

BIMM 143

Structural Bioinformatics II

Lecture 12

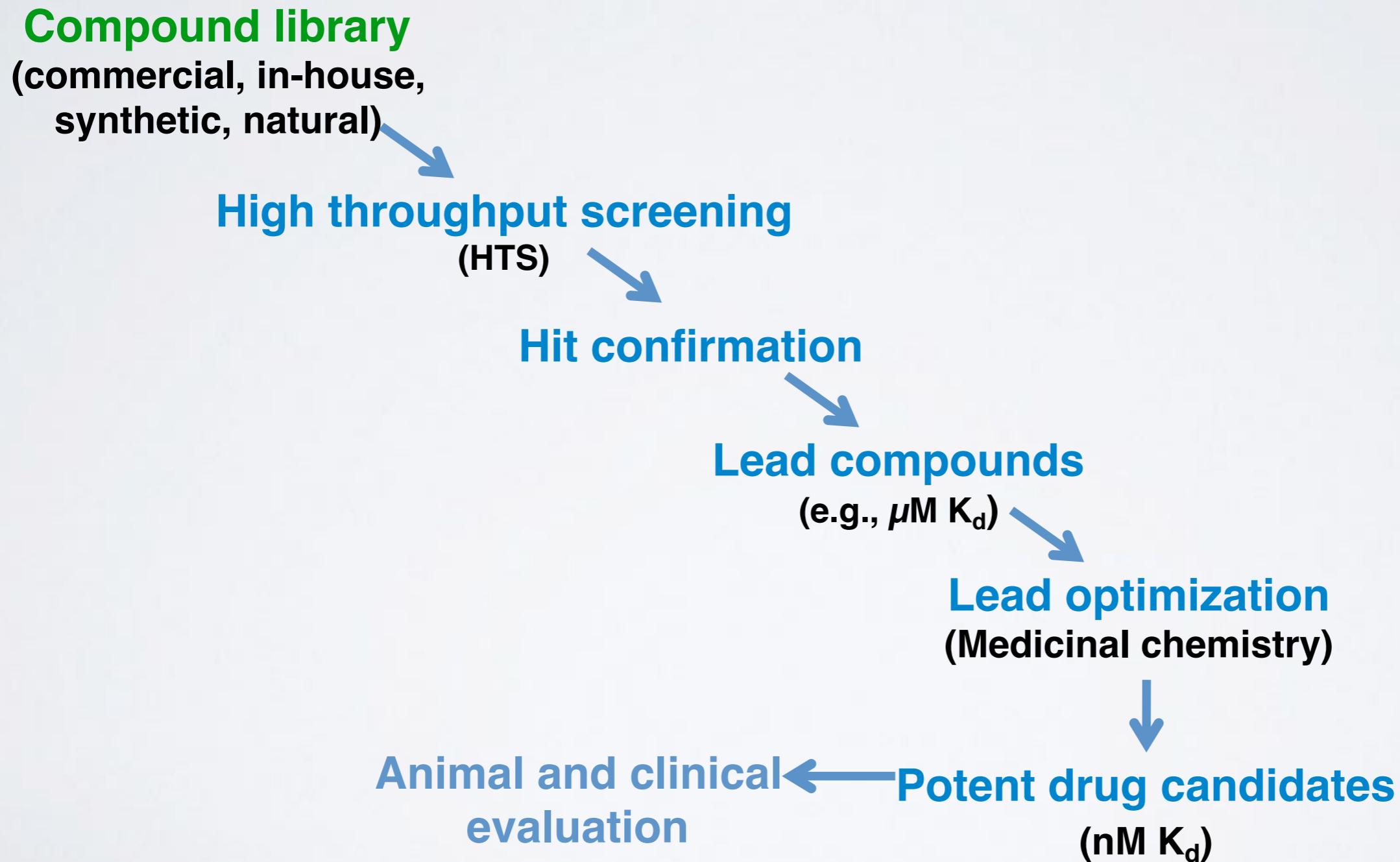
Barry Grant
UC San Diego

<http://thegrantlab.org/bimm143>

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**
 - drug discovery & Predicting functional dynamics

THE TRADITIONAL EMPIRICAL PATH TO DRUG DISCOVERY



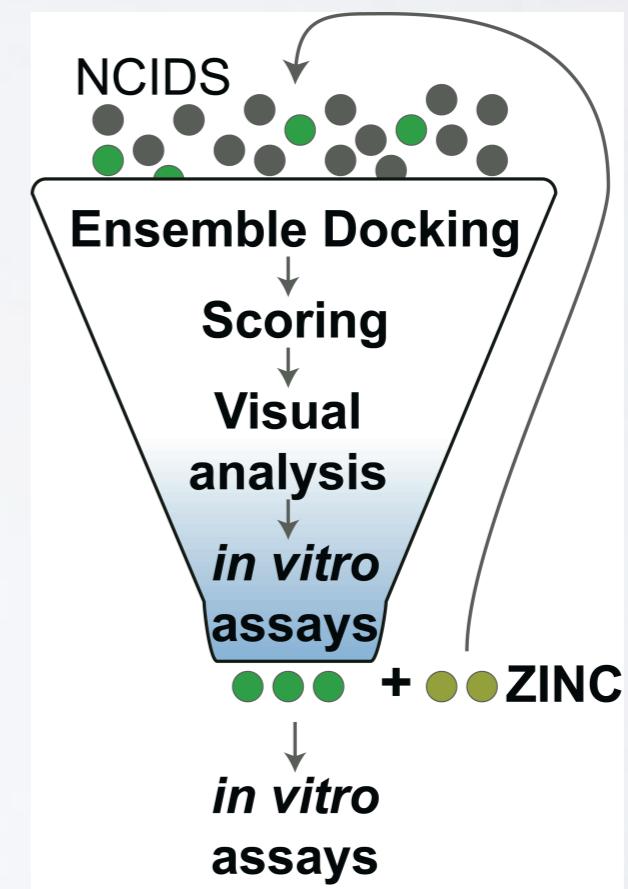
COMPUTER-AIDED LIGAND DESIGN

Aims to reduce number of compounds synthesized and assayed

Lower costs

Reduce chemical waste

Facilitate faster progress



Two main approaches:

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

Two main approaches:

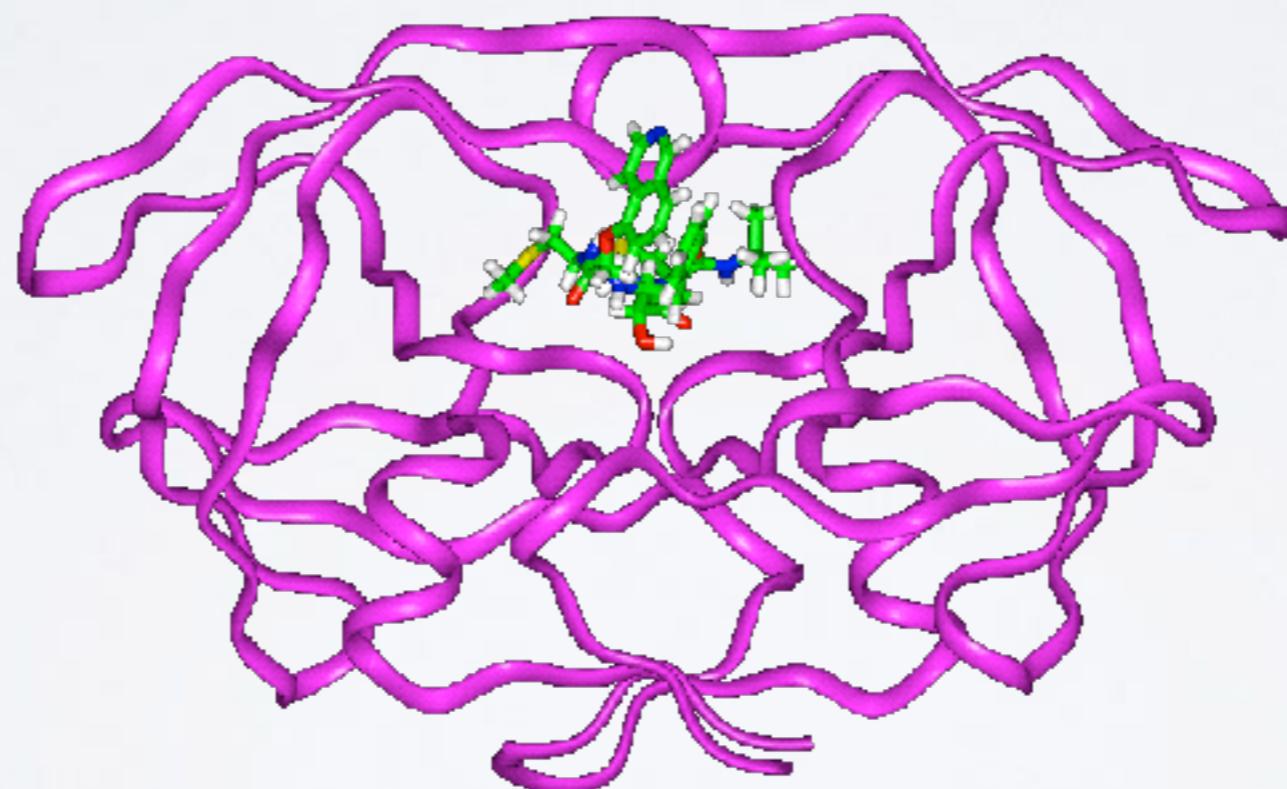
(1). Receptor/Target-Based

(2). Ligand/Drug-Based

SCENARIO I:

RECEPTOR-BASED DRUG DISCOVERY

Structure of Targeted Protein Known: **Structure-Based Drug Discovery**



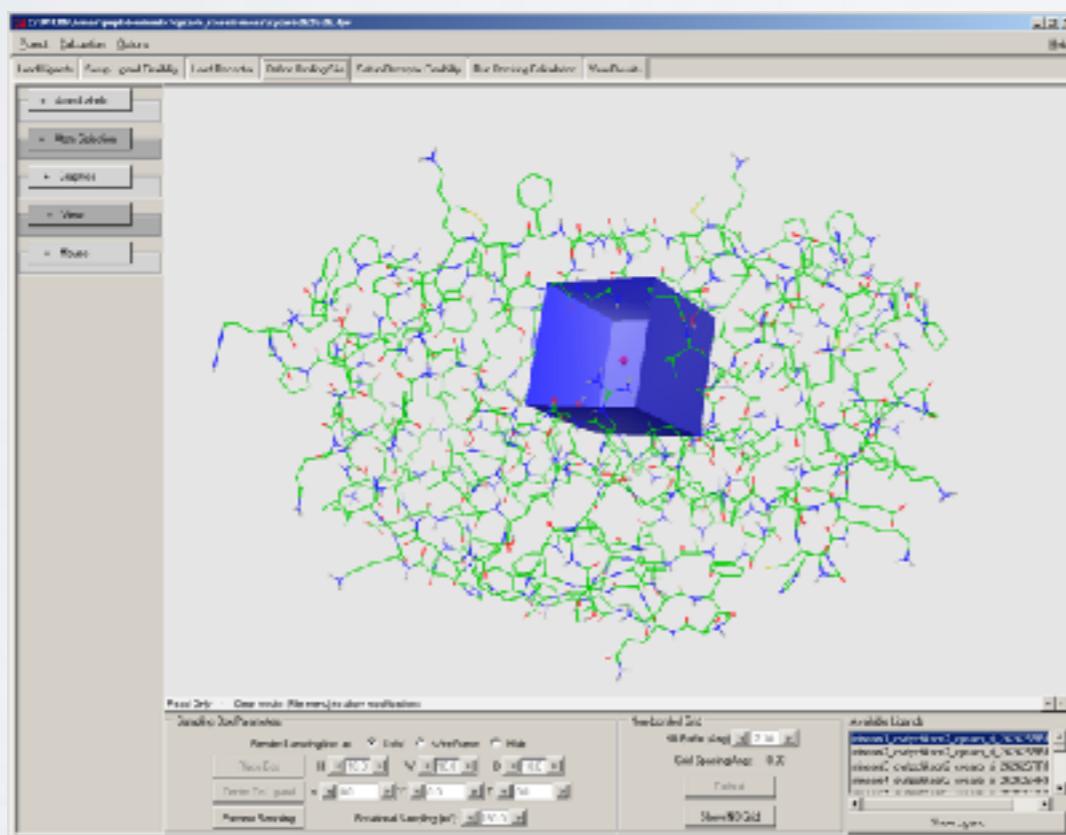
HIV Protease/KNI-272 complex

PROTEIN-LIGAND DOCKING

Structure-Based Ligand Design

Docking software

Search for structure of lowest energy



Potential function

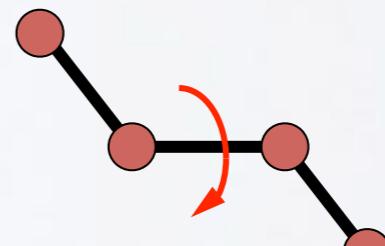
Energy as function of structure



VDW

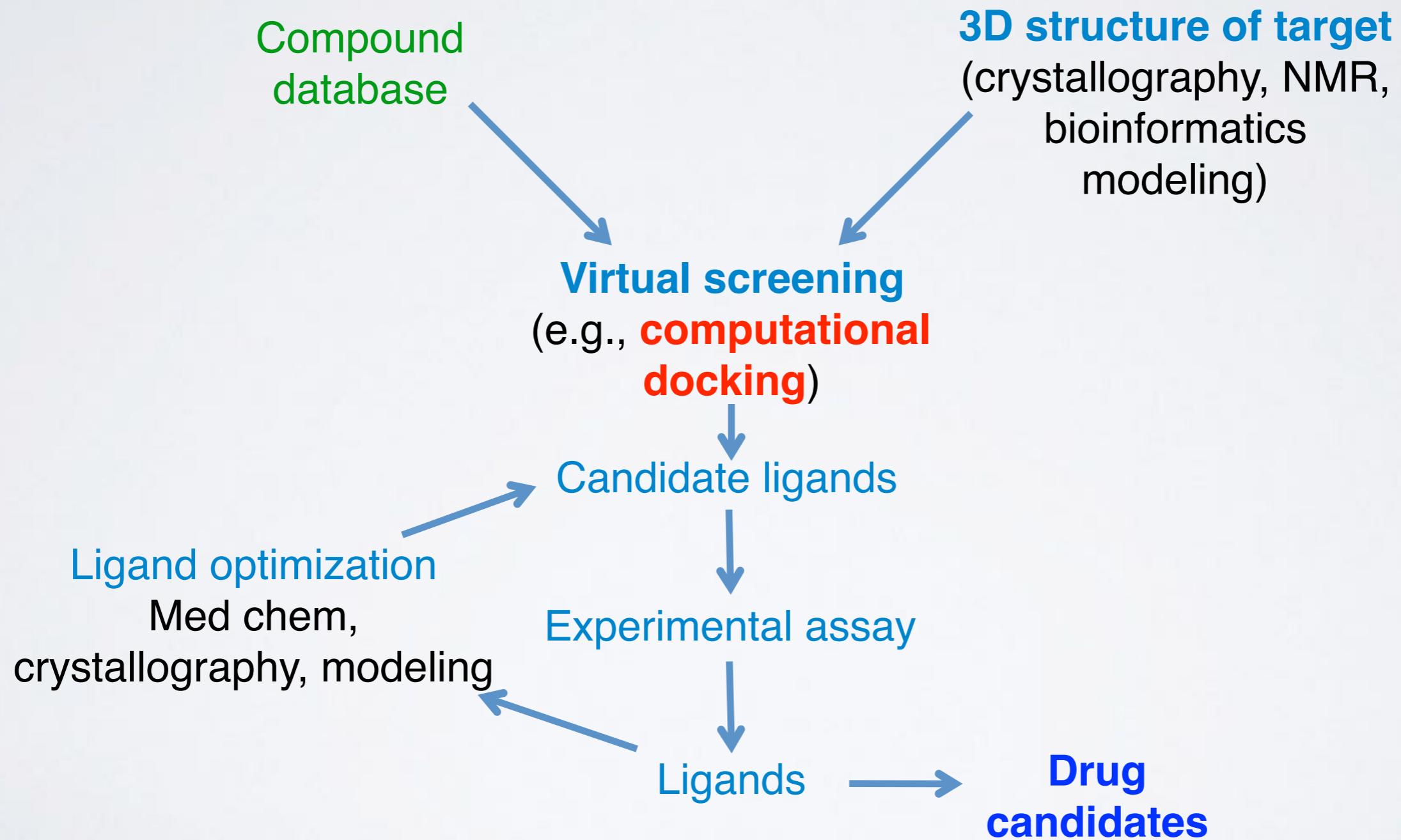


Screened Coulombic



Dihedral

STRUCTURE-BASED VIRTUAL SCREENING



COMPOUND LIBRARIES



The screenshot shows the Mybridge H Under™ website. At the top, there's a banner with a landscape image of a bridge over water. Below the banner, the header includes "HOME", "BIOLOGICAL SCREENING", "BIOACTIVE LIBRARIES", "CONTACT", "ABOUT US", and "JOBS". A sidebar on the left has sections for "Search our sites", "Mybridge", "Libraries", "Databases", "Submit your screening project to us", "Our Services", and "Contact Us". The main content area features a heading "Mybridge H Under™" and a sub-section "The pre-selected diverse screening library includes identifying potential drug leads easy, universal, and cost effective." It also contains a "Sample Quality Data from your screens" section and a "Reduced time to synthesis every hit" section. At the bottom, there's a "Ready to Screen" section with a small image of a plate and some text.



The screenshot shows the NIH Molecular Libraries Small Molecule Repository website. The header reads "NIH MOLECULAR LIBRARIES SMALL MOLECULE REPOSITORY" and "A Roadmap Initiative". The main menu includes "HOME", "MSMR Project", "MSMR", "MSMR Control", "Sample Screening", "Sample Arrays", "Information", "MLPNC Controls", "MLPNC Details", and "Robust Compounds". A "Registered Users Logon" link is also present. The central content area has a "Welcome" section with text about the repository's mission and a "Behind the Scenes" section with a photograph of a scientist in a lab. The footer includes "NIH Molecular Libraries Small Molecule Repository", "BioFocus, a Galapagos company", and "NIH Roadmap Project supporting 'From Pathways to Discovery' in the 21st Century".



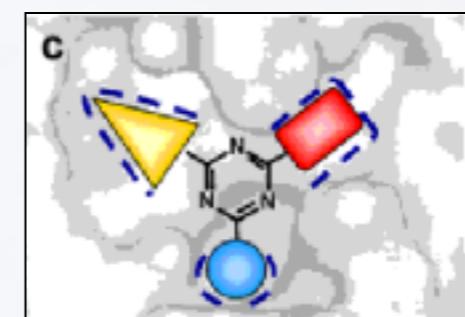
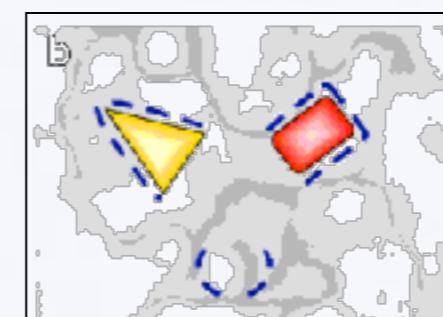
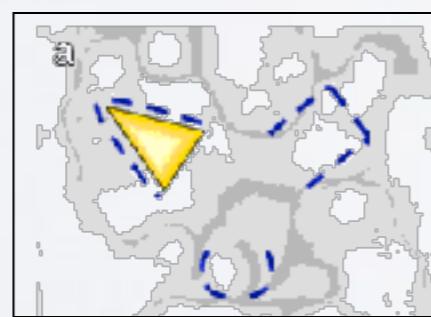
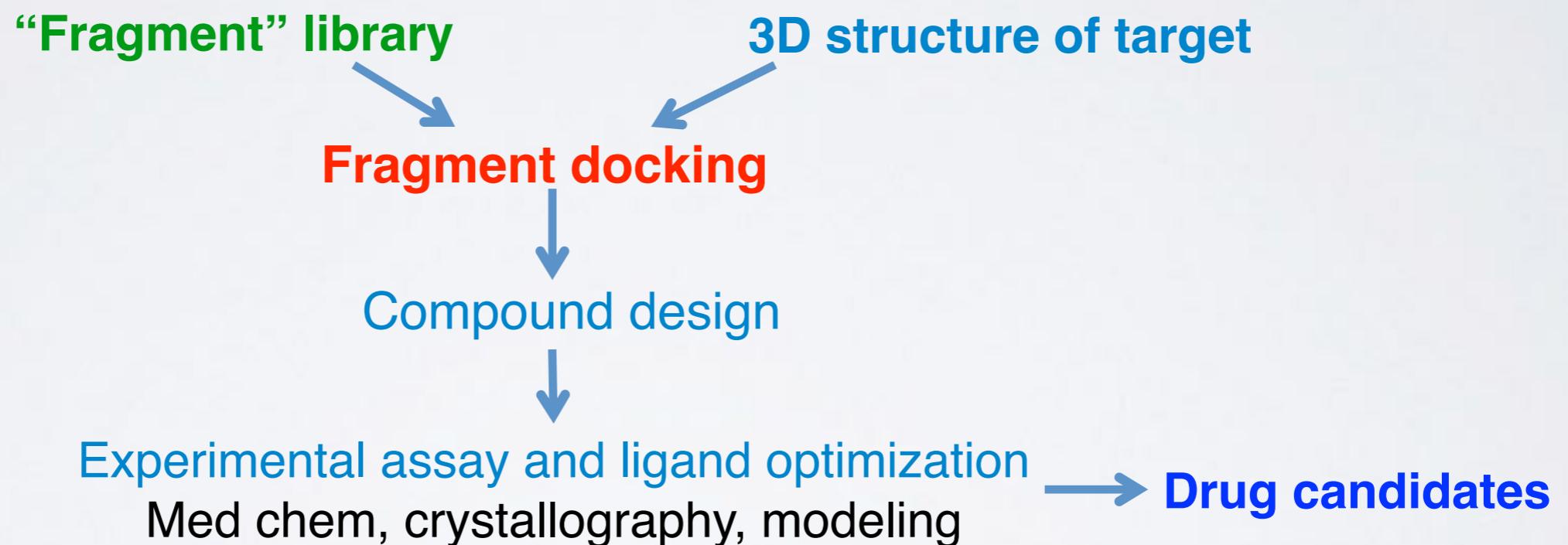
The screenshot shows the Pittsburgh Molecular Libraries Screening Center (PMLSC) website. The header includes "University of Pittsburgh" and "Pittsburgh Molecular Libraries Screening Center". The main menu lists "HOME", "HISTORY", "PERSONNEL", "SCREENING TECHNOLOGY", "COMPOUND LIBRARIES", "INSTRUMENTATION", "HTS GUIDELINES", "ADVISORY PANEL ASSAY PROTOCOLS", "PUBLISHED REPORTS", "LIBRARY", "DATA ANALYSIS/STATISTICS", "EDUCATIONAL ACTIVITIES", "PUBLICATIONS", "LINKS", "CONTACTS", and "Keyword Search". The central content area features a large image with the text "BIG DISCOVERIES FROM SMALL MOLECULES" and a "Welcome" section. The footer includes "NIH Roadmap Project", "NIH", "NIH Clinical Collection", "NIH Molecular Libraries Screening Center", "University of Pittsburgh", "Carnegie Mellon University", "Last Update 3/1/2007", and "Copyright © 2007, Office of the Senior Vice President for Health Sciences, University of Pittsburgh. All rights reserved".

