THE TRADITIONAL EMPIRICAL PATH TO DRUG DISCOVERY

- Compound library (commercial, in-house, synthetic, natural)
- High throughput screening (HTS)
- Hit confirmation
- Lead compounds (e.g., μM $K_d$)
- Lead optimization (Medicinal chemistry)
- Potent drug candidates (nM $K_d$)
- Animal and clinical evaluation
- Lead optimization (Medicinal chemistry)
- Potent drug candidates (nM $K_d$)
- Animal and clinical evaluation

COMPUTER-AIDED LIGAND DESIGN

Aims to reduce number of compounds synthesized and assayed
- Lower costs
- Reduce chemical waste
- Facilitate faster progress

NEXT UP:
- Overview of structural bioinformatics
  - Major motivations, goals and challenges
- Fundamentals of protein structure
  - Composition, form, forces and dynamics
- Representing and interpreting protein structure
  - Modeling energy as a function of structure
- Example application areas
  - Predicting functional dynamics & drug discovery
Two main approaches:
(1). Receptor/Target-Based
(2). Ligand/Drug-Based

**SCENARIO 1:**
**RECEPTOR-BASED DRUG DISCOVERY**

Structure of Targeted Protein Known: Structure-Based Drug Discovery

HIV Protease/KNI-272 complex

**PROTEIN-LIGAND DOCKING**

Structure-Based Ligand Design

Docking software
Search for structure of lowest energy

Potential function
Energy as a function of structure

- VDW
- Screened Coulombic
- Dihedral
STRUCTURE-BASED VIRTUAL SCREENING

Virtual screening (e.g., computational docking) → Candidate ligands

Ligand optimization → Experimental assay → Drug candidates

Compound database → 3D structure of target (crystallography, NMR, bioinformatics modeling)

Experimental assay and ligand optimization → Drug candidates

Candidate ligands

Virtual screening

FRAGMENTAL STRUCTURE-BASED SCREENING

“Fragment” library → 3D structure of target

Fragment docking → Compound design

Experimental assay and ligand optimization → Drug candidates

Drug candidates


COMPOUND LIBRARIES

Commercial (in-house pharma) → Government (NIH) → Academia

Multiple non active-site pockets identified

Small organic probe fragment affinities map multiple potential binding sites across the structural ensemble.
Ensemble docking & candidate inhibitor testing

Top hits from ensemble docking against distal pockets were tested for inhibitory effects on basal ERK activity in glioblastoma cell lines.

Ensemble computational docking

Compound effect on U251 cell line

Ensemble docking & candidate inhibitor testing

Proteins and Ligand are Flexible

COMMON SIMPLIFICATIONS USED IN PHYSICS-BASED DOCKING

Quantum effects approximated classically
Protein often held rigid
Configurational entropy neglected
Influence of water treated crudely

Two main approaches:
(1). Receptor/Target-Based
(2). Ligand/Drug-Based
Hand-on time!

https://bioboot.github.io/bggn213_f17/lectures/#12

You can use the classroom computers or your own laptops. If you are using your laptops then you will need to install [VMD](https://vmd.scripps.edu/) and [MGLTools](https://mgltools.scripps.edu/)

Two main approaches:

1. **Receptor/Target-Based**
2. **Ligand/Drug-Based**

---

**Scenario 2**

*Structure of Targeted Protein Unknown: Ligand-Based Drug Discovery*

- e.g. MAP Kinase Inhibitors

Using knowledge of existing inhibitors to discover more

**Why Look for Another Ligand if You Already Have Some?**

- Experimental screening generated some ligands, but they don’t bind tightly enough
- A company wants to work around another company’s chemical patents
- An high-affinity ligand is toxic, is not well-absorbed, difficult to synthesize etc.
LIGAND-BASED VIRTUAL SCREENING

- Compound Library
- Known Ligands
  - Molecular similarity
    - Machine-learning
    - Etc.
  - Candidate ligands
  - Optimization
    - Med chem, crystallography, modeling
  - Assay
  - Actives
    - Potent drug candidates

CHEMICAL SIMILARITY

- LIKAND-BASED DRUG-DISCOVERY
  - Compounds
    - (available/synthesizable)
  - Compare with known ligands
  - Different
    - Don’t bother
  - Similar
    - Test experimentally

