Complex Structure Modeling

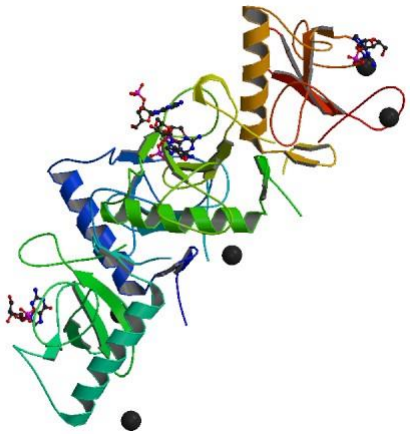
Agricultural Bioinformatics Research Unit,
Graduate School of Agricultural and Life Sciences,
The University of Tokyo
Tohru Terada

Today's topics

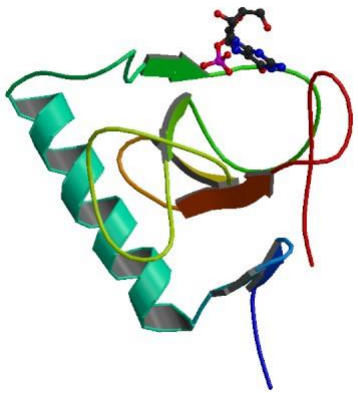
- Notes for conducting simulations
- Protein-protein docking
- Protein-small molecule docking
  - Exercise
- Applications of molecular simulation

Notes 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

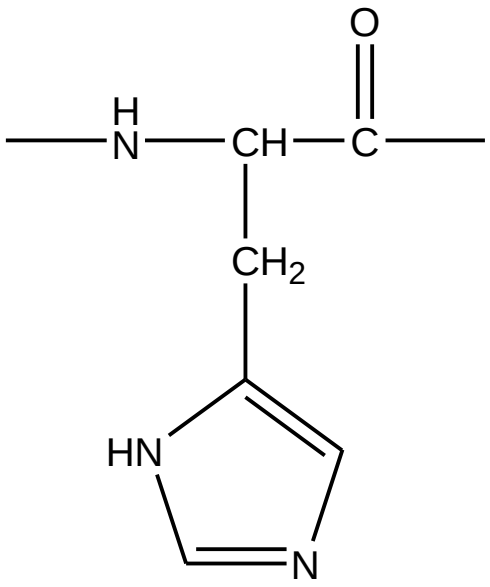


Biological Unit

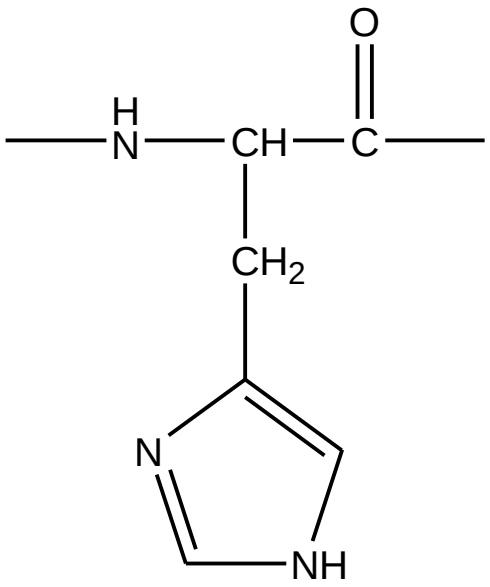
Notes for conducting simulations (2)

- 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.

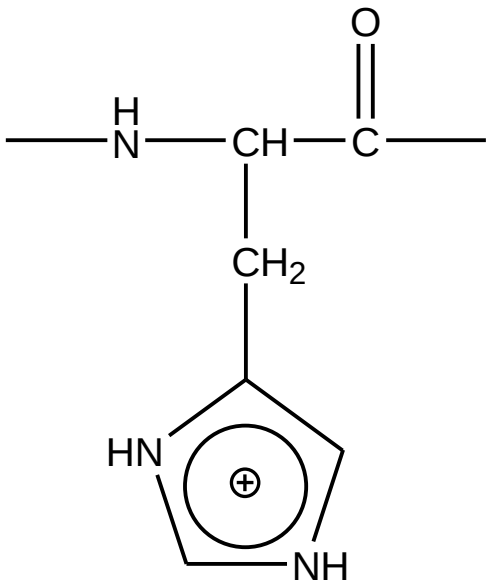
Protonation states of His



Protonation at δ



Protonation at ε



Protonation at δ and ε

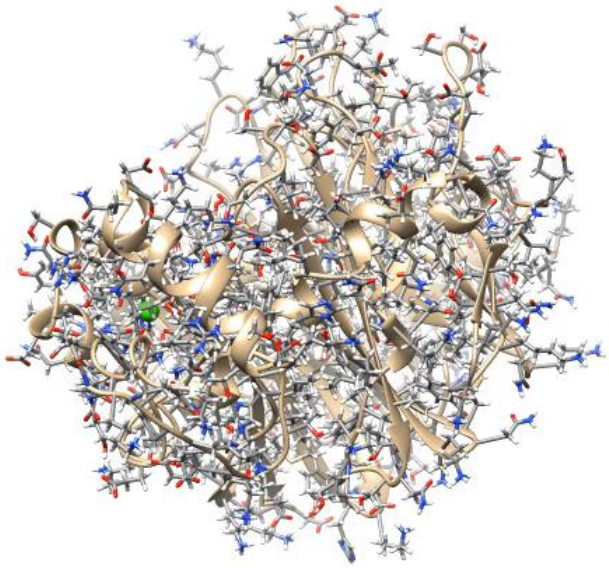
- pK_a of His side chain is close to neutral.
- You can determine the protonation state based on the hydrogen bond network where His is involved.

Notes 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

An example of a soluble protein

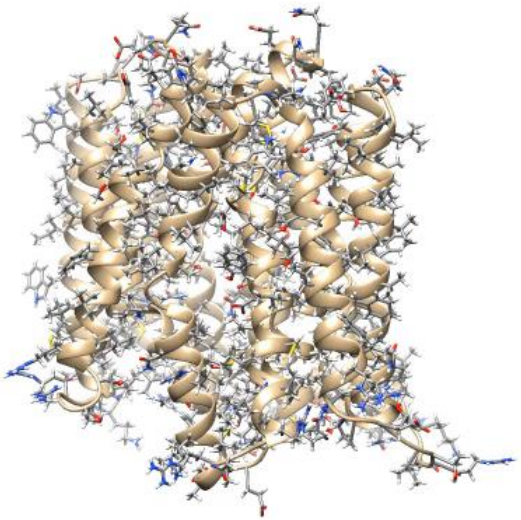


Levansucrase (PDB ID: 1PT2)

Simulations of membrane proteins

- Membrane proteins are embedded in solvated lipid bilayers.
- Lipid models
  - Prokaryote: phosphatidylethanolamine
  - Eukaryote: phosphatidylcholine etc.
- Database of protein orientations in membrane
  - Orientations of Proteins in Membranes (OPM) database
  - http://opm.phar.umich.edu/
- Initial structure generator
  - http://www.charmm-gui.org/

An example of a membrane protein



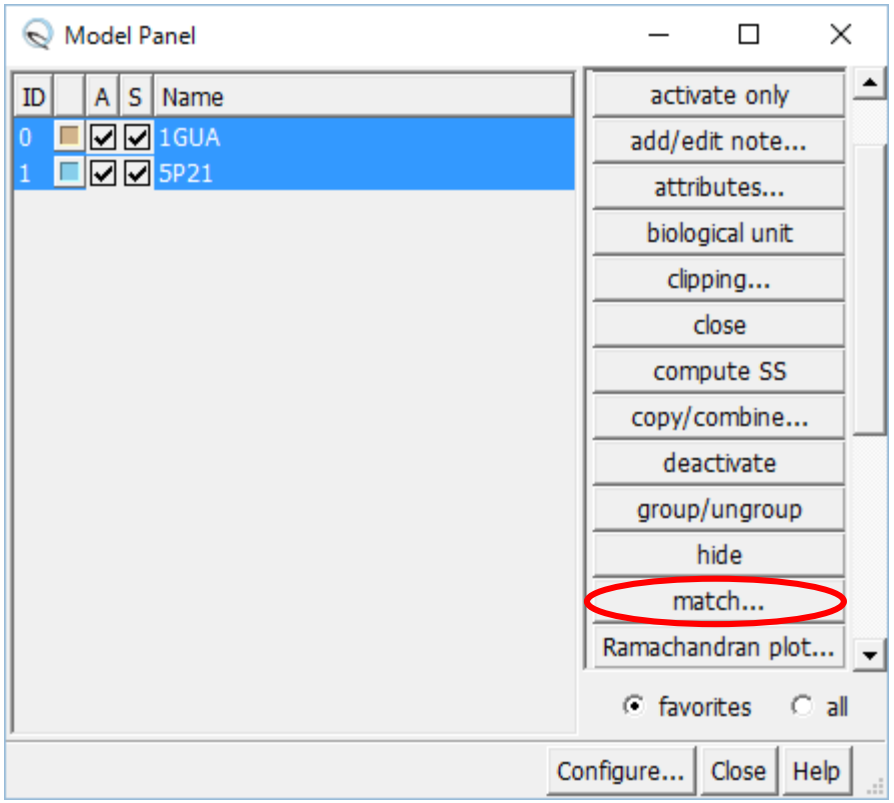
E. coli multidrug transporter MdfA (PDB ID: 4ZOW)

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
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.

