Computational Biochemistry

Lecture 7 Docking

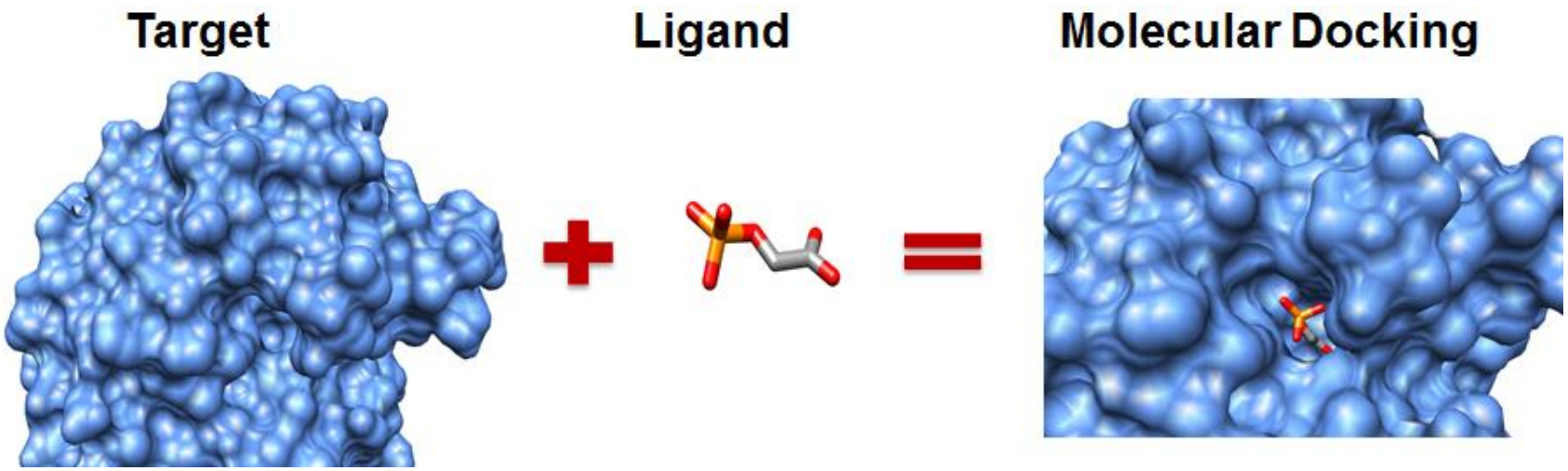
Ligand Docking (Handle with Care!)

Accuracy – Ability to discriminate binders from non-binders (Scoring)

– Ability to identify bound conformation (Internal Energies)

– Ability to identify binding site (Search Algorithm)

Efficiency – Conformation searching and pose searching are inversely proportional to ligand flexibility (Smaller is Better)



Molecular docking challenge

- How to approximate complex physical and thermodynamic interactions?
- Employ rigid or flexible structures for ligand and receptor (Side-chains or Back-bone flexible)
- Treat with full atomic detail or simplified models?
- Which docking energy function is best?

Molecular docking challenge

A successful docking application needs :

1. **Search algorithm**

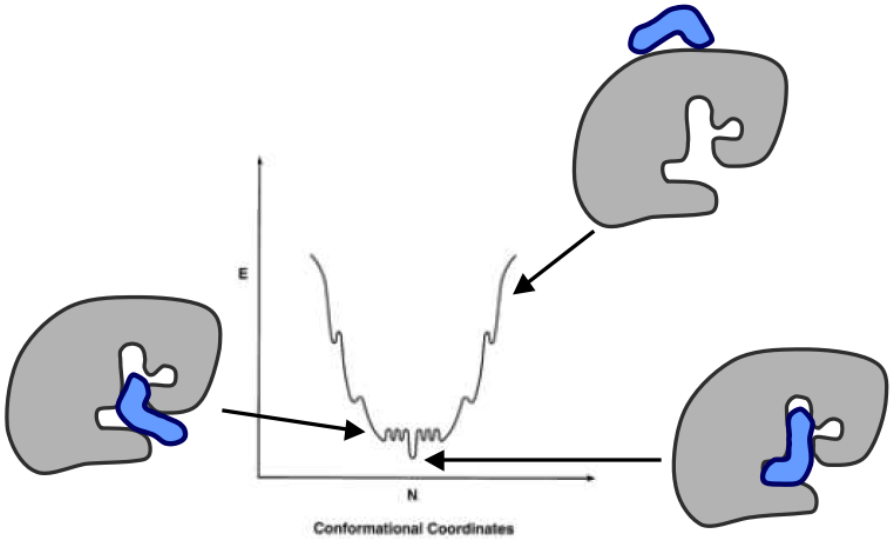
   A method to explore the ligand-receptor conformation space for plausible poses.

2. **Scoring Function**

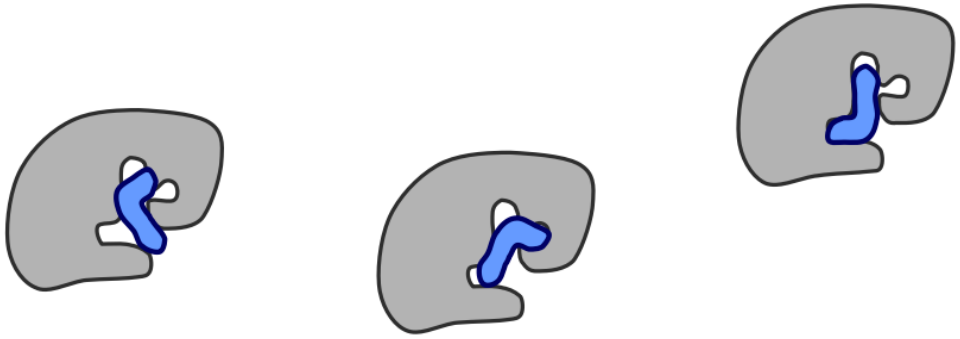
   A method to relatively order those plausible poses.

The Molecular Docking Challenge

Where is the binding site?



What is the optimum pose?



Factors Affecting Binding

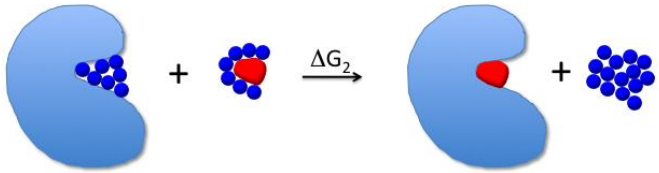
Electrostatic Interactions (relatively long-range, proportional to 1/R):
hydrogen bonds, salt bridges, charge-charge

Dispersive Interactions (short range)
Van der Waals attractions (proportional to $1/R^6$)
Van der Waals repulsions (proportional to $1/R^{12}$)

Hydrophobic contacts (short range)
depend on displacing solvent from the binding site

Tight binding requires both the correct shape of **interacting surfaces** and **polarities**.

The binding affinity is the energetic difference between the bound and free states which requires solvation and entropy to be considered



Specificity is driven by shape and hydrogen bond complementarity (easy to quantify)
Affinity is driven by hydrophobic and entropic effects (hard to quantify)

Applications of Docking

Estimating the binding affinity:

- Searching for lead structures (drug candidates) for protein targets
- Comparing a set of inhibitors
- Estimating the influence of modifications in lead structures
- *De Novo* ligand design
- **Design of targeted combinatorial libraries**
- **Virtual screening**

Predicting the molecular complex:

- Understanding the binding mode / principle
- Optimizing lead structures
- Determining ligand positions in crystal structures
- Determining specificity

Approximations in Docking

Eliminate explicit waters (what about desolvation?)

Approximate *desolvation*

Eliminate dynamics (what about entropy?)

Approximate *entropy*

Employ general force field (what about precision?)

Approximate *enthalpy*

Ignore the unbound state (what about ΔG?)

Approximate ΔG

Scoring Functions

Instead of using: $\Delta G_{Binding} = \Delta G_{Complex} - \Delta G_{Ligand} - \Delta G_{Receptor}$

Develop a "scoring function" that attempts to empirically reproduce experimental interaction energies, without rigorous attention to physics

Use: $\Delta G_{Binding} \approx \sum_{\text{interactions}} f_i E_i$

The interactions (E_i) might include:

hydrogen bonds
electrostatic interactions
hydrophobic contacts
solvent exclusion volume, among others...

Each contribution has an adjustable weighting factor (f_i).

Scoring Functions General or Specific?

In determining the weighting factors (f_i) the developer must choose how broadly or how narrowly the scoring function is to be applied.

$$\Delta G_{Binding} \approx \sum_{\text{interactions}} f_i E_i$$

Is the function to be used for all classes of interactions? Or only some? For protein-protein only, or protein-drug only, or only for a particular class of drug?

There are many Scoring Functions. The AutoDock 3 function is:

$$\Delta G = f_{ELEC} \sum_{i,j} \left(\frac{q_i q_j}{\varepsilon_{R_{ij}} R_{ij}} \right) + f_{VDW} \sum_{i,j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right) + f_{HBOND} \sum_{i,j} \chi_t \left(\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right) + f_{SOL} \sum_{i,j} (S_i V_j + S_j V_i) e^{\left(\frac{-R_{ij}^2}{2\sigma^2} \right)}$$

The *f* coefficients are determined empirically from a multi-linear regression (MLR) to a set of protein–ligand complexes with known binding constants.

