Texas A&M HPRC Short Course Series

Drug Docking with Schrodinger

Xin Yang

DIVISION OF RESEARCH
TEXAS A&M UNIVERSITY

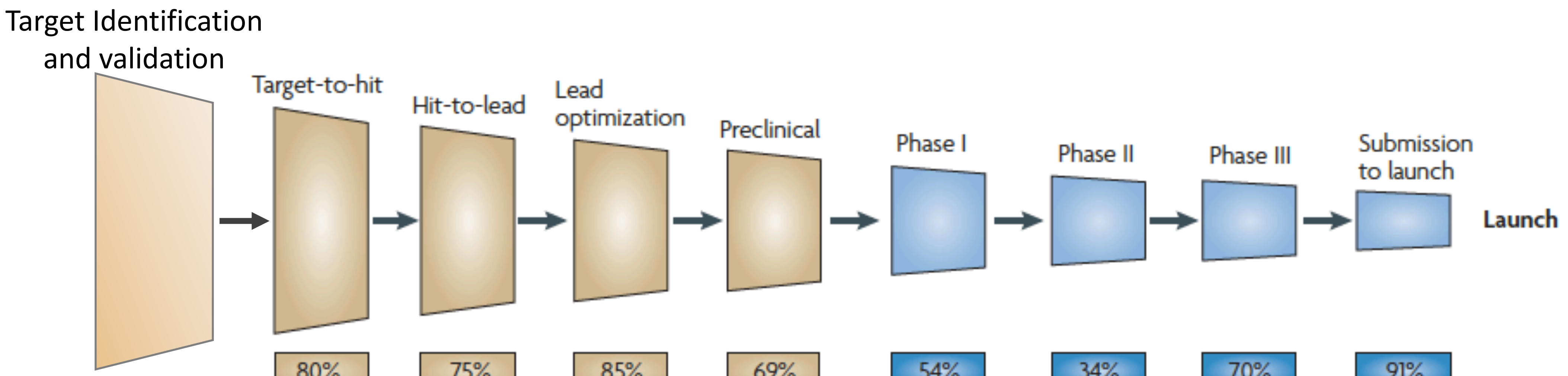
Outline

10:00 -10:20 Intro to Molecular Modeling in Drug Discovery
10:20 -11:10 Hands-on Session 1 – Structure Preparation with Maestro

11:10-11:30 Basics of Structural Based Virtual Screening
11:30-12:15 Hand-on Session 2 – Docking with Glide

12:15-12:30 Covalent Docking with Covdock & Wrap-up

A Tough Road of Drug Discovery

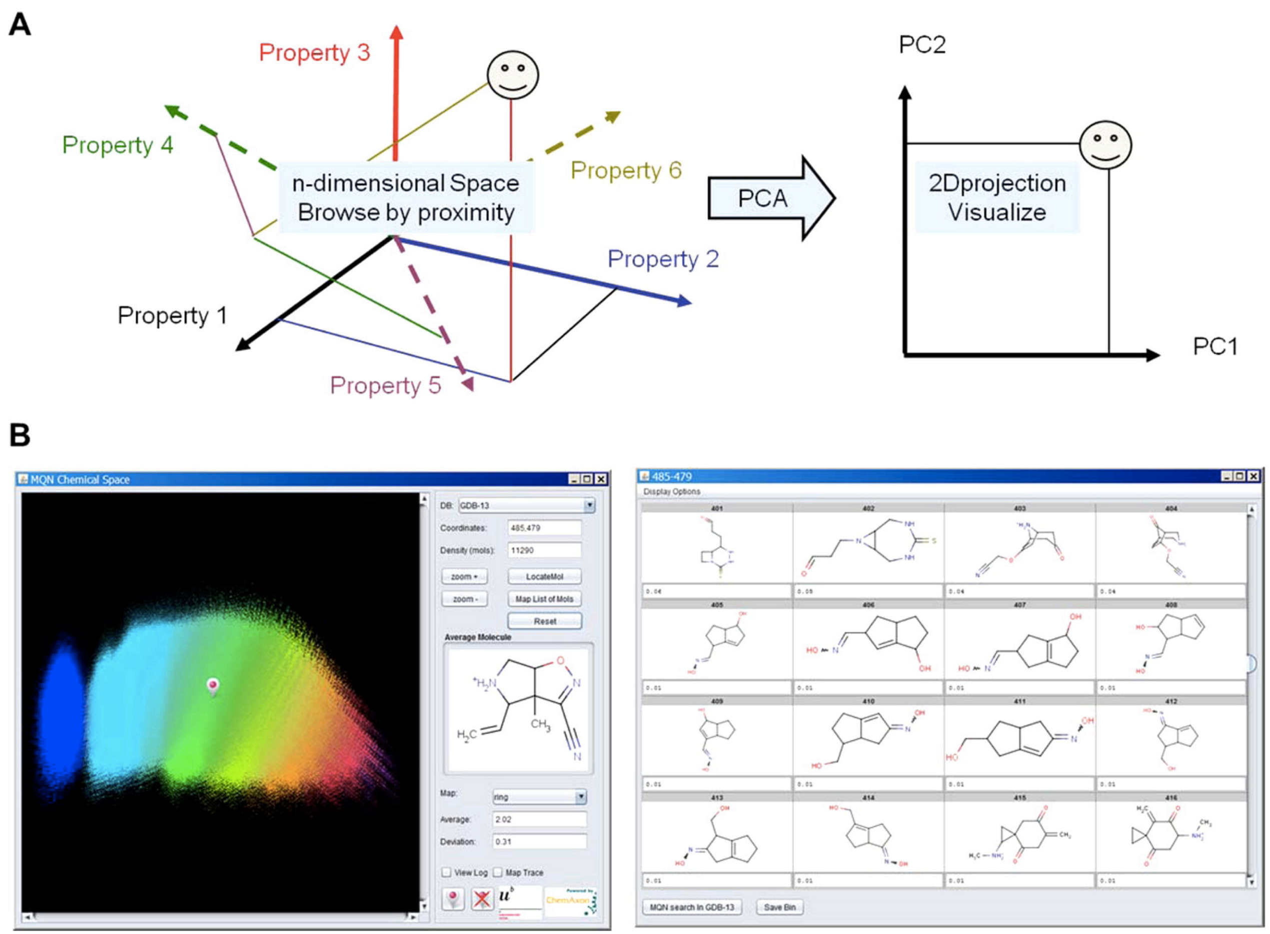


	Target-to-hit	Hit-to-lead	Lead optimization	Preclinical	Phase I	Phase II	Phase III	Submission to launch	Launch
p(TS)	80%	75%	85%	69%	54%	34%	70%	91%	
***WIP* needed for 1 launch**	**24.3**	**19.4**	**14.6**	**12.4**	**8.6**	**4.6**	**1.6**	**1.1**	**1**
Cost per WIP per Phase	$1	$2.5	$10	$5	$15	$40	$150	$40	
Cycle time (years)	1.0	1.5	2.0	1.0	1.5	2.5	2.5	1.5	**13.5 years!!!**
Cost per launch (out of pocket)	**$24**	**$49**	**$146**	**$62**	**$128**	**$185**	**$235**	**$44**	**$873**
% Total cost per NME	3%	6%	17%	7%	15%	21%	27%	5%	
Cost of capital	11%								**1.8 billions!!!**
Cost per launch (capitalized)	**$94**	**$166**	**$414**	**$150**	**$273**	**$319**	**$314**	**$48**	**$1,778**

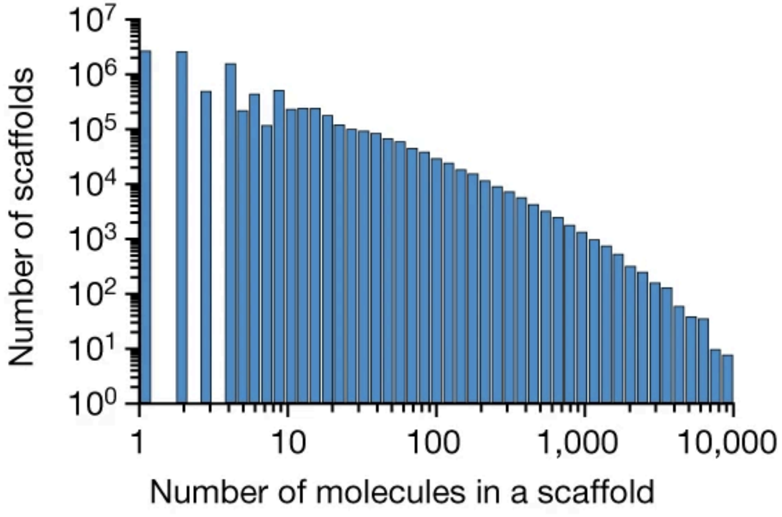
☐ Discovery ☐ Development

Figure adapted from Paul SM et al., Nat. Rev. Drug Discov., 2010, 9, 203-214

CADD - Larger chemical space, new hits



- Chemical diversity (scaffolds) increases with large chemical space searching
- As screening decks expand there will be more tighter binders that could be found