CHEMICAL FINGERPRINTS

- BINARY STRUCTURE KEYS

<table>
<thead>
<tr>
<th>Molecule 1</th>
<th>Molecule 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>phenyl</td>
<td>phenyl</td>
</tr>
<tr>
<td>methyl</td>
<td>methyl</td>
</tr>
<tr>
<td>ketone</td>
<td>ketone</td>
</tr>
<tr>
<td>carboxylate</td>
<td>carboxylate</td>
</tr>
<tr>
<td>amide</td>
<td>amide</td>
</tr>
<tr>
<td>aldehyde</td>
<td>aldehyde</td>
</tr>
<tr>
<td>chlorine</td>
<td>chlorine</td>
</tr>
<tr>
<td>fluorine</td>
<td>fluorine</td>
</tr>
<tr>
<td>ethyl</td>
<td>ethyl</td>
</tr>
<tr>
<td>naphthyl</td>
<td>naphthyl</td>
</tr>
<tr>
<td>S-S bond</td>
<td>S-S bond</td>
</tr>
<tr>
<td>alcohols</td>
<td>alcohols</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

CHEMICAL SIMILARITY FROM FINGERPRINTS

- Tanimoto Similarity
  - (or Jaccard Index), \( T = \frac{N_I}{N_U} = 0.25 \)

Intersection: \( N_I = 2 \)

Union: \( N_U = 8 \)
Pharmacophore Models
Φάρμακο (drug) + Φορά (carry)

A 3-point pharmacophore

Bulky hydrophobe

5.0 ±0.3 Å
3.2 ±0.4 Å
2.8 ±0.3 Å

Aromatic

Molecular Descriptors
More abstract than chemical fingerprints

Physical descriptors
- molecular weight
- charge
- dipole moment
- number of H-bond donors/acceptors
- number of rotatable bonds
- hydrophobicity (log P and clogP)

Topological
- branching index
- measures of linearity vs interconnectedness

Etc. etc.

A High-Dimensional “Chemical Space”
Each compound is at a point in an n-dimensional space
Compounds with similar properties are near each other

Apply multivariate statistics and machine learning for descriptor-selection. (e.g. partial least squares, support vector machines, random forest, deep learning etc.)

Approved drugs and clinical candidates
- Catalogue approved drugs and clinical candidates from FDA Orange Book, and USAN applications
- Small molecules and biotherapeutics
Drug properties

LIPINSKI'S RULE OF FIVE

Lipinski's rule of five states that, in general, an orally active drug has no more than one violation of the following criteria:

• Not more than 5 hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms)
• Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
• A molecular mass less than 500 daltons
• An octanol-water partition coefficient log P not greater than 5

Rules for drug discovery success

• Set of approved drugs or medicinal chemistry compounds and their targets can be used to derive rules for drug discovery success (or failure):
  • What features make a successful drug target?
  • What features make a protein druggable by small molecules?
  • What features of a compound contribute to good oral bioavailability?
  • What chemical groups may be associated with toxicity?
Examples

**Target prediction models**

- Active compounds from ChEMBL can be used to train target prediction models
- Variety of methods used
  - Multi-Category Naïve Bayesian Classifier (e.g., ChEMBL)
  - Chemical similarity between ligand sets (e.g., SEA)
  - 3D similarity between ligands (e.g., SwissTargetPrediction)
  - Protein and ligand descriptors (e.g., Proteochemometric models)
- Open source tools available for many methods
  - E.g., Scikit-learn with RDKit

Examples at: https://github.com/chembl/mychembl/blob/master/ipython_notebooks

---

**CAUTIONARY NOTES**

- “Everything should be made as simple as it can be but not simpler”
  A model is never perfect. A model that is not quantitatively accurate in every respect does not preclude one from establishing results relevant to our understanding of biomolecules as long as the biophysics of the model are properly understood and explored.

- Calibration of the parameters is an ongoing and imperfect process
  Questions and hypotheses should always be designed such that they do not depend crucially on the precise numbers used for the various parameters.

- A computational model is rarely universally right or wrong
  A model may be accurate in some regards, inaccurate in others. These subtleties can only be uncovered by comparing to all available experimental data.
SUMMARY

• Structural bioinformatics is computer aided structural biology

• Described major motivations, goals and challenges of structural bioinformatics

• Reviewed the fundamentals of protein structure

• Introduced both physics and knowledge based modeling approaches for describing the structure, energetics and dynamics of proteins computationally

• Introduced both structure and ligand based bioinformatics approaches for drug discovery and design

NEXT UP:

› Overview of structural bioinformatics
  • Major motivations, goals and challenges

› Fundamentals of protein structure
  • Composition, form, forces and dynamics

› Representing and interpreting protein structure
  • Modeling energy as a function of structure

› Example application areas
  • Predicting functional dynamics & drug discovery

PREDICTING FUNCTIONAL DYNAMICS

• Proteins are intrinsically flexible molecules with internal motions that are often intimately coupled to their biochemical function
  – E.g. ligand and substrate binding, conformational activation, allosteric regulation, etc.

