Small-molecule docking and virtual screening

CS 279 / BIOE 279 / BIOPHYS 279 / BMDS 245 / CME 279

Nov. 10 and 12, 2026

Ron Dror

- No class on Tues., Nov. 17
- Project kickstart: Monday, Nov.16, 11 AM - 12 PM (via Zoom and recorded)
  - Recommended for those who wish to write/modify code but need help getting started
  - Reminder: Coding is not required for the project. If you prefer, you can do a project using existing software.

See “Project Guide” on website

- Guidance on selecting a project topic, as well as suggestions for software, datasets, and other resources
- Reminder: very short project proposal due Monday, Nov. 16

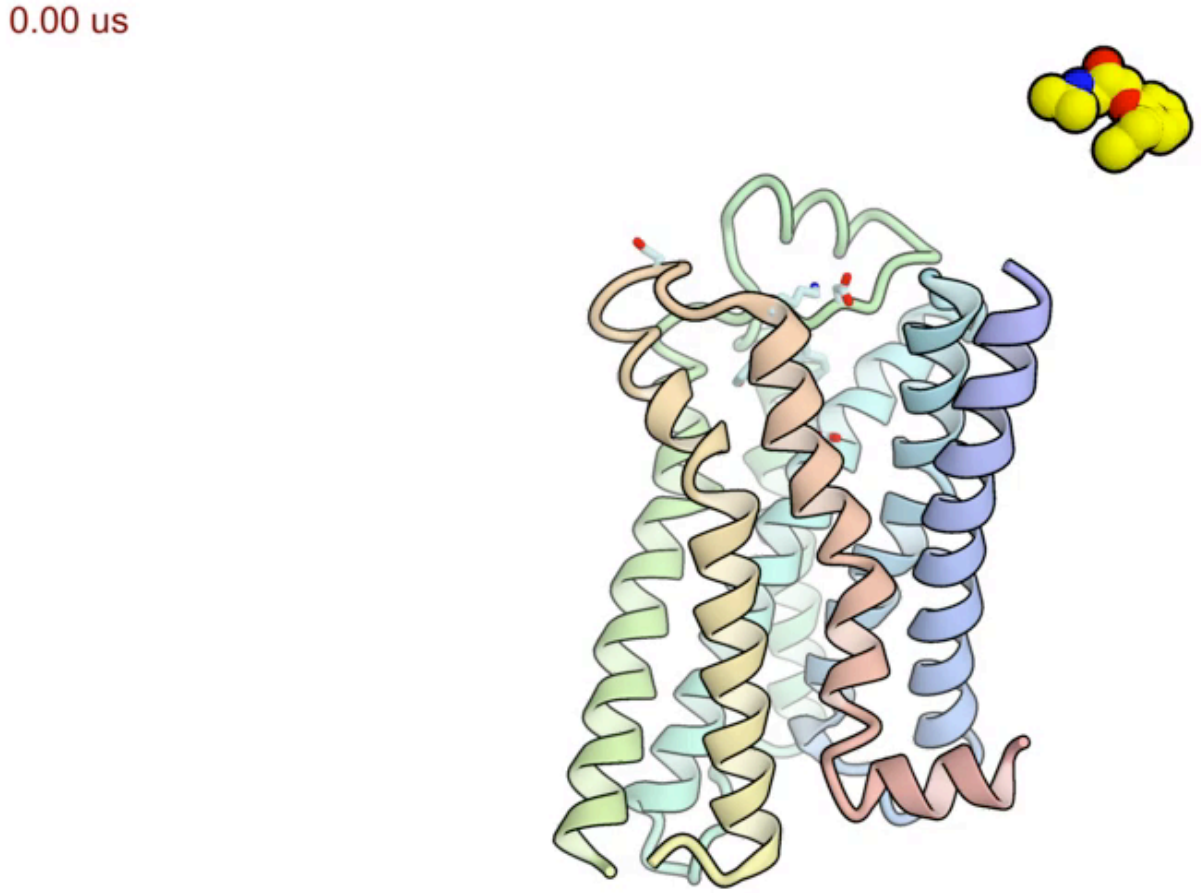
Outline

- Goals of small-molecule docking
- Defining binding *affinity* (strength)
- Computing binding affinity: Simplifying the problem
- Standard small-molecule docking methodology
- Virtual screening
- Alternative methods and current research directions

Goals of small-molecule docking (also known as ligand docking or molecular docking)

A drug binding to its target

(Most drug targets are proteins)

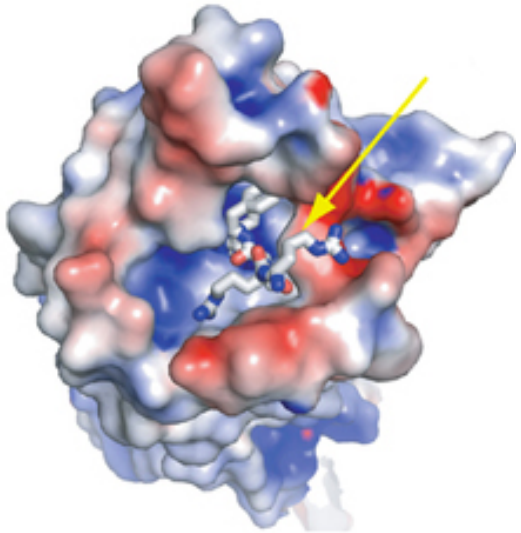


Beta-blocker alprenolol binding to an adrenaline receptor

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
- *Small-molecule docking* addresses two problems:
  - Given a small-molecule 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 3D coordinates of each ligand atom when bound)
  - How *tightly* does a given small molecule bind a given protein (or other macromolecule)?