Commercial
(in-house pharma)

Government (NIH)

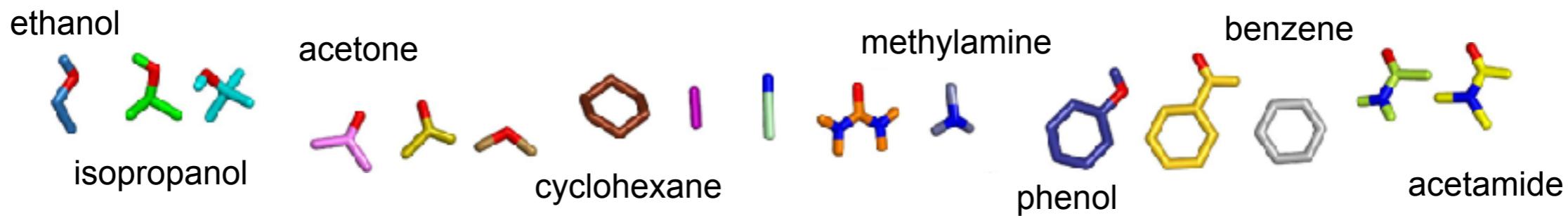
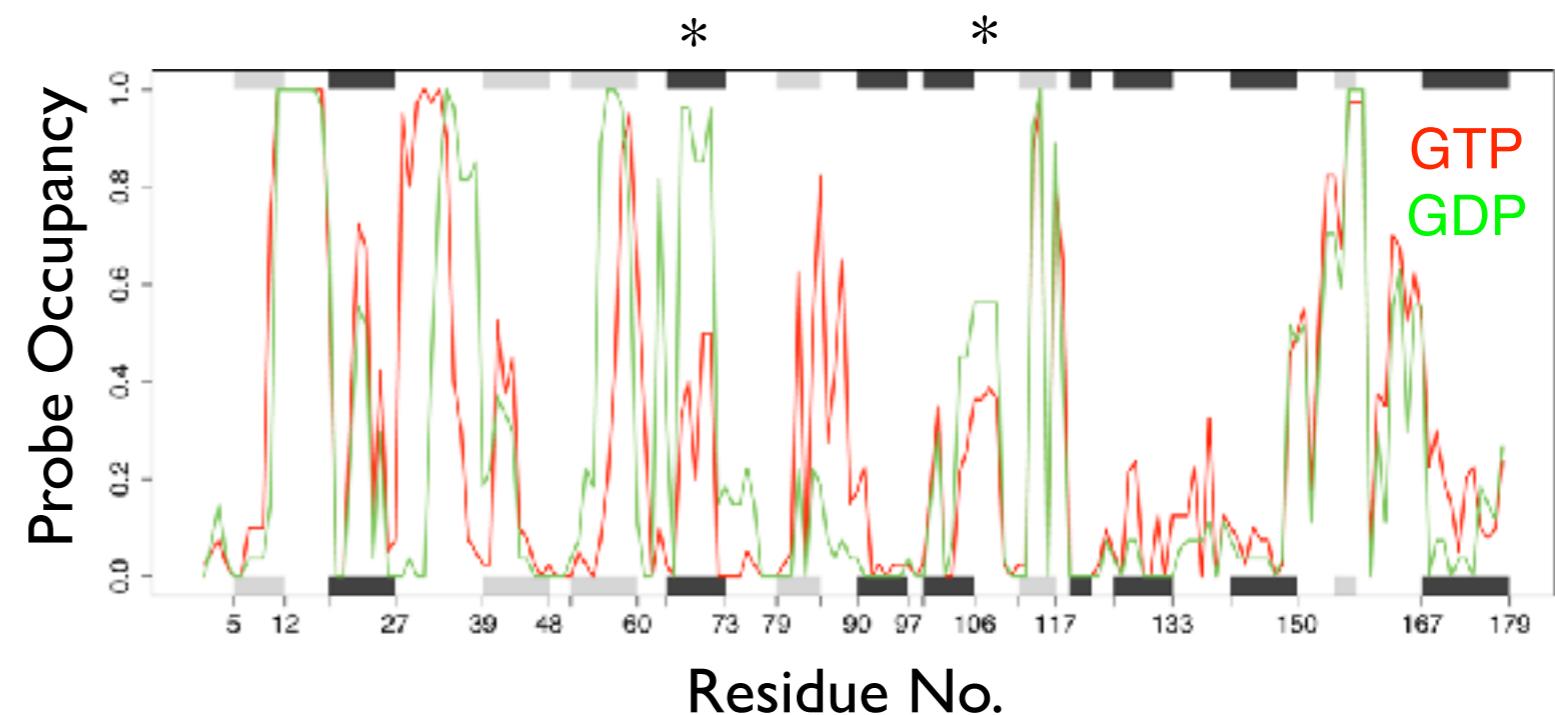
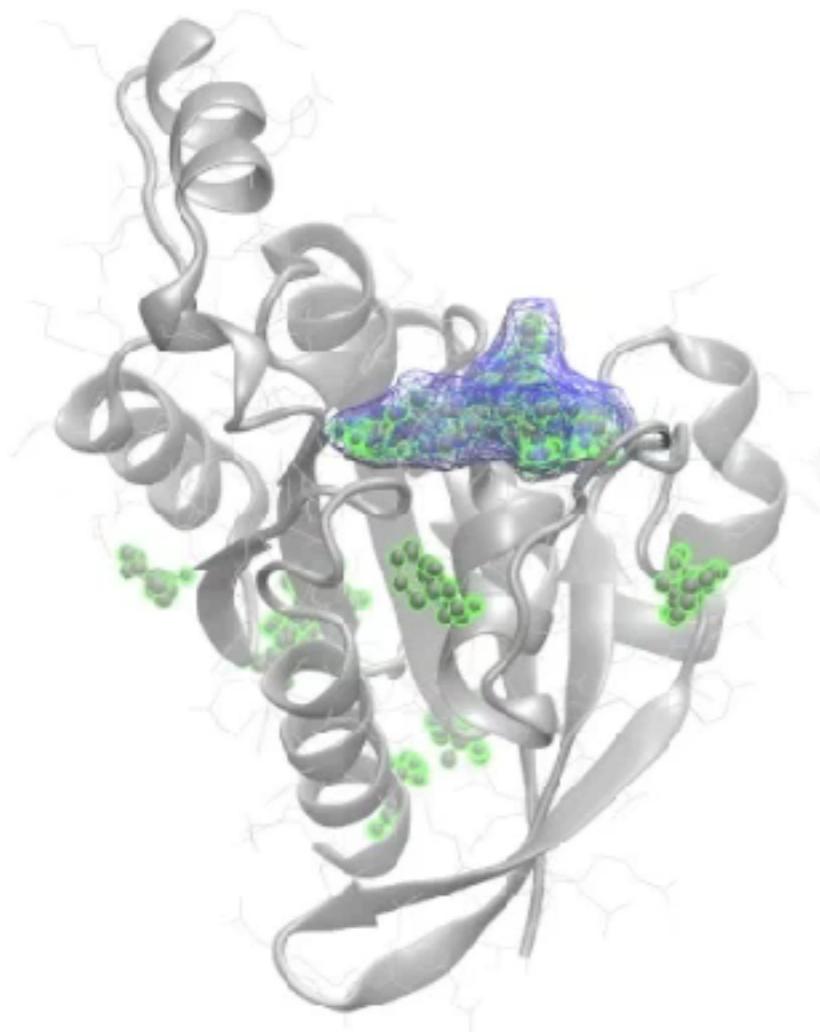
Academia

