Section: Sat, March 20, 2021 10am-10:50am

INTRODUCTION TO COMPUTER-AIDED DRUG DESIGN (CADD)

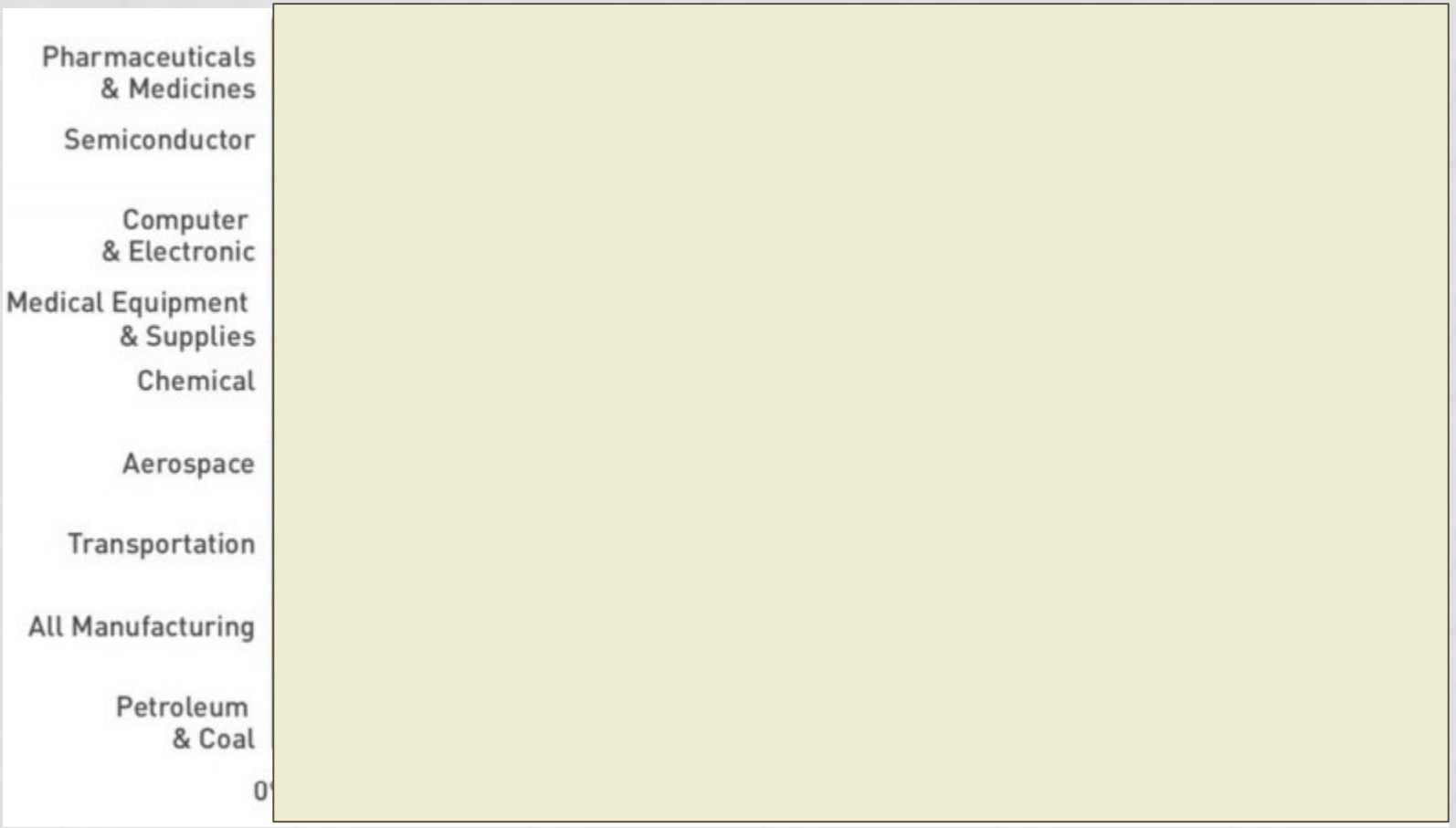
BENJAMIN RUDSHTEYN
POSTDOCTORAL RESEARCH FELLOW IN CHEMISTRY AT COLUMBIA UNIVERSITY
BR2575@COLUMBIA.EDU

IS THIS THE CLASS YOU ARE LOOKING FOR?

- Computer-aided drug design (CADD) uses the tools of computational chemistry and data science to aid the process of drug (medication; not the bad kind) discovery.
- Almost everything you learned about the lock and key model of protein-ligand binding is wrong. Concepts include,
  - the drug discovery pipeline and associated costs,
  - ligand-based vs. structure-based drug design,
  - lock and key vs. allostery
  - hit Identification and lead optimization,
  - Lipinski's rule of five,
  - structure-function relationships,
  - high-throughput virtual screening,
  - agonist vs. antagonist vs. inverse agonist,
  - decoys,
  - selectivity and side effects,
  - drug docking
  - molecular dynamics
  - more advanced computational methods.
- **Prerequisites**
  High school level biology/chemistry in grades 10-12

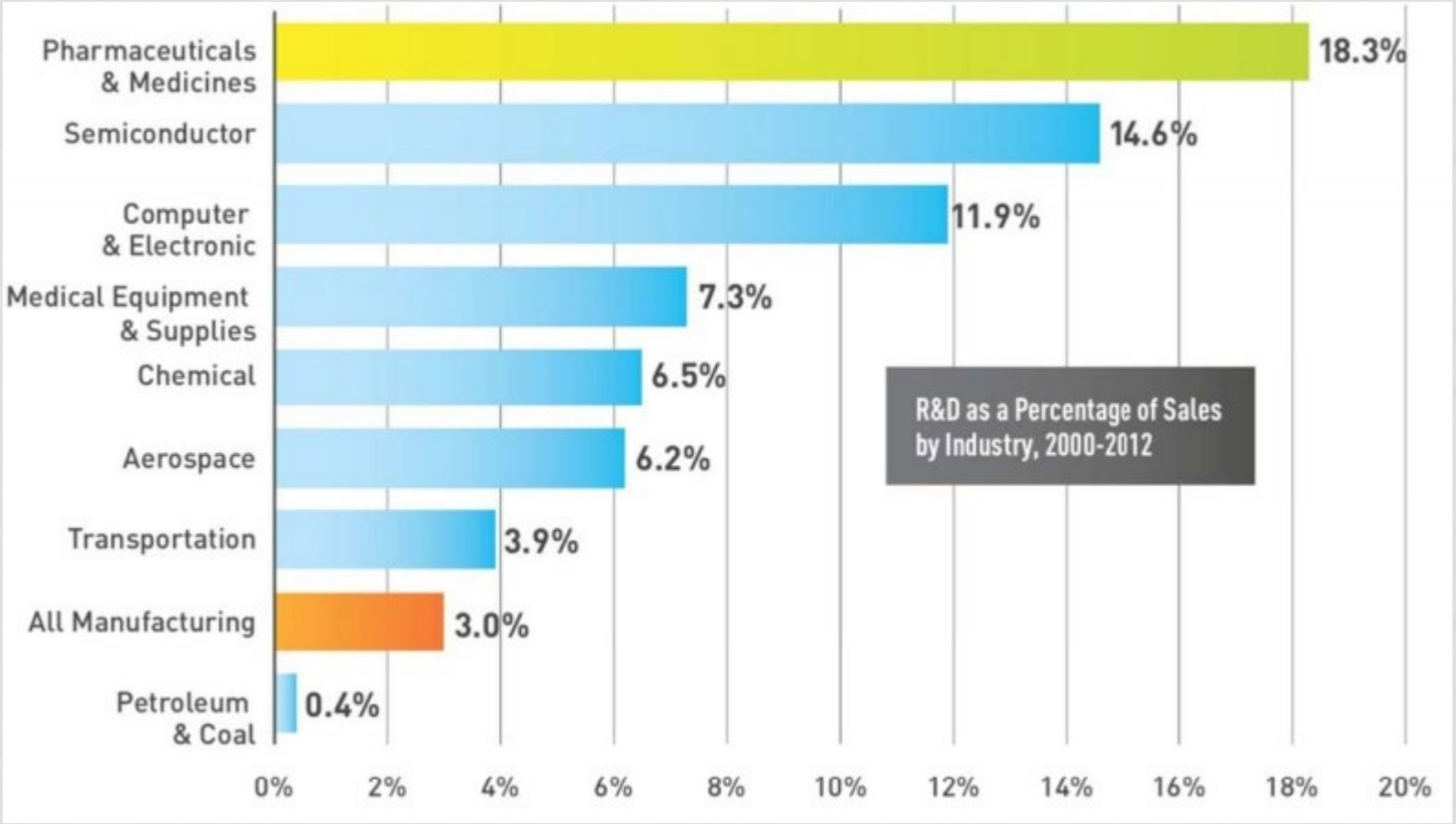
  AP Biology and/or Chemistry are/is a huge plus.