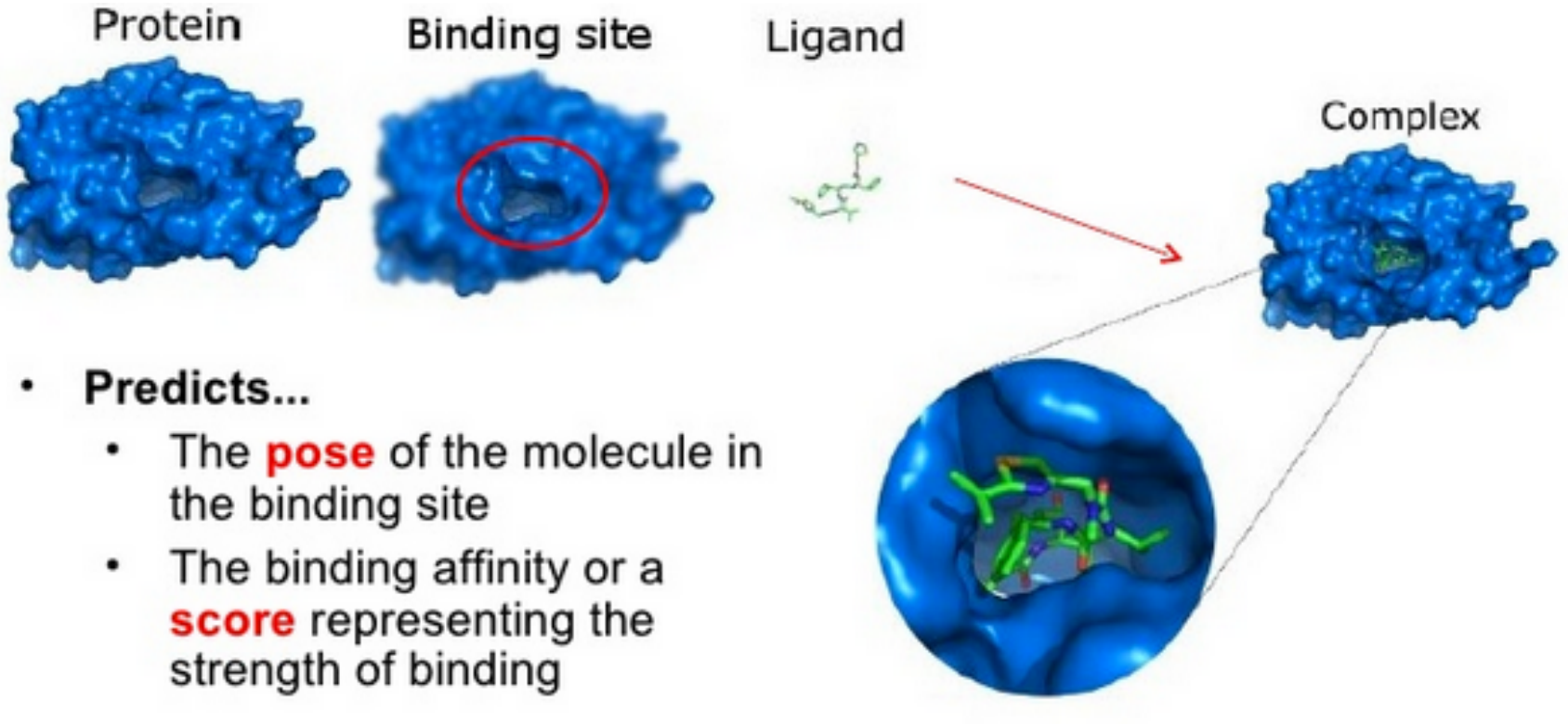


http://www.nih.gov/researchmatters/october2012/images/structure_l.jpg

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

http://www.slideshare.net/baoilleach/proteinligand-docking-13581869

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 of ligand per liter of solute, typically water)
  - This is the most common way to express affinity
- The difference ΔG 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 ΔG, one can compute the fraction of time the ligand will be bound

Binding affinity: Clarifications

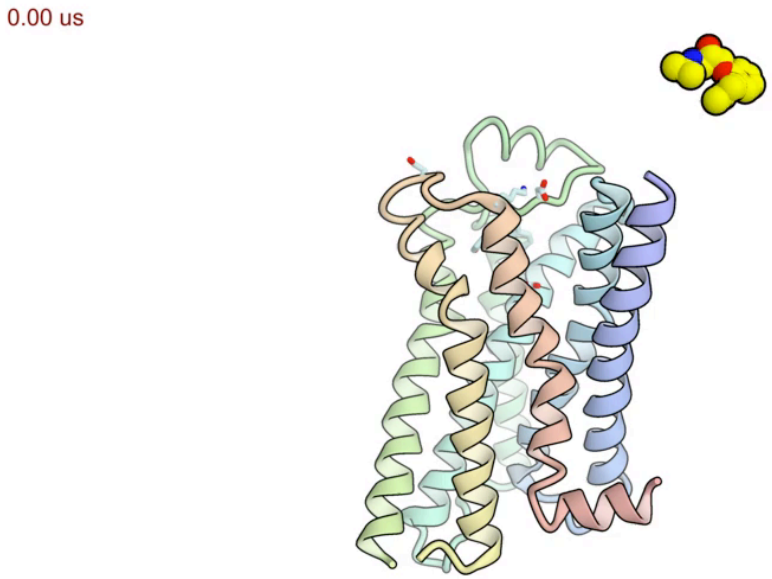
- Binding affinity is different from “how long the ligand remains bound” (the off-rate) or “how quickly the ligand binds” (the on-rate)
  - Binding affinity is a ratio of the on-rate and off-rate; you can’t calculate it from either one alone
  - These rates are also of interest in drug discovery, and 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 *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 one 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, 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 an assumed 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)

