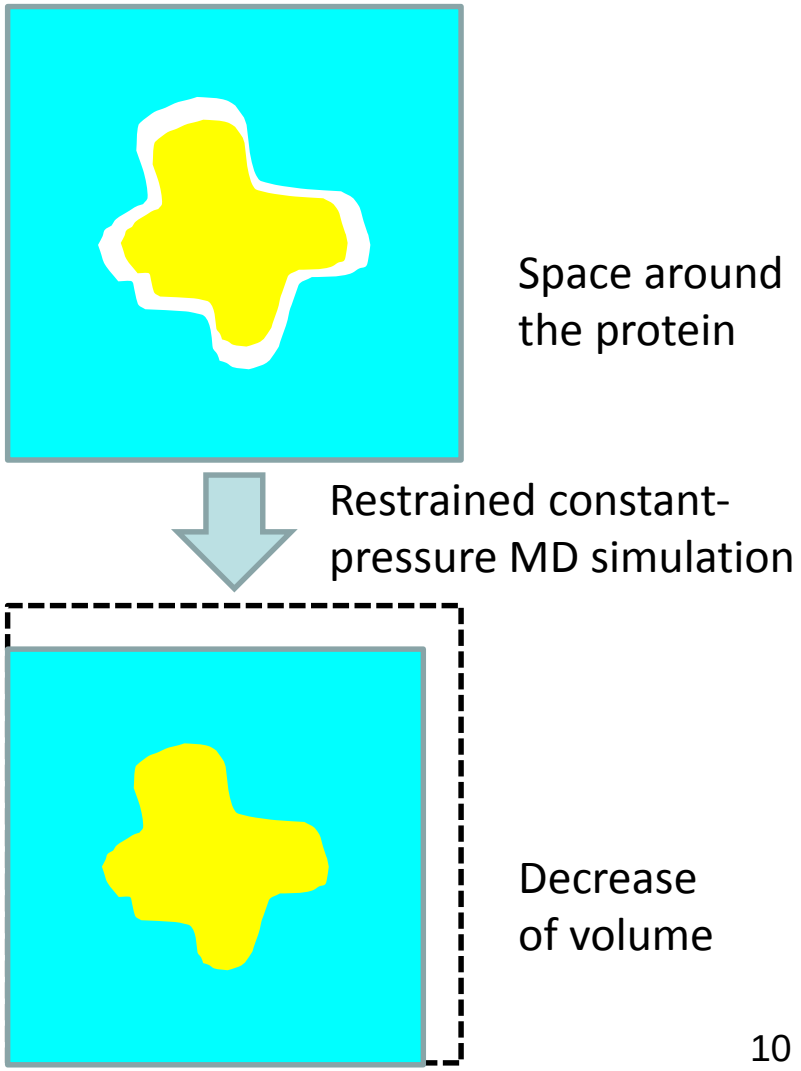
Preparation of the initial structure (3)

- Obtain ligand force field parameters
  - Ligand force field parameters are not included in the molecular dynamics software. It is necessary to make them by yourself or to obtain them from Amber Parameter Database.*
- Solvate the system
  - For an accurate and efficient simulation using the PME method, solvate system in a rectangular water box.
  - Add counterions to neutralize the system.

*http://www.pharmacy.manchester.ac.uk/bryce/amber

Equilibration

- In the initial structure, there is a space between the protein and the water.
- It is necessary to optimize water arrangement by performing a constant-pressure MD simulation.
- During the simulation, positions of protein atoms are restrained to their initial position and the restraints are gradually relaxed.



Complex structure modeling

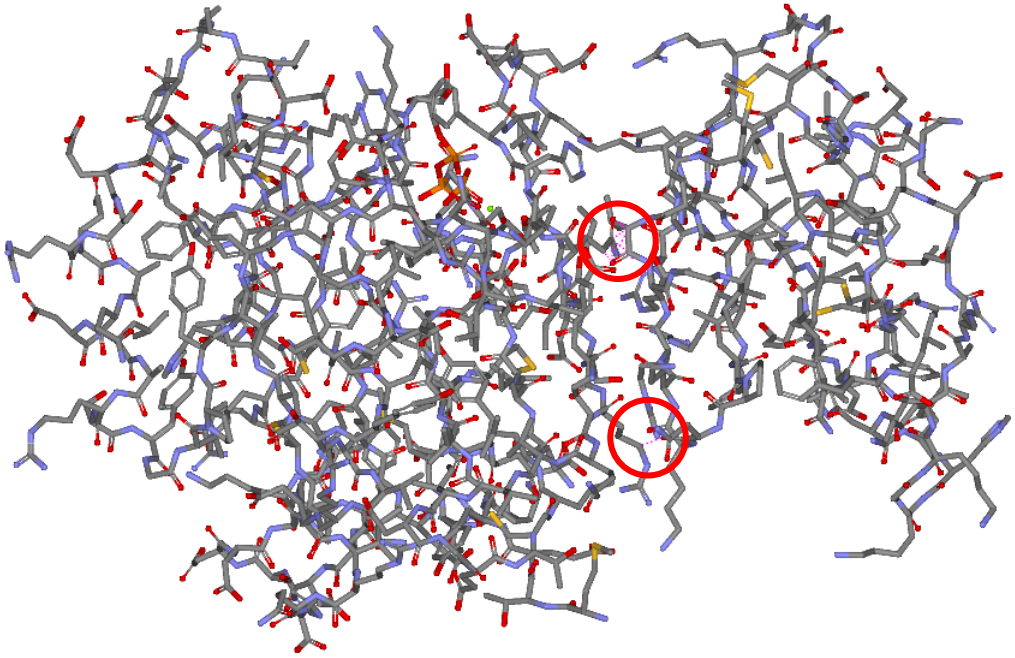
- Predicts protein-protein or protein-small molecule complex structure.
- If an experimental structure of similar complex is available, you should try following methods:
  - Homology modeling
  - Structure superposition
- If not, try
  - Docking simulation

Structure superposition (1)

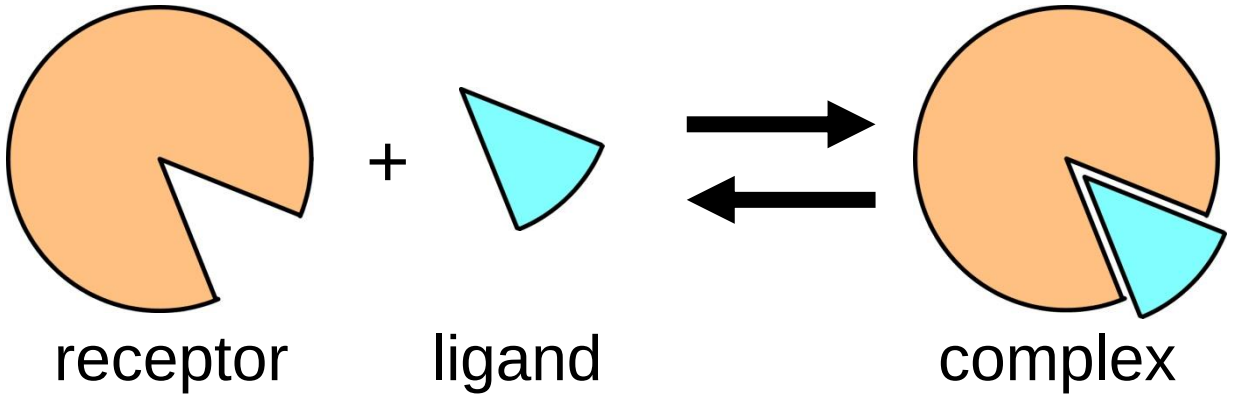
1. Start Discovery Studio 3.0 Client.
2. Choose “File”→“Open URL” from the menu, set ID to “1GUA” (complex of Rap1A and Ras binding domain of Raf-1), and click “Open.”
3. Choose “File”→“Insert From”→“URL”, set ID to “5P21” (Ras), and click “Open.”
4. Click “Macromolecules” button, expand “Align Sequences and Structures”, click “Align Structures” in the Align by Structure Similarity section, and click “Run.”

Structure superposition (2)

1. Result is displayed in a new Molecule Window. Change Display Style to Line.
2. Hide Rap1A structure.
3. Choose "Structure"→ "Monitor"→ "Intermolecular Bumps" to display bumps between proteins.

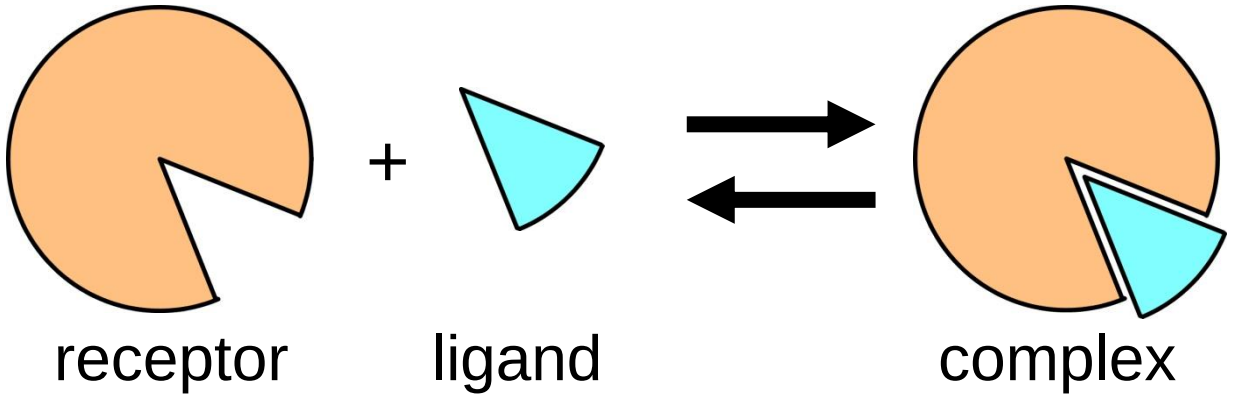


Docking simulation



- Dock a ligand into ligand-binding site on the surface of a receptor protein.
- Different methods are used depending on the type of ligand (protein or small molecule).

