Ligand docking

CS/CME/BioE/Biophys/BMI 279
Oct 17, 2019
Possu Huang
Ron Dror
Problem definition

• A *ligand* is any molecule that binds to a protein
  – We’ll also use *ligand* to refer to any molecule that *might* bind to a protein (e.g., any candidate drug, DNA, ATP, or macromolecule, such as another protein). A ligand/receptor pair. (but generally associated with small molecules)

• *Ligand docking* addresses at least three problems:
  – Given a ligand known to bind a particular protein, what is its binding *pose* (that is, the location, orientation, and internal conformation of the bound ligand—basically, the position of each ligand atom when bound)
  – How *tightly* does a ligand bind a given protein?
  – Understanding how a ligand (like a drug) may affect the protein

A ligand will bind its target protein in the target protein’s binding-pocket (like a key in a lock). This pocket will generally have complementary shape and electrostatic environment suitable for interaction with the ligand. We can define properties of the ligand-protein bind such as the binding position or orientation of the ligand within the pocket, how tightly (how long) the ligand binds to the pocket, and how the interaction of the ligand and protein affects the structure and function of the protein itself.

Generally a prediction problem

What are the molecular driving forces for binding?

electrostatics + hydrogen bonding + hydrophobicity (entropy) and....

shape complementarity (Van der Waals forces)!

DOI: 10.1126/science.aaa5372
How do we measure how tightly a ligand binds to a protein?

- **Binding affinity** quantifies the binding strength of a ligand to a protein (or other target)
  - Conceptual definition: if we mix the protein and the ligand (with no other ligands around), what fraction of the time will the protein have a ligand bound?
  - This depends on ligand concentration, so we assume that the ligand is present at some standard concentration.
  - Binding affinity is usually expressed as either:
    - The difference $\Delta G$ in free energy of the bound state (all atomic arrangements where the protein is ligand-bound) and the unbound state (all atomic arrangements where the protein is not ligand-bound)
      - Again, assume standard concentration of ligand
      - From $\Delta G$, one can compute the fraction of time the ligand will be bound
    - A dissociation constant ($K_d$), which is (roughly) the ligand concentration at which half the protein molecules will have a ligand bound
Biochemist’s view of the process

\[ [A] + [B] \rightleftharpoons [AB] \]

the brackets [ ] means “the concentration of”

when it reaches chemical equilibrium,
the rate of association and dissociation is equal

rate of formation of AB from A and B = \( k[A][B] \)
rate of dissociation of AB into A and B = \( k_{-1}[AB] \)

At equilibrium \( k[A][B] = k_{-1}[AB] \)

When reactions reach a state of equilibrium, the rate of formation is the same as the rate of dissociation, and the concentration of A, B, and AB, do not change.

Two constants can be derived:

Ka, called association constant = \( \frac{[AB]}{[A][B]} = \frac{k}{k_{-1}} \) with unit M\(^{-1}\)

Kd, called dissociation constant = \( \frac{[A][B]}{[AB]} = \frac{k_{-1}}{k} \) with unit M

Conversion between free energy and equilibrium constant is:

\( \Delta G = -RT\ln(K) \)
Which one is harder to bind?

Amino acids are much harder to bind, as there are many degrees of freedom. In other words, there are many orientations an amino acid can adopt, but only one orientation may be correct for binding. Forcing the amino acid into a specific orientation becomes entropically unfavorable.

Steroids have many rings, which do not rotate - therefore they tend to be easier to bind, as there is only one conformation the molecule can adopt.
Which one is harder to bind?

HepB contains many stable surface markers that can act as binding targets

HIV markers change rapidly, making the binding problem significantly harder
Which one is harder to bind?

Amino acid is harder to bind, due to 1) one binding molecules as opposed to many binding molecules on HepB and 2) again, more orientations available (HepB markers are within a larger protein, fixing the degree of conformations)
Which one is harder to bind?

Having many binding sites increases the probability of the target being bound.

rcsb.org 2qij

http://www.bio.iitb.ac.in/~ashutosh/research.html
Direct approach to computing binding affinity

- Run a really long molecular dynamics (MD) simulation in which a ligand binds to and unbinds from a protein many times.
- Directly observe the fraction of time the ligand is bound.

Drawbacks - remember that simulations take a long time to run. You likely won't be able to simulate the binding interaction for a long time, so resolving rate constants (k,k-1) is not feasible from these interactions. That being said, if the actual binding interaction occurs within a small window, you can resolve important protein-ligand interactions that can inform better design for future ligands or binding proteins.
Problem Statement: imagine you are a company with many candidate molecules (drugs) of interest. How will you screen your candidates?

Building the docking algorithm

- scoring function (analogous to an energy function)
- computational sampling of interface
- computational sampling of protein flexibility
- experimental data (ex. knowing prior drugs that we know bind to our protein of interest)

Scoring Function (heuristics are used here to simplify the problem, ex. using physical shape to model interaction)

- molecular mechanics force field
- knowledge-based force field
- empirical scoring function

Computational Sampling (interface)

- stochastic search for conformations at the interface
- human-based sampling decisions
- shaped-based matching (vector matching)
- MD simulate it (brute force, also computationally expensive)
- (expensive) grid-search, search through all possible conformations

In class exercise

How to build a docking algorithm?