Costs of Pharmaceuticals



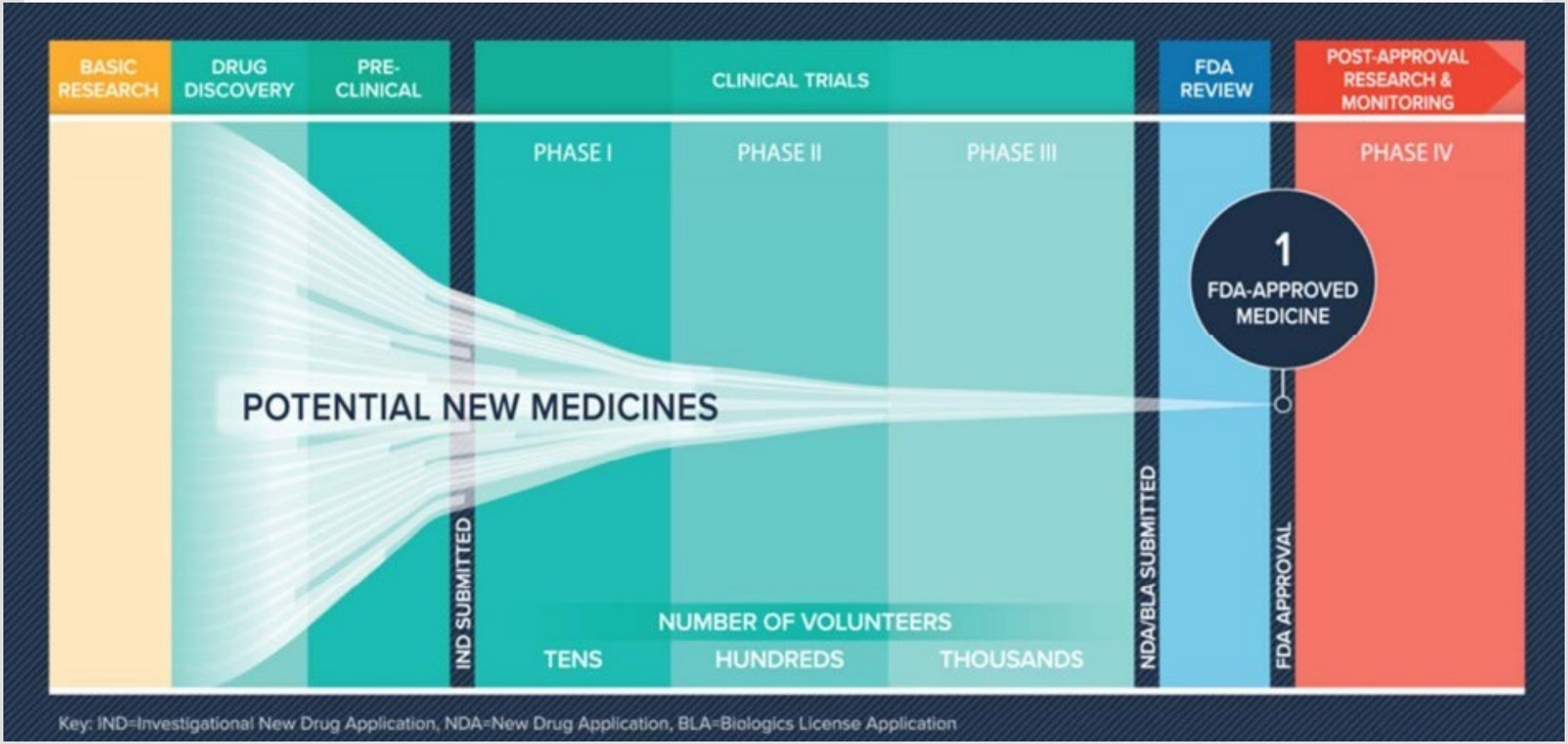
https://theskepticalchemist.com/process-costs-drug-development/

Costs of Pharmaceuticals



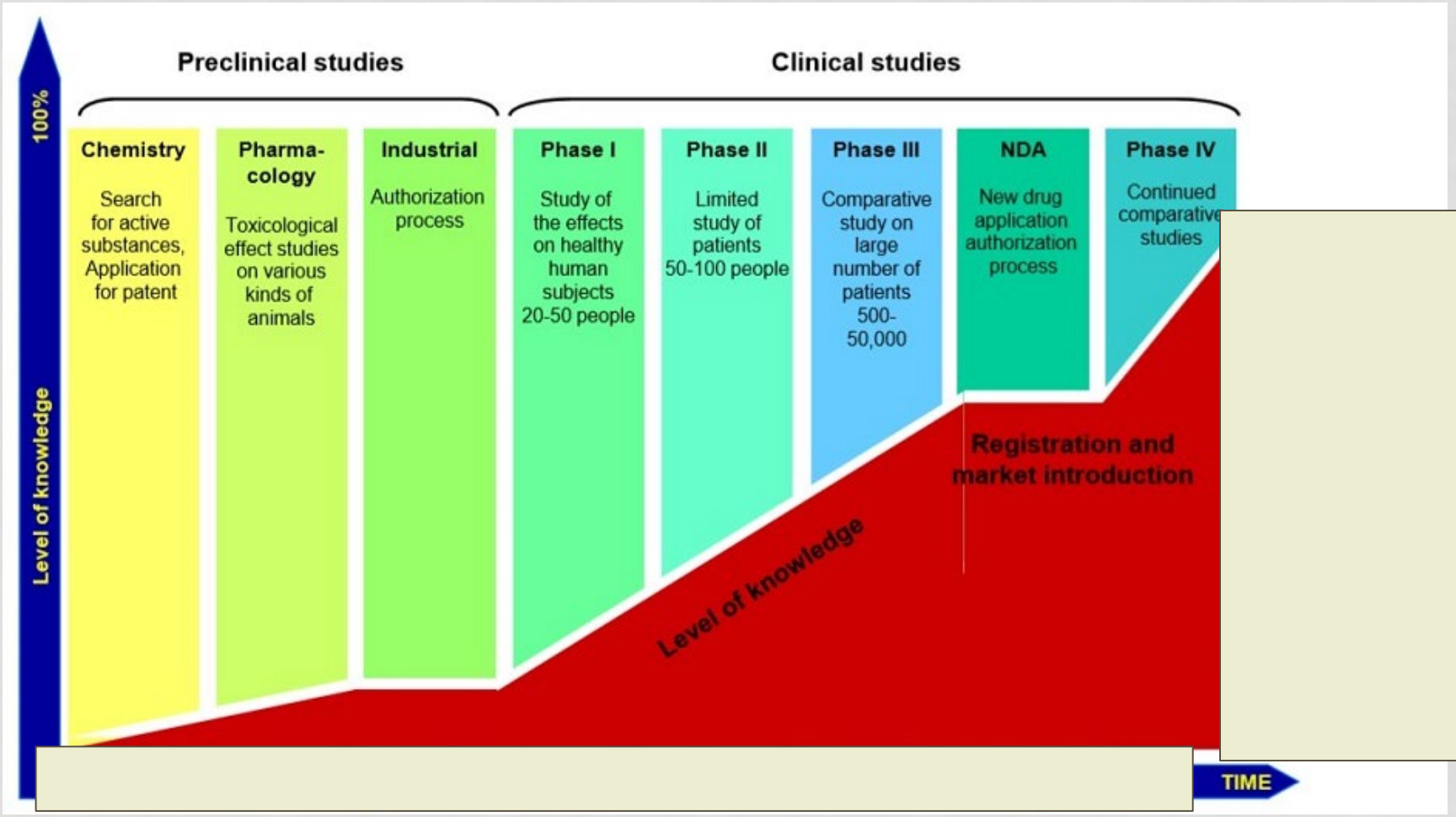
https://theskepticalchemist.com/process-costs-drug-development/

Fail Fast, Fail Cheap



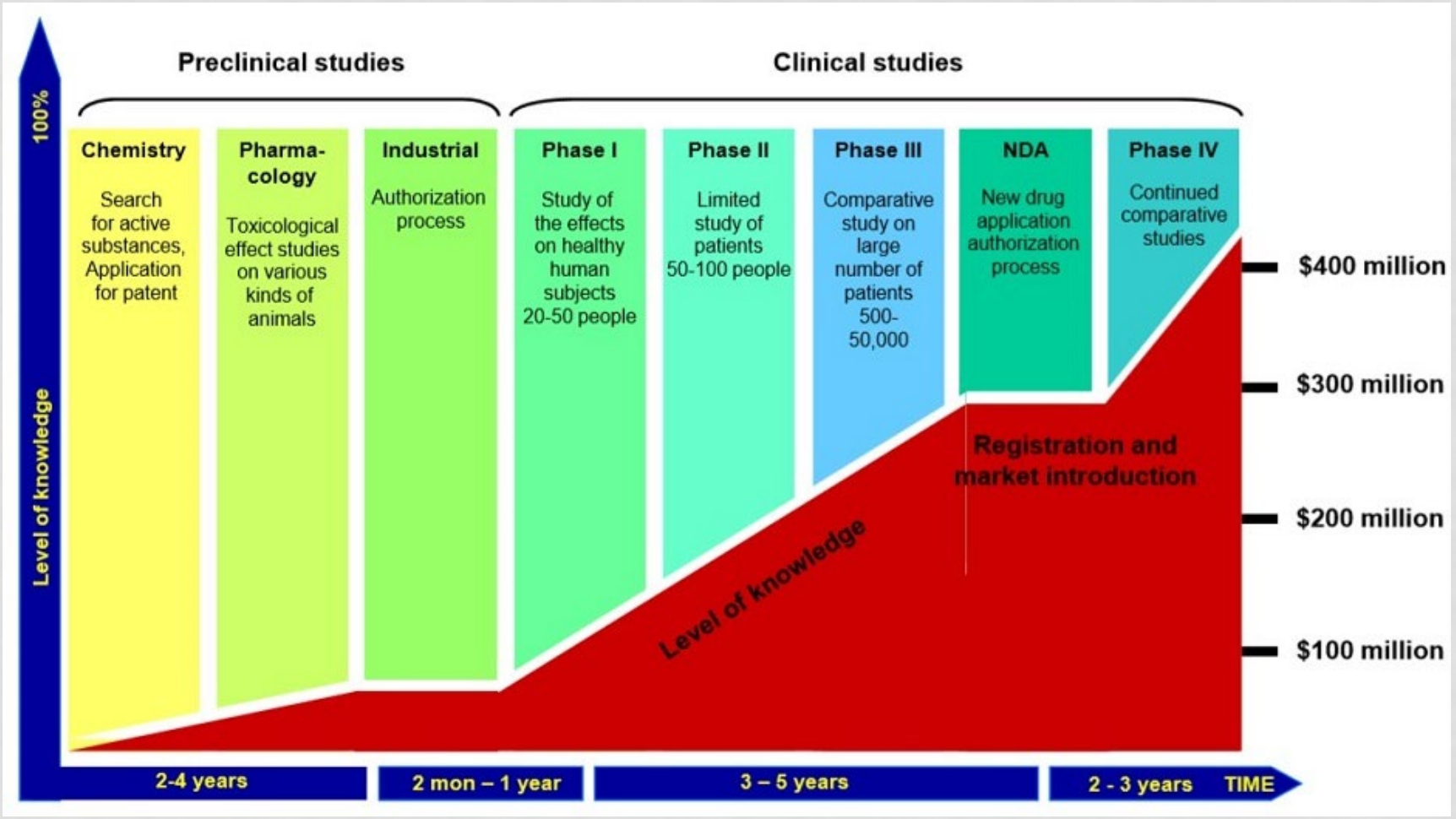
https://theskepticalchemist.com/process-costs-drug-development/