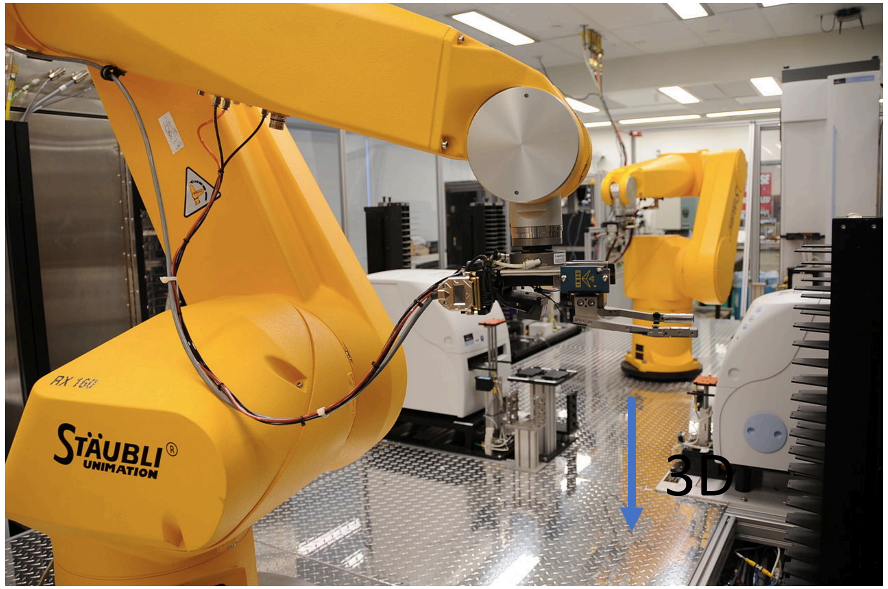


Acc. Chem. Res. 2015, 48, 3, 722–730
Nature 2019, 566, 224–229

Methods for Hit Identification

Hit: A small molecule that is known to bind to a target in drug discovery.

High Throughput Screening (HTS)



https://en.wikipedia.org/wiki/High-throughput_screening

High Throughput Virtual Screening (HTVS)

ligand-based

- 2D Fingerprint searching
- 2D/3D pharmacophore
- 2D/3D QSAR Models

structure-based 3D

- **Docking**
- pharmacophore screening

Virtual Screening

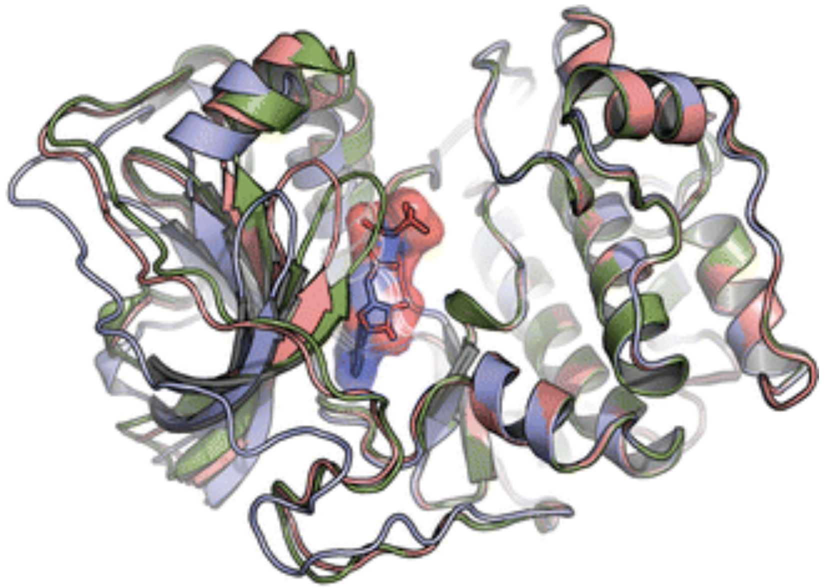


Figure from *J. Am. Chem. Soc.* 2013, 135, 15, 5819–5827

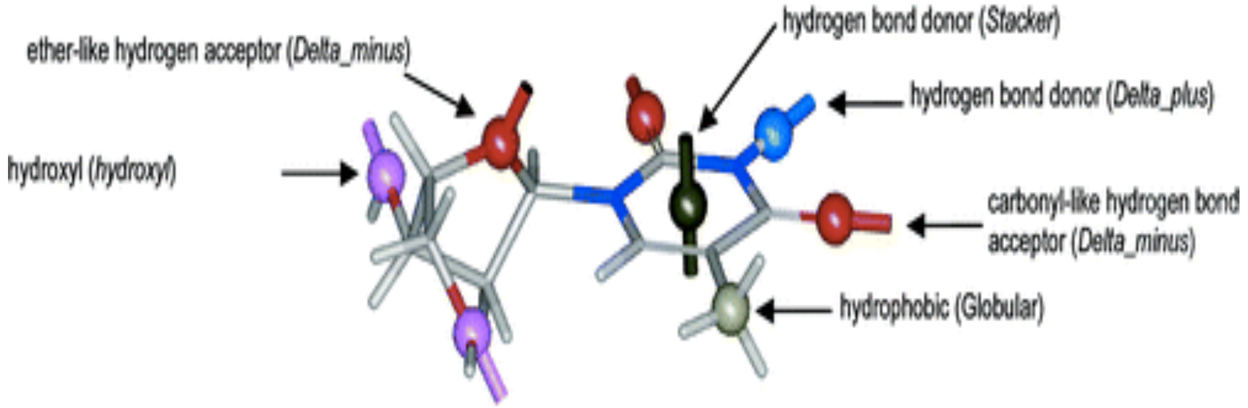


Figure from *J. Chem. Inf. Model.* 2007, 47, 3, 1097–1110

Structure-based Drug Design

- known target structure
- known ligand binding site
- (optional) bound ligands/hits

Ligand-based Drug Design

- known hits
- (optional) active conformation

Protein Target

- Crystal structure

  [RCSB Protein Data Bank (PDB)]

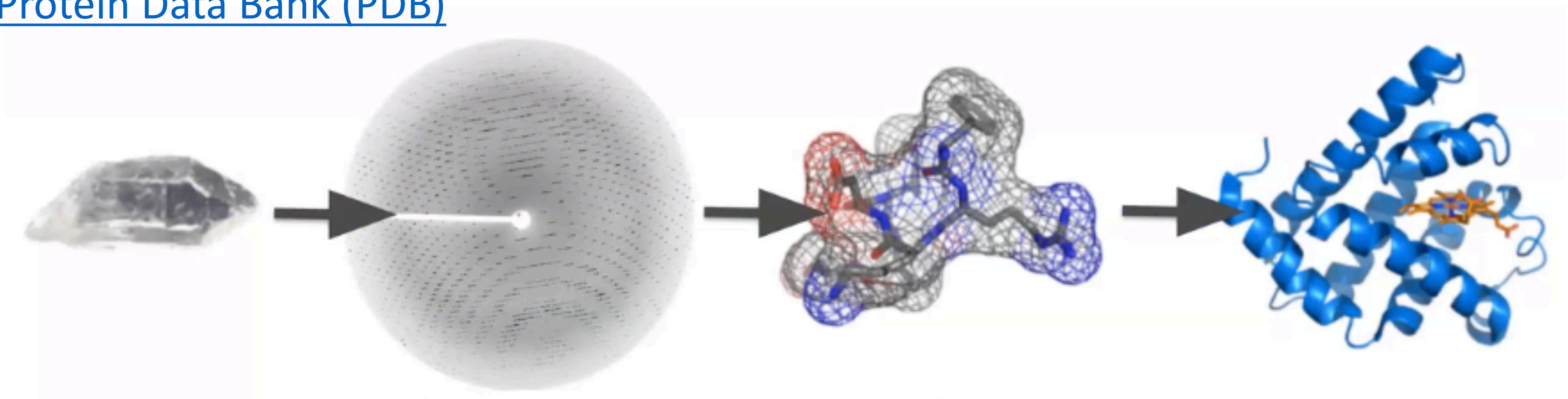
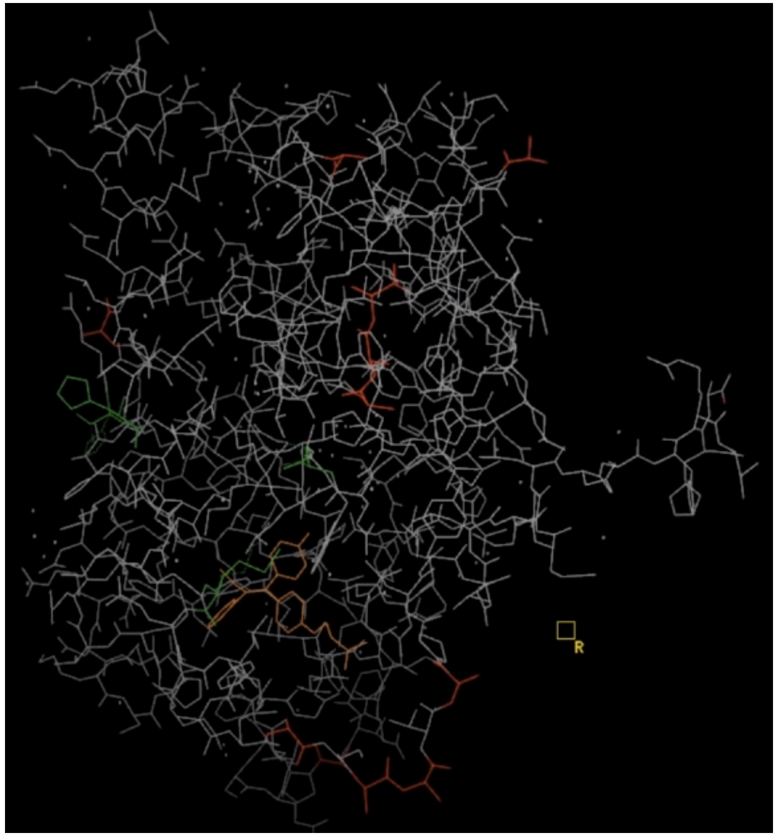