FRAGMENTAL STRUCTURE-BASED SCREENING



Multiple non active-site pockets identified

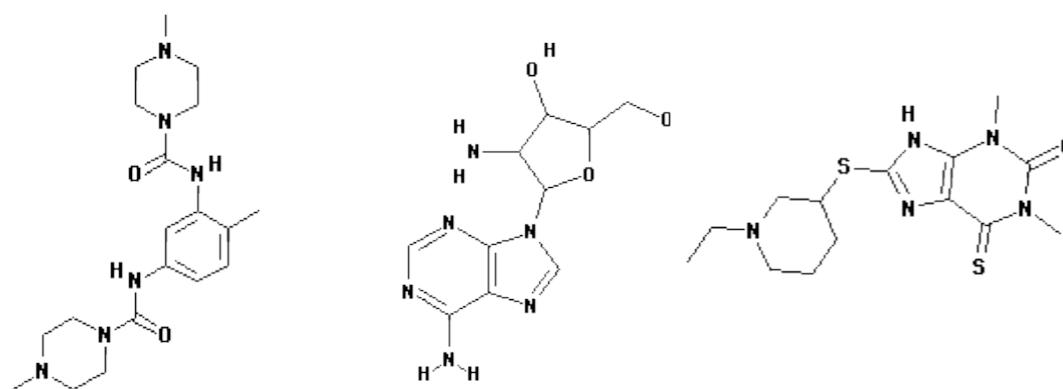
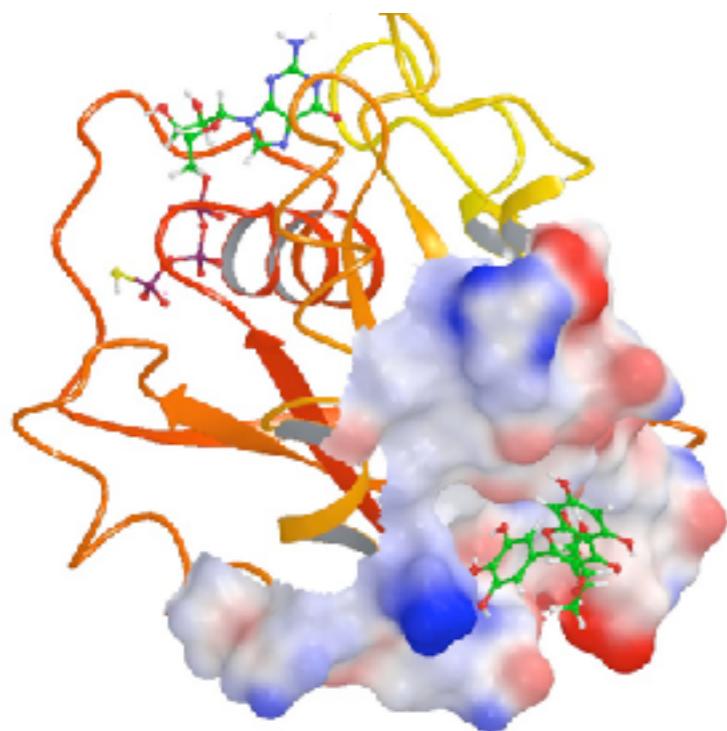
Small organic probe fragment affinities map multiple potential binding sites across the structural ensemble.



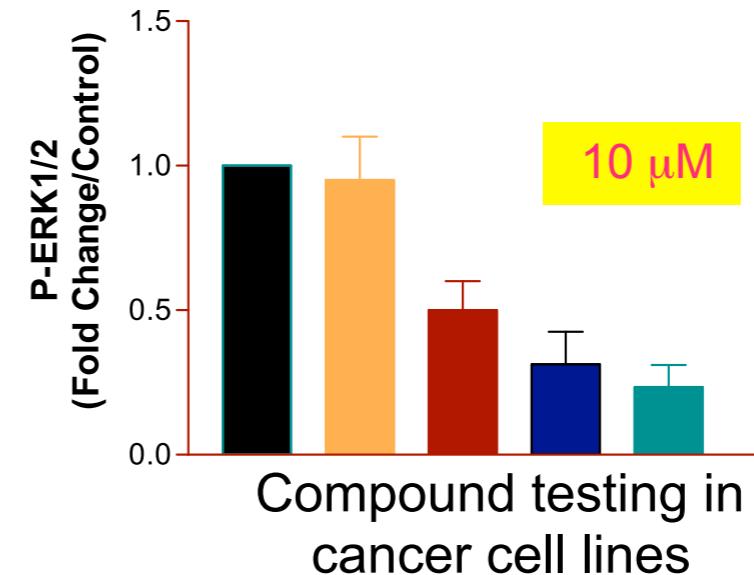
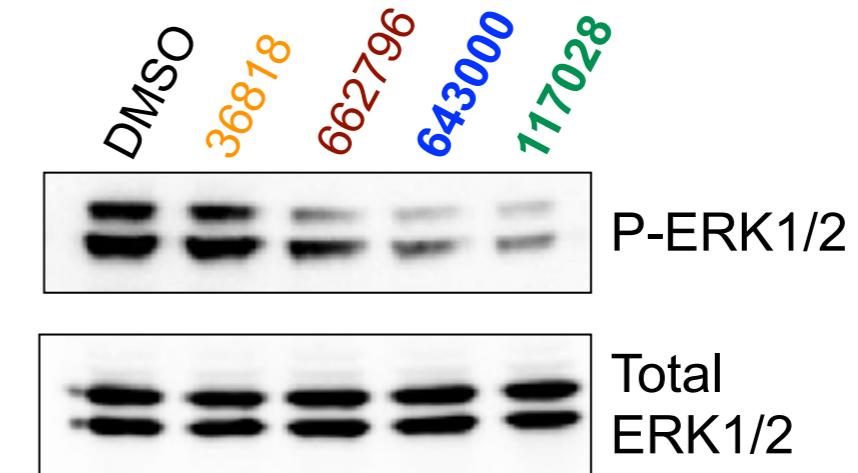
Ensemble docking & candidate inhibitor testing

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

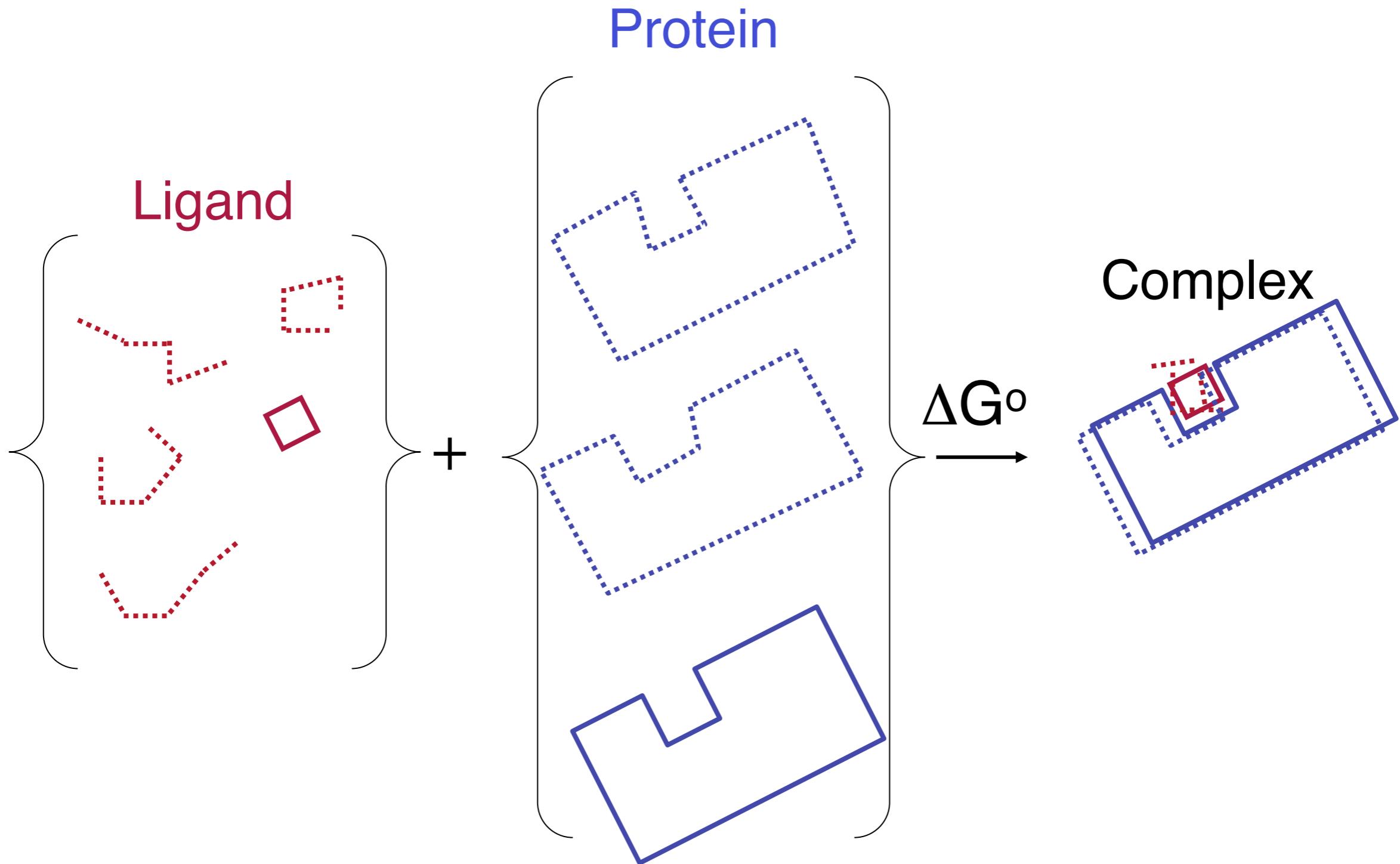
Ensemble computational docking



Compound effect on U251 cell line



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!