Drug Discovery Costs



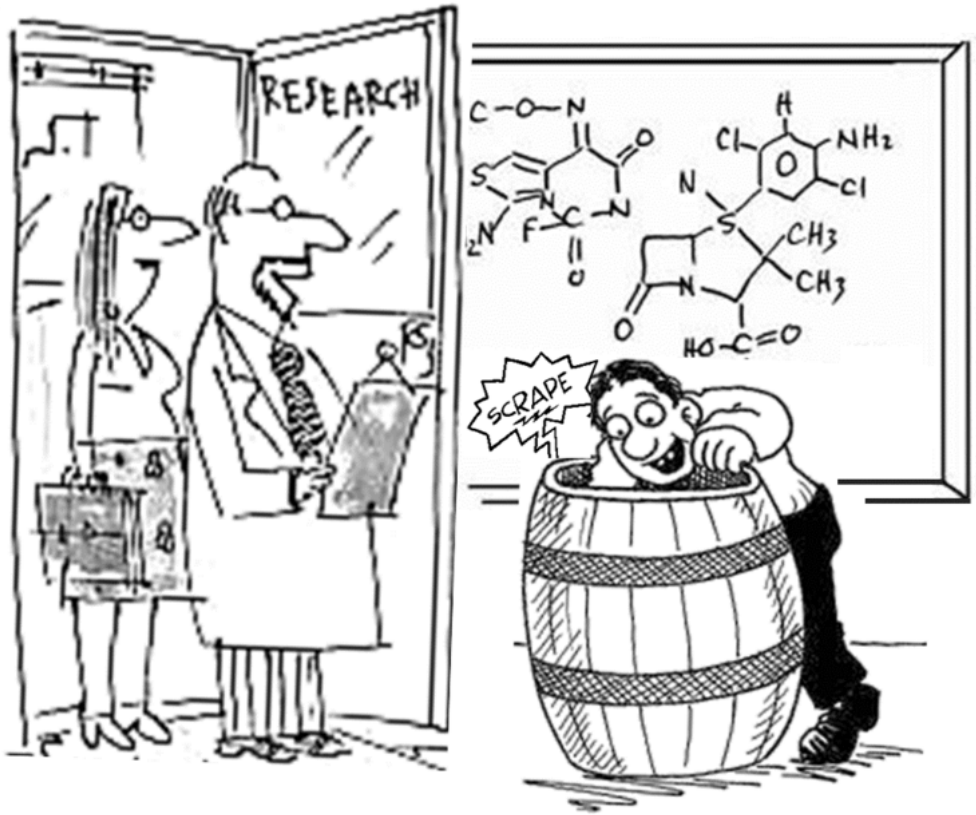
https://theskepticalchemist.com/process-costs-drug-development/

Drug Discovery Costs



https://theskepticalchemist.com/process-costs-drug-development/

Why is it so expensive? All the "easy" drugs have been found already.



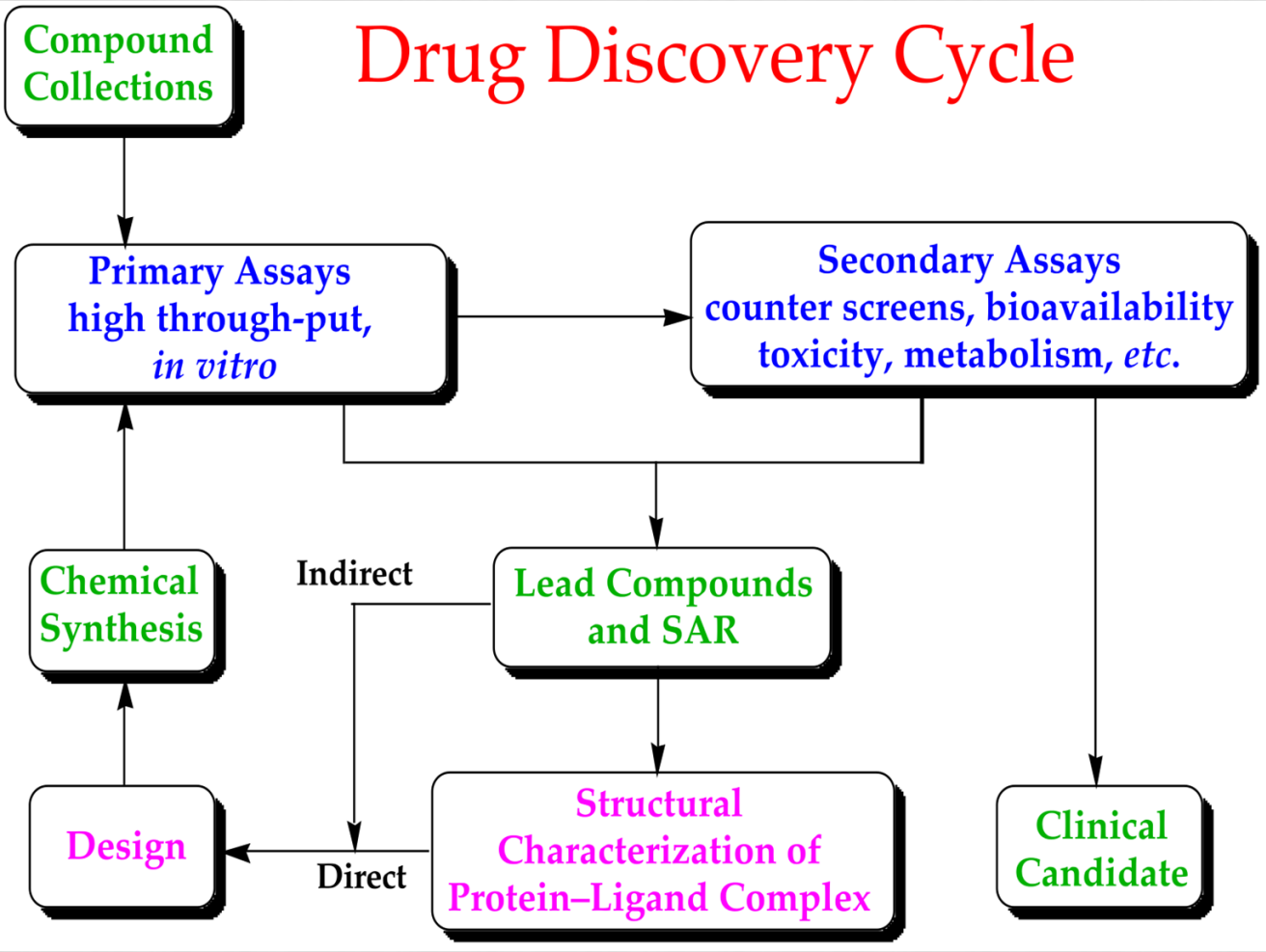
"And this is where we find all our new drug targets."

https://www.linkedin.com/pulse/pharmas-broken-business-model-part-2-scraping-barrel-kelvin-stott?trk=portfolio_article-card_title

What is a Drug Exactly?

- What do you think a drug does?
- If you were going to take a medicine to make your headache go away, what would you want this medicine to do?
- Drug: a medicine or other substance which has a physiological effect when ingested or otherwise introduced into the body. (Oxford Languages)
- What if the drug does other things too?
- Are these effects good or bad?
- What are these effects called?

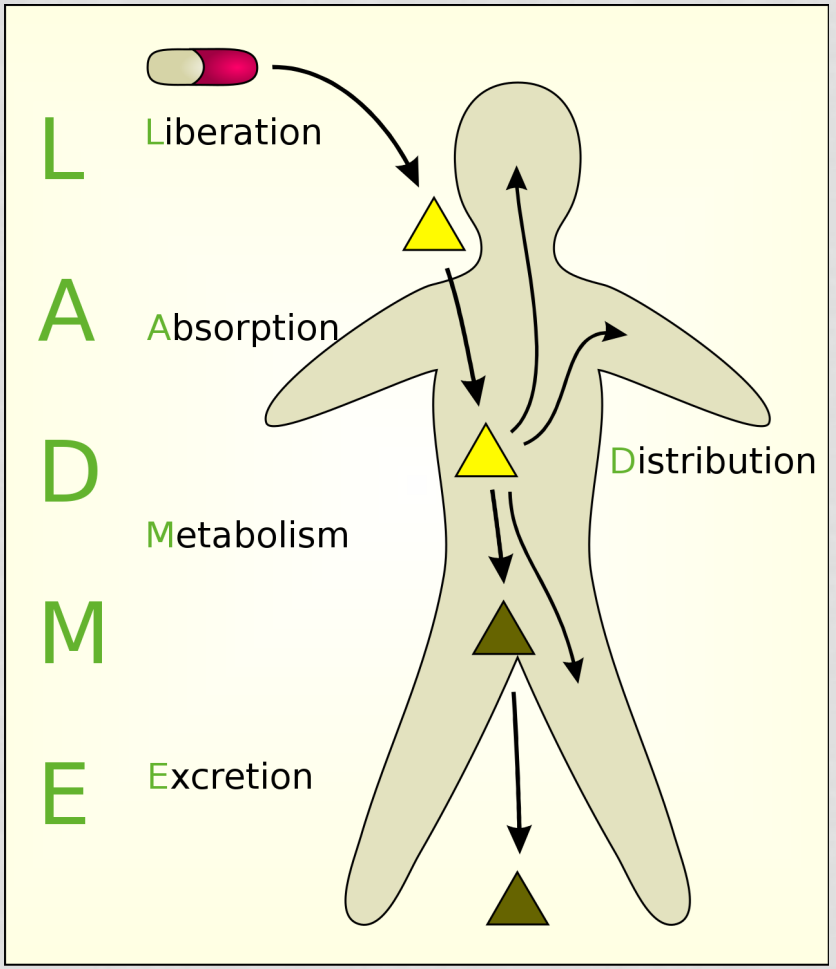
How Drug Discovery Works Without the Computers