Figure from https://www.schrodinger.com/webinars/archives/1248/virtual-screening/469153

- NMR
- Homology Model
- Cryoelectron Microscopy (cryo-EM)

Protein Preparation

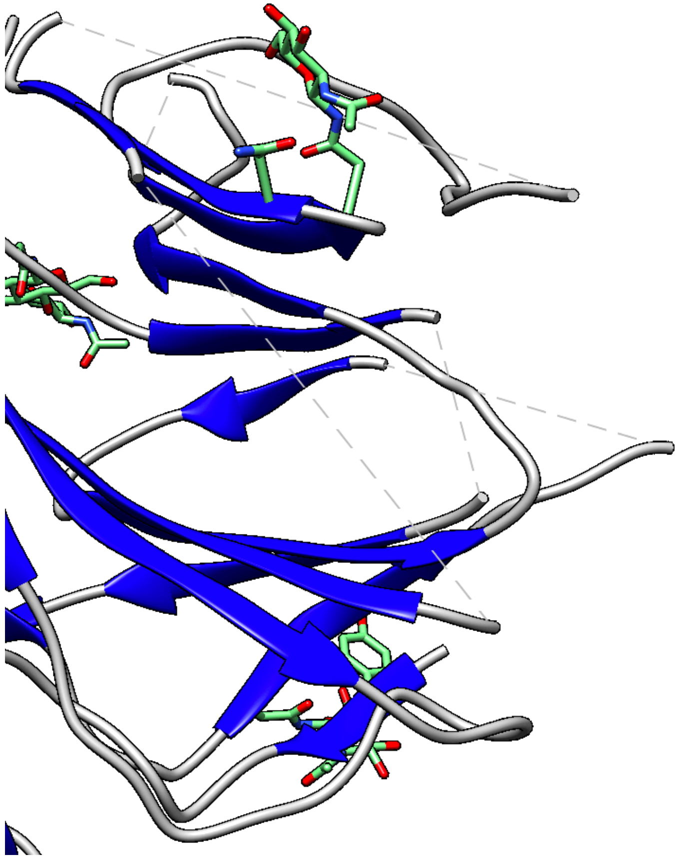
- Typical PDB structure is not suitable for immediate use
  - it typically contains heavy atoms, co-crystalized ligand, water molecules, metal ions, cofactors, …
  - may be multimeric, need to be reduced to a single unit
  - limited resolution, eg. it’s difficult to distinguish carbonyl oxygen and secondary amine nitrogen’s of amide
  - may have incorrect bond orders, assignment of charge state, orientation of groups



Color protein with **PDB conversion Status**
Gray: standard residue connect by standard
Red: standard residue with missing atoms
orange: nonstandard residues, HET groups
green: residue with an alternate conformation

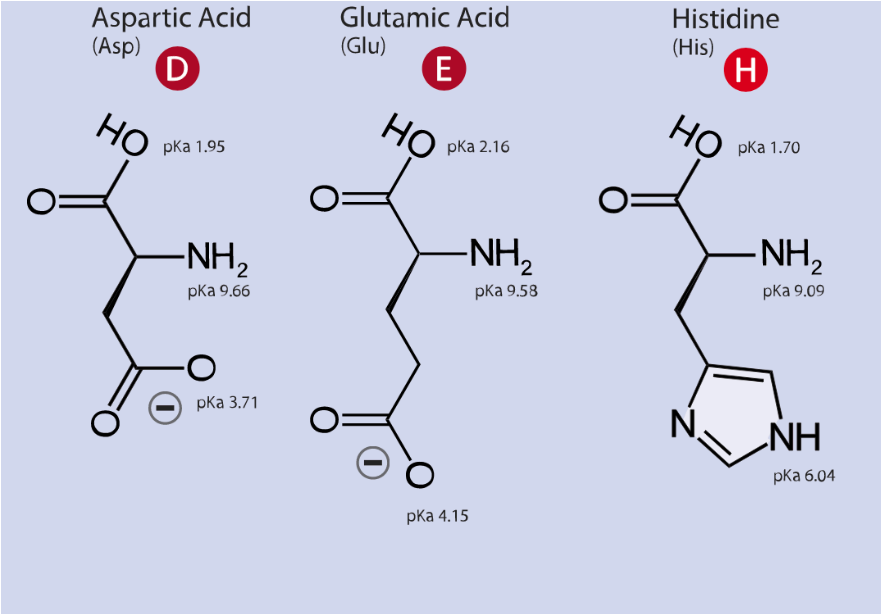
Protein Preparation – Missing atoms

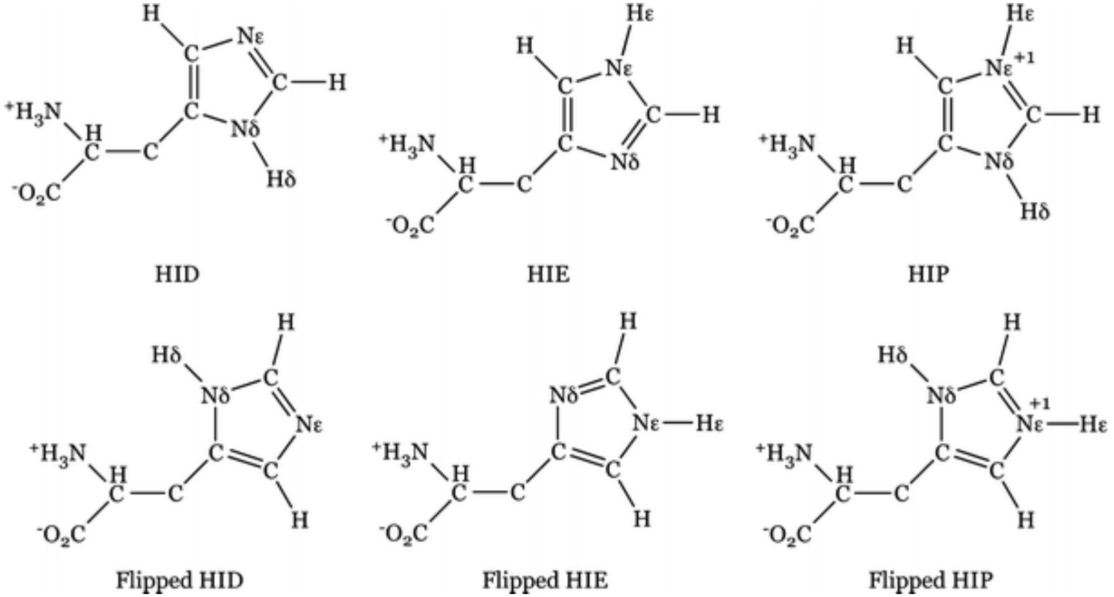
- Missing atoms
  - Hydrogens are not included
  - Entire side chains may be missing
  - There are a number of utilities to fill in missing atoms/sidechain
- Missing segments
  - More complicated to fix
  - Normally requires homology modeling to obtain reasonable results if more than a few residues are missing



Protein Preparation – Protonation states

- ASP, GLU and HIS





Adapted from https://commons.wikimedia.org/wiki/File:Amino_Acids.svg
Dancojocari / CC BY-SA (https://creativecommons.org/licenses/by-sa/3.0)

https://link.springer.com/article/10.1007/s10822-013-9643-9

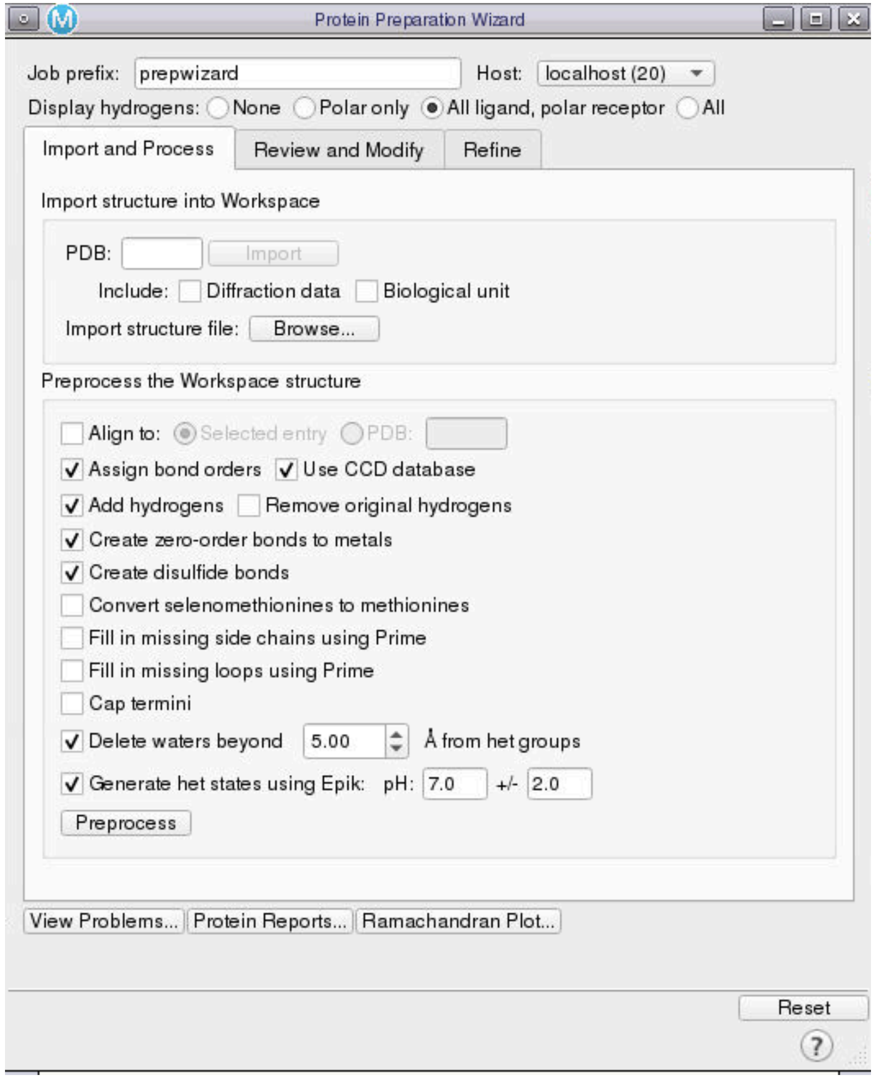
Protein Preparation Wizard in Schrodinger

