Ligand docking and virtual screening

CS/BioE/CME/Biophys/BMI 279
Nov. 14 and 16, 2023
Ron Dror
Outline

• Goals of ligand docking
• Defining binding *affinity* (strength)
• Computing binding affinity: Simplifying the problem
• Standard ligand docking methodology
• Virtual screening
• Alternative methods and current research directions
Goals of ligand docking
A drug binding to its target
(The great majority of drug targets are proteins)

Beta-blocker alprenolol binding to an adrenaline receptor

Dror et al., *PNAS* 2011
Problem definition

• A ligand is any molecule that binds to a target macromolecule (e.g., a protein or RNA drug target)
  – We’ll also use ligand to refer to any molecule (e.g., any candidate drug) that might bind to a given macromolecule

• Ligand docking addresses two 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 (or other macromolecule)?

Why is docking useful in drug discovery?

- **Virtual screening**: Identifying drug candidates by considering large numbers of possible ligands
- **Ligand optimization**: Modifying a drug candidate to improve its properties
  - Docking can predict the candidate molecule’s binding pose, which helps envision how modifying that molecule would change its binding strength and/or alter its effect on the target protein
  - Docking can predict binding strengths of related candidate molecules
Ligand docking: a graphical summary

- Predicts...
  - The **pose** of the molecule in the binding site
  - The binding affinity or a **score** representing the strength of binding
Defining binding *affinity* (strength)
How do we specify 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 can be expressed in two ways

- A dissociation constant ($K_D$), which is (roughly) the ligand concentration at which half the protein molecules will have a ligand bound
  - For example, a “1 nanomolar (1 nM) binder” is a ligand that will occupy the binding site half the time at a concentration of 1 nM (i.e., $10^{-9}$ moles per liter)
  - This is the most common way to express affinity
- The difference $\Delta G$ in free energy of the bound state (all atomic arrangements where the protein has a ligand bound) and the unbound state (all other atomic arrangements)
  - Typical units are kcal/mol or kJ/mol
  - Again, assume standard concentration of ligand
  - From $\Delta G$, one can compute the fraction of time the ligand will be bound
Binding affinity: Clarifications

• Binding affinity is different from “how long the ligand remains bound” (the off-rate) or “how quickly the ligand binds” (the on-rate)
  – Binding affinity is a ratio of the on-rate and off-rate; you can’t calculate it from either one alone
  – These rates are also of interest in drug discovery, and predicting them is a different (and even more challenging) computational problem

• Binding affinity is different from “how strong are the inter-atomic forces between the ligand and the target when the ligand is bound”
  – Binding affinity also depends a great deal on what happens when the ligand isn’t bound (e.g., how favorable are the interactions of the ligand and the binding pocket with water)
Computing binding affinity: Simplifying the problem
A hypothetical direct approach to computing binding affinity

- Run a really long molecular dynamics (MD) simulation in which a ligand binds to \textit{and} unbinds from a protein many times.
- Directly observe the fraction of time the ligand is bound.
This direct approach rarely works

• It is so computationally intensive that we usually cannot do it for even a single ligand, let alone millions
  – The toughest part is the unbinding (dissociation)
    • Drug molecules usually take seconds to hours to unbind from their targets.
    • Microsecond-timescale molecular dynamics simulations usually take days.
  – We’d have to simulate many cycles of binding and unbinding.

• It is also limited by force field accuracy
  – Most molecular mechanics force fields are less accurate for small-molecule ligands than for proteins
Question to discuss

• How would you compute a binding affinity?
  – Suppose you’re given the structure of a target protein, and you want to compute the affinity of a particular ligand to that protein
  – To simplify the problem a bit, you may also assume that you’re given the binding pose
Standard ligand docking
(most common method to predict ligand binding affinity)

• Ligand docking is a fast, heuristic approach with two key components
  – A *scoring function* that very roughly approximates the binding affinity of a ligand to a protein given a binding pose
  – A *search method* that searches for the best-scoring binding pose for a given ligand
Standard ligand docking
(most common method to predict ligand binding affinity)