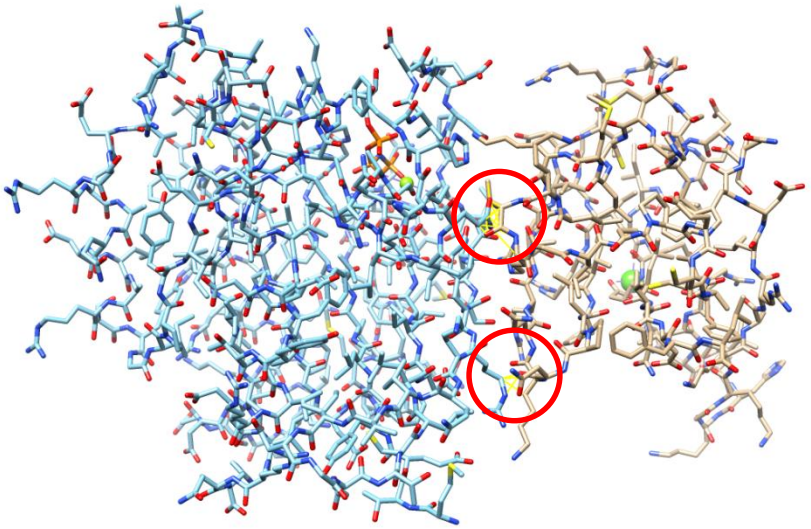


Structure superposition (2)

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.

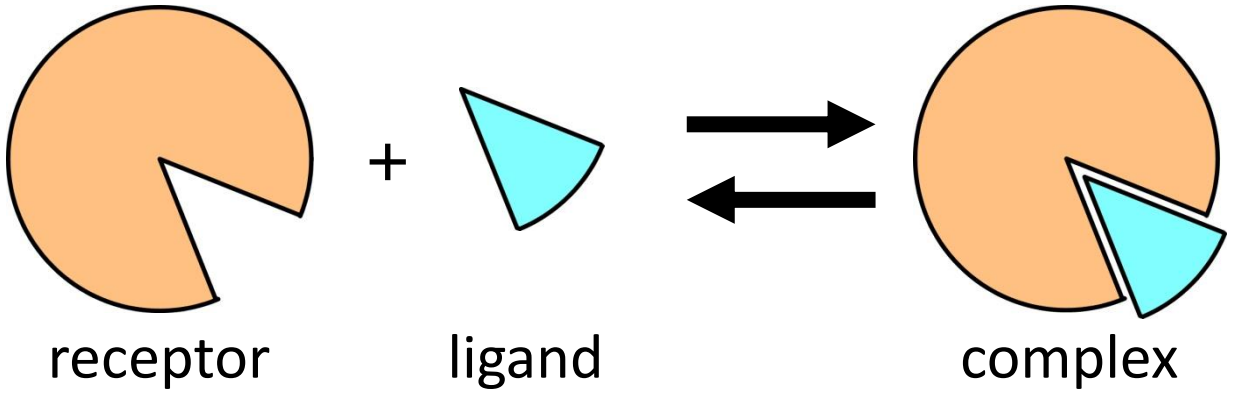
Structure superposition (3)

10. Select 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 is given by:

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

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

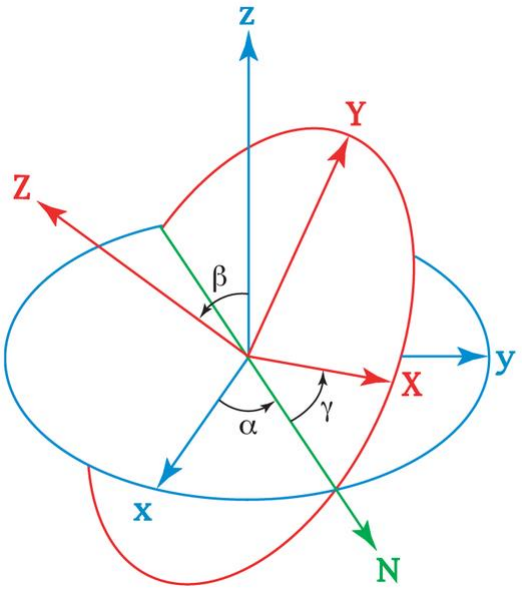
- The complex structure is predicted as the one having the lowest ΔG_{bind} in the candidate structures generated by docking the ligand into all the possible sites on the protein surface in all the possible orientations and conformations.

Calculation of binding free energy

- When calculation is done for a small number of the targets:
  - e.g. calculation for the crystal structure of a complex
  - MM-PB/SA method
  - Free-energy perturbation, thermodynamic integration
- When calculation is done for a large number of the targets:
  - e.g. docking simulation
  - Score function

Protein-protein docking

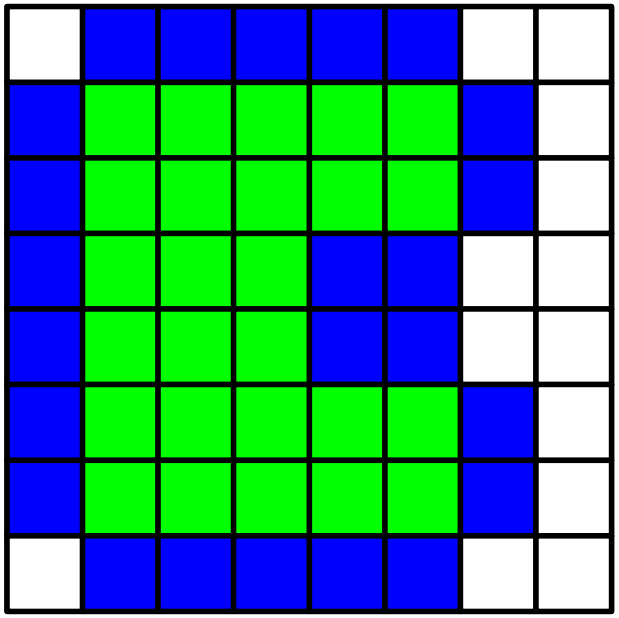
- Both of receptor and ligand are treated as rigid bodies. Conformational changes upon complex formation are not considered.
- Three translational and three rotational degrees of freedom of ligand are considered.
  - Rotation is described with Euler angle.
- Shape complementarity is important.



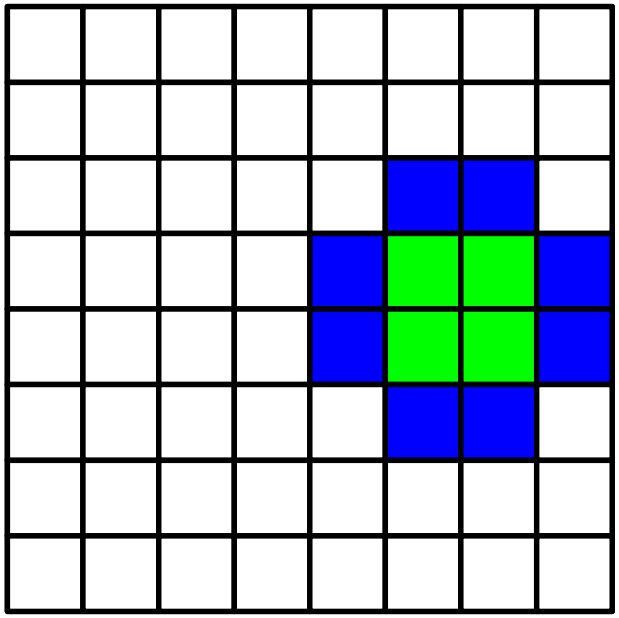
http://en.wikipedia.org/wiki/Euler_angles

Shape complementarity (1)

Receptor



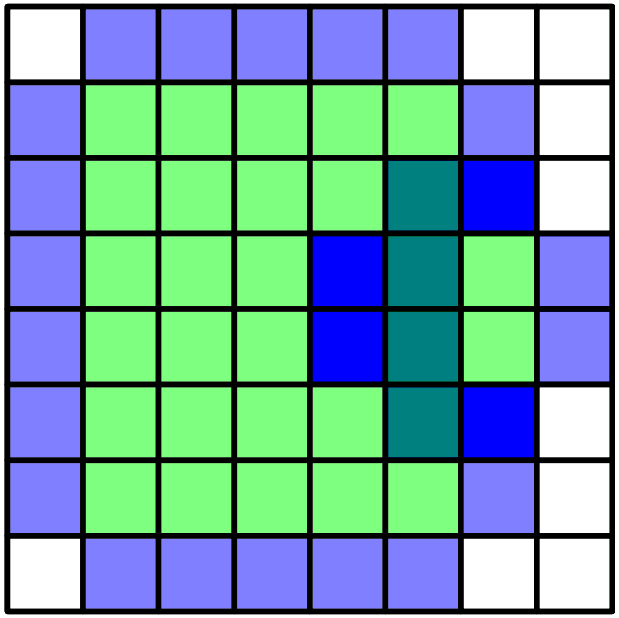
Ligand



= 1 (solvent accessible surface layer)

