Molecular Modeling and Simulation
On June 14, 2016

Complex Structure Modeling

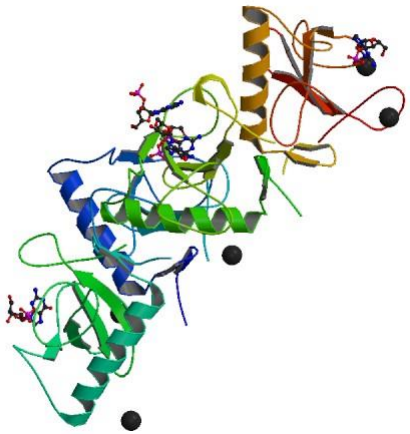
Agricultural Bioinformatics Research Unit,
Graduate School of Agricultural and Life Sciences,
The University of Tokyo
Tohru Terada

Today's topics

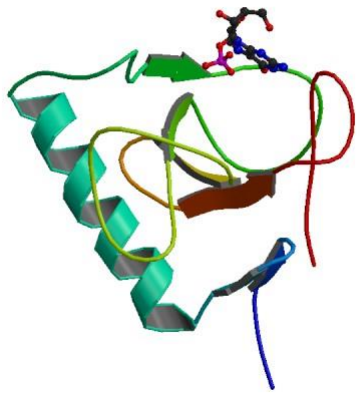
- Note for conducting simulations
- Protein-protein docking
- Protein-small molecule docking
  - Assignment
- Perspectives of molecular simulation

Note for conducting simulations (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

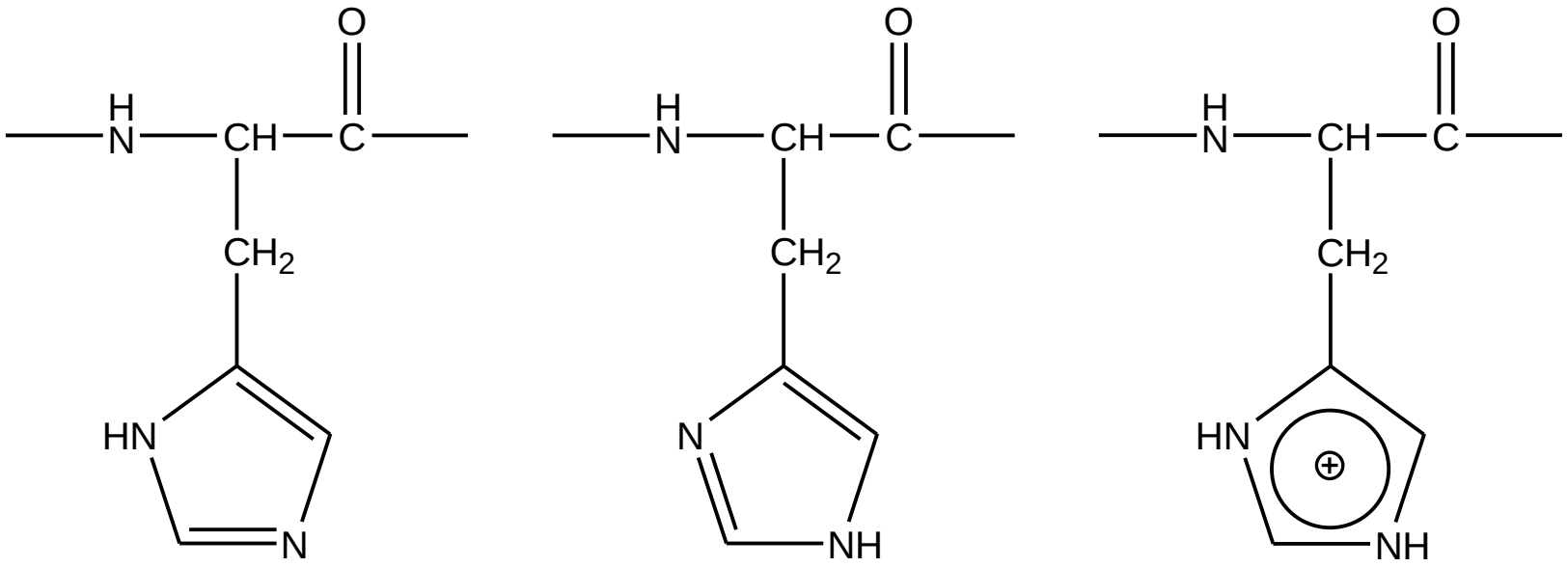


Biological Unit

Note for conducting simulations (2)

- Add missing atoms and residues 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.

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.

Note for conducting simulations (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 e.g. Amber Parameter Database.*
- Add counterions to neutralize the system, when using PME.
- The equilibration run should be sufficiently long.
  - The length should be at least 1 ns.
  - During the equilibration run, use position restraints to avoid large deviation from the initial structure and gradually reduce the restraining force.

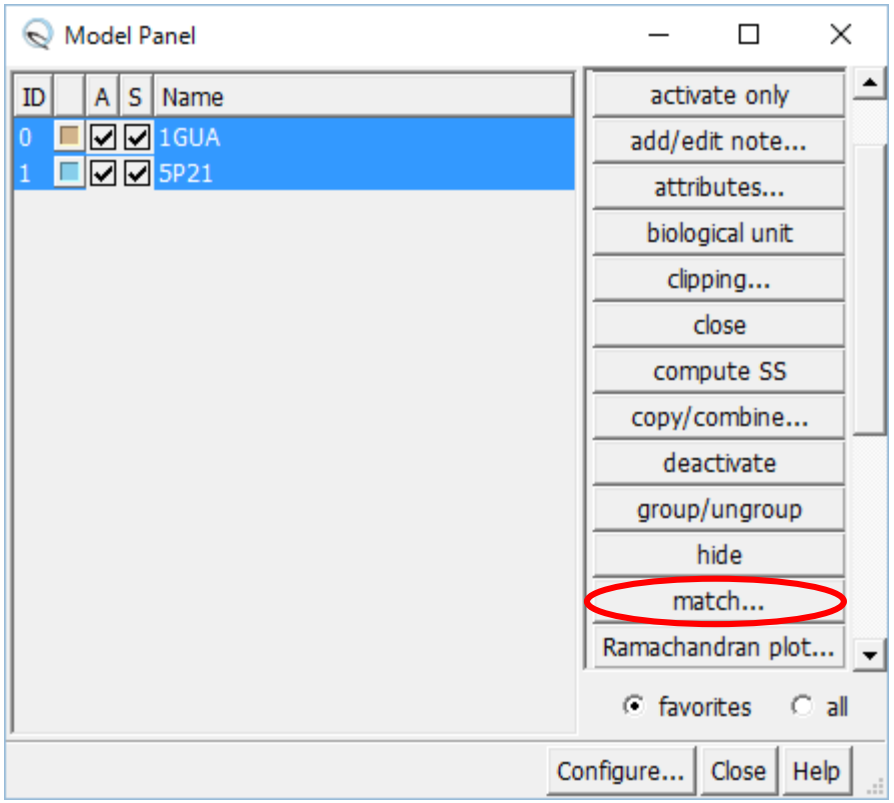
*http://www.pharmacy.manchester.ac.uk/bryce/amber

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 UCSF Chimera 1.10.1
2. Choose “File” → “Fetch by ID” to open the complex structure between Rap1A and the Ras-binding domain of Raf-1 with PDB ID of “1GUA.”
3. Open the uncomplexed structure of Ras with PDB ID of “5P21.”
4. Choose “Favorites” → “Model Panel,” select “1GUA” and “5P21,” and click “match” and “OK.”
   – RMSD value is displayed at the bottom of the main window.

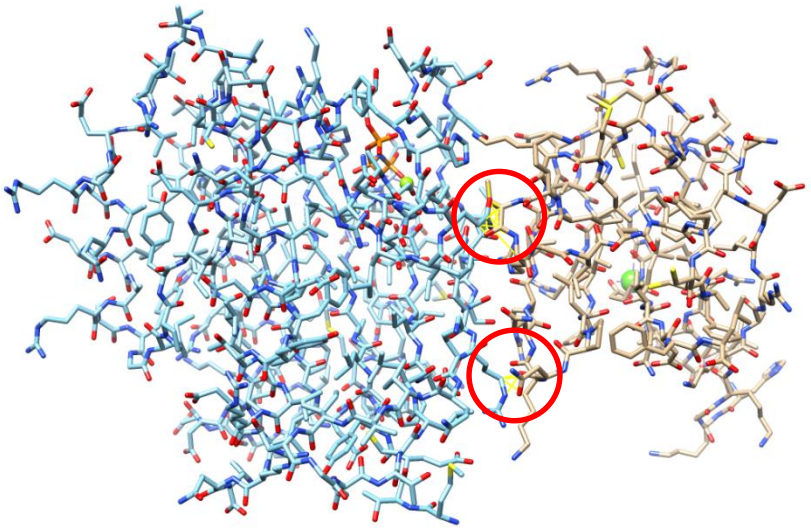


Structure superposition (2)

5. Choose “Select” → “Chain” → “A” → “1GUA” to select the Rap1A atoms and delete them by choosing “Actions” → “Atoms/Bonds” → “delete.”
6. Delete water molecules after selecting them by choosing “Select” → “Residue” → “HOH.”
7. Change the representation to sticks
8. Choose “Tools” → “Structure Analysis” → “Find Clashes/Contacts”
9. Choose “Select” → “Chain” → “A” to select the Ras atoms and click the “Designate” button in the Find Clashes/Contacts window.

