Schrödinger

Docking and Virtual Screening with Glide

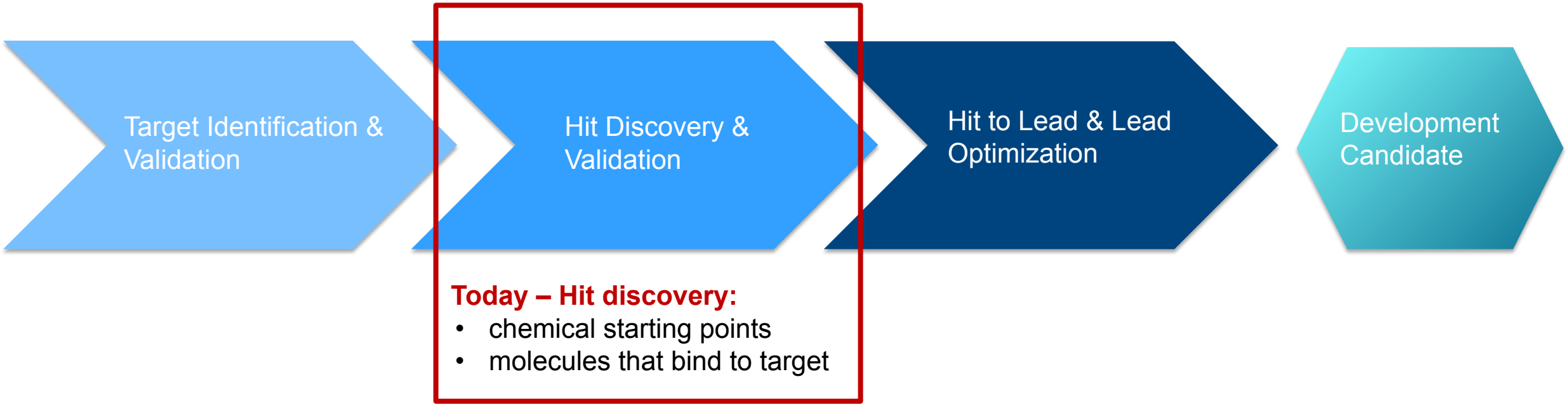
Fredericks National Laboratory
October 2022

Agenda

- ***9:00 – 9:05 Introduction & set-up***
- ***9:05 – 9:35 Presentation***: Principles of virtual screening with docking (Loukas Petridis)
- ***9:35 – 9:45 Break***
- ***9:45 - 10:30 Hands-on tutorial:*** Maestro (Kat Bay)
  - Maestro basics
  - Protein preparation
  - Grid generation
  - Ligand preparation
- ***10:30 – 10:40 Break***
- ***10:40 – 11:35 am Hands-on tutorial***: Glide (Loukas Petridis)
  - Redocking cognate ligand
  - Screening of library
  - Analyze results
- ***11:35 – 12:30 Questions & Answer session*** (Deren, Kat & Loukas):
  - Questions by attendees

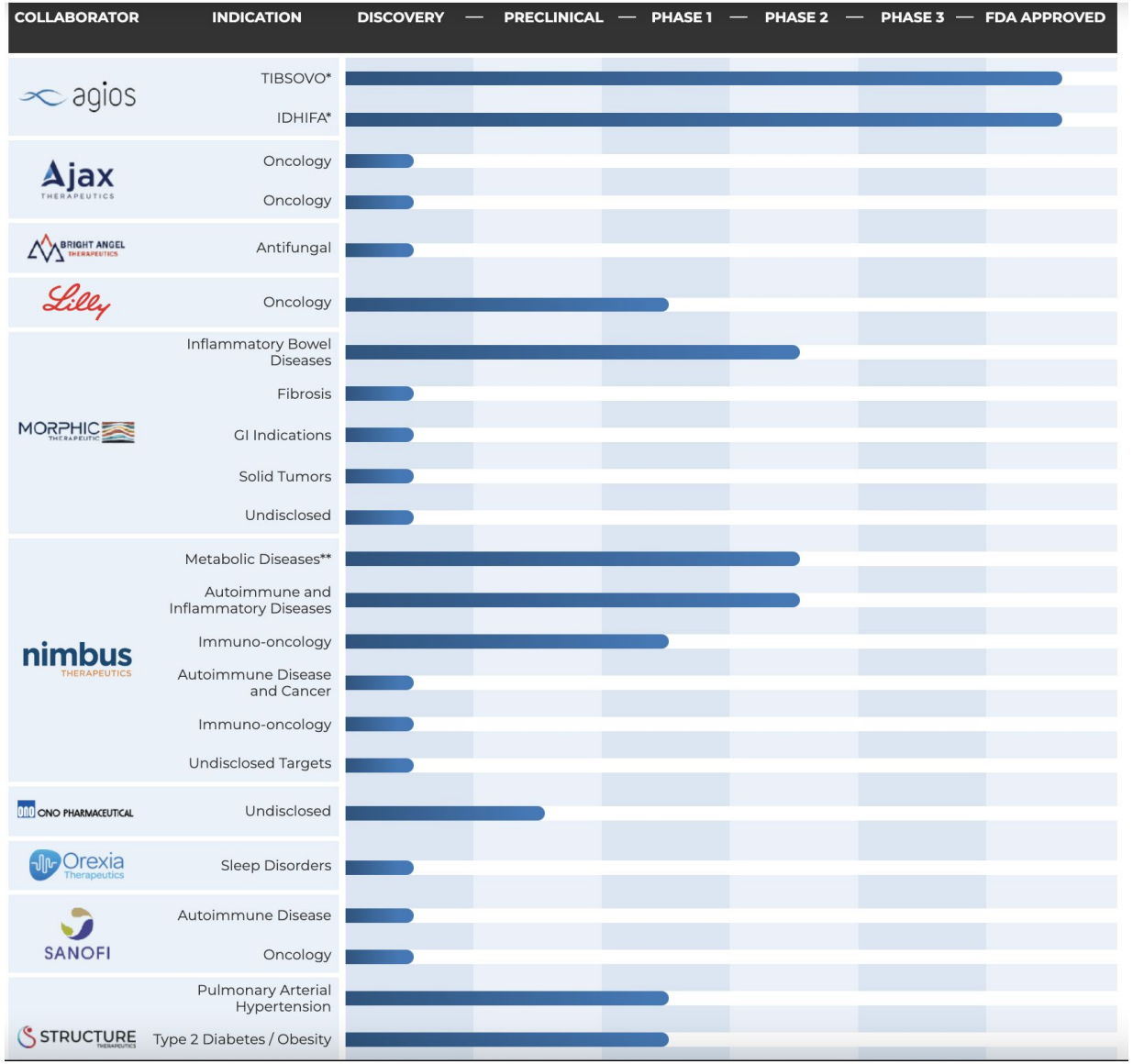
Introduction to Hit Discovery

Schrödinger offers tools to help all along the pre-clinical drug discovery process

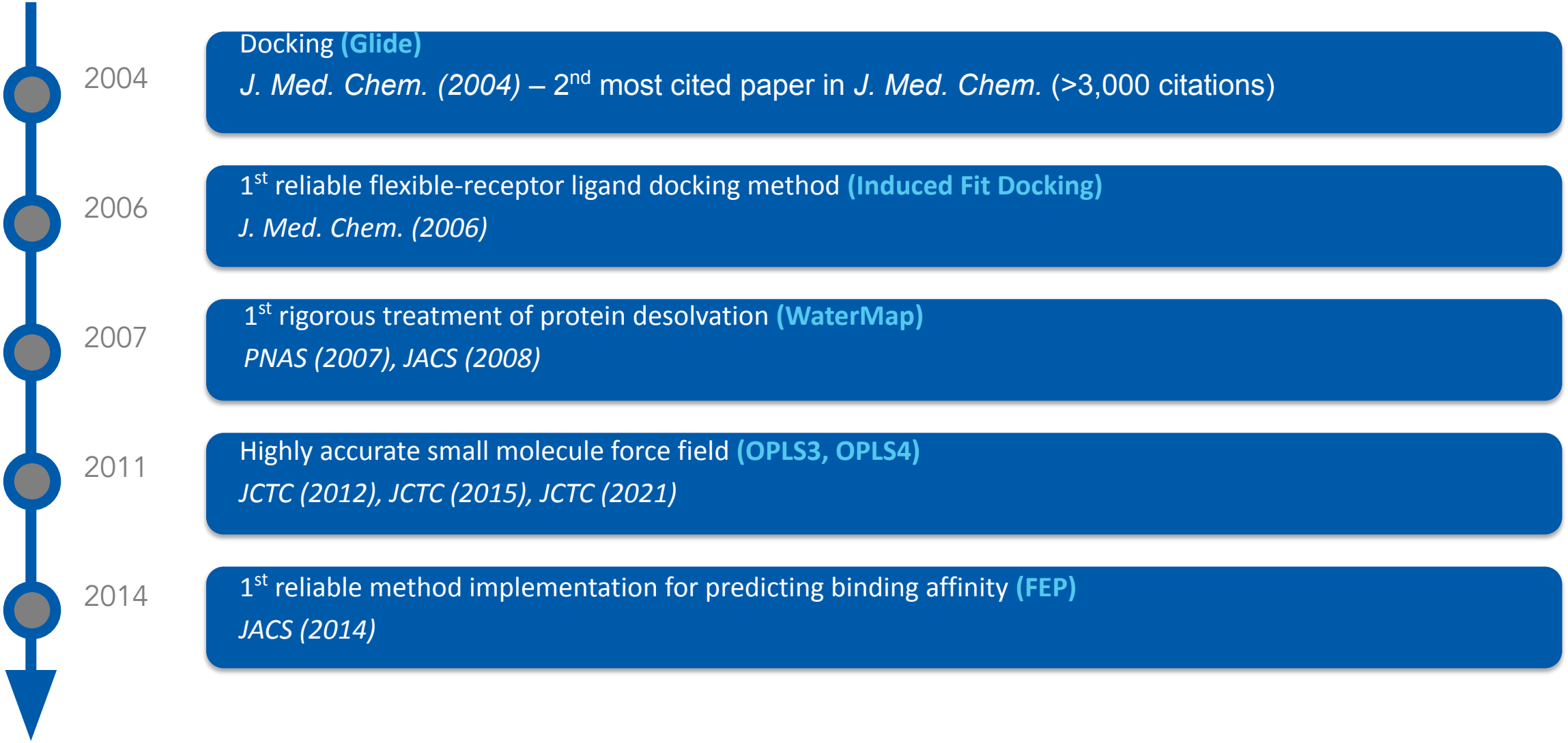


Schrodinger: physics-based technology validated across hundreds of targets in real-world drug discovery projects