• To predict the binding affinity of a ligand:
  – Docking software searches through poses of the ligand to find the pose with the best score
  – That pose is the predicted pose of the ligand, and its score is the predicted affinity
• Here affinity is expressed as a binding energy: the lower the score, the more tightly the ligand binds
Standard ligand docking
(most common method to predict ligand binding affinity)

1. Sample potential candidate binding poses

2. Score candidate poses

-2 kcal/mol

-10 kcal/mol

-3 kcal/mol

-8 kcal/mol

Predicted binding energy (minimum across candidate poses)
Predicted binding pose

3. Select the best-scoring pose

Note that for docking to run reasonably quickly, one needs a good search strategy for the sampling step. One might iterate between generating candidate poses and scoring them.

Figure credit:
Ayush Pandit and Joe Paggi
Ligand docking is approximate!

- For example, most ligand docking methods assume that the target protein is rigid and don’t explicitly consider water molecules.
- In reality, protein mobility, ligand mobility, and water molecules all play a major role in determining binding affinity.
  - Docking is approximate but useful.
  - The term *scoring function* is used instead of *energy function* to emphasize the highly approximate nature of the scoring function.
**Docking software (a partial list)**

<table>
<thead>
<tr>
<th>Program</th>
<th>Country of Origin</th>
<th>Year Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>AADS</td>
<td>India</td>
<td>2011</td>
</tr>
<tr>
<td>ADAM</td>
<td>Japan</td>
<td>1994</td>
</tr>
<tr>
<td>AutoDock</td>
<td>USA</td>
<td>1990</td>
</tr>
<tr>
<td>AutoDock Vina</td>
<td>USA</td>
<td>2010</td>
</tr>
<tr>
<td>BetaDock</td>
<td>South Korea</td>
<td>2011</td>
</tr>
<tr>
<td>DARWIN</td>
<td>USA</td>
<td>2000</td>
</tr>
<tr>
<td>DIVALI</td>
<td>USA</td>
<td>1995</td>
</tr>
<tr>
<td>DOCK</td>
<td>USA</td>
<td>1988</td>
</tr>
<tr>
<td>DockVision</td>
<td>Canada</td>
<td>1992</td>
</tr>
<tr>
<td>EADock</td>
<td>Switzerland</td>
<td>2007</td>
</tr>
<tr>
<td>eHiTS</td>
<td>UK</td>
<td>2008</td>
</tr>
<tr>
<td>EUDOC</td>
<td>USA</td>
<td>2001</td>
</tr>
<tr>
<td>FDS</td>
<td>UK</td>
<td>2003</td>
</tr>
<tr>
<td>FlexE</td>
<td>Germany</td>
<td>2001</td>
</tr>
<tr>
<td>FlexX</td>
<td>Germany</td>
<td>1996</td>
</tr>
<tr>
<td>FLIPDock</td>
<td>USA</td>
<td>2007</td>
</tr>
<tr>
<td>FLOG</td>
<td>USA</td>
<td>1994</td>
</tr>
<tr>
<td>FRED</td>
<td>UK</td>
<td>2003</td>
</tr>
<tr>
<td>FTDOCK</td>
<td>UK</td>
<td>1997</td>
</tr>
<tr>
<td>GEMDOCK</td>
<td>Taiwan</td>
<td>2004</td>
</tr>
<tr>
<td>Glide</td>
<td>USA</td>
<td>2004</td>
</tr>
<tr>
<td>GOLD</td>
<td>UK</td>
<td>1995</td>
</tr>
<tr>
<td>Hammerhead</td>
<td>USA</td>
<td>1996</td>
</tr>
<tr>
<td>ICM-Dock</td>
<td>USA</td>
<td>1997</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lead finder</th>
<th>Country of Origin</th>
<th>Year Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigandFit</td>
<td>USA</td>
<td>2003</td>
</tr>
<tr>
<td>LigDockCSA</td>
<td>South Korea</td>
<td>2011</td>
</tr>
<tr>
<td>LIGIN</td>
<td>Germany</td>
<td>1996</td>
</tr>
<tr>
<td>LUDI</td>
<td>Germany</td>
<td>1992</td>
</tr>
<tr>
<td>MADAMM</td>
<td>Portugal</td>
<td>2009</td>
</tr>
<tr>
<td>MCDOCK</td>
<td>USA</td>
<td>1999</td>
</tr>
<tr>
<td>MDock</td>
<td>USA</td>
<td>2007</td>
</tr>
<tr>
<td>MolDock</td>
<td>Denmark</td>
<td>2006</td>
</tr>
<tr>
<td>MS-DOCK</td>
<td>France</td>
<td>2008</td>
</tr>
<tr>
<td>ParDOCK</td>
<td>India</td>
<td>2007</td>
</tr>
<tr>
<td>PhDock</td>
<td>USA</td>
<td>2003</td>
</tr>
<tr>
<td>PLANTS</td>
<td>Germany</td>
<td>2006</td>
</tr>
<tr>
<td>PRO_LEADS</td>
<td>UK</td>
<td>1998</td>
</tr>
<tr>
<td>PRODOCK</td>
<td>USA</td>
<td>1999</td>
</tr>
<tr>
<td>ProPose</td>
<td>Germany</td>
<td>2004</td>
</tr>
<tr>
<td>PSI-DOCK</td>
<td>China</td>
<td>2006</td>
</tr>
<tr>
<td>PSO@AUTO DOCK</td>
<td>Germany</td>
<td>2007</td>
</tr>
<tr>
<td>PythDock</td>
<td>South Korea</td>
<td>2011</td>
</tr>
<tr>
<td>Q-Dock</td>
<td>USA</td>
<td>2008</td>
</tr>
<tr>
<td>QXP</td>
<td>USA</td>
<td>1997</td>
</tr>
<tr>
<td>rDock</td>
<td>UK</td>
<td>2013</td>
</tr>
<tr>
<td>SANDOCK</td>
<td>UK</td>
<td>1998</td>
</tr>
<tr>
<td>SFDOCK</td>
<td>China</td>
<td>1999</td>
</tr>
<tr>
<td>SODOCK</td>
<td>Taiwan</td>
<td>2007</td>
</tr>
<tr>
<td>SOFTDocking</td>
<td>USA</td>
<td>1991</td>
</tr>
<tr>
<td>Surflex</td>
<td>USA</td>
<td>2003</td>
</tr>
<tr>
<td>SYSDOC</td>
<td>USA</td>
<td>1994</td>
</tr>
<tr>
<td>VoteDock</td>
<td>Poland</td>
<td>2011</td>
</tr>
<tr>
<td>YUCCA</td>
<td>USA</td>
<td>2005</td>
</tr>
</tbody>
</table>