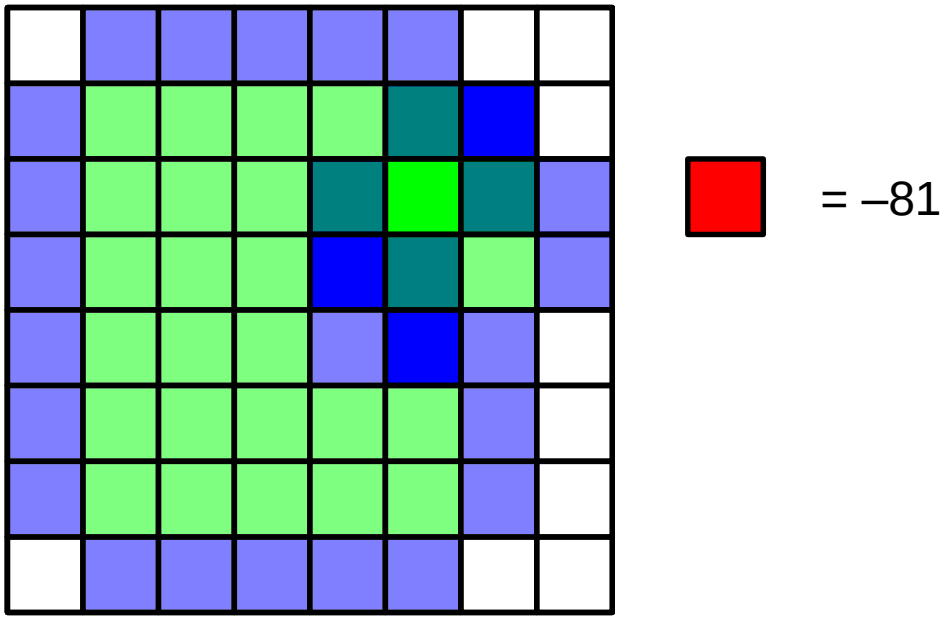
= 9*i* (solvent excluding surface layer)

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)

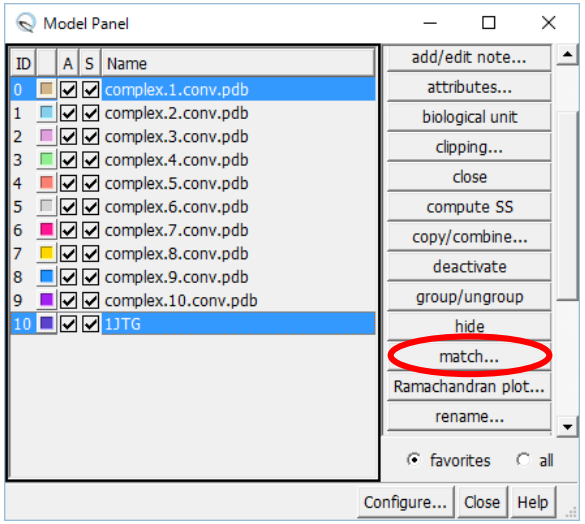
5. You can specify residues that are not involved in the binding or that are involved in the binding.
   Do not click “submit” in this lecture.

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 “Favorites”→ “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

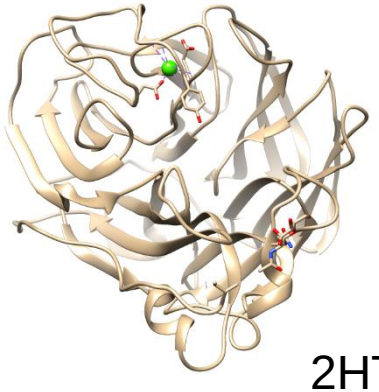
- AutoDock Vina
  - http://vina.scripps.edu/
- DOCK
  - http://dock.compbio.ucsf.edu/
- Glide
  - http://www.schrodinger.com/Glide
- GOLD
  - http://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/
- AutoDock Vina and DOCK are free, but Glide and GOLD incur a charge.
- Only the translational,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

Hands-on training

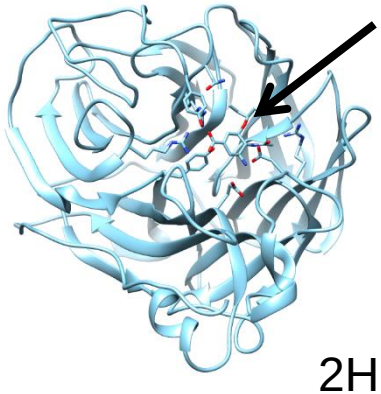
- Dock an inhibitor into N1 neuraminidase using AutoDock Vina.
  1. Obtain a crystal structure of N1 neuraminidase
  2. Generate the structure of the inhibitor
  3. Detect cavities on the protein surface
  4. Perform a docking simulation
  5. Analyze the result

1. Obtaining a crystal structure

1. Start Chimera
2. Choose “File” → “Fetch by ID” to open “2HTY.”
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 “2HTY_A.pdb” on the Desktop.
5. After choosing “File” → “Close session”, fetch “2HU4” and save the coordinates of chain A as “2HU4_A.pdb” on the Desktop.



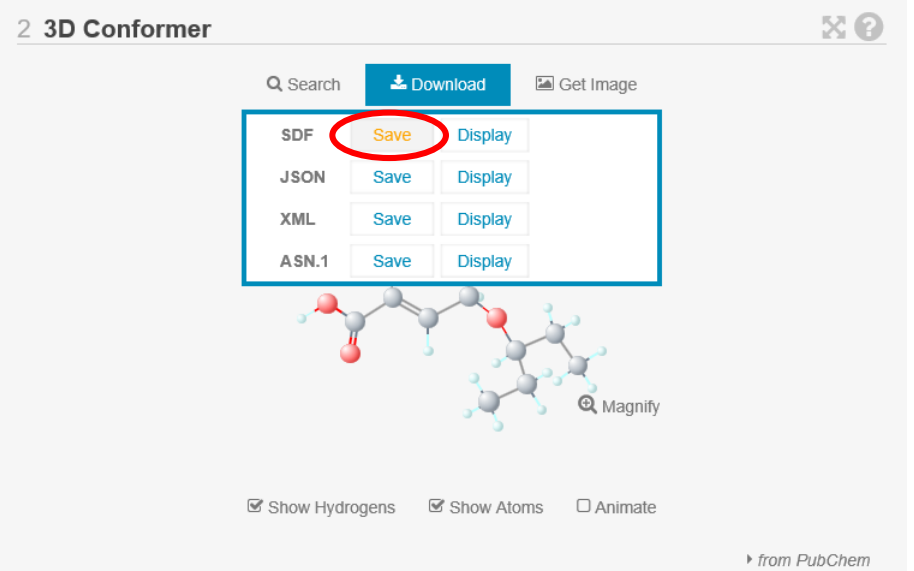
2HTY chain A



2HU4 chain A

2. Generating inhibitor structure (1)

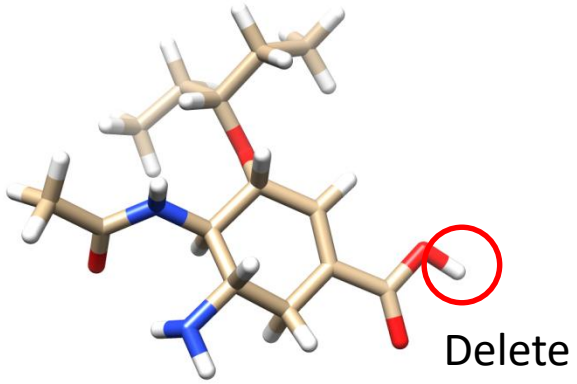
1. Access PubChem at https://pubchem.ncbi.nlm.nih.gov/
2. Enter “oseltamivir carboxylate” and “Go.”
3. Ascertain that CID is 449381.
4. Save 3D Conformer in SDF format.
5. Open the file in UCSF Chimera.

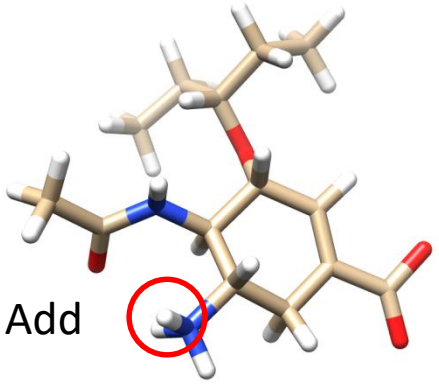


File name is:
Structure3D_CID_449381.sdf

2. Generating inhibitor structure (2)

6. Select the hydrogen atom of the carboxy 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.



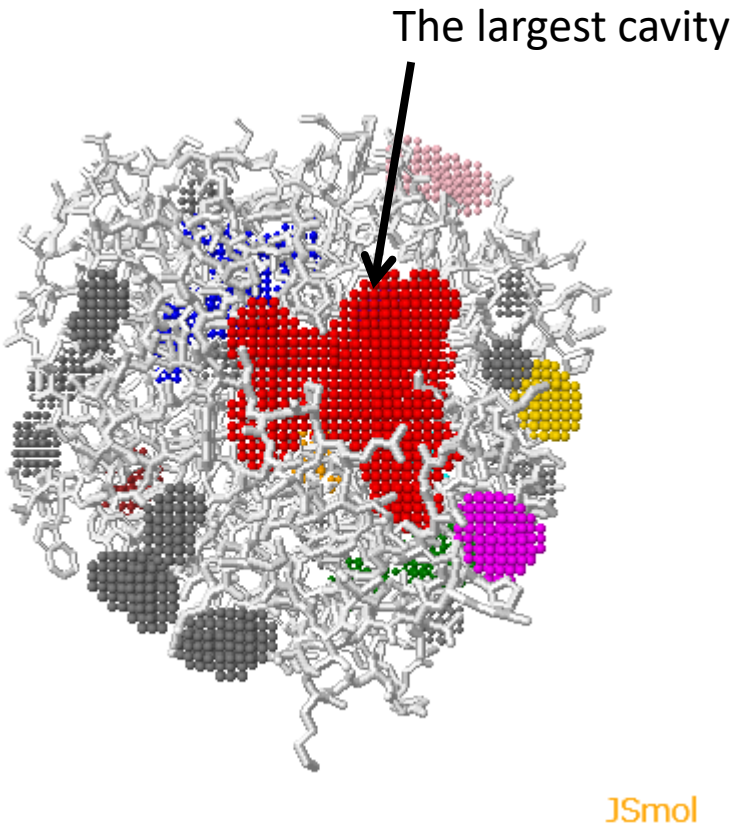


Compound library

- Available Chemicals Directory (ACD)
  - A commercial compound library
  - http://accelrys.com/products/collaborative-science/databases/sourcing-databases/biovia-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/
  - Put “2hty” in INPUT PDBID
  - Center of the largest cavity: (–0.260, 78.567, 112.835)