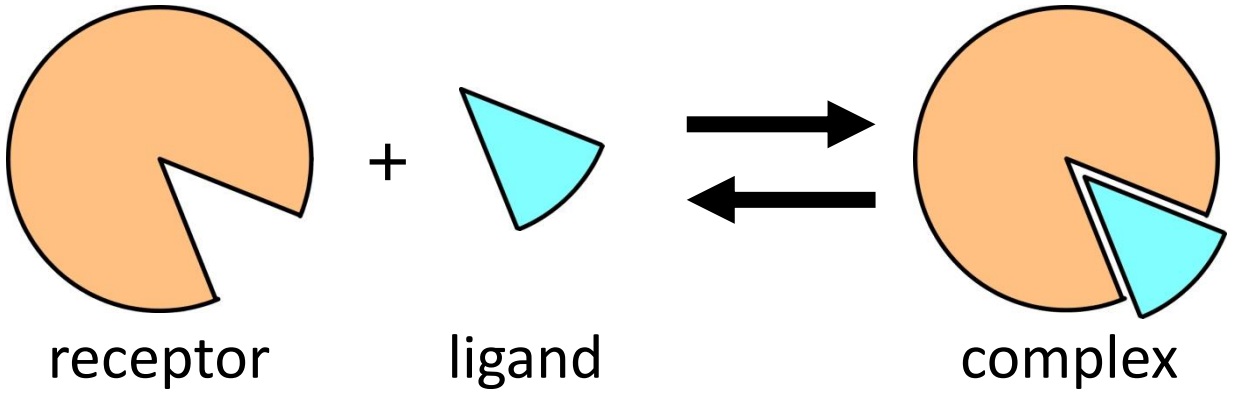
Structure superposition (3)

10. Choose the “second set of designated atoms” option for Check designated atoms against.
11. Choose “Select” → “Chain” → “B” to select the Raf atoms, click the “Designate selection as second set” button, and click “OK.”
    - Clashing atoms are indicated by yellow lines.



Ras Raf

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

- The complex structure has the minimum free energy.
- Free energy difference between the complex and the uncomplexed states.

$$\Delta G^{\circ}_{\text{bind}} = G^{\circ}_{\text{complex}} - \left(G^{\circ}_{\text{receptor}} + G^{\circ}_{\text{ligand}} \right)$$

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

- The complex structure is predicted as the one having the lowest ΔG_{bind} in the candidate structures generated by docking the ligand into all the possible sites on the protein surface in all the possible orientations and conformations.

Components of binding free energy

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

$$G = E + PV - TS$$

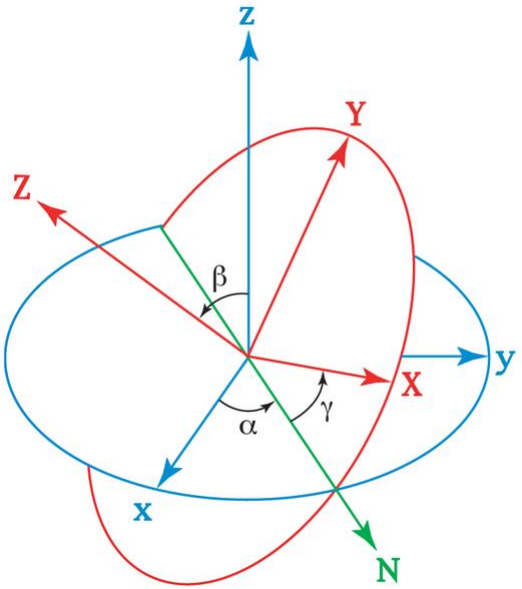
$$\Delta G^{\circ}_{bind} \approx \Delta E - T\Delta S = \Delta E_{int} + \Delta E_{desolv} - T\left(\Delta S_{conf} + \Delta S_{wat}\right)$$

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

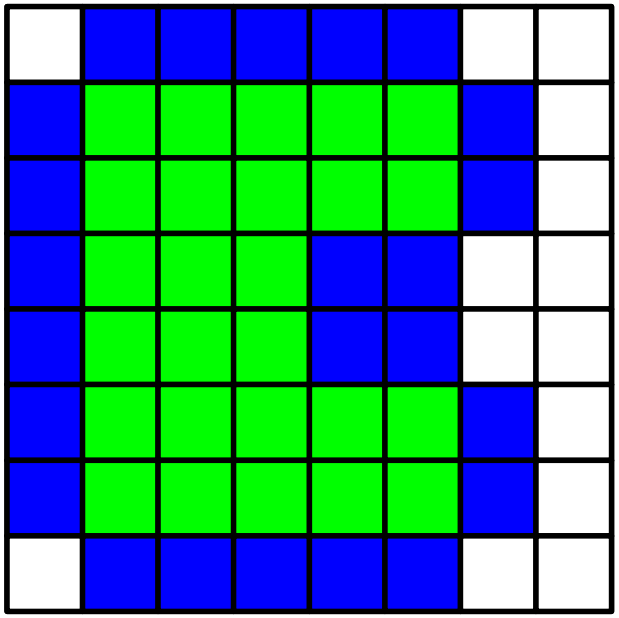
- Both of receptor and ligand are treated as rigid bodies. Conformational changes upon complex formation are not considered.
- Three translational and three rotational degrees of freedom of ligand are considered.
  - Rotation is described with Euler angle.
- Shape complementarity is important.



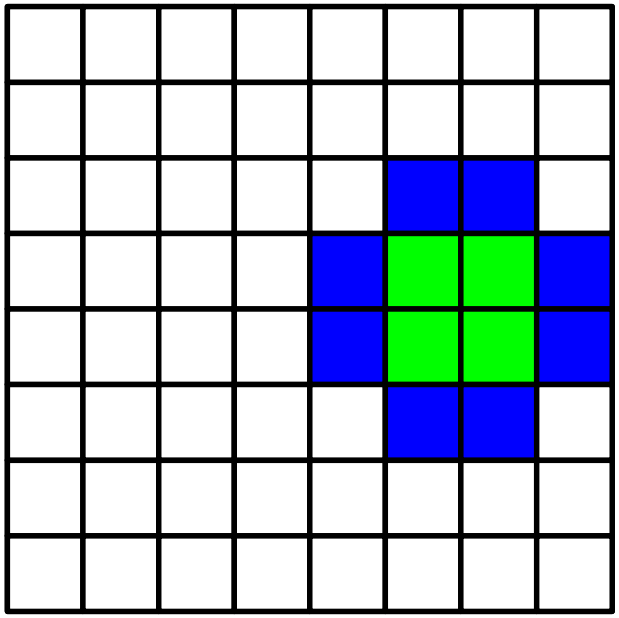
http://en.wikipedia.org/wiki/Euler_angles

Shape complementarity (1)

Receptor



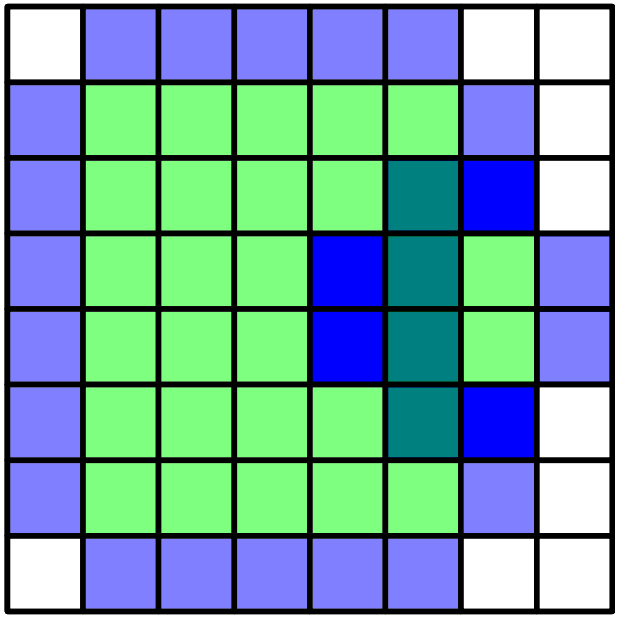
Ligand



= 1 (solvent accessible surface layer)