- **Import and Process Tab:** fix common problems
  - Protonation
  - Missing Side Chain
  - Missing Loops

- **Review and Modify:** Remove Unwanted Molecule
  - counterions, artifact of crystallography, waters
  - Biologically relevant

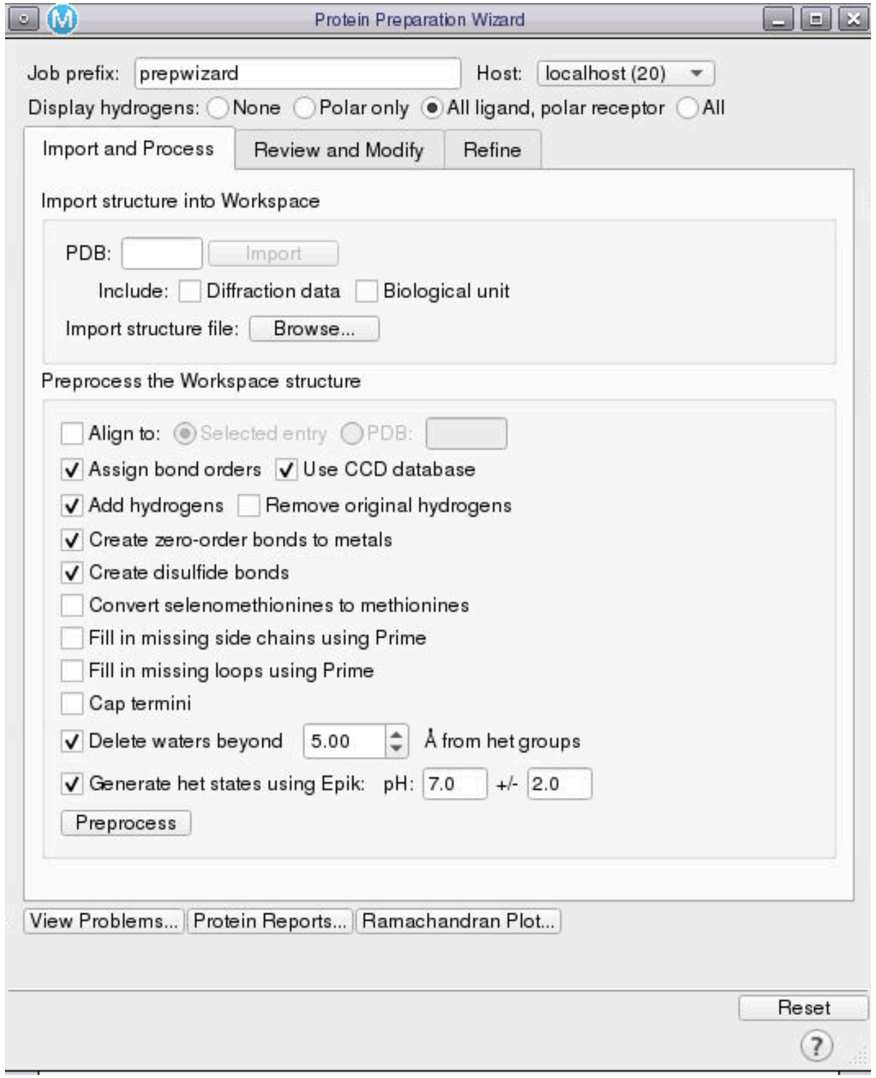
- **Refine:** Optimize your structure
  - Hydrogen bonded optimization
  - Remove waters?
  - Restrained Minimization

- **View Problems…**



Protein Preparation Wizard in Schrodinger

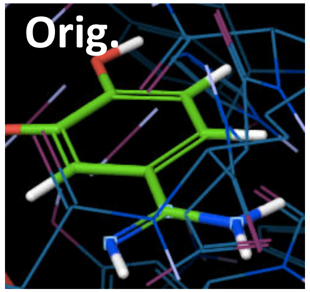
- **Import structure**
  - From RCSB website:
    - Diffraction data: for refining data with Primax
    - Biological unit: merge into a single entry
  - From local PDB files
- **Preprocess options**
  - Align one protein to another protein
  - Correct metal ionization states to ensure proper formal charge and force field treatment
  - Add sulfur bond between sulfur atoms that are within 3.2 Å of each other
  - Convert selenomethionines to methionine
  - Protein refinement with Prime
  - Cap protein termini with ACE and NMA residues
  - Remove water molecules at the user's discretion

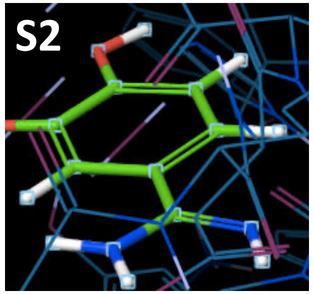


Protein Preparation Wizard in Schrodinger

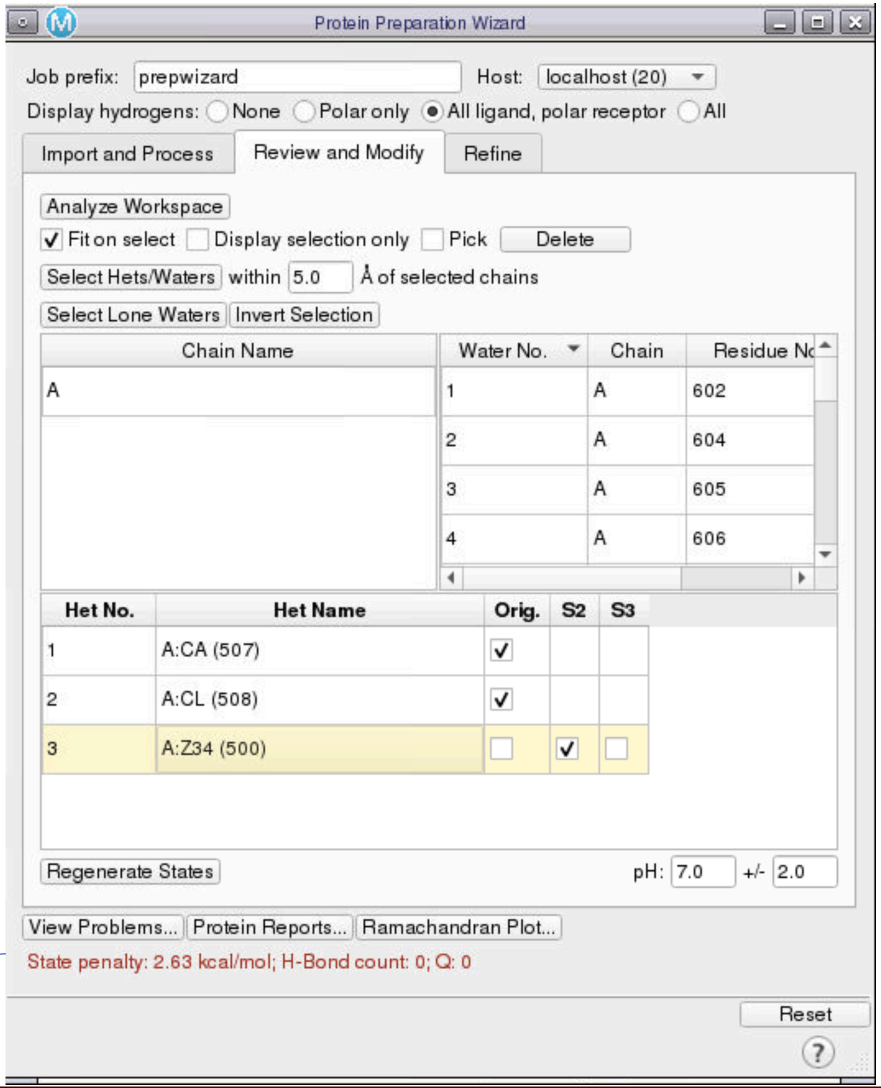
Review and Modify Tab

- **Analyze workspace**
- Delete waters (bulk water, water away from binding site, …)
- Correct the ionization and tautomeric states of listed HET groups
  - Generate States: run an epic job at the target PH range