Binding free energy



$$G^{\circ}_{\text{complex}} - \left(G^{\circ}_{\text{receptor}} + G^{\circ}_{\text{ligand}}\right) + RT \ln \frac{[\text{complex}]}{[\text{receptor}][\text{ligand}]}$$

$$\Delta G^{\circ}_{\text{bind}} - RT \ln K_{\text{D}} = 0$$

$$K_{\text{D}} = \exp\left(\Delta G^{\circ}_{\text{bind}} / RT\right)$$

Binding free energy is related with dissociation constant.

Components of binding free energy

- Free energy is the sum of potential energy, volume-dependent term, and entropy-dependent term.
  - Receptor-ligand interaction: $\Delta E_{\text{int}} < 0 \rightarrow$stabilizing
  - Desolvation: $\Delta E_{\text{desolv}} > 0 \rightarrow$destabilizing
  - Restriction on the conformational flexibility: $\Delta S_{\text{conf}} < 0 \rightarrow$destabilizing
  - Release of bound water: $\Delta S_{\text{wat}} < 0 \rightarrow$stabilizing

$$G = E + PV - TS$$

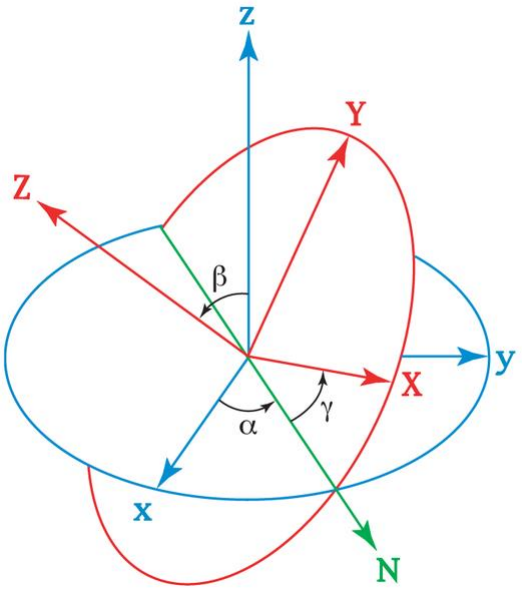
$$\Delta G^{\circ}_{\text{bind}} \approx \Delta E - T\Delta S = \Delta E_{\text{int}} + \Delta E_{\text{desolv}} - T\left(\Delta S_{\text{conf}} + \Delta S_{\text{wat}}\right)$$

Calculation of binding free energy

- Energy method
  - Considers only change in potential energy.
  - Ignores effects of solvation and conformational entropy.
- MM-PB/SA method
  - Calculates the free energy from potential energy, solvation energy derived from Poisson-Boltzmann equation and surface area model, and conformational entropy obtained from vibrational analysis.
- Free-energy perturbation method
  - Calculates free-energy change by the substitution of a functional group.
  - Gives an accurate result only when the structural difference caused by the substitution is very small.
- Score function

Protein-protein docking

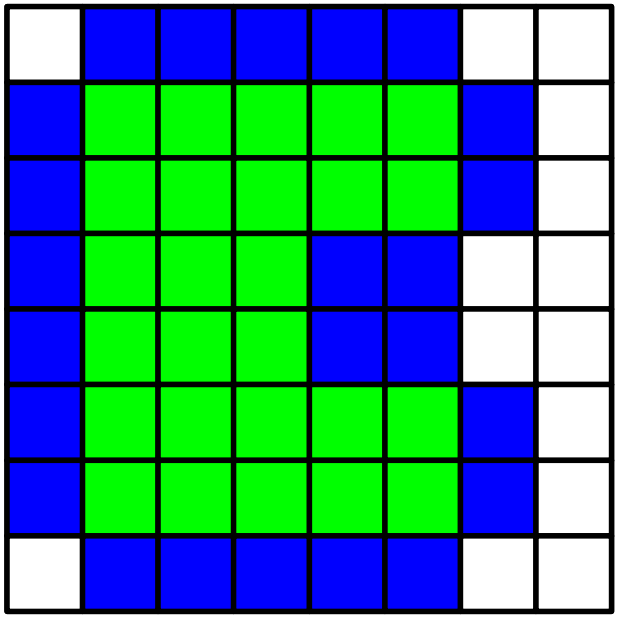
- Both of receptor and ligand are treated as rigid bodies. Conformational changes upon complex formation are not considered.
- Three translational and three rotational degrees of freedom of ligand are considered.
  - Rotation is described with Euler angle.
- Shape complementarity is important.



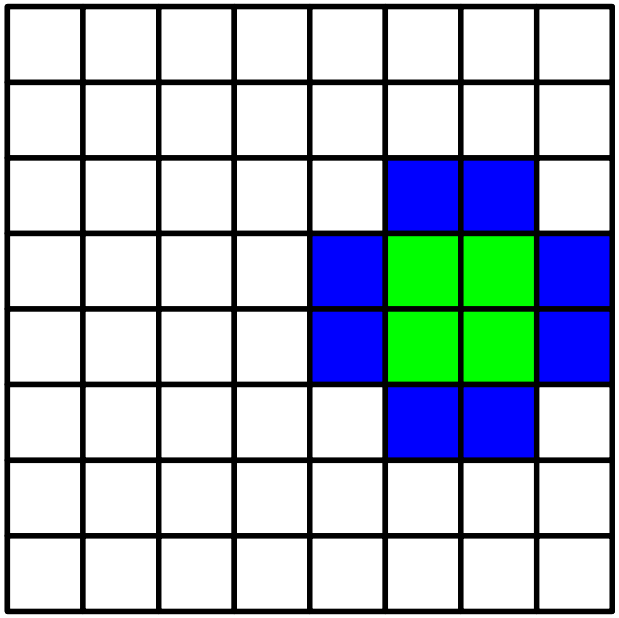
http://en.wikipedia.org/wiki/Euler_angles

Shape complementarity (1)

Receptor



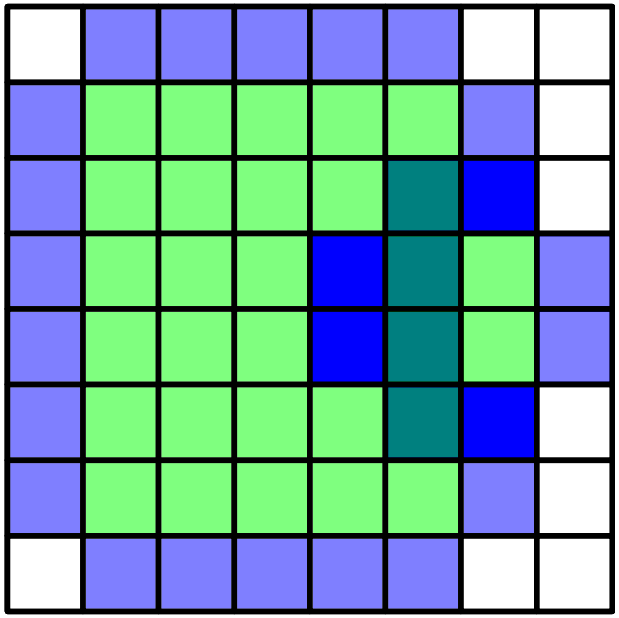
Ligand



= 1 (solvent accessible surface layer)

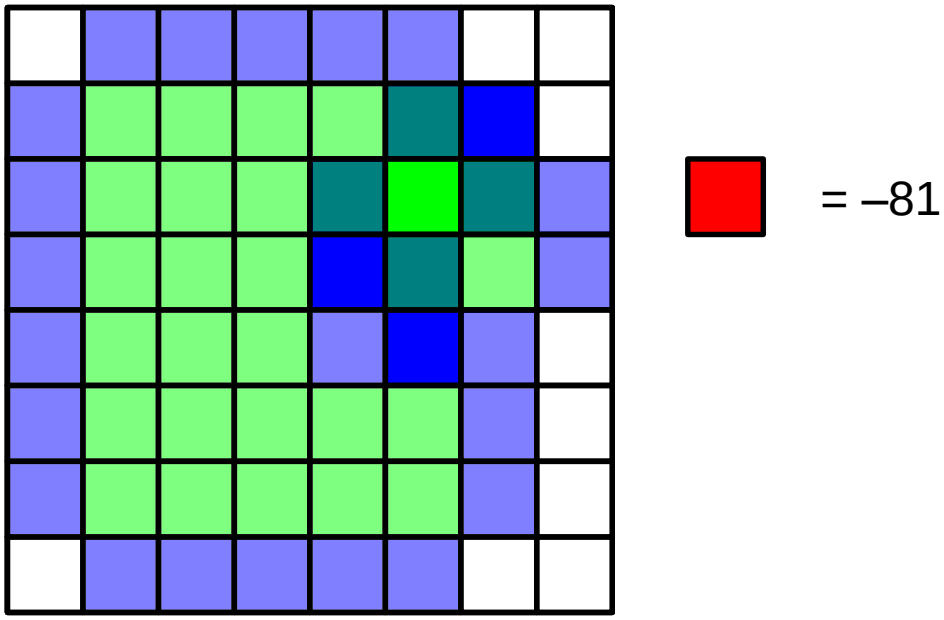
= 9*i* (solvent excluding surface layer)

Shape complementarity (2)



Calculate product of scores for each grid.
Real part of sum of products = Docking score = 4

Shape complementarity (3)



Calculate product of scores for each grid.

Real part of sum of products = Docking score = 3–81 = –78

Efficient calculation

- Generalization

  $$S(a,b,c) = \sum_{x,y,z} f(x,y,z)\,g(x+a, y+b, z+c)$$

  Find ligand position (*a*, *b*, *c*) that maximizes *S*.
- *S* can be efficiently calculated with fast Fourier transform (FFT).