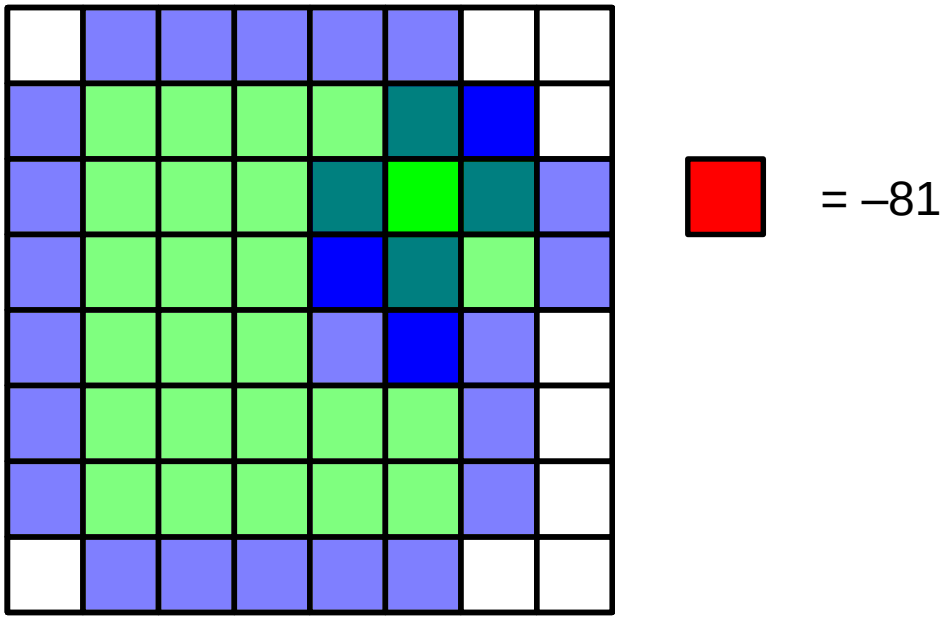
= 9*i* (solvent excluding surface layer)

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

- DOT

  http://www.sdsc.edu/CCMS/DOT/
- FTDock

  http://www.sbg.bio.ic.ac.uk/docking/ftdock.html
- GRAMM-X

  http://vakser.bioinformatics.ku.edu/resources/gramm/grammx
- HEX

  http://hex.loria.fr/
- ZDOCK

  http://zlab.umassmed.edu/zdock/index.shtml

An application of ZDOCK (1)

1. Access http://zdock.umassmed.edu/
2. Put the PDB ID of β-lactamase, “1ZG4,” in the textbox of Input Protein 1
3. Put the PDB ID of β-lactamase inhibitory protein, “3GMU,” in the textbox of Input Protein 2
4. Enter your e-mail address and click “Submit.”

An application of ZDOCK (2)

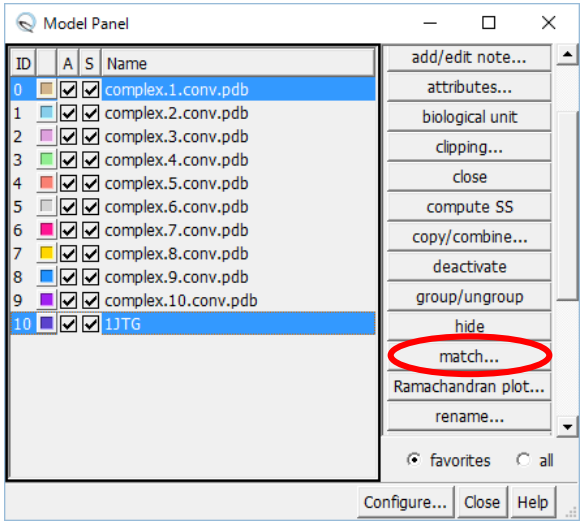
5. You can specify residues that are not involved in the binding or that are involved in the binding.
   Do not click “submit” in this lecture.

An application of ZDOCK (3)

6. When the calculation is finished, you will receive an e-mail containing the URL of the calculation result page.
7. Download top_preds.zip from the web page of this lecture and unzip it on the desktop.
8. The top_preds folder contains ten PDB files named from complex.1.pdb to complex.10.pdb in the order of the score. Because they cannot be displayed in Chimera, convert them by using conv.pl available on the web page of this lecture.
   - Ten more files named from complex.1.conv.pdb to complex.10.conv.pdb are generated

An application of ZDOCK (4)

9. Start Chimera and open complex.1.cov.pdb, …, and complex.10.conv.pdb.
10. Open the experimental complex structure, 1JTG.
11. Choose “Favotites”→ “Model Panel,” select complex.1.conv.pdb and 1JTG, and click “match” and “OK.”
12. Delete chains C and D of 1JTG



Which is the closest to the experimental structure among complex.1.cov.pdb, …, and complex.10.conv.pdb?

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.

Docking software

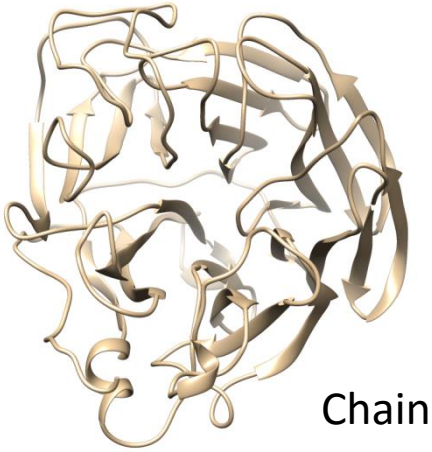
- AutoDock Vina
  - http://vina.scripps.edu/
- DOCK
  - http://dock.compbio.ucsf.edu/
- Glide
  - http://www.schrodinger.com/Glide
- GOLD
  - http://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/
- AutoDock Vina and DOCK are free, but Glide and GOLD incur a charge.
- Only the translational, rotational, and dihedral degrees of freedom of small molecules are considered. Proteins are treated as rigid bodies.

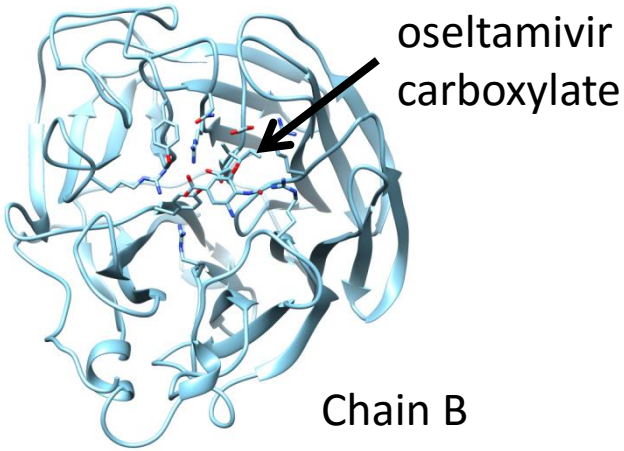
Hands-on training

- Dock an inhibitor into N1 neuraminidase using AutoDock Vina.
  1. Obtain a crystal structure of N1 neuraminidase
  2. Generate the structure of the inhibitor
  3. Detect cavities on the protein surface
  4. Perform a docking simulation
  5. Analyze the result

1. Obtaining a crystal structure

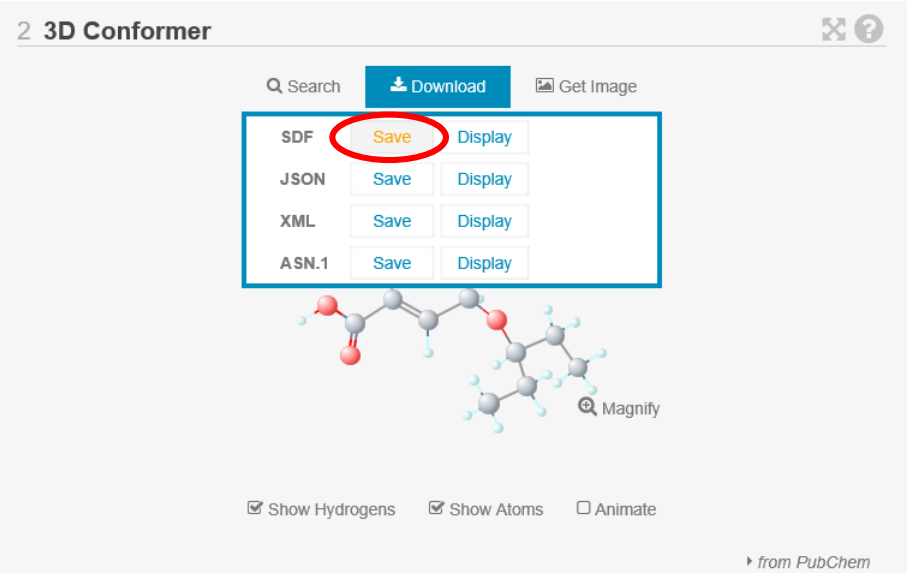
1. Start Chimera
2. Choose “File” → “Fetch by ID” to open “2HU0.”
3. Choose “Select” → “Chain” → “A” to select chain A.
4. Choose “File” → “Save PDB,” check the checkbox of “Save selected atoms only,” and save the coordinates as “2HU0_A.pdb” on the Desktop.
5. Similarly, save the coordinates of chain B as “2HU0_B.pdb” on the Desktop.