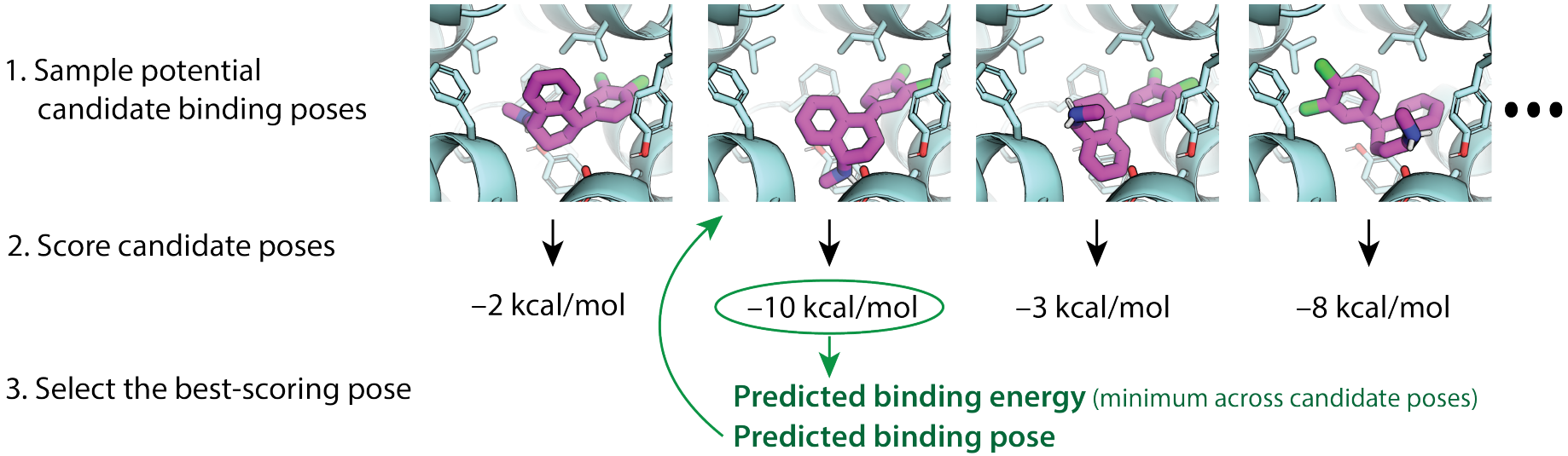


Figure by Ayush Pandit and Joe Paggi
Paggi et al., *Annual Review of Biochemistry*, 93:389-410, 2024

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

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 flexibility, ligand flexibility, 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)

Program	Country of Origin	Year Published
AADS	India	2011
ADAM	Japan	1994
AutoDock	USA	1990
AutoDock Vina	USA	2010
BetaDock	South Korea	2011
DARWIN	USA	2000
DIVALI	USA	1995
DOCK	USA	1988
DockVision	Canada	1992
EADock	Switzerland	2007
eHiTS	UK	2006
EUDOC	USA	2001
FDS	UK	2003
FlexE	Germany	2001
FlexX	Germany	1996
FLIPDock	USA	2007
FLOG	USA	1994
FRED	UK	2003
FTDOCK	UK	1997
GEMDOCK	Taiwan	2004
Glide	USA	2004
GOLD	UK	1995
Hammerhead	USA	1996
ICM-Dock	USA	1997
Lead finder	Canada	2008
LigandFit	USA	2003
LigDockCSA	South Korea	2011
LIGIN	Germany	1996
LUDI	Germany	1992
MADAMM	Portugal	2009
MCDOCK	USA	1999
MDock	USA	2007
MolDock	Denmark	2006
MS-DOCK	France	2008
ParDOCK	India	2007
PhDOCK	USA	2003
PLANTS	Germany	2006
PRO_LEADS	UK	1998
PRODOCK	USA	1999
ProPose	Germany	2004
PSI-DOCK	China	2006
PSO@AUTODOCK	Germany	2007
PythDock	South Korea	2011
Q-Dock	USA	2008
QXP	USA	1997
rDock	UK	2013
SANDOCK	UK	1998
SFDOCK	China	1999
SODOCK	Taiwan	2007
SOFTDocking	USA	1991
Surflex	USA	2003
SYSDOC	USA	1994
VoteDock	Poland	2011
YUCCA	USA	2005

Glide and GOLD are often considered the best commercial docking software packages. They're expensive, but you have free access through Stanford.

DOCK is free and very powerful in the hands of experts. AutoDock Vina and AutoDock are free and easier to use.

http://en.wikipedia.org/wiki/Docking_(molecular)

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:

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

Friesner et al., Journal of Medicinal Chemistry 47:1739-49 (2004)

  - 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

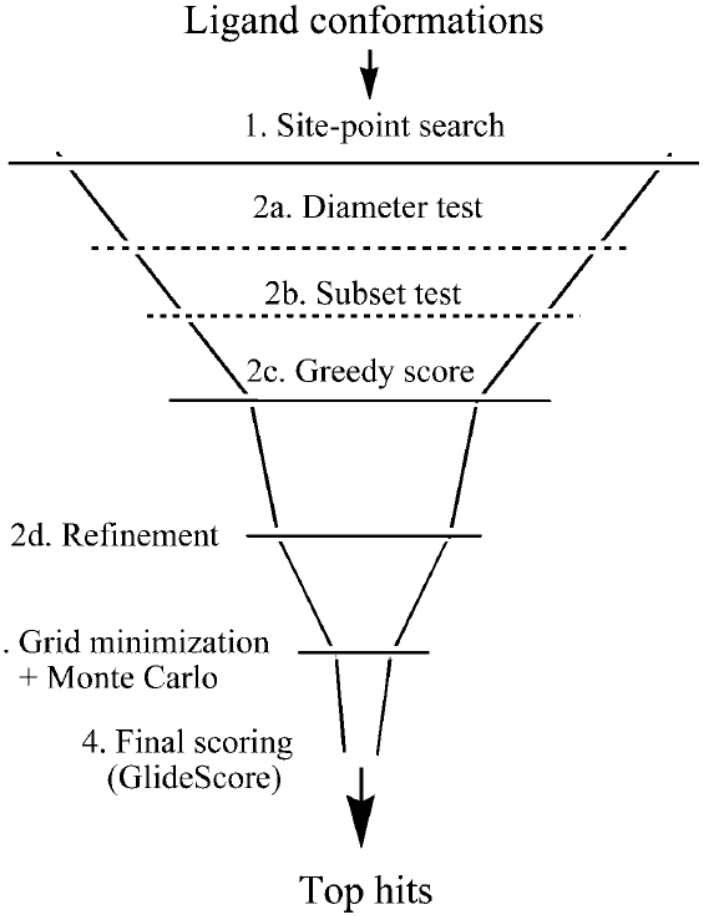
Optional material

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: discovery ligands that bind to a target
- 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