Because the *f* coefficients are not based on physics, scoring functions are considered empirical

Electrostatics

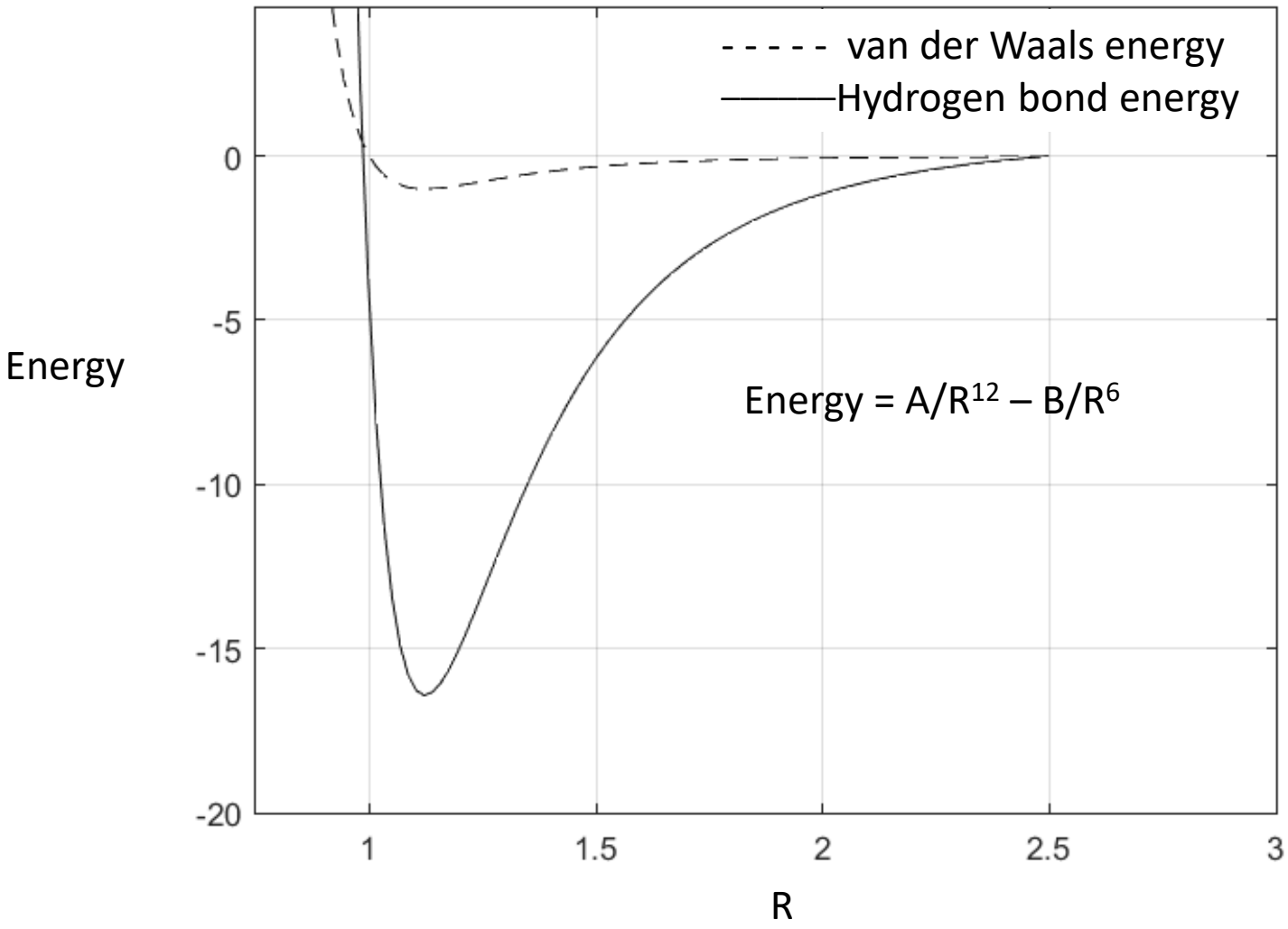
- Electrostatic interaction energy is calculated as a sum of interactions between partial atomic charges, using Coulombs law

$$V_{Electrostatic} = \frac{q_i q_j}{4\pi\varepsilon R_{ij}}$$

If the quantity of ligands is low, we can derive charges specific to the ligands.

What if the quantity of ligands is high? Like hundreds of thousands of them?

Modeling van der Waals Interactions (Lennard-Jones Potentials)



Van der Waals Interactions

Non-bonded interactions that are not electrostatic (e.g. between atoms in noble gas) are labeled van der Waals interactions

Contains **dispersion and repulsion components**

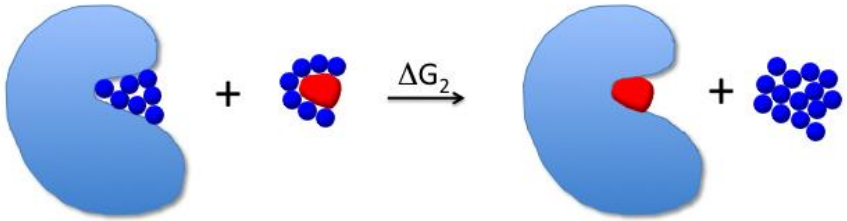
Dispersion interactions always attractive. Arise from instantaneous dipoles that occur during fluctuations within the molecular electron cloud

Repulsive interactions occur at short-range. Also labeled exchange, or overlap, forces. They occur between electrons with the same spin that must not occupy same region in space (Pauli exclusion principle)

Homework, what values of A and B result in a Lennard-Jones function with a minimum at 3.5 A with a minimum energy of -1 kcal/mol? Hint, use Excel to create A/R^{12} and $-B/R^{6}$ and vary A and B until you find a solution

Desolvation Energy

$$f_{SOL} \sum_{i,j} (S_i V_j + S_j V_i) e^{\left(-\frac{R_{ij}^2}{2\sigma^2}\right)}$$



i = index of atoms in the ligand,
j = index of atoms in the receptor

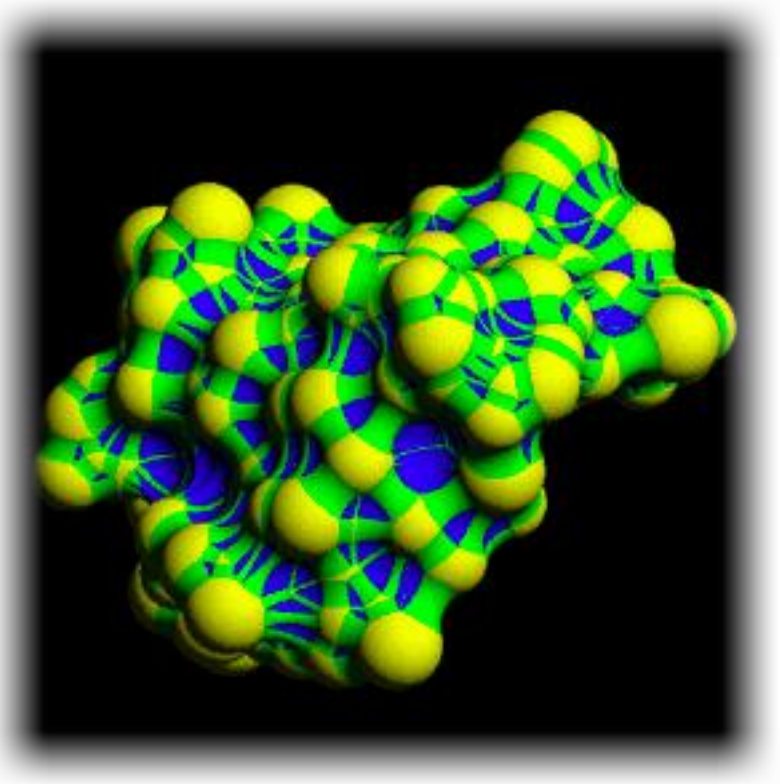
S_i = solvation term for atom *i* — a function of the polarity of atom *i*

Vi = atomic fragmental volume of atom *i* — volume of atom *i* shielded by the interaction

R_{ij} = distance between atom *i* and atom *j* (in Å)