2. Generating inhibitor structure (1)

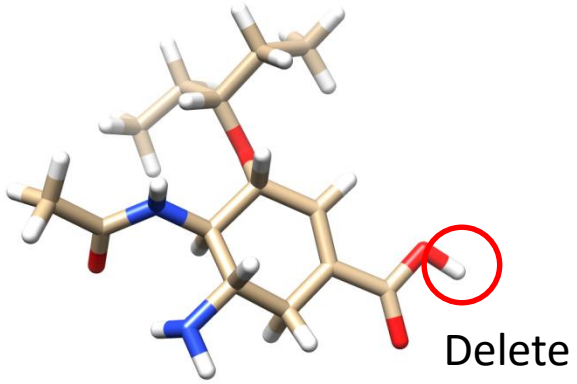
1. Access PubChem at https://pubchem.ncbi.nlm.nih.gov/
2. Enter “oseltamivir carboxylate” and “Go.”
3. Ascertain that CID is 449381.
4. Save 3D Conformer in SDF format.
5. Open the file in UCSF Chimera.

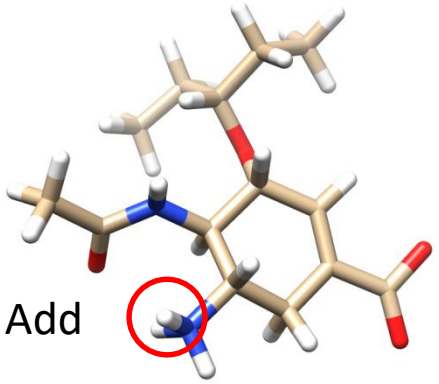


File name is:
Structure3D_CID_449381.sdf

2. Generating inhibitor structure (2)

6. Select the hydrogen atom of the carboxyl group by clicking on the atom while pressing the Ctrl key.
7. Delete the selected atom by using “Actions” → “Atoms/Bonds” → “delete.”
8. Add hydrogen to the amino group by using “Tools” → “Structure Editing” → “AddH.”
9. Calculate charges by using “Tools” → “Structure Editing” → “Add Charge.”
   – The net charge is +0.
10. Optimize the structure by using “Tools” → “Structure Editing” → “Minimize Structure.”
11. Choose “File” → “Save Mol2” to save the coordinates as “ose.mol2” on the Desktop.



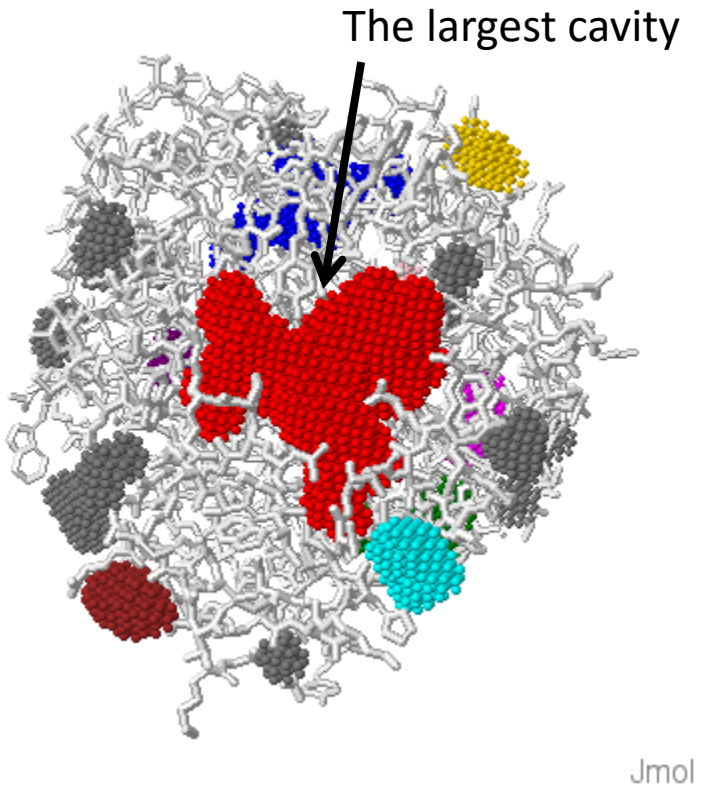


Compound library

- Available Chemicals Directory (ACD)
  - A commercial compound library
  - http://accelrys.com/products/databases/sourcing/available-chemicals-directory.html
- DrugBank
  - Database of drugs and their targets
  - http://www.drugbank.ca/
- PubChem
  - Database managed by NCBI
  - https://pubchem.ncbi.nlm.nih.gov/
- ZINC
  - Database managed by USCF
  - http://zinc.docking.org/

3. Cavity detection

- Detect cavities on the protein surface to which small molecules bind
  - Here, we use the GHECOM server at http://strcomp.protein.osaka-u.ac.jp/ghecom/
  - Center of the largest cavity: (–0.561, 78.515, 112.190)



Cavity detection software

- SURFNET
  - http://www.ebi.ac.uk/thornton-srv/software/SURFNET/
  - 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
  - Detects cavities on the protein surface and ranks them based on the interaction energy with CH_4 probe.

4. Docking simulation (1)

1. Make a folder named “docking” on the Desktop.
2. Move 2HU0_A.pdb, 2HU0_B.pdb, and ose.mol2 to the folder.
3. Download vina.exe from the web page of this lecture and save it on the Desktop.
4. Choose “File” → “Close Session” in Chimera.
5. Choose “File” → “Open” to open 2HU0_A.pdb and ose.mol2 in the docking folder.

4. Docking simulation (2)

6. Choose “Tools” → “Surface/Binding Analysis” → “Dock Prep.”
7. Select “2HU0_A” and “449381” in Molecules to prep, uncheck the check box of Write Mol2 file, and click “OK.”
8. Click “OK” in the Add Hydrogens for Dock Prep window.
9. Choose “Amber ff14SB” for Standard residues and “AM1-BCC” for Other residues in the Assign Charges for Dock Prep window and click “OK.”
10. Choose “+0” for Specify Net Charges and click “OK.”

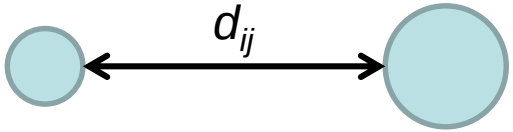
4. Docking simulation (3)

11. Choose “Tools” → “Surface/Binding Analysis” → “AutoDock Vina.”
12. Set “ose.pdbqt” in the docking folder to Output File.
13. Select “2HU0_A.pdb” and “CID 449381” as Receptor and Ligand, respectively.
14. Unfold Receptor search volume options and set “–0.561 78.515 112.190” and “25 25 25” to Center and Size, respectively.
15. Unfold Executable location and select “Local” and set “vina.exe” on the Desktop to Path.
16. Click “OK” to start the simulation.

Score function

- Complex structure is optimized to minimize *c*:

$$c = \sum_X w_X \sum_{i<j} h\left(d_{ij}\right)$$



  - *X* is gauss1, gauss2, repulsion, hydrophobic, or hydrogen bonding.
- Binding free energy is calculated as:

$$s = \frac{c - c_{\text{intra1}}}{1 + w_{\text{rot}} N_{\text{rot}}}$$

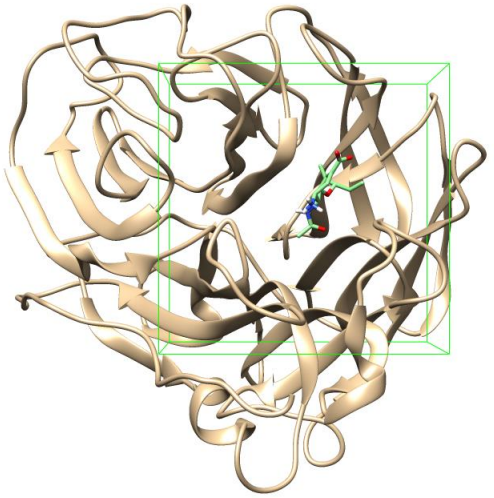
c_{intra1} : *c* for intramolecular interaction of the model with the smallest *c*

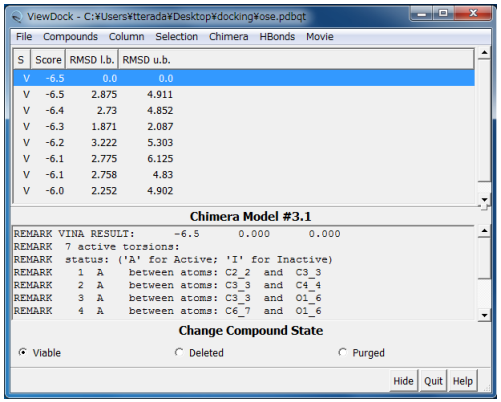
- w_x is optimized so that *s* correlates with the experimental value of ΔG_{bind}.