Cavity detection software

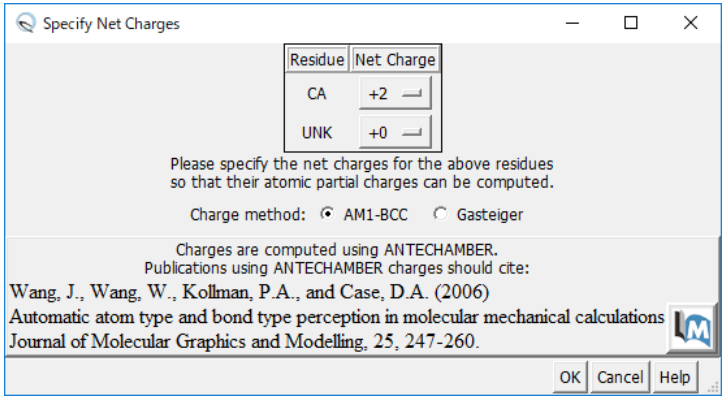
- SURFNET
  - http://www.ebi.ac.uk/thornton-srv/software/SURFNET/
  - Detects “gap regions” on the protein surface.
- PASS
  - http://www.ccl.net/cca/software/UNIX/pass/overview.shtml
  - Detects cavities on the protein surface and ranks them.
- Q-SiteFinder
  - Detects cavities on the protein surface 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 and move 2HTY_A.pdb, 2HU4_A.pdb, and ose.mol2 to the folder.
2. Download vina.exe from the web page of this lecture and save it on the Desktop.
3. Choose “File” → “Close Session” in Chimera.
4. Choose “File” → “Open” to open 2HTY_A.pdb
5. Choose “Select” → “Residues” → “NDG” to select the sugar and delete it by choosing “Actions” → “Atoms/Bonds” → “delete.”
6. Delete water molecules (residue name: HOH) in a similar manner.

4. Docking simulation (2)

7. Choose “File” → “Open” to open ose.mol2.
8. Choose “Tools” → “Surface/Binding Analysis” → “Dock Prep.”
9. Make sure that “2HTY_A” and “449381” are selected in Molecules to prep, uncheck the check box of Write Mol2 file, and click “OK.”
10. Click “OK” in the Add Hydrogens for Dock Prep window.
11. Select “Amber ff14SB” for Standard residues and “AM1-BCC” for Other residues in the Assign Charges for Dock Prep window and click “OK.”
12. Select “+0” for Specify Net Charges and click “OK.”
    (You can ignore the warning.)



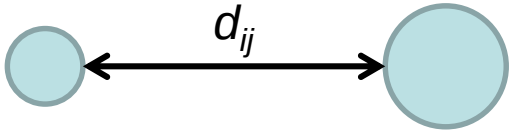
4. Docking simulation (3)

13. Choose “Tools” → “Surface/Binding Analysis” → “AutoDock Vina.”
14. Set “ose.pdbqt” in the docking folder to Output File.
15. Select “2HTY_A.pdb” and “449381” as Receptor and Ligand, respectively.
16. Unfold Receptor search volume options and set “–0.260 78.567 112.853” and “20 20 20” to Center and Size, respectively.
17. Unfold Executable location and select “Local” and set “vina.exe” on the Desktop to Path.
18. Click “OK” to start the simulation. (It will take about one minute.)

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:

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

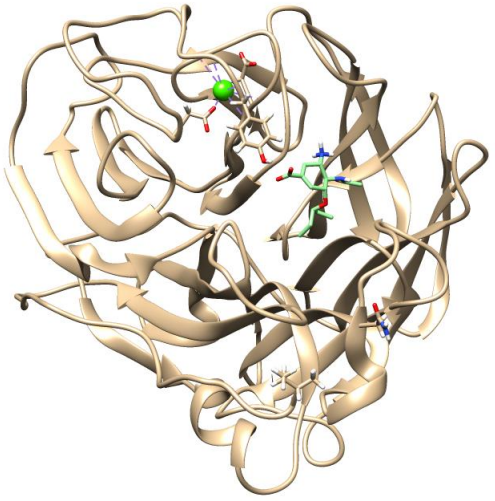
c_{intra1} : *c* for intramolecular interaction of the model with the smallest *c*

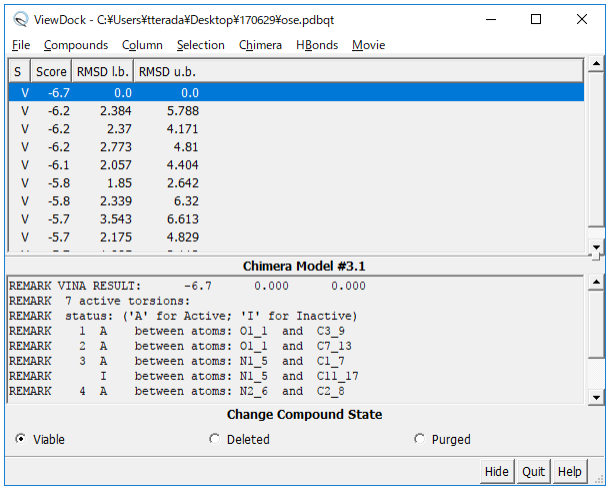
- w_x is optimized so that *s* correlates with the experimental value of ΔG_{bind}.

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

5. Analysis of the result (1)

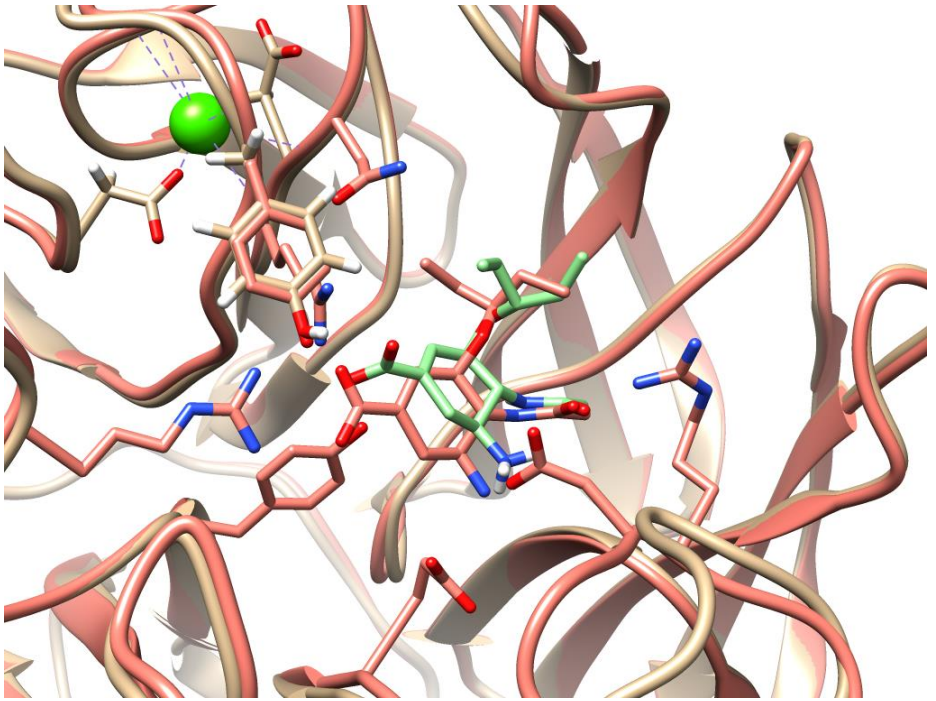
- The result is displayed when the calculation is completed.
- The docking scores are listed in the ViewDock window in the order of the score.
- Clicking on the row of the list changes docking pose displayed.





5. Analysis of the result (2)

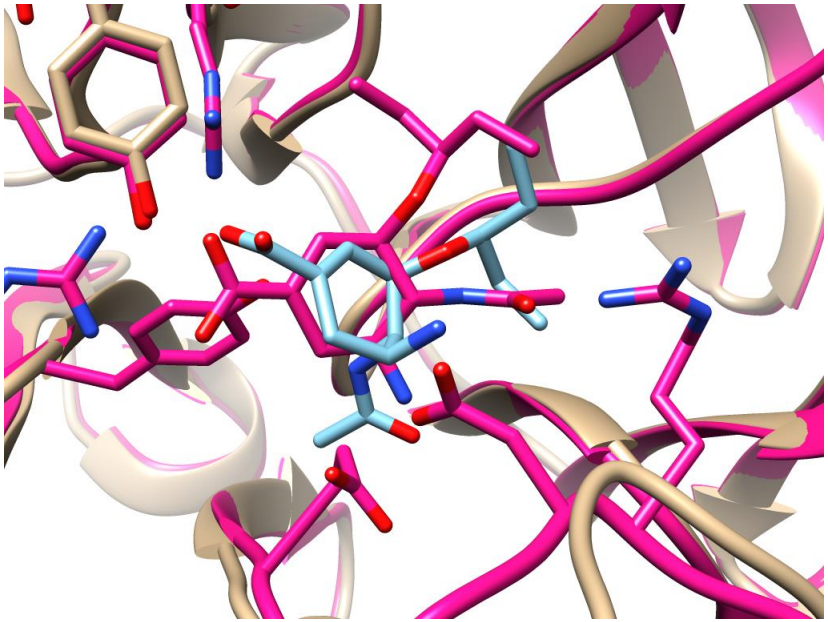
- Compare with the crystal structure.
  1. Open 2HU4_A.pdb.
  2. Choose “Favorites” → “Model Panel” to hide the model of 449381.
  3. Select “2HTY_A.pdb” and “2HU4_A.pdb” in Model Panel and click “match” and “OK.”
- Which pose is the closest to the crystal structure?