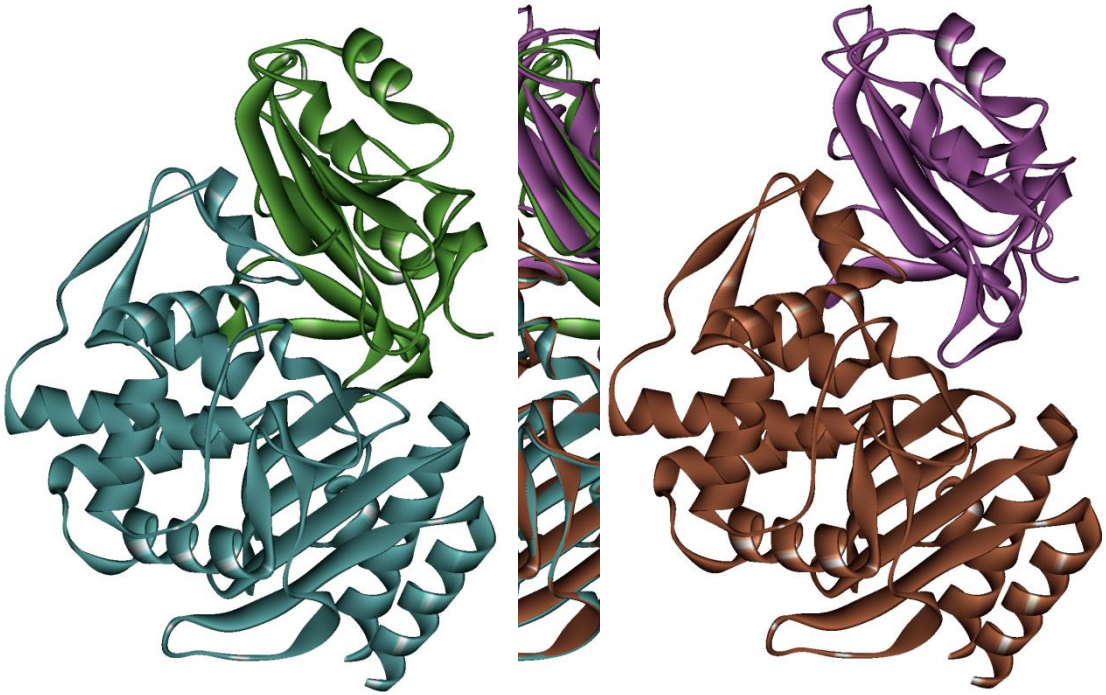
  $$\tilde{S}(h,k,l) = \tilde{f}(h,k,l)\,\tilde{g}(h,k,l)$$
- *S* is calculated for different ligand orientation.
- It is possible to calculated electrostatic and other interactions in a similar manner.

Docking software

- FTDock
  http://www.sbg.bio.ic.ac.uk/docking/ftdock.html
- ZDock
  http://zlab.bu.edu/zdock/index.shtml
- HEX
  http://www.loria.fr/~ritchied/hex/
- DOT
  http://www.sdsc.edu/CCMS/DOT/
- GRAMM-X
  http://vakser.bioinformatics.ku.edu/resources/gramm/grammx

An application of ZDock

- Complex of TEM-1 β-lactamase and its inhibitor
  - β-lactamase: 1ZG4 (receptor)
  - Inhibitor: 3GMU (ligand)



Top ranked model

Experiment (1JTG)

Protein-small molecule docking

- Find the ligand-binding site on the surface of the receptor protein. Then, dock the ligand into the site.
- Search the conformational space of ligand for the free-energy minimum “pose” by translating and rotating the ligand and rotating all the rotatable bonds in the ligand.
- Usually, the receptor atoms are not moved. The receptor is treated as a rigid body.

Empirical score function (1)

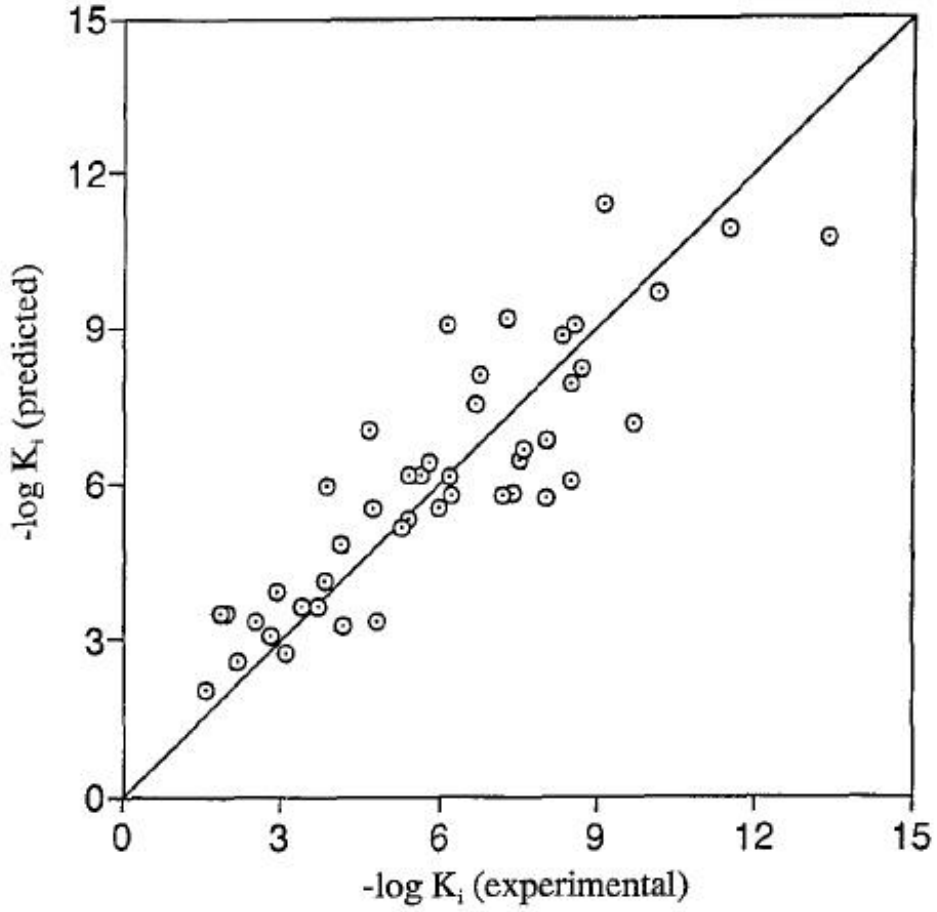
- Ludi

$$\Delta G^{\circ}_{\text{bind}} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{\text{h-bonds}} f(\Delta R, \Delta\alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionicint.}} f(\Delta R, \Delta\alpha) + \Delta G_{\text{lipo}} A_{\text{lipo}} + \Delta G_{\text{rot}} N_{\text{rot}}$$

  - Binding free energy change is expressed as the sum of the hydrogen-bond term, ionic-interaction term, lipophilic-interaction term and the loss of free energy due to freezing of internal degrees of freedom in the ligand.
  - Coefficients ΔG_x were determined by fitting the calculated free-energy values to the experimental data of 45 protein-small molecule complexes.

Böhm (1994) *J. Comput.-Aided Mol. Des.* **8**, 243.

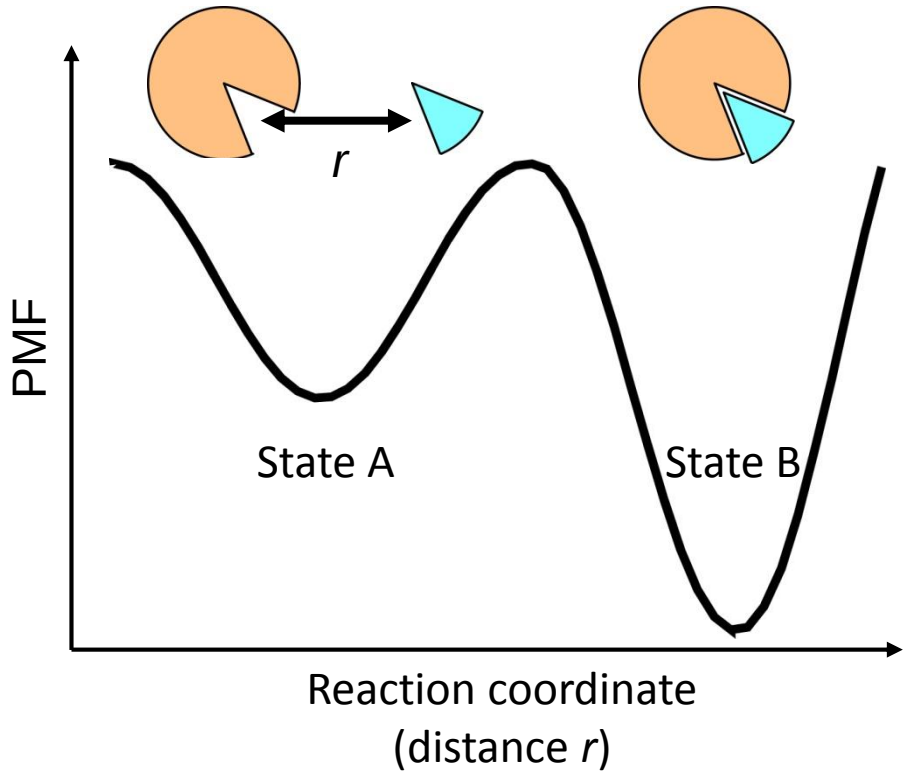
Empirical score function (2)



Böhm (1994) *J. Comput.-Aided Mol. Des.* **8**, 243.

Statistical potential

- Potential of mean force（Pmf）
  - Plot of free-energy along the reaction coordinate.



$$\Delta G^{\circ}_{\mathrm{bind}} + RT \ln \frac{[\mathrm{B}]}{[\mathrm{A}]} = 0$$

$$\Delta G^{\circ}_{\mathrm{bind}} = -RT \ln \frac{[\mathrm{B}]}{[\mathrm{A}]} = -RT \ln p_{\mathrm{B}} / p_{\mathrm{A}}$$

Statistical potential

- Potential of mean force（Pmf）
  - Plot of free-energy along the reaction coordinate.
  - Related with probability distribution function.
  - Probability distribution as a function of the distance between protein and ligand atoms, $p_{ij}(r)$, was calculated for each pair of atom types *i* and *j* using 77 complex structures.