Trott & Olson *J. Comput. Chem.* **31**, 455 (2010).

5. Analysis of the result (1)

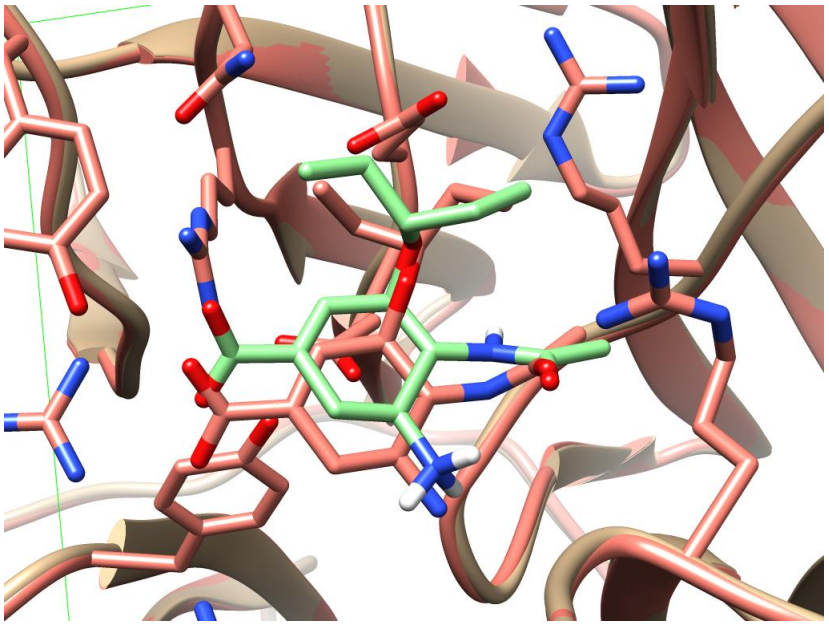
- The result is displayed when the calculation is completed.
- The docking scores are listed in the ViewDock window in the order of the score.
- Clicking on the row of the list changes docking pose displayed.





5. Analysis of the result (2)

- Compare with the crystal structure.
  1. Open 2HU0_B.pdb.
  2. Choose “Favorites” → “Model Panel” to hide the model of 449381.
  3. Select “2HU0_A.pdb” and “2HU0_B.pdb” in Model Panel and click “match” and “OK.”
- Which pose is the closest to the crystal structure?

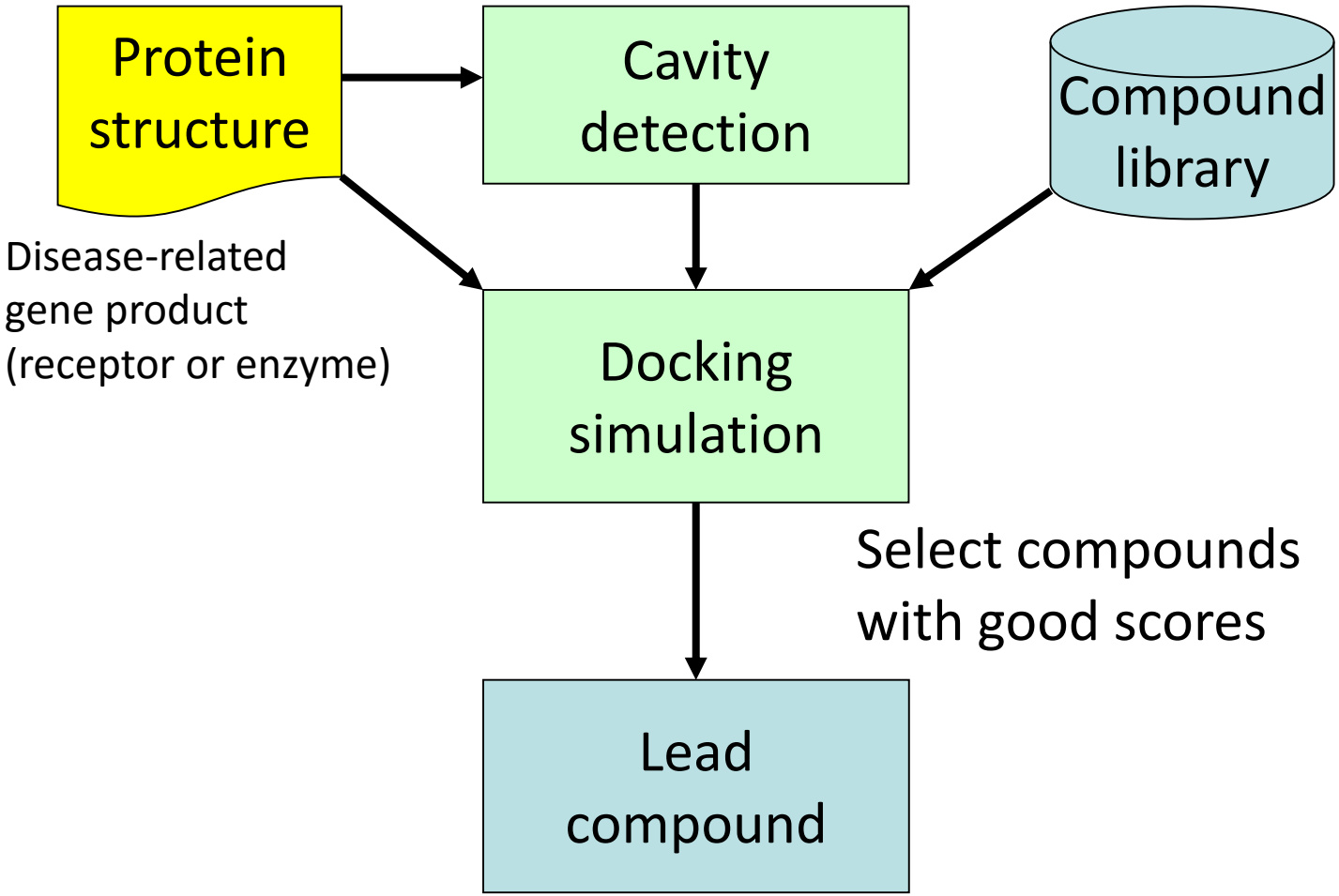


Comparison between the 4-th ranked pose and the crystal structure

Application to drug discovery

- A high-throughput screening (HTS), which searches compound libraries for compounds tightly binding to the target protein, is used to find drug candidates in the field of drug discovery.
- It costs huge sums of money to establish compound libraries and to measure the affinity of each compound to the protein.
- The affinity can be estimated by reproducing the protein-compound binding in a computer, *i.e.*, by docking simulation. → virtual screening

Virtual screening



Current status of molecular simulations (1)

- Possible
  - Tertiary structure prediction of a protein using a template structure with high sequence similarity
  - Folding simulation of a small protein
  - Refinement of a near-native model toward the native structure
  - Protein-protein and protein-small molecule docking using the native or near-native structures
  - Prediction of binding free energy between a protein and a small molecule using the native complex structure
  - Reproduction of thermal fluctuations and fast (< microseconds) motions of a protein using the native or a near-native structure

Current status of molecular simulations (2)

- Difficult
  - Tertiary structure prediction of a large protein using folding simulation
  - Refinement of an inaccurate model toward the native structure
  - Protein-protein and protein-small molecule docking using inaccurate models
  - Prediction of binding free energy between proteins
  - Prediction of binding free energy between a protein and a small molecule using a model structure
  - Reproduction of slow (> milliseconds) motions