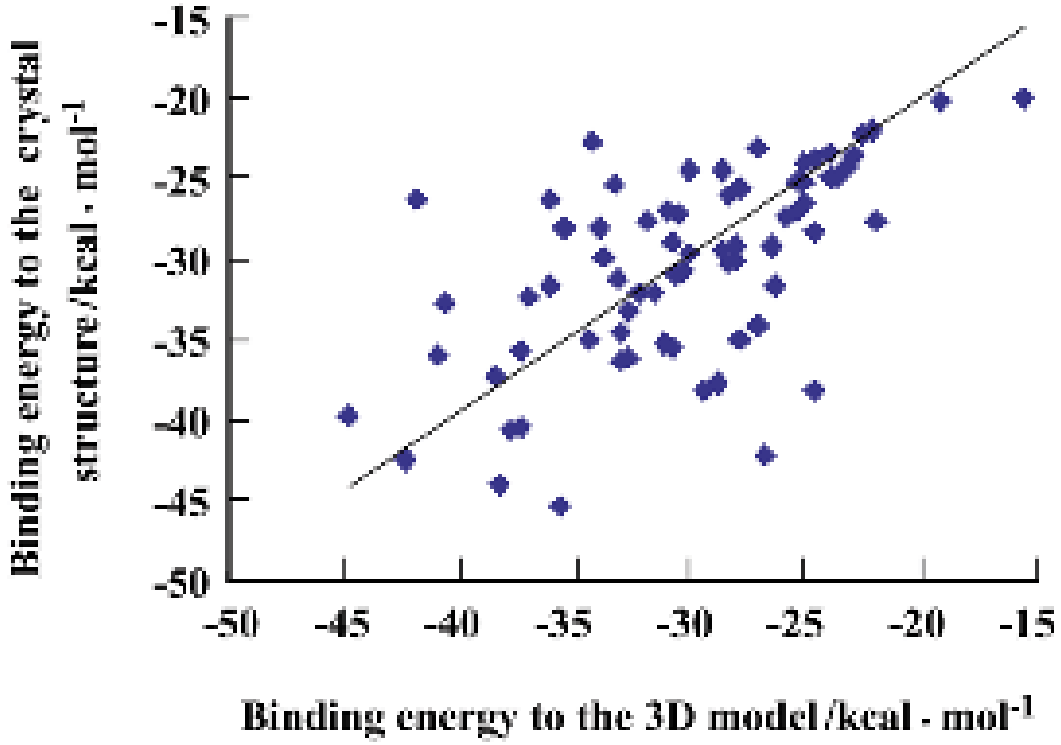
σ = distance constant (3.5 Å)

Surface Representation



- Each atomic sphere is given the van der Waals radius of the atom
- Rolling a Probe Sphere over the Van der Waals Surface leads to the Solvent Reentrant Surface or **Connolly surface**

Empirical Fitting to Experimental Data (choosing "*f*")



Creating a scoring function for docking:

1) Assemble a set of known 3D structures for protein-ligand complexes with known binding energies

2) Adjust "f" coefficients until the predicted interaction energy gives a best fit to the experimental data

Types of docking

Rigid docking

Lock and Key

In rigid docking, both the internal geometry of the receptor and ligand is kept fixed and docking is performed.

Flexible docking

Induced fit

An enumeration on the rotations of one of the molecules (usually smaller one) is performed. Every rotation the surface cell occupancy and energy is calculated; later the most optimum pose is selected.

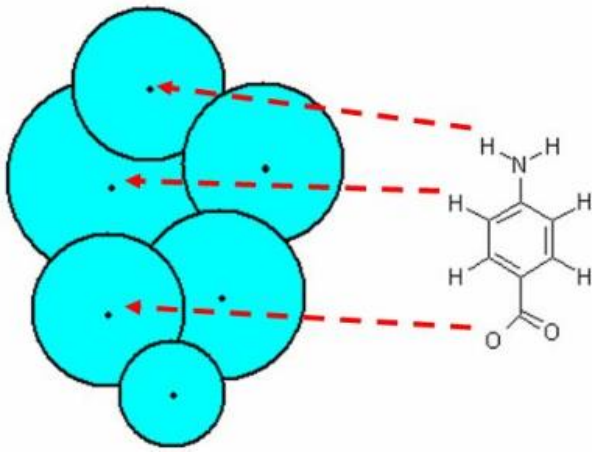
Blind docking v.s. Binding-site based docking

- Blind docking (Global docking) refers to docking a ligand to the whole surface of a protein without any prior knowledge of the target pocket.

- Blind docking was introduced for the detection of possible binding sites and modes of peptide ligands by scanning the entire surface of protein targets.

Rigid docking

- Historically the first approaches
- Protein and ligand are fixed.
- Search for the relative orientation of two molecules with lowest energy.
  - Protein-protein docking
  - Both molecules usually considered rigid
  - First apply steric constraints to limit search space and the examine energetics of possible binding conformations.



Flexible docking

A more accurate view of this process was first presented by Koshland in the **induced fit model.**

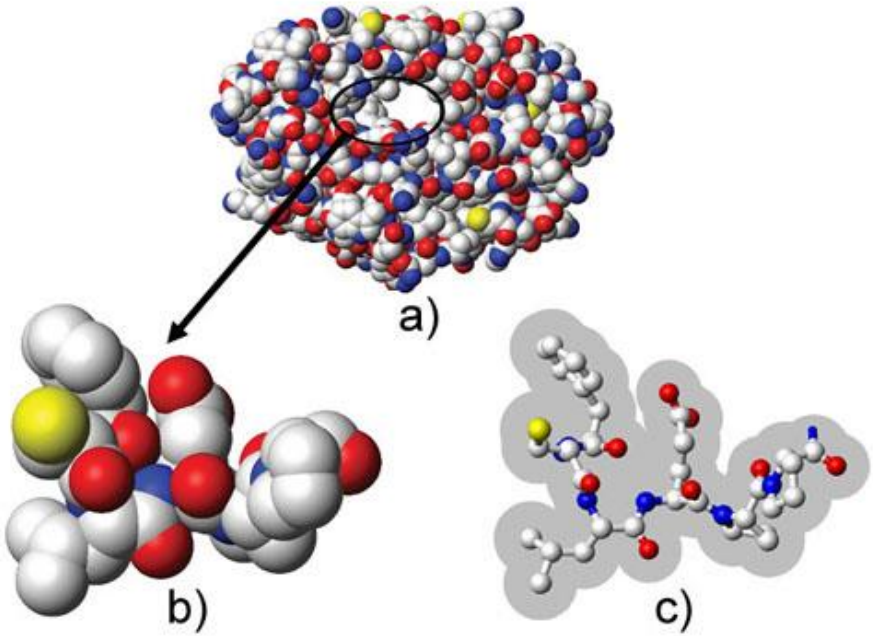
In this model the 3D structure of the ligand and the receptor *adapt to each other during the binding process*.

It is important to note that not only the structure of the ligand but also the structure of the receptor changes during the binding process. This occurs because the introduction of a ligand modifies the chemical and structural environment of the receptor.

Treating Induced Fit: Soft Receptors

Soft receptors can be easily generated by reducing the van der Waals repulsive ($1/R^{12}$) contributions to the total energy score.

This makes the receptor "softer", thus allowing, for example, a larger ligand to fit in a binding site determined experimentally for a smaller molecule.



a) van der Waals representation of a target receptor.

b) Close up image of a section of the binding site with normal van der Waals properties.

c) Same section of the binding site as shown in b) but with reduced radii for the atoms in the receptor.

This type of soft representation allows ligand atoms to enter the grey shaded area without incurring a high energetic penalty.

http://cnx.org/content/m11456/latest/

Treating Induced Fit: Side Chain Rotations

Rotations around single bonds, such as in side chains is a “natural” way to model induced fit.

Selection of which torsion angles to permit to rotate is usually the most difficult part of this method because it requires a considerable amount of a priori knowledge of alternative binding modes for a given receptor.

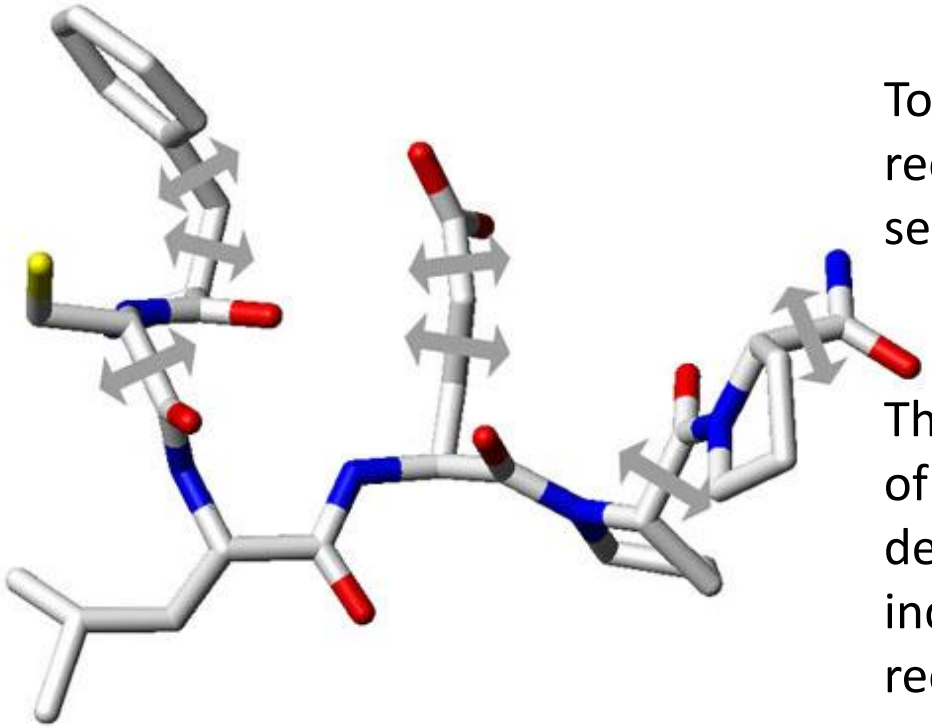
Alternatively, probable side chain orientations may be selected from rotamer libraries

The principle problem with this method is that is adds significantly to the time required for the calculation because of the exponential number of permutations of side chain rotamers in a binding site

http://cnx.org/content/m11456/latest/

Treating Induced Fit: Side Chain Rotations

Stick representation of a section of a binding site



To approximate the flexibility of the receptor it is possible to carefully select a few degrees of freedom.