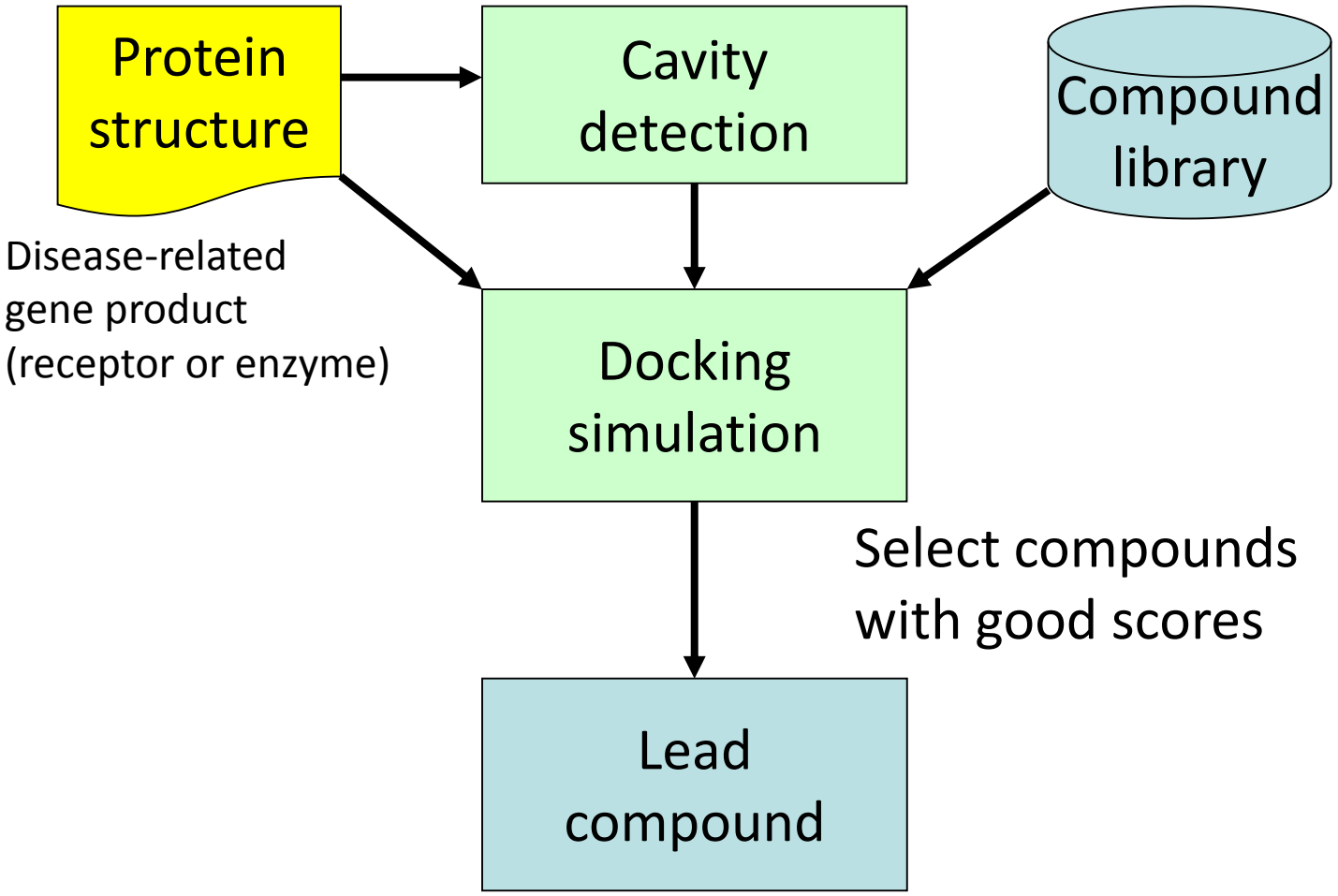
$$\Delta G^{\circ}_{\mathrm{bind}}(r) = -RT \ln p(r) / p_{\mathrm{bulk}}(r) \approx -RT \sum_{k,l} \ln p_{ij}(r_{kl}) / p^{ij}_{\mathrm{bulk}}(r)$$

Muegge & Martin (1999) *J. Med. Chem.* **42**, 791.

Application to drug discovery

- In drug discovery, high-throughput screening (HTS) is used to efficiently and exhaustively search the compound library for drug candidates that tightly bind to the target protein.
- It costs huge amount of money to establish the compound library and binding-assay system.
- It is possible to evaluate the affinity of a ligand to the protein by docking simulation.→virtual screening

Virtual screening



Compound library

- Available Chemicals Directory (ACD)
  - Database of commercially available compounds.
  - http://accelrys.com/products/databases/sourcing/available-chemicals-directory.html
  - Includes about 3,870,000 compounds.
- ZINC
  - 'Ready-to-dock' 3D-structure database provided by USCF.
  - http://zinc.docking.org/
  - Includes about 21,000,000 compounds.
- PubChem
  - Provided by NCBI.
  - http://pubchem.ncbi.nlm.nih.gov/
  - Includes about 57,000,000 compounds.

Cavity detection

- A ligand binds to the cavity on the surface of a protein.
- SURFNET
  - http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html
  - Detects “gap regions” on the protein surface.
- PASS
  - http://www.ccl.net/cca/software/UNIX/pass/overview.shtml
  - Detects cavities on the protein surface and ranks them.
- Q-SiteFinder
  - http://www.bioinformatics.leeds.ac.uk/qsitefinder/
  - Detects cavities on the protein surface and ranks them based on the interaction energy with CH_4 probe.

Docking software

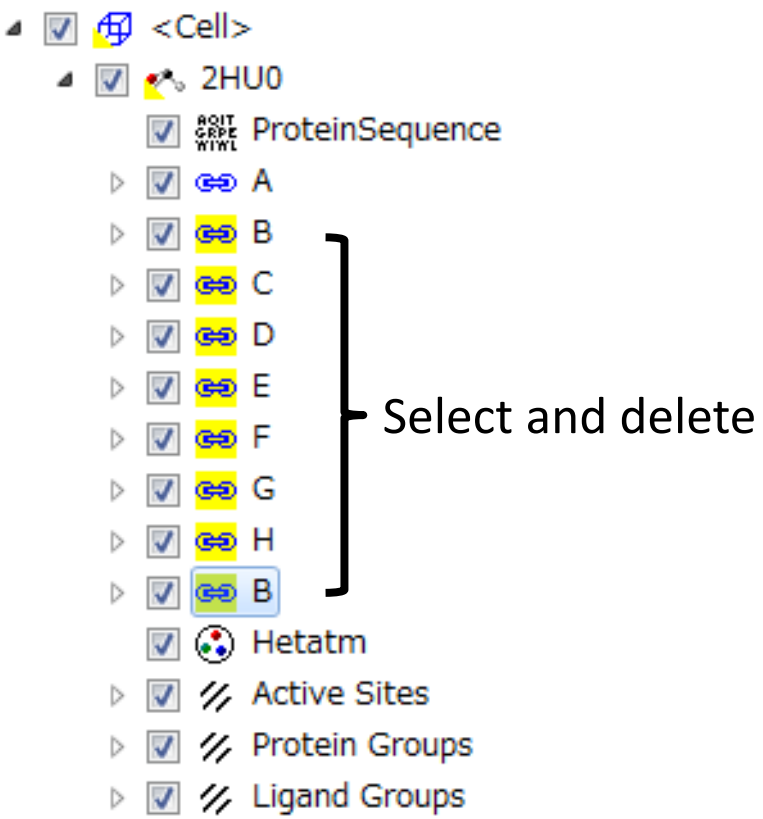
- DOCK
  - http://dock.compbio.ucsf.edu/
  - Matches ligand atoms with spheres that represent the cavity.
- AutoDock
  - http://autodock.scripps.edu/
  - Optimizes empirical free-energy score with genetic algorithm (GA).
- GOLD
  - http://www.ccdc.cam.ac.uk/products/life_sciences/gold/
  - Optimizes score function with GA.
- Only the translational, rotational, and torsional degrees of freedom of the ligand are considered and the flexibility of the protein is not considered.

Practice of docking simulation

- Dock an inhibitor to N1 neuraminidase using Discovery Studio 3.0 Client.
  1. Obtain crystal structure of N1 neuraminidase.
  2. Detect cavity.
  3. Obtain structure data of the inhibitor.
  4. Perform docking simulation.
  5. Analyze the result.

1. Structure of receptor

1. Open the structure of N1 neuraminidase (PDB ID: 2HU0).
   - The B chain in this structure binds oseltamivir (trade name: Tamiflu).
2. Select B–H chains and delete them.
3. Change Display Style to Line.

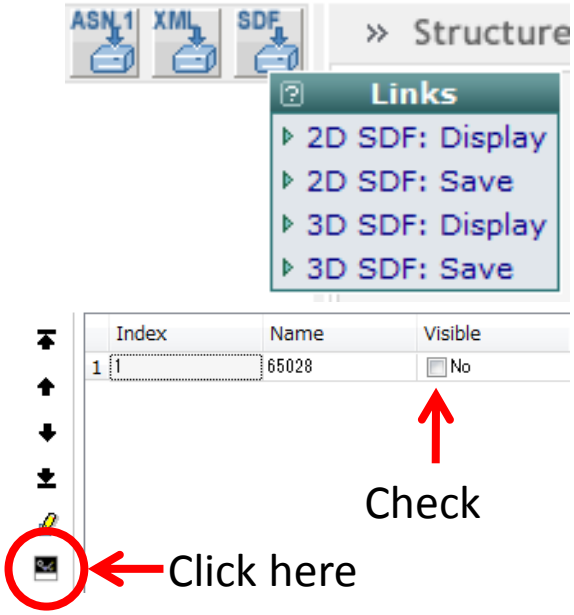


2. Cavity detection

1. Apply “charmm27” force field to the protein.
2. Click “Receptor-Ligand Interactions” button, expand “Define and Edit Binding Site”, and click “Define Receptor: 2HU0.”
3. Click “From Receptor Cavities” in the Define Site section.→Cavities are displayed.