Comparison between the 5-th ranked pose and the crystal structure

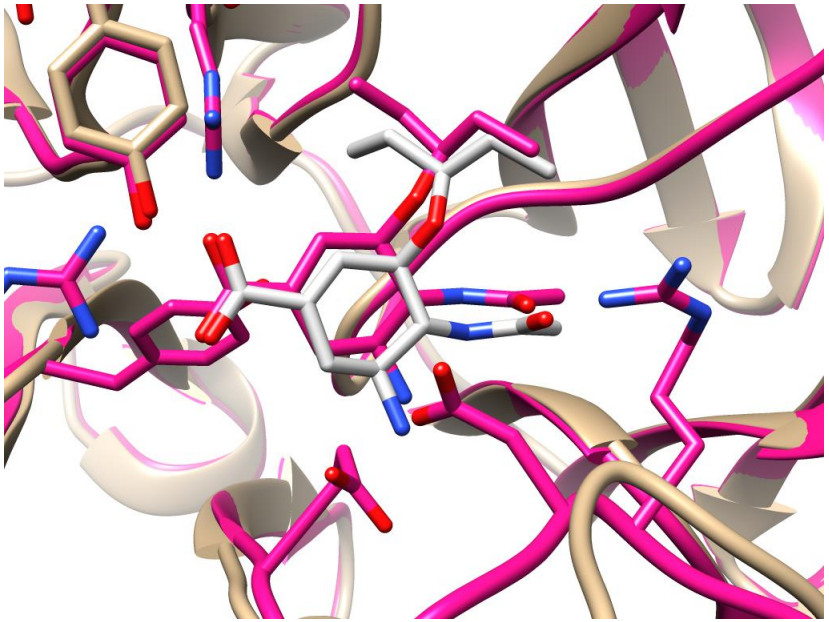
Results from Glide

Top ranked pose



Light blue: top ranked pose
Pink: 2HU4

5-th ranked pose

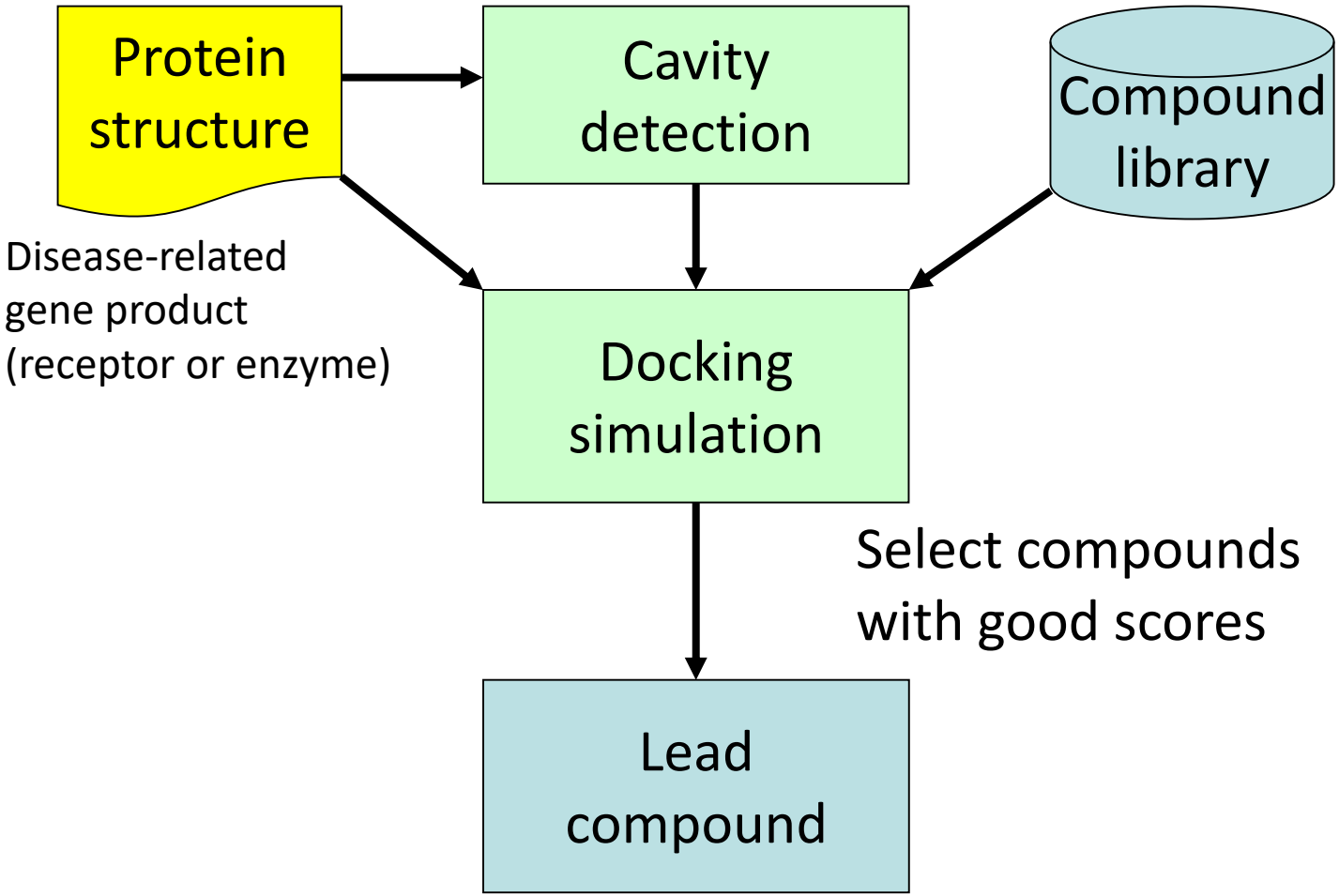


Gray: 5-th ranked pose
Pink: 2HU4

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

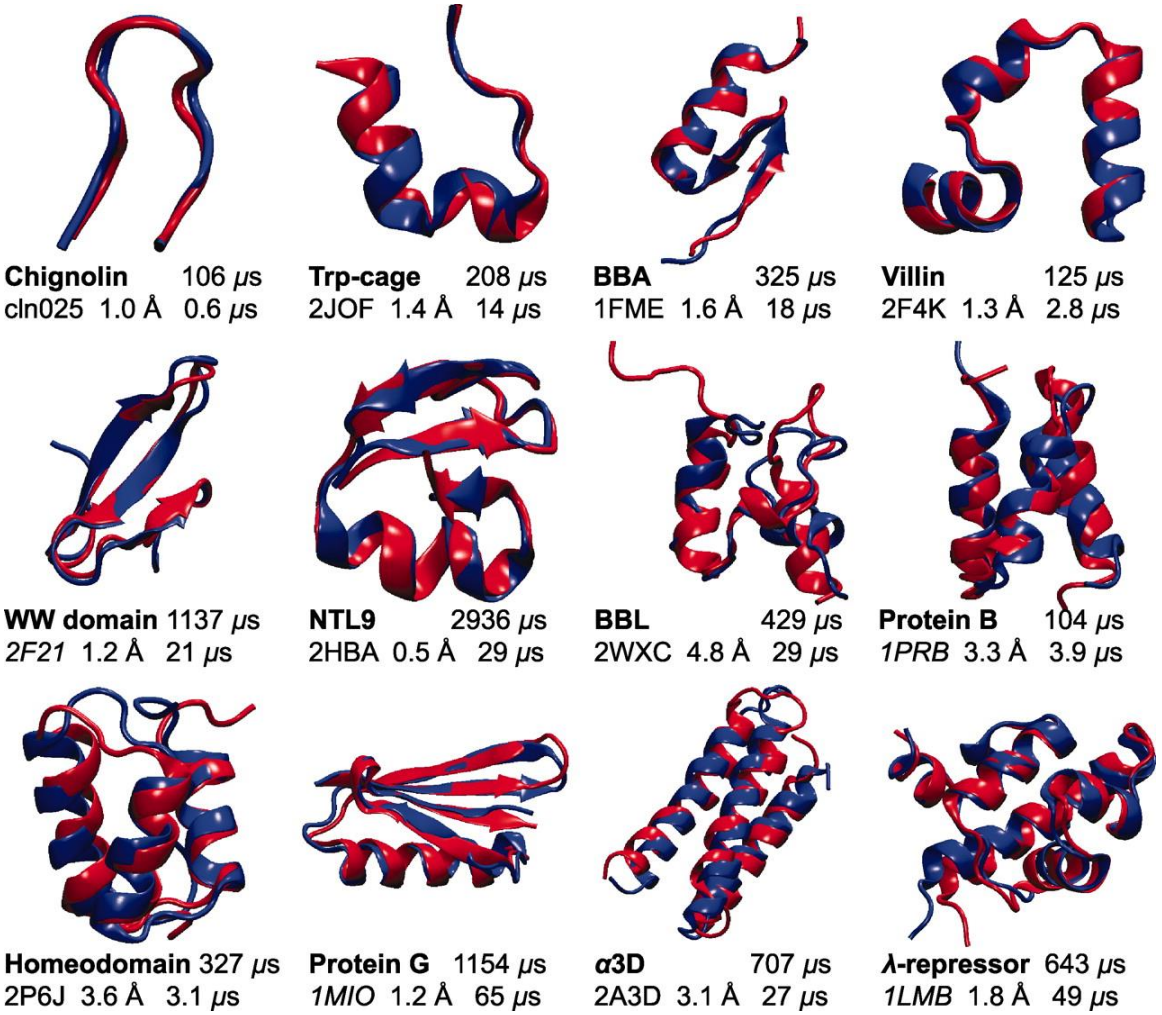
Virtual screening



Applications of molecular simulations

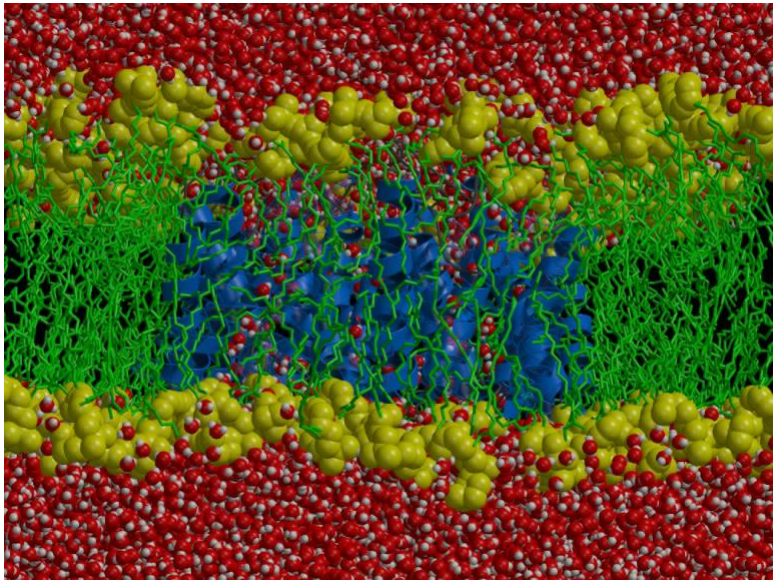
- Reproducing the dynamics
  - MD simulations of chromatin
  - Folding simulations
  - An MD simulation of Aquaporin