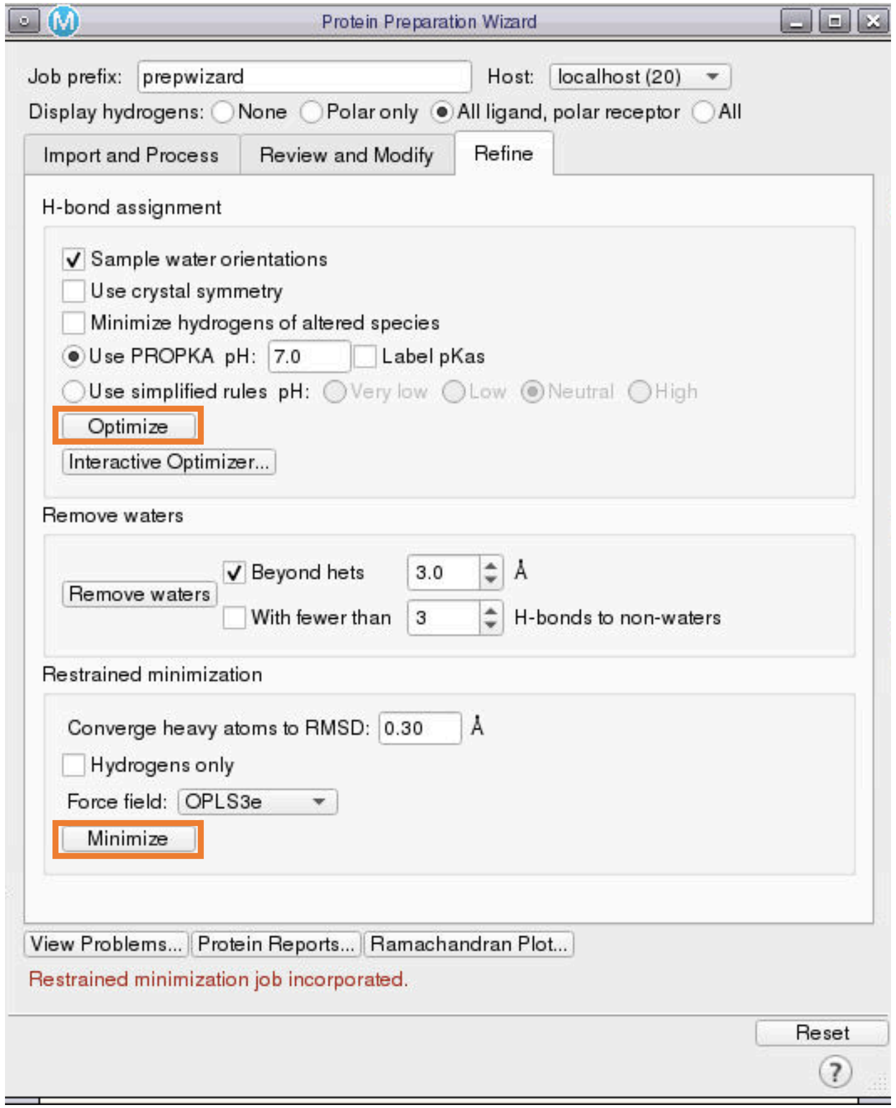
  - Display state penalty



Protein Preparation Wizard in Schrodinger

Refine Tab

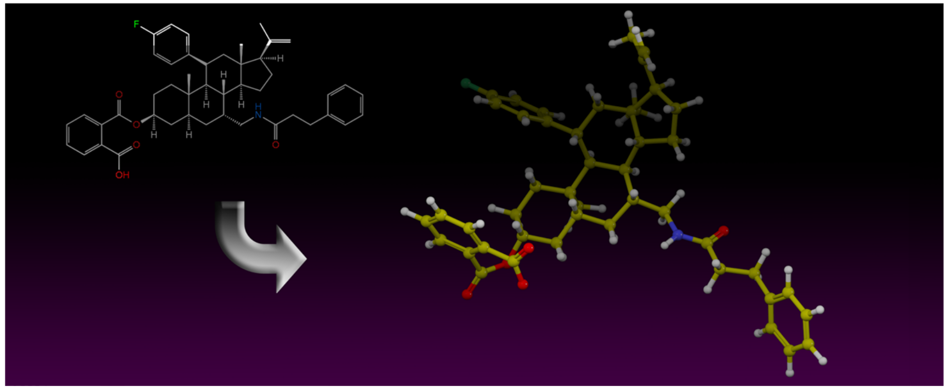
- **Optimizing the H-bonding network by**
  - reorienting water, amide groups, imidazole ring, …
  - Use crystal symmetry: important when part of of structure is present in the asymmetric unit
  - Two options for pH
    - PROPKA
    - simplified rules
      - very low: protonate ASU, GLU, HIS
      - low: protonate HIS
      - neutral: normal biological state
      - high: deprotonate cystines
  - Optimize H-bond interactively
- **Remove waters with less than a specified number of H-bond**
- **Restrained minimization**



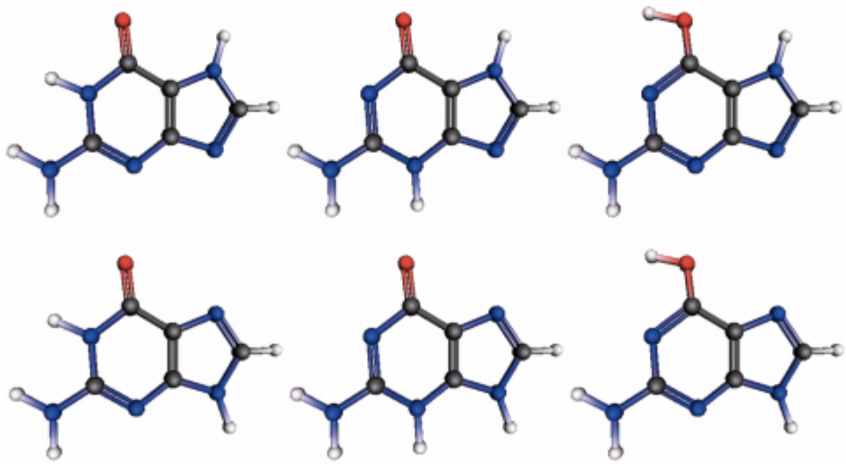
Ligand Preparation

- Take 2D or 3D structures and produce low energy 3D structures
- Generate reasonable atomic coordinates for a ligand dataset

  - tautomeric states
  - ionization states
  - ring conformations
  - stereoisomers
  - conformers



https://www.schrodinger.com/ligprep



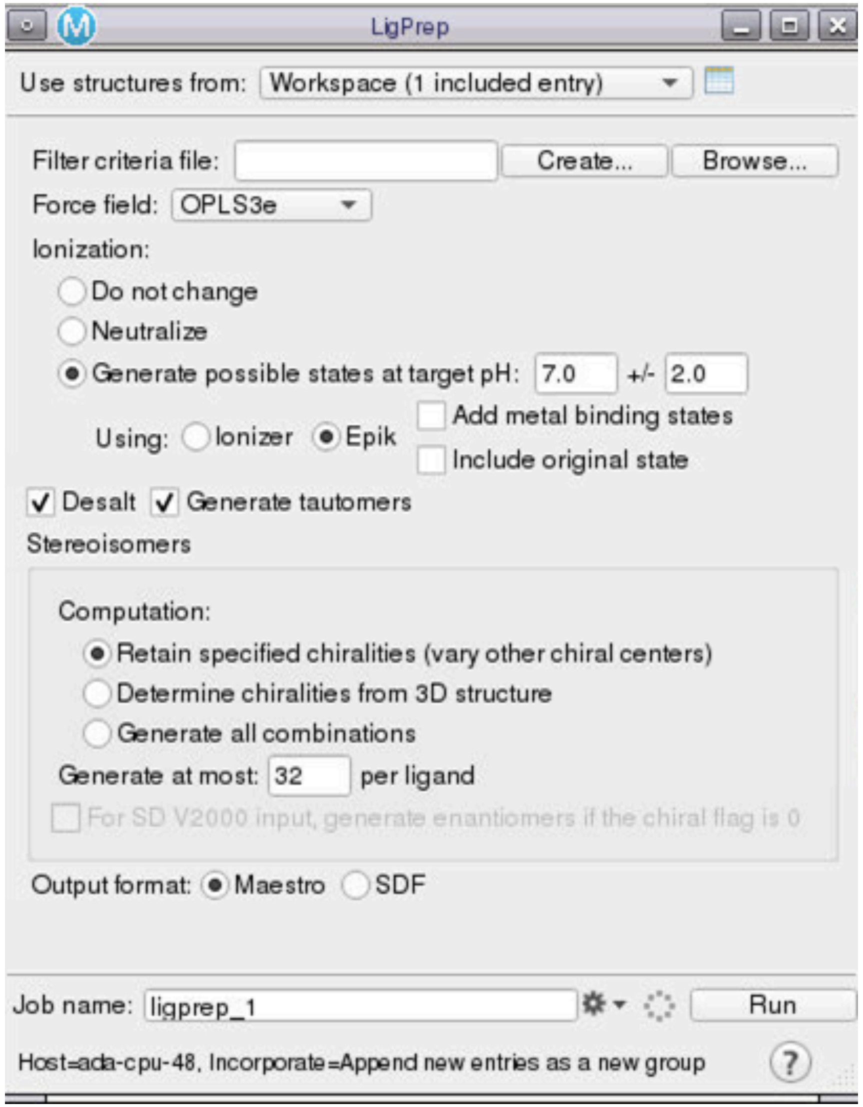
Generation of multiple tautomerics forms of the ring system in a guanine ligand