research collaborations



Schrödinger's commitment to make scientific breakthrough



Schrödinger's commitment to make scientific breakthrough

2017 — Predict effect of mutations in proteins **(Protein FEP)**
J Mol Biol

2018 — Aggregation-prone regions in proteins **(AggScore)**
Proteins

2021 — Xray and CryoEM structure refinement **(Phenix / OPLS3a)**
Structure

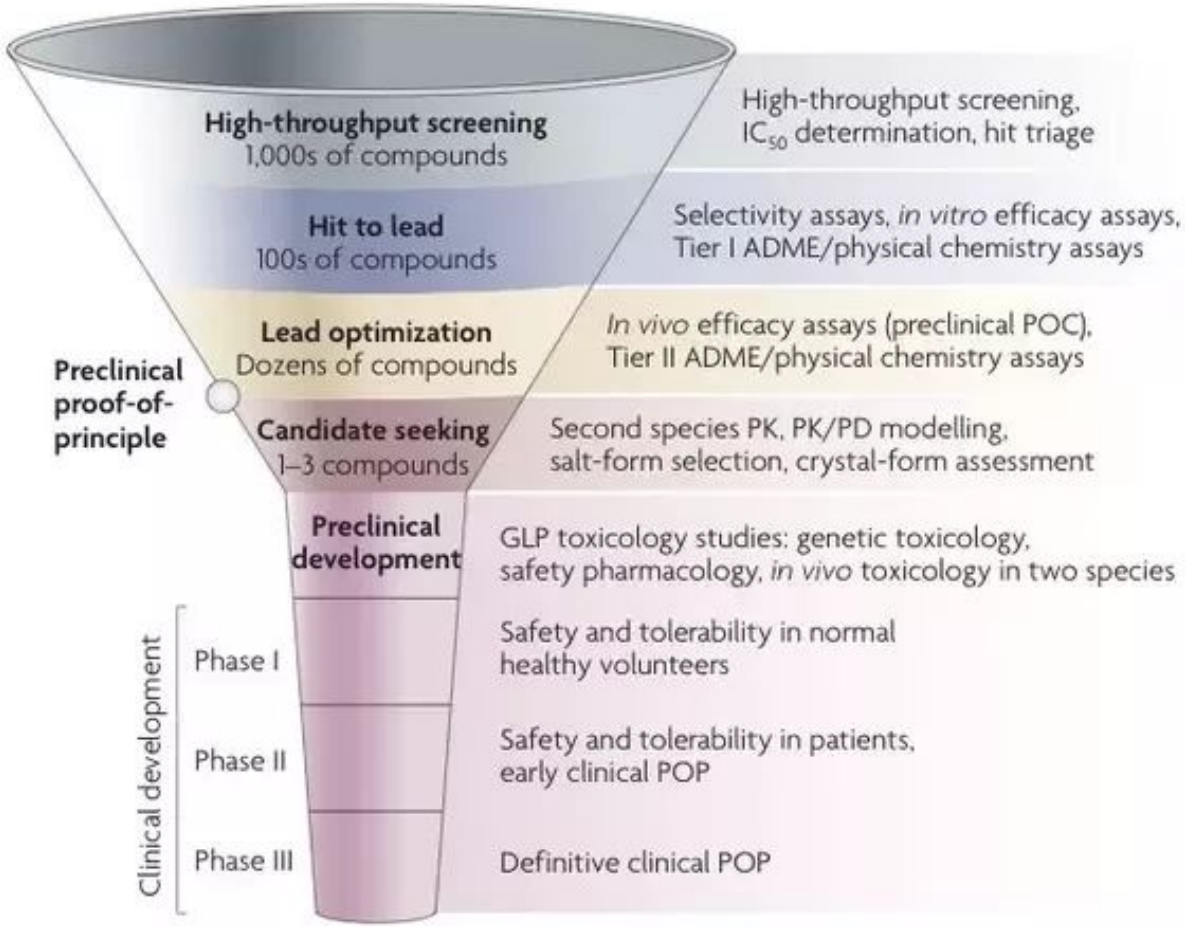
2021 — Large scale Docking with Deep learning **(Active Learning Docking)**
JCTC

2021 — Exploring large chemical space for lead optimization **(Autodesigner)**
Chemrxiv

The drug discovery funnel

10^{60} molecules estimated in Lipinski space[1]

Small number of chemical series typically examined in typical discovery projects

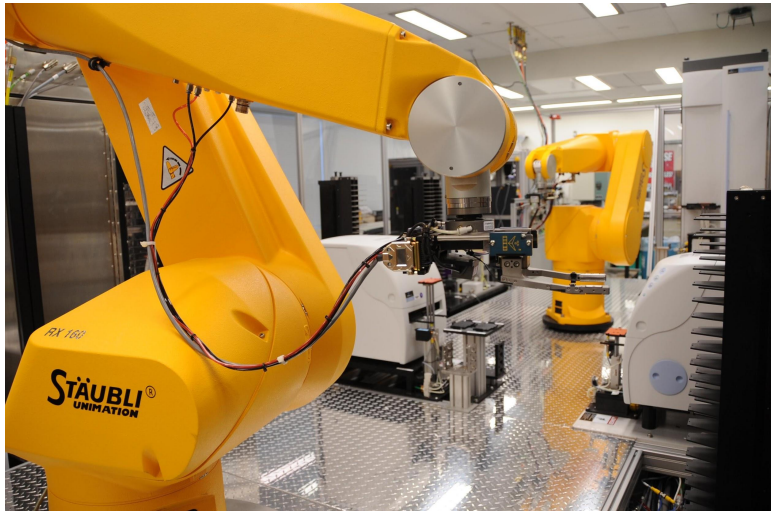


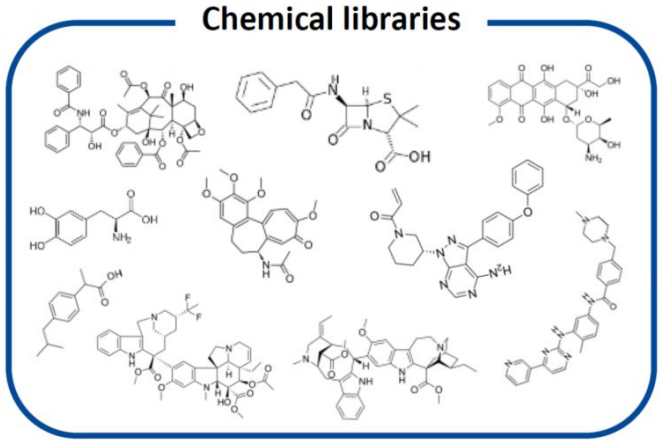
Find a small molecule that inhibits (or activates) a protein, resulting in a **therapeutic effect in a disease**.