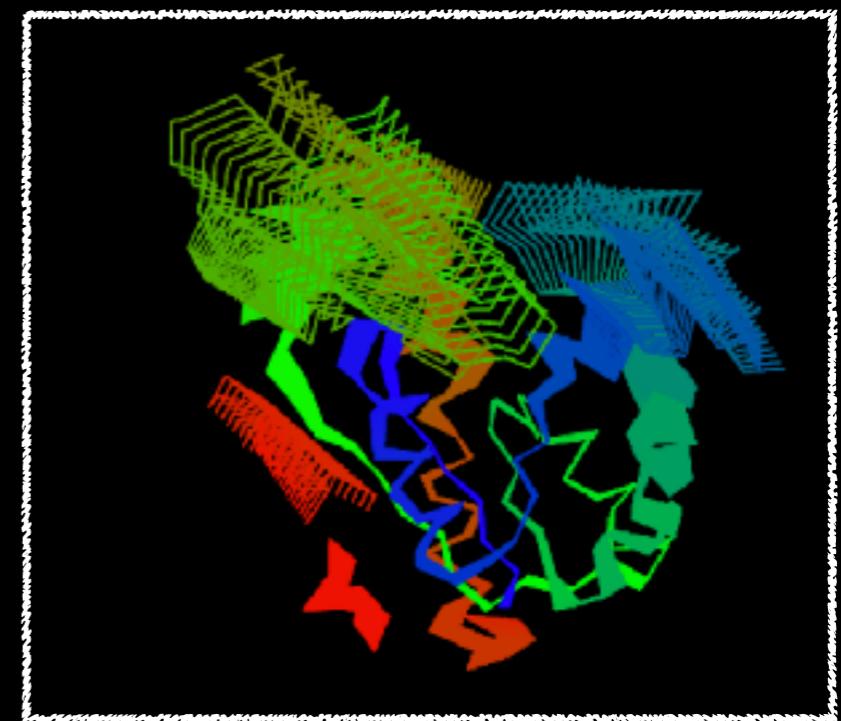
https://bioboot.github.io/bimm143_S18/lectures/#12

You can use the classroom computers or your own laptops. If you are using your laptops then you will need to install **VMD** and **MGLTools**

Bio3D view()

- If you want the 3D viewer in your R markdown you can install the development version of Bio3D
- For **MAC**:

```
> download.file("https://tinyurl.com/bio3d-mac", "bio3d.tar.gz")
> install.packages("bio3d.tar.gz", repos = NULL)
```

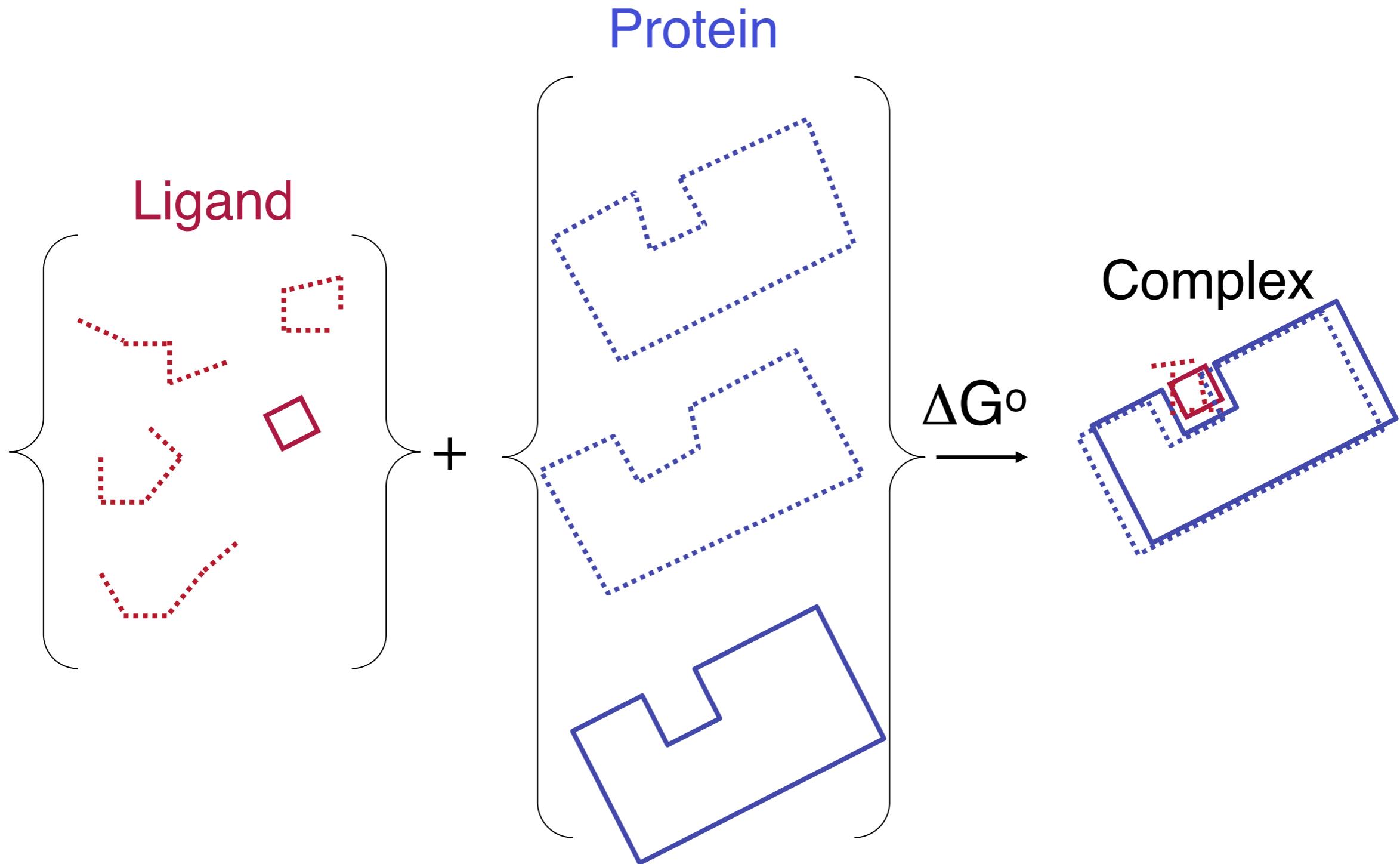


- For **Windows**:

```
> install.packages("https://bioboot.github.io/bggn213_S18/class-
material/bio3d_2.3-4.9000.zip", repos = NULL)
```

[See: Appendix I in Lab Sheet]

Proteins and Ligand are Flexible



HTTP://129.177.232.111:3848/PCA-APP/

HTTP://BIO3D.UCSD.EDU/PCA-APP/

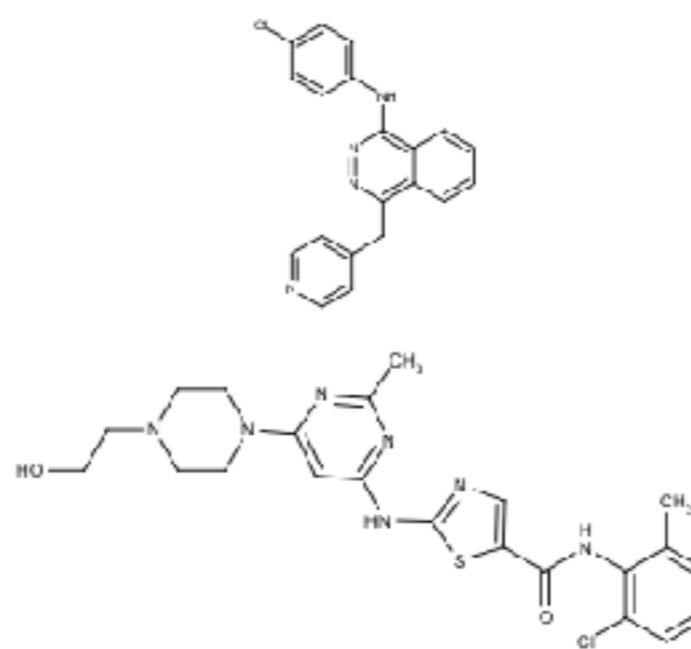
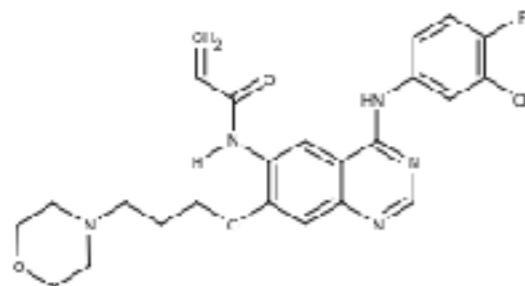
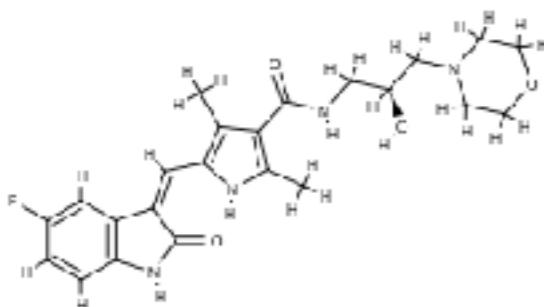
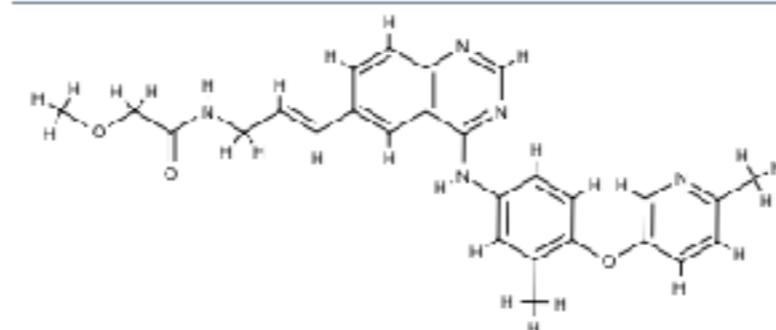
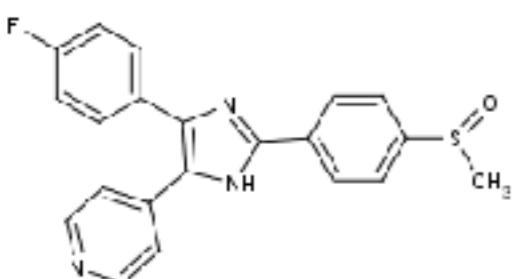
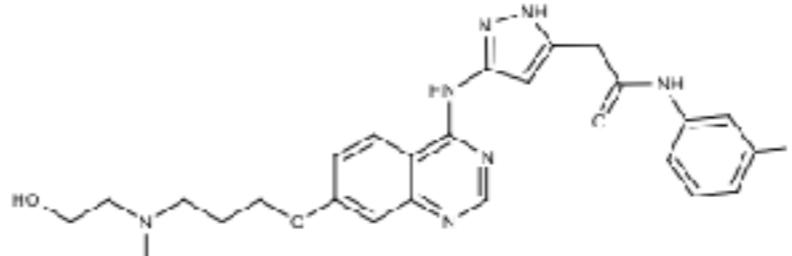
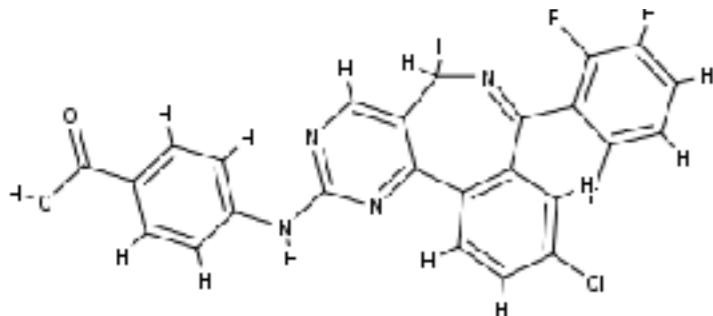
Two main approaches:

- (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

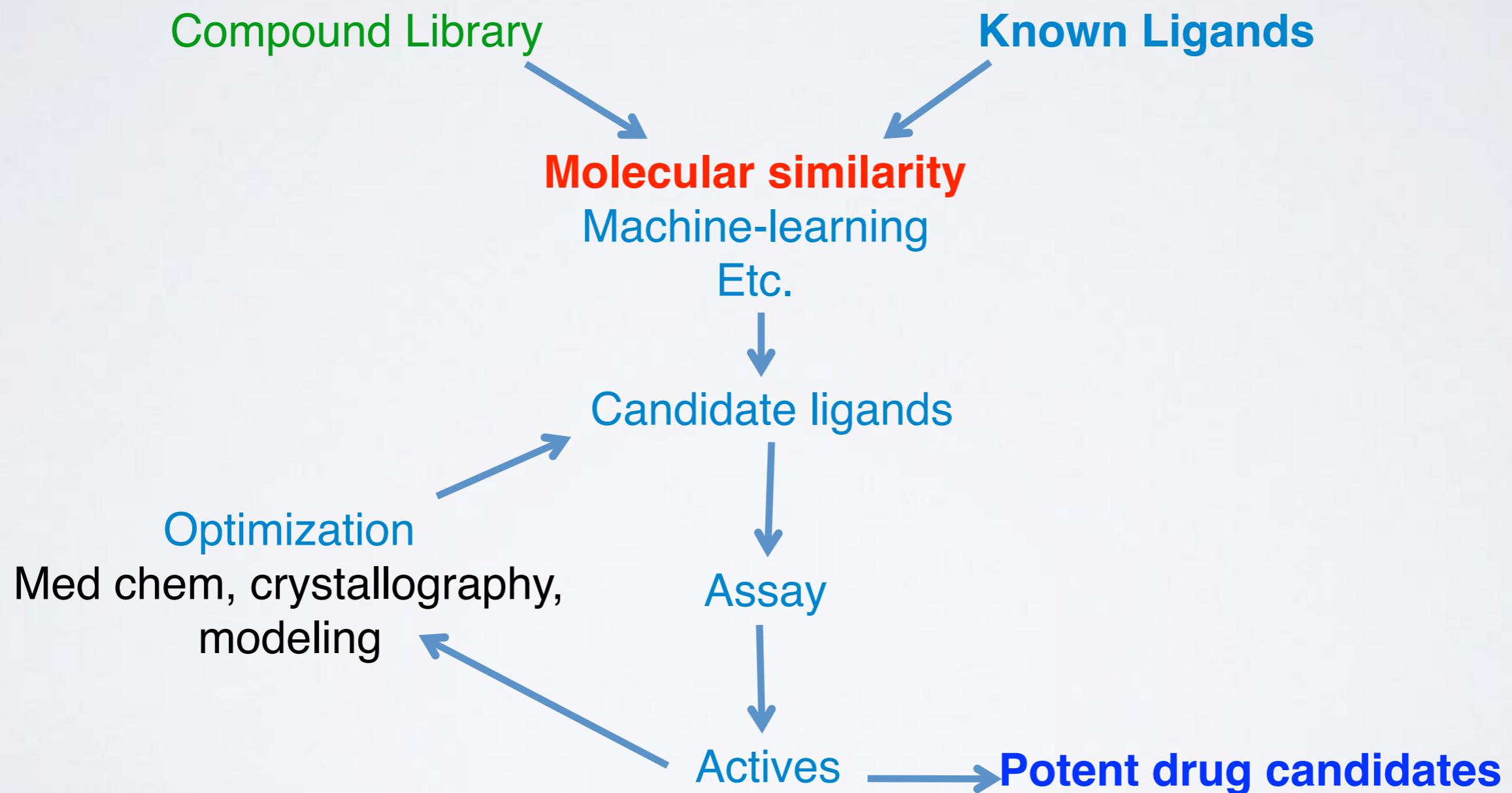
Why Look for Another Ligand if You Already Have Some?

Experimental screening generated some ligands, but they don't bind tightly enough

A company wants to work around another company's chemical patents

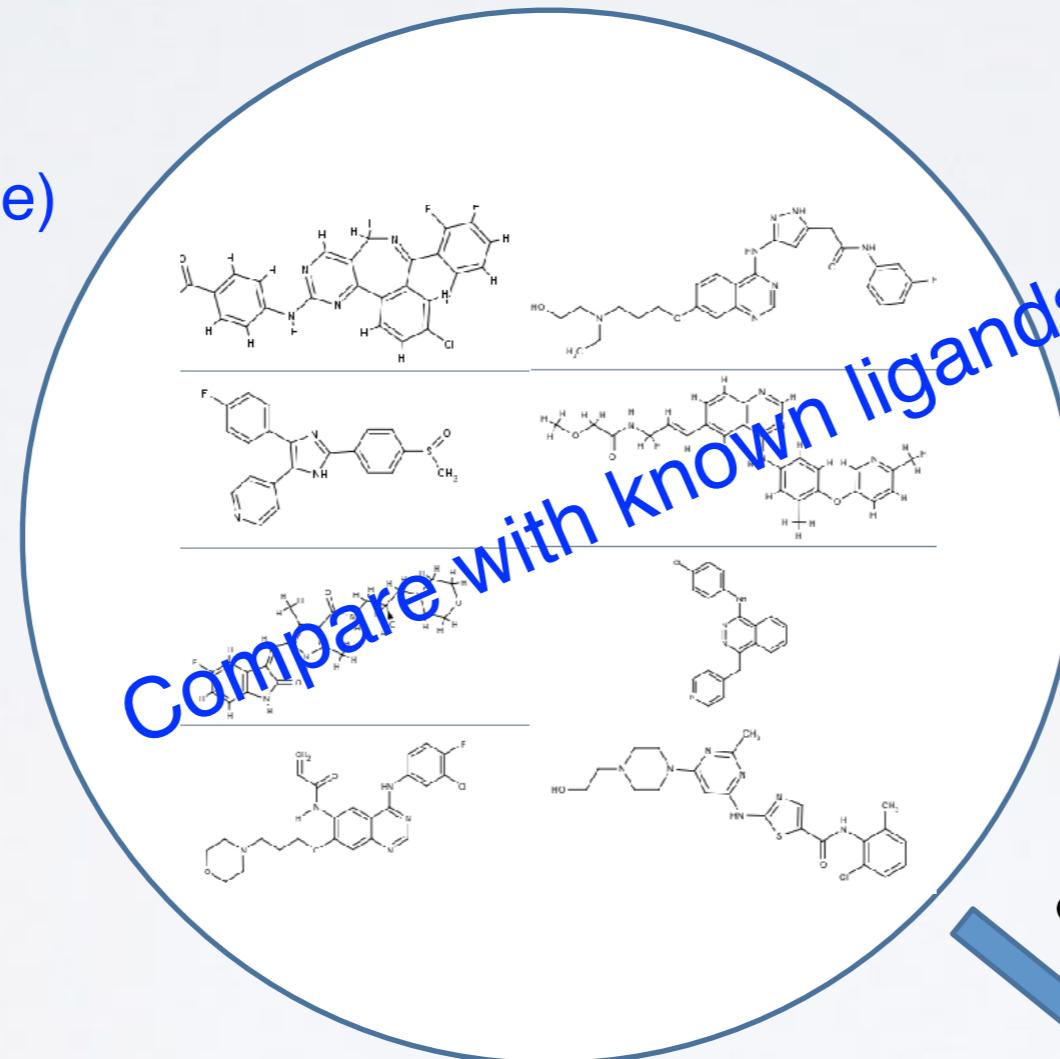
An high-affinity ligand is toxic, is not well-absorbed, difficult to synthesize etc.

LIGAND-BASED VIRTUAL SCREENING



CHEMICAL SIMILARITY LIGAND-BASED DRUG-DISCOVERY

Compounds
(available/synthesizable)



