Drug Design

An Industrial Perspective

Joseph Corkery
OpenEye Scientific Software

Drug Discovery: A Little History

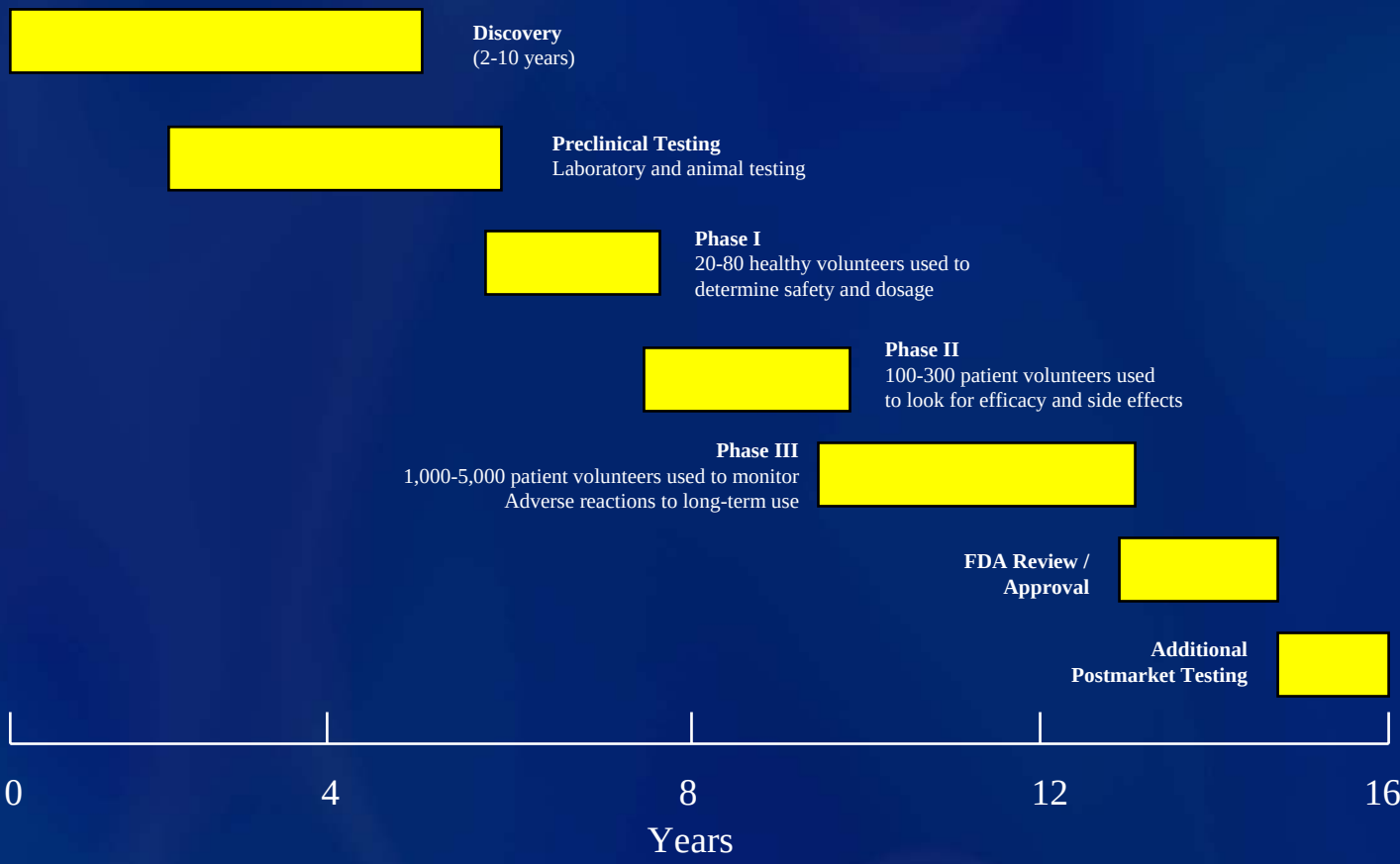
- "Folk" medicine
  - Digoxin (digitalis) from Foxglove plant
- Serendipity
  - Penicillin
  - Sildenafil (Viagra)
- Brute Force (Screen everything you can)
  - Most drugs developed during the 50s-70s
  - Many are derived from natural products
    - Example: Taxol (yew tree)
    - Companies often paid employees to bring back plants from exotic vacation destinations
- Design
  - HIV reverse transcriptase and HIV protease inhibitors

Drug Development: Overview

- Drug development is very expensive
  - Recent estimates put the cost of developing a single drug at ~$800 million
  - Much of this cost is an amortization of all the other drugs that fail during development
- Drug development is very time consuming
  - Recent estimates put the time to market for a single drug in the range from 8-15 years
- Drug development is a high-risk game
  - For every 250 lead candidates only 1 will make it to FDA approval
  - Seven out of every ten drugs brought to market never generate enough revenue to recover the cost of development *

* PhRMA: Pharmaceutical Spending Facts and Figures 2004
http://www.phrma.org/publications/publications/2005-08-30.1290.pdf

Drug Development: Overview



Compound Success Rates by Stage

5,000 – 10,000 screened

↓

250 enter preclinical testing

↓

5 enter clinical testing

↓

1 approved by FDA

Source: PhRMA, based on data from *Center for the Study of Drug Development*, Tufts University, 1995

Challenges to Making a Drug?

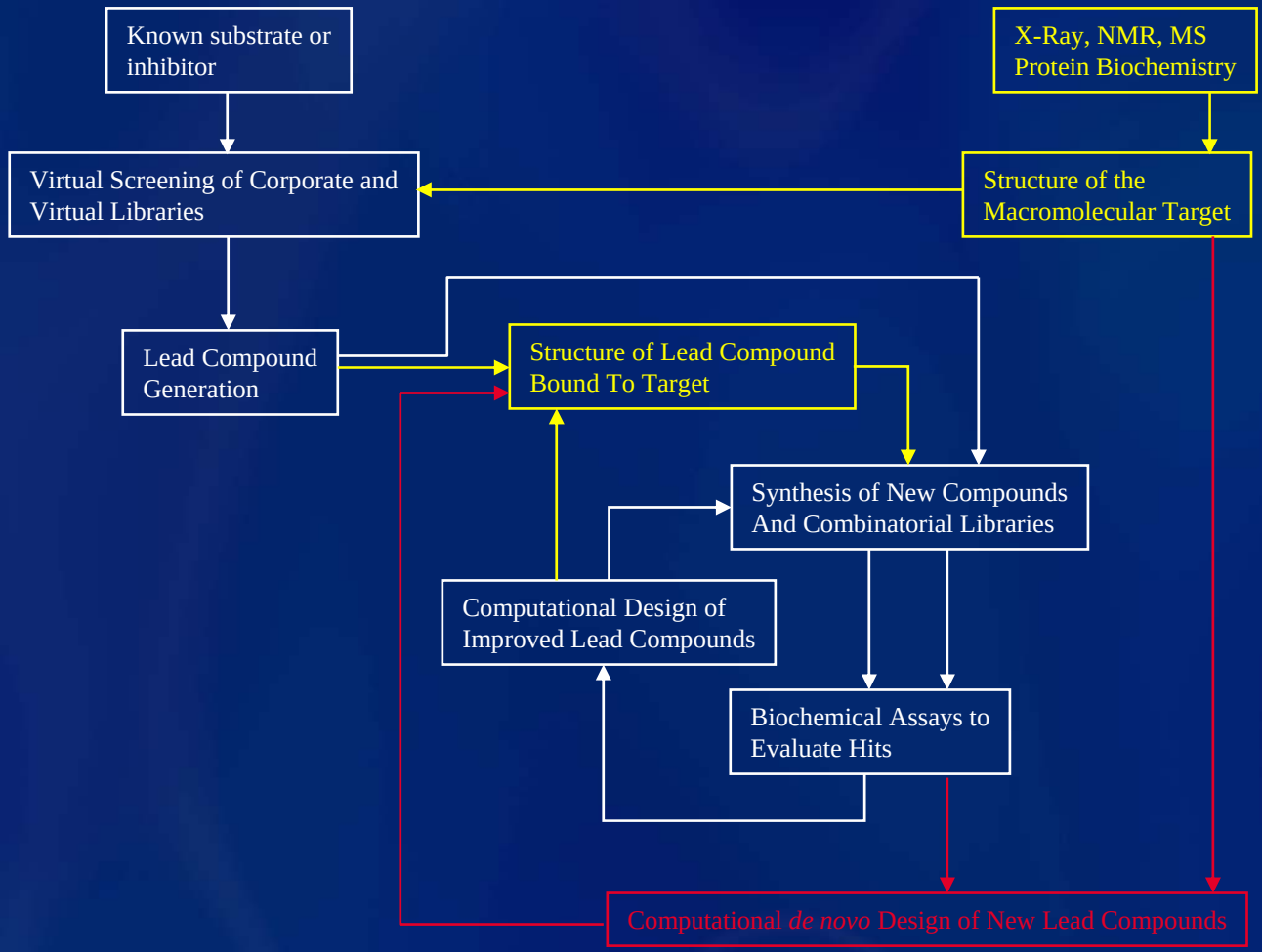
- Selection and determination of a feasible receptor target
- Finding an appropriate “key” for the “lock”
- Synthetic feasibility
- Biological effectiveness
- Bioavailablity (ADME)
- Toxicity
- Patents
- Financial

Drug Discovery

- Disease / Target Selection (1 – 5 years)
- Lead Discovery (0.5 – 1 year)
- Lead Optimization (2 – 4 years)
- Preclinical Testing / Formulation (1 -2 years)

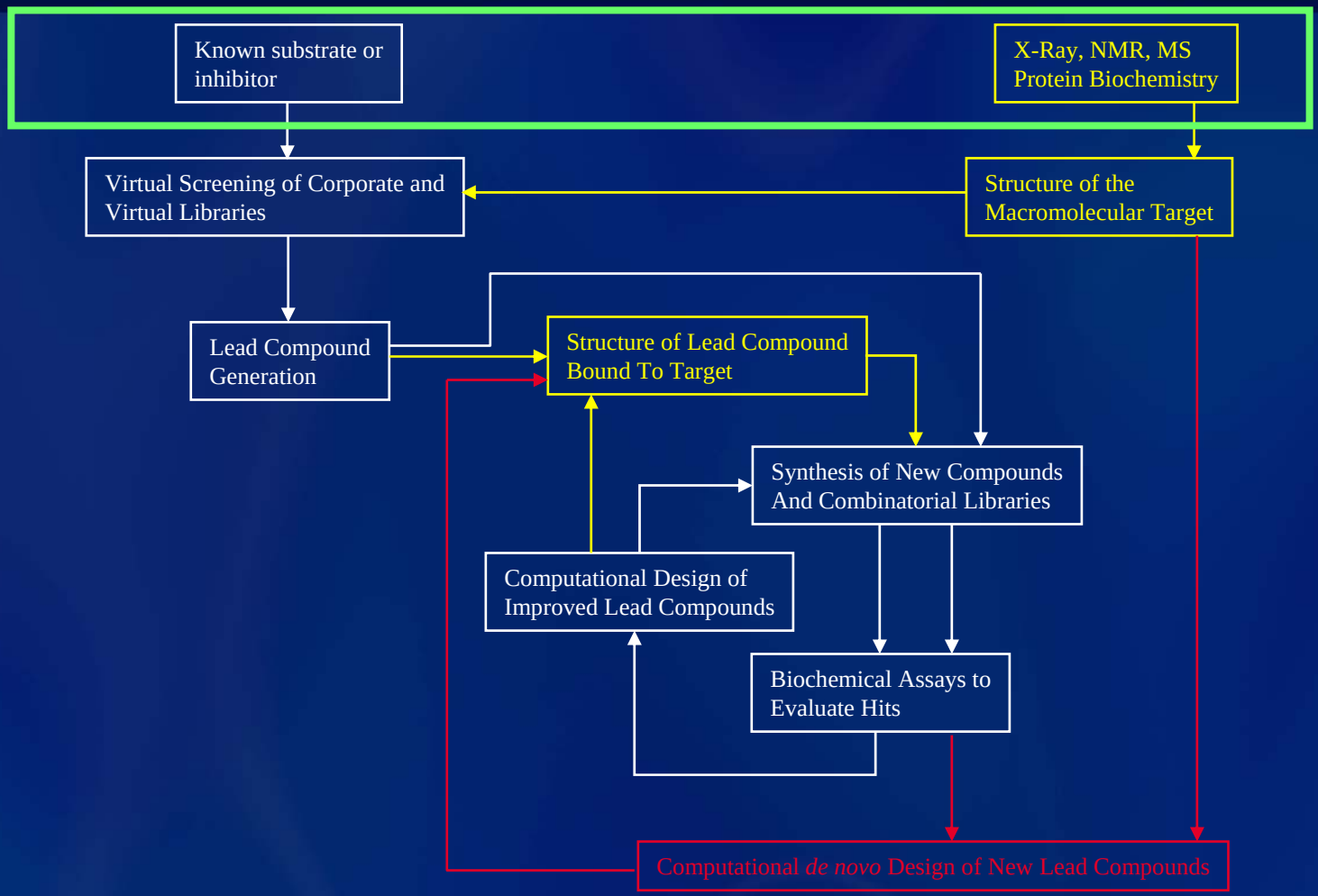
Source: ShipleGroup: Life Sciences Sales Training

Drug Design: Process



Adapted from Figure 22.1: Bourne, Wessig. Structural Bioinformatics.

Disease / Target Selection



Adapted from Figure 22.1: Bourne, Wessig. Structural Bioinformatics.