Nature Reviews Drug Discovery 6, 636-649

High-throughput screen to identify Hits

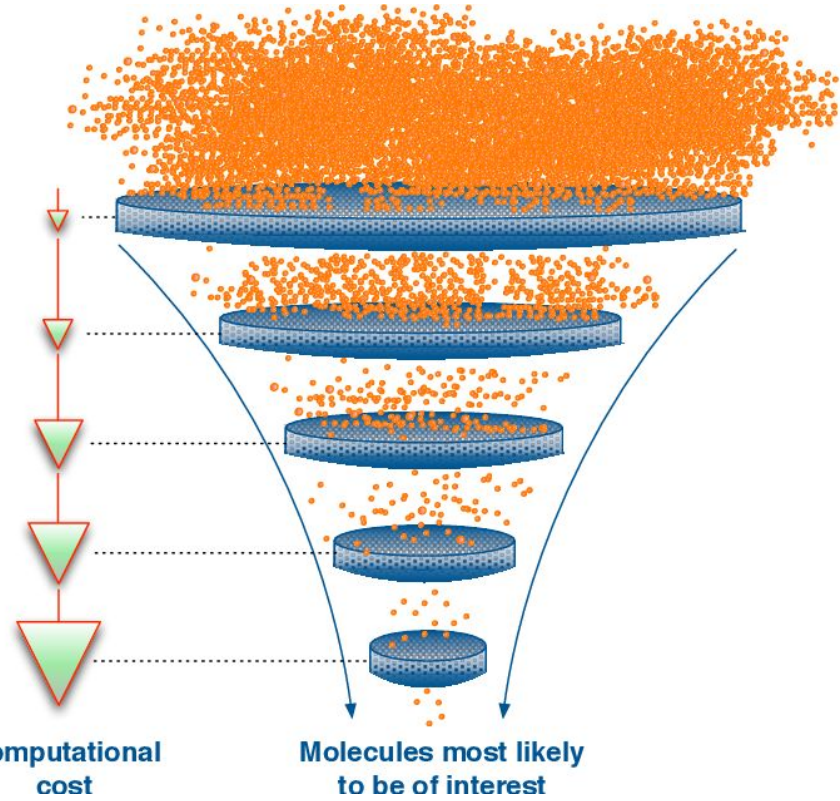
Experimentally





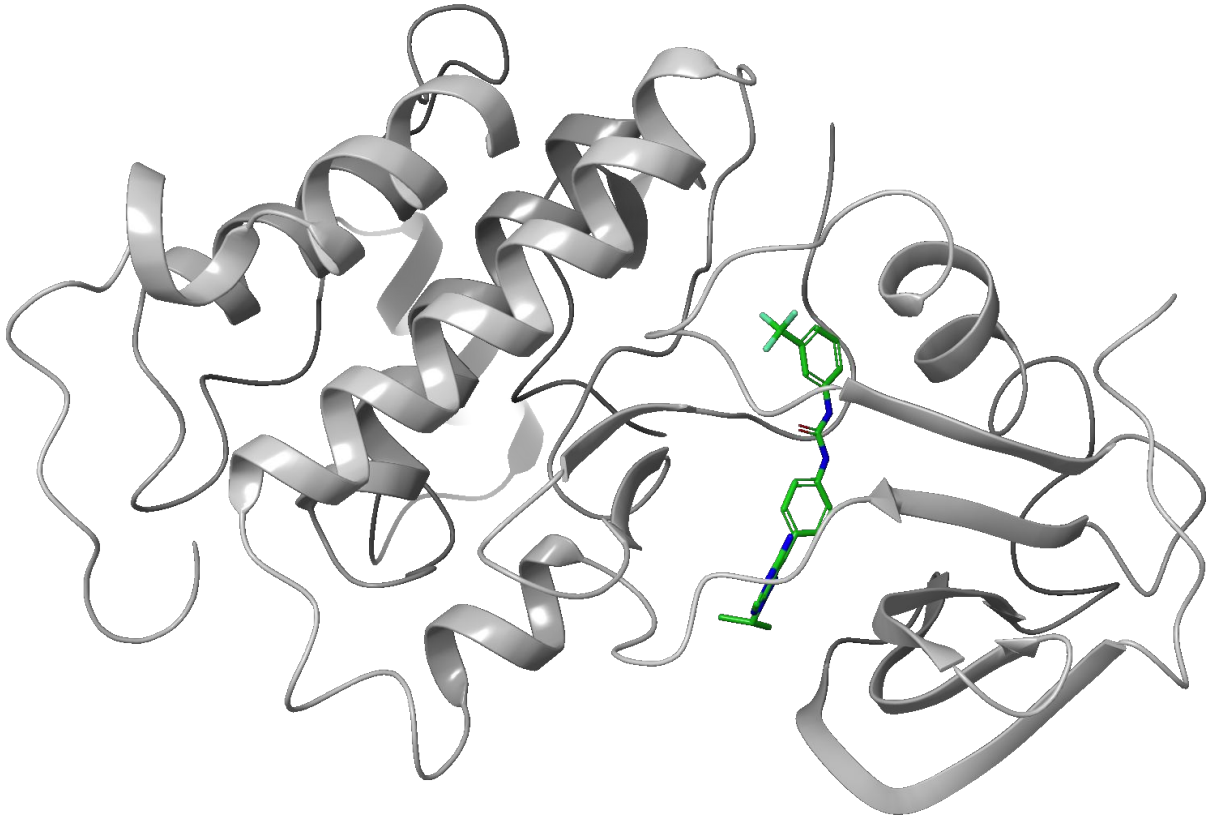
Identify “**actives**”
Bind target

Computationally

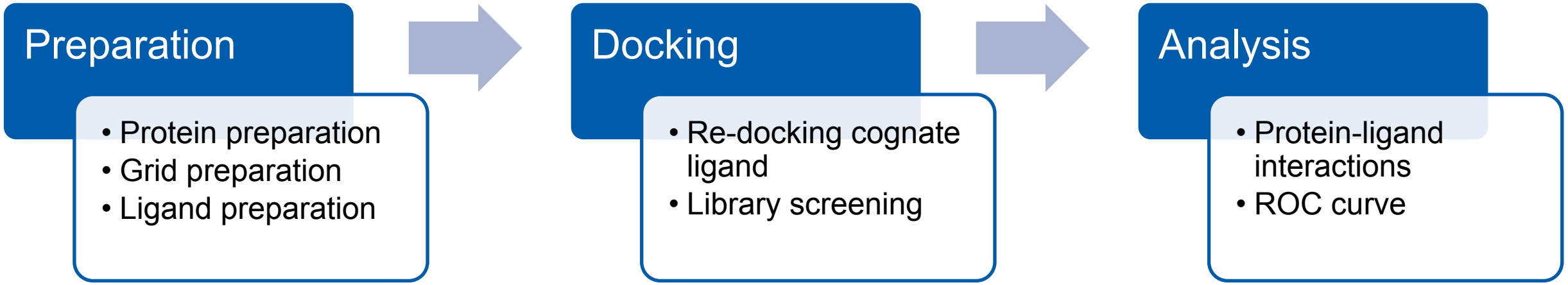


Structure-based hit discovery

- Find a (small) molecule that **binds** to the target.
- Use 3D information on target-ligand interactions
- Physics-based = Requires only one datum: the 3D structure of the target