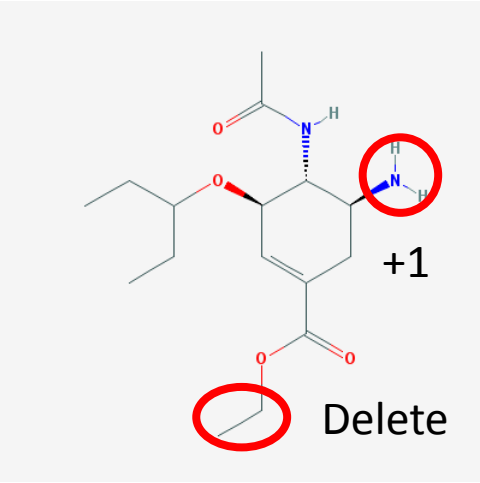
3. Structure of ligand (1)

1. Access PubChem (http://pubchem.ncbi.nlm.nih.gov/), enter “oseltamivir” in the query box and click “GO.”
2. Click the hit with CID 65028.
3. Save the structural data in 3D SDF on Desktop.
4. Open it with Discovery Studio 3.0.
5. Change the molecule’s name to “oseltamivir” in Molecule tab of Data Table.



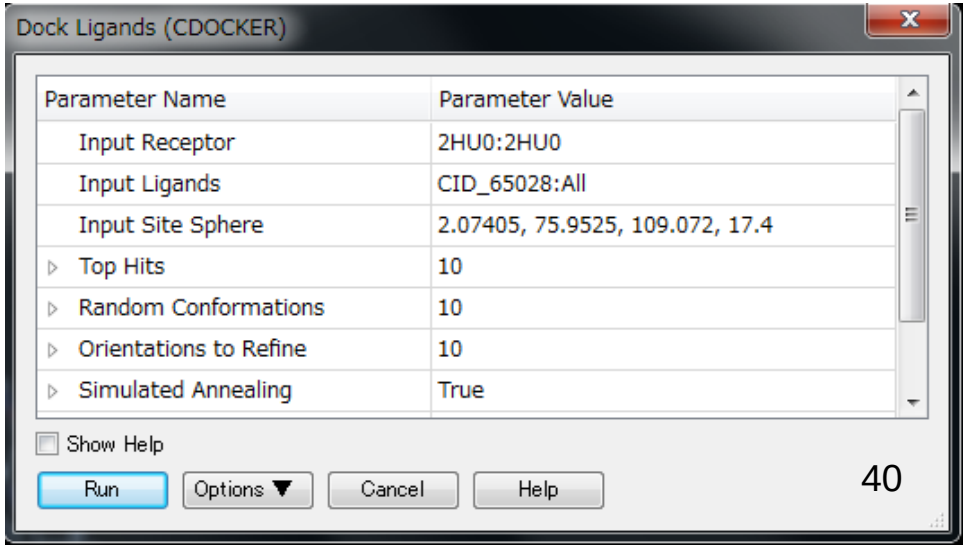
3. Structure of ligand (2)

6. Since the ethyl group is removed in the liver, delete it from the structure.
7. Select atoms in the carboxyl group, and choose “Chemistry”→“Bond”→ “Partial Double” from the menu.
8. Select the nitrogen atom of the NH_2 group and choose “Chemistry”→ “Charge”→“+1” to change the charge to +1. (A hydrogen atom is automatically added.)
9. Apply “CHARMm” force field to the molecule.
10. Expand “Run Simulations”, click “Minimization”, and click “Run.”



4. Docking simulation

1. Activate the Molecule Window in which 2HU0 is displayed.
2. Click “Receptor-Ligand Interactions” button, expand “Dock Ligands”, and click “Dock Ligands (CDOCKER)” in Docking Optimization section.
3. Set Input Receptor, Input Ligands as shown here and click “Run.”



CDOCKER

- Developer
  - C. L. Brooks III, M. Vieth, et al.
  - Wu *et al. J. Comput. Chem.* **24**, 1549 (2003).
- Potential energy function
  - CHARMm
- Optimization method
  - Simulated annealing (SA) and energy minimization
  - In SA, the interaction energy is evaluated with a grid-based method.
  - In energy minimization, interaction energy is calculated by using the potential energy function.

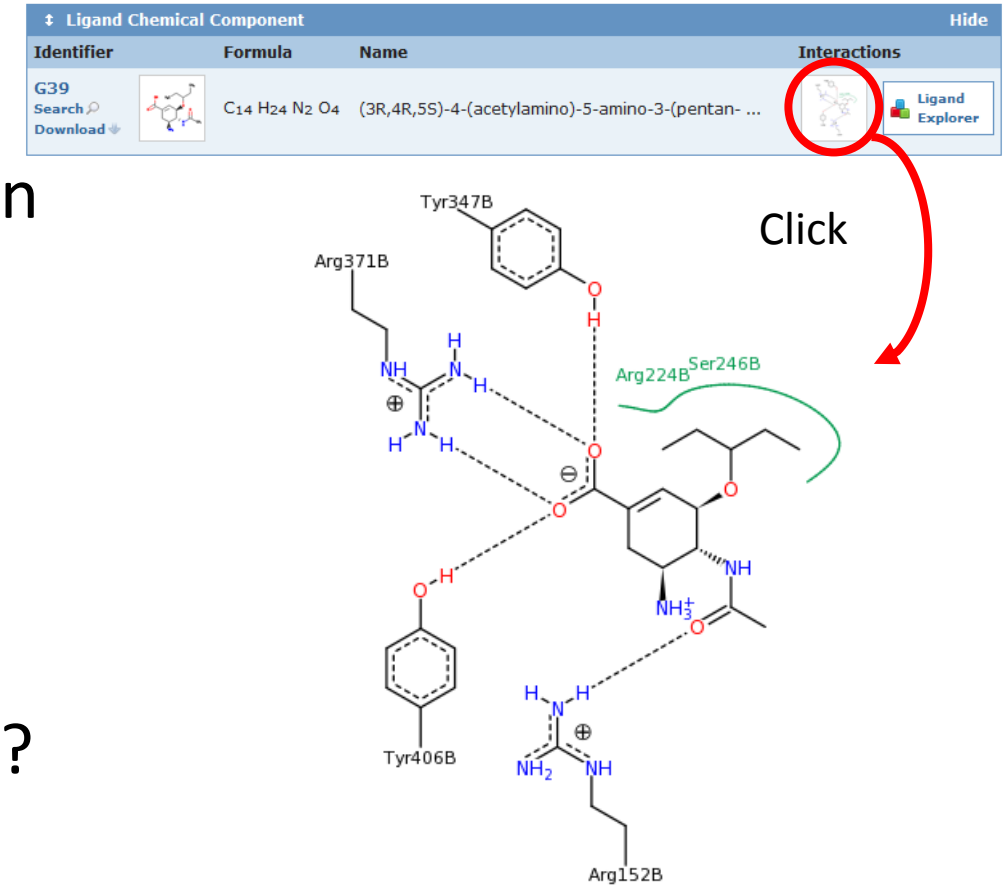
5. Analysis of the result

1. When the calculation has finished, the result is shown in a new Molecule Window. Uncheck Visibility Locked of 2HU0 in Data Table.
2. Hide all the binding site indicators (Site 1–11).
3. Choose “Chemistry”→“Hydrogens”→“Hide.”
4. Docking poses are sorted in the descending order of –CDOCKER_ENERGY values below the second raw of Data Table.

	Index	Name	Visible	Tagged	Visibility Locked
1	1	2HU0	☑ Yes	☐ No	☐ No
2	2	oseltamivir	☑ Yes	☐ No	☐ No
3	3	oseltamivir	☐ No	☐ No	☐ No
4	4	oseltamivir	☐ No	☐ No	☐ No

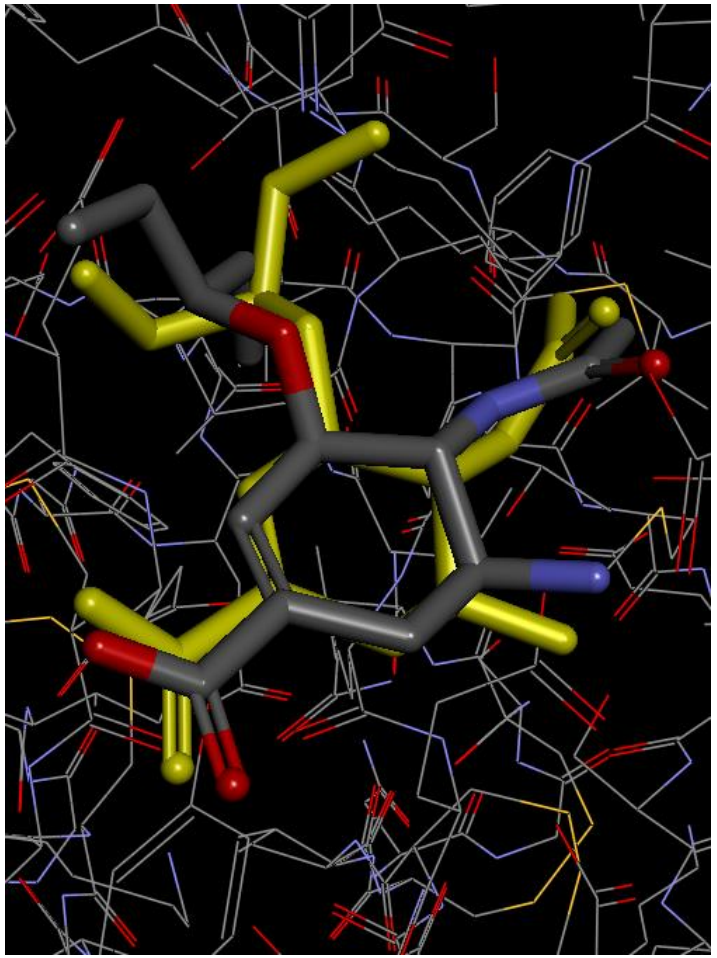
Comparison with experiment (1)

- Interactions between the ligand and the protein are illustrated in the Summary page of 2HU0 at the RCSB site.
- Which pose shows similar interactions to those in the experimental structure?