These are usually the torsional angles of side chains that have been determined to be critical in the induced fit effect for a specific receptor.

In this example the selected torsional angles are represented by arrows.

http://cnx.org/content/m11456/latest/

Treating Induced Fit
Multiple Receptor Conformations

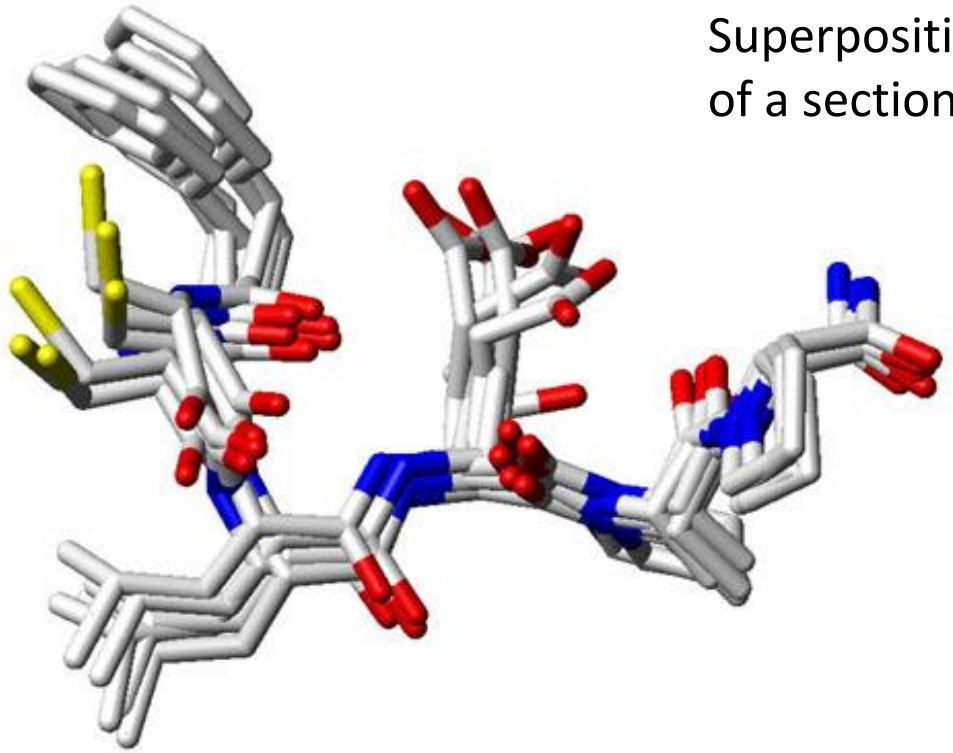
One possible way to represent a flexible receptor is to use multiple static receptor structures. This concept reflects the idea that proteins in solution do not exist in a single minimum energy static conformation but are in fact constantly jumping between low energy conformational sub-states.

In this way, the best description for a protein structure is that of a conformational ensemble of slightly different protein structures coexisting in a low energy region of the potential energy surface.

Thus, the binding process can be thought of not as an induced fit model as described by Koshland in 1958, but more like a selection of a particular sub-state from the conformational ensemble that best complements the shape of a specific ligand.

http://cnx.org/content/m11456/latest/

Treating Induced Fit
Multiple Receptor Conformations

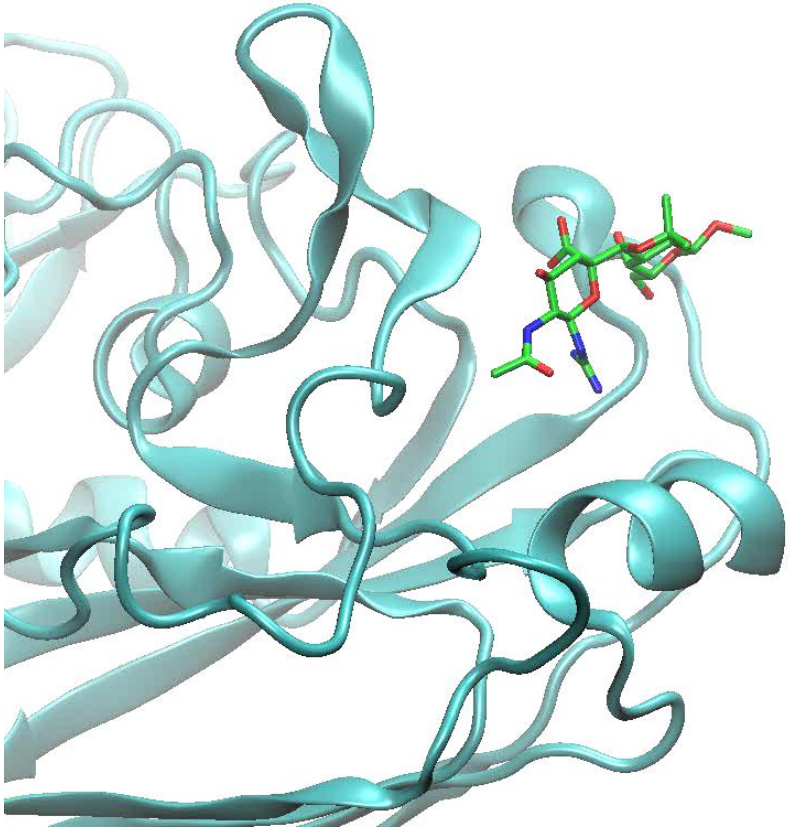


Superposition of multiple conformers of a section of a binding site.

These can be either considered individually as rigid representatives of the conformational ensemble or can be combined into a single representation that preserves the most relevant structural information.

Use MD after Docking

Performing an MD simulation of the docked complex provides a method to detect poor poses.



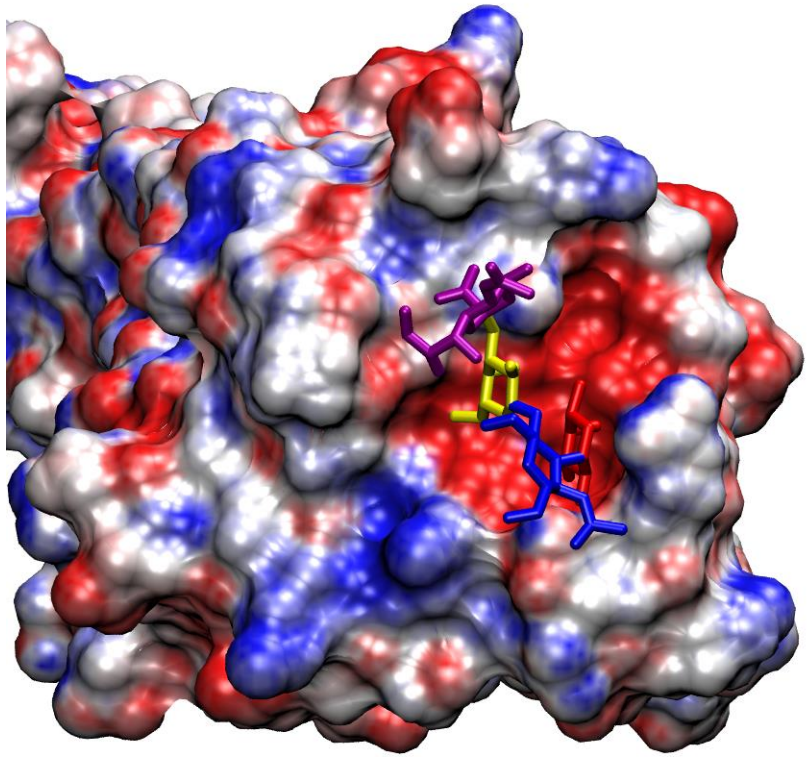
Docking Conclusions

- Scoring functions are very approximate!
- Docking is better at identifying possible or impossible ligands, than at ranking a group of possible ligands
- Docking accuracy is enhanced if the binding site is already known
- It is often difficult to confidently identify the best poise – so try to combine with additional data – such as NMR chemical shifts, or point mutagenesis
- Although approximate, docking is often the only alternative to crystallography

Virtual screening example

Goal for this study:

- Find small molecules attached to the natural ligand, and expect to enhance binding affinity or specificity.



Virtual screening

- Molecular docking is the core of virtual screening.
- It aims at prediction of the modes and affinities of non-covalent binding between a pair of molecules. Oftentimes, the molecules consist of a macromolecule (the receptor) and a small molecule (the ligand).
- The multidimensional search space of the ligand includes the degrees of freedom of its translation, rotation, and torsions of flexible bonds that may exist within it.

AutoDock Vina

AutoDock Vina[3]:

- AutoDock Vina achieves an approximately two orders of magnitude speed-up compared to AutoDock4.
- Scoring function in AutoDock Vina does not require atomic partial charge information.
- Preparation for running AutoDock Vina is simpler than that with AutoDock4.

1. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of computational chemistry* **2010,** *31* (2), 455-461.

ZINC database

ZINC[1]

1. A free database of commercially-available compounds for virtual screening, and a great source of input small molecules for AutoDock.

2. Contains over **18 million** compounds in ready-to-dock, 3D formats.

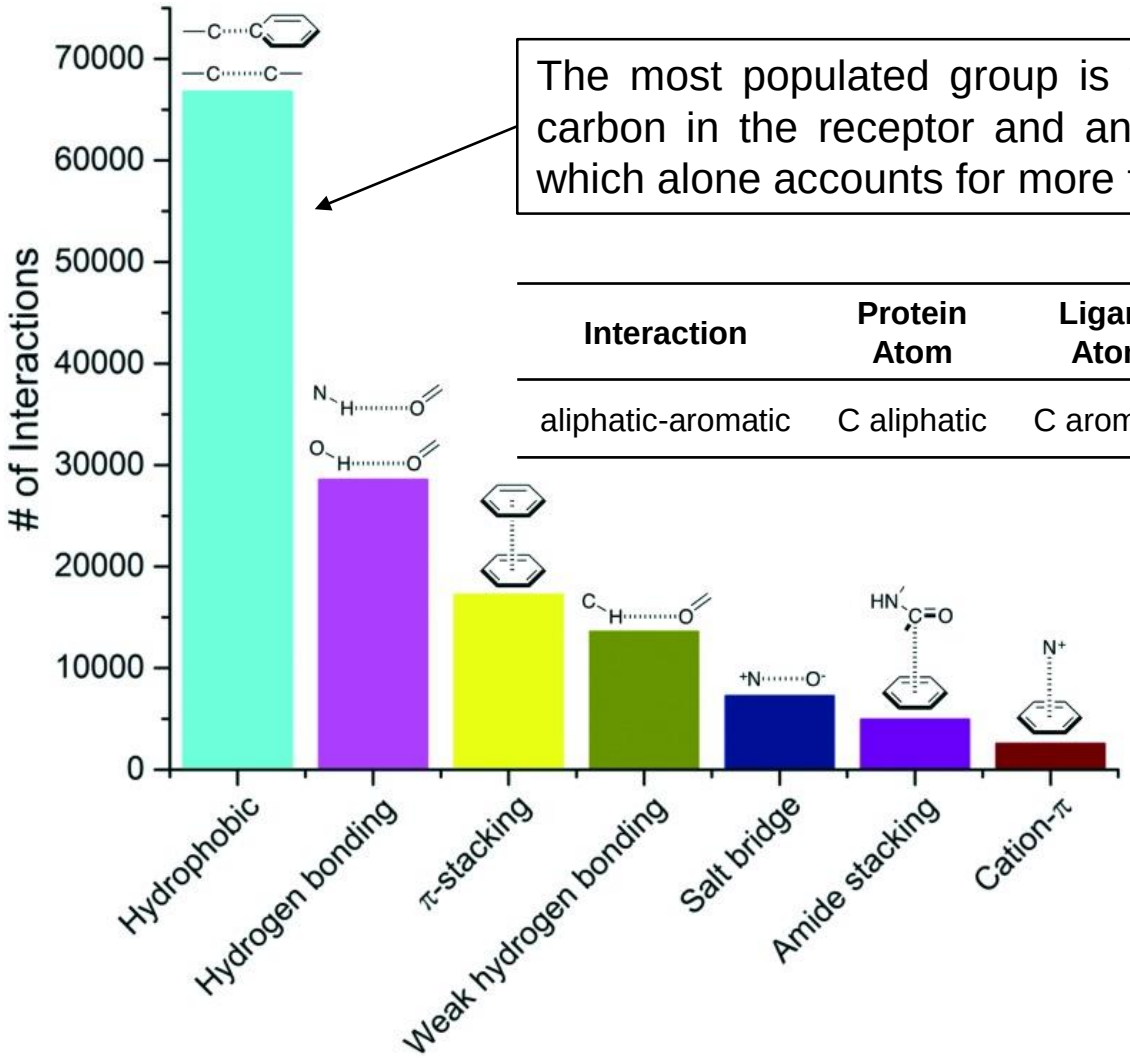
3. Provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF).

4. Works together with AutoDock for small molecule virtual screening.

1. Irwin and Shoichet, *J. Chem. Inf. Model*. **2005**, *45* (1):177-82.

Non-covalent Protein-ligands Interactions

Frequency distribution of the most common non-covalent interactions observed in protein–ligands extracted from the PDB[1]



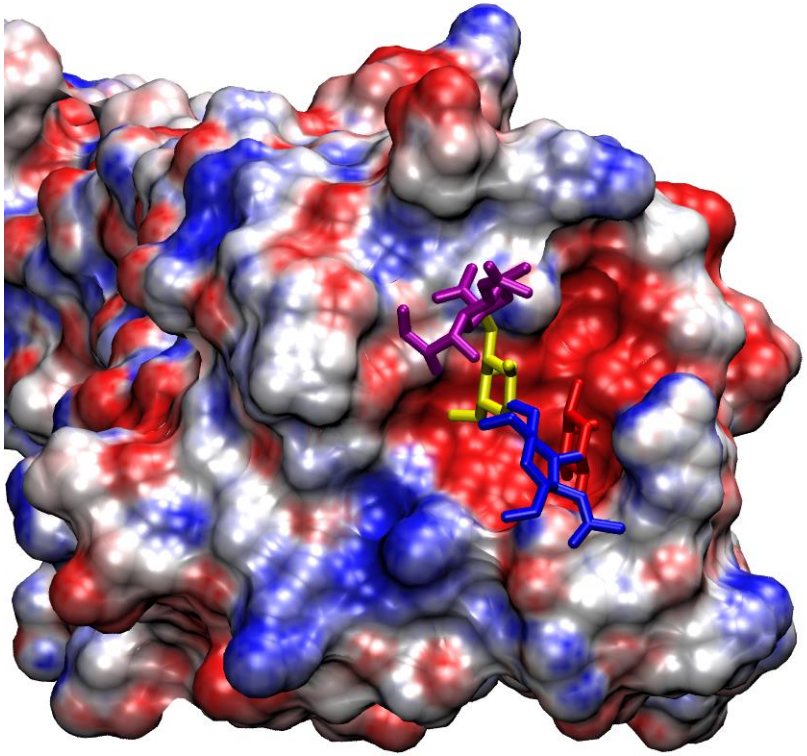
The most populated group is the one formed by an aliphatic carbon in the receptor and an aromatic carbon in the ligand, which alone accounts for more than 42 000 interactions

Interaction	Protein Atom	Ligand Atom	Distance (Å)	# of Interactions
aliphatic-aromatic	C aliphatic	C aromatic	≤ 4.0Å	42443

1. Ferreira de Freitas, R.; Schapira, M., *Medchemcomm* **2017,** *8* (10), 1970-1981.

Virtual screening example

- Top 2 of the most common non-covalent interactions observed in protein–ligands extracted from protein databank are hydrophobic and hydrogen bonding interactions
- 45% of these ligands have aromatic rings.



Non-covalent Protein-ligands Interactions

Our targets are **SMALL AROMATIC** molecules:

1. ≤ 15 heavy atoms
   SMALL -- no more than 2 phenyl rings

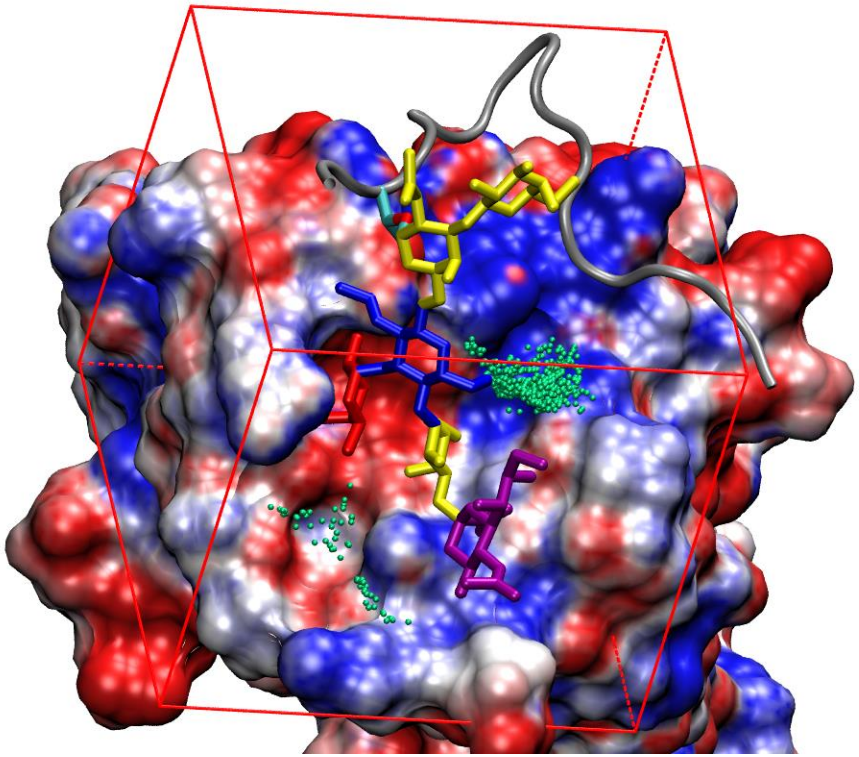
2. ≥ 5 aromatic carbon atoms
   AROMATIC -- at least 1 aromatic ring

ZINC molecule database contains over 18 million small molecules.