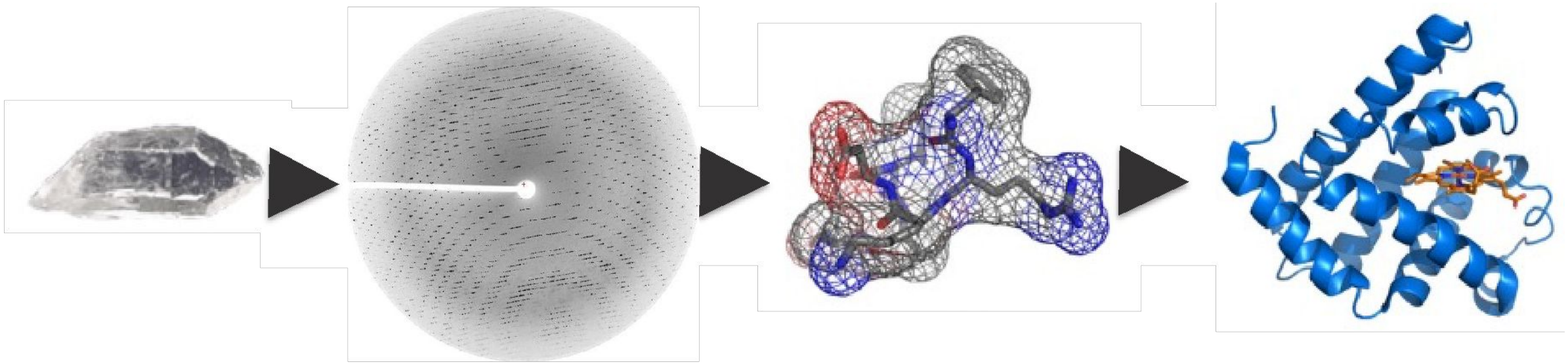


Computer-aided structure-based drug discovery workflow



Preparing Structures for Molecular Modeling

From a crystal to a model



protein crystal

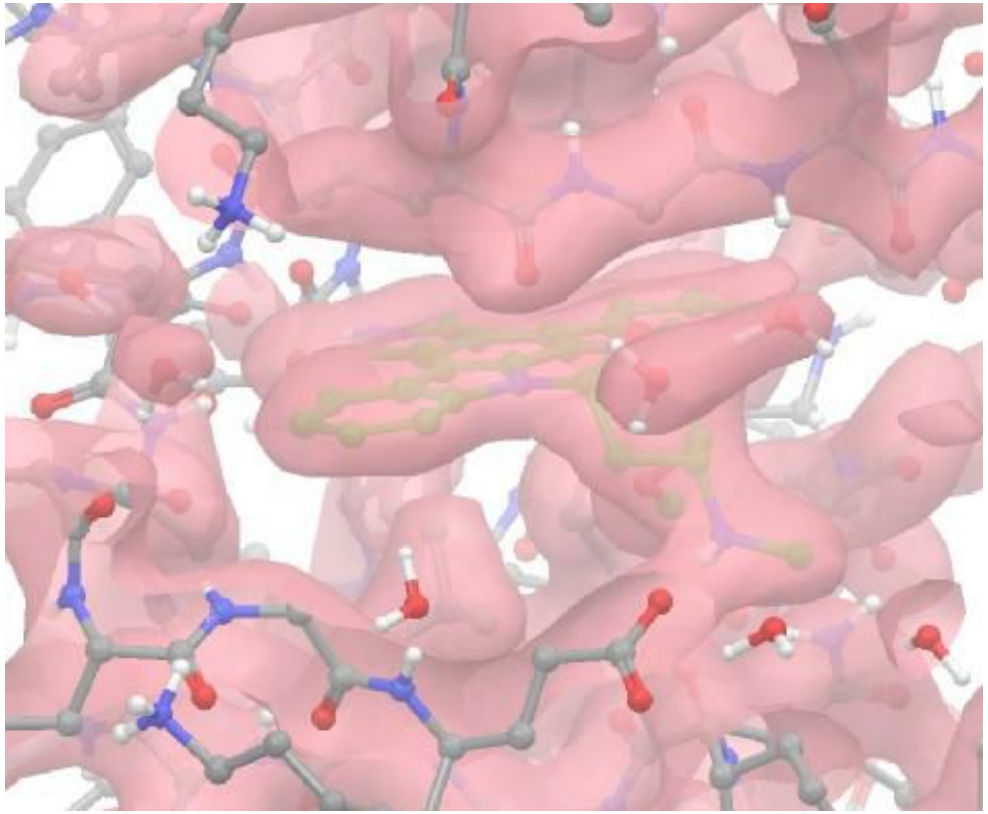
diffraction pattern

electron density map

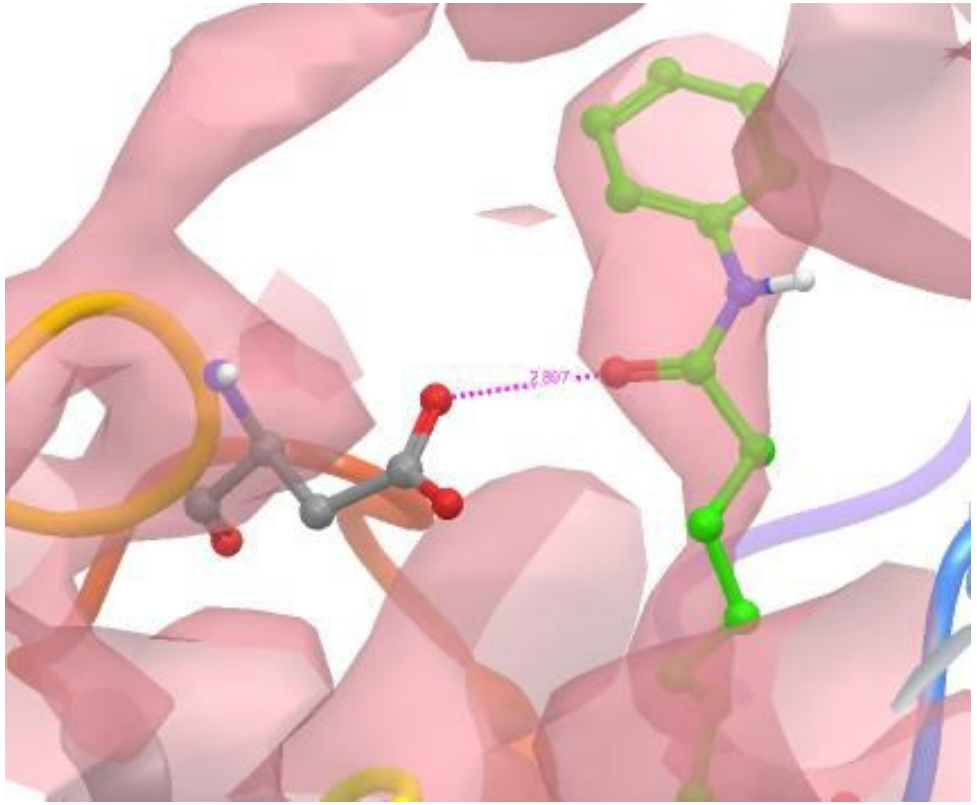
atomic model

Adapted from: http://www.scistyle.com/

Not all experimental data contain the same amount of information

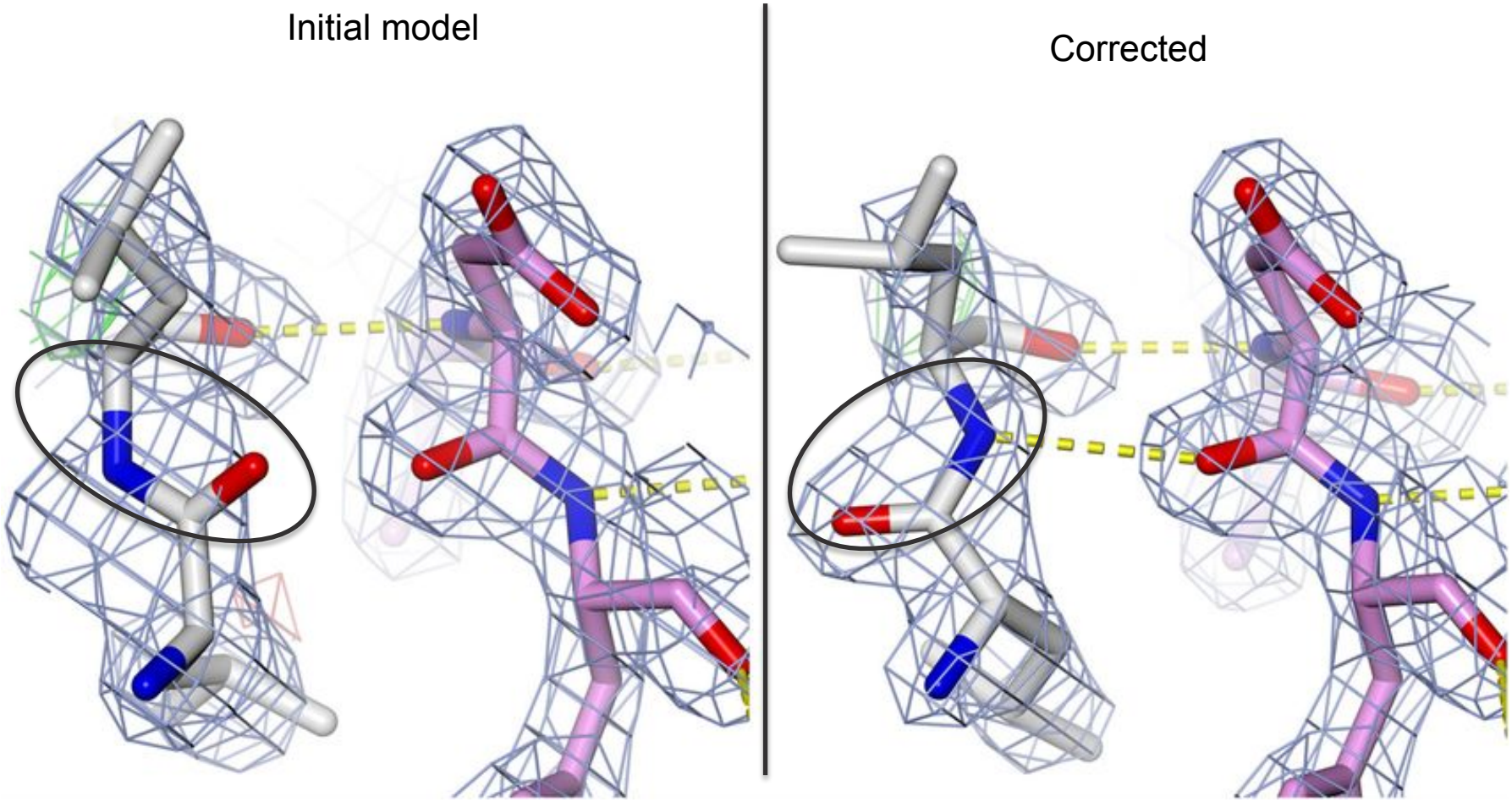


The ligand density is relatively unambiguous.



Density is missing, which may result in misleading information.

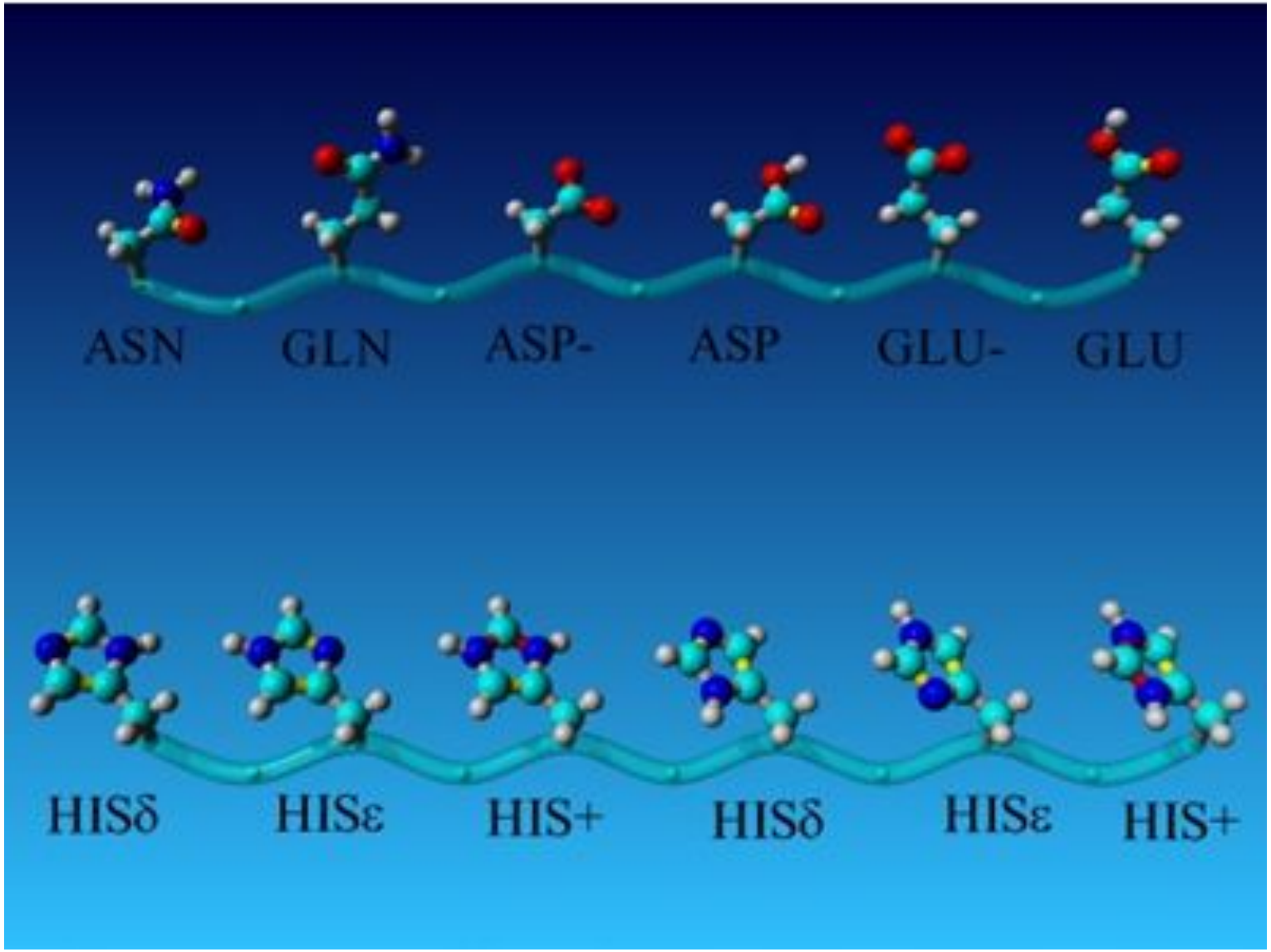
Low-resolution X-ray can be ambiguous



Peptide plane flip of Leu to allow a HB to the carbonyl of Asp

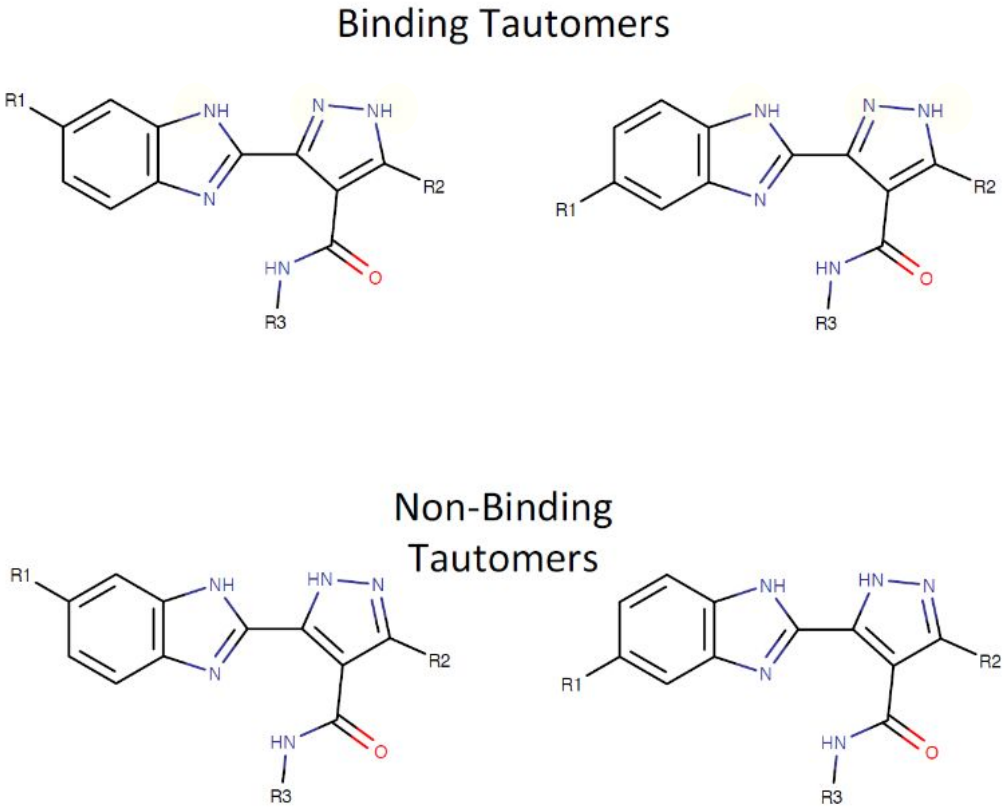
Difficult to resolve H in X-ray crystallography or CryoEM