https://en.wikipedia.org/wiki/Drug_design#Computer-aided_drug_design

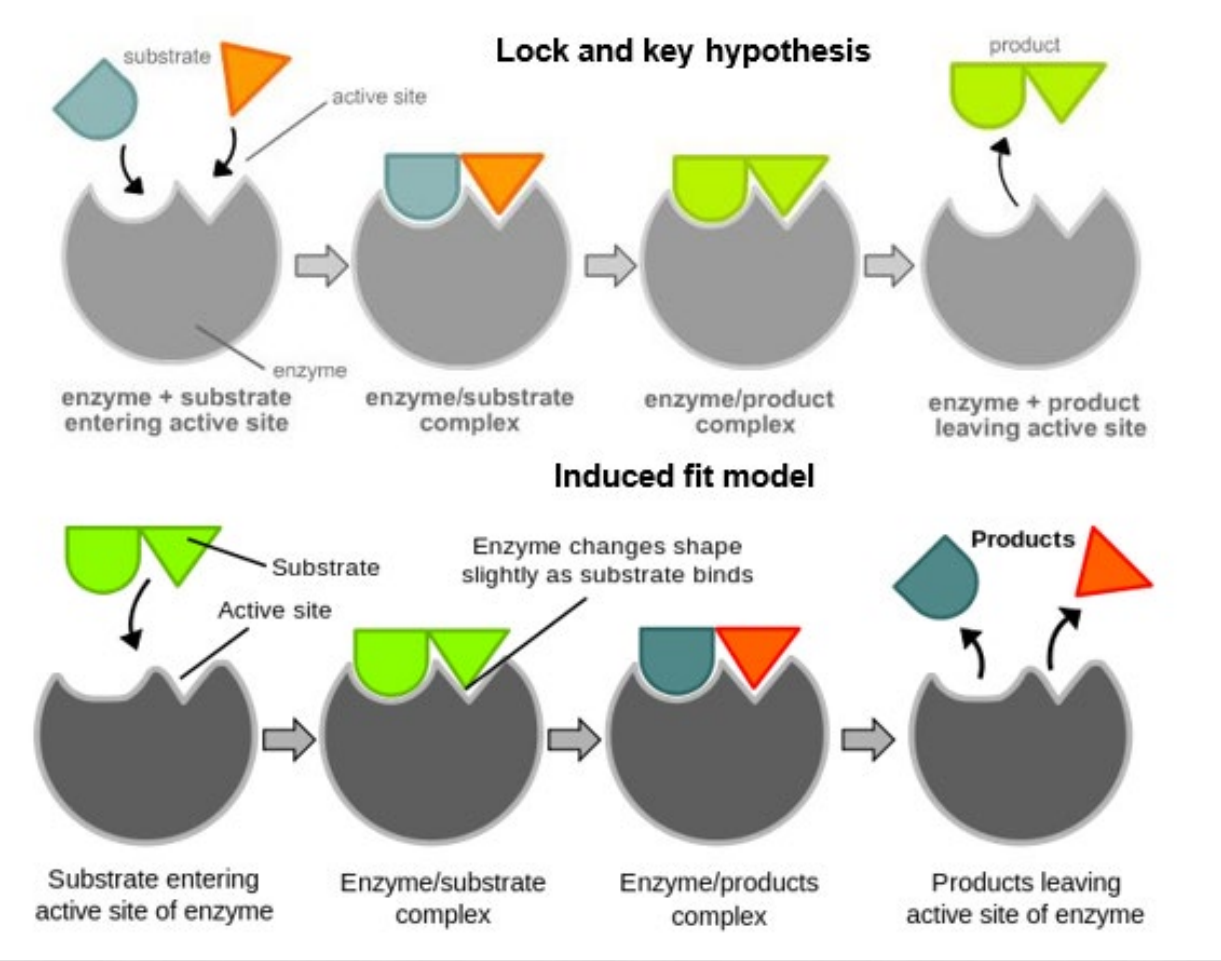
Kinds of Drug Design

- Drug design is all about design of a molecule i.e. the ligand or drug that will bind tightly to its *intended* target, usually the protein, that also follows (L)ADME(T)
  - is bioavailable,
  - has appropriate metabolic half-life,
  - is so selective it does not bind to other proteins i.e. no side effects,
  - is not **toxic**



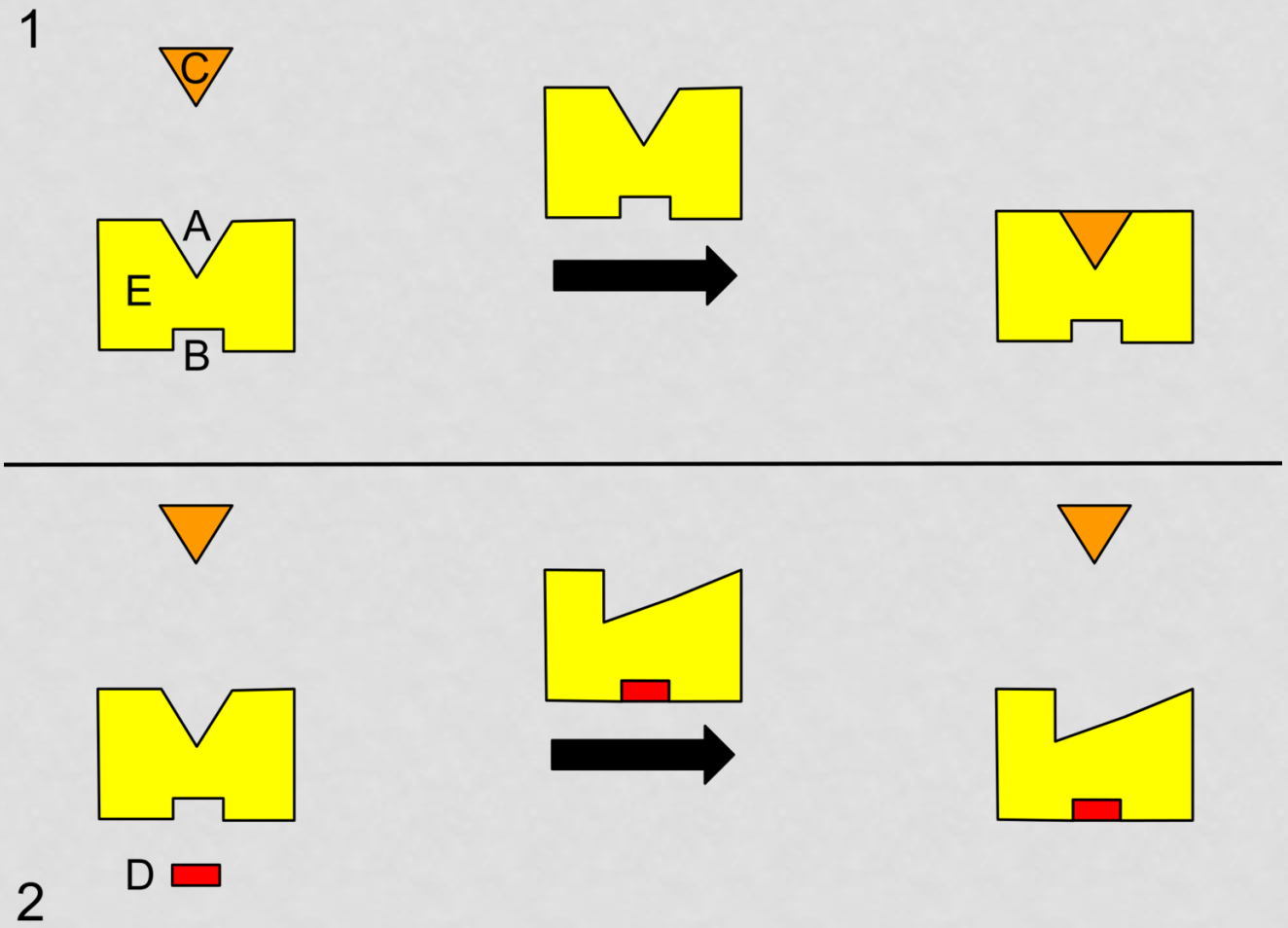
https://en.wikipedia.org/wiki/Drug_design
https://en.wikipedia.org/wiki/ADME