After filtering, we have 300,006 potential choices.

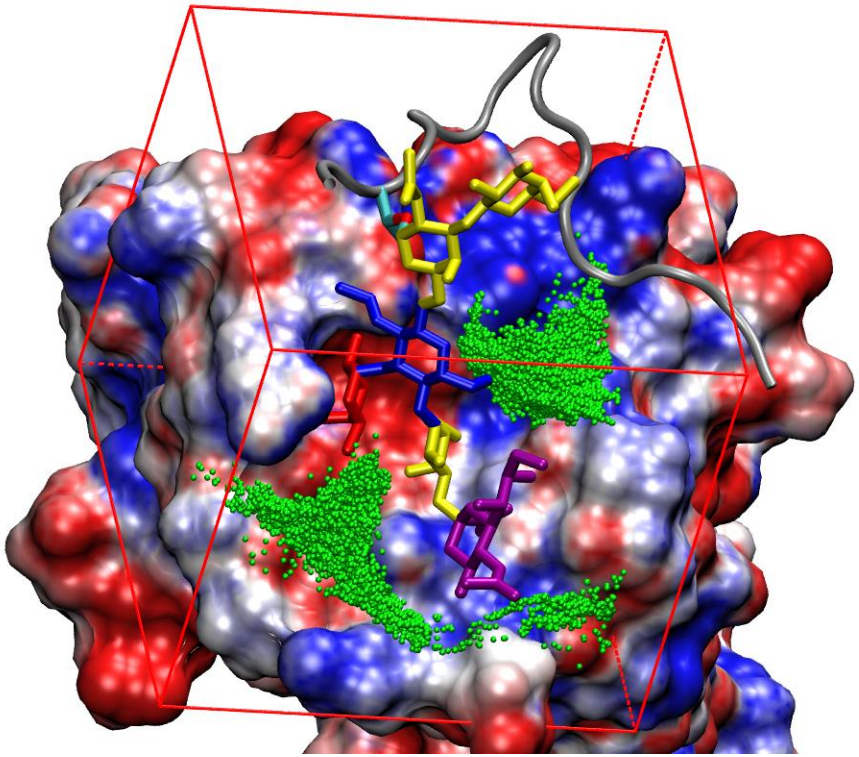
Virtual Screening Near Glycan

Autodock-vina energy score
-6.9 ~ -6.0 kcal/mol



Total number of 682

Autodock-vina energy score
-5.9 ~ -5.0 kcal/mol

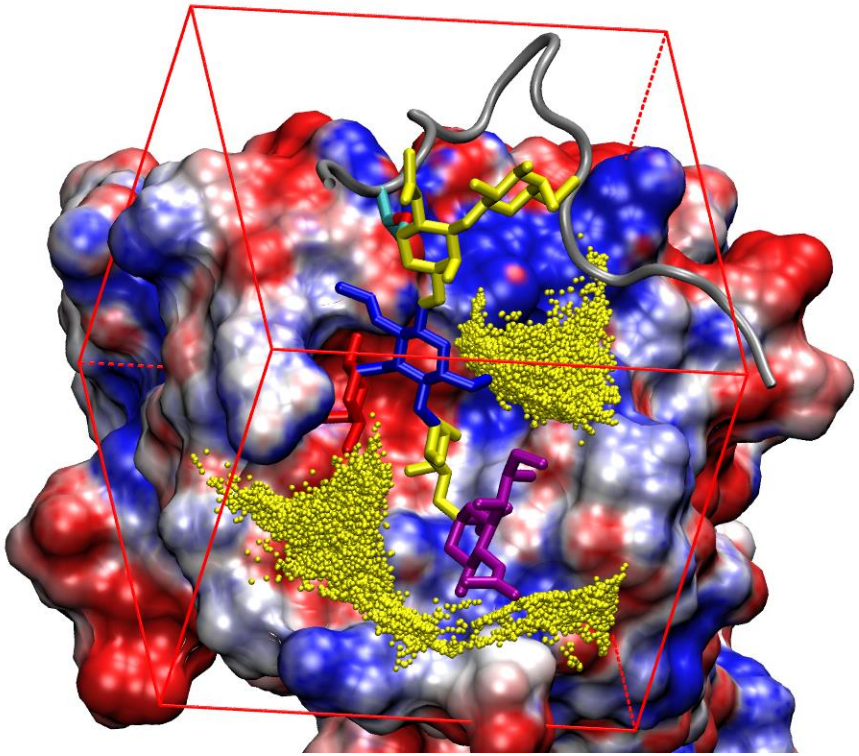


Total number of 73,874

- Grid box centers at the geometric center of pyranose ring for GlcNAc in sLe^x.
- The size of the grid box is 24Å × 24Å × 24Å.

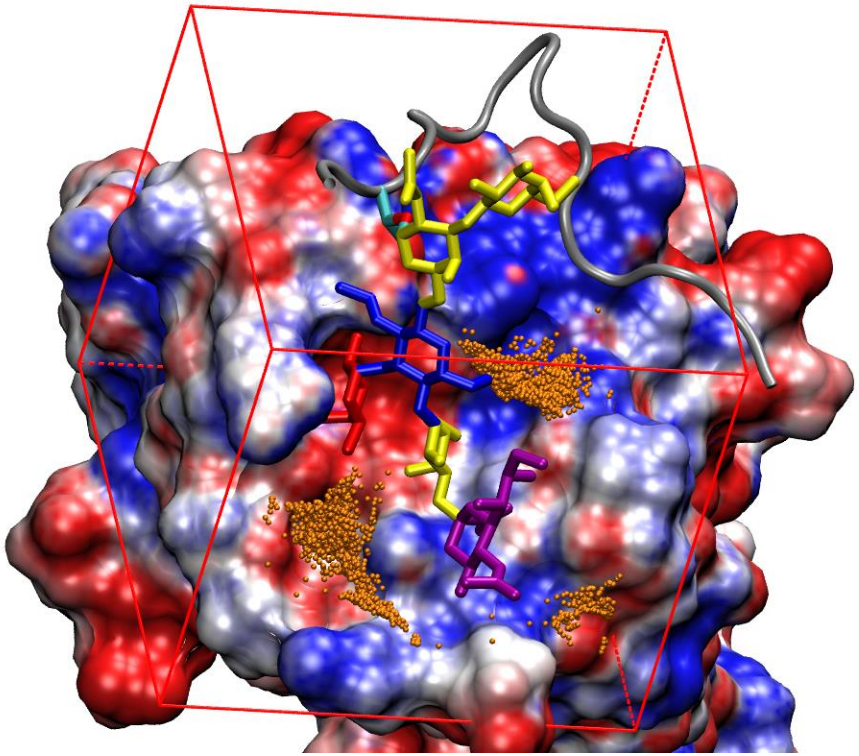
Virtual Screening Near Glycan

Autodock-vina energy score
-4.0 ~ -3.0 kcal/mol



Total number of 222,245

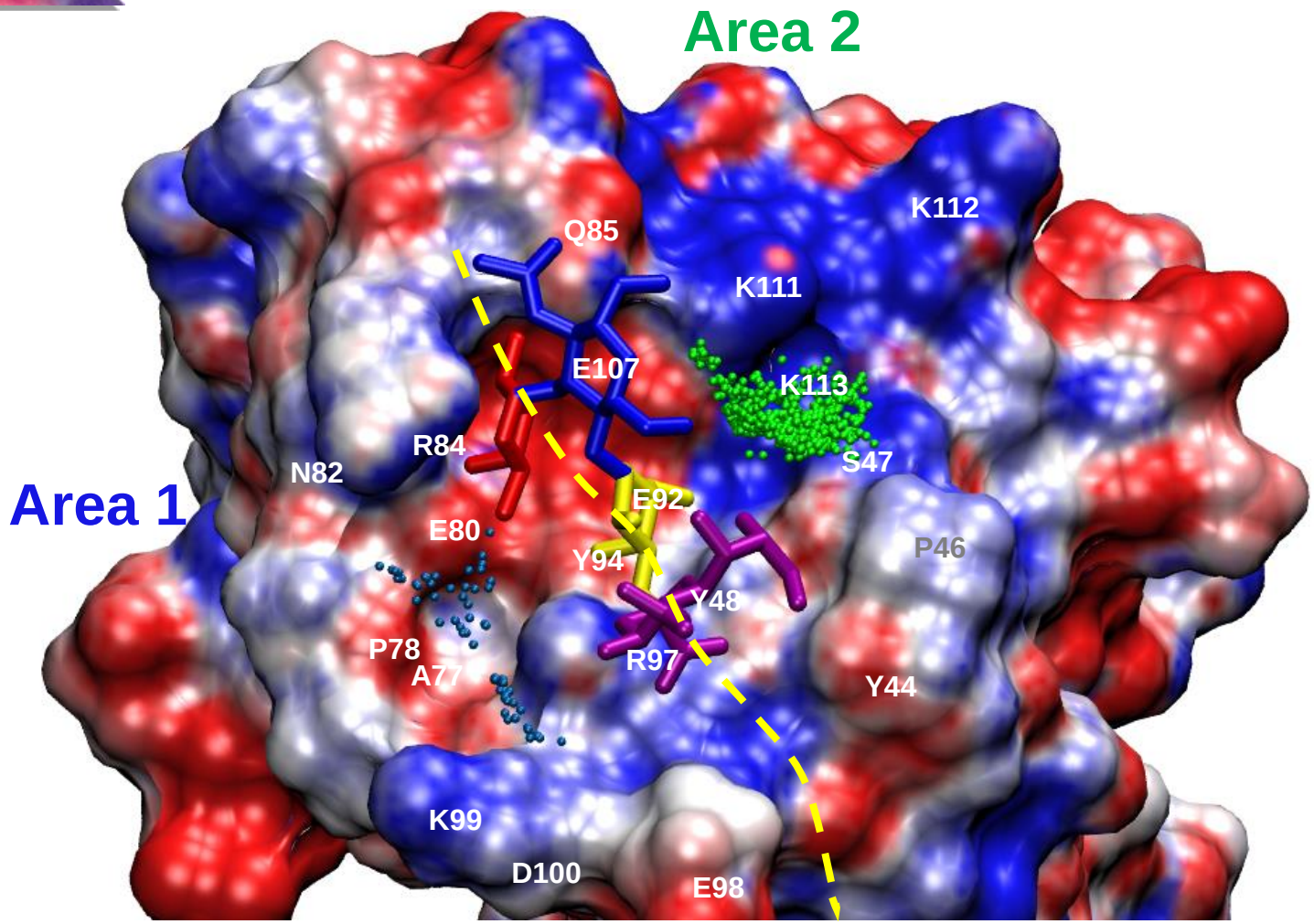
Autodock-vina energy score
-5.0 ~ -4.0 kcal/mol