- pH-dependent protonation states
- tautomeric states (chemical compounds that readily interconvert)



Ligand chemical diversity can affect its binding to receptors

- LigPrep generates chemical *species*
  - protonation states (differ in charge)
  - Tautamers (differ in the position of the protons)
- Species are important because they influence binding of a ligand to the receptor



Protein & ligand preparation wizards improve the initial structures

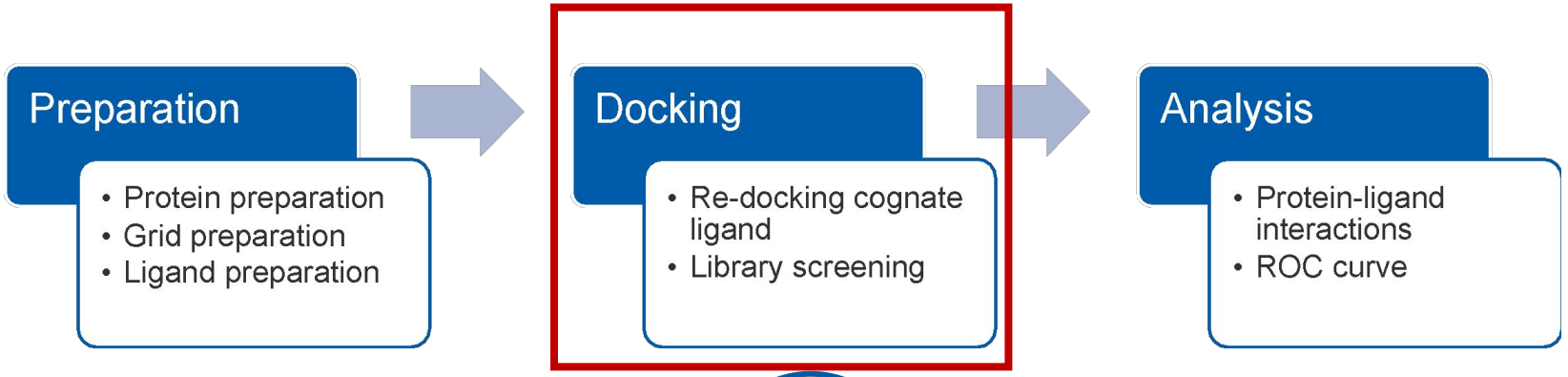
Protein Preparation

- Protonation
- Missing side chains
- Missing loops
- Hydrogen-bond optimization
- Restrained minimization

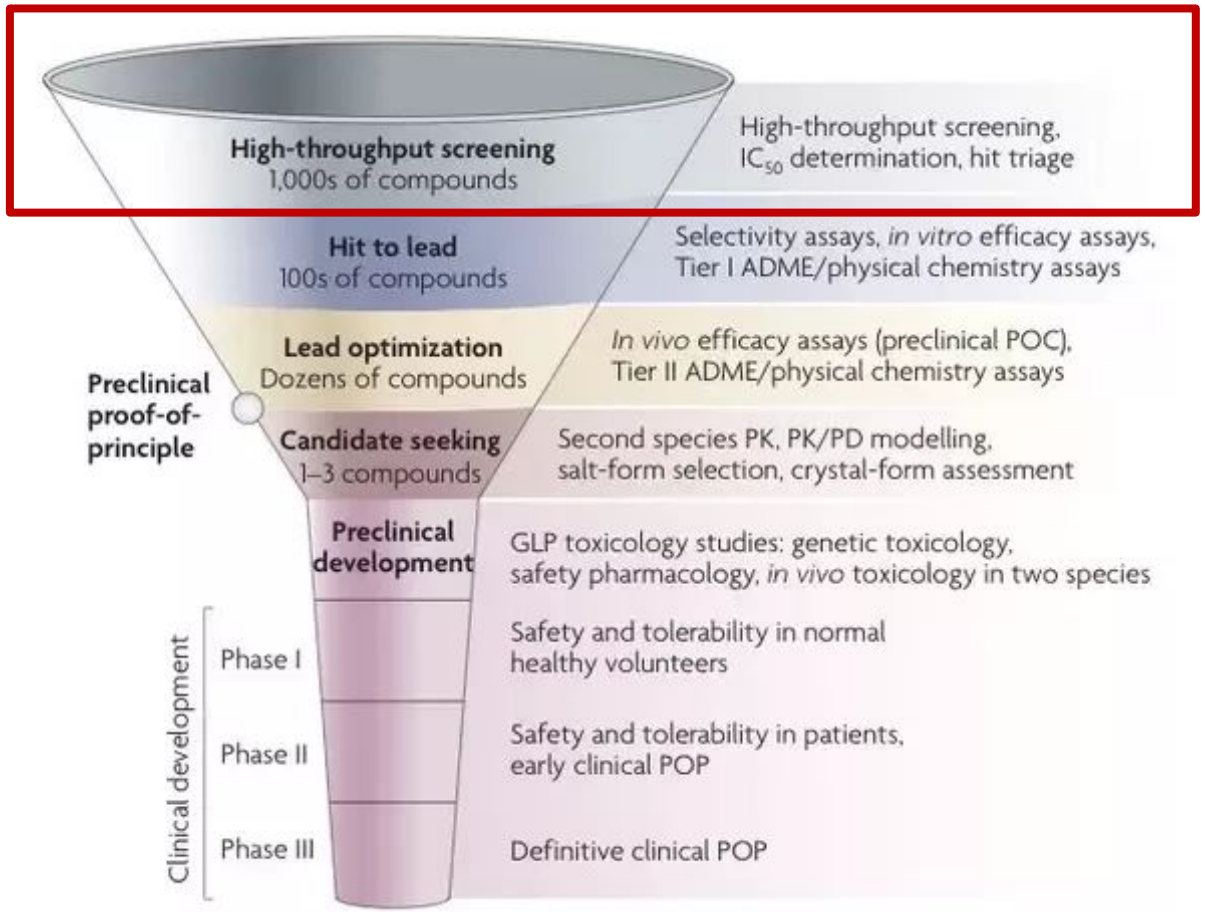
Ligand Preparation

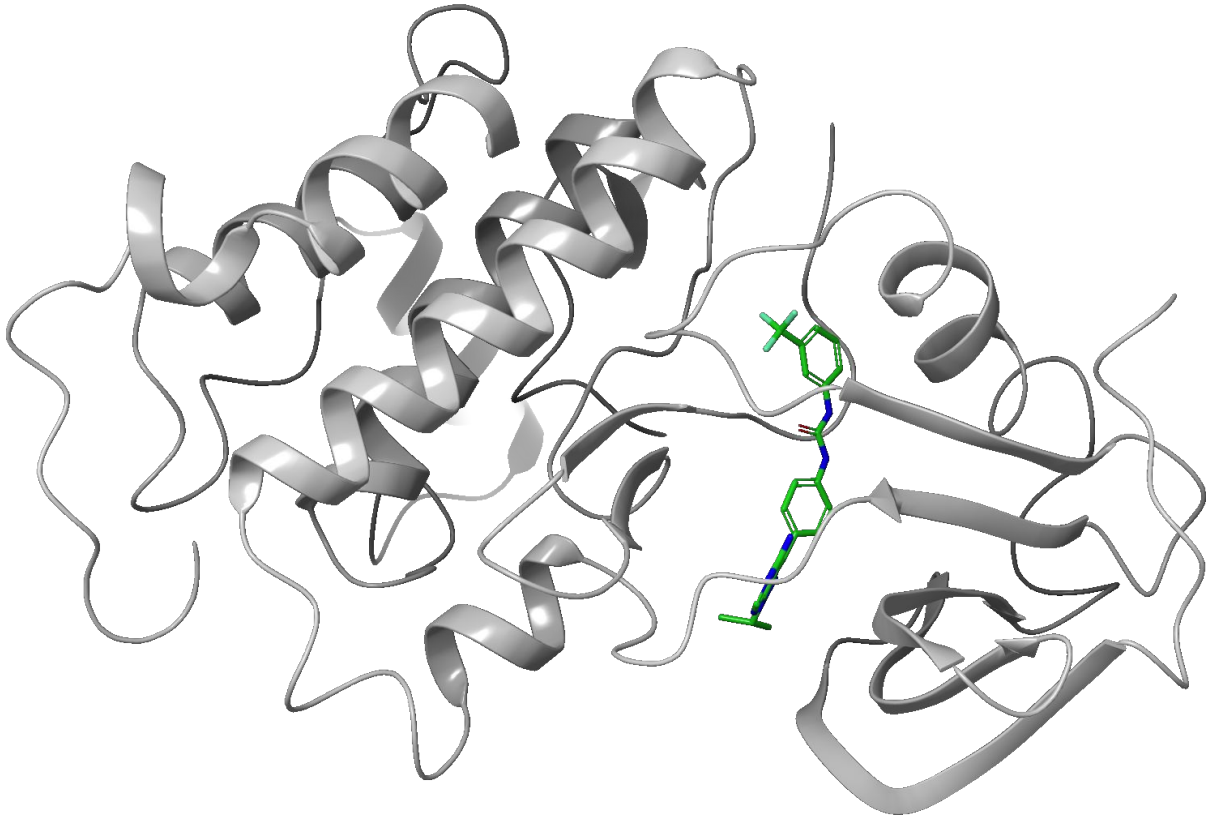
- Chemical and structural diversity from a single input structure.
- Energy minimized 3D structures.
- Tautomeric and ionization states
- Stereoisomers

Screening compounds using Glide



Structure-based hit discovery

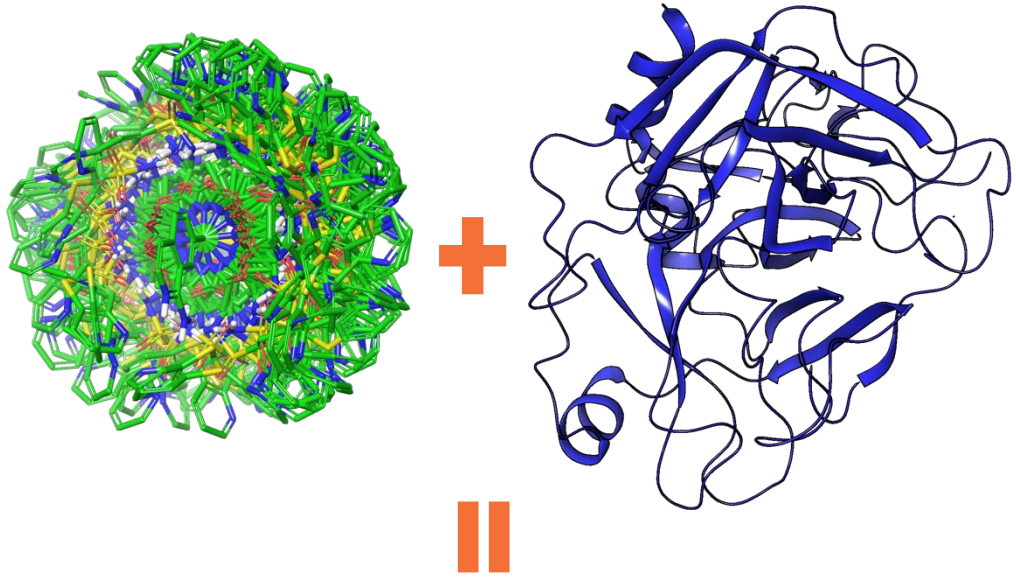




- Find a molecule that **binds** to the target.
- Use 3D information on target-ligand interactions
- Physics-based = Requires only one datum: the 3D structure of the target