• Thus knowledge of dynamics can provide a deeper understanding of the mapping of structure to function
  – Molecular dynamics (MD) and normal mode analysis (NMA) are two major methods for predicting and characterizing molecular motions and their properties

MOLECULAR DYNAMICS SIMULATION

• Use force-field to find Potential energy between all atom pairs
• Move atoms to next state
• Repeat to generate trajectory

McCammon, Gelin & Karplus, Nature (1977)
[ See: https://www.youtube.com/watch?v=ui1ZysMFcKk ]
Divide time into discrete (~1fs) time steps ($\Delta t$) (for integrating equations of motion, see below)

$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet$ $t$

At each time step calculate pair-wise atomic forces ($F(t)$) (by evaluating force-field gradient)

Nucleic motion described classically

$$\frac{F}{m} \frac{d^2 \mathbf{r}}{dt^2} = - \nabla V(\mathbf{r})$$

Empirical force field

$$V(\mathbf{r}) = \sum_{i<j} \frac{1}{4} \sum_{i \neq j} \mathbf{r}_{ij} \times \mathbf{F}_{ij}$$

Use the forces to calculate velocities and move atoms to new positions (by integrating numerically via the “leapfrog” scheme)

$$v(t + \Delta t) = v(t - \frac{\Delta t}{2}) + \frac{m}{\Delta t} F(t)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + v(t) \Delta t$$

REPEAT, (iterate many, many times… 1ms = $10^{12}$ time steps)
MD Prediction of Functional Motions

Simulations Identify Key Residues Mediating Dynamic Activation

Example Application of Molecular Simulations to GPCRs

Proteins Jump Between Many, Hierarchically Ordered “Conformational Substates”

Yao and Grant, Biophys J. (2013)

Yao and Grant, Journal of Biological Chemistry (2016)

H. Frauenfelder et al., Science 229 (1985) 337
MOLECULAR DYNAMICS IS VERY EXPENSIVE

Example: F$_1$-ATPase in water (183,674 atoms) for 1 nanosecond:

$\Rightarrow 10^6$ integration steps
$\Rightarrow 8.4 \times 10^{11}$ floating point operations/step

$[n(n-1)/2$ interactions]

Total: $8.4 \times 10^{17}$ flop
(on a 100 Gflop/s cpu: ca 25 years!)

... but performance has been improved by use of:

- multiple time stepping ca. 2.5 years
- fast multipole methods ca. 1 year
- parallel computers ca. 5 days
- modern GPUs ca. 1 day
- (Anton supercomputer ca. minutes)

COARSE GRAINING: NORMAL MODE ANALYSIS (NMA)

- MD is still time-consuming for large systems
- Elastic network model NMA (ENM-NMA) is an example of a lower resolution approach that finishes in seconds even for large systems.

NMA models the protein as a network of elastic strings

Atomistic

COARSE GRAINED

COARSE GRAINING: NORMAL MODE ANALYSIS (NMA)

- 1 bead / 1 amino acid
- Connected by springs

Hand-on time!

https://bioboot.github.io/bgg213_f17/lectures/#12

Focus on section 4 exploring PCA and NMA apps
Key forces affecting structure:

- H-bonding
- Van der Waals
- Electrostatics
- Hydrophobicity

\[ AE = \frac{A}{r^2} - \frac{B}{r^6} \]

\[ 2.6 \, \text{Å} < d < 3.1 \, \text{Å} \]

\[ 150^\circ < \theta < 180^\circ \]
Key forces affecting structure:

- H-bonding
- Van der Waals
- Electrostatics
- Hydrophobicity

(Some time called IONIC BONDS or SALT BRIDGES)

\[ E = \frac{K q_1 q_2}{Dr} \]

\( E \) = Energy
\( k \) = constant
\( D \) = Dielectric constant (vacuum = 1; \( H_2O = 80 \))
\( q_1 \) & \( q_2 \) = electronic charges (Coulombs)
\( r \) = distance (Å)

The force that causes hydrophobic molecules or nonpolar portions of molecules to aggregate together rather than to dissolve in water is called Hydropobicity (Greek, “water fearing”). This is not a separate bonding force; rather, it is the result of the energy required to insert a nonpolar molecule into water.

SUMMARY

- Structural bioinformatics is computer aided structural biology
- Described major motivations, goals and challenges of structural bioinformatics
- Reviewed the fundamentals of protein structure
- Introduced both physics and knowledge based modeling approaches for describing the structure, energetics and dynamics of proteins computationally
- Introduced both structure and ligand based bioinformatics approaches for drug discovery and design