J. Chem. Inf. Model. 2009, 49, 6, 1535–1546

Ligand Preparation with LigPrep

- **Import structures:** from project, SD, SMILES format ….
- **Filter criteria:**
  - properties
  - general attributes: MW, number of atoms, …
  - functional group counts
- **Force field**
- **Generate ionization states:**
  - Ionizer
  - Epik (recommended)
- **Desalt**: removes extra water molecule or counter ions that are present in ligand files that are originate from some structure databases
- **Generate tautomers**: keto-enol, sulfur/nitrogen, histidine, DNA base tautomerization
- **Stereoisomers**

LigPrep takes about 1-2 seconds on average to process a ligand.
Result in difference in Epik State penalty (kcal/mol)



Hands-on Session 1

Structure Visualization and Preparation with Maestro

1. Creating Projects and Importing Structures
2. Preparing Protein Structures (**Protein Preparation Wizard**)
3. Preparing Ligand Structures (**LigPrep**)
4. Visualizing Protein-Ligand Complexes (configuration bar, Ligand Interaction Diagram)

Steps of structure based virtual screening

Docking	Scoring	Refining
explore poses of ligand in the binding site	quantify the poses with a function	improve poses and select compounds

Docking

- Rigid receptor docking
- Induced fit docking
- Covalent docking
- …

Scoring

- Docking Score
- Glide Score
- Emodel
- …

Refining

- RMSD
- Enrichment
- Receiver operator characteristic plots
- …

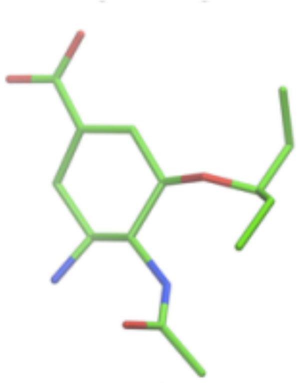
https://www.schrodinger.com/webinars/archives/1248/virtual-screening/469153

https://www.schrodinger.com/

Docking fits ligands to a rigid receptor in a pose

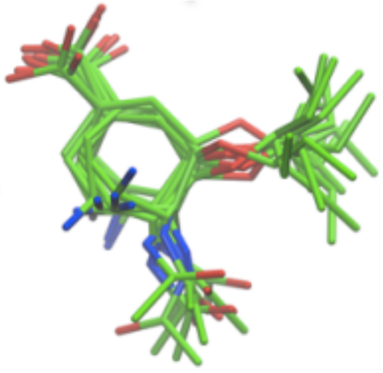
Docking

Search for the best-scoring binding pose for a given ligand
Rigid receptor docking with Glide HTVS, SP, XP
Receptor is rigid
Ligand is flexible



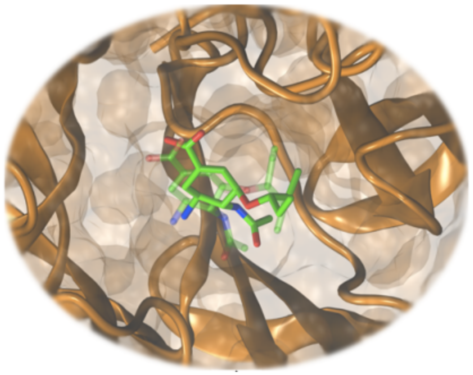
prepared ligand

Ligand sampling →



collection of ligand conformations

docking →



ensemble of receptor-ligand poses

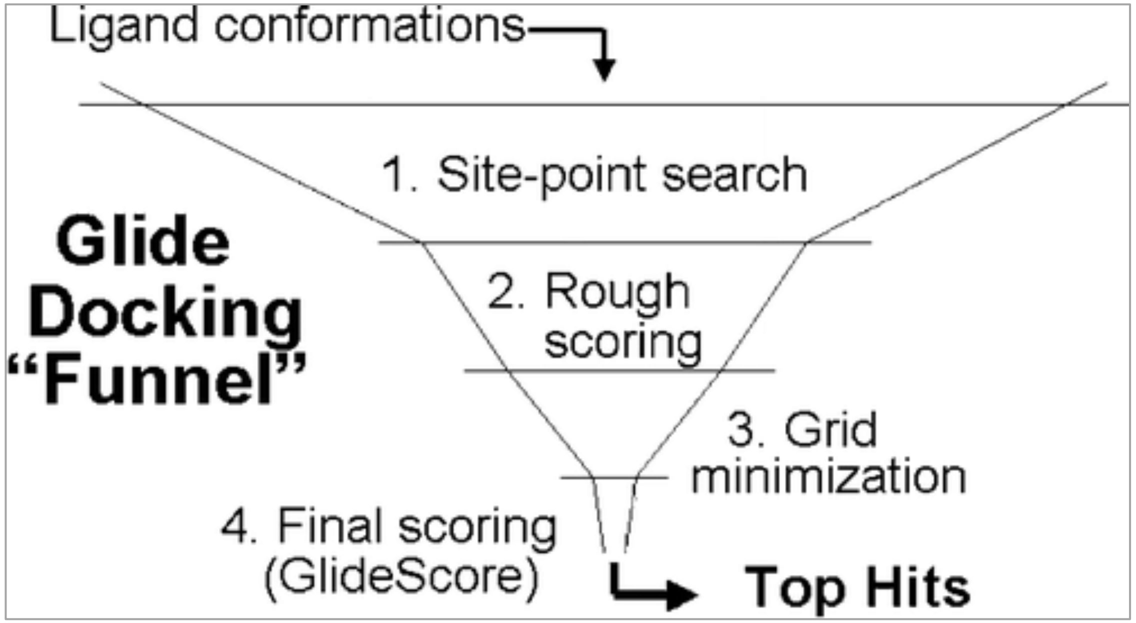
Figures from *J. Chem. Inf. Model.* 2013, 53, 11, 3097–3112

https://www.schrodinger.com/

Ligand Docking

- Procedure
  - Prepare the protein
    - Missing atoms/side chains
    - Protonation state
    - Flexible side chains
  - Prepare the ligand
    - Protonation state
  - Create a docking grid
    - Specify where to dock the ligand
  - Dock the ligand(s)
  - Scoring
  - Refinement

- Glide docking hierarchy

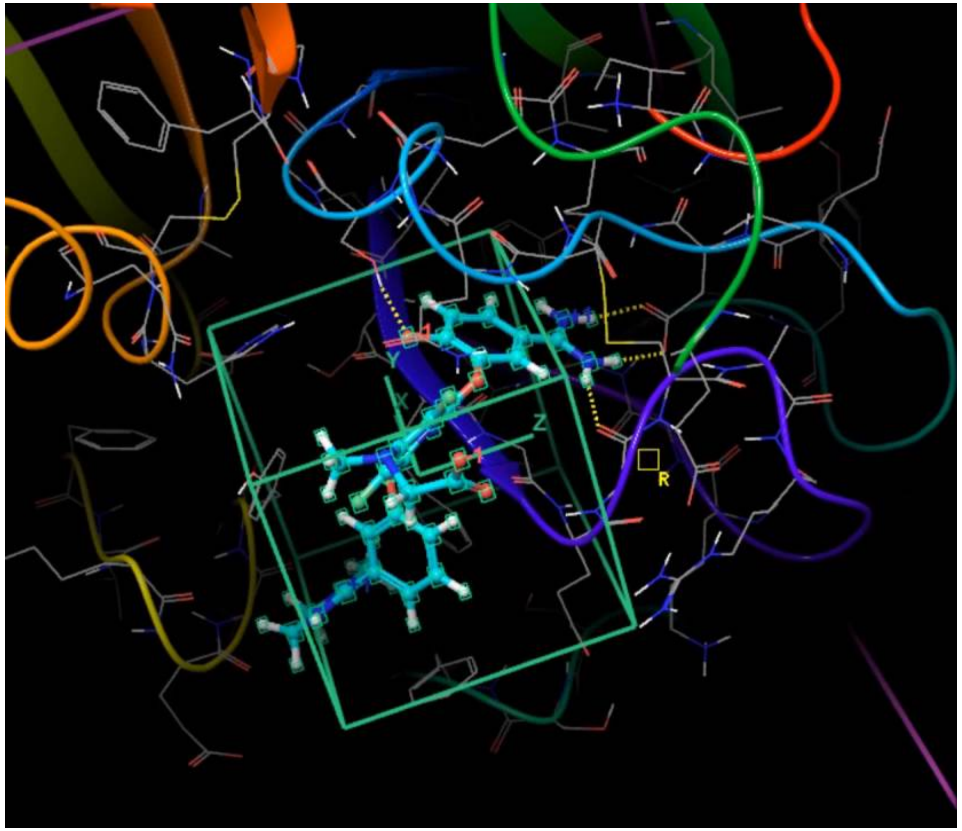


Glide docking "funnel", showing the Glide docking hierarchy.