Comparison with experiment (2)

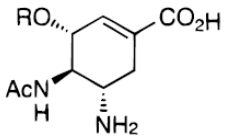
- Since the B chain of 2HU0 binds oseltamivir, the pose is directly compared with the experimental one by superimposing the B chain on the receptor protein.
- The fifth-ranked pose is very close to the experimental one.
- Note that the energy difference between the top-ranked and fifth-ranked poses is small.



Exercise

- This table lists the activity of the analogues tested during the development of oseltamivir. (Oseltamivir acid is **6h**.)
- Dock one of the analogues to N1 neuraminidase.
- Discuss the difference in the docking pose and the energy.

Table 1. Influenza Neuraminidase Inhibition and Plaque Reduction by Carbocylic Analogues



6

R	compd	enzyme[a] IC_{50} (nM)	plaque[b] EC_{50} (nM)
H	**8**	6300	ND[c]
CH_3	**6a**	3700	ND
CH_3CH_2	**6b**	2000	
$CH_3CH_2CH_2$	**6c**	180	ND
$CH_3CH_2CH_2CH_2$	**6d**	300	ND
$(CH_3)_2CHCH_2$	**6e**	200	ND
$CH_3CH_2(CH_3)CH$*	**6f**	10	80
	(*R*)-isomer		
	6g	9	135
	(*S*)-isomer		
$(CH_3CH_2)_2CH$	**6h**	1	16
$(CH_3CH_2CH_2)_2CH$	**6i**	16	ND
	2	150	2500
	3	1	15

[a] NA. [b] H1N1, A/ws. [c] ND = not determined.

Kim *et al. J. Am. Chem. Soc.* **119**, 681 (1997).

The State of molecular simulation

- Feasible
  - Folding simulation of a small protein
  - Refinement of accurate models
  - Reproduction of thermal fluctuation and fast (up to microseconds order) motions
- Difficult
  - Folding simulation of a large protein
  - Refinement of inaccurate models
  - Reproduction of slow motions
  - Cell-scale simulation

Time scale of protein dynamics

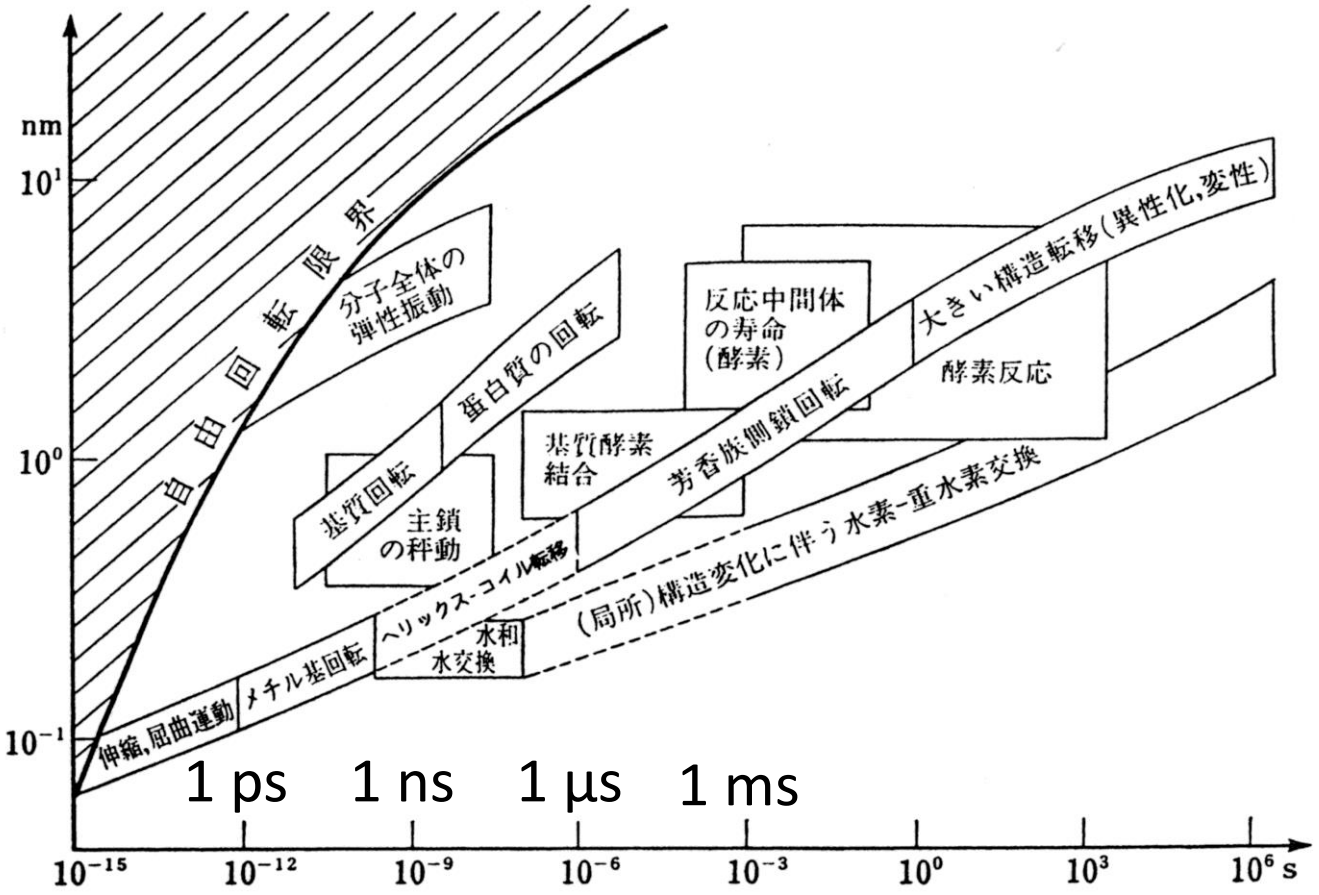
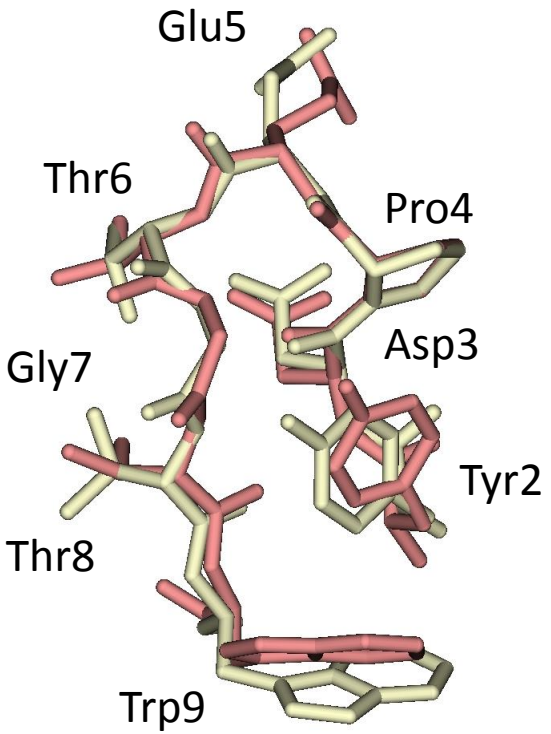


図 4. 17 タンパク質分子の構造変化や化学反応事象の時間-空間相関プロット

時間軸と空間軸のケタ数の違いに注目.

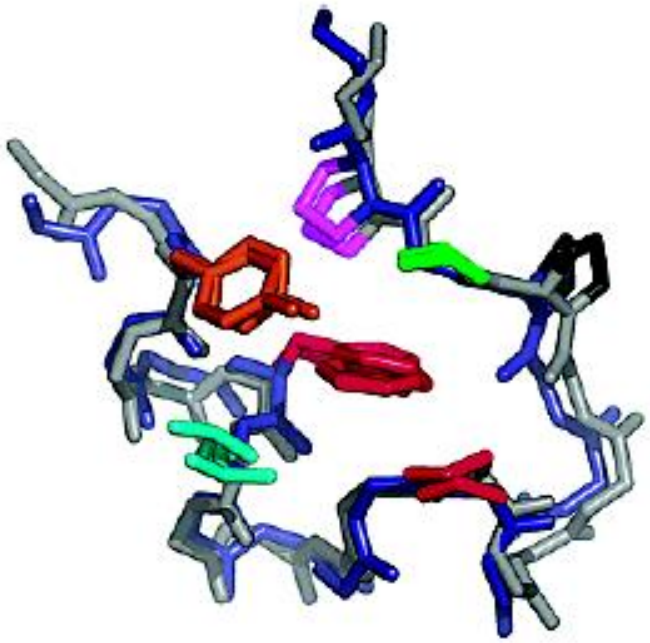
永山國昭 「生命と物質　生物物理学入門」より引用

Folding simulations



Yellow: NMR, Pink: Simulation

Satoh *et al. FEBS Lett.* **580**, 3422 (2006).

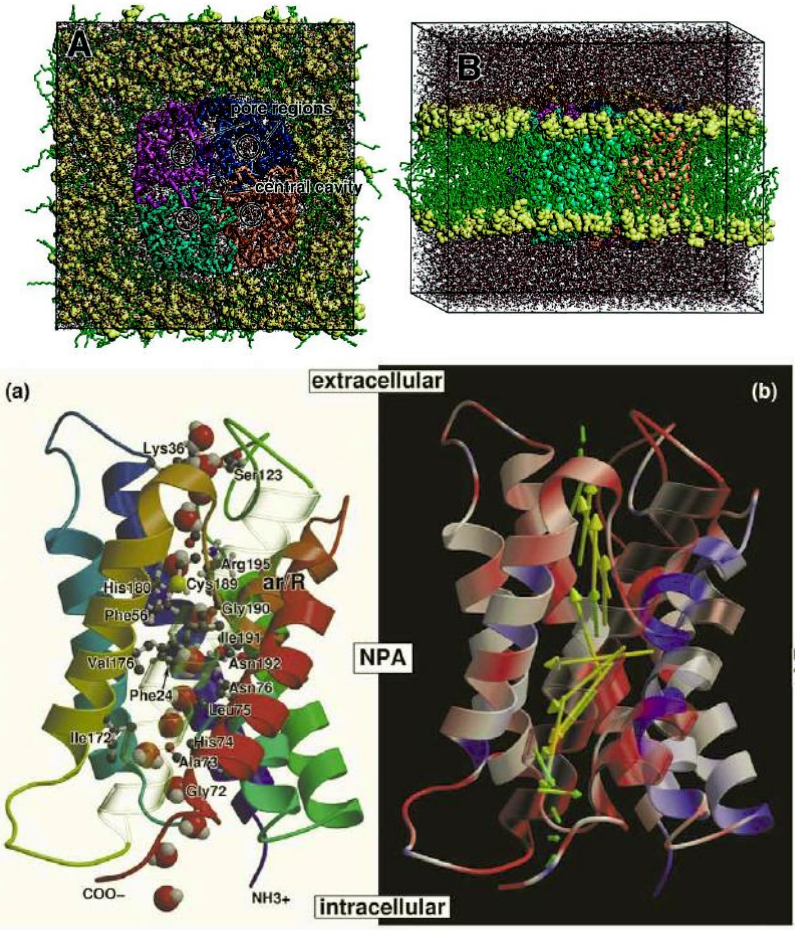


Gray: NMR, Blue: Simulation

Simmerling *et al. J. Am. Chem. Soc.* **124**, 11258 (2002).