Lock and Key Models



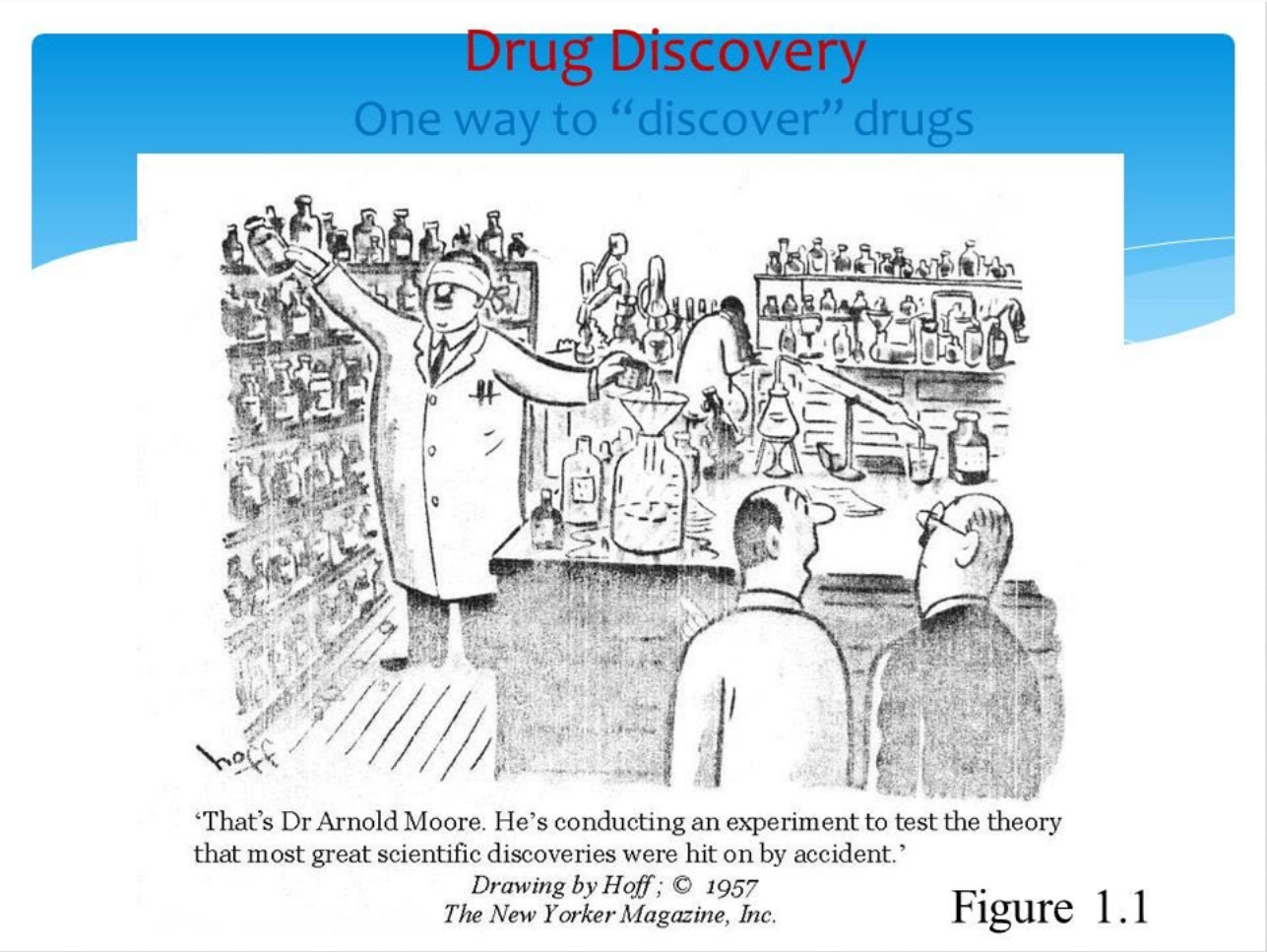
also conformational selection
https://alevelbiology.co.uk/notes/biological-catalysts-enzymes/

Allosteric Interactions



https://en.wikipedia.org/wiki/Allosteric_regulation

Random Approach to Drug Discovery



https://slideplayer.com/slide/5813563/19/images/2/Drug+Discovery+One+way+to+discover+drugs.jpg

Other Kinds of Drug Discovery

- Structure-based drug design: drug design that relies on the knowledge of the three-dimensional structure of the biomolecular target is known. (more expensive, more accurate)
- Ligand-based drug design: drug design that relies on knowledge of other molecules that bind to the biological target of interest. (less expensive, less accurate)

https://en.wikipedia.org/wiki/Drug_design

Where Do the Computers Come in?

- Hit identification using virtual screening (structure- or ligand-based design).
- Hit to lead (H2L, lead generation) is a stage in early drug discovery where small molecule hits from a high throughput (virtual) screen (HT(V)S) are evaluated and undergo limited optimization to identify promising lead compounds.
- Hit-to-lead optimization of affinity and selectivity (structure-based design, QSAR, etc.)
- Lead optimization of other pharmaceutical properties while maintaining affinity

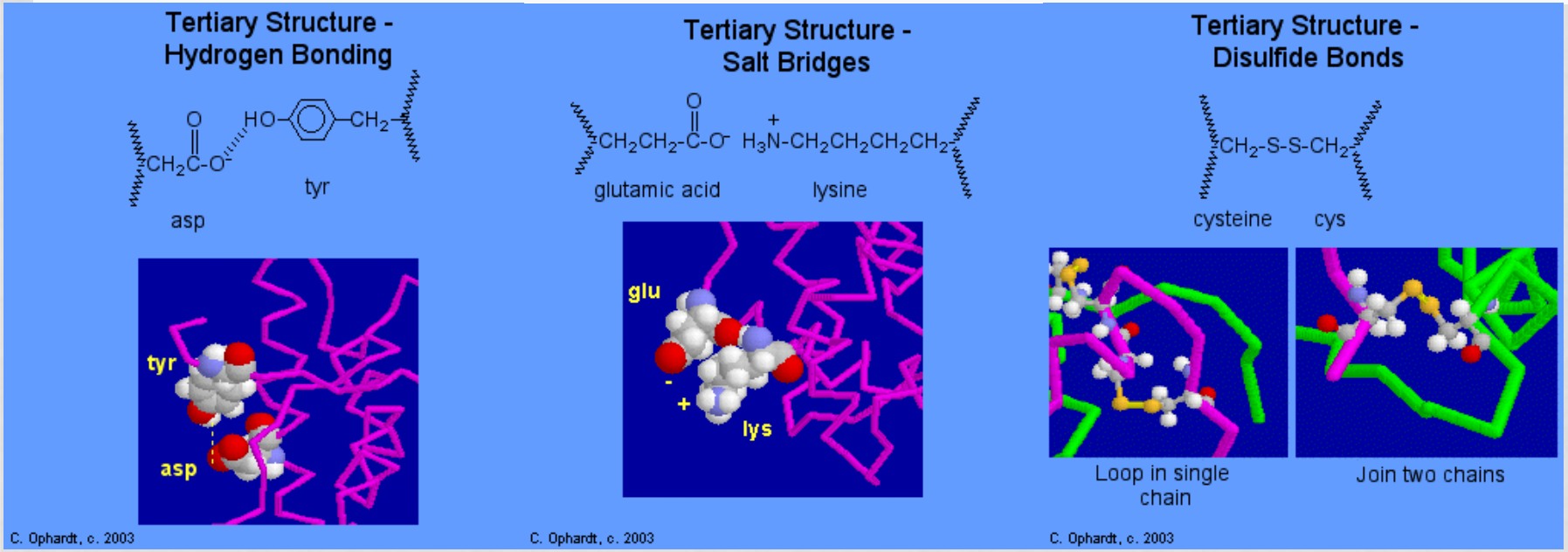
https://en.wikipedia.org/wiki/Drug_design
https://en.wikipedia.org/wiki/Hit_to_lead

Refresher: Tertiary Structure

- The tertiary structure is the final specific geometric shape that a protein assumes.
- There are four types of bonding interactions between "side chains" including
  - hydrogen bonding
  - salt bridges
  - disulfide bonds
  - non-polar hydrophobic interactions
- Often what drives more favorable binding is the molecule ejecting more unstable water molecules from the protein pocket while maintaining these properties.

http://chemistry.elmhurst.edu/vchembook/567tertprotein.html

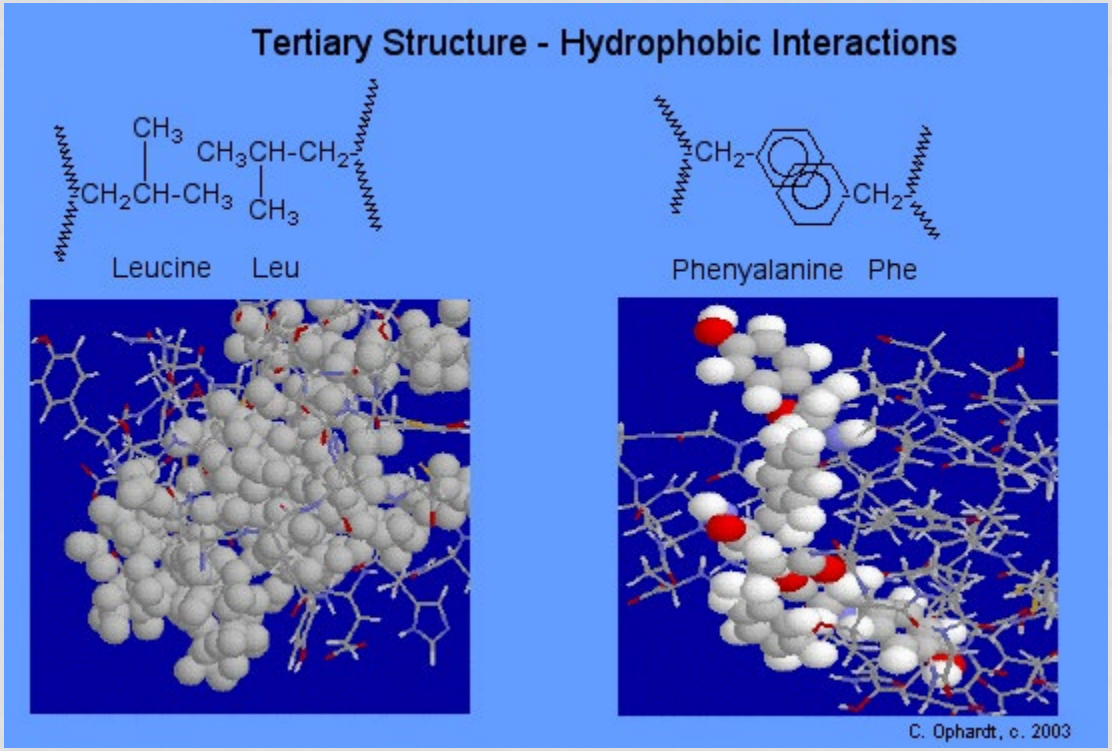
Polar Interactions



- Tyr is **donating** a hydrogen bond to Asp which is **accepting**

http://chemistry.elmhurst.edu/vchembook/567tertprotein.html

Nonpolar Interactions



http://chemistry.elmhurst.edu/vchembook/567tertprotein.html

How Hydrophobic is the Drug? Octanol-Water Partition Coefficient

- The partition coefficient measures how hydrophilic ("water-loving") or hydrophobic ("water-fearing") a chemical substance is.
- Partition coefficients are useful in estimating the distribution of drugs within the body.
- Hydrophobic drugs with high octanol-water partition coefficients are mainly distributed to hydrophobic areas such as lipid bilayers of cells.
- Conversely, hydrophilic drugs (low octanol/water partition coefficients) are found primarily in aqueous regions such as blood serum.