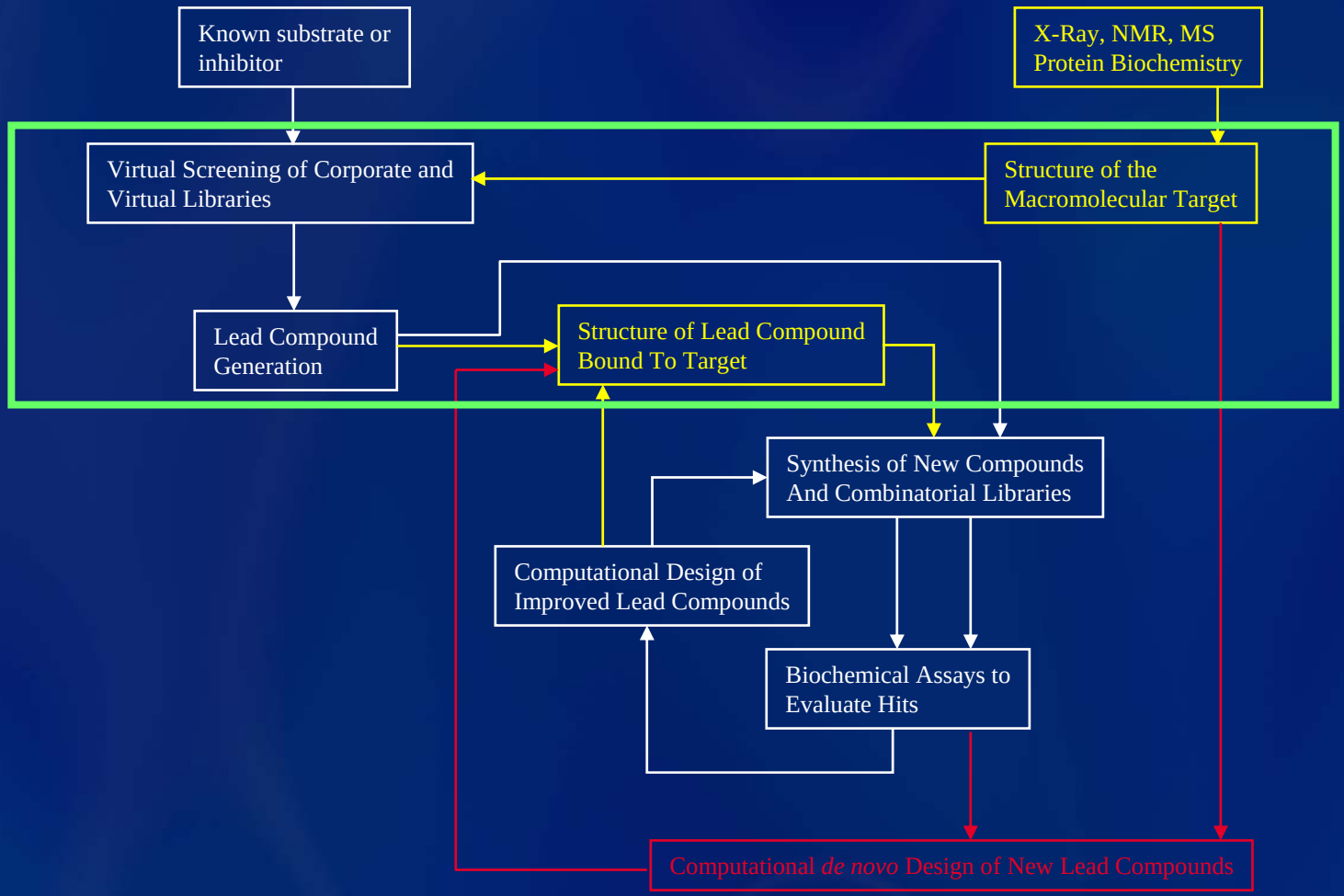
Disease / Target Selection

- Many financial considerations as many drugs never recoup investment costs.
- Best targets from a financial standpoint are chronic conditions with a high societal prevalence such as:
  - Asthma and Allergy
  - Diabetes
  - Hypertension
  - Joint Disease (arthritis)
  - Lipid Disorders (cholesterol and triglycerides)
  - Mental Disorders (depression, ADHD, etc.)
  - Ulcerative and Reflux Disease
- For rare and low prevalence diseases, pharmaceutical companies may apply for *orphan drug* status.
  - Government provides significant tax and marketing incentives
  - http://www.fda.gov/orphan/designat/

Disease / Target Selection: Cont'd

- In addition to financial considerations, there are many biological considerations:
  - The biological process has to be known such that a specific target protein can be selected
  - Must have the ability to produce and purify the protein of interest
  - Must be able to assay the *in vitro* activity of the protein in the setting of individual compounds
  - Must be able to scale-up the assay to allow for compound screening

Lead Discovery



Adapted from Figure 22.1: Bourne, Wessig. Structural Bioinformatics.

Lead Discovery

- Iterative process of screening large numbers of compounds for biological activity
- Screening and testing compounds can be quite expensive
- Use “virtual screening” to reduce the number of compounds actually screened

Virtual Screening

- Process by which computational tools are used to reduce the number actual compounds screened
- Pharmaceutical companies and chemical vendors have very large databases of chemical compounds (millions) available for screening
- Goal is to reduce the number needed to screen while:
  - Increasing the probability of biological activity
  - Increasing the probability of oral absorption
  - Decreasing the probability of toxicity
  - Decreasing the number of false positives

Filtering: Lipinski's "Rule of Five"

- Seminal paper published in 1997 describing commonly found features of orally active drugs
- Features
  - <= 5 hydrogen bond donors
  - <= 10 hydrogen bond acceptors
  - Molecular weight < 500
  - LogP < 5
    - LogP is a measurement of the hydrophobicity / hydrophilicity of a given compound
    - Actual values are frequently not known
    - Frequently predicted using sum-of-fragment methods, neural networks trained on known structures with measured values, or a combination of these methods
      - Examples include xLogP and cLogP
- Widely extended to include many other properties to help identify "drug-like" molecules

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Advanced Drug Delivery Reviews. 15 January 1997: 23(1). pp 3-25.

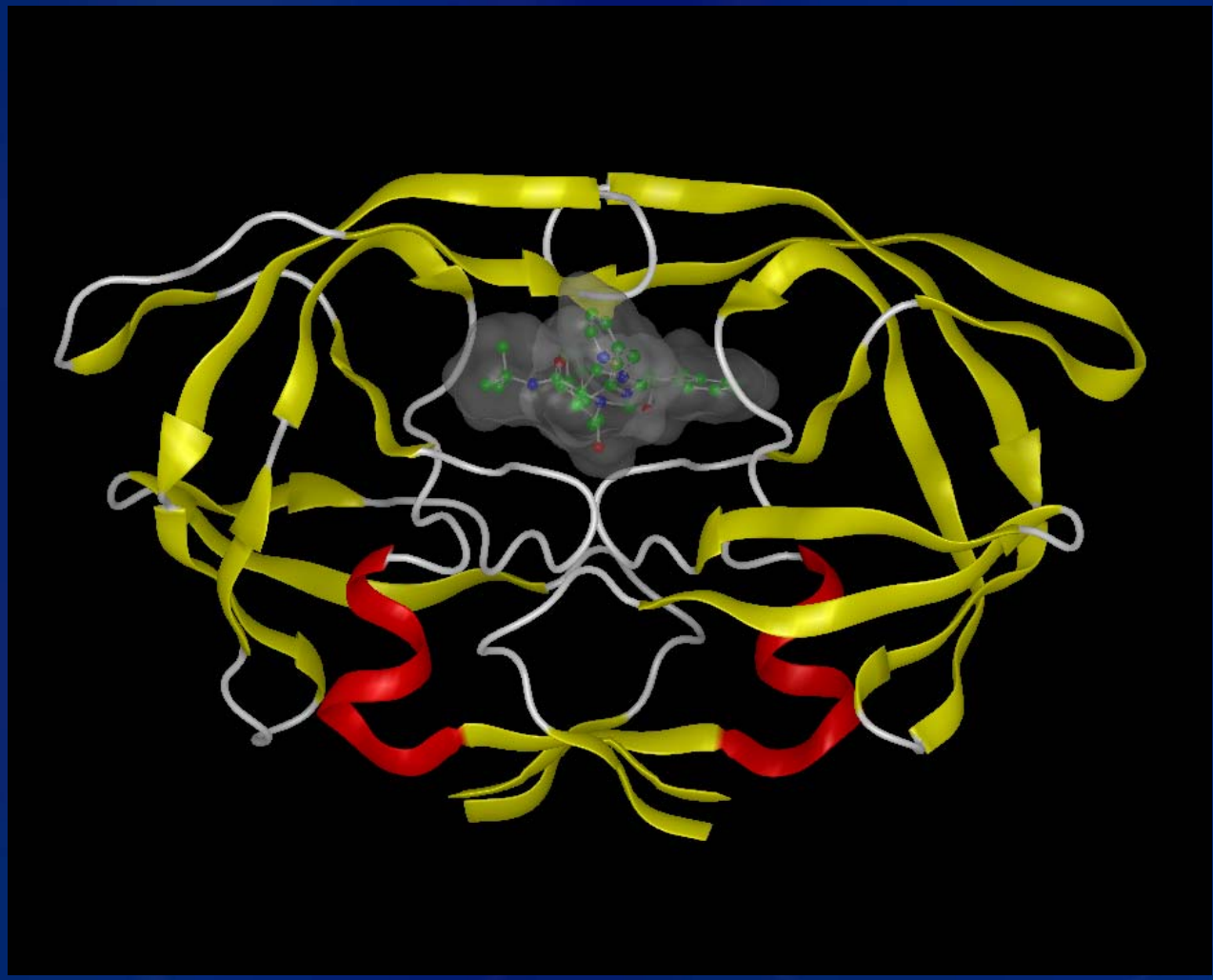
Post-Filtering

- Structure-based design
  - Requires 3D structure of target protein
  - Requires knowledge of location of ligand binding site
  - Example: HIV protease
- Ligand-based design
  - Requires known substrates or inhibitors of target
  - Example: HIV reverse transcriptase
- Random screening
  - Requires time, money, patience, and serendipity

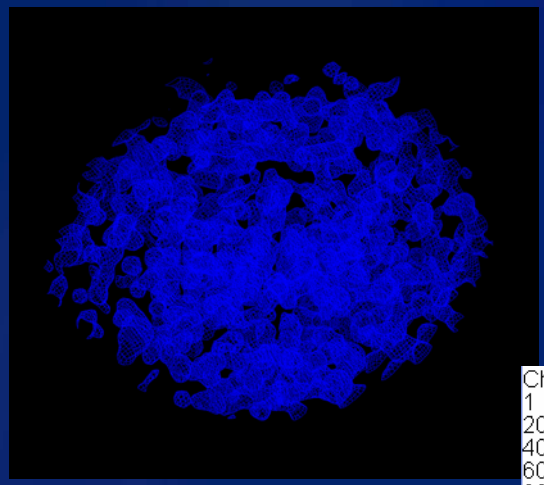
Structure-Based Design Example: HIV Protease

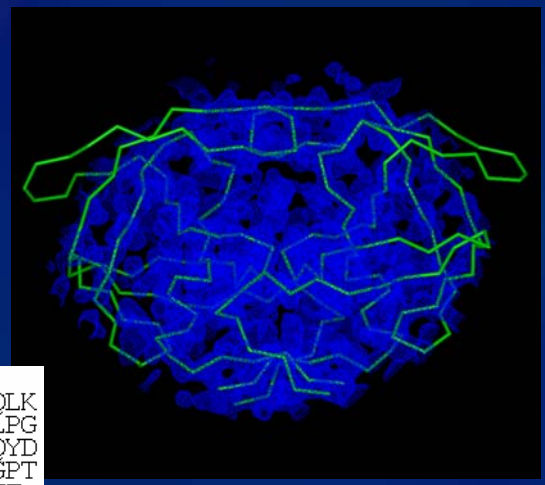
- HIV-1 protease was one among many potential proteins involved in the life-cycle of the HIV-1 virus
- HIV-1 protease was determined to be unique to HIV and absolutely required for replication
- As a result, there was a significant effort made to rapidly solve the crystal structure of this enzyme which helped elucidate the mechanism
- Using this information, a large number of protease inhibitors have been developed to date

Structure-Based Design Example: HIV Protease Inhibitor (Indinavir)

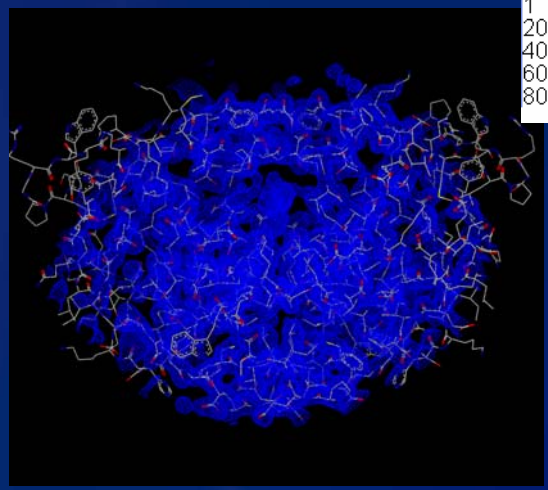


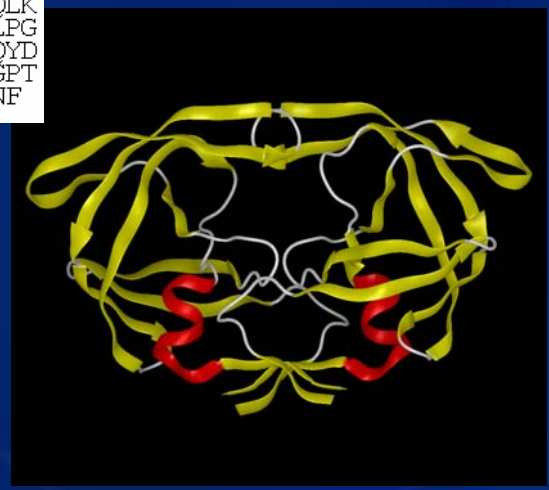
Structure-Based Design: Structure Determination



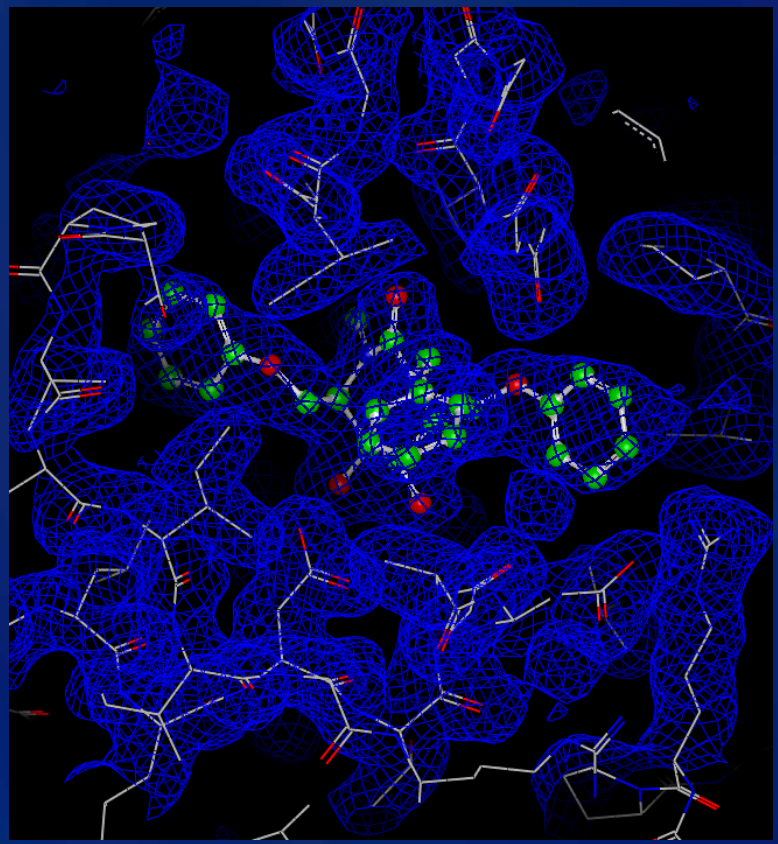


Chain A
1 PQITLWQRPLVTIKIGGQLK
20 EALLDTGADDTVLEEMSLPG
40 RWKPKMIGGIGGFIKVRQYD
60 QILIEICGHKAIGTVLVGPT
80 PVNIIGRNLLTQIGCTLNF
Chain B
1 PQITLWQRPLVTIKIGGQLK
20 EALLDTGADDTVLEEMSLPG
40 RWKPKMIGGIGGFIKVRQYD
60 QILIEICGHKAIGTVLVGPT
80 PVNIIGRNLLTQIGCTLNF