An MD simulation of Aquaporin

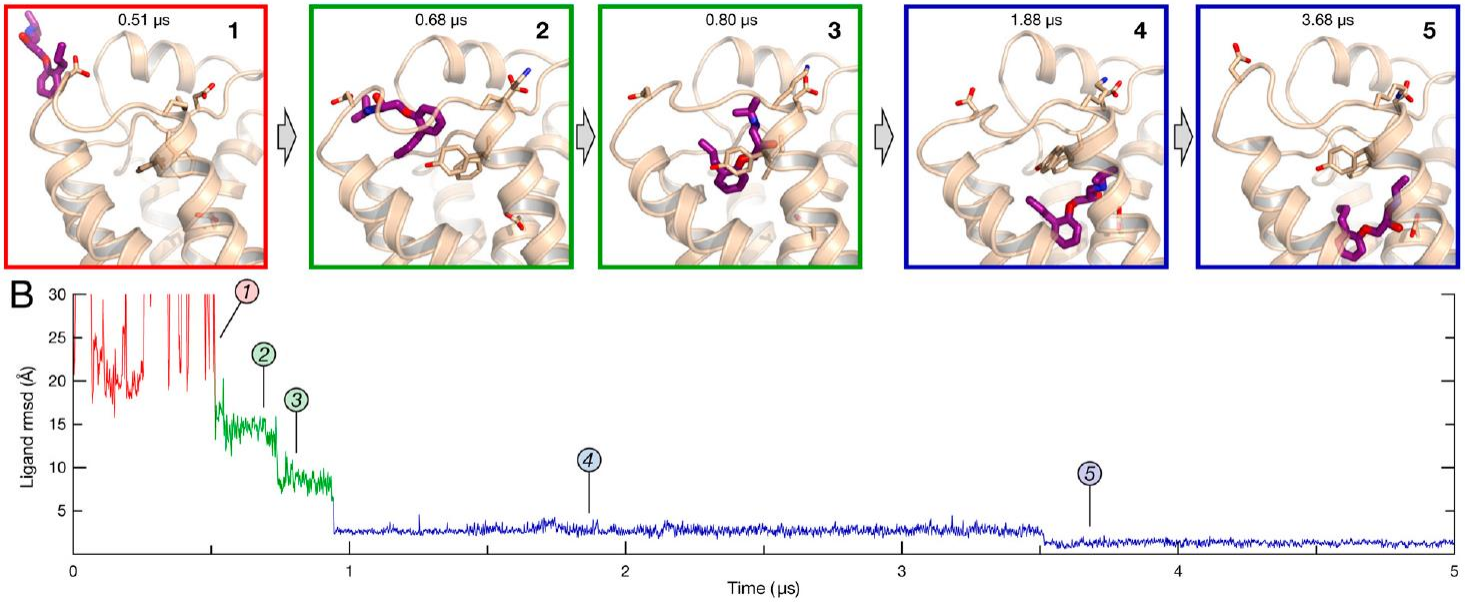
- The protein is embedded in a lipid bilayer and water molecules are arranged on both sides of the membrane.
- Water permeation rate
  Expt.: 3×10^9 sec^{-1}
  Simulation: 16 / 10 ns
  $\rightarrow 1.6 \times 10^9$ sec^{-1}



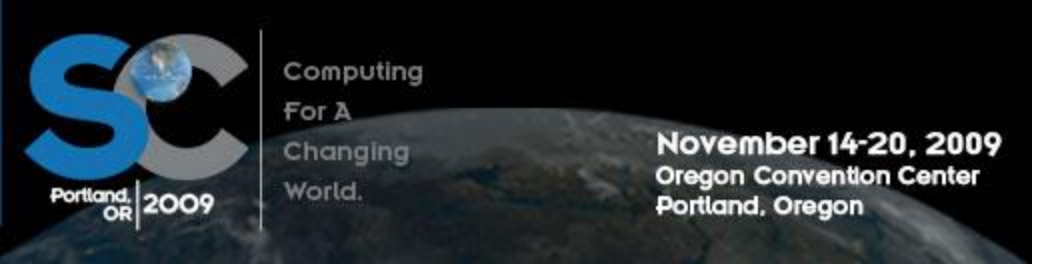
de Groot & Grubmuller *Science* **294**, 2353 (2001).
de Groot & Grubmüller *Curr. Opin. Struct. Biol.* **15**, 176 (2005).

Ligand-binding simulation

- MD simulations of binding of the beta-blocker drugs, alprenolol, etc., to its receptor, β_2-adrenergic receptor.
- Binding rate constant
  - Experiment: $1.0 \times 10^7\ M^{-1}\ s^{-1}$
  - Simulation: $3.1 \times 10^7\ M^{-1}\ s^{-1}$



Dror *et al. PNAS* **108**, 13118 (2011).



http://sc09.supercomputing.org/

Millisecond-Scale Molecular Dynamics Simulations on Anton

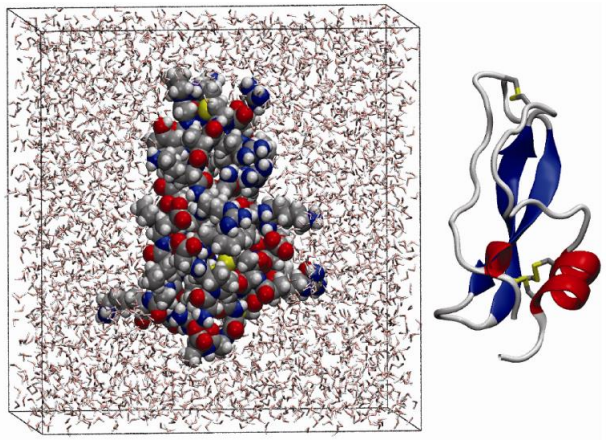
David E. Shaw*, Ron O. Dror, John K. Salmon, J.P. Grossman, Kenneth M. Mackenzie, Joseph A. Bank, Cliff Young, Martin M. Deneroff, Brannon Batson, Kevin J. Bowers, Edmond Chow, Michael P. Eastwood, Douglas J. Ierardi, John L. Klepeis, Jeffrey S. Kuskin, Richard H. Larson, Kresten Lindorff-Larsen, Paul Maragakis, Mark A. Moraes, Stefano Piana, Yibing Shan, and Brian Towles

D. E. Shaw Research, New York, NY 10036, USA

* *Correspondence:* David.Shaw@DEShawResearch.com

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.



Shaw's approach

- They developed a special purpose system for MD simulation named Anton.
- They can conduct a MD simulation of 23,558-atom system at the speed of 16.4 μs per day using 512 Anton nodes.
- The simulation speed of a PC cluster is at most 100 ns per day.

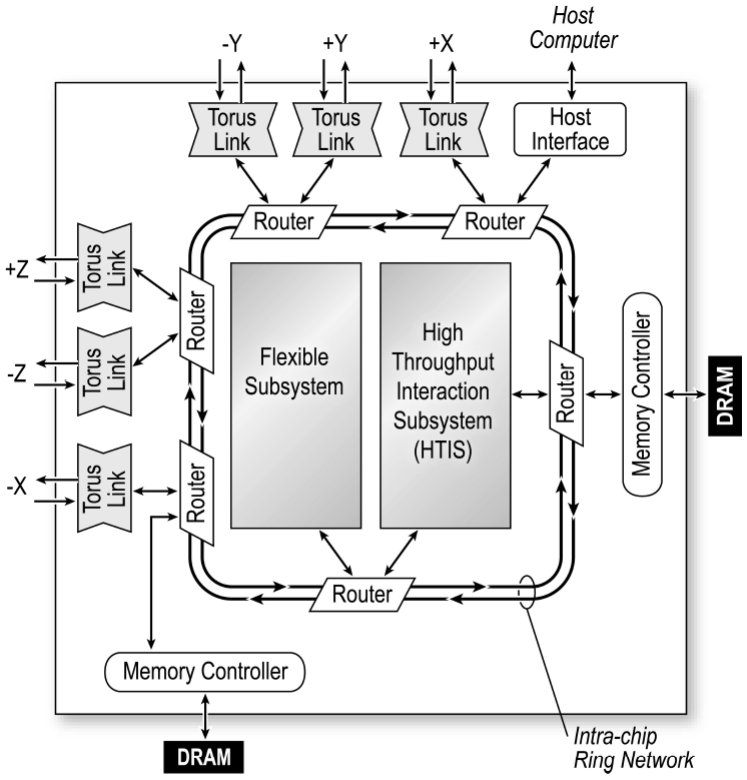
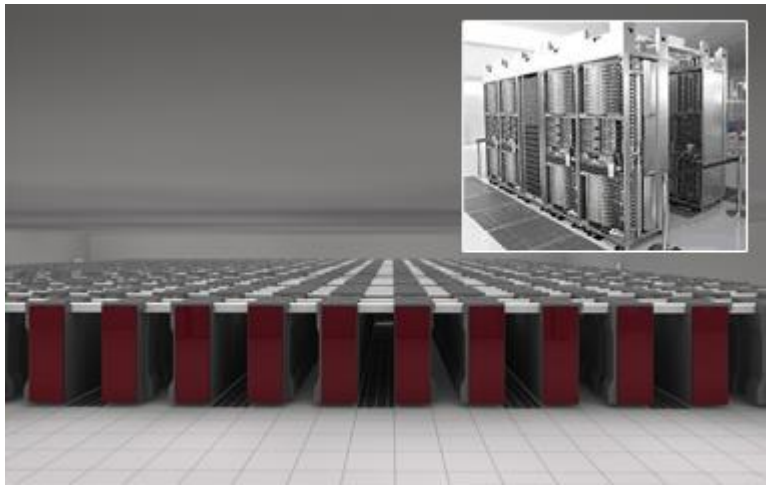
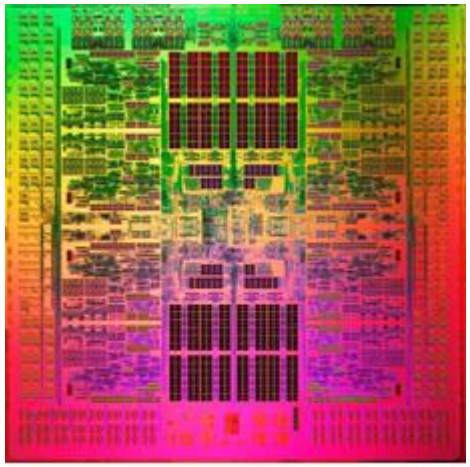