Total number of 11,527

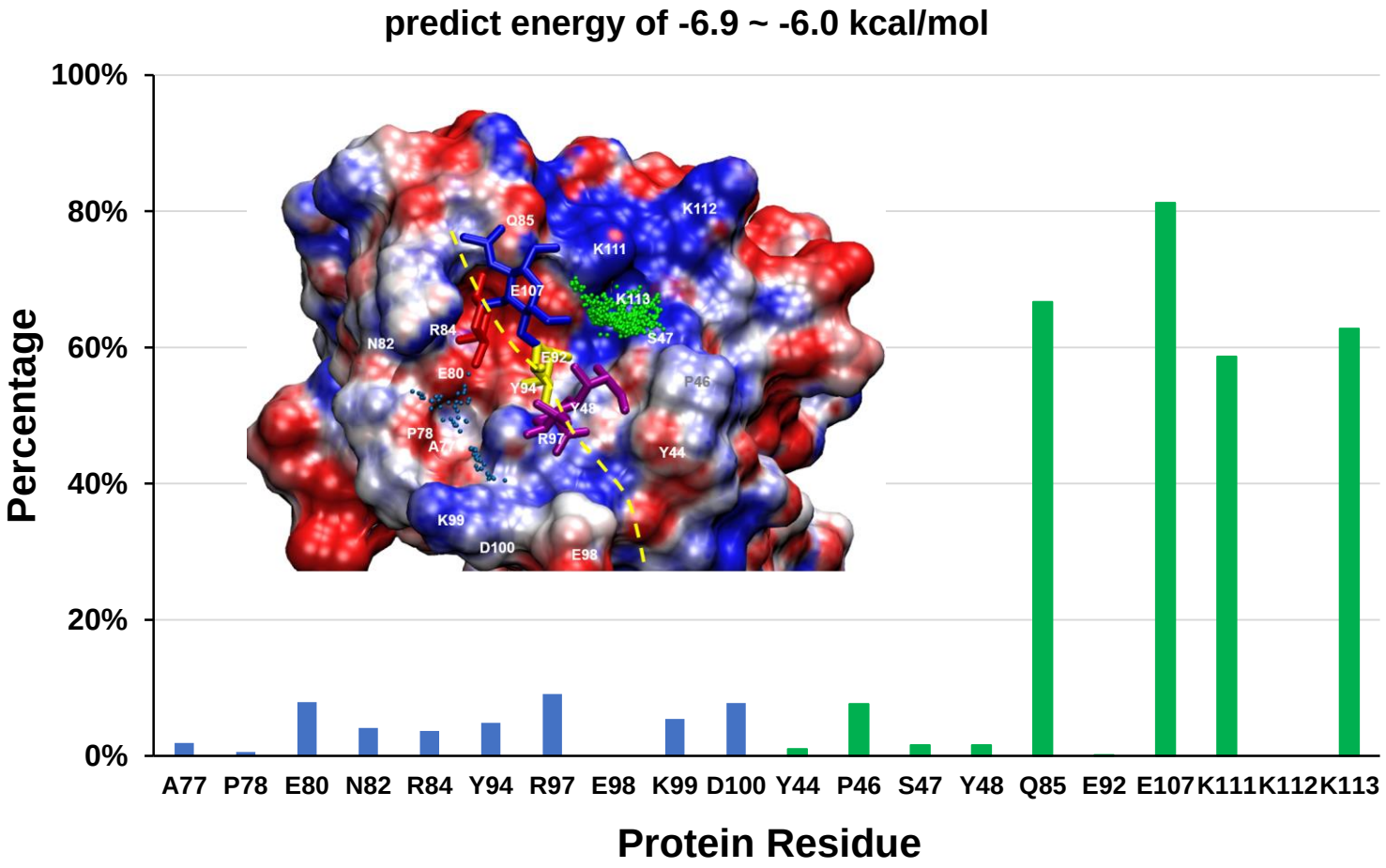
- Grid box centers at the geometric center of pyranose ring for GlcNAc in sLe^x.
- The size of the grid box is 24Å × 24Å × 24Å.

Virtual Screening Results



Area1										Area2										
77	78	80	82	84	94	97	98	99	100	44	46	47	48	85	92	107	111	112	113
A	P	E	N	R	Y	R	E	K	D	Y	P	S	Y	Q	E	E	K	K	K

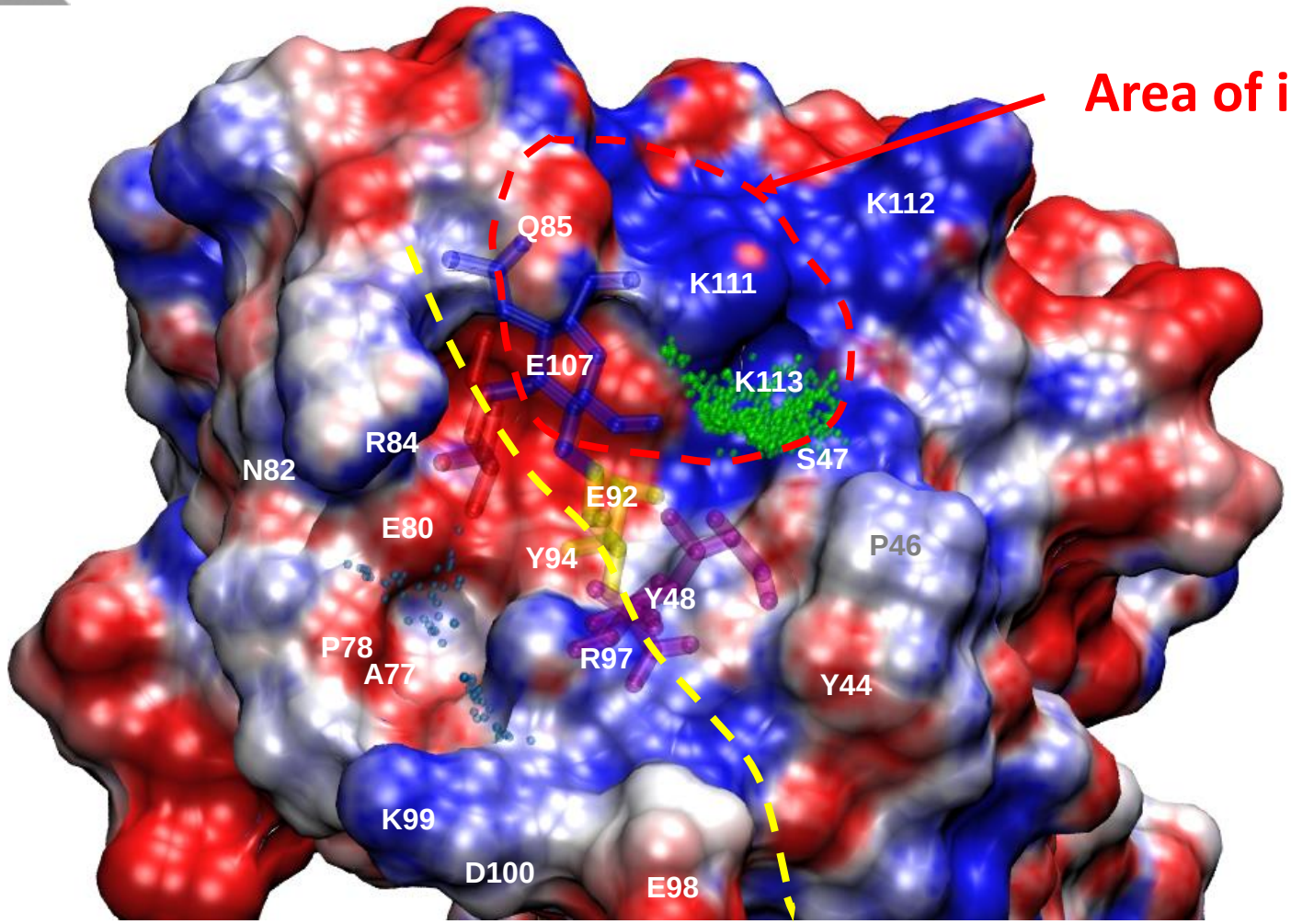
Virtual Screening Results



Q85, E107, K111, and K113 are in contact with 60% ~ 80% of the ligands.