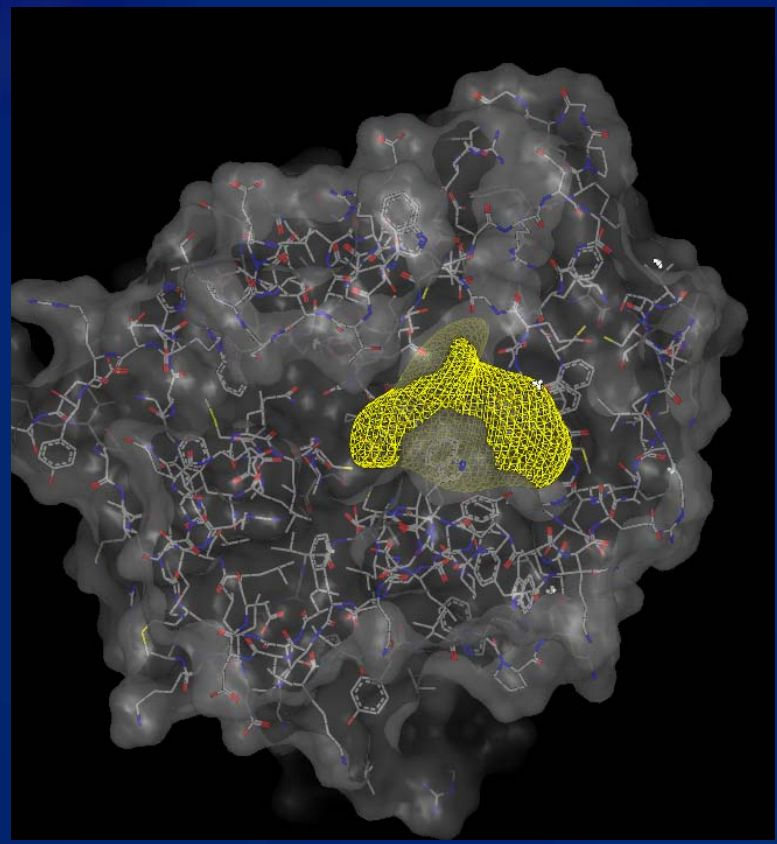




Structure-Based Design: Binding Site Determination



Co-crystallization of ligand

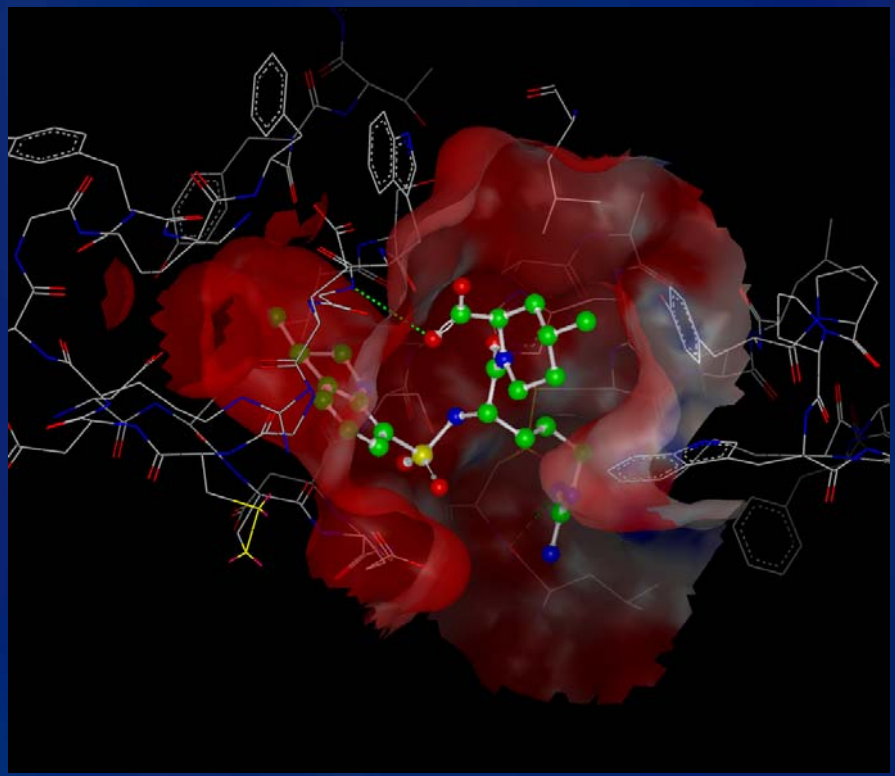


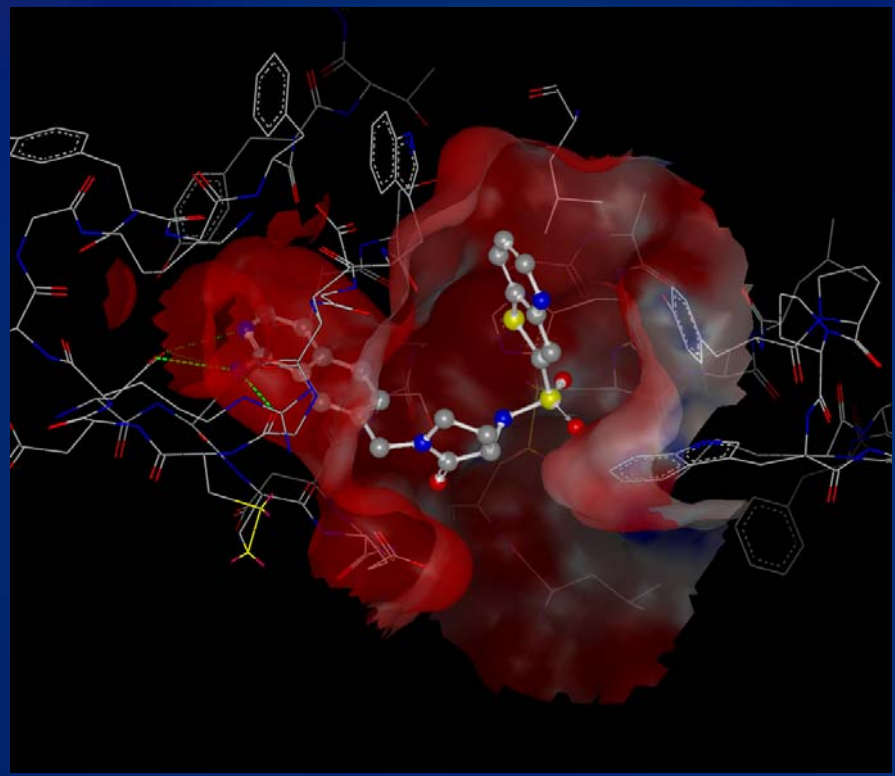
Computational methods

Structure-Based Design: Docking

- Docking is a widely used methodology for selecting compounds for initial screening
- It is now possible to systematically search all potential docking poses of a given compound in a reasonable period of time
- Docking's primary limitations:
  - Scoring functions
  - Protein flexibility

Docking: Continued





Ligand-Based Design

- Select potential leads based on other compounds that are known bind to the protein
- Natural substrate
  - Looking for compounds that bind in a similar fashion thus preventing binding of the natural substrate
- Known inhibitors / competitors' drugs
  - Looking for compounds that bind in a similar fashion but with stronger binding capability and sufficiently different structure / chemistry to avoid patent infringement

Ligand-Based Design Example: HIV Reverse Transcriptase

- HIV is a retrovirus, meaning that it's genetic material is RNA (not DNA) and that it contains an enzyme *reverse transcriptase* which converts RNA to DNA

- When HIV infects a human T-cell, it's RNA is converted to DNA (via *reverse transcriptase*) which is then incorporated into the genetic material of the cell enabling continued reproduction of the virus

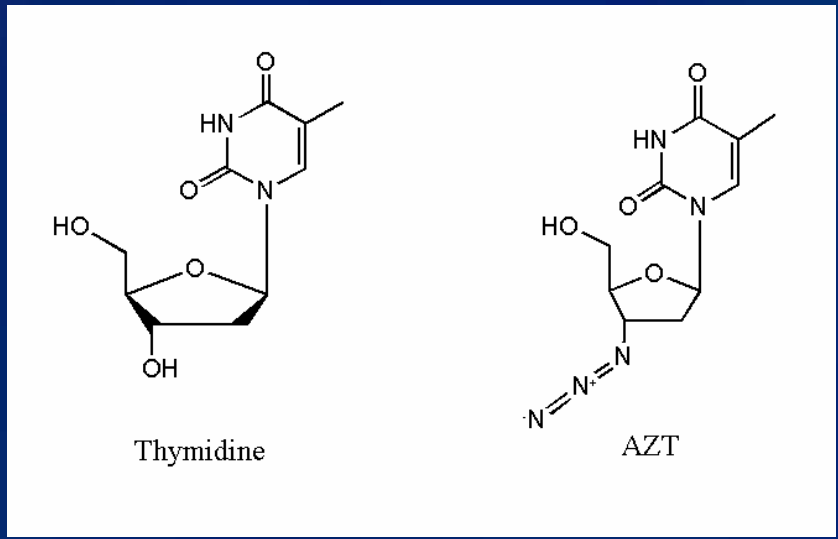
- Given this knowledge, *reverse transcriptase* was an obvious target for drug design

Ligand-Based Design Example: HIV Reverse Transcriptase Cont'd

- As *reverse transcriptase* produces DNA, its substrates were obvious: the four DNA nucleosides – adenosine, cytidine, guanosine, thymidine

- It was hypothesized that creation of an analogue of one of these nucleosides might inhibit *reverse transcriptase* thus preventing viral reproduction

Ligand-Based Design Example: HIV Reverse Transcriptase Cont'd



Ligand-Based Design Methodologies

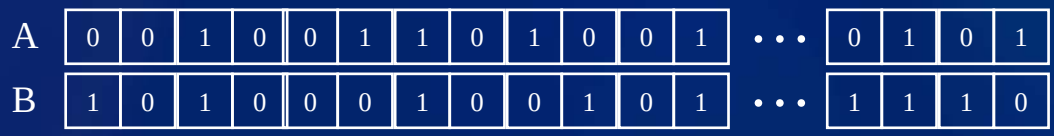
- 2D Similarity
  - Fingerprint-based
- 3D Similarity
  - Pharmacophores
  - Shape
  - Electrostatics

2D Similarity: Fingerprints

- A molecular descriptor is a numerical or binary value for a given structure
  - 1D: logP, molecular weight, num atoms
  - 2D: graph-based, functional groups
- A fingerprint is a collection of molecular descriptors for a given molecule
- Can be time consuming to calculate, so are often calculated once and stored in a database

2D Similarity: Fingerprints

- Fingerprint searching is fast with binary fingerprints



- Similarity is assessed by calculating a Tanimoto coefficient

$$T(a,b) = \frac{N(a \cap b)}{N(a) + N(b) - N(a \cap b)}$$

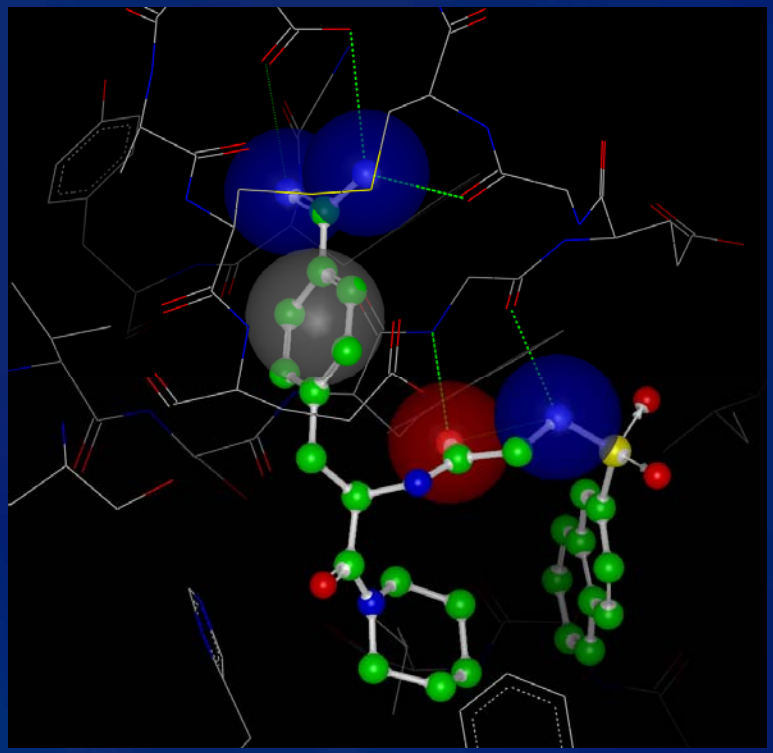
- Tanimoto has a range [0,1]