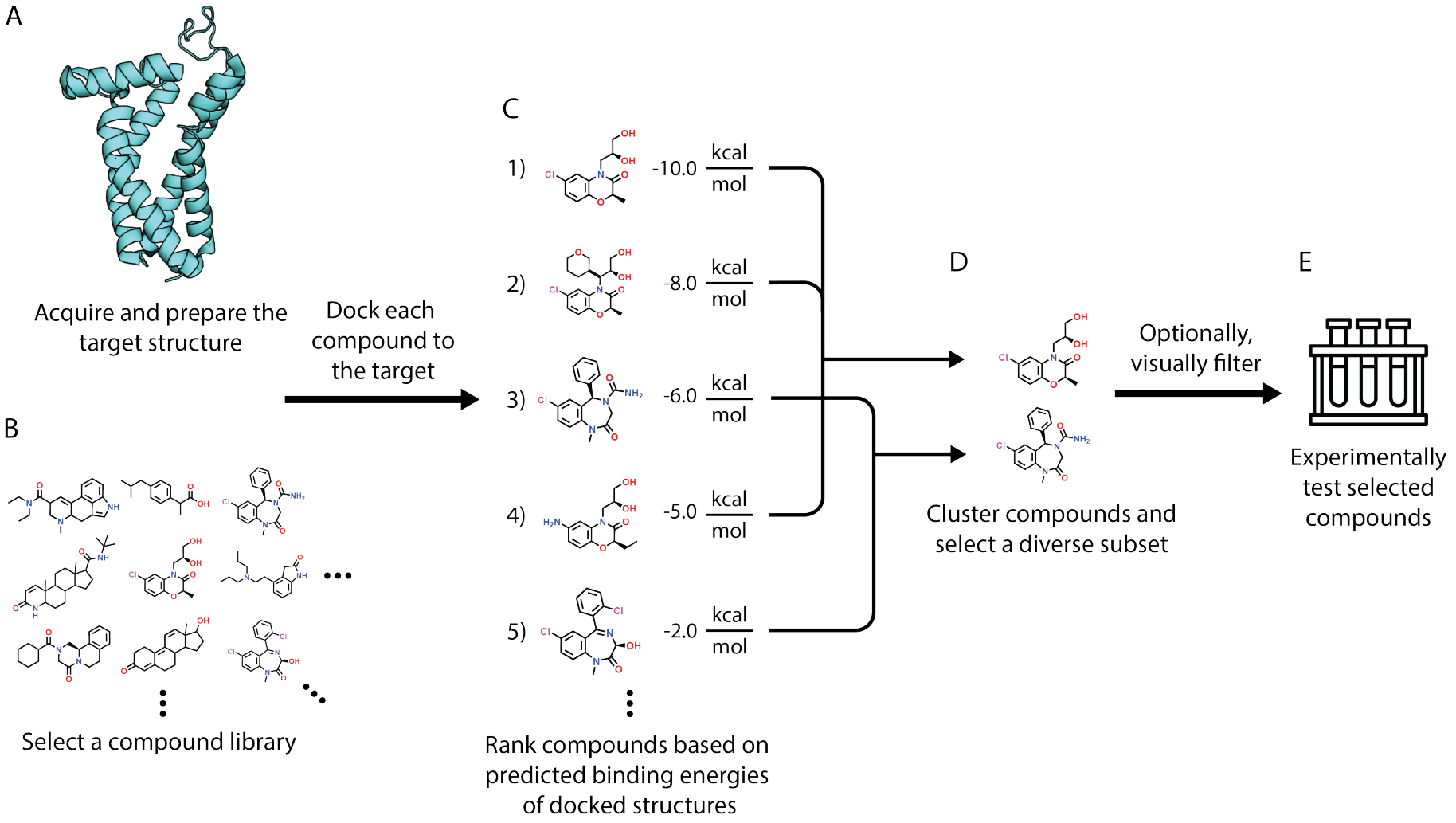


Figure by Ayush Pandit and Joe Paggi
Paggi et al., Annual Review of Biochemistry, 93:389-410, 2024

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
- 15–20 years ago, only a few million compounds were easily available from vendors



REAL Space: >100 billion compounds
xREAL: trillions of compounds

- Now it’s 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

How well does standard docking work?