  - Ligand-binding simulation
- Free-energy calculation
  - MM-PBSA method
  - Free-energy perturbation and its application

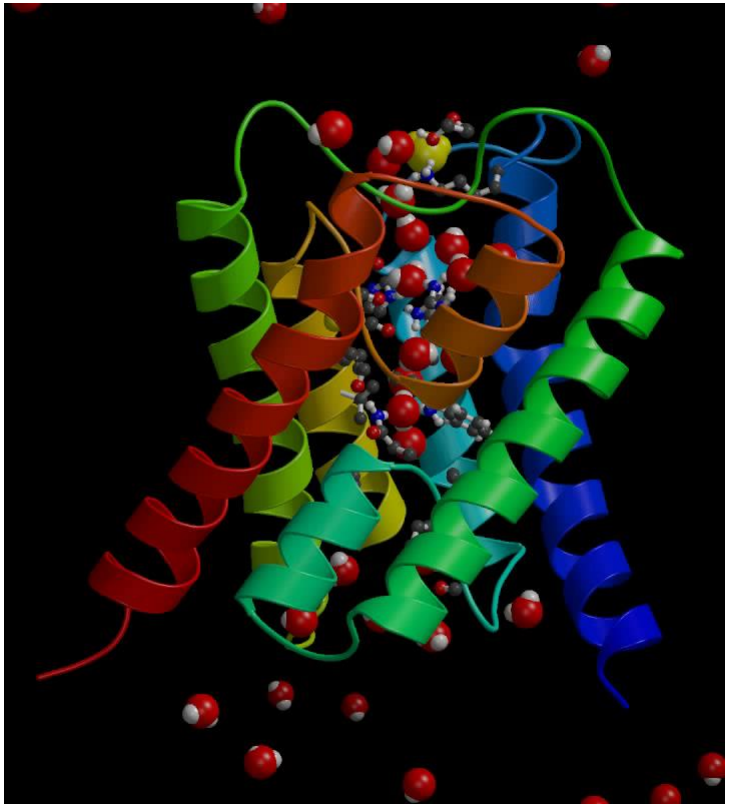
Folding simulations



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

An MD simulation of Aquaporin



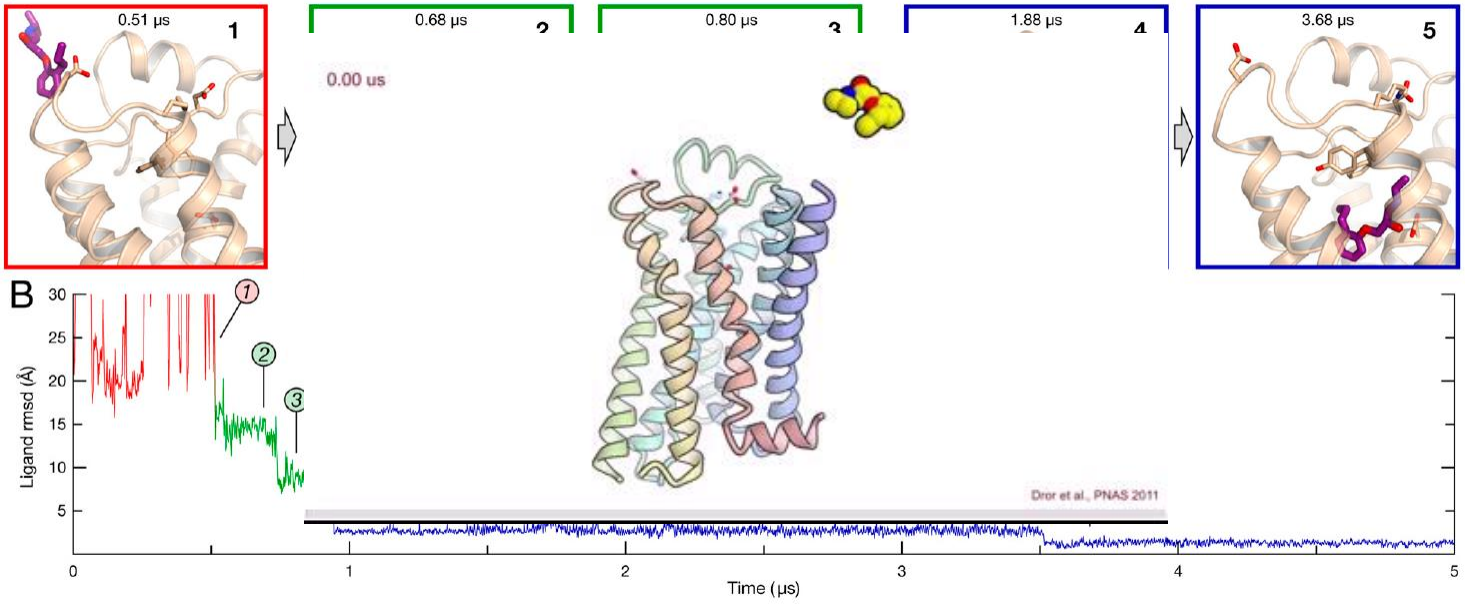


- Water permeation rate
  - Expt.: 3×10^9 sec^{-1}
  - Simulation: 16 / 10 ns
    $\rightarrow 1.6 \times 10^9$ sec^{-1}

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: $3.1 \times 10^7\ M^{-1}\ s^{-1}$



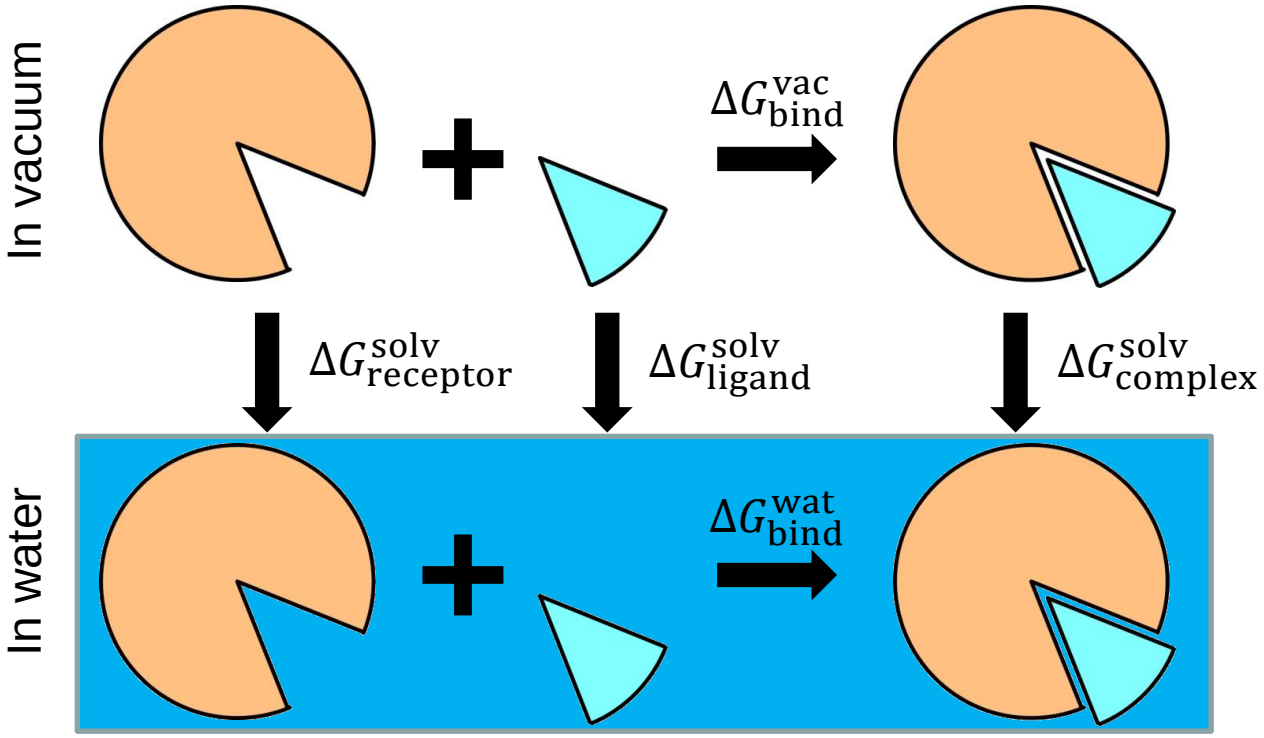
Dror *et al. PNAS* **108**, 13118 (2011).