Time scale of protein dynamics

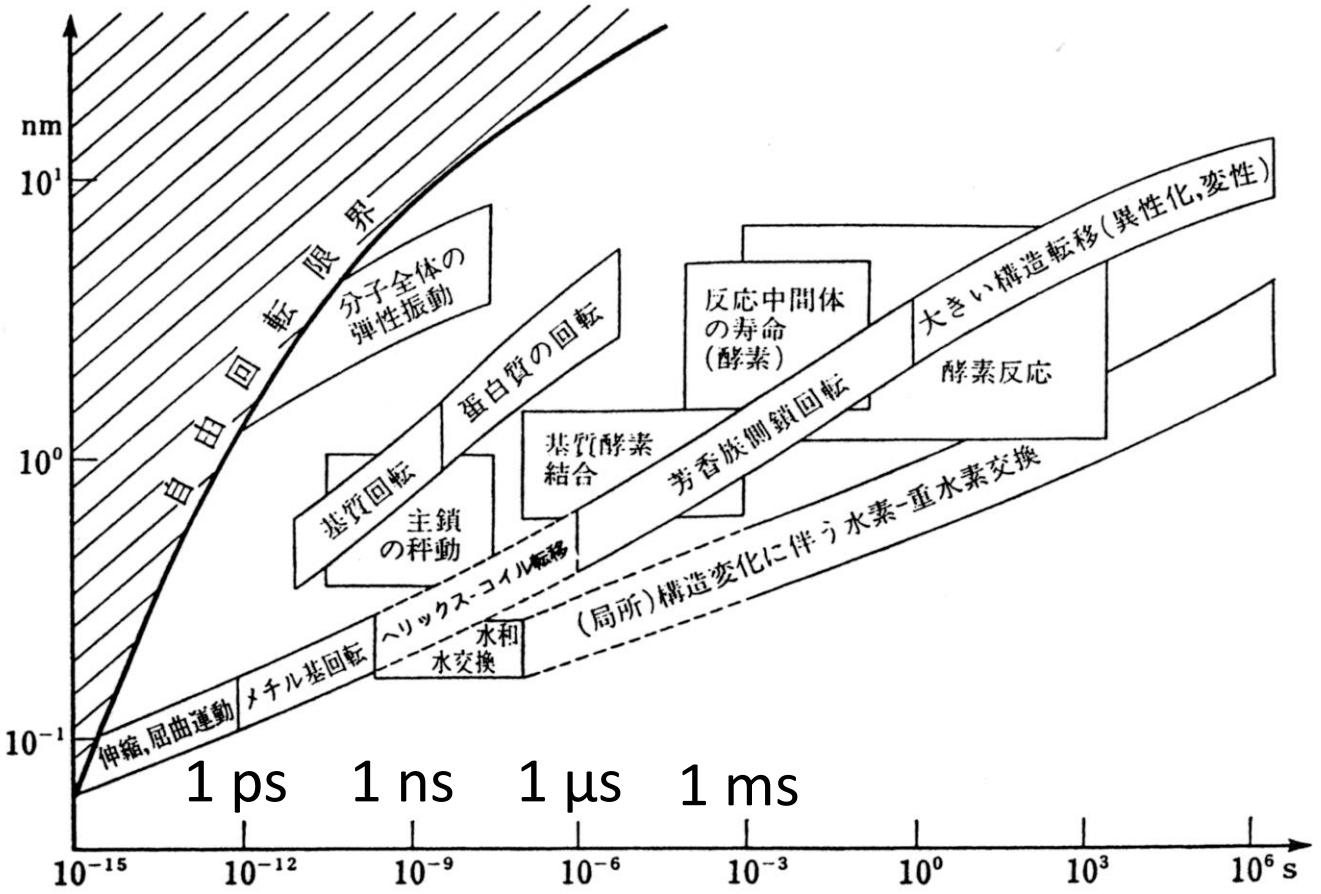
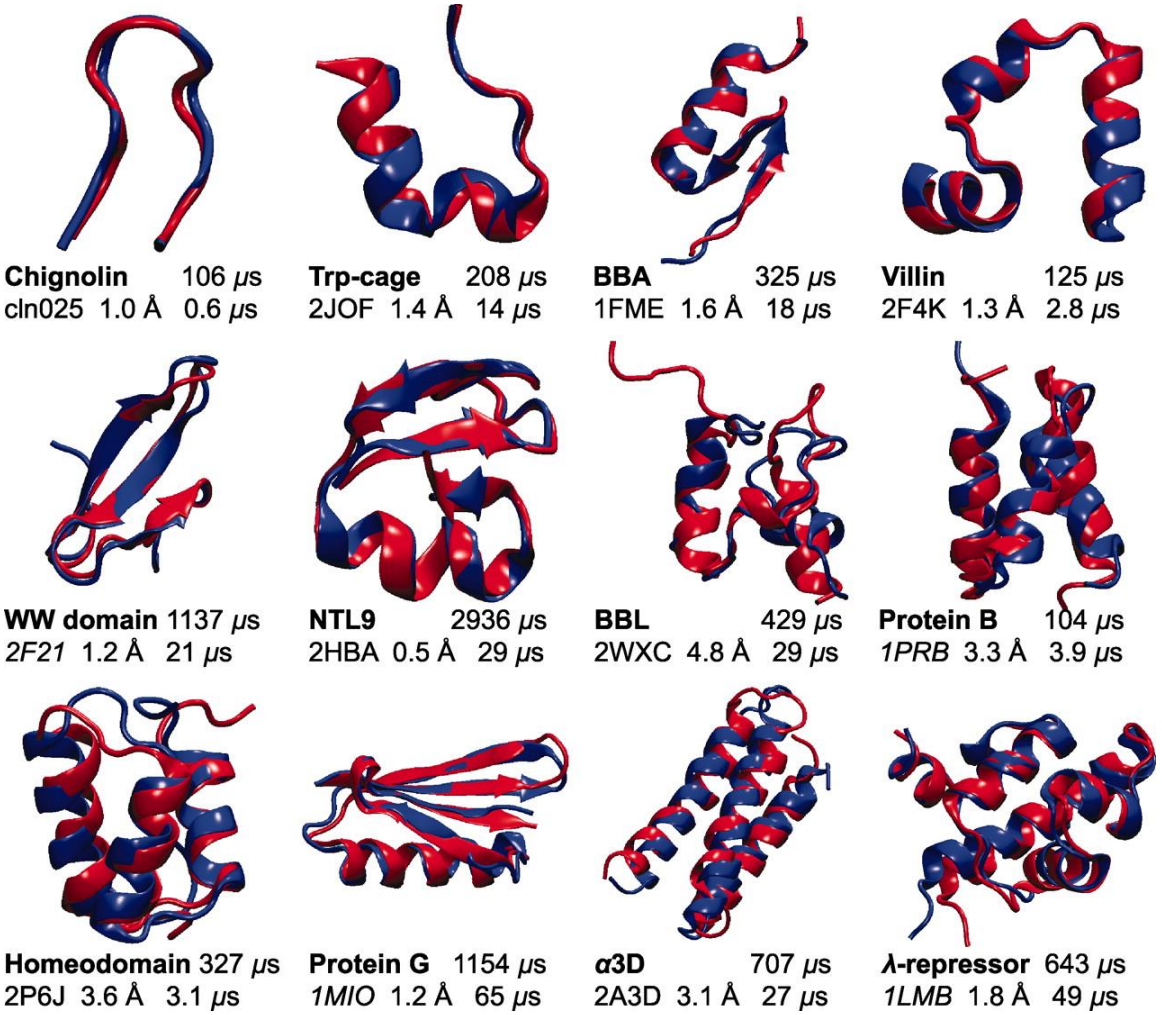


図 4. 17 タンパク質分子の構造変化や化学反応事象の時間-空間相関プロット

時間軸と空間軸のケタ数の違いに注目.