3D Similarity: Pharmacophores

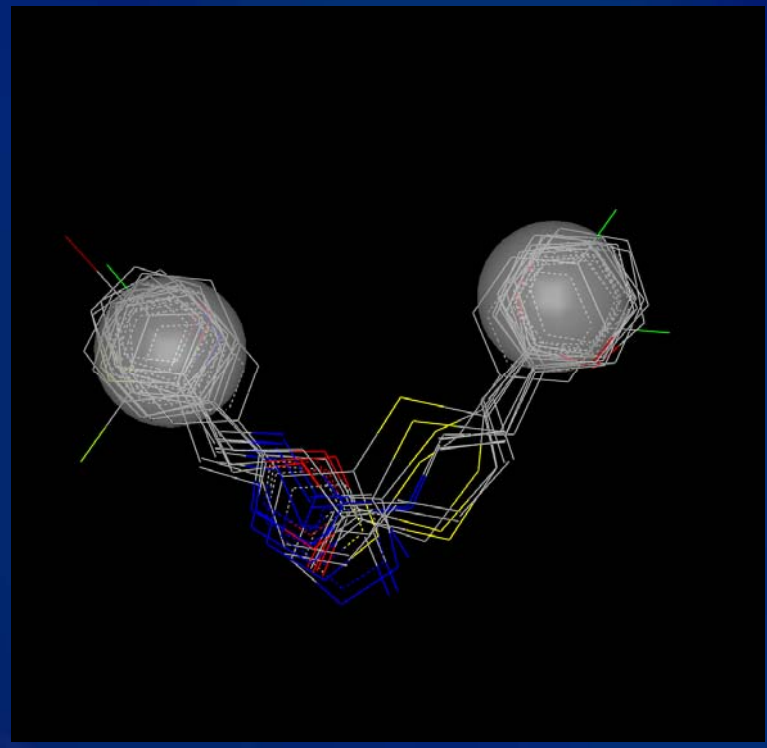
- Traditional definition is the minimum functionality a molecule has to contain in order to exhibit activity
- Typically defined in terms of atoms or centers which can interact with a receptor and are categorized into six types:
  - Hydrogen bond acceptors
  - Hydrogen bond donors
  - Anion
  - Cations
  - Aromatic ring centers
  - Hydrophobic ring centers

3D Similarity: Pharmacophores

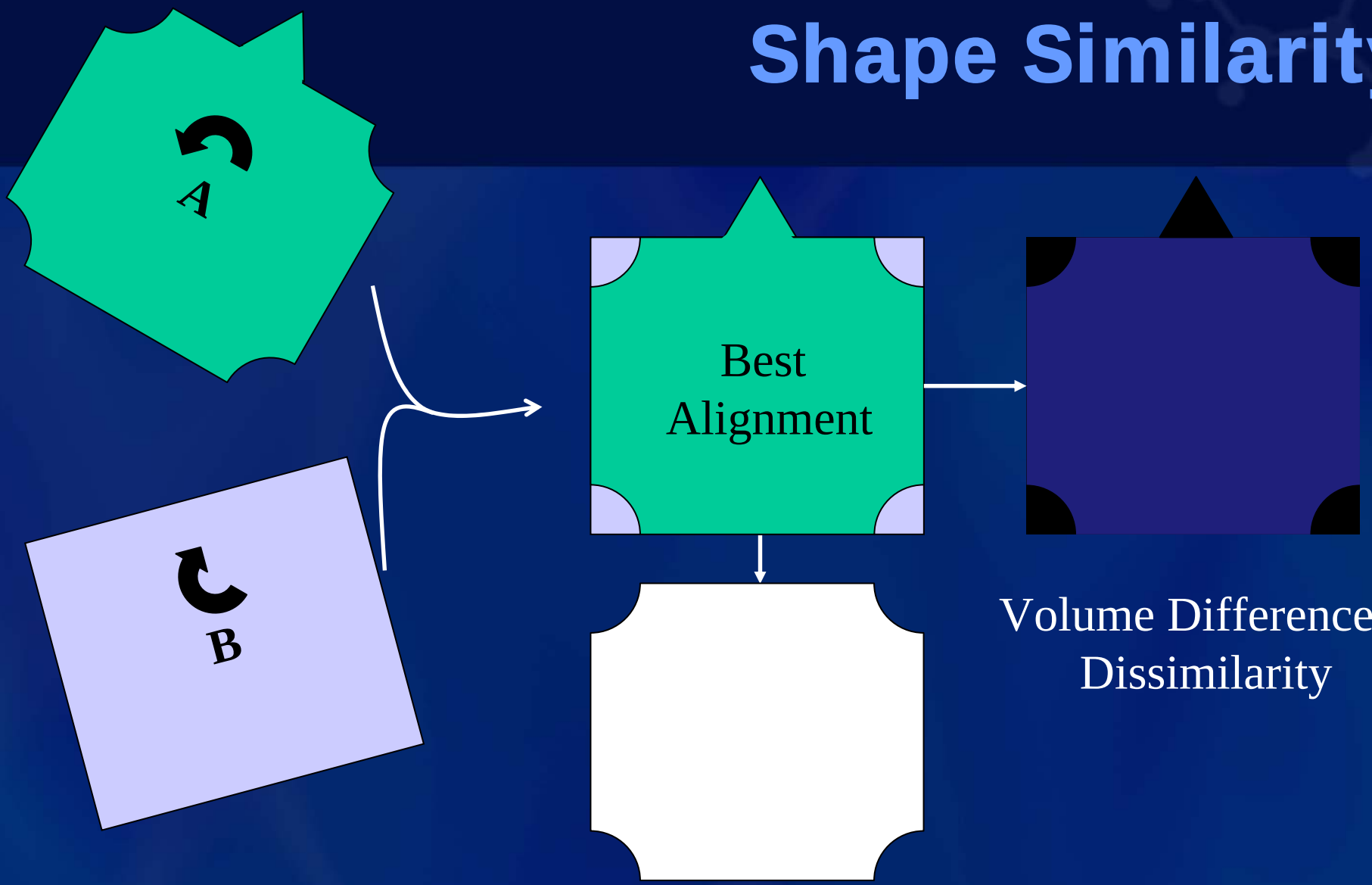
Receptor based



Ligand based



Shape Similarity



Shape: Mathematically...

$$\Delta^2 = \int_{\text{all space}} (\chi_1 - \chi_2)^2 dv$$

$\chi_1 = defines \quad shape \quad 1$

$\chi_2 = defines \quad shape \quad 2$

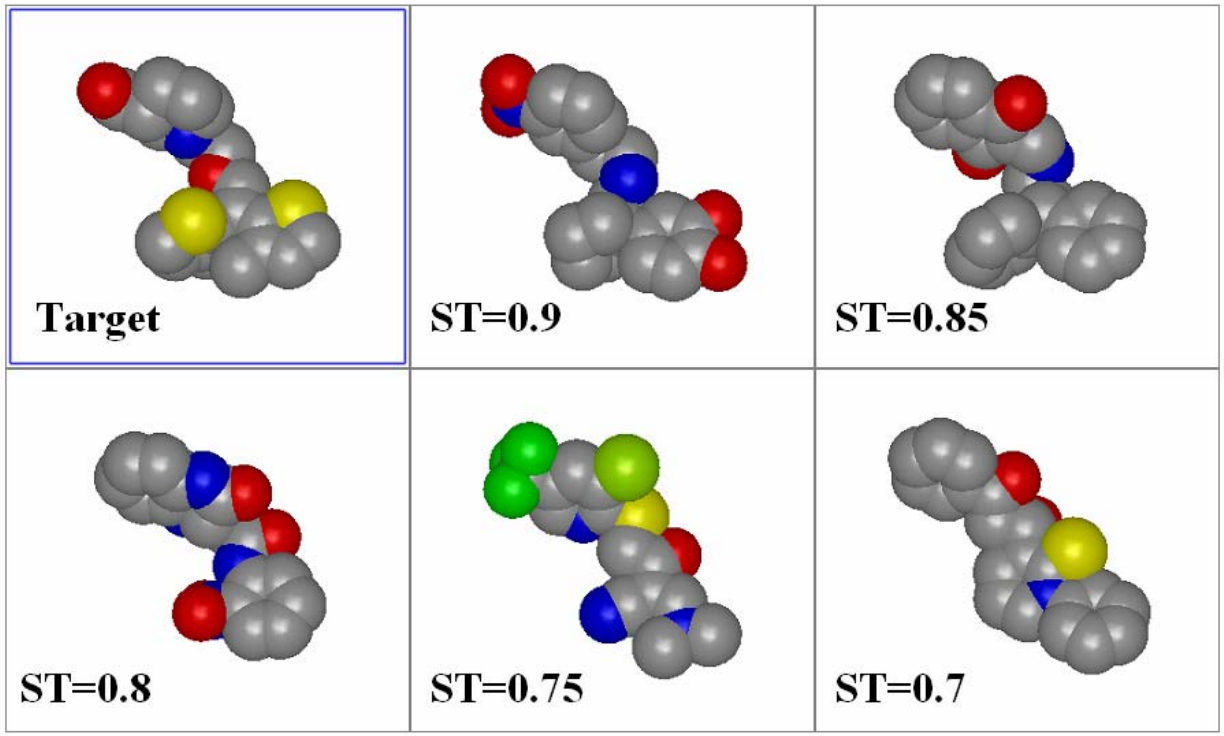
$\Delta = shape \quad distance$

Shape: Tanimoto (ST)

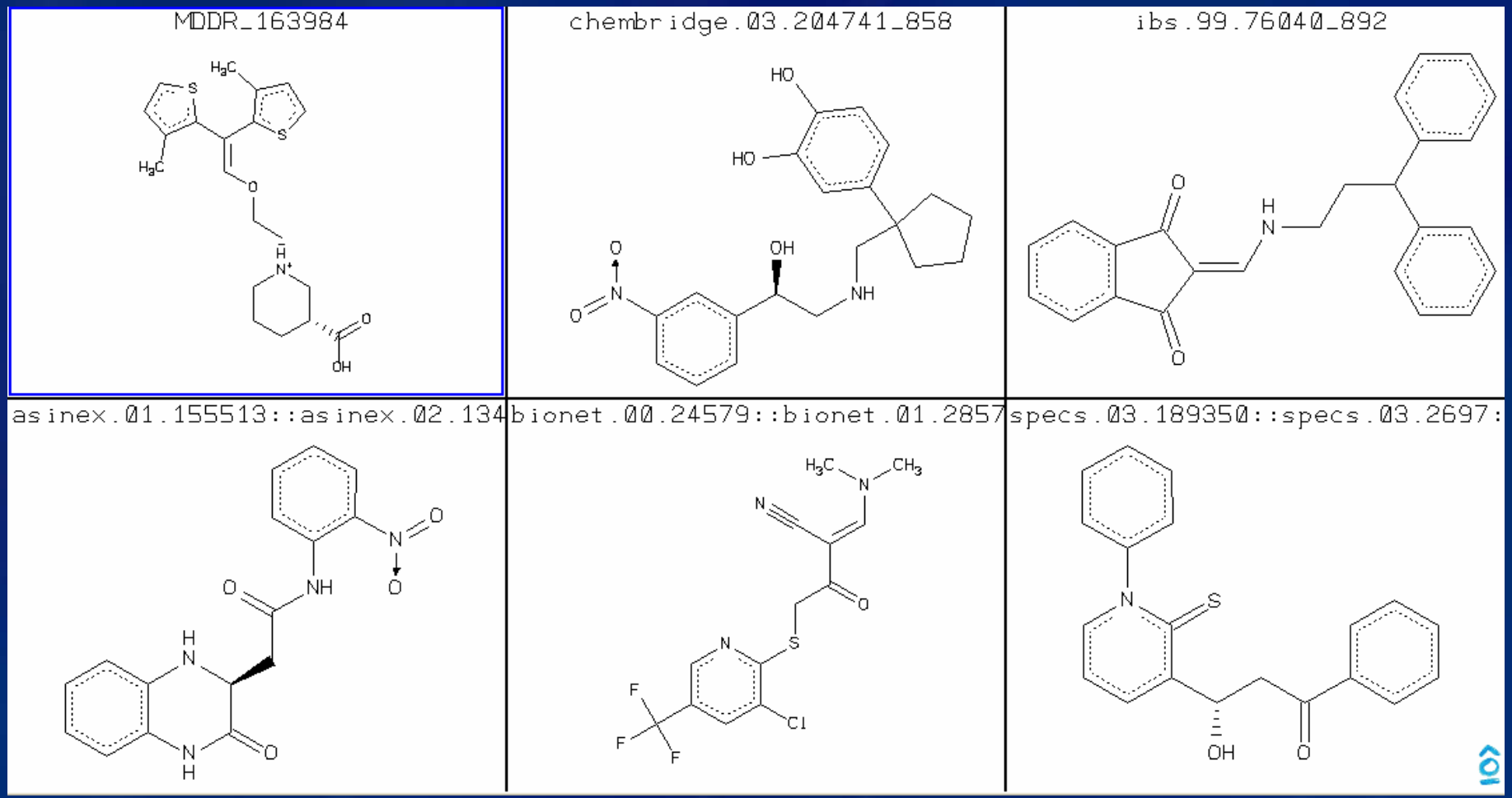
$$\frac{Overlap(A,B)}{Overlap(A,A)+Overlap(B,B)-Overlap(A,B)}$$

- Larger Tanimoto = More similar = Better
- Smaller Tanimoto = Less similar = Worse
- Range = [0,1]
- Value understood, e.g. T>0.75 = Shape similar