Free-energy calculation methods

Method	Accuracy	Decomposition	Computational cost
Free-energy perturbation	◎	×	Large (requires many intermediate states)
Thermodynamic integration	○	○	Large (requires many intermediate states)
MM-PBSA	△	○	Small (do not use intermediate state)

MM-PBSA method (1)

- Calculate $\Delta G_{\text{bind}}^{\text{wat}}$ based on the following cycle.



$$\Delta G_{\text{bind}}^{\text{wat}} = \Delta G_{\text{bind}}^{\text{vac}} + \Delta G_{\text{complex}}^{\text{solv}} - \left(\Delta G_{\text{receptor}}^{\text{solv}} + \Delta G_{\text{ligand}}^{\text{solv}}\right)$$

MM-PBSA method (2)

- Perform MD simulations for the free and complex states and calculate average energies and entropies.

$$\Delta G_{\text{bind}}^{\text{vac}} = E_{\text{complex}}^{\text{R-R}} + E_{\text{complex}}^{\text{R-L}} + E_{\text{complex}}^{\text{L-L}} - \left(E_{\text{free}}^{\text{R-R}} + E_{\text{free}}^{\text{L-L}}\right) - T\left[S_{\text{complex}}^{\text{RL}} - \left(S_{\text{free}}^{\text{R}} + S_{\text{free}}^{\text{L}}\right)\right]$$

- Calculate solvation free energies using the Poisson-Boltzmann/Surface Area method and calculate their averages ($\Delta G_{\text{complex}}^{\text{solv}}$, $\Delta G_{\text{receptor}}^{\text{solv}}$, and $\Delta G_{\text{ligand}}^{\text{solv}}$).

Free-energy perturbation (1)

- Free energy
  - $G = -RT\log Z,\ Z = \int d\mathbf{r}\exp[-E(\mathbf{r})/RT]$
- Free-energy difference between states 0 and 1
  - $\Delta G = G_1 - G_0 = -RT\log Z_1/Z_0$
  - $\Delta G = -RT\log\frac{\int d\mathbf{r}\exp[-E_1(\mathbf{r})/RT]}{\int d\mathbf{r}\exp[-E_0(\mathbf{r})/RT]} =$

    $-RT\log\frac{\int d\mathbf{r}\exp\left[-\left(E_1(\mathbf{r})-E_0(\mathbf{r})\right)/RT\right]\exp[-E_0(\mathbf{r})/RT]}{\int d\mathbf{r}\exp[-E_0(\mathbf{r})/RT]} =$

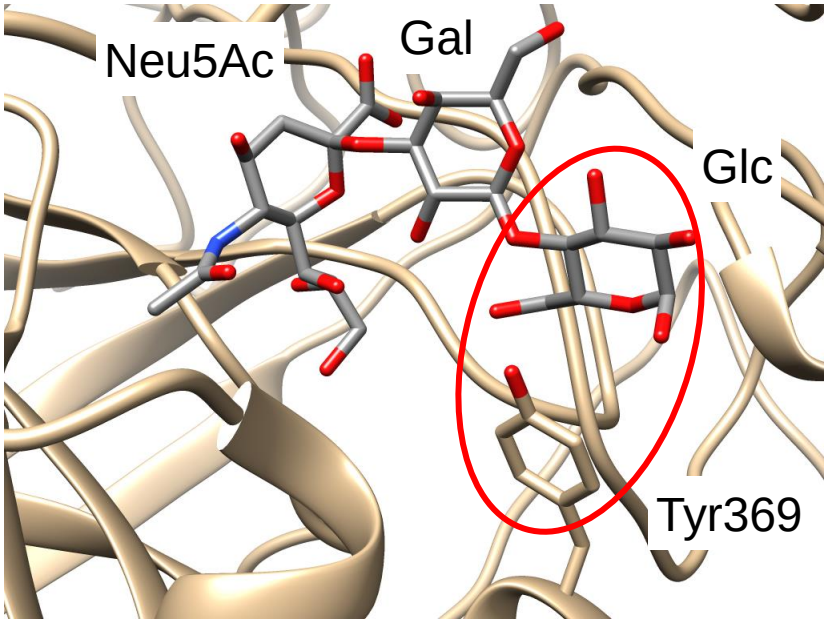
    $-RT\log\langle\exp[-\Delta E(\mathbf{r})/RT]\rangle_0$

Free-energy perturbation (2)

- This method does not work well if $\Delta E(\mathbf{r})$ is large.
- Intermediate states interpolating the initial and final state are considered.
  - $E_i(\mathbf{r}) = (1-\lambda_i)E_0(\mathbf{r}) + \lambda_i E_1(\mathbf{r})$
    $(0 = \lambda_0 < \lambda_1 < \cdots < \lambda_{N-1} < \lambda_N = 1)$
  - $\Delta G = \sum_{i=1}^{N}(G_i - G_{i-1})$
  - $G_i - G_{i-1} =$
    $-RT\log\left\langle \exp\left[-\left(E_i(\mathbf{r}) - E_{i-1}(\mathbf{r})\right)/k_{\mathrm{B}}T\right]\right\rangle_{i-1}$

Interactions between a virus protein and its receptor sugar chain

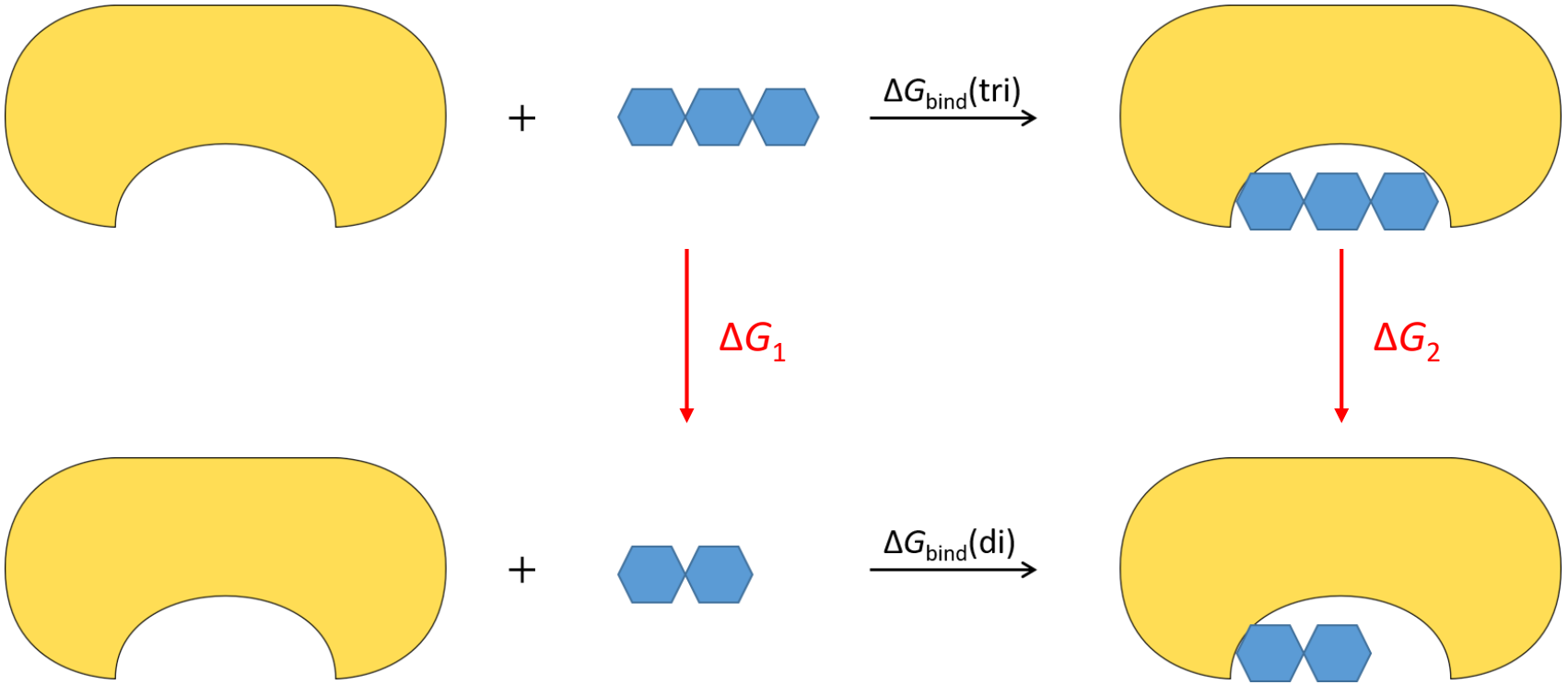
- A protein from Mumps virus
- It binds to a sugar chain on the surface of host cells.
- It has been said that the Neu5Ac residue at the non-reducing end of the receptor sugar chain is important for the interaction with the protein.
- The crystal structure revealed the Glc residue at the third position is also important for the complex formation through interaction with Tyr369.
- We calculated the contribution of Glc and Tyr369 to ΔG_{bind}.