Virtual screening - docking

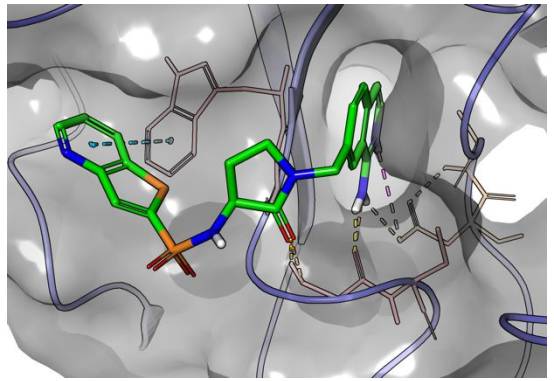
- Treat protein as rigid and ligand as flexible
  - Lock & key model
- Physics based model for Docking and Scoring steps



Docking → **Scoring**

What possible poses can my ligand adopt in the binding site?

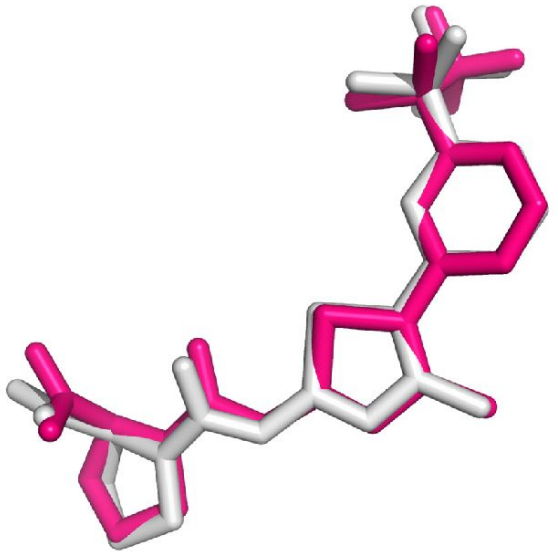
Which of those poses makes the best interactions with the receptor?



A “pose” is a ligand conformation oriented relative to the receptor

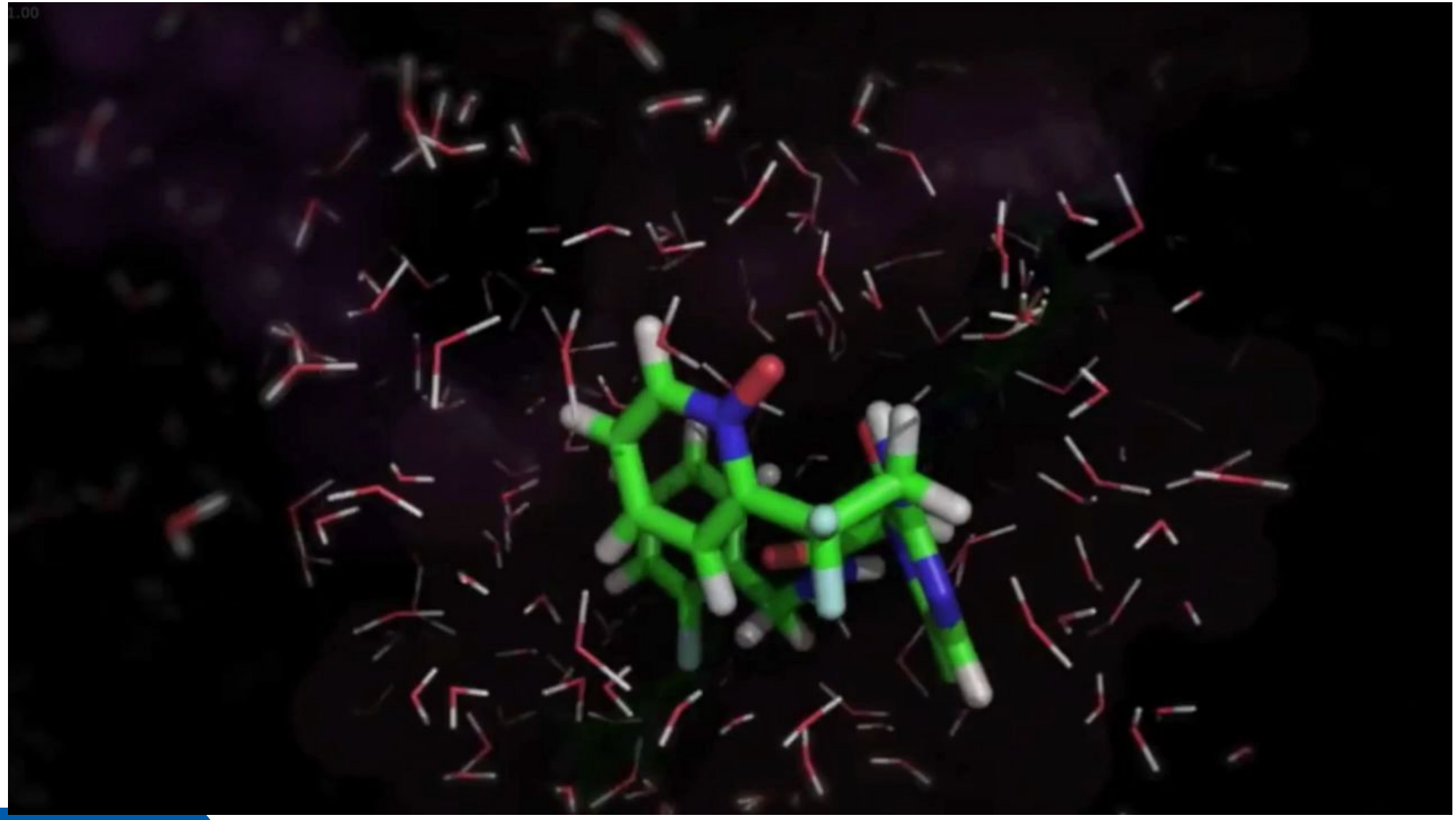
Re-docking: sampling efficacy of Glide can retrieve near-native poses in most cases

- **Re-docking**: the native ligand is docked to the target, and the ligand pose is compared to the crystal structure
- Average RMSD between the docked and cocrystallized poses = 1.17 Å,
- An acceptable binding mode with RMSD < 2 Å could be found in 80% of the cases.
- Typically use reduced vdW for ligand non-polar atoms to soften interactions



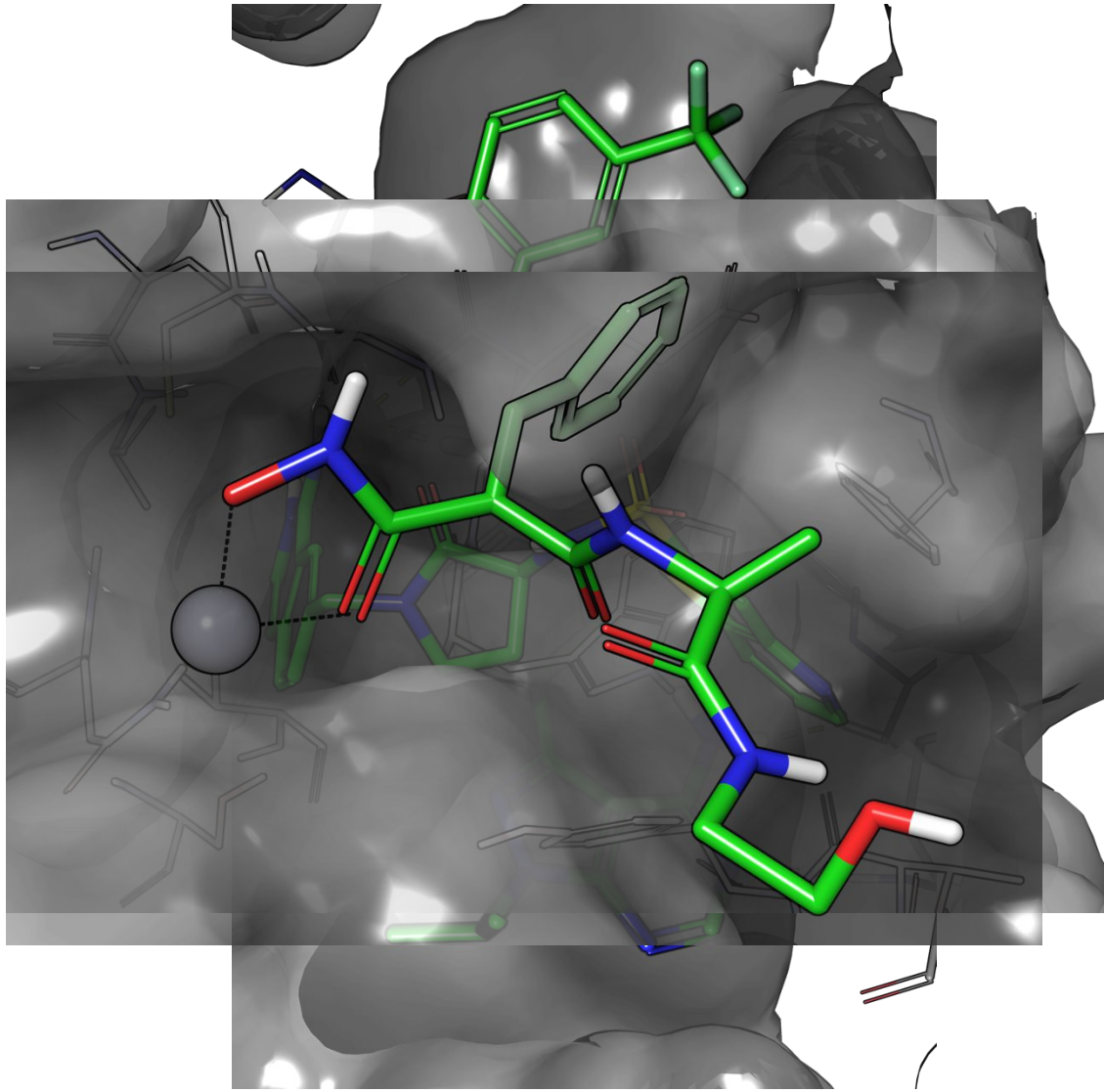
Percent of Poses with RMSD < X	WScore	Glide SP
< 1.5 Å	87	83
< 2.0 Å	92	90
< 2.5 Å	97	93
< 3.0 Å	98	95
# of complexes that failed to dock	0	4