**Most popular (based on citations 2001–2011):**

- AutoDock
- GOLD
- DOCK
- FlexX
- Glide
- Glide
- QXP

---

Sousa et al., *Current Medicinal Chemistry* 2013

Optional material

Standard ligand docking methodology
Scoring functions

• Scoring functions used for docking typically capture chemists’ intuition about what makes a ligand–target interaction energetically favorable. For example:
  – Hydrogen bonding
  – Hydrophobic interactions
• Parameters are fit based on known ligand–target structures and affinities
• These scoring functions are (very rough) attempts to approximate the binding free energy
  – By contrast, molecular mechanics force fields give potential energy associated with a particular arrangement of atoms
Example: Glide scoring function

• Glide (widely used commercial docking software) uses the following “GlideScore” function:

\[ \Delta G_{\text{bind}} = C_{\text{lipo-lipo}} \sum f(r_{lr}) + \]
\[ C_{\text{hbond-neut-neut}} \sum g(\Delta r) h(\Delta \alpha) + \]
\[ C_{\text{hbond-neut-charged}} \sum g(\Delta r) h(\Delta \alpha) + \]
\[ C_{\text{hbond-charged-charged}} \sum g(\Delta r) h(\Delta \alpha) + \]
\[ C_{\text{max-metal-ion}} \sum f(r_{lm}) + C_{\text{rotb}} H_{\text{rotb}} + \]
\[ C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + \]
\[ C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms} \]

- The first term rewards contacts between hydrophobic atoms of the ligand and protein, and is a function of the distance between them.
- The next three terms reward specific kinds of hydrogen bonds, and are functions of both distance and angle for each hydrogen bond.