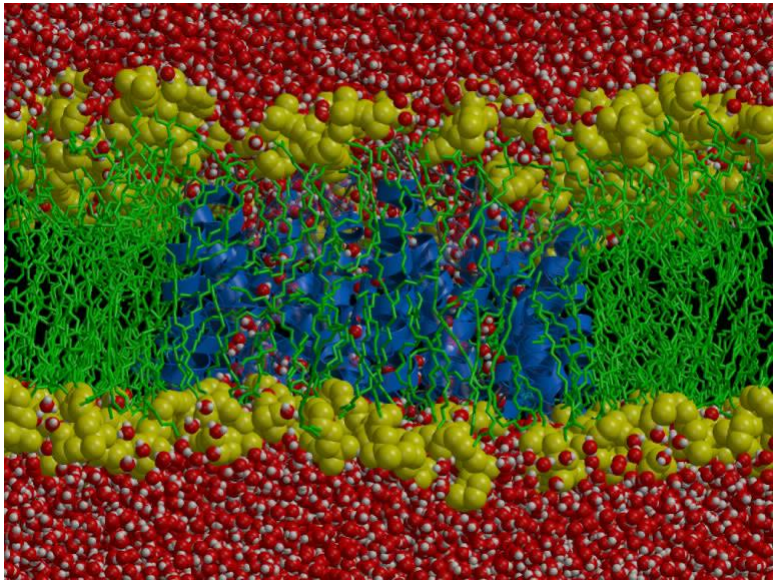
永山國昭 「生命と物質　生物物理学入門」より引用

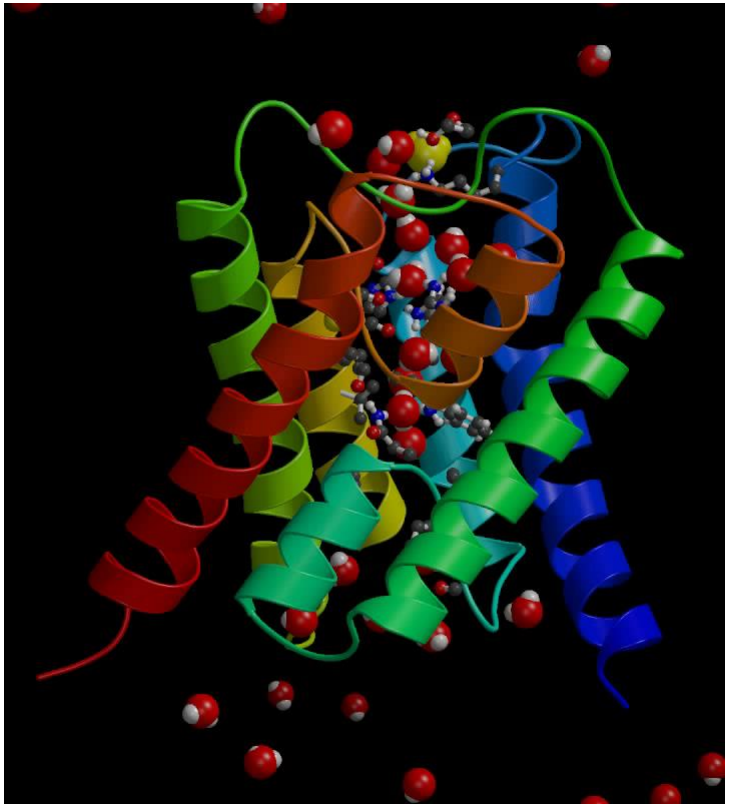
Folding simulations



Lindorff-Larsen *et al. Science* **334**, 517 (2011).

An MD simulation of Aquaporin



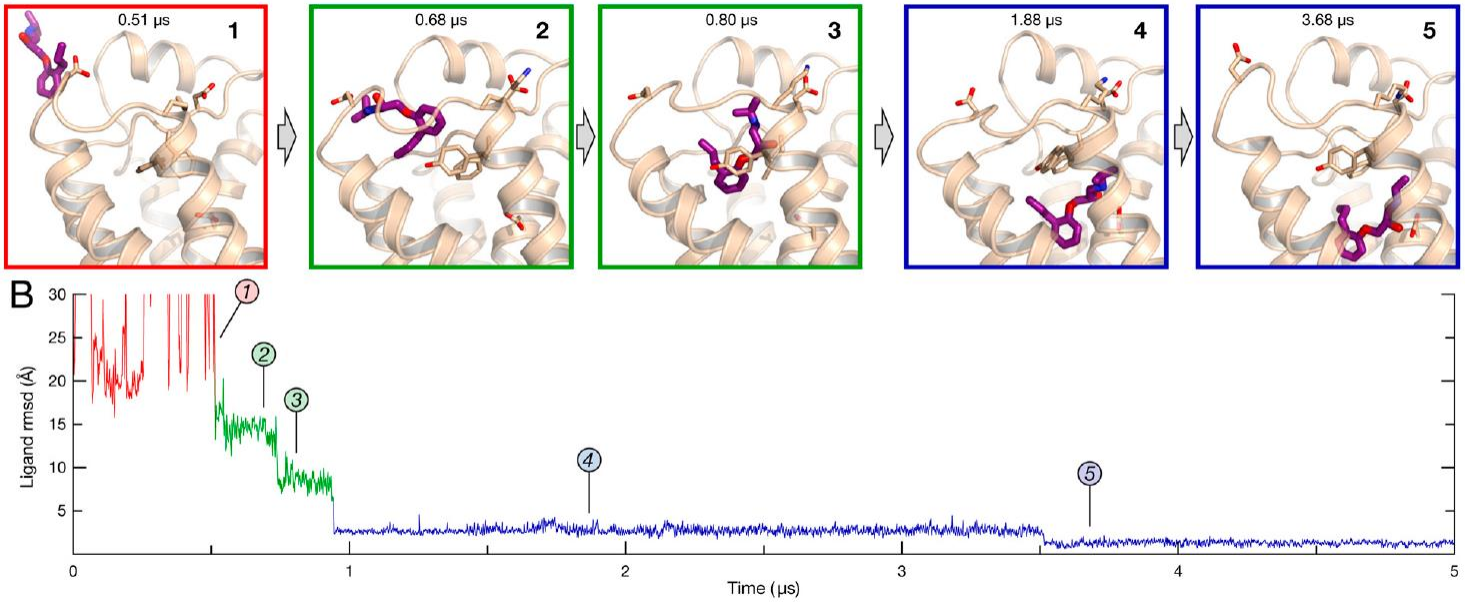


- Water permeation rate
  - Expt.: 3×10^9 sec^{-1}
  - Simulation: 16 / 10 ns
    $\rightarrow 1.6 \times 10^9$ sec^{-1}

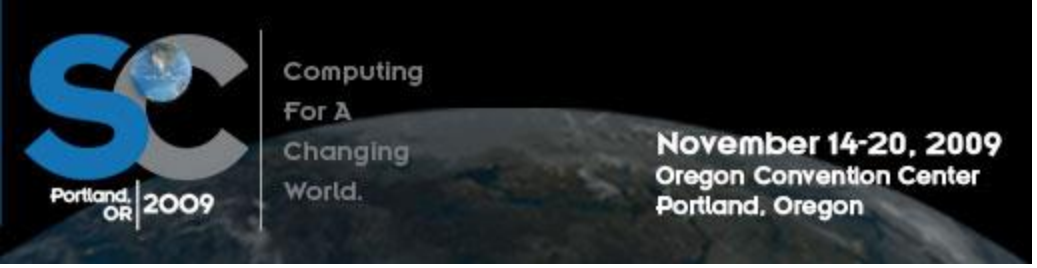
de Groot & Grubmüller, *Science* **294**, 2353 (2001).

Ligand-binding simulation

- MD simulations of binding of antagonistic drugs such as alprenolol to the β_2-adrenergic receptor.
- Binding rate constant
  - Expt.:$1.0 \times 10^7\ M^{-1}\ s^{-1}$
  - 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

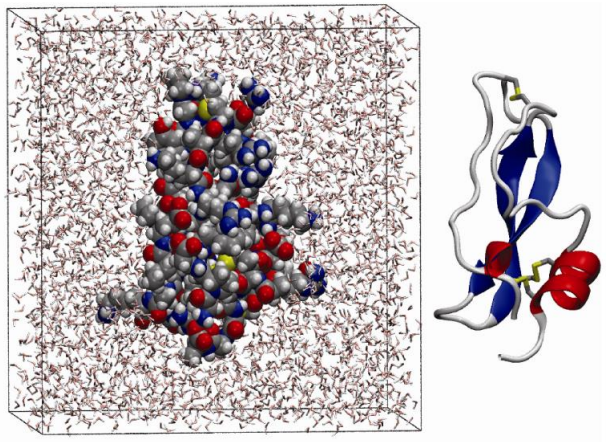
David E. Shaw*, Ron O. Dror, John K. Salmon, J.P. Grossman, Kenneth M. Mackenzie, Joseph A. Bank, Cliff Young, Martin M. Deneroff, Brannon Batson, Kevin J. Bowers, Edmond Chow, Michael P. Eastwood, Douglas J. Ierardi, John L. Klepeis, Jeffrey S. Kuskin, Richard H. Larson, Kresten Lindorff-Larsen, Paul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