https://en.wikipedia.org/wiki/Partition_coefficient

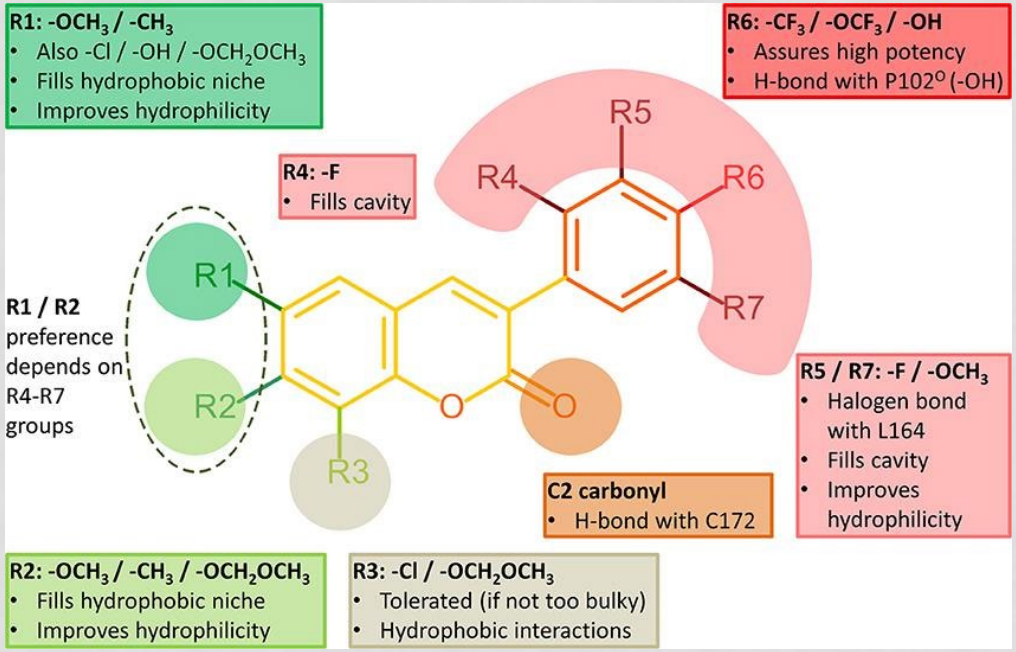
Lipinski's Rule of Five

1. No more than **5** hydrogen bond donors (the total number of nitrogen–hydrogen and oxygen–hydrogen bonds).
2. No more than **10** hydrogen bond acceptors (all nitrogen or oxygen atoms).
3. A molecular mass less than **500** Daltons (molecular weight units)
4. An octanol-water partition coefficient (log P) that does not exceed **5** (becoming too hydrophobic).

https://en.wikipedia.org/wiki/Drug_design
https://en.wikipedia.org/wiki/Hit_to_lead

Structure-Function Relationships

- (Quantitative) structure–activity relationship = (Q)SAR
- Relationship between the chemical structure of a molecule and its biological activity.
- Modification of the effect or the potency of a bioactive compound (typically a drug) by changing its chemical structure using chemical synthesis.

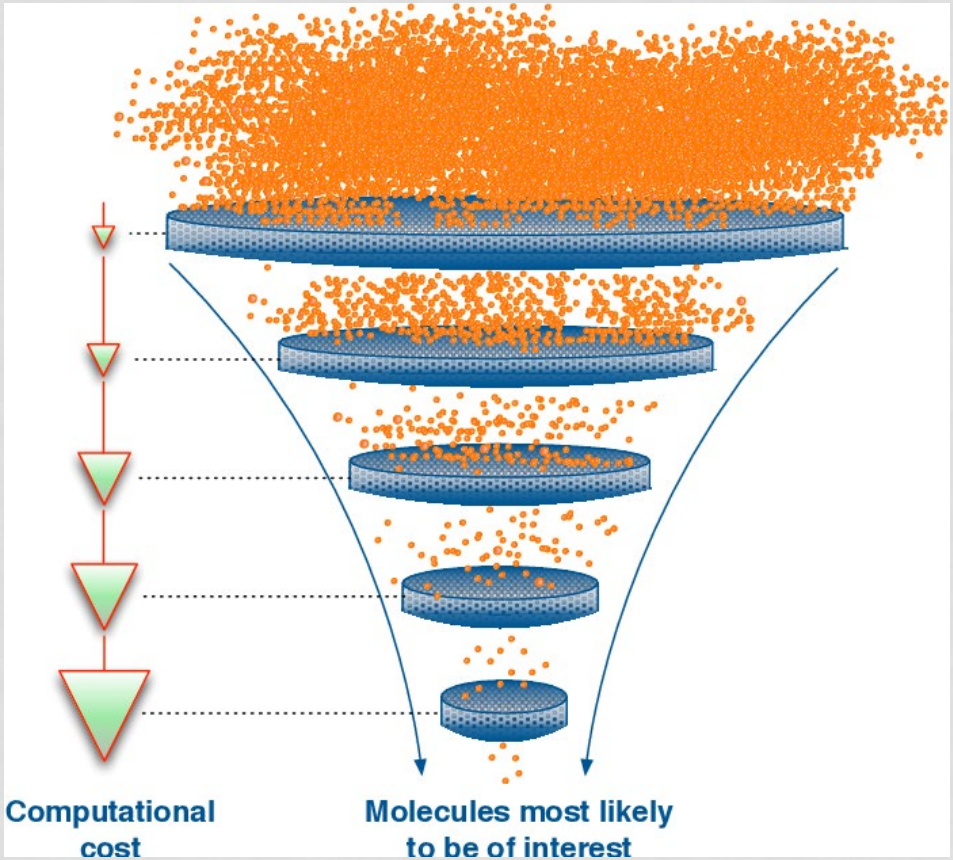


https://www.frontiersin.org/articles/10.3389/fchem.2018.00041/full

https://en.wikipedia.org/wiki/Structure%E2%80%93activity_relationship

High-Throughput Virtual Screening

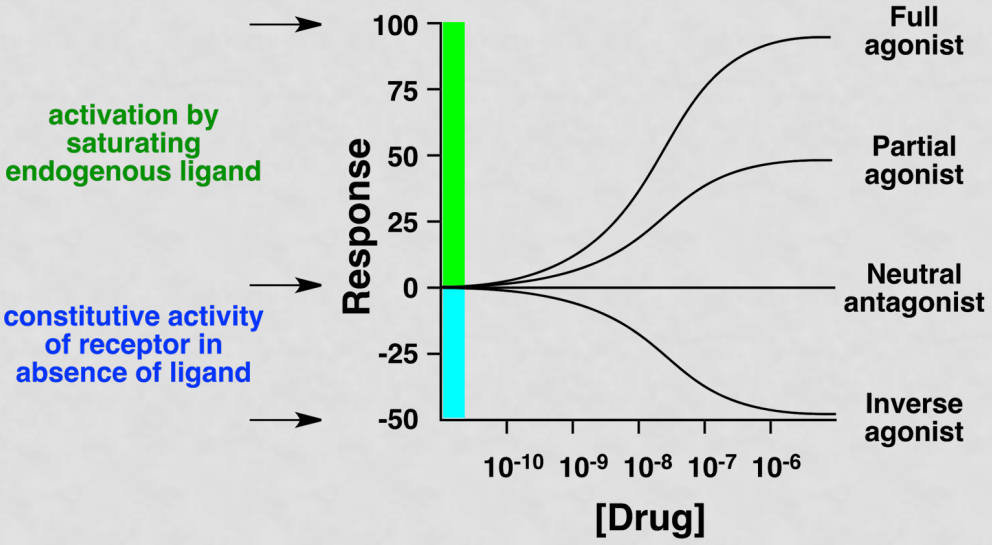
- Using progressively more expensive methods to screen through a large library of possible drug molecules to decide which are worth testing with the most expensive theoretical methods/experiments



https://www.annualreviews.org/doi/abs/10.1146/annurev-matsci-070214-020823
https://www.semanticscholar.org/paper/What-Is-High-Throughput-Virtual-Screening-A-from-Pyzer-Knapp-Suh/e142490c28bb6b954632e85ed6001eb312d001e7

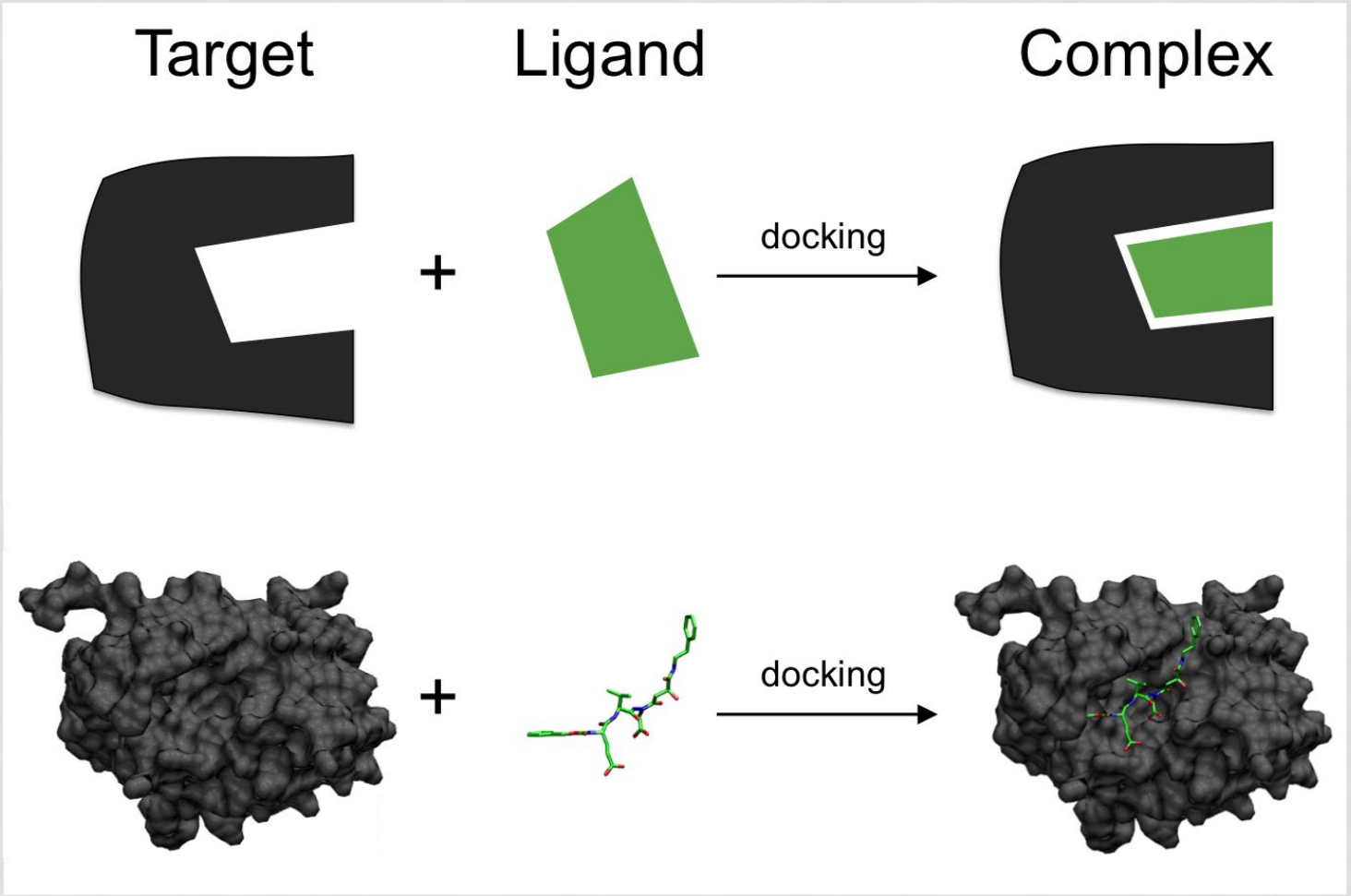
Agonist vs. Antagonist vs. Inverse Agonist