• Glide uses many additional terms as well.

Search methods

• Docking software searches for the best-scoring pose for each ligand
• The search space is huge, because one needs to consider all combinations of ligand position, ligand orientation, and ligand conformation (shape)
• To search this space efficiently, docking software typically employs either or both:
  – Hierarchical methods in which one uses approximate measures to identify promising groups of poses, then evaluates subgroups in more detail
  – Monte Carlo methods
Example: Glide search

• Glide uses a hierarchical search method
• It first identifies a set of “reasonable” conformations for each ligand, by varying internal torsion angles
• For each ligand, it scans possible positions and orientations, using a rough measure of fit to binding pocket
• The most promising approximate poses undergo further “refinement”
• Candidate poses are ranked by the scoring function

Friesner et al., *J Med Chem* 47:1739, 2004
Virtual screening
Virtual screening: the basics

• Goal: identify ligands that bind to a target—particularly ligands that are very different from any known binder

• Typical process
  – Select a virtual library of chemical compounds
  – Use docking to estimate the affinity of each
  – Buy or make the compounds with the best predicted affinities and do experiments to test how well they bind
  – Optional: Optimization of experimentally validated binders by testing related chemical compounds
Virtual screening: the basics

A

Acquire and prepare the target structure

B

Dock each compound to the target

Select a compound library

C

1) \(-10.0 \text{ kcal/mol}\)
2) \(-8.0 \text{ kcal/mol}\)
3) \(-6.0 \text{ kcal/mol}\)
4) \(-5.0 \text{ kcal/mol}\)
5) \(-2.0 \text{ kcal/mol}\)

Rank compounds based on predicted binding energies of docked structures

D

Cluster compounds and select a diverse subset

E

Optionally, visually filter

Experimentally test selected compounds

Figure credit:
Ayush Pandit and Joe Paggi
New: “Ultra-large” virtual libraries

• In virtual screening, one typically uses libraries of compounds that can be easily ordered from vendors, so that one can easily test the top-ranked ones

• A few years ago, a few million compounds were available from vendors

• Now it’s billions or trillions
  – Thanks to the advent of the make-on-demand approach (pioneered by Enamine in Ukraine)
  – Idea: gigantic library of compounds that have not yet been made but that vendor can make quickly and cheaply with high probability

• This has increased the utility of virtual screening
  – A few million compounds can be tested experimentally by “high-throughput screening” robots, but this doesn’t scale to billions and requires that all the compounds be synthesized in advance
Alternatives methods and current research directions

Optional material
MD-based approaches

MD-based approaches

• It turns out that one can compute binding affinities by molecular dynamics simulation without waiting for ligands to spontaneously dissociate (unbind) and bind

• Instead, in “alchemical” methods such as free energy perturbation (FEP), one performs a series of simulations in which the ligand gradually “dematerializes” from its bound position and “materializes” in an unbound position. This works because binding affinity does not depend on the binding pathway.

  – These methods currently represent the most accurate way to predict binding affinities, at least for comparing binding energies of chemically similar ligands, which is how they’re typically used

    • One can determine a difference in binding affinity between two similar ligands by “mutating” one ligand into the other in simulation.

  – These methods assume a known binding pose for each ligand

  – These methods are still very expensive computationally and thus cannot be used on large numbers of ligands
The absolute approach (Figure 5A) determines the change in free energy on moving a ligand from solution to the bound state ($\Delta G_{\text{bind}(a)}$) and requires simulations that progressively decouple the bound ligand from the binding site and the free ligand from the solvent. The decoupled bound ligand and the decoupled free ligand are equivalent, so the free energy of transfer between these two states is 0. To avoid delinquent wandering of the ligand as all interactions to the protein are turned off, it may be necessary to constrain the position of the decoupled ligand.

Alternatively, relative alchemical perturbation methods (Figure 5B) calculate the free energy difference between the binding of a pair of ligands, $\Delta \Delta G_{\text{bind}(a \rightarrow b)}$, by performing sets of simulations that transform one ligand into the other in the binding site and also in solution.

