Texas A&M HPRC Short Course Series
Drug Docking with Schrodinger

Xin Yang
Outline

10:00 - 10:20 Intro to Molecular Modeling in Drug Discovery
10:20 - 11:05 Hands-on Session 1 – Structure Visualization with Maestro

11:05 - 11:25 Basics of Quantum Mechanical Simulation
11:25 - 12:10 Hand-on Session 2 – Docking with Glide

12:10 - 12:30 Covalent Docking with Covdock & Wrap-up
# Cost of Drug Discovery

<table>
<thead>
<tr>
<th>Pharmaceutical company</th>
<th>Number of drugs approved</th>
<th>Average R&amp;D spending per drug (in $ Millions)</th>
<th>Total R&amp;D spending from 1997-2011 (in $ Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AstraZeneca</td>
<td>5</td>
<td>$11,790.93</td>
<td>$58,955</td>
</tr>
<tr>
<td>GlaxoSmithKline</td>
<td>10</td>
<td>$8,170.81</td>
<td>$81,708</td>
</tr>
<tr>
<td>Sanofi</td>
<td>8</td>
<td>$7,909.26</td>
<td>$63,274</td>
</tr>
<tr>
<td>Roche Holding</td>
<td>11</td>
<td>$7,803.77</td>
<td>$85,841</td>
</tr>
<tr>
<td>Pfizer</td>
<td>14</td>
<td>$7,727.03</td>
<td>$108,178</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>15</td>
<td>$5,885.65</td>
<td>$88,285</td>
</tr>
<tr>
<td>Eli Lilly &amp; Co.</td>
<td>11</td>
<td>$4,577.04</td>
<td>$50,347</td>
</tr>
<tr>
<td>Abbott Laboratories</td>
<td>8</td>
<td>$4,496.21</td>
<td>$35,970</td>
</tr>
<tr>
<td>Merck &amp; Co Inc.</td>
<td>16</td>
<td>$4,209.99</td>
<td>$67,360</td>
</tr>
<tr>
<td>Bristol-Meyers Squibb Co.</td>
<td>11</td>
<td>$4,152.26</td>
<td>$45,675</td>
</tr>
<tr>
<td>Novartis</td>
<td>21</td>
<td>$3,983.13</td>
<td>$83,646</td>
</tr>
<tr>
<td>Amgen Inc.</td>
<td>9</td>
<td>$3,692.14</td>
<td>$33,229</td>
</tr>
</tbody>
</table>

A Tough Road of Drug Discovery

Figure adapted from Paul SM et al., Nat. Rev. Drug Discov., 2010, 9, 203-214
R&D productivity in pharmaceutical industry

The number of drugs approved per billion dollars in US R&D spending is slowing over time.

Yearly number of NMEs is decreasing

A new molecular entity (NME) is a drug that contains an active moiety that has never been approved by the FDA or marketed in the US.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>215</td>
<td>207</td>
<td>162</td>
<td>146</td>
</tr>
<tr>
<td>Average per year</td>
<td>43</td>
<td>41</td>
<td>32</td>
<td>29</td>
</tr>
</tbody>
</table>

Source: EFPIA (2010a)

Moreover, only 8% of NMEs will successfully make it from the point of candidate selection (preclinical stage) to launch.

https://www.ohe.org/
Methods for Hit Identification

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

**Hit identification strategies:**

- Known: 43%
- Random Screen: 29%
- SBDD: 14%
- Directed Screen: 1%
- Fragment Screen: 5%
- DEL: 8%

**Target classes:**

- Kinases: 30%
- Other enzymes: 23%
- GPCR’s: 17%
- Epigenetic: 9%

**Disease areas:**

- Oncology: 30%
- CNS/Pain: 18%
- Infection: 14%
- Metabolic: 12%

*J. Med. Chem. 2018, 61, 21, 9442–9468*
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

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 Through-put Screening (HTS)

High Through-put Virtual Screening (HTVS)

- Fingerprint searching
- Shape-based
- Docking
- Pharmacophore screening

ligand-based
- 2D pharmacophore
- 3D pharmacophore

structure-based
- 3D

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

Structure-based Drug Design
- known target structure
- known ligand binding site
- (optional) bound ligands/hits

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

Ligand-based Drug Design
- known hits
- (optional) active conformation

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

• Crystal structure
  RCSB Protein Data Bank (PDB)