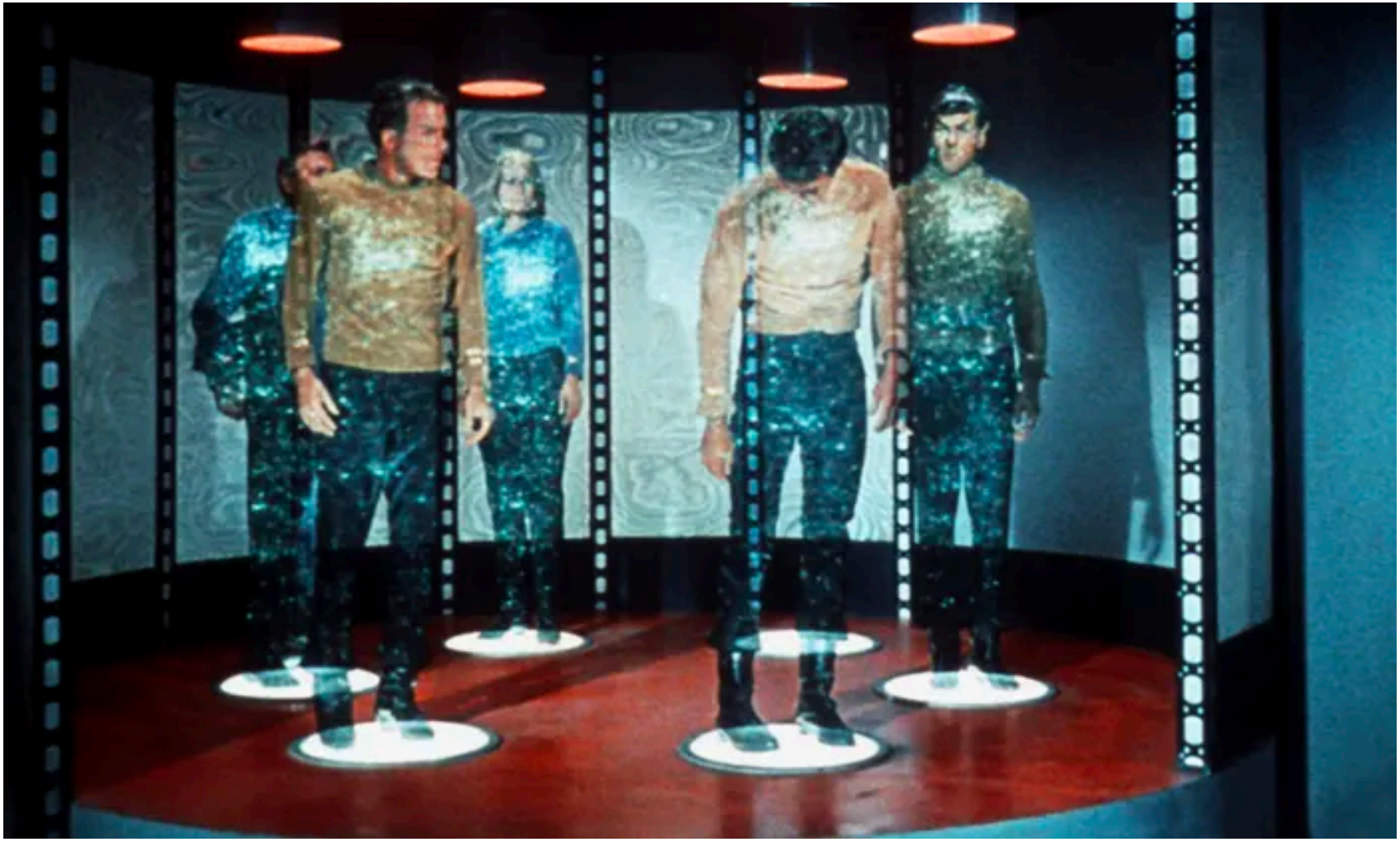
- The best standard docking protocols:
  - Predict a reasonably accurate pose (for ligands that do in fact bind the target protein) about half the time
    - Most of the time, some highly-ranked pose is reasonably accurate, but it may not be ranked first
  - Only give very rough affinity estimates
  - Provide very useful (though far from perfect) results for virtual screening
    - Many recent, experimentally verified success cases
    - Virtual screening is different from affinity estimation because, in virtual screening, missing some actual binders in the library is fine as long as one can identify other binders with reasonable confidence.

More information in Paggi et al., *Annual Review of Biochemistry*, 93:389-410, 2024
For rigorous (though outdated) comparisons of docking packages, see Leach et al., J *Med Chem* 49:5851 (2006); Warren et al., *J Med Chem* 49:5912 (2006)

Alternative methods and current research directions

Optional material

MD-based approaches



https://www.theguardian.com/technology/2015/jan/23/german-scientists-teleporter-transporter-3d-printing-star-trek

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.*

MD-based approaches

- Alchemical methods currently represent the most accurate way to predict binding affinities, particularly 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.
- Limitations
  - Assume a known binding pose for each ligand
  - Still very expensive computationally