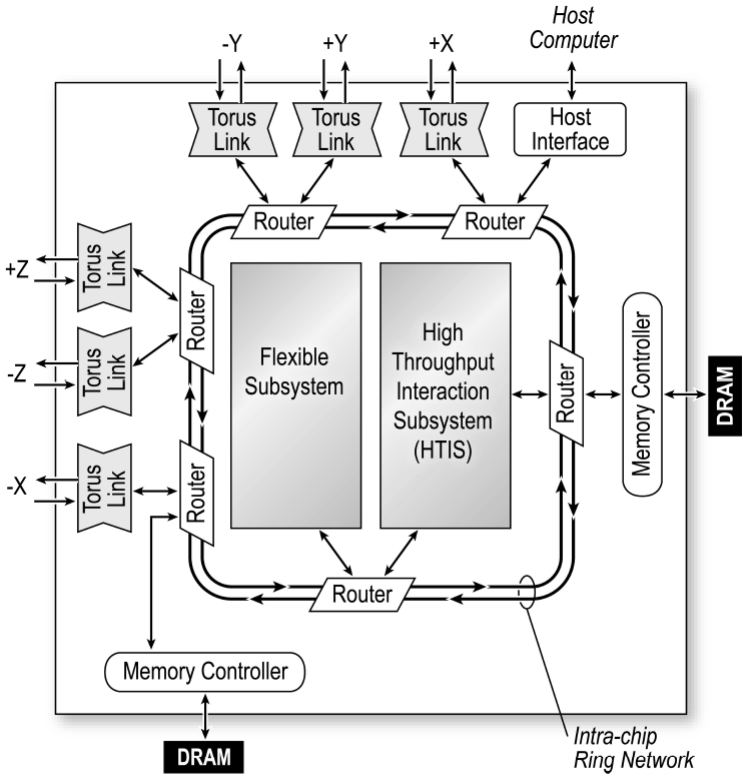
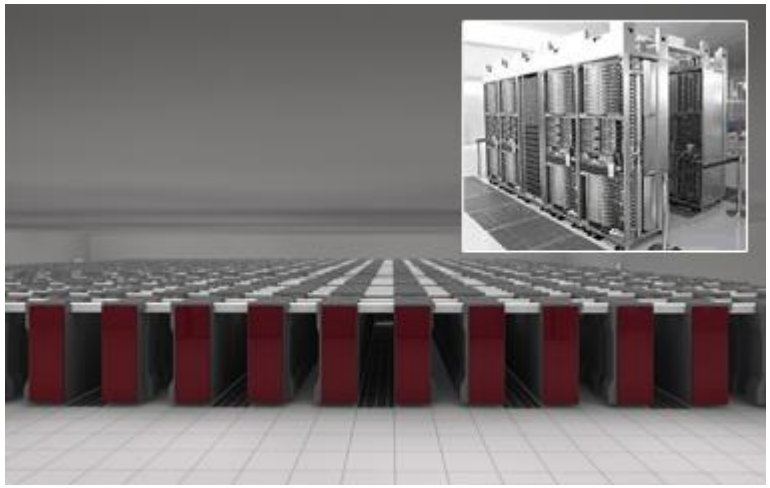
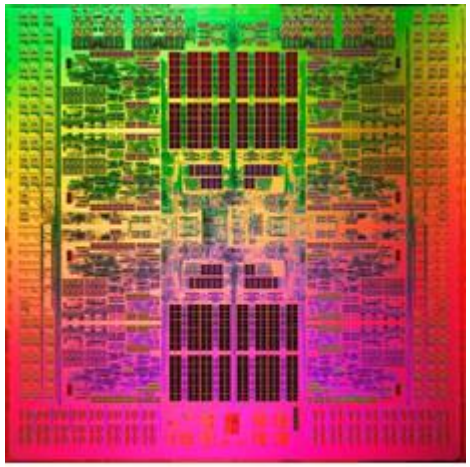


Figure 2: Anton ASIC block diagram.

The K supercomputer

- 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/

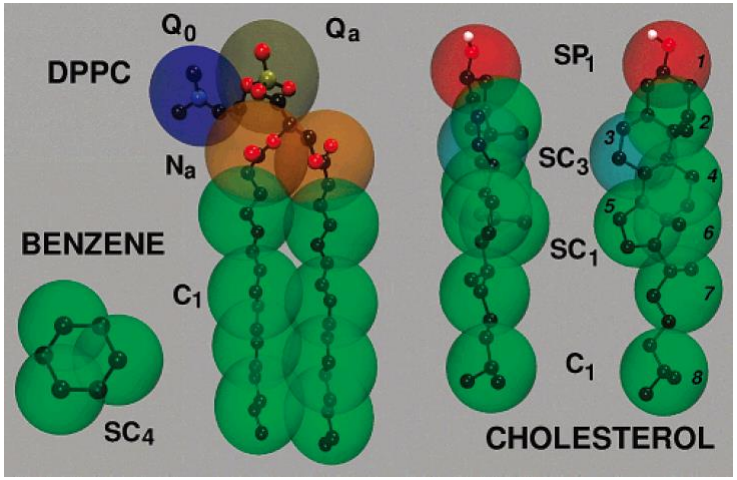
Coarse-grained (CG) model

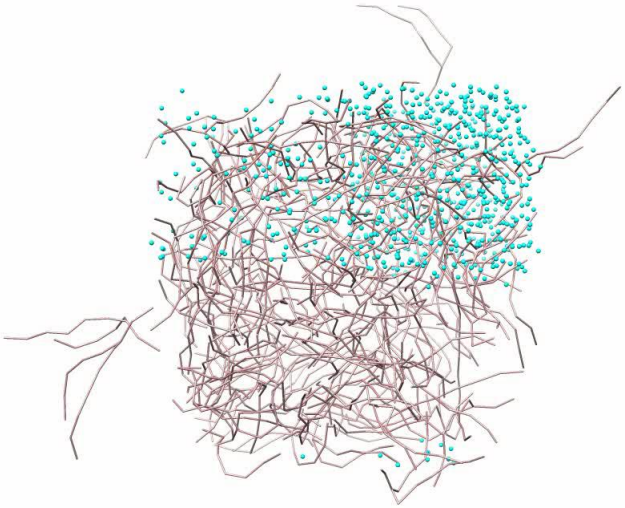
- In the MD simulation, all of the details of the dynamics, including the bond-stretching motions, are reproduced.
- Such detailed information is not necessary.

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

The MARITINI force field

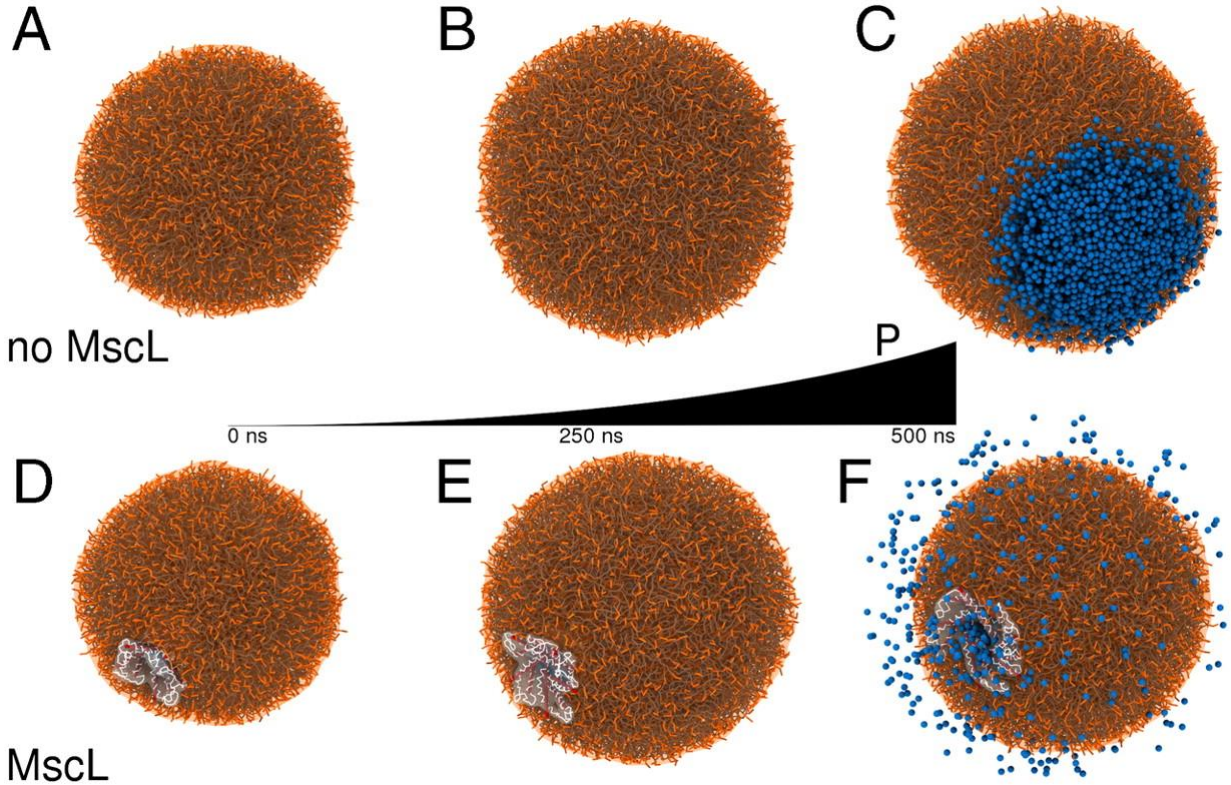
- Developed by Marrink's group.
- Maps four non-hydrogen atoms into one particle.
- Force field parameters were determined so as to reproduce free energies of hydration, vaporization, and partitioning between water and organic phases.
- Time step is 30 fs. The effective time is 4-fold longer.
- The movie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





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.

Exercise

- What is the rank and the score of the pose closest to the crystal structure in the docking of oseltamivir carboxylate to chain A of 2HU0?
- Create a figure of the closest pose superimposed on the crystal structure.
- Create a similar figure for the first ranked pose. Discuss why this pose has such a good score.

Assignment submission instructions

- Put the results and discussion in the body of email.
- Send the two figures (image files) as attachments.
- Put “Molecular modeling assignments” in the Subject field of the email.
- Be sure to put your name and ID card number (if you are a student) in the body of the email.
- Send the email to Prof. Tohru Terada (tterada@iu.a.u-tokyo.ac.jp).