J. Med. Chem. 2004, 47, 7, 1739–1749

Binding Pocket – Grid Generation

- Utilities to suggest binding sites – such as Schrödinger's SiteMap
- Use binding site from crystal structures with a bound ligand (cognate ligand)
- Binding Pocket Grid
  - Bounding box where docking is performed
  - Too small
    - ligands won't dock
    - miss good ligands
  - Too big
    - increase computational cost substantially
    - miss good binding poses
- Is the binding pocket rigid or flexible?
  - Molecular dynamics simulations can be used to investigate the stability of the binding pocket



Steps of structure based virtual screening

Docking → **Scoring** → Refining

A scoring function very **roughly** approximates the **binding affinity** of a ligand to a protein given a binding pose.

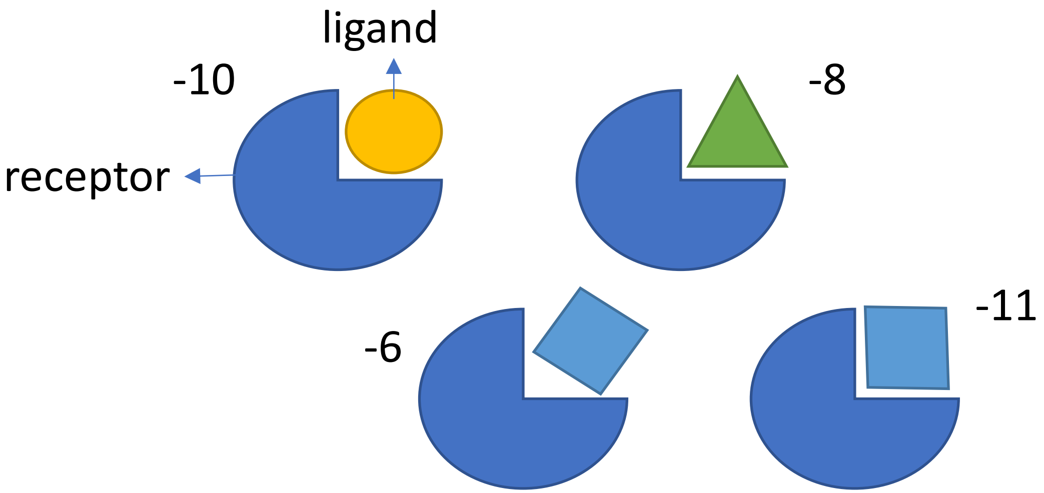


Illustration of binding pose ensembles

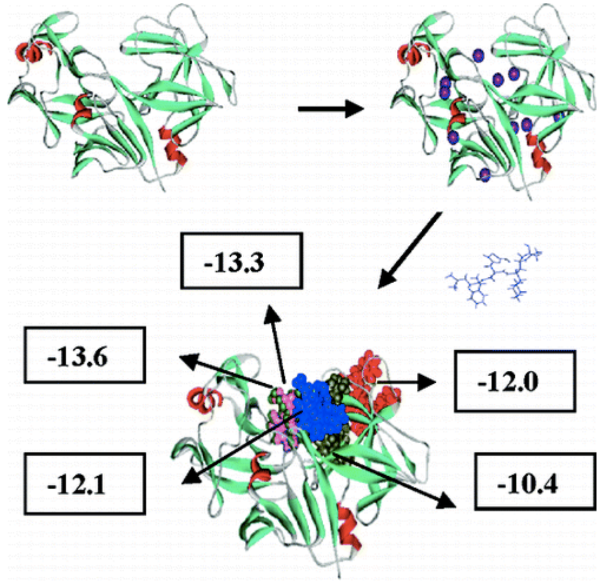


Figure from *J. Chem. Inf. Model.* 2011, 51, 10, 2515–2527

Scoring evaluates the ligand pose

Docking → **Scoring** → Refining

- Do not correlate with IC_{50}, K_d, EC_{50}, etc
- More negative the score, the better
- Are optimized to give good enrichment
  - Separate good from bad ligands
  - Limit the number of ligands that need to be investigated further

https://www.schrodinger.com/webinars/archives/1248/virtual-screening/469153

GlideScore and Emodel

- Glidescore

  rank-order compounds to separate compounds that bind strongly (actives) from those that don't (inactives)

Scoring Function	Computing Time	When to Use
SP	5 – 20 sec/molecule	First pass virtual screening on large databases/hit generation
XP	3-5 min/molecule	Refinement of a smaller dataset for lead optimization

- SP seeks to minimize false negatives while XP seeks to minimize false positives
- The XP scoring function includes more stringent terms for modeling desolvation, hydrophobic effects, and charged interactions

- Emodel
  - primarily defined by protein-ligand coulomb-vdW energy with a small contribution from GlideScore
  - Choose the best-docked structure for each ligand

J. Med. Chem. 2004, 47, 1739-49.

Glide Docking SP

GScore = 0.05*vdW + 0.15*Coul + Lipo + Hbond + Metal + Rewards + RotB + Site

Components	Description
VdW	Van der Waals energy. This term is calculated with reduced net ionic charges on groups with formal charges, such as metals, carboxylates, and guanidiniums.
Coul	Coulomb energy. This term is calculated with reduced net ionic charges on groups with formal charges, such as metals, carboxylates, and guanidiniums.
Lipo	Lipophilic term, which is a pairwise term in SP but is derived from the hydrophobic grid potential for XP. Rewards favorable hydrophobic interactions.
HBond	Hydrogen-bonding term. This term is separated into differently weighted components that depend on whether the donor and acceptor are neutral, one is neutral and the other is charged, or both are charged.
Metal	Metal-binding term. Only the interactions with anionic or highly polar acceptor atoms are included. If the net metal charge in the apo protein is positive, the preference for anionic or polar ligands is included; if the net charge is zero, the preference is suppressed.
Rewards	Rewards and penalties for various features, such as buried polar groups, hydrophobic enclosure, correlated hydrogen bonds, amide twists, and so on. This category covers all terms other than those explicitly mentioned.
RotB	Penalty for freezing rotatable bonds.
Site	Polar interactions in the active site. Polar but non-hydrogen-bonding atoms in a hydrophobic region are rewarded.

Glide User Manual, Schrödinger Software Release 2020-3

Glide Docking XP (Extra Precision)

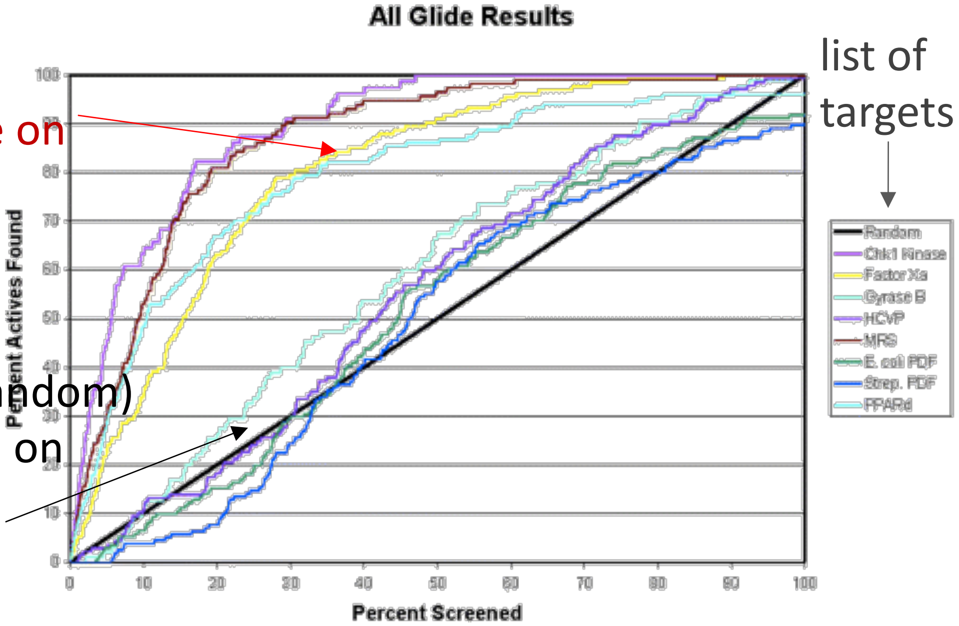
- Increase computational cost
- Glide SP with additional Extra Precision terms
- Anchor fragments of the docked ligand, typically rings, are chosen from the set of SP poses and the molecule is re-grown bond by bond from these anchor positions
- Rewards occupancy of well-defined hydrophobic pockets by hydrophobic ligand groups which is often under-estimated
- Includes improvements to the scoring of hydrogen bonds as well as detection of buried polar groups, and detection of pi-cation and pi-pi stacking interactions