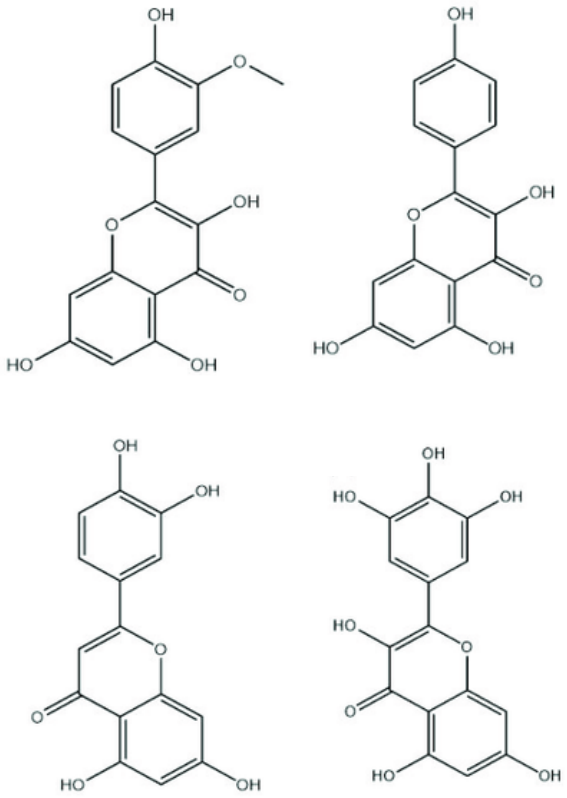
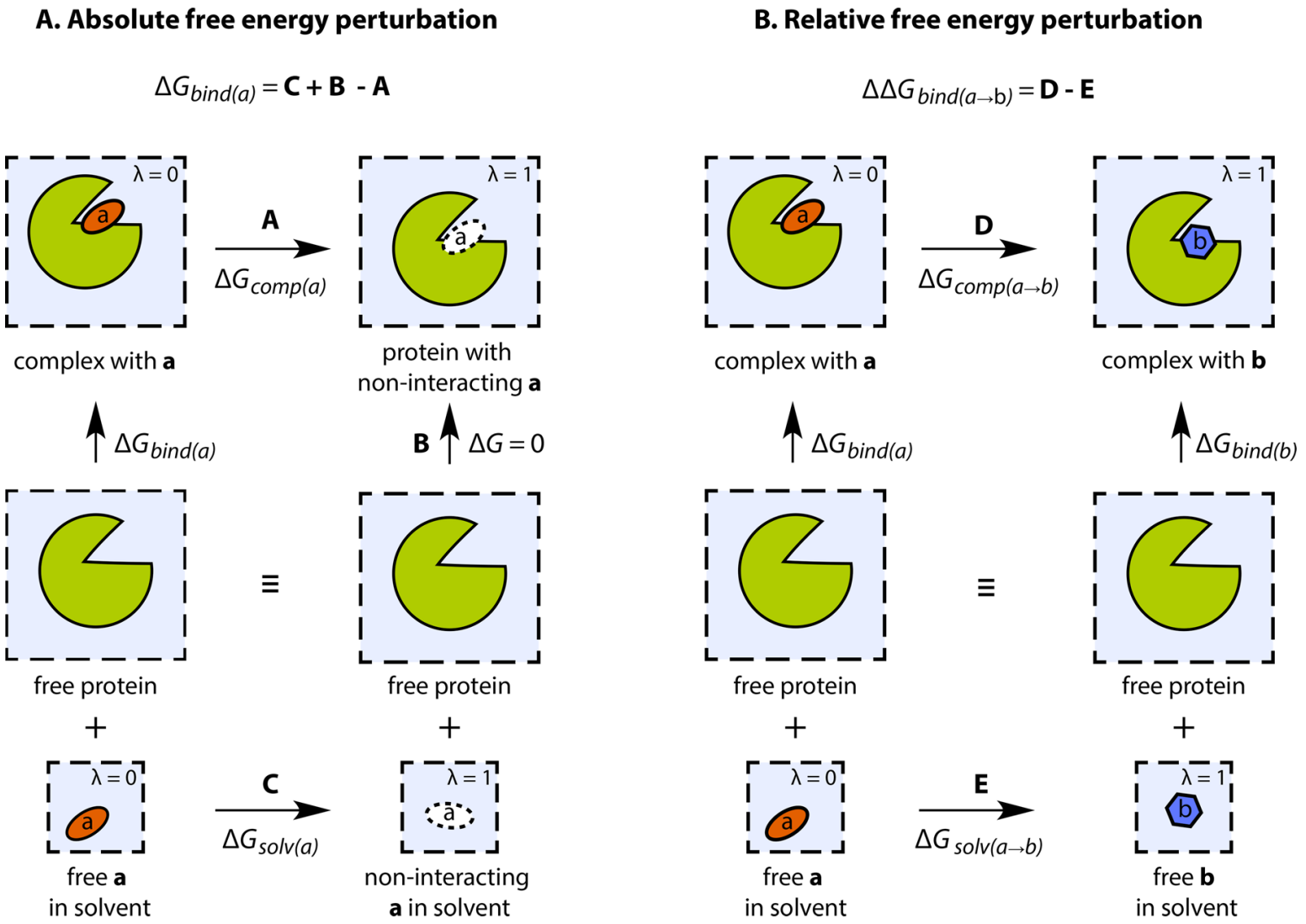


Image: https://www.mdpi.com/2543-6031/91/6/35

MD-based approaches

Calculating binding energy of a single ligand

Calculating difference in binding energies of two similar ligands



Optional material

From Williams-Noonan et al., *Journal of Medicinal Chemistry*, 2018, 61:638–649

AlphaFold 3 predicts binding poses for small-molecule ligands

- Should one use AF3-like methods or standard physics-based docking for pose prediction?
  - If one does't have an experimentally determined structure of the target, or if one doesn't know in advance where on the target a ligand binds, AF3-like methods are likely the best choice
  - When one has an experimentally determined structure of the target and knows the binding pocket for the ligand—a common scenario in drug discovery—the choice is less clear

Can one use AlphaFold for affinity prediction and virtual screening?

- AF3 does not predict affinity and was not designed to do so
  - AF3 will produce a binding pose for any given ligand
  - AF3 does give confidence measures for the predicted binding pose of each ligand, but these typically correlate only weakly with affinity
- Some recent AF3-like methods (Boltz-2, AQAffinity, etc.) are trained to predict affinity in addition to structure
  - The accuracy of the predicted affinities is highly debated

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

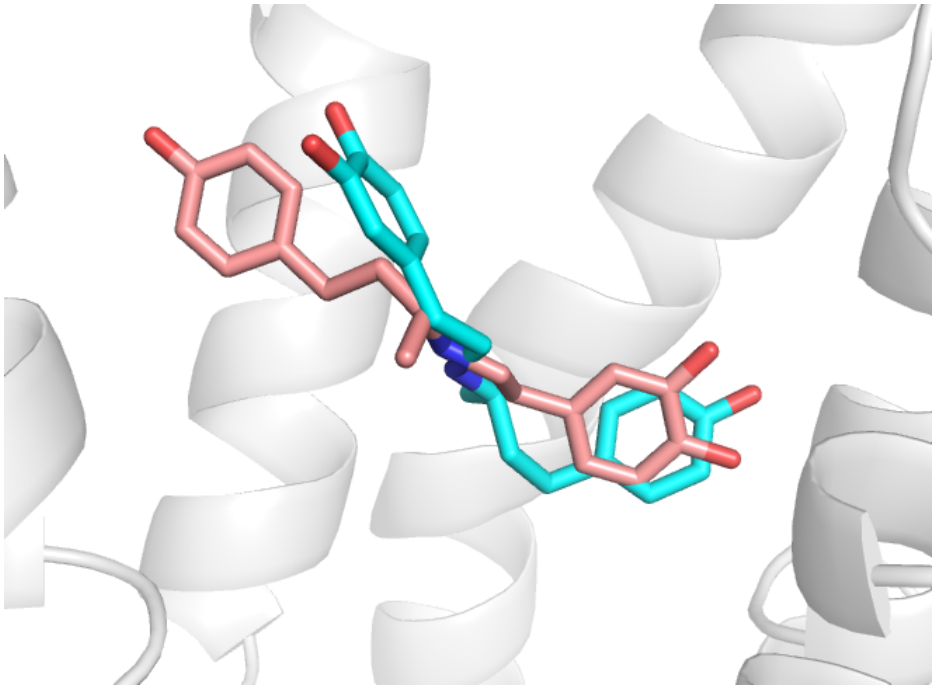
- The idea is to fit general functional forms (as described by large neural networks), rather than assuming specific functional forms based on approximations to physics