Shape Similarity



Chemical Dissimilarity



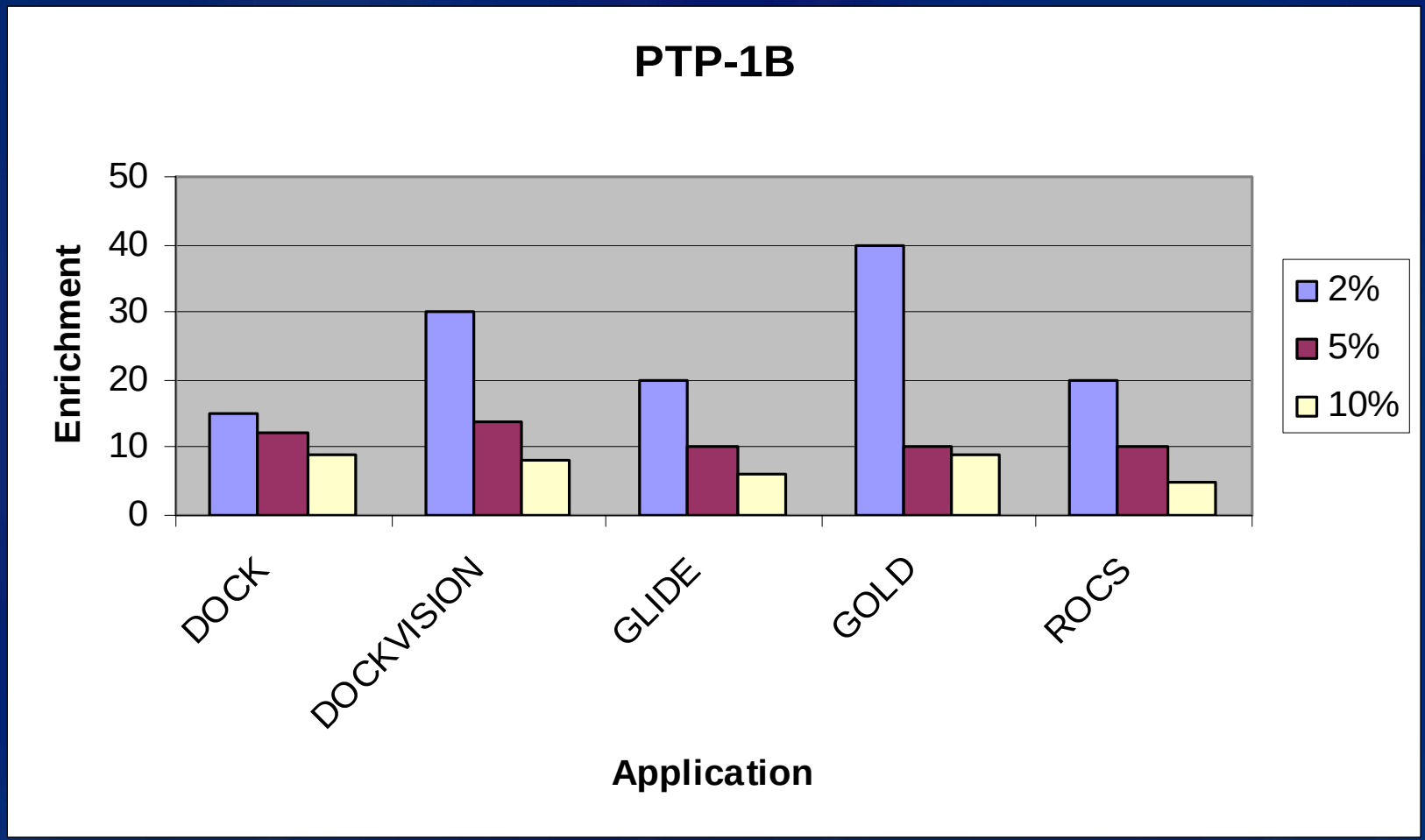
Docking: Revisited

- How does ligand-based design compare with structure-based design?
- Compared OpenEye's ligand shape-based search tool (ROCS) with common docking programs
- Docking results obtained from published papers

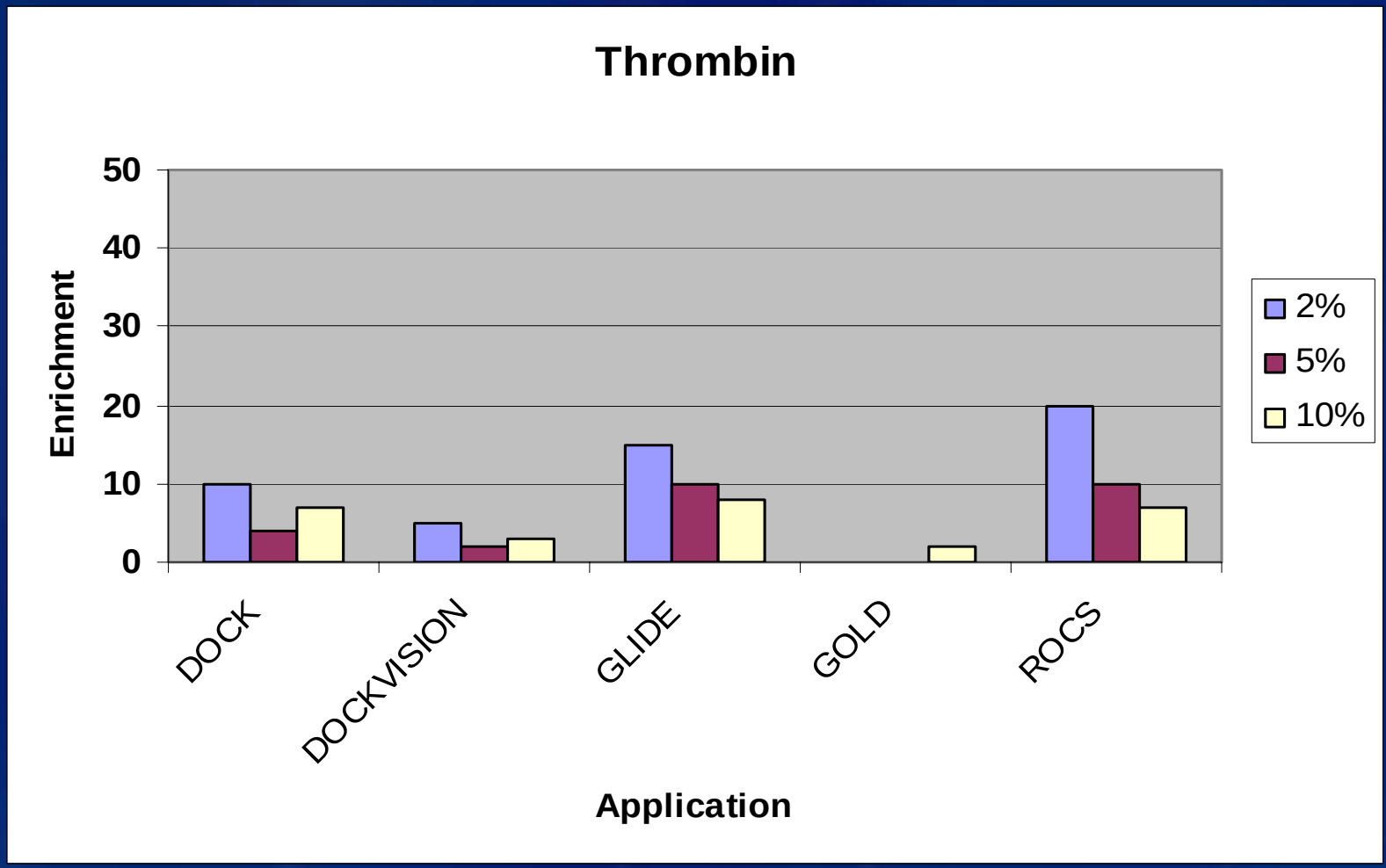
Docking: First Study

- Cummings *et al., J. Med. Chem.*, **2005**, *48*, 962
  - compared 4 programs on 3 public and 2 proprietary datasets
  - DOCK, Dockvision, GOLD, GLIDE
  - HIV-PR, PTP-1B, thrombin
  - Decoys randomly selected from MDDR
  - Decoys and actives publicly available

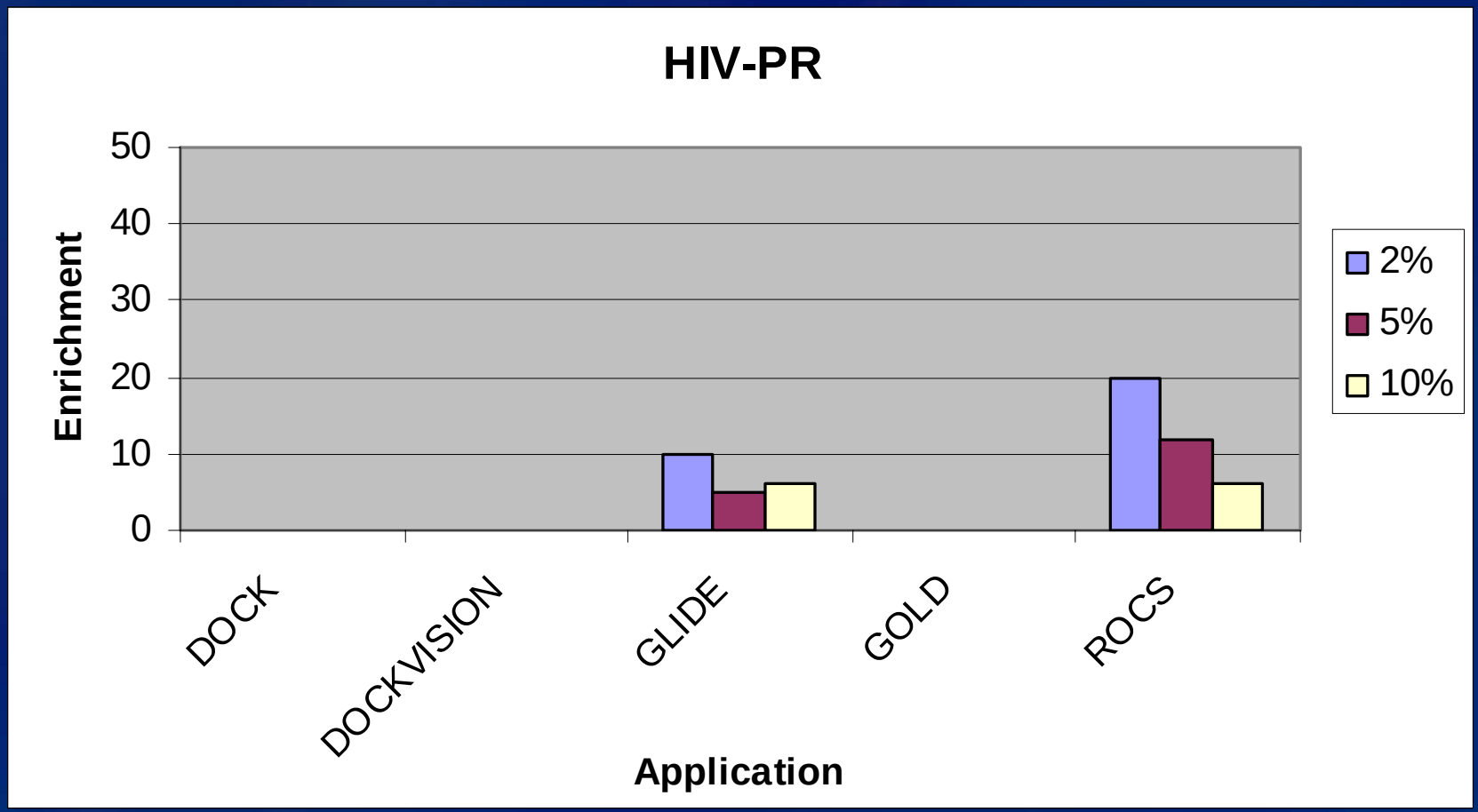
PTP-1B


PTP-1B
Enrichment
50
40
30
20
10
0
DOCK
DOCKVISION
GLIDE
GOLD
ROCS
Application
2%
5%
10%

Thrombin


Thrombin
Enrichment
50
40
30
20
10
0
DOCK
DOCKVISION
GLIDE
GOLD
ROCS
Application
2%
5%
10%

HIV-Protease



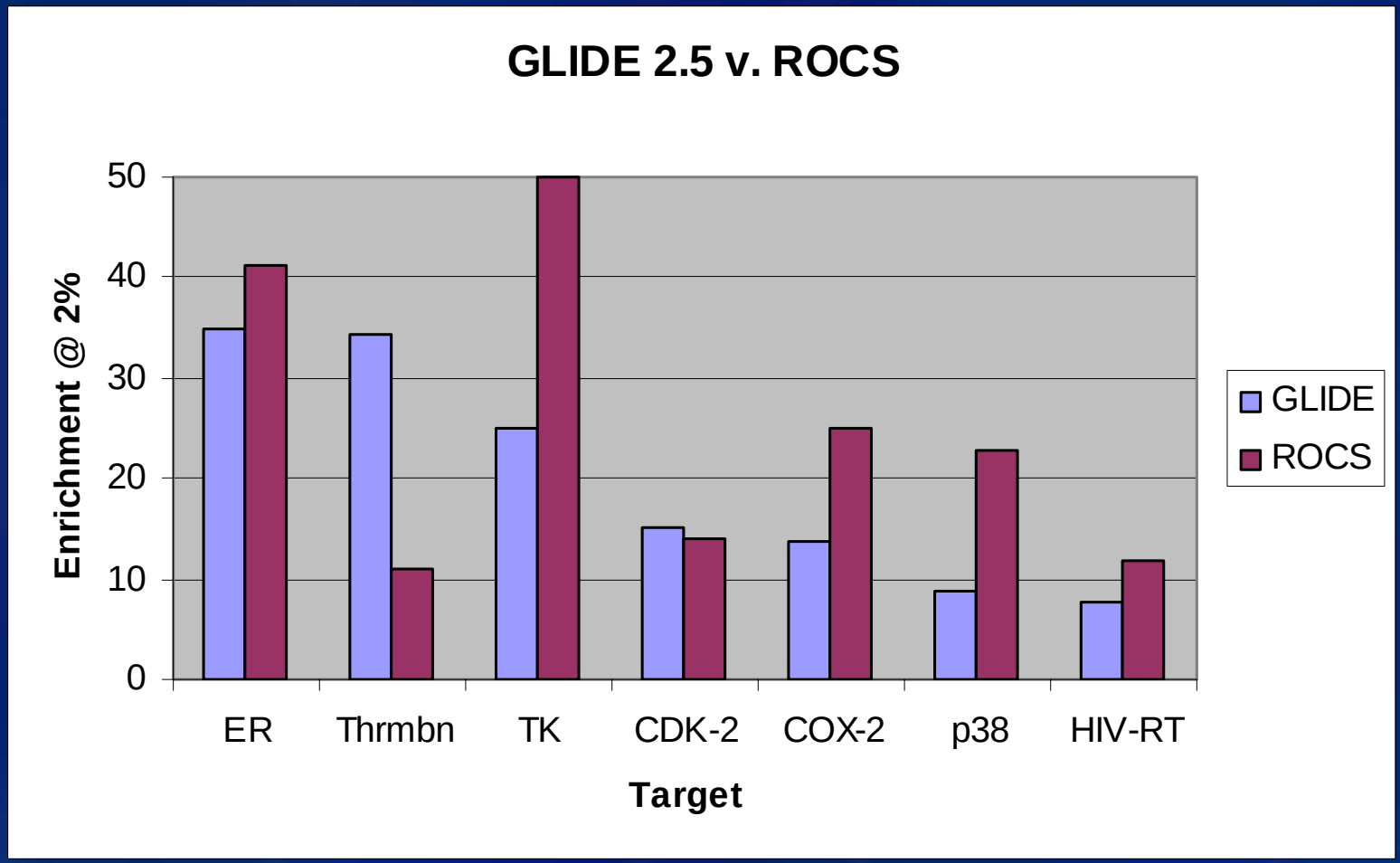
First Study: Conclusion

- Docking tools have inconsistent performance
- Ligand-based shape techniques shows consistent performance