Predicting binding affinity is a highly complex problem



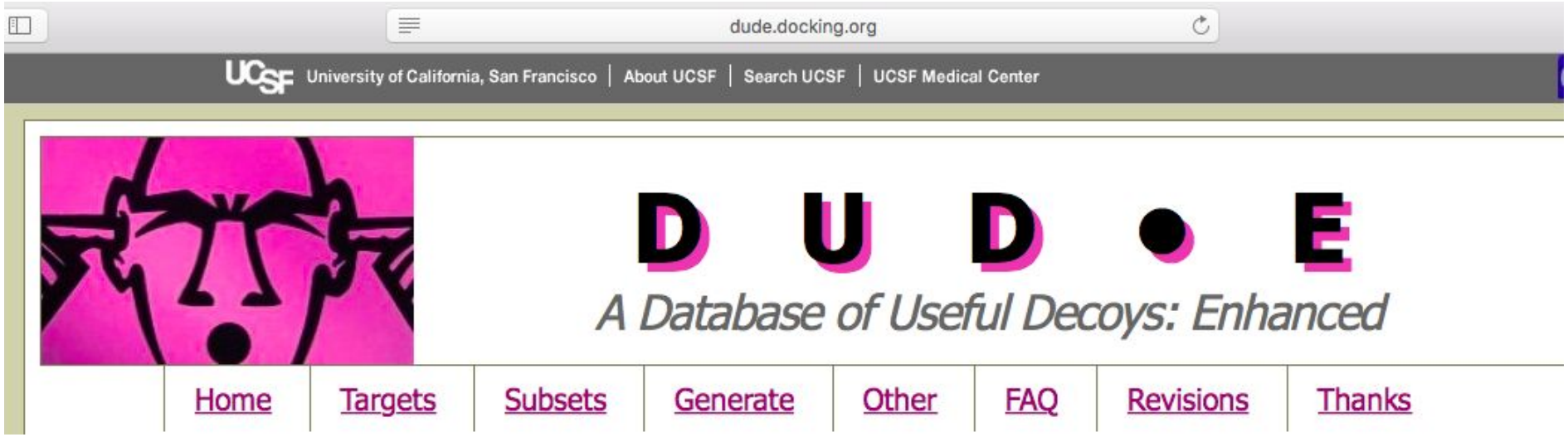
Empirical rank ordering of compounds from docking

- Rank top-ranked pose per compound by a scoring function
- Empirical scoring functions encode physical rational for protein-ligand binding (GlideScore)
  - Lipophilic complementarity rewards
  - Hydrogen bond rewards
  - Metal-ligand rewards
  - Desolvation penalties

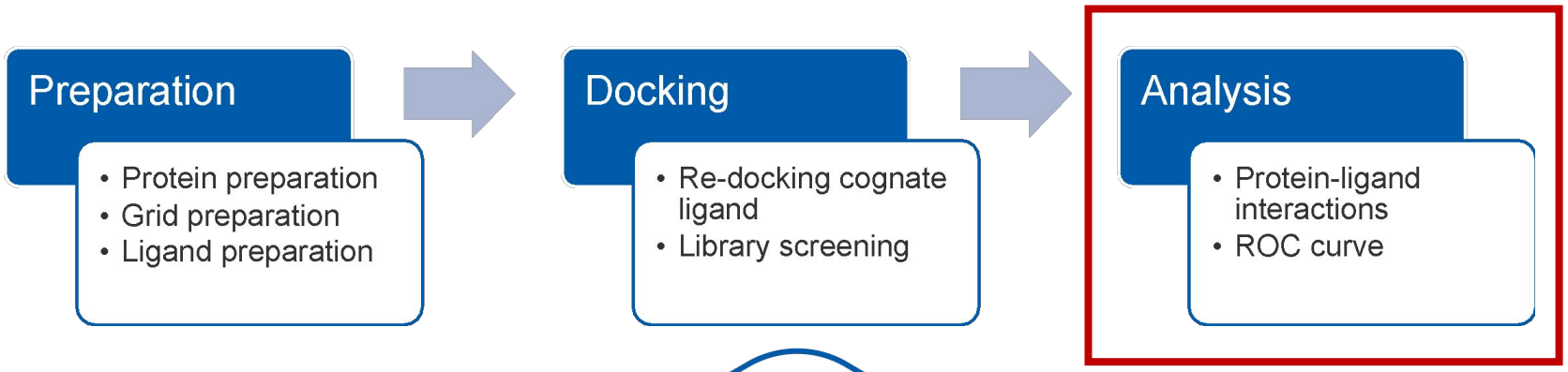


Asses docking using the Directory of Useful Decoys - Enhanced

- Retrospectively verify accuracy of predictions
- DUD-E database **http://dude.docking.org**
- **Actives** are molecules that bind the target and change its function
- **Decoys** are (assumed) non-binders
- **~50 decoys per active having similar physico-chemical properties but dissimilar 2-D topology**
- Today: we will screen 15 actives and 45 decoys, to ensure tutorial finishes in a timely manner.
- Larger screens are recommended

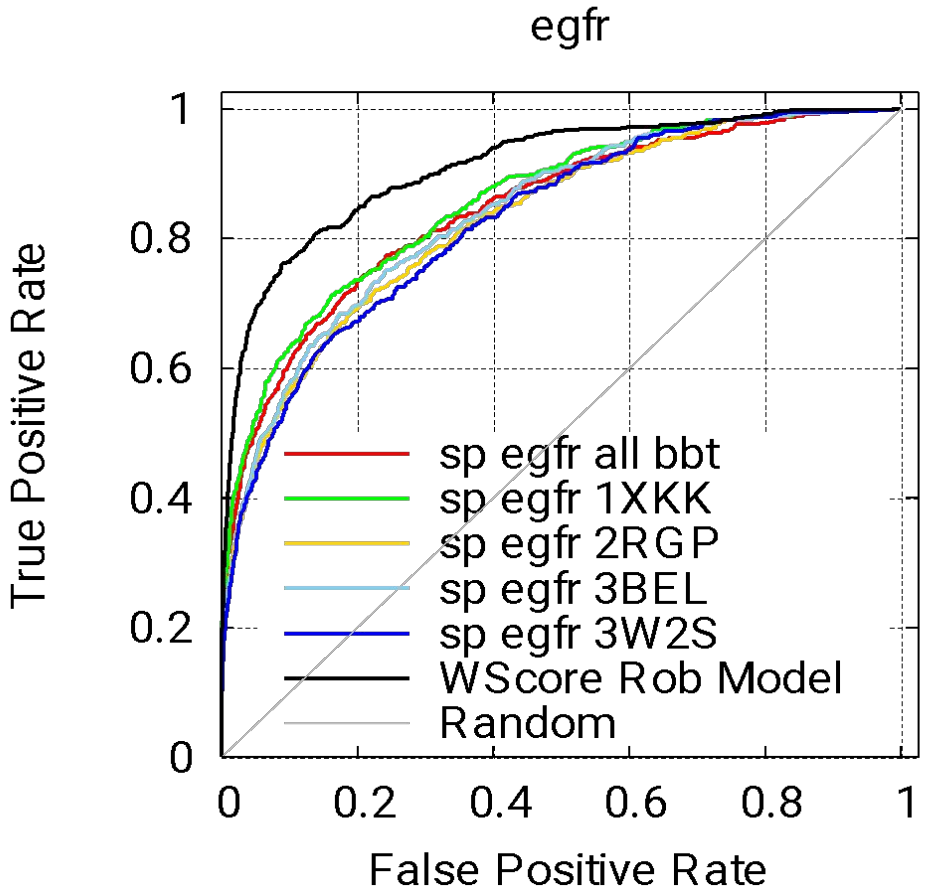


Analysis of docking



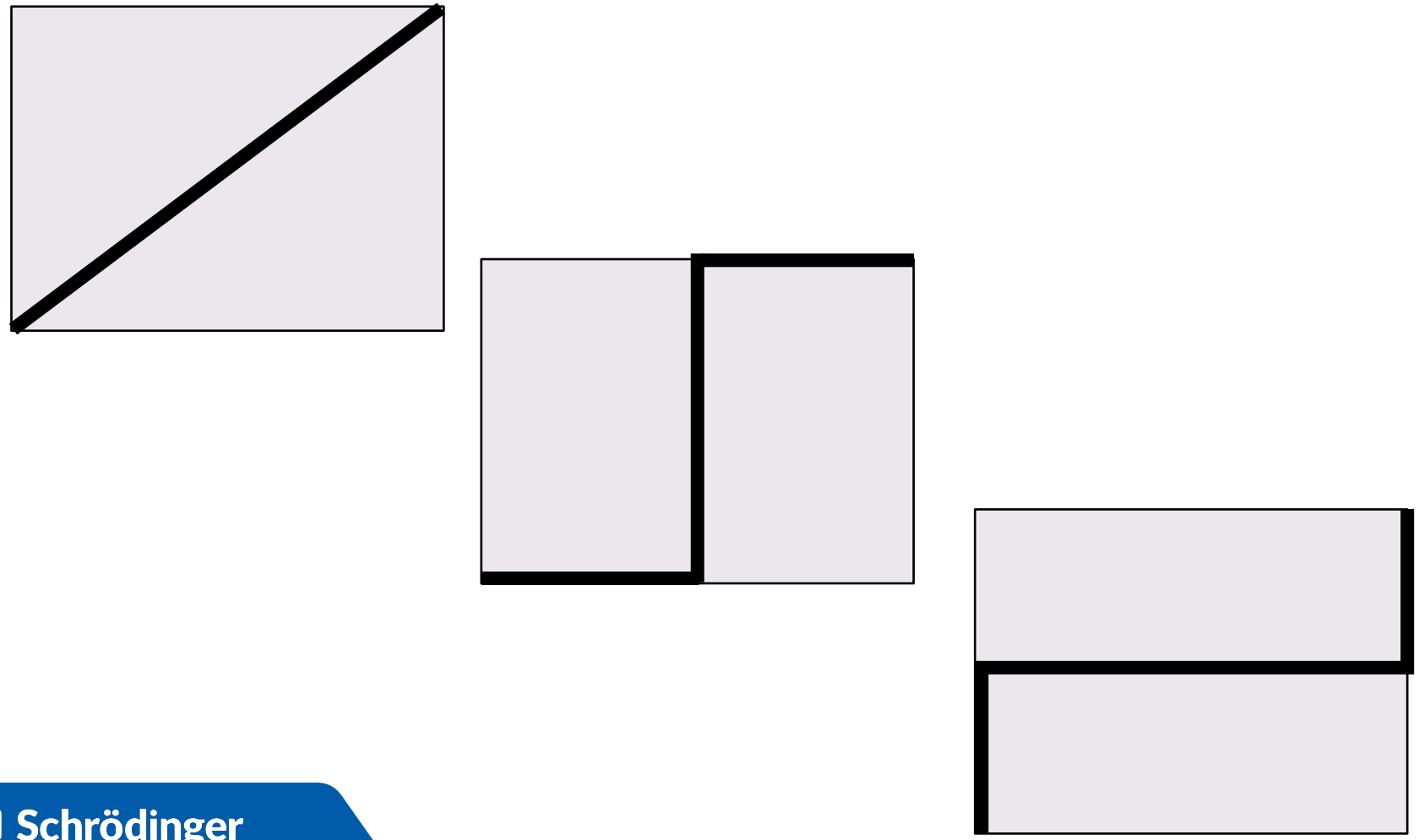
Metrics to assess docking performance (retrospectively)