Sugar chain:
Neu5Acα(2→3)Galβ(1→4)GlcNAc

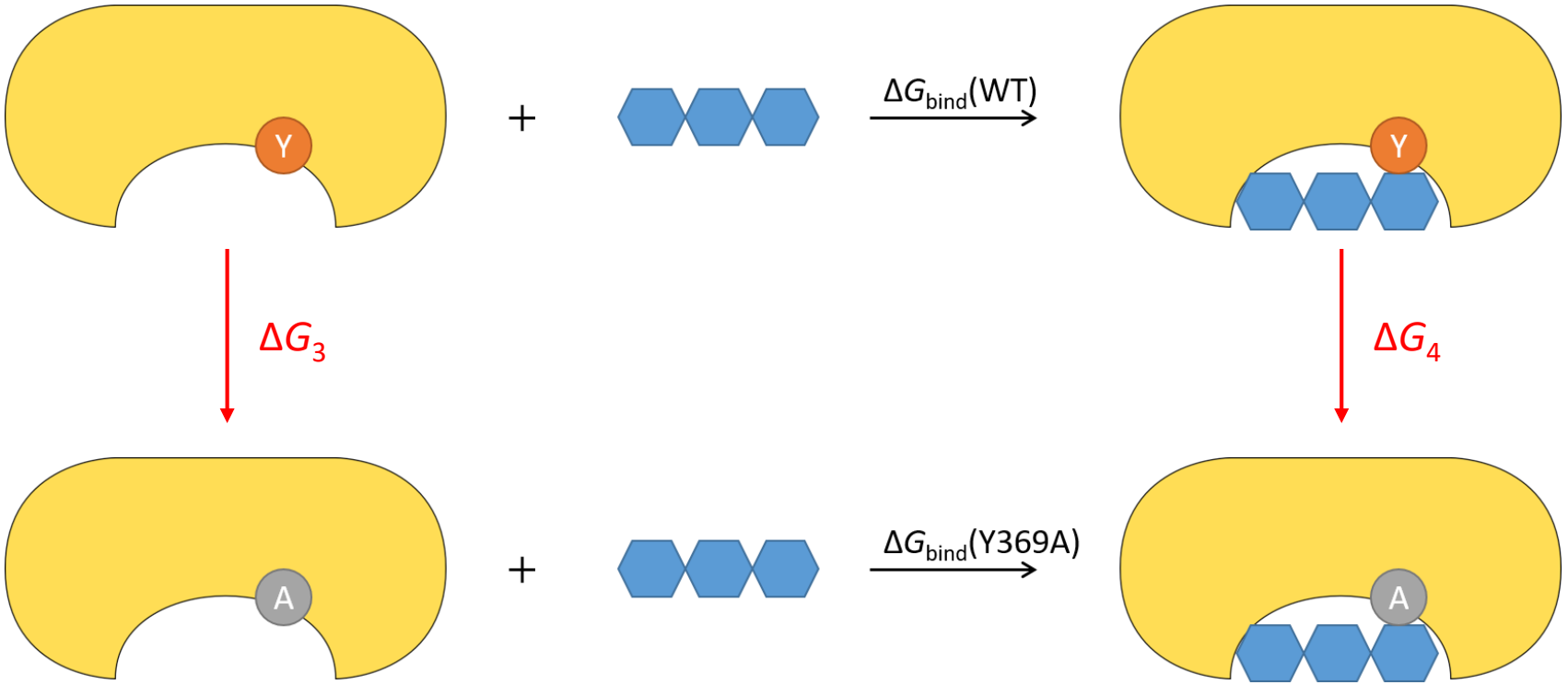
Calculation of the contribution of Glc

$\Delta\Delta G = \Delta G_{bind}(tri) - \Delta G_{bind}(di)$



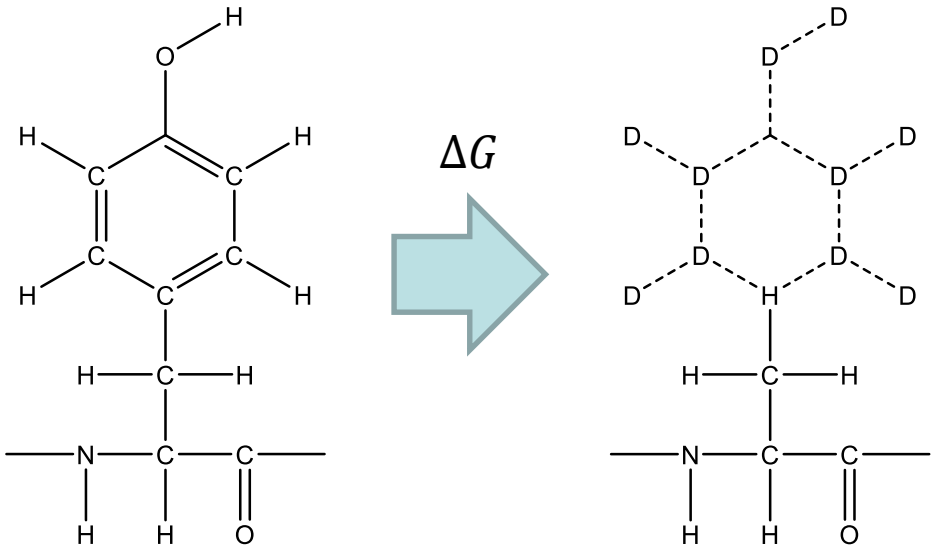
Calculation of the contribution of Tyr369

$\Delta\Delta G = \Delta G_{bind}(WT) - \Delta G_{bind}(Y369A)$



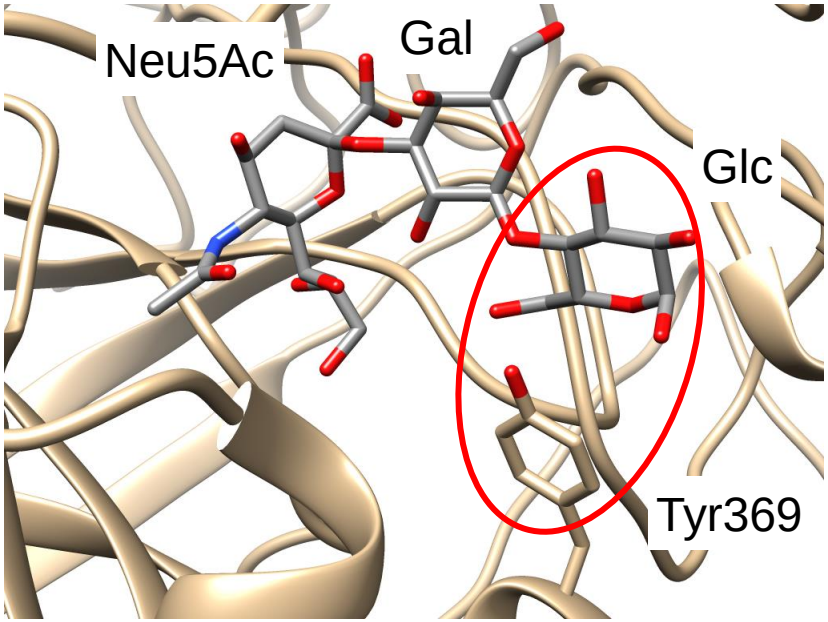
Alchemical transformation

- Free-energy change of alchemical transformation is calculated.
- In the calculation of the free-energy difference between the wild type and the Y369A mutant, $E_1(\mathbf{r})$ gives the energy of the model where the Cγ atom of Y369 is replaced with H and the other atoms of *p*-hydroxyphenyl group are replaced with dummy atoms that do not have nonbonded interactions with other atoms.



Results

- Contribution of Glc
  - $\Delta G_1 = -251.2 \pm 0.8$ kJ mol^{-1}
  - $\Delta G_2 = -234.6 \pm 1.2$ kJ mol^{-1}
  - $\Delta\Delta G = -16.6 \pm 1.9$ kJ mol^{-1}
  - Glc contributes to the stabilization of the complex structure.
- Contribution of Tyr369
  - $\Delta G_3 = 57.6 \pm 0.8$ kJ mol^{-1}
  - $\Delta G_4 = 64.8 \pm 1.2$ kJ mol^{-1}
  - $\Delta\Delta G = -7.2 \pm 2.0$ kJ mol^{-1}
  - Tyr369 contributes to the stabilization of the complex structure.



Experiment:
K_D(di) = 56 µM
K_D(tri) = 500 µM

$\Delta\Delta G = -5.2$ kJ mol^{-1}

Kubota *et al.*, *PNAS* **113**, 11579 (2016).

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 2HTY?
- Create a figure of the closest pose superimposed on the crystal structure.
- Create a similar figure for the top ranked pose. Discuss why this pose has such a good score.

Instructions for submission

- Put the results and discussion in the body of email.
- Send the two figures (image files) as attachments.
- Put “Molecular modeling” in the Subject field.
- 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).