(Notes will be provided online later)

Computational Sampling (protein flexibility/conformation)

- rigid body
- side-chain rotation
- increase tolerance between two protein structures, allow for greater clashing (ex. soften Van Der Waals)

Experimental Data

- binding affinity/rates derived experimentally (from current protein and analogues)
- observe effect of mutations at proposed target sites
- run through a proteolytic digest and observe response (for some interactions, the protein-protein interaction will stabilize the interface, adding digest enzymes will cut around the interface)
- co-evolution at the protein-protein interface
How well does predictive docking work?

• The best docking protocols currently:
  – Predict a reasonably accurate pose (for ligands that do in fact bind the target protein) about half the time for rigid targets (the “easy” cases)
  • In these cases, one of the highly-ranked poses is usually close to the correct one
  – Provide useful, but far from perfect results, when ranking ligands
  • Tend to work best when comparing closely related ligands
  – Are not particularly useful when it comes to quantitatively estimating binding free energies

Can we trust (small molecule) docking results?

Best algorithm only hits target ~ 50% of the time

it depends... and maybe not
What if the goals of docking is to design?

- with small molecules
  - improve the functionality of a drug
  - create a sensor
  - make an enzyme
  - change the specificity of an enzyme

- with proteins
  - make an inhibitor (e.g. antibody drugs; protein based drugs are also called biologics)
  - make an activator (e.g. molecule that binds GPCR)
  - create a sensor
  - create materials (ala LEGO)
Enzymes are good (at what they do)

Enzymes speed up chemical reactions by reducing the activation energy required for the reaction to start (i.e., the input energy required to kick off a reaction).

slide from Zev Bryant
Enzymes lower the activation barrier of reactions
transition state stabilization hypothesis

Enzymes can reduce the energy barrier by providing a local environment more favorable for the reaction. In other words, they make the energy of the transition state lower (notice the difference in energy between ST and EST).

There is a transition state, that molecule S adopts (in the process of reacting, here labeled by ST and EST). Creating proteins that target transition state analogs of the catalyzed transition state (EST) can result in enzyme-like proteins that catalyze binding/reaction.
Catalytic antibodies follow this view and are generated using transition state analogs.

Schultz and Lerner, 1980s
Current enzyme design workflow also borrow from the idea of binding and stabilizing transition state analogs.

Example of a computational theozyme model:

1. Find a scaffold in PDB that can support the geometry of the blue side chains.
2. Redesign the positions in yellow.
3. Experimentally improved versions have mutations found in green and blue positions.

Reference: [https://www.nature.com/articles/nchem.2596](https://www.nature.com/articles/nchem.2596)
Currently the best enzyme derived from computational design has $10^9$ rate enhancement after experimental optimization.

Catalytic efficiency, indicating a quality of an enzyme’s efficiency to convert substrate to product and the efficiency to bind substrate productively.

<table>
<thead>
<tr>
<th>Enzyme*</th>
<th>$k_{\text{cat}}$ (s$^{-1}$)</th>
<th>$K_m$ (μM)</th>
<th>$k_{\text{cat}}/K_m$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_{\text{cat}}/k_{\text{uncat}}$</th>
<th>$(k_{\text{cat}}/K_m)/K_{\text{Lys}}$</th>
<th>$S_\dagger$</th>
<th>$pK_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA95.5-8F</td>
<td>10.8 ±0.6</td>
<td>320±36</td>
<td>34,000</td>
<td>1.7×$10^9$</td>
<td>1.5×$10^{10}$</td>
<td>480</td>
<td>6.2</td>
</tr>
<tr>
<td>RA95.5-8†</td>
<td>0.36</td>
<td>230</td>
<td>1,600</td>
<td>5.5×$10^7$</td>
<td>7.0×$10^8$</td>
<td>14</td>
<td>7.6</td>
</tr>
<tr>
<td>RA95.5‡</td>
<td>0.0043</td>
<td>270</td>
<td>16</td>
<td>6.6×$10^5$</td>
<td>6.9×$10^6$</td>
<td>3</td>
<td>7.6</td>
</tr>
<tr>
<td>RA95.0‡</td>
<td>0.00005</td>
<td>300</td>
<td>0.17</td>
<td>4.8×$10^3$</td>
<td>2.3×$10^4$</td>
<td>0.4</td>
<td>8.1</td>
</tr>
</tbody>
</table>

You are not responsible for the values nor the kinetic parameters.

rate enhancement over uncatalyzed reaction
Lastly, new trends in thinking about protein structures

Tertiary Motifs

The majority of structural elements interact with each other following some kind of “motif”

PNAS, vol. 113 no. 47
quaternary assembly (motif based docking)

Other representative methods are FFT, PatchDock, RosettaDock

Nature volume 535, pages 136–139 (07 July 2016) doi:10.1038/nature18010
Suddenly, protein-protein interface design seems to be a “solved” problem…

But only if you can design on both sides and have lots of stable parts available

Huang, Boyken and Baker, Nature 2016