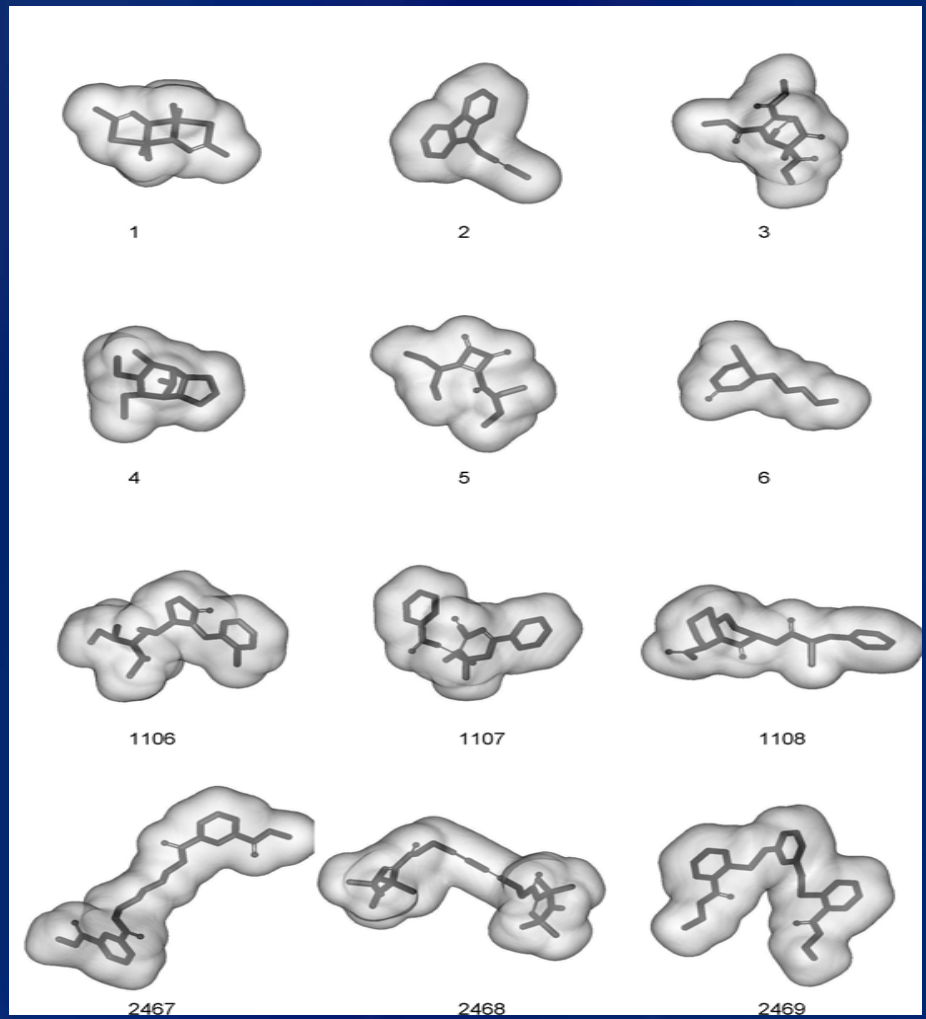
Docking: Second Study

- Halgren *et al.*, *J. Med. Chem.*, **47**, 1750 (2004)
  - GLIDE 2.5
  - Thymidine kinase, ER, CDK-2, p38, HIV-PR, thrombin, thermolysin, COX-2, HIV-RT
  - Decoys chosen to match a property profile
- No compounds publicly available

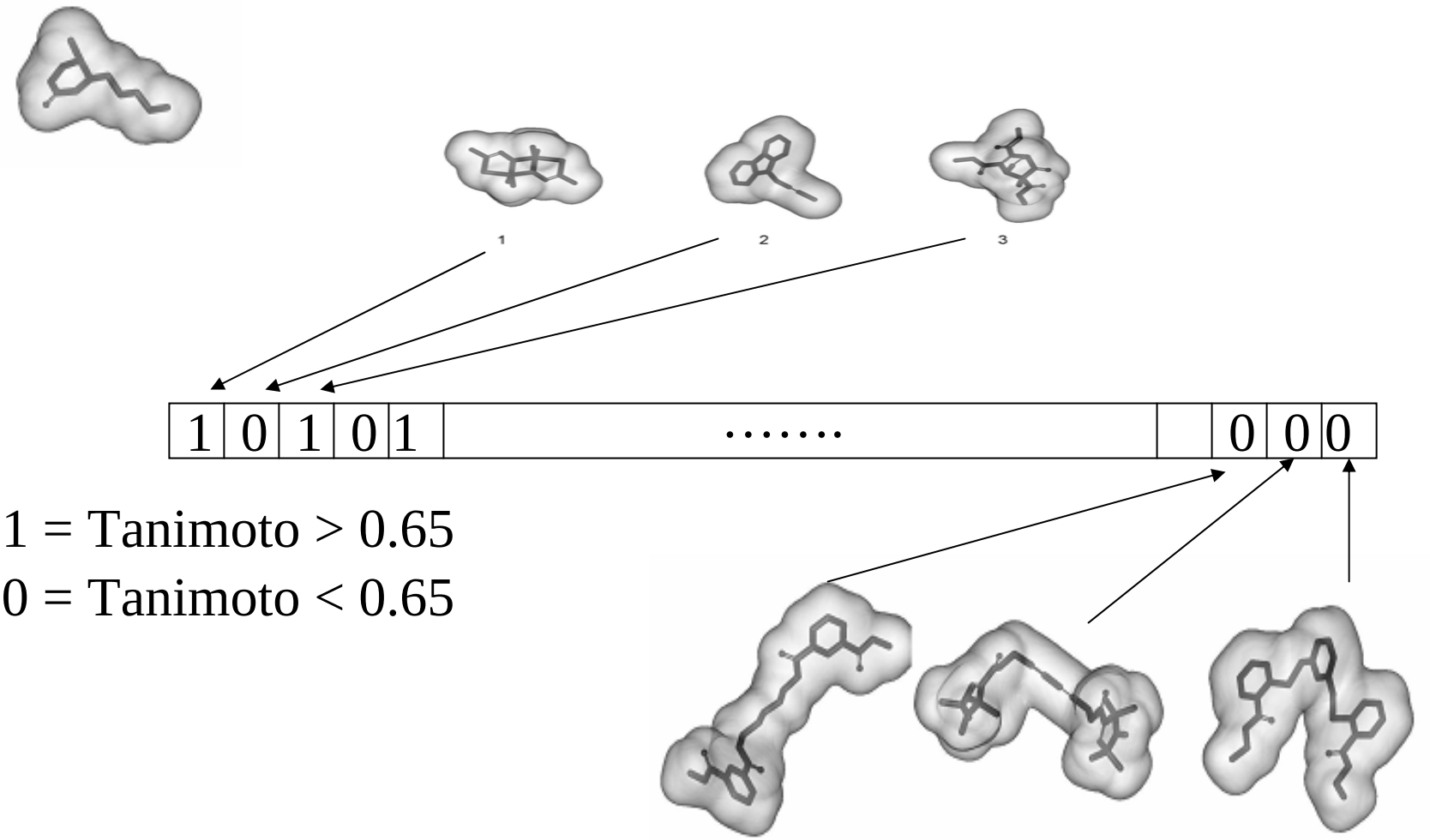
ROCS & GLIDE


GLIDE 2.5 v. ROCS
Enrichment @ 2%
50
40
30
20
10
0
ER
Thrmbn
TK
CDK-2
COX-2
p38
HIV-RT
Target
GLIDE
ROCS

Representative Shapes



Binary Shape Fingerprints



1 = Tanimoto > 0.65

0 = Tanimoto < 0.65

Electrostatic Similarity

$$\Delta^2 = \int_{\text{all space}} (\chi_1 - \chi_2)^2 dv$$

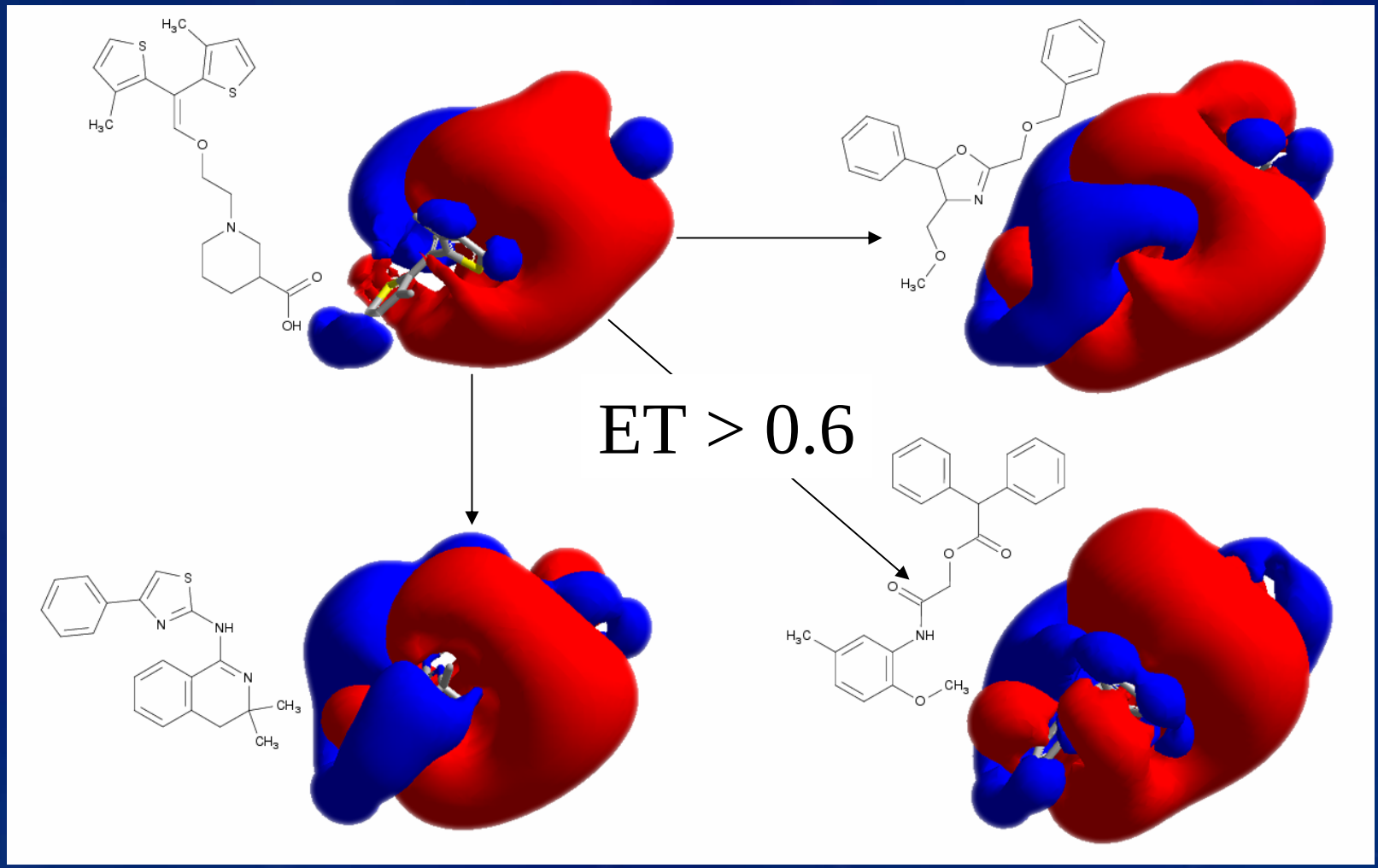
$$\Theta^2 = \int_{\text{all space}} (\phi_1 - \phi_2)^2 dv$$

$\phi_1 = defines \quad potential \quad field \quad 1$

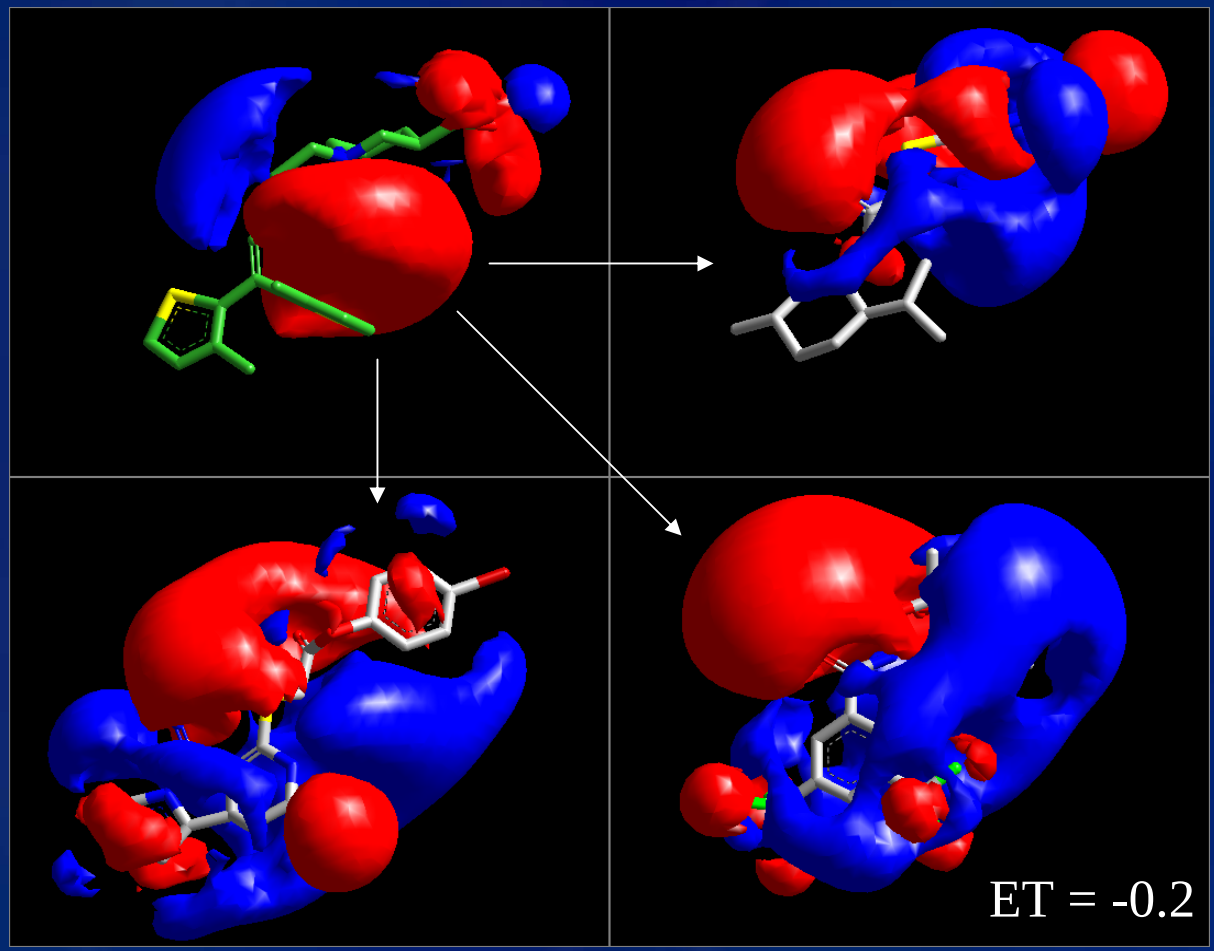
$\phi_2 = defines \quad potential \quad field \quad 2$