- Agonist: binds to a receptor and activates the receptor to produce a biological response.
- Antagonist (blockers): blocks or dampens a biological response by binding to and blocking a receptor.
- Inverse agonists: binds to the same receptor as an agonist but induces a pharmacological response opposite to that of the agonist.



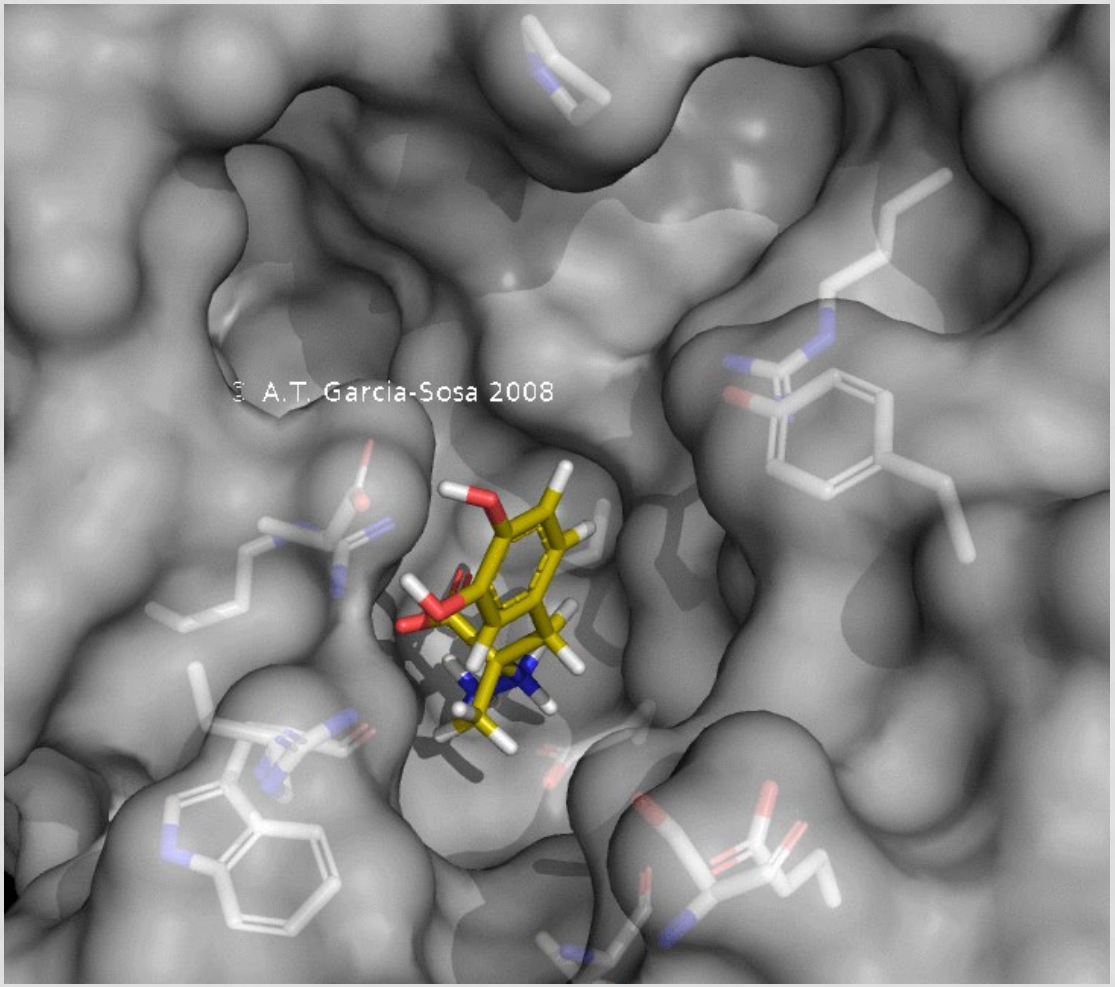
https://en.wikipedia.org/wiki/Inverse_agonist

Drug Docking



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

Docking of compounds exploring the binding sites of H5N1 avian influenza neuraminidase.



http://hermes.chem.ut.ee/~alfx/animationH5N1.html

Drug Docking: Typical Scoring Function

$$\Delta G_{\text{bind}} = -RT \ln K_{\text{d}}$$

$$K_{\text{d}} = \frac{[\text{Ligand}][\text{Receptor}]}{[\text{Complex}]}$$

$$\Delta G_{\text{bind}} = \Delta G_{\text{desolvation}} + \Delta G_{\text{motion}} + \Delta G_{\text{configuration}} + \Delta G_{\text{interaction}}$$

where:

- desolvation – enthalpic penalty for removing the ligand from solvent
- motion – entropic penalty for reducing the degrees of freedom when a ligand binds to its receptor
- configuration – conformational strain energy required to put the ligand in its "active" conformation
- interaction – enthalpic gain for "resolvating" the ligand with its receptor

https://en.wikipedia.org/wiki/Drug_design

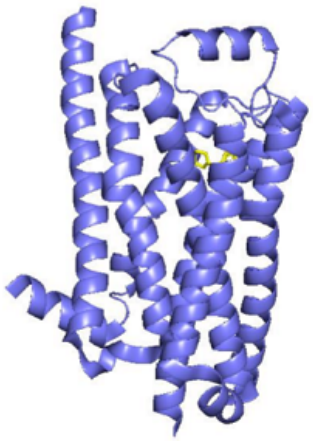
Docking and Decoys

- Test/train the docking algorithm on some target proteins with known strong binding molecules
- "An imperfect scoring function can mislead by predicting incorrect ligand geometries or by selecting nonbinding molecules over true ligands. These false-positive hits may be considered 'decoys'."
- Docking can be static or induced fit.

Video:
https://commons.wikimedia.org/w/index.php?title=File%3ADocking_GPCR_example.webm

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1557646/

What if Structure Has to Relax?



Crystal Structures

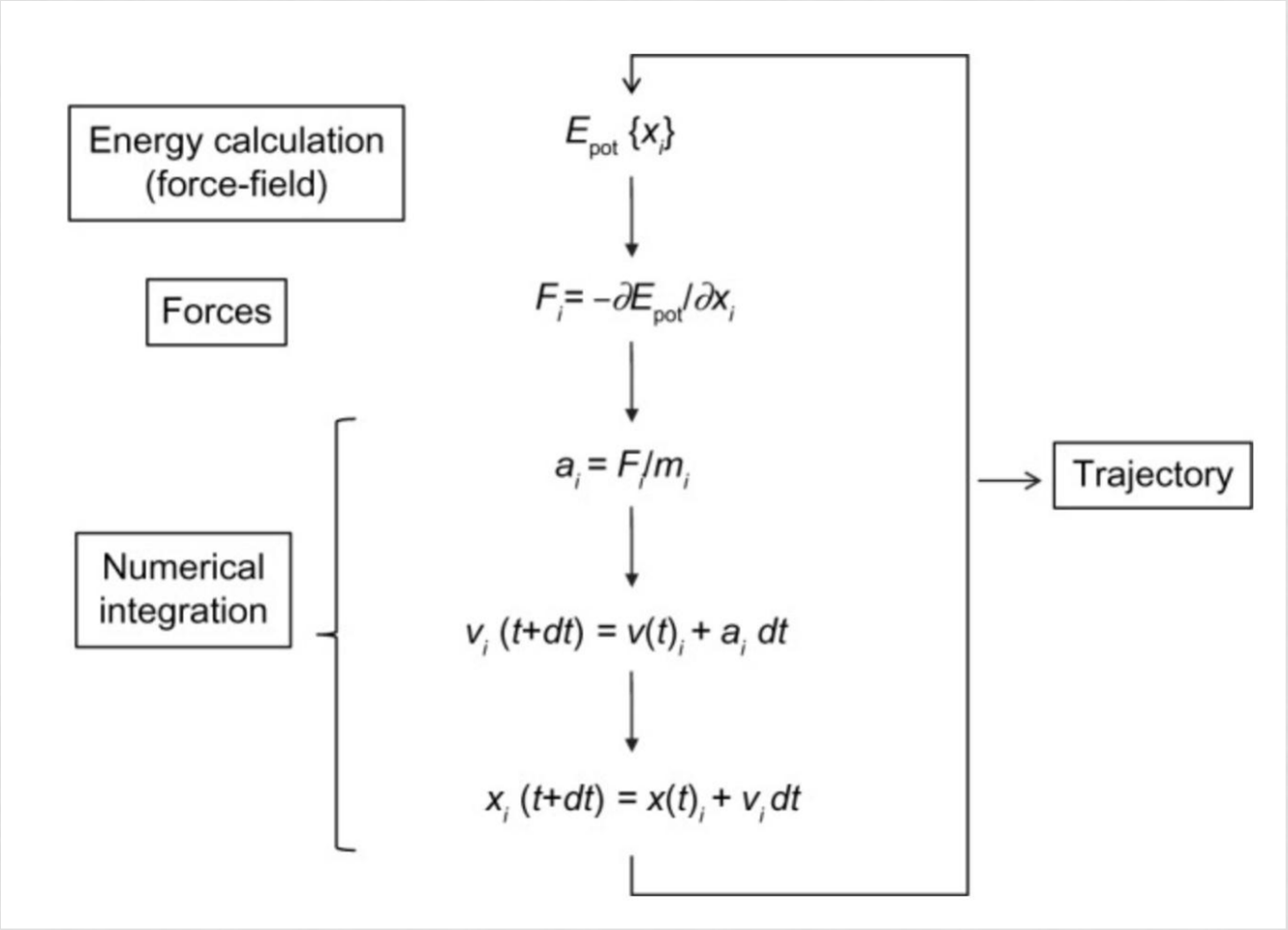
Protein Structure Databank (PDB)