Figure 2: Anton ASIC block diagram.

The K supercomputer

- Shared use starts on October. (http://www.aics.riken.jp)
- It has more than 80,000 Fujitsu CPUs capable of performing 1.28 $\times 10^{11}$ floating point calculations per second (128 GFLOPS), and can perform 10^{16} floating point calculations per second (10 PFLOPS) in total.





SPARC64™ VIIIfx

http://jp.fujitsu.com/about/tech/k/

Accuracy of force field parameters

Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain

Peter L. Freddolino,*† Feng Liu,* Martin Gruebele,*†‡§ and Klaus Schulten†‡
*Center for Biophysics and Computational Biology, †Beckman Institute, ‡Department of Physics, and §Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

ABSTRACT All-atom molecular dynamics (MD) simulations of protein folding allow analysis of the folding process at an unprecedented level of detail. Unfortunately, such simulations have not yet reached their full potential both due to difficulties in sufficiently sampling the microsecond timescales needed for folding, and because the force field used may yield neither the correct dynamical sequence of events nor the folded structure. The ongoing study of protein folding through computational methods thus requires both improvements in the performance of molecular dynamics programs to make longer timescales accessible, and testing of force fields in the context of folding simulations. We report a ten-microsecond simulation of an incipient downhill-folding WW domain mutant along with measurement of a molecular time and activated folding time of 1.5 microseconds and 13.3 microseconds, respectively. The protein simulated in explicit solvent exhibits several metastable states with incorrect topology and does not assume the native state during the present simulations.

Further improvement is necessary.

Freddolino *et al. Biophys. J.* **94**, L75 (2008).

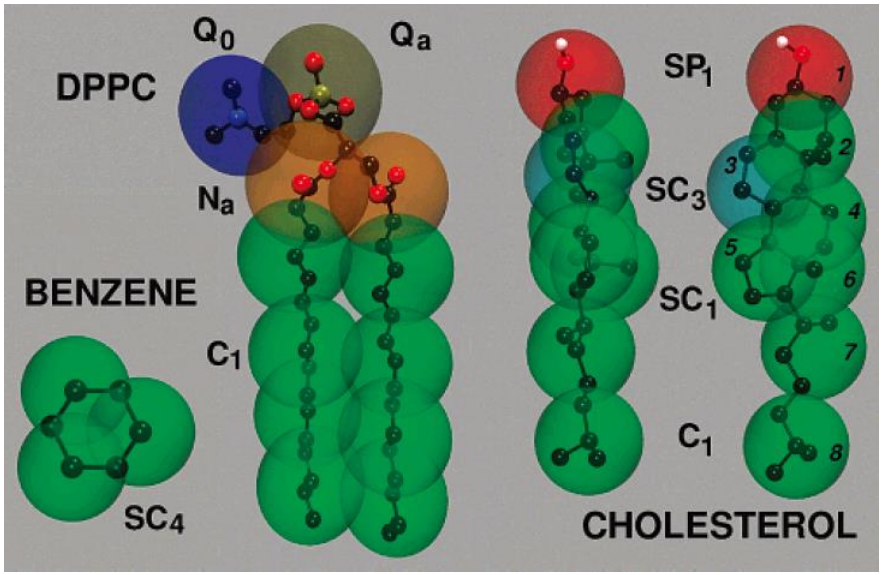
Coarse-grained (CG) model

- In the MD simulation, all of the details of the dynamics, including the bond-stretching motions, are reproduced.
- Such detailed information is not necessary.

- Coarse-graining of a molecule
  - Allows use of a longer time step.
  - Reduces the computational cost of the calculation of interaction.

MARITINI force field

- Developed by Marrink's group.
- Maps four non-hydrogen atoms into one particle.
- Force field parameters were determined so as to reproduce free energies of hydration, vaporization, and partitioning between water and organic phases.
- Time step is 30 fs. The effective time is 4-fold longer.

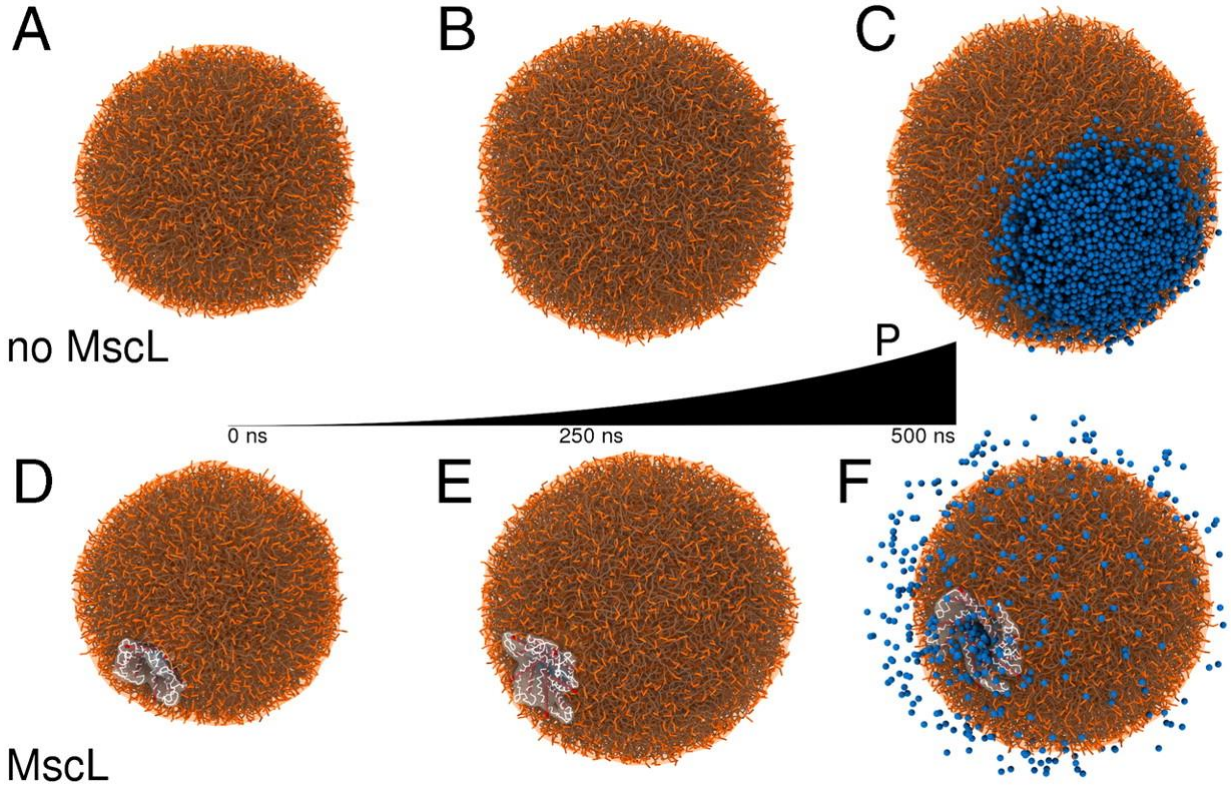


A simulation of lipid bilayer

- 128 DSPC (distearoyl-phosphatidylcholine) molecules are randomly arranged in a cube of edge length 77 Å.
- After energy minimization, 768 CG water particles, each of which corresponds to four water molecules, are arranged in the cube.
- With the time step of 30 fs, 900,000-step constant-*NPT* simulation (effective time of 108 ns) were performed at 323 K and at 1 bar.
- Download membrane.tpr, membrane.trr from the lecture's page. Visualize it with UCSF Chimera.
    - Choose "Tools"→"MD/Ensemble Analysis"→"MD Movie."
    - Set Trajectory format to "GROMACS", Run input (.tpr) to "membrane.tpr", and Trajectory (.trr) to "membrane.trr."
    - Click "OK."

A simulation of liposome

- Increase of the interior pressure causes the burst of a liposome.
- When a mechano-sensitive channel (MscL) is embedded in its membrane, water is released through the channel and the liposome does not burst.



Louhivuori *et al. PNAS* **107**, 19856 (2010).

Perspectives

- It will become possible to perform simulations for longer (milliseconds to seconds) time by further improvement of computer performance.
  - Further improvement of the accuracy of the potential energy function is necessary.
- It will become possible to perform cell-scale simulations by increased size of the computer.
  - Development of multi-scale methods that combine all-atom and coarse-grained models is necessary.

How to send your report

- Use PowerPoint to create your report.
- Report should include the results and discussion of the exercise.
- Send the PowerPoint file to tterada@iu.a.u-tokyo.ac.jp.
- Subject of the e-mail should be “Molecular modeling” and write your name and ID card number in the body of the e-mail.