- Aim of docking is Enrichment – increase the rate of hit discovery
- Aim of docking is NOT to predict accurately binding affinity
- Metrics related to enrichment
  - Receiver operator characteristic (ROC) plots
    - Area under curve metric (AUC)
    - The higher the better
  - Enrichment factor (EF)
    - Focus on predictions of the the highest-ranked ligands
    - The higher the better



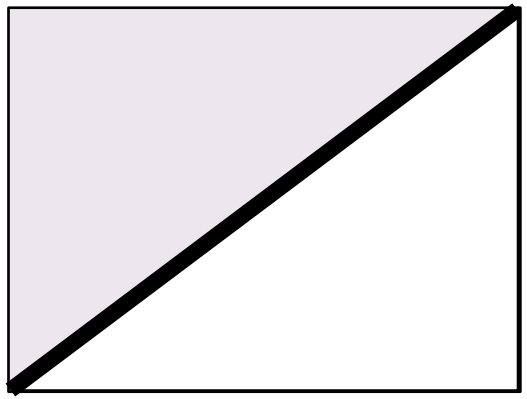
An Aside…Problems with AUC as a Metric

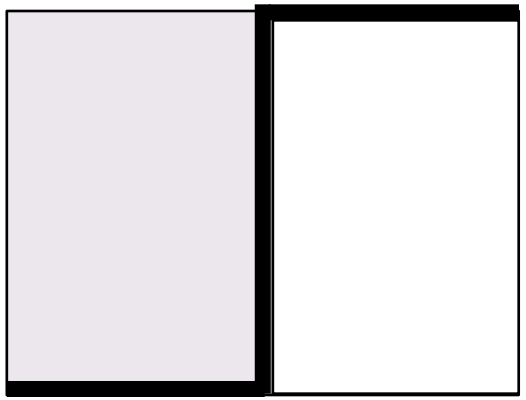
- Which of the below is better?

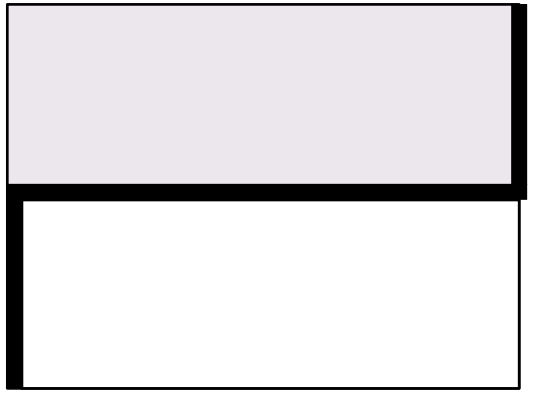


An Aside…Problems with AUC as a Metric

- Which of the below is better?
- All 3 have the same area under the curve

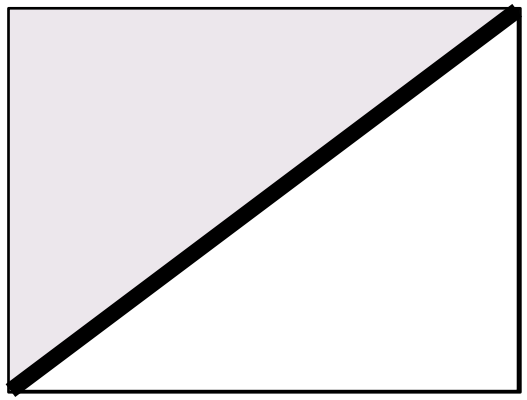




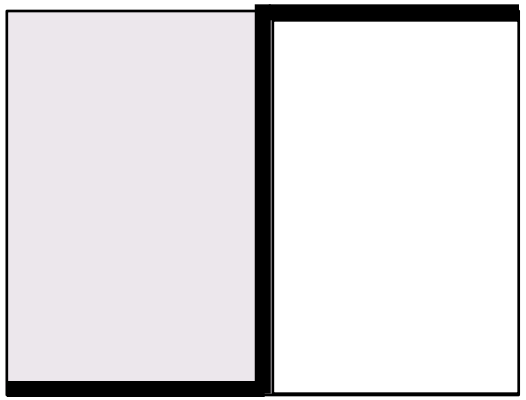


An Aside…Problems with AUC as a Metric

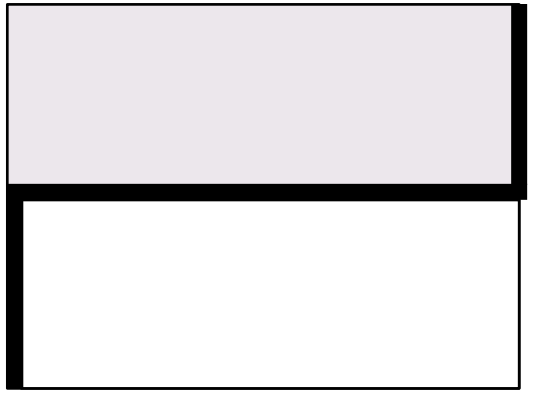
- Which of the below is better?
- All 3 have the same area under the curve



Random



Worthless

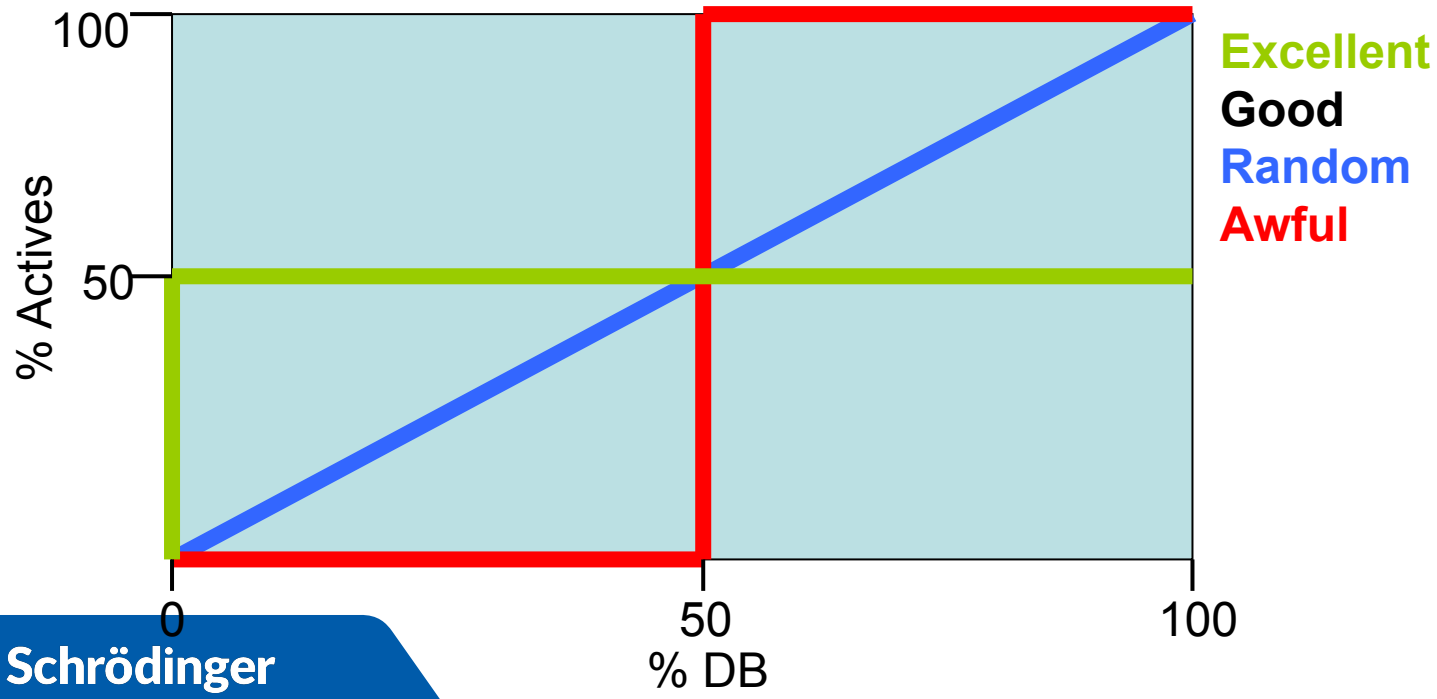


Excellent

Early enrichment is another useful metric

- Green, blue, and red curves all have same area under the ROC curve (AUC)
- Enrichment Factor (1%) would distinguish green as the best

$$Enrichment\ Factor\ (1\%) = \frac{(number\ of\ actives\ in\ top\ 1\%\ of\ ranked\ ligands\ in\ \boldsymbol{docking})}{(number\ of\ actives\ in\ top\ 1\%\ of\ ranked\ ligands\ in\ \boldsymbol{random\ selection})}$$



Summary

- We learned how to prepare a protein and the ligands, and why it is important
- We learned how to generate a protein grid for docking
- We performed docking using Glide
  - Cognate ligand
  - Library screen
- Analyzed the data
  - Protein-ligand interactions
  - RMSD to native pose
  - ROC, area under the curve and Enrichment Factor

- **QUESTIONS?**

Thank you