Molecular Dynamics Simulations

- New Binding Sites
- Free Energy
- Dynamics
- Drug Design

https://www.preprints.org/manuscript/201908.0271/v1

Molecular Dynamics (MD): Classical Physics with Special Potential Energy Function



Potential Energy Function

Levitt

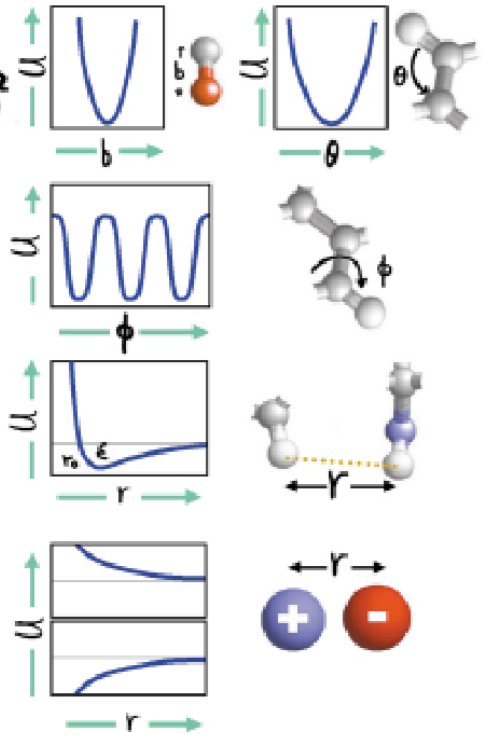
Nat. Struc. Bio. **2001,** *8*, 392

$$U = \sum_{\text{All Bonds}} \tfrac{1}{2} K_b (b - b_o)^2 + \sum_{\text{All Angles}} \tfrac{1}{2} K_\theta (\theta - \theta_o)^2$$

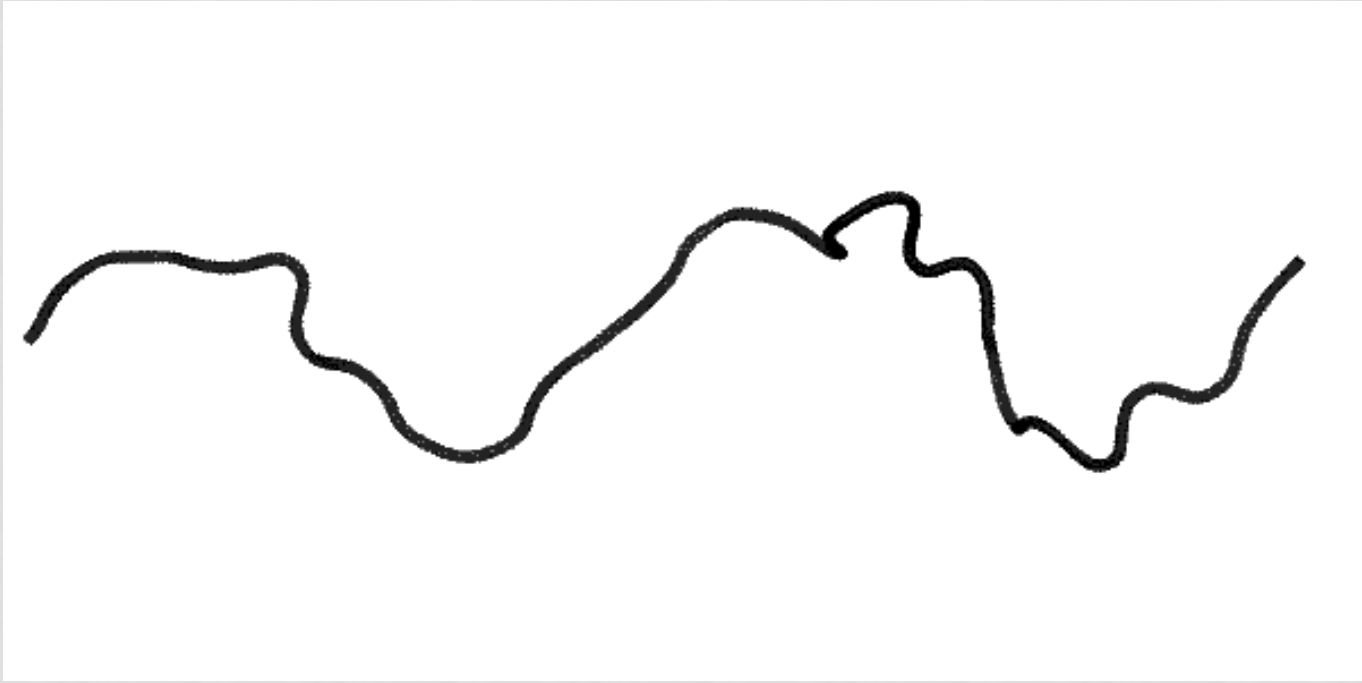
$$+ \sum_{\text{All Torsion Angles}} K_\phi [1 - \cos(n\phi + \delta)]$$

$$+ \sum_{\text{All nonbonded pairs}} \epsilon \left[\left(\frac{r_o}{r}\right)^{12} - 2\left(\frac{r_o}{r}\right)^{6} \right]$$

$$+ \sum_{\text{All partial charges}} 332 q_i q_j / r$$

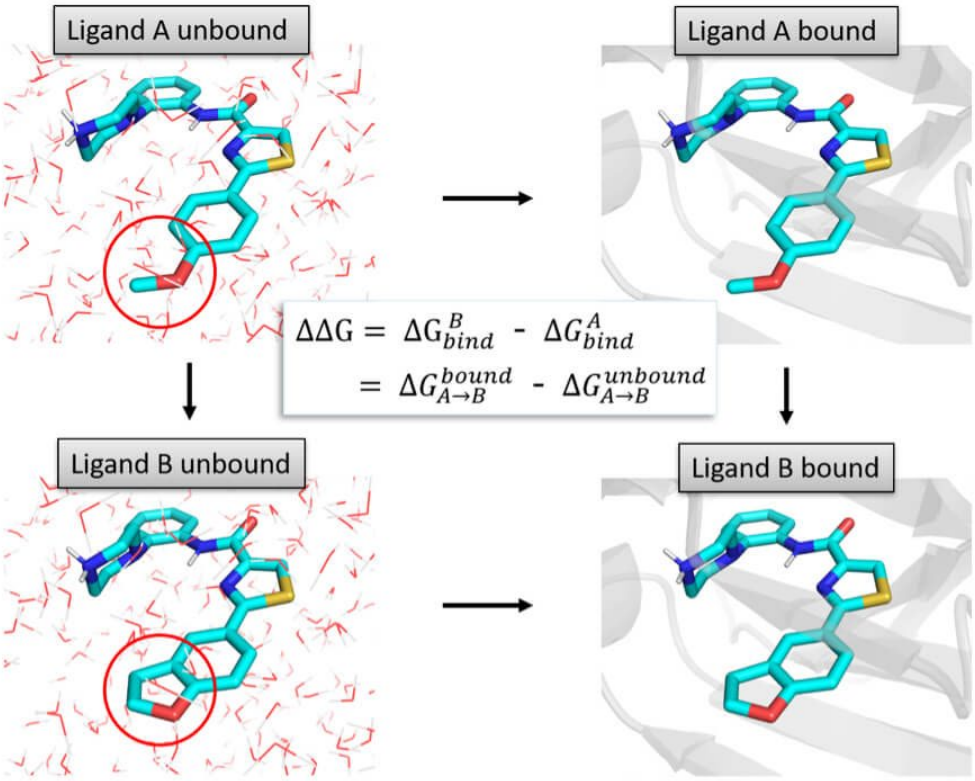


Protein Folding Example



https://medium.com/proteinqure/welcome-into-the-fold-bbd3f3b19fdd

Free Energy Perturbation (FEP)



https://www.xtalpi.com/en/core-tech/xfep/

Conclusions

- Drug discovery is very expensive and laborious.
- Drug molecules have to not only bind very well to the protein (potency; focus of the CADD), but also satisfy many other (pharmacophore) characteristics.
- Computer-aided drug design (CADD) offers a cost-saving approach (in comparison to experiments) to identify which molecules are worth studying further (screening).
- The expense of calculations is often directly related to their accuracy.
- Hierarchy of methods: ligand-based screening < rigid docking (classic lock + key) < induced fit docking (modern lock + key) < molecular dynamics approaches
- These methods use aspects of tertiary structure like hydrogen bonds, ionic bonds, hydrophobic interactions, and disulfide bridges.

Questions?
Interested in getting involved in research?

br2575@columbia.edu

In terms of getting started with computational chemistry research, I would recommend:

1. Taking rigorous courses in high school
2. Getting involved in the Regeneron (formerly Intel?) science fair high school program if possible
3. Getting involved in research in college.
4. However, there are some pre-college programs in computational chemistry/biology like:
    a. http://cbd.cmu.edu/education/pre-college/index.html
    b. http://www.discobio.pitt.edu/
    c. https://cns.utexas.edu/tides/k-12/high-school-summer-research-academy