- A variant of this approach is to calculate standard physics-based docking scores for a subset of the compounds in the library, and then use the results to train a fast machine learning model to screen the remainder of the compounds (often referred to as “active learning”)

“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 a query ligand to previously characterized ligands
  - These approaches, which date back many decades, are a type of machine learning. Recent versions use modern machine learning methods.
  - They generally work well when one has experimentally characterized ligands similar to the query ligand

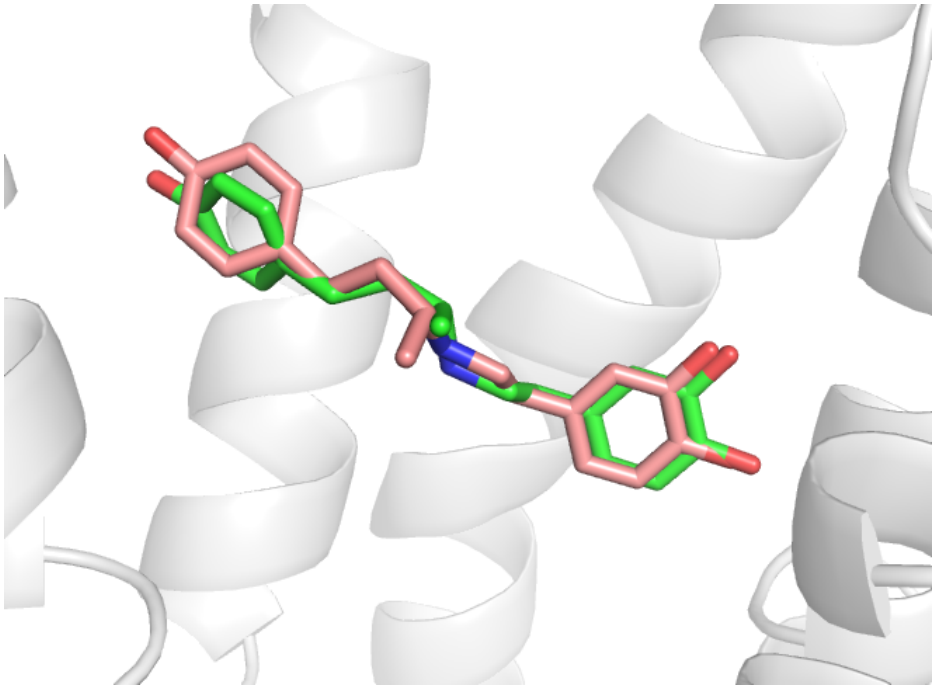
Another machine learning approach: experimental information on unrelated ligands can substantially improve docking predictions



Experimental structure
Computational prediction (docking)

Beta agonist dobutamine bound to
β1-adrenergic receptor

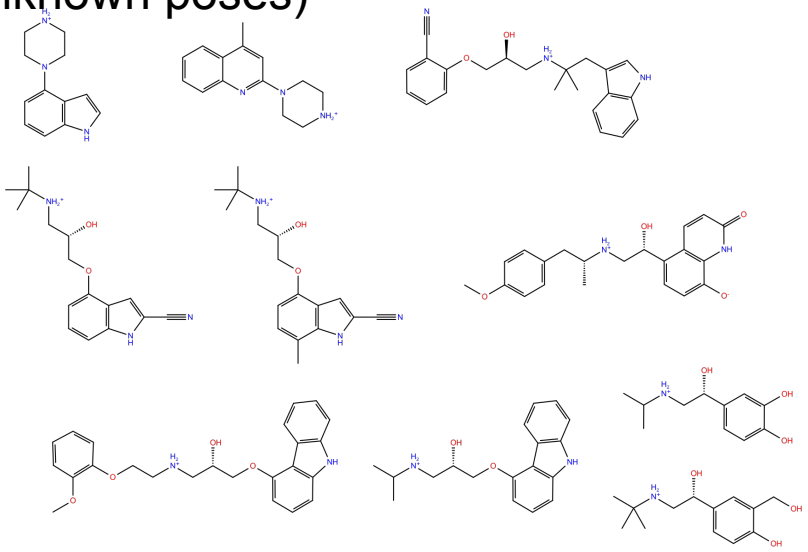
Another machine learning approach: experimental information on unrelated ligands can substantially improve docking predictions



Beta agonist dobutamine bound to β1-adrenergic receptor

Experimental structure
Computational prediction (ComBind)