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

Figure from https://www.schrodinger.com/webinars/archives/1248/virtual-screening/469153
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
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: use together with OPLS2001 force field
  - Protein refinement with Prime
  - Cap protein termini with ACE and NMA residues
  - Remove water molecules at the user's discretion
Protein Preparation Wizard in Schrödinger

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

```
<table>
<thead>
<tr>
<th>Hot No.</th>
<th>Hot Name</th>
<th>Orig.</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A:CA (597)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A:CL (508)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A:Zn (500)</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
```

Display state penalty: 2.63 kcal/mol; H-Bond count: 3; Q: 0
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 automated optimization
  - 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 tautomeric forms of the ring system in a guanine ligand

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)

User guide for running Schrodinger on Ada and Terra:
https://hprc.tamu.edu/wiki/SW:Schrodinge
Steps of structure based virtual screening

- **Docking**
  - Rigid receptor docking
  - Induced fit docking
  - Covalent docking
  - ...

- **Scoring**
  - Docking Score
  - Glide Score
  - Emodel
  - ...

- **Filtering Assessing**
  - RMSD
  - Enrichment
  - Receiver operator characteristic plots
  - ...

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


https://www.schrodinger.com/
Docking fits ligands to a rigid receptor in a pose

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


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.

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 were 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
- Filtering

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

- 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

GlideScore and Emodel

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

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

- 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

# Glide Docking SP

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

<table>
<thead>
<tr>
<th>Components</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VdW</td>
<td>Van der Waals energy. This term is calculated with reduced net ionic charges on groups with formal charges, such as metals, carboxylates, and guanidiniums.</td>
</tr>
<tr>
<td>Coul</td>
<td>Coulomb energy. This term is calculated with reduced net ionic charges on groups with formal charges, such as metals, carboxylates, and guanidiniums.</td>
</tr>
<tr>
<td>Lipo</td>
<td>Lipophilic term, which is a pairwise term in SP but is derived from the hydrophobic grid potential for XP. Rewards favorable hydrophobic interactions.</td>
</tr>
<tr>
<td>HBond</td>
<td>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.</td>
</tr>
<tr>
<td>Metal</td>
<td>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.</td>
</tr>
<tr>
<td>Rewards</td>
<td>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.</td>
</tr>
<tr>
<td>RotB</td>
<td>Penalty for freezing rotatable bonds.</td>
</tr>
<tr>
<td>Site</td>
<td>Polar interactions in the active site. Polar but non-hydrogen-bonding atoms in a hydrophobic region are rewarded.</td>
</tr>
</tbody>
</table>
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
Filtering refines the ligand evaluation

Docking

• Docking performance
  - Good performance on these targets
  - Poor (near random) performance on these targets

Scoring

Assessing

• Screening compounds for further evaluation

J. Med. Chem. 2006, 49, 20, 5912-5931
Hands-on Session 2

Structure-Based Virtual Screening Using Glide

1. Virtual Screening Prerequisites
2. Creating Projects and 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
- A covalent bond is formed between the target and the inhibitor
- 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 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.
Examples of Bond Formation in Covalent Inhibitors

A. Michael Reactions
–The ligand, L is the electrophile with the Michael acceptor group (O=C=C=C), a Cysteine residue in the receptor R is the nucleophile
• EGFR inhibitor, Afatinib

B. Nucleophilic Addition
–The double bond is the electrophile and a Cysteine residue is the nucleophile
• HCV Protease inhibitor, Boceprevir

C. β-lactam Ring Opening
–The four-membered β-lactam is the electrophile, while the nucleophile is a Serine residue
• Penicillin Binding Protein 1B inhibitor, Penicillin

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
- **More details in the following two Schrödinger papers and documentation**
  - Pose Prediction and Scoring
  - Virtual Screening
  - Covalent Docking User Manual, Schrödinger 2020-3 release
Challenges for Covalent Inhibitor Design

• Safety concern (nonspecific reactivity) of covalent drug
  - Off target binding
  - Undergo metabolism to form highly reactive intermediates may lead to adverse effects such as tissue injury and immune response
  - Pre-existing reactive electrophilic functionality in the parent structure (penicillin)

• Inadequate characterization techniques
  - For irreversible inhibitors in particular, drug activity is governed by reaction kinetics rather than conventional binding thermodynamics
  - Residence time and percentage of receptor occupancy leading to a pharmacological effect

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

Mol. Inf. 2018, 37, 1800062
Running Schrödinger on HPRC

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