$\Theta = Electrostatic\ Distance$

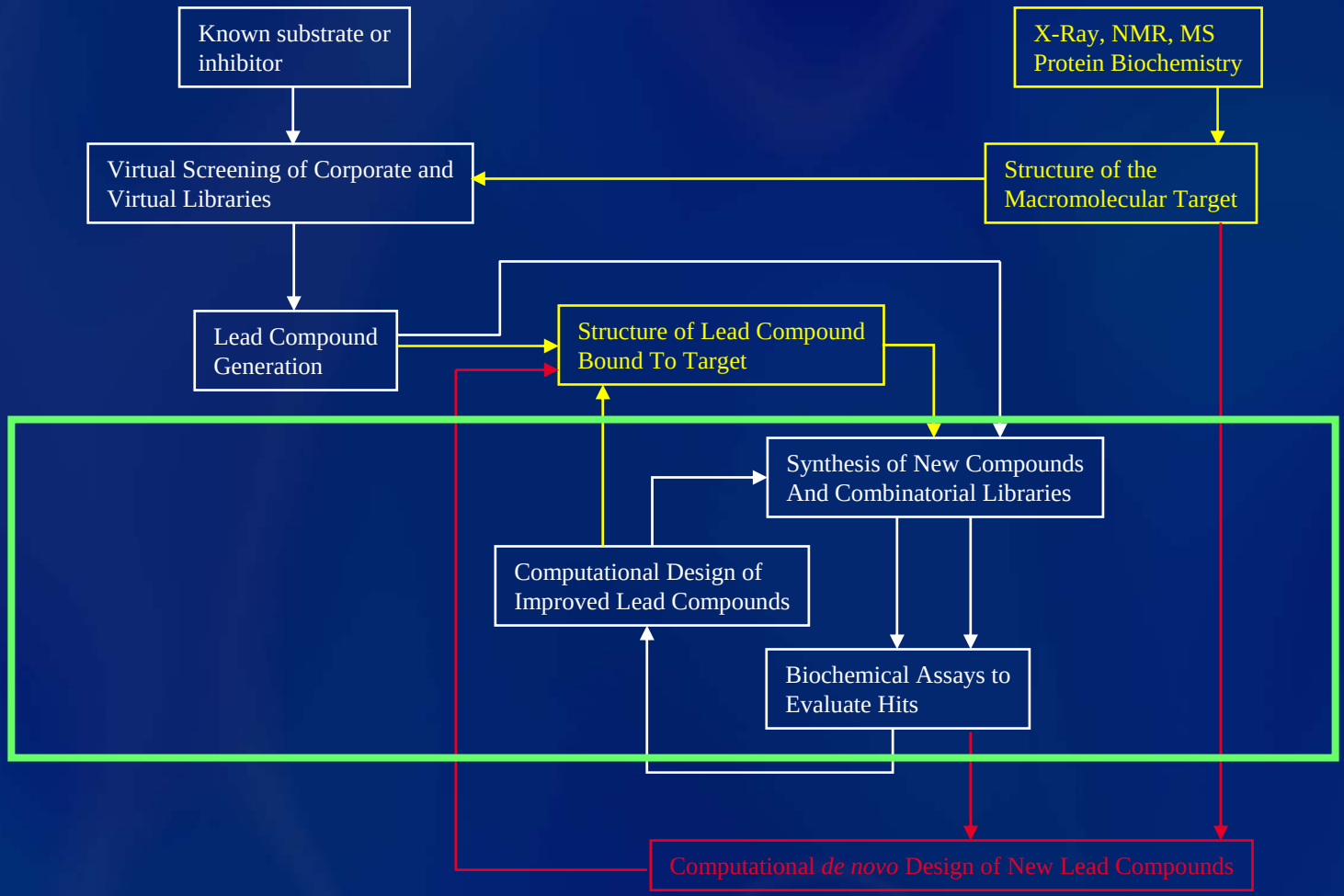
Electrostatic Tanimoto: Good



Electrostatic Tanimoto: Bad



Lead Optimization



Adapted from Figure 22.1: Bourne, Wessig. Structural Bioinformatics.

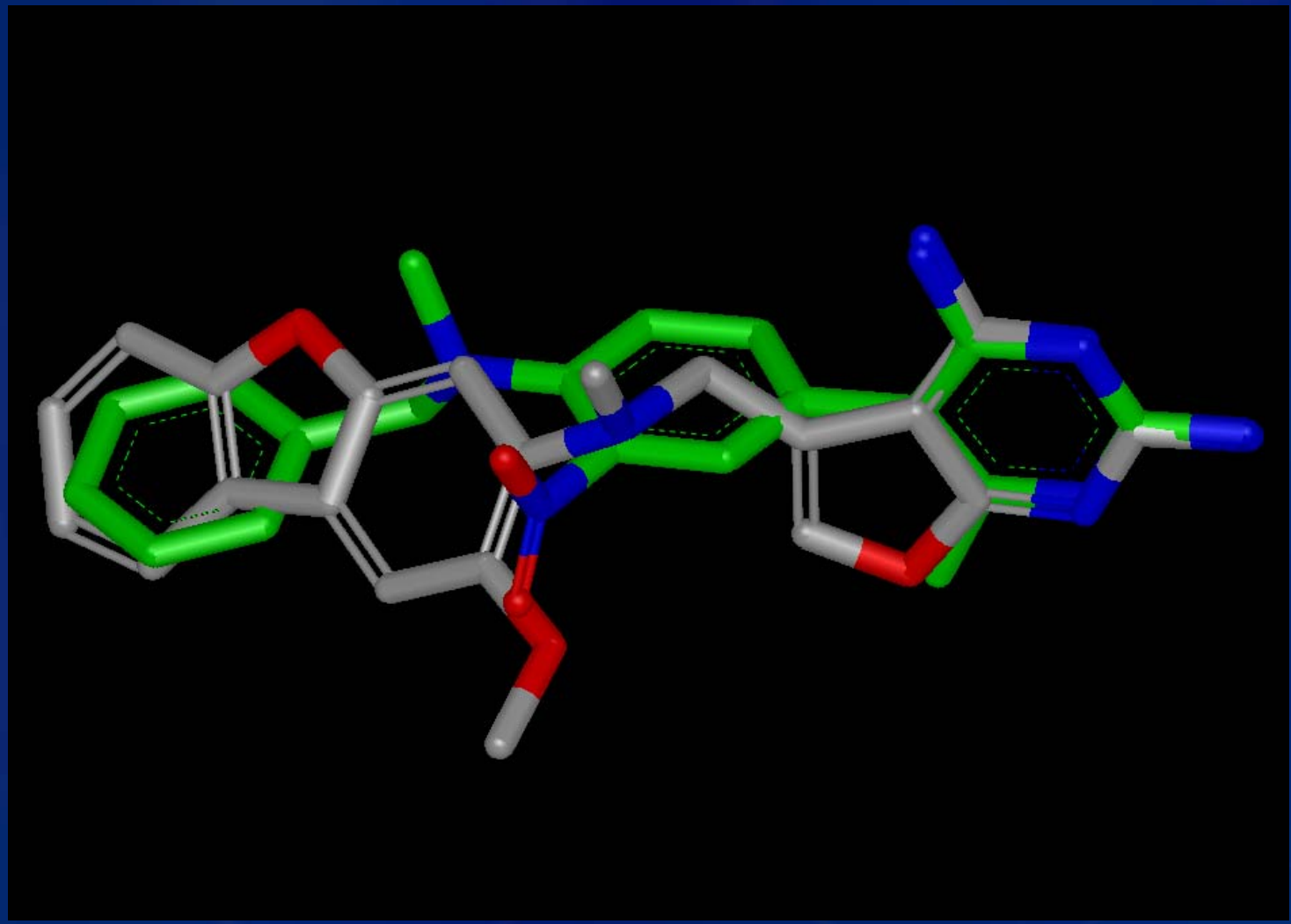
Lead Optimization

- Leads discovered in the first stage are rarely potent enough to be considered drug candidates
- Initial candidates must be optimized for binding affinity and other “drug-like” properties
- Medicinal chemists play a large role in making optimizations based on their experience
- Molecular graphics tools are particularly important in structure based projects to help maximize receptor interactions

Lead Optimization

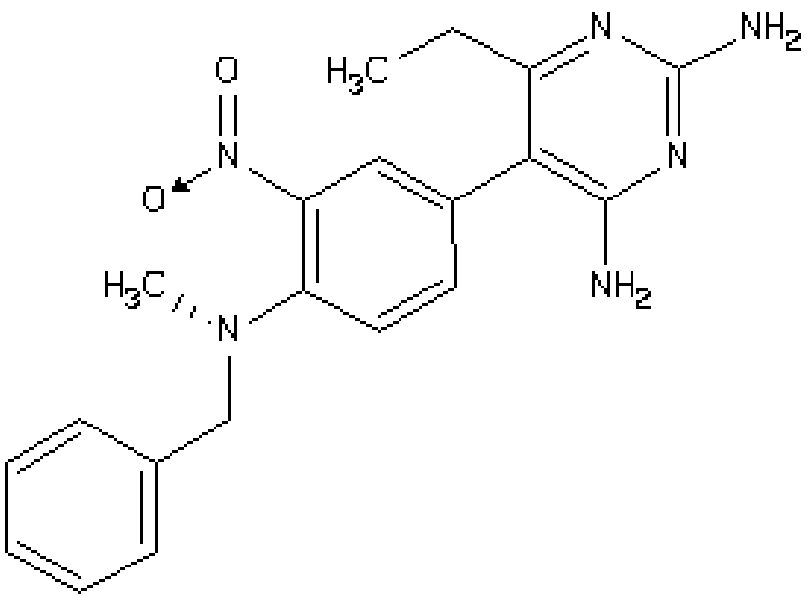
- Look for compounds with same shape (isosteres) but with different chemical and/or electrostatic profiles
- Fix the important parts, but change the rest
- Look for un- or under-utilized regions in the receptor