**Expected Accuracy of Alchemical Perturbation Methods**

For a binding affinity value measured using a high quality experimental method, such as isothermal titration calorimetry (ITC), the noise is said to be within 0.3−0.5 kcal/mol, which suggests the RMSD from experiment for an absolute alchemical perturbation calculation cannot exceed this level of accuracy. Therefore, even alchemical calculations that are run for an extremely long time using supremely high-quality force are expected to deviate by at least 0.3 kcal/mol RMSD from experiment.

Absolute free energy calculations often give binding free energies ($\Delta G$) within 2−3 kcal/mol RMSD of experimental determinations for uncharged ligands and up to 4 kcal/mol RMSD for flexible or charged ligands. Comparatively, relative free energy calculations, which yield $\Delta \Delta G$ values, are typically more accurate because the initial and final states are generally similar (i.e., the perturbation might be performed between two molecules having a common core and differing in only a single functional group). Similarity between initial and final states reduces sampling errors and provides greater accuracy.

When translating errors in free energy calculations to a corresponding effect on the dissociation equilibrium constant ($K_D$), one could question whether an RMSD of 2−3 kcal/mol from the experimental binding free energy is useful, given this level of error corresponds to quite large deviations in predicted $K_D$. Indeed with errors of this magnitude, it is likely that the rank order of a series of ligands may not be reliably predicted using alchemical methods.

There are a number of sources of error in perturbation methods including, but not limited to, the quality of the force field parameters used, the treatment of electrostatic interactions, incomplete sampling, and methodological simplifications. Often, the largest sources of error associated with alchemical free energy methods come down to inadequacies in the force field and suboptimal sampling of the energy landscape between states A and B. In the case of charged ligands, it has been shown that the free energy of binding is influenced by a variety of factors, including the size and shape of the periodic cell, the treatment of electrostatic interactions, and the presence of counterions.

“Ligand-based” approaches

• If one has experimentally measured affinity values for many ligands at a particular target, one can ignore the target structure entirely and simply make affinity predictions based on similarity of query ligand to previously characterized ligands

• These approaches, which date back many decades, are a type of machine learning

• They generally work well only when one has experimentally characterized ligands similar to the query ligand
Current research area:
Machine learning approaches for virtual screening

• Both academic research groups and companies are working on deep learning approaches to develop more accurate scoring functions

• The idea is to fit general functional forms (as described by large neural networks), rather than assuming specific functional forms based on approximations to physics

• A variant of this approach is to do reasonably accurate, time consuming calculations for a subset of the compounds in the library, and then use the results to predict binding affinities of other compounds with faster ligand-based methods
Another machine learning approach: experimental information on unrelated ligands can substantially improve docking predictions.
Another machine learning approach: experimental information on unrelated ligands can substantially improve docking predictions.

Prediction informed by the fact that the compounds below bind this target (in unknown poses)

Beta agonist dobutamine bound to β1-adrenergic receptor

Paggi, …, Dror, PNAS 2021
Compounds that bind to the same target often form similar interactions with the binding pocket

- We thus predict their poses simultaneously
  - Without requiring any similarity between ligands
  - Without requiring shared interactions
- We learn the likelihood of a given set of ligand poses (one pose per ligand)
A similar approach (ComBindVS) improves virtual screening

Average performance across 102 targets in DUD-E benchmark set

Note: All ligands screened are very different from known binders.

Paggi, ..., Dror, PNAS 2021
Current research area: Generative models

• Instead of learning a scoring function, one can learn to directly generate ligand binding poses or even ligands themselves

• Ligand binding poses:
  – Given a 3D structure of a protein and the 2D structure of a ligand, generate ligand coordinates (e.g., using a diffusion model)
  – Given only the sequence of a protein the 2D structure of a ligand, generate coordinates for both. This is essentially a generalization of RoseTTAFold or AlphaFold 2 to include ligands

• Ligands
  – Given a 3D structure of a protein binding pocket, generate ligands that bind tightly to that pocket