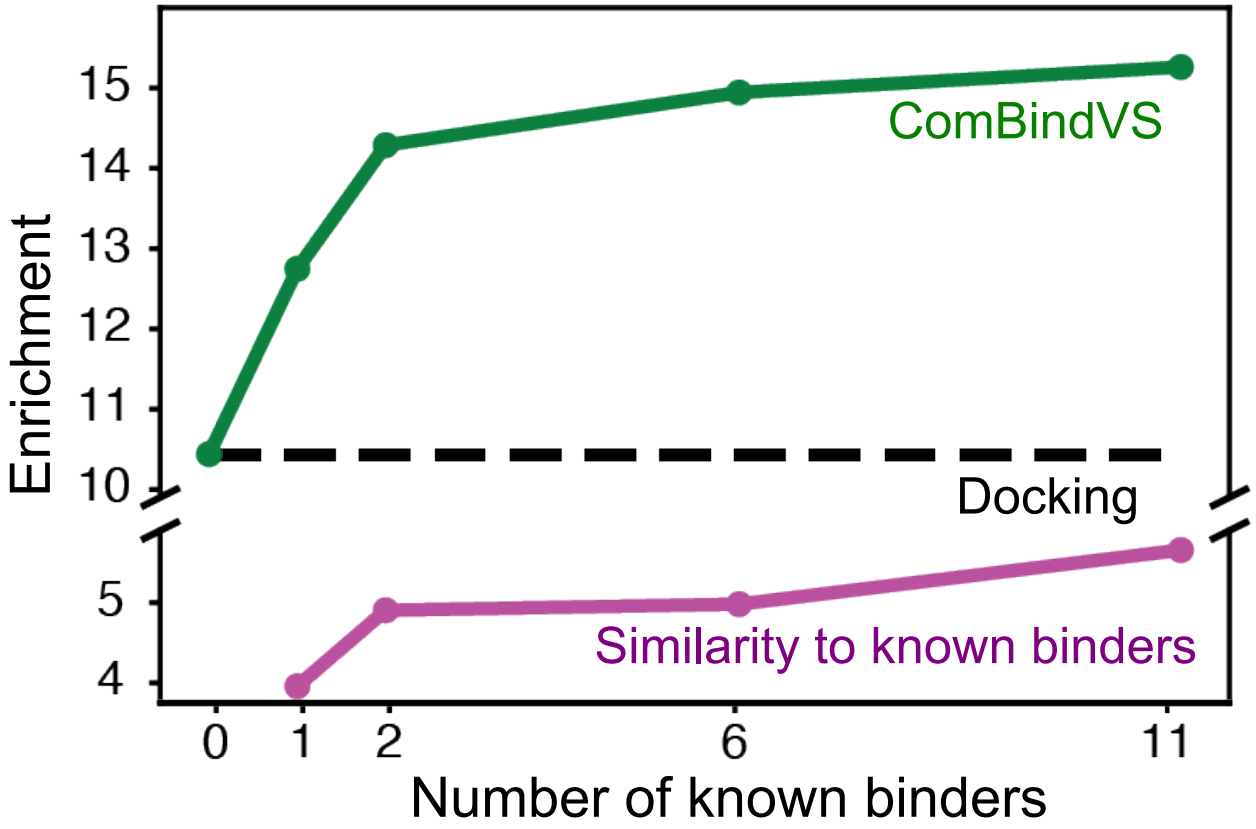
Prediction informed by the fact that the compounds below bind this target (in unknown poses)



Paggi, …, Dror, *PNAS* 2021

A similar approach (ComBindVS) improves virtual screening

Average performance across 102 targets in DUD-E benchmark set



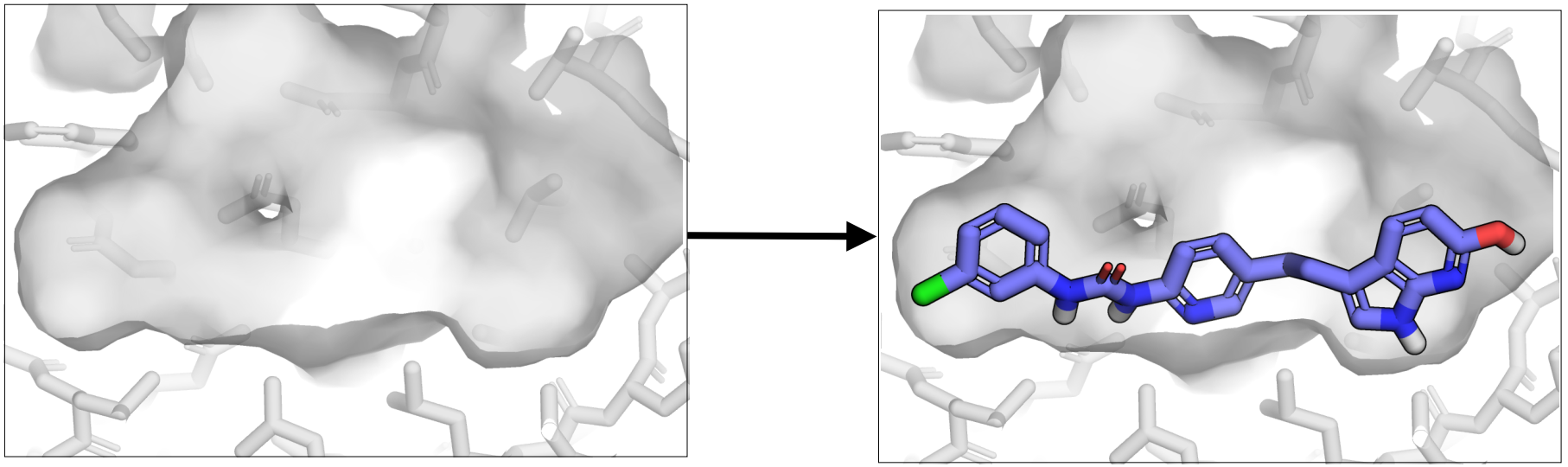
Note: All compounds screened are very different from known binders

Paggi, …, Dror, *PNAS* 2021

Current research area: Generative models for ligands

Input: Protein structure

Output: Small-molecule ligand



- Goal: given a protein structure and a specified binding pocket, generate small molecules that bind to that pocket with high affinity
- Very active research area, with many recent papers
  - Many of these use denoising diffusion methods, in which atoms can change element types during denoising
  - A challenge is ensuring that the generated ligands have drug-like properties (e.g., selective, soluble, non-toxic)