Glide User Manual, Schrödinger Software Release 2020-3

Filtering refines the ligand evaluation

Docking → Scoring → Refining

- Docking performance



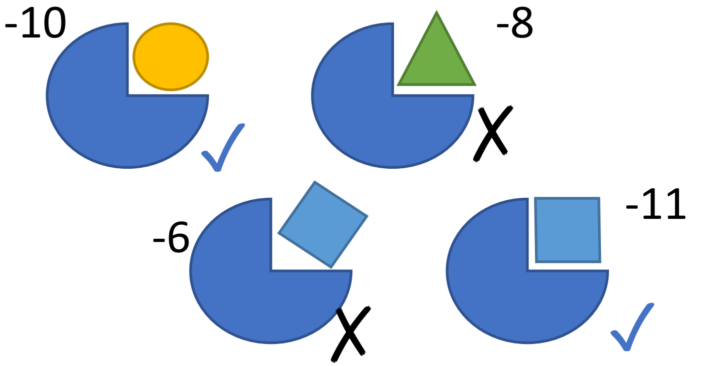
Good performance on these target

Poor (near random) performance on these target

list of targets

J. Med. Chem. 2006, 49, 20, 5912–5931

- Screening compounds for further evaluation



✓ evaluate with higher-level of simulations

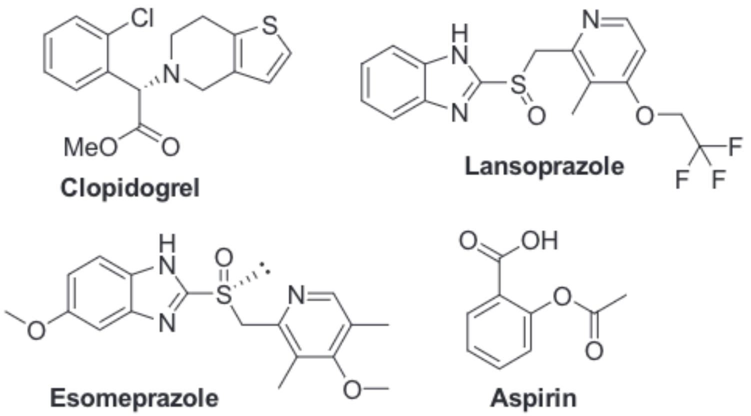
Hands-on Session 2

Structure-Based Virtual Screening Using Glide

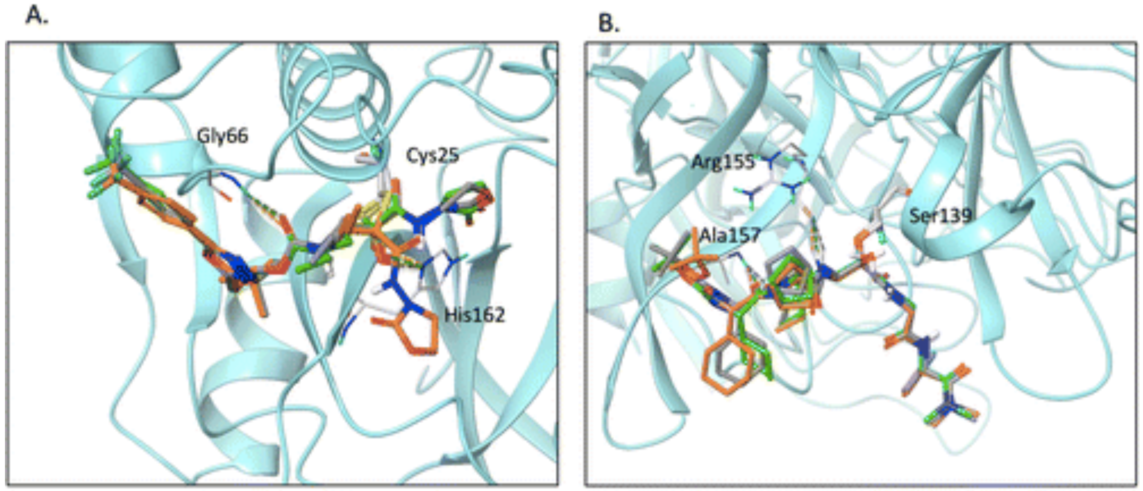
1. Virtual Screening Prerequisites
2. Importing Structures
3. Generating a Receptor Grid
4. Docking the Cognate Ligand and Screening Compounds
5. Analyzing Results and Binding-Site Characterization

Covalent Docking

- Nearly 30% of the marketed drugs targeting enzymes known to act by covalent inhibition
- The inhibition can be either reversible or irreversible
- Covalent inhibitors derive their activity not only from the formation of a covalent bond between the target and the ligand but also from stabilizing non-covalent forces in the binding pocket



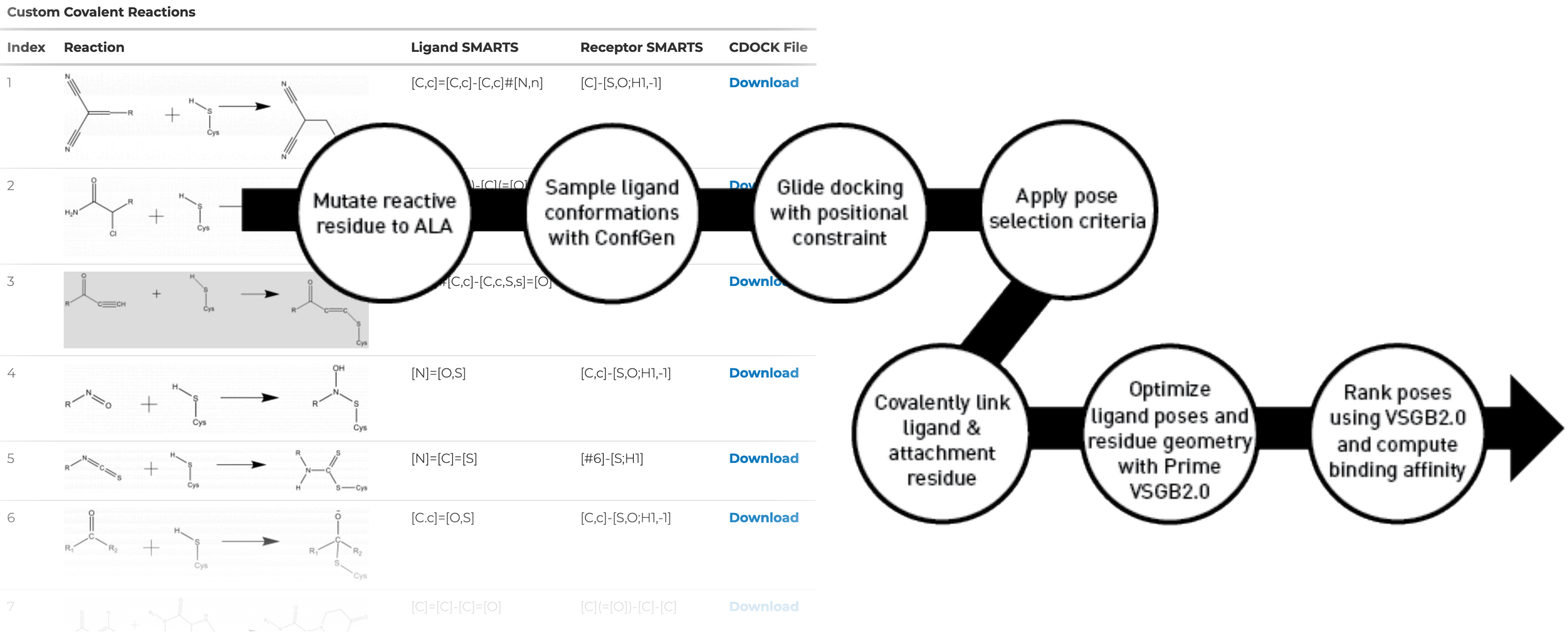
Examples of drug act through covalent mechanisms.



Examples of covalent complexes. A) Cathepsin K structure (PDB ID 1YT7) with the cocrystal ligand, B) HCV NS3 protease structure (PDB ID 2F9U) with the cocrystal ligand.

Eur. J. Med. Chem. **138**, 96–114 (2017).
J. Chem. Inf. Model. 2014, 54, 1941−1950

Covalent Docking



https://www.schrodinger.com/newsletters/introducing-covdock-covalent-docking

CovDock Uses Glide & Prime

- Main steps
  - Conventional non-covalent docking of pre-reactive species **(Glide)**
  - Formation of covalent attachment (via a number of different mechanisms)
  - Structural refinement of the covalent complex **(Prime)**

- Output: **cdock affinity, prime energy, ligand reaction site**

- Speed
  - Pose selection (default) protocol: 1~2 hour per ligand
  - Virtual screening protocol: 10x faster than default protocol

- Challenges for covalent docking
  - Bond formation, bond cleavage and bond rearrangements all require an explicit treatment of electronic degrees of freedom and, hence, a quantum mechanics (QM) approach.

Drug Docking with Schrodinger: wrap-up

Compound Library Preparation

- Obtain Compound Library
- Filter Library
- LigPrep Library

Receptor Grid Preparation

- Prepare protein
- Generate Grids
- Test Grids

Virtual Screening

Post Processing

Compile Virtual Screening Result

↓

First Cut

↓

Secondary Scoring

↓

Second Cut

↓

Visual Inspection

↓

Final Cut

↓

Generate Buy List

Running Schrodinger on HPRC

Schrödinger is a restricted software.

Usage of this software is restricted to subscribers of the Laboratory for Molecular Simulation (LMS).

Running Schrodinger on Ada and Terra, please refer to: https://hprc.tamu.edu/wiki/SW:Schrodinger.

More about the Schrodinger: documentation, training

The LMS also holds license for:

- Discovery Studio
- MOE
- Amber
- Material Studio
- Gaussian
- ADF
- Molpro
- Chemissian
- NBO
- AIMALL Professional

Need Help? Contact the HPRC Helpdesk

Website:	hprc.tamu.edu
Email:	help@hprc.tamu.edu
Telephone:	(979) 845-0219

Help us, help you -- we need more info

- Which Cluster (Terra, Ada)
- NetID (NOT your UIN)
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used if any
- Module(s) loaded if any
- Error messages
- Steps you have taken, so we can reproduce the problem