Different

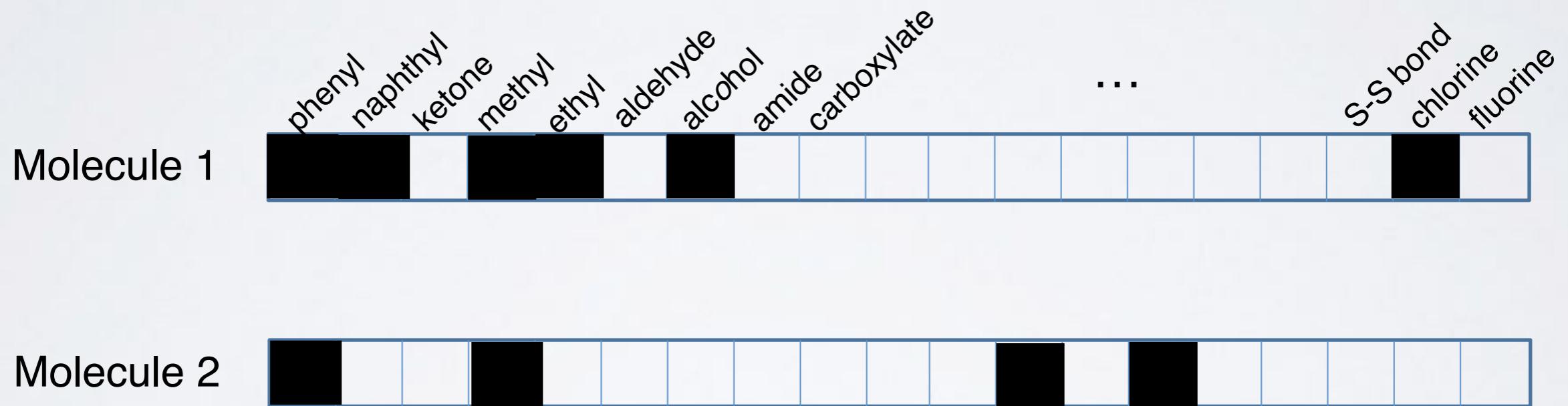
Don't bother

Similar

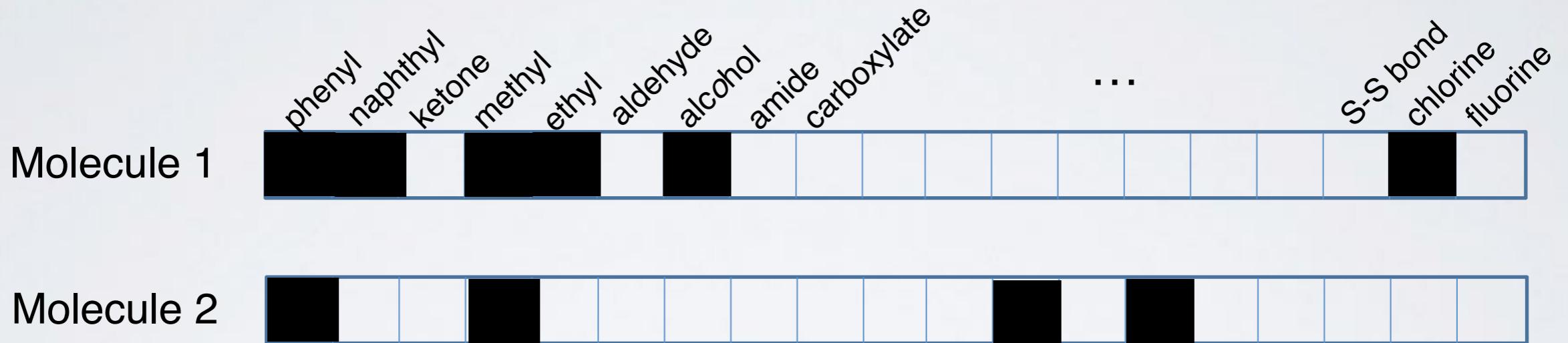
Test experimentally

CHEMICAL FINGERPRINTS

BINARY STRUCTURE KEYS



CHEMICAL SIMILARITY FROM FINGERPRINTS



Tanimoto Similarity
(or Jaccard Index), T

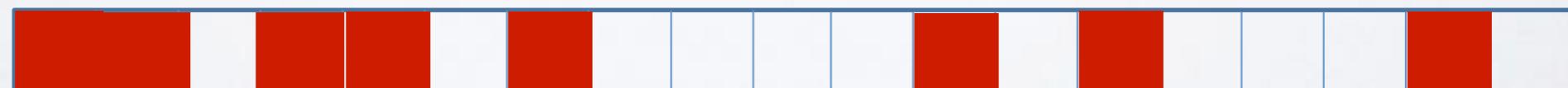
$$T \equiv \frac{N_I}{N_U} = 0.25$$

Intersection



$N_I=2$

Union

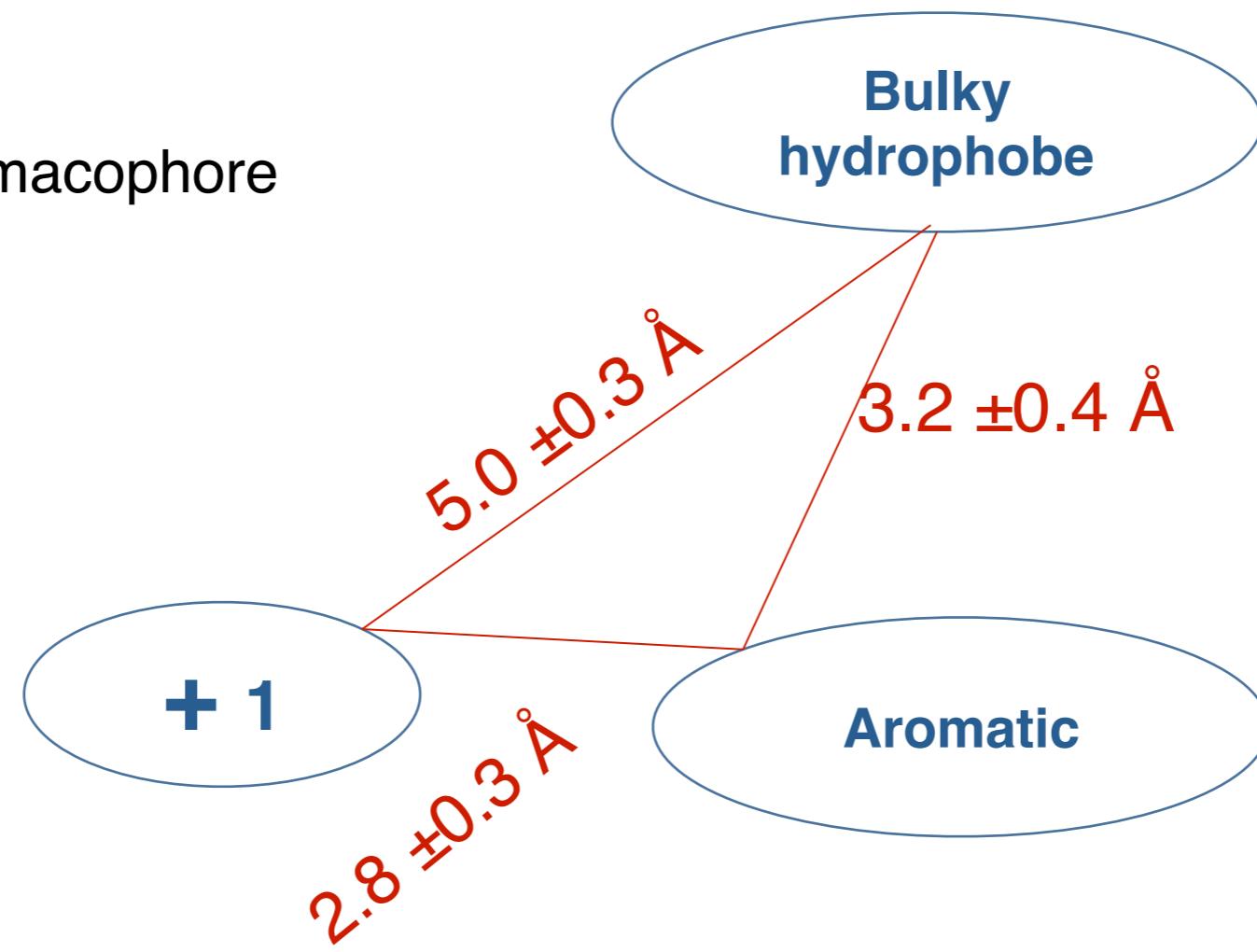


$N_U=8$

Pharmacophore Models

Φάρμακο (drug) + Φορά (carry)

A 3-point pharmacophore



Molecular Descriptors

More abstract than chemical fingerprints

Physical descriptors

molecular weight

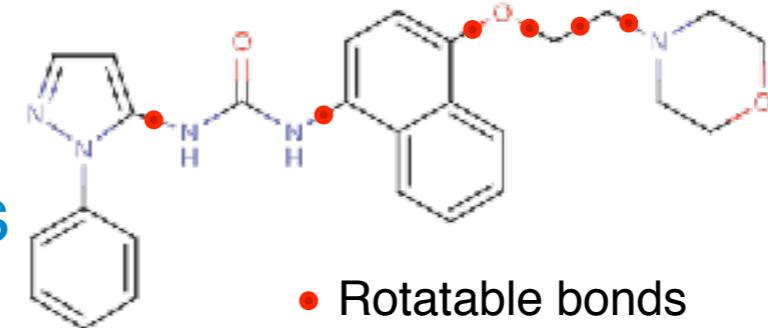
charge

dipole moment

number of H-bond donors/acceptors

number of rotatable bonds

hydrophobicity ($\log P$ and $c\log P$)



• Rotatable bonds

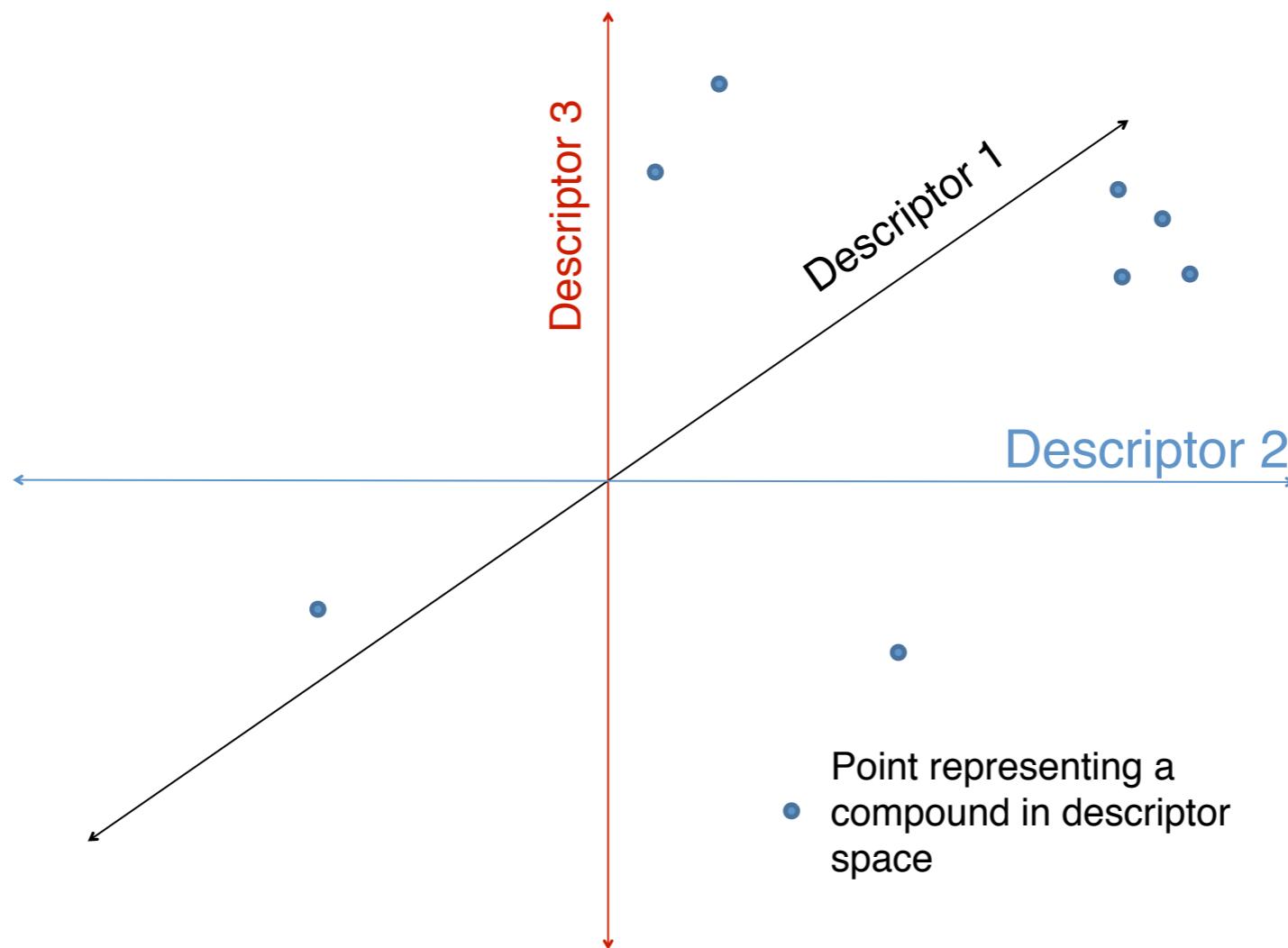
Topological
branching index
measures of linearity vs interconnectedness