Scaffold Hopping: Shape



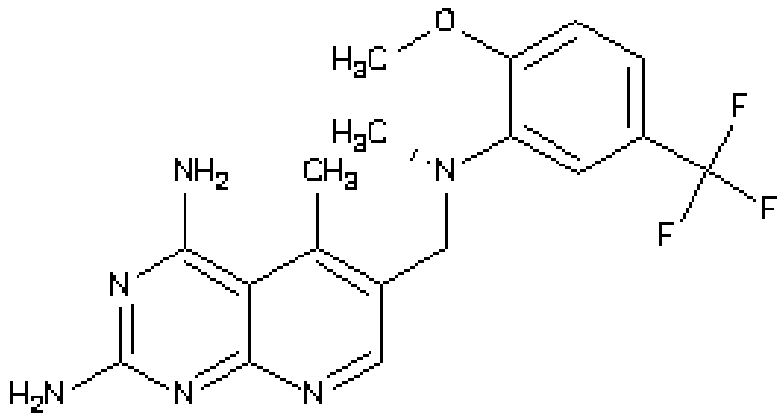
Scaffold Hopping: Shape

IC50 1.6 uM



Query

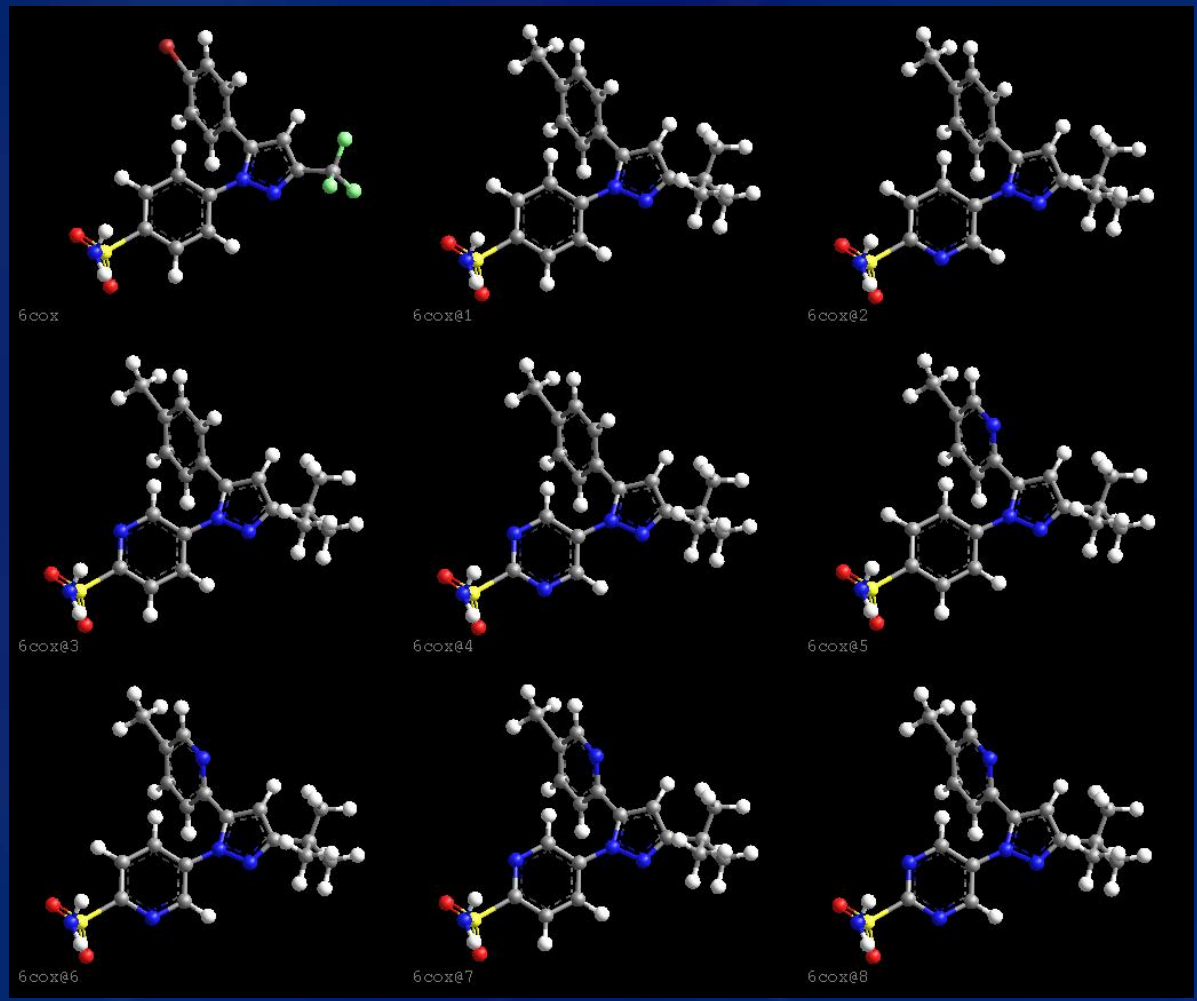
IC50 93 nM



Result

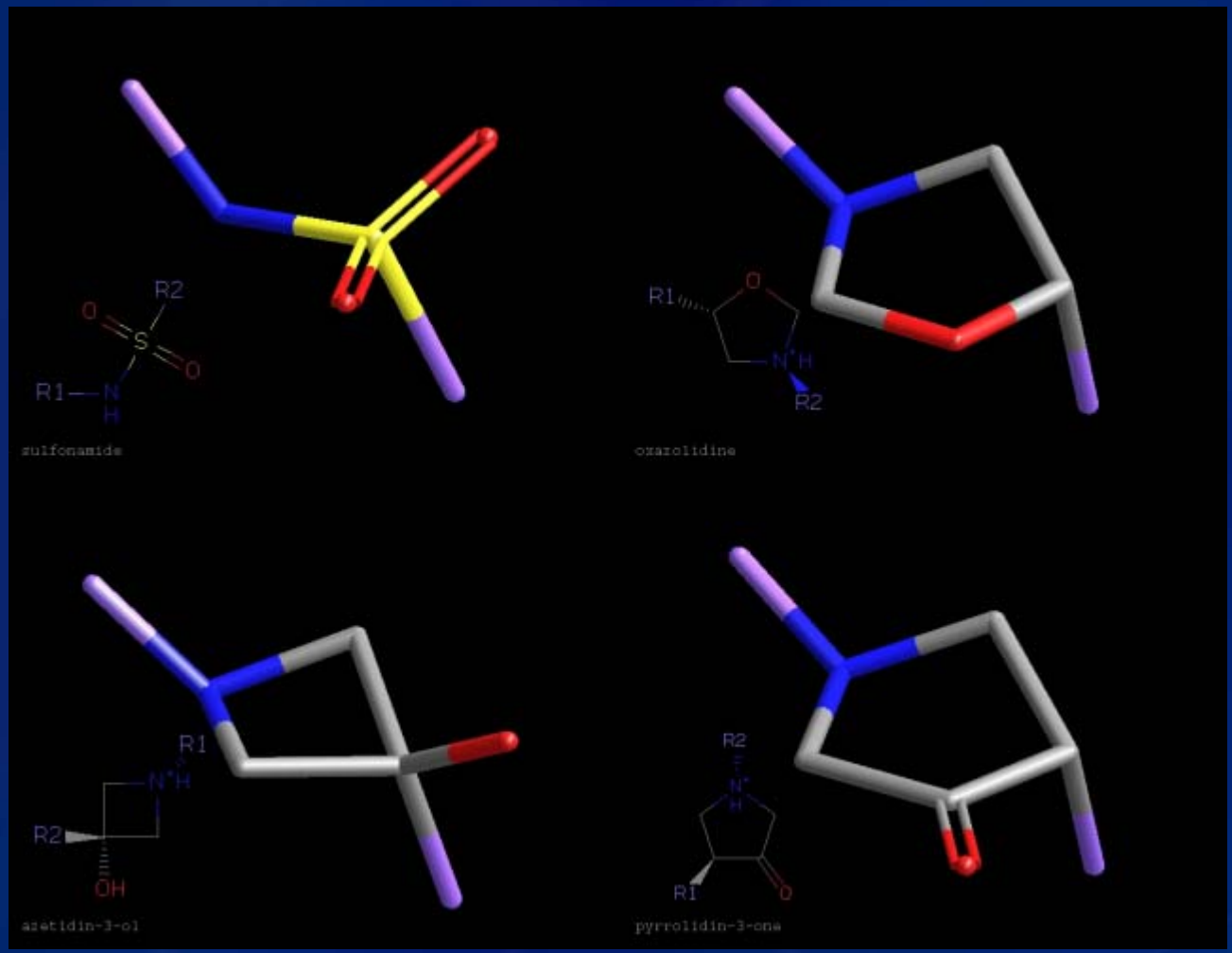
2dsim 0.55

Electrostatic Modification

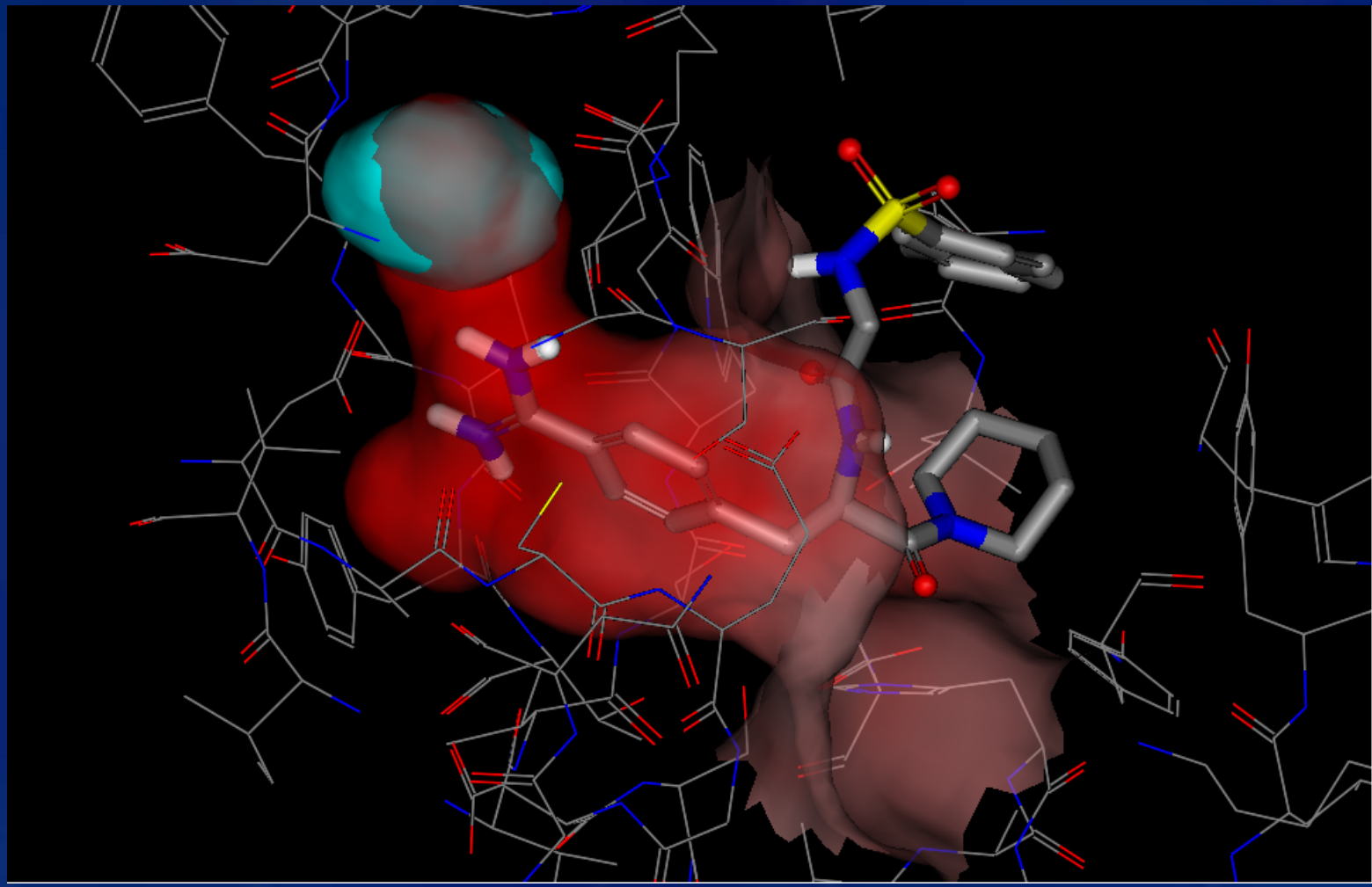


Electrostatic Modification

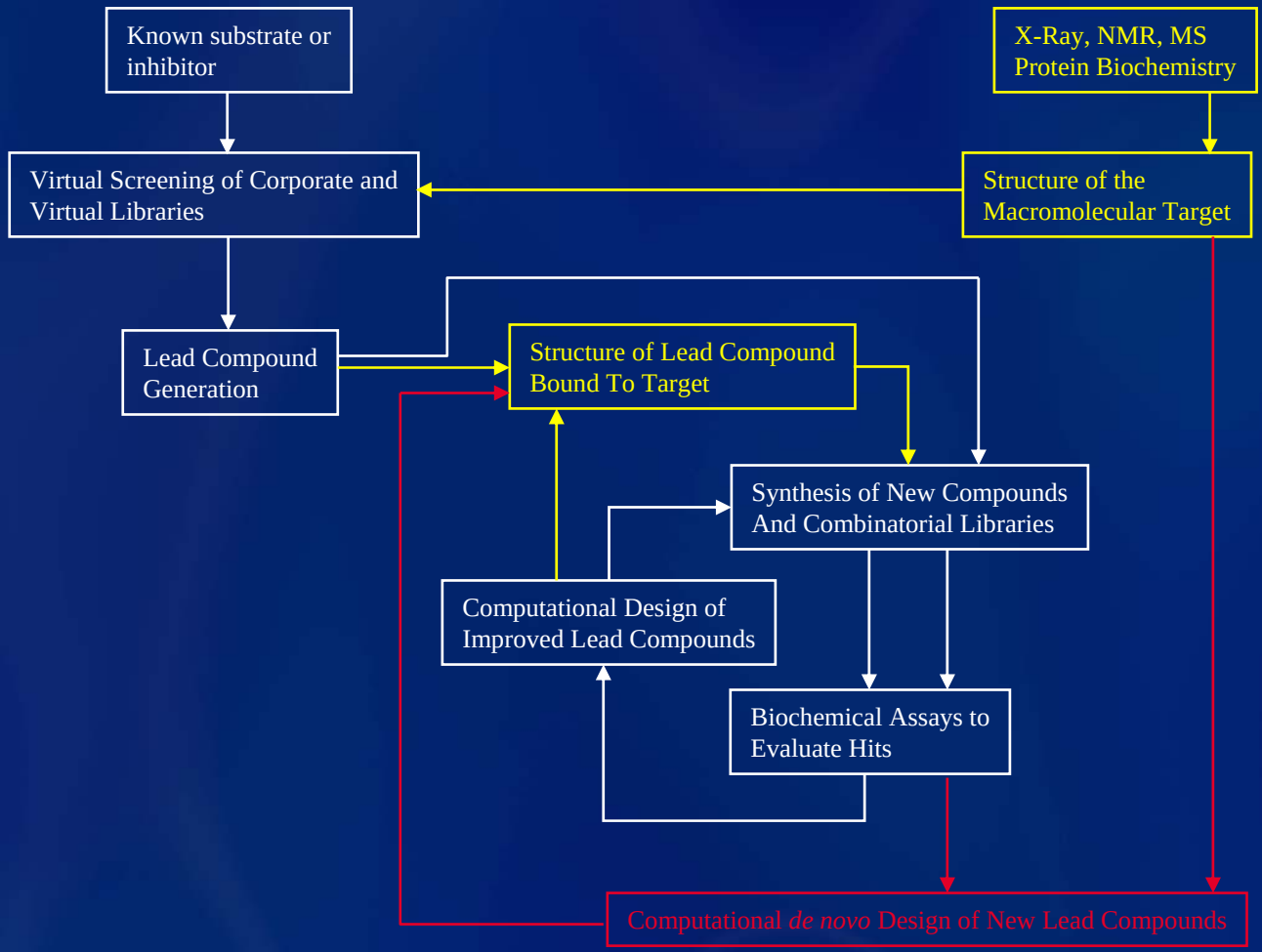
Scaffold Substitution



Void Volumes

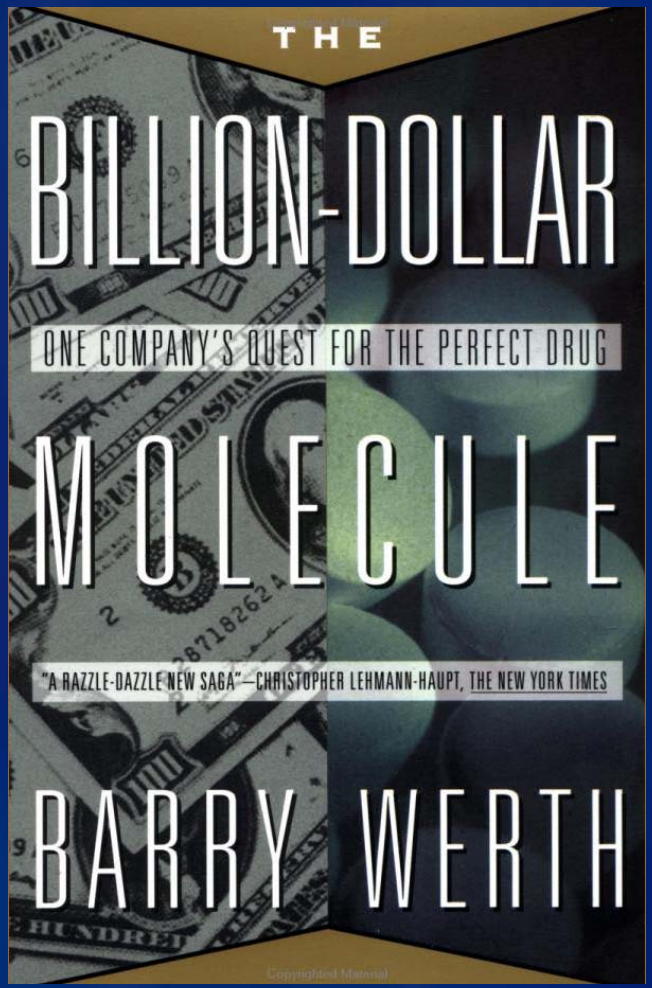


Drug Discovery



Adapted from Figure 22.1: Bourne, Wessig. Structural Bioinformatics.

Fun Reading



The Billion-Dollar Molecule
by Barry Werth

Story of the creation of Vertex Pharmaceuticals and their efforts to develop a blockbuster drug using rational drug design methods

*Image obtained from Amazon.com

Present your class project!

- OpenEye's 7th Annual CUP Meeting
  - March 6th to 8th, Santa Fe, NM
- Poster submissions still welcome (contrary to what it says on website)
- http://www.eyesopen.com