This area can be considered as the area of interest.

Virtual Screening Results



Area1

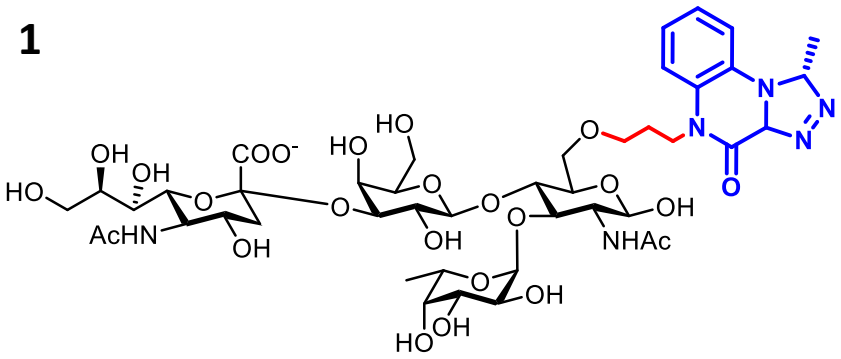
77	78	80	82	84	94	97	98	99	100
A	P	E	N	R	Y	R	E	K	D

Area2

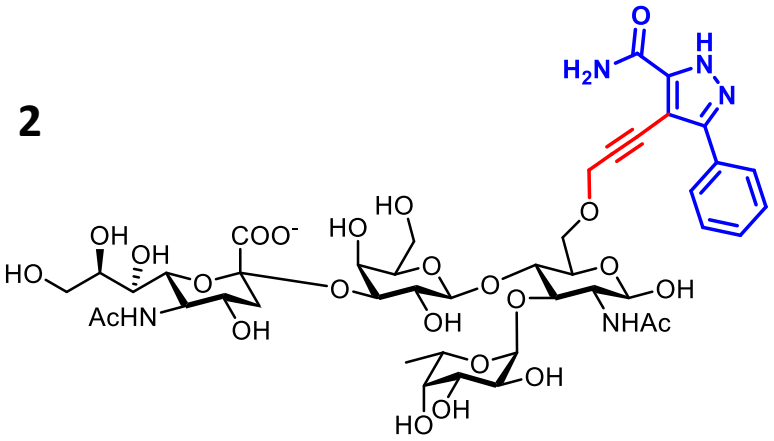
44	46	47	48	85	92	107	111	112	113
Y	P	S	Y	Q	E	E	K	K	K

Identified ZINC Molecules

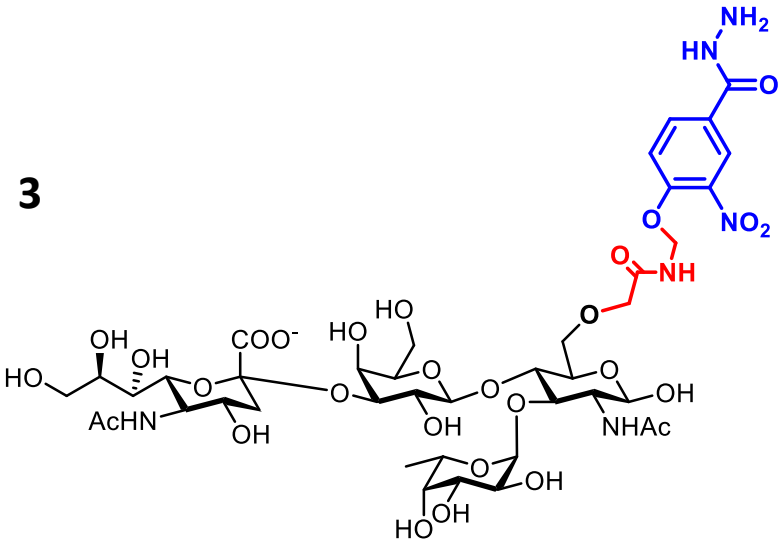
1



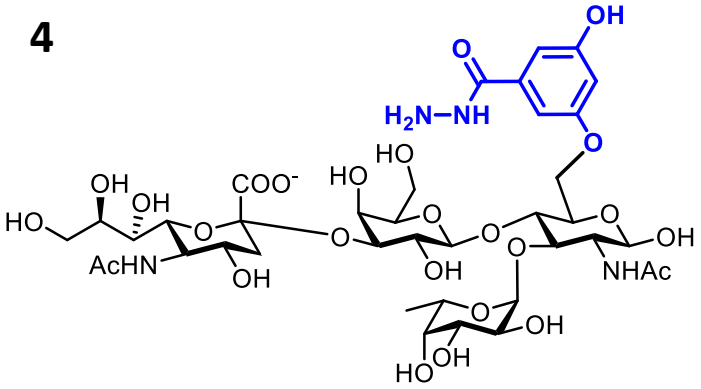
2



3



4

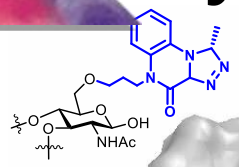


ZINC molecule

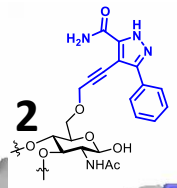
Added linkage moiety

Dynamics of E-selectin/sLe^x-ZINC Complexes

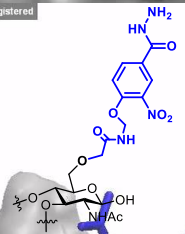
1



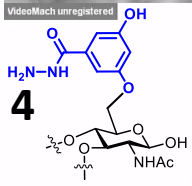
2



3



4



ZINC molecule in original docked position

Binding Interaction Energy for Ligand3

- From the MD simulations of E-selectin complexes, ZINC molecule in **ligand3** could stay stably in its original docked position.

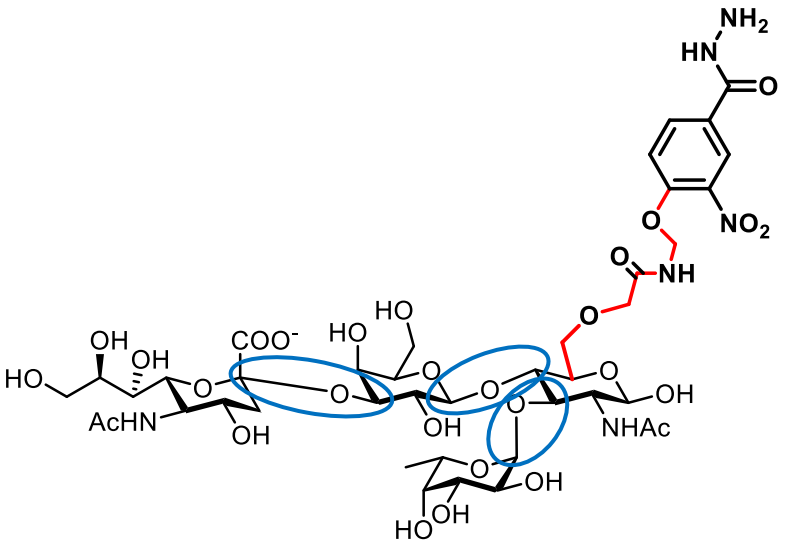
Theoretical binding energies for wild type and modified ligand (kcal/mol)

	E-selectin/sLe^x	E-selectin/sLe^x-ZINC
$\Delta G_{MM/GBSA}{}^{*}$	-18.8	-23.4
$-T\Delta S_{RTV(all)}$	19.5	19.6
$-T\Delta S_q{}^{C}$	**2.4**	**7.7**
$\Delta G_{Theoretical}$	3.1 ± 1.5	3.9 ± 1.1

*To incorporate with parameters for ZINC molecules, $\Delta G_{MM/GBSA}$ for E-selectin/sLex was recalculated with a different surface area algorithm.

- ZINC molecule in ligand3 interacts with K111, E107, P46, E92, and S47.
- ZINC molecule in ligand3 has strong H-bond interaction with E107.

Conformational Entropic Penalty in sLe^x-ZINC



Dihedral angles for entropic penalty in sLe^x.
Additional dihedral angles in sLe^x-ZINC.

- There are **6** dihedral angles for calculating conformational entropy for sLe^x.
- After attaching ZINC molecule to GlcNAc, **8** additional dihedral angles need to be considered. Therefore, the entropic penalty for sLe^x-ZINC is higher.