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, PCA, 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

ChEMBL wellcome trust

EBI > Databases > Small Molecules > ChEMBL Database > Home

Search ChEMBL... Compounds Targets Assays Documents Activity Source Filter

Ligand Search Target Search Browse Targets **Browse Drugs** Browse Drug Targets Drug Approvals About

Downloads... 10 records per page Search: Show / hide columns

Parent Molecule	Synonyms	Phase	Research Codes	Applicants	USAN Stem	USAN Year	First Approval	ATC Code	Icon
	Elosulfase Alfa (INN, USAN)	4		Biomarin Pharmaceutical Inc.	-ase	2012	2014		
CHEMBL2108676									
	Tasimelteon (FDA, INN, USAN)	4	BMS-214778 VEC-162	Vanda Pharmaceuticals Inc	-melteon	2007	2014		
CHEMBL2103822									
	Apremilast (FDA, INN, USAN)	4	CC-10004	Celgene Corp	-ast	2005	2014	L04AA32	
CHEMBL514800									
	Flortetaben F-18 (FDA) Flortetaben F18 (USAN)	4	BAY-949172 UNII-TLA7312TOI	Piramal Imaging Sa		2013	2014		
CHEMBL1908906									
	Droxidopa (FDA, INN, USAN)	4	DOPS L-DOPS	Chelsea Therapeutics Inc	-dopa	2008	2014		
CHEMBL2108676									

Drug properties

	Drug Type	Rule of Five	First in Class	Chirality	Prodrug	Oral	Parenteral	Topical	Black-Box Warning	Availability Type	
synthetic small molecule		5									prescription only
natural product-derived										over-the-counter	
inorganic					chirally pure					discontinued	
polymer		Ingredient-related (USANs, candidates and approved drugs)				Product-related (approved drugs only)					
monoclonal antibody											
enzyme											
peptide/protein											
oligonucleotide											
oligosaccharide											

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?

Druggability prediction

https://www.ebi.ac.uk/chembl/druggability/domain/32655

**View cavities
(and ligands)
on structure**

↓

Domain Details:

PDB	1eeo	SCOP
Gene	P18031	View Protein Summary
Description	Tyrosine-protein phosphatase non-receptor type 1	
Fold	Phosphotyrosine protein phosphatases II	
Superfamily	Phosphotyrosine protein phosphatases II	
Family	Higher-molecular-weight phosphotyrosine protein phosphatases	
Other PDB(s)	1eeo:A - px32655 (Tractable: 1, Druggable: 0, Ensemble: -0.96)	

Average Druggability Scores:

Tractable	Druggable	Ensemble
0.97	0.02	-0.93

Tractable/Druggable ranges from low:0 to high:1. Ensemble ranges from low:-1 to high:+1.

Site Druggability Details:

Reset:	Site 1	Site 2	Site 3	Site 4
Druggable	0.00	0.00	0.00	0.00
Confidence	0.73	0.96	0.96	0.96
Tractable	1.00	0.00	0.00	0.00
Confidence	0.92	0.86	0.83	0.86
Ensemble	-0.96	-0.99	-0.98	-0.99
Volume [Å ³]	1535.2	1318.36	1446.61	1454.2
Buried Surface [%]	71.3	65.25	72.27	64.08
Show Site	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Show Residues	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>

Ligand PTR (CHEMBL286939):

Green :Druggable, Yellow :Tractable, Pink :Undruggable

Use | Privacy | Cookies | EBI Funding | Contact EBI | © European Bioinformatics Institute 2012. EBI is an Outstation of the European Molecular Biology Laboratory.

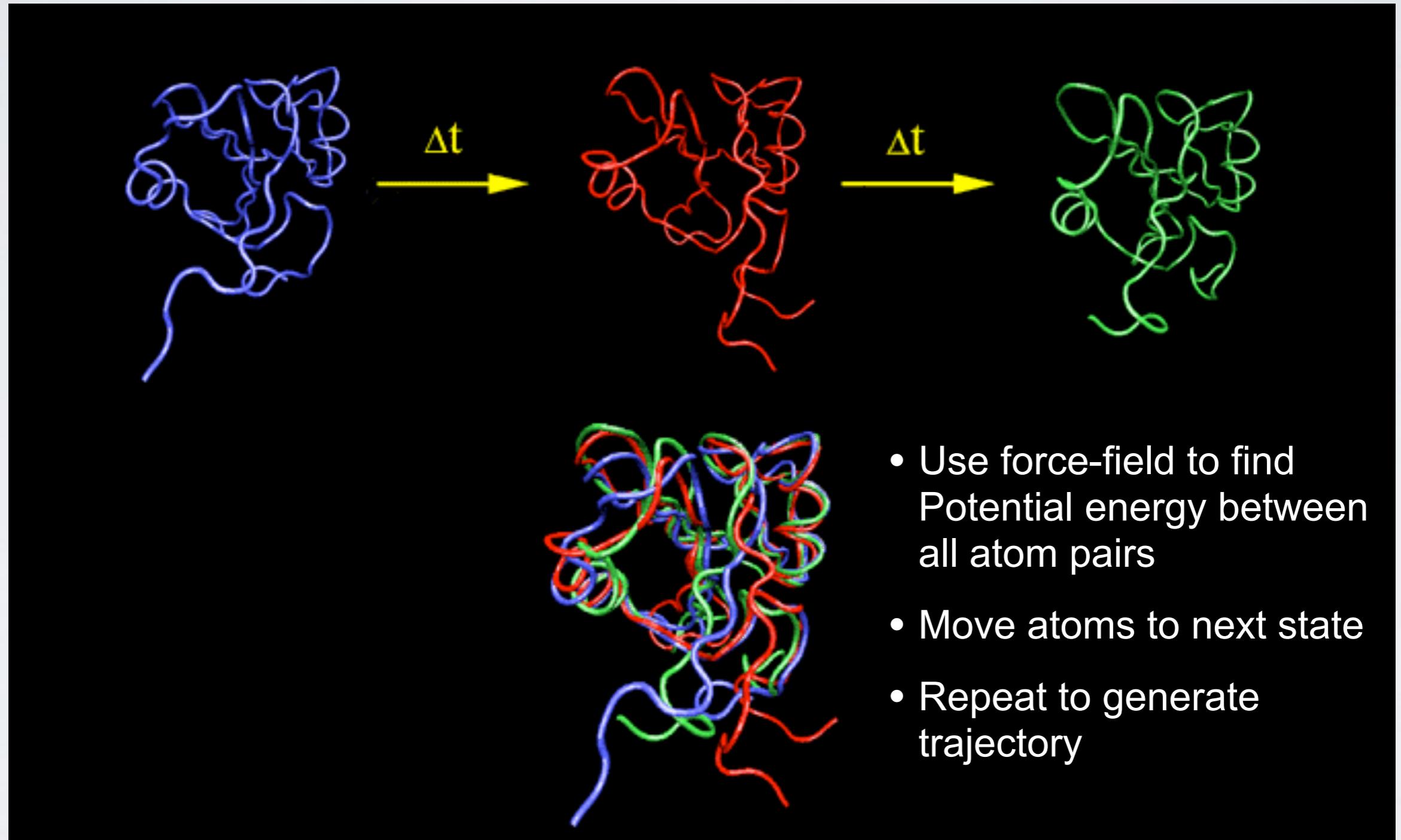
NEXT UP:

- ▶ **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**
 - Drug discovery & predicting functional dynamics

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



McCammon, Gelin & Karplus, *Nature* (1977)

[See: <https://www.youtube.com/watch?v=ui1ZysMFcKk>]

KEY CONCEPT: POTENTIAL FUNCTIONS
DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION
OF ITS **STRUCTURE**

Two main approaches:

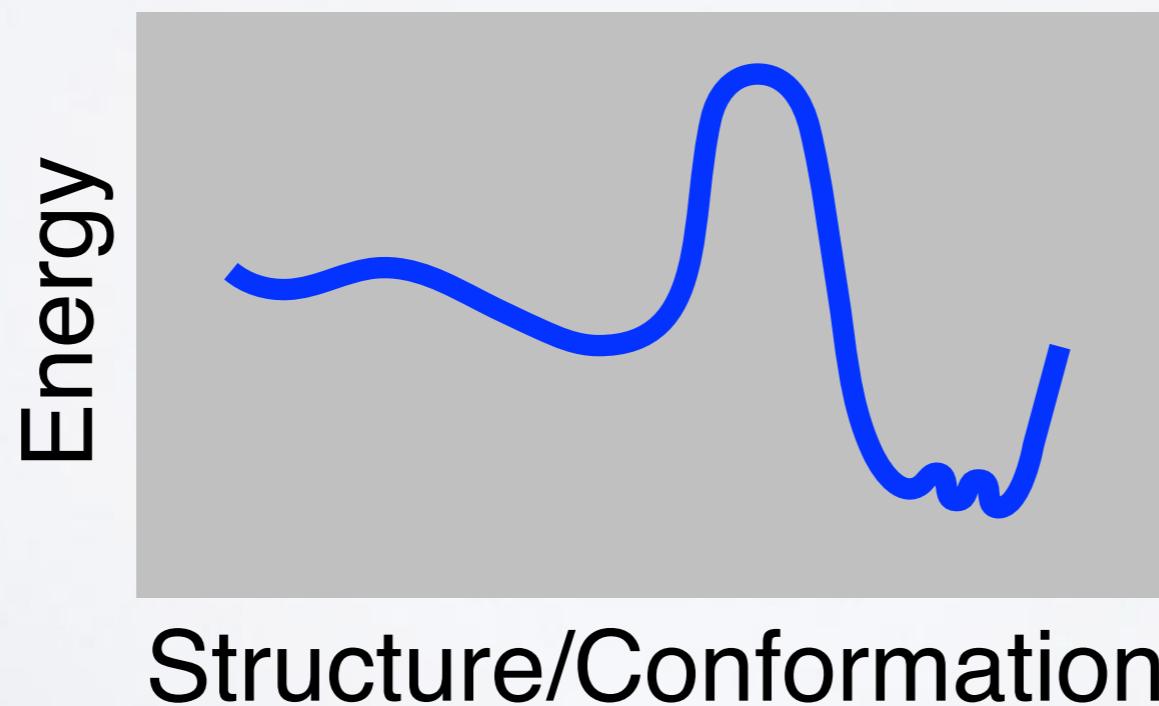
- (1). Physics-Based
- (2). Knowledge-Based

KEY CONCEPT: POTENTIAL FUNCTIONS

DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION
OF ITS **STRUCTURE**

Two main approaches:

- (1). Physics-Based
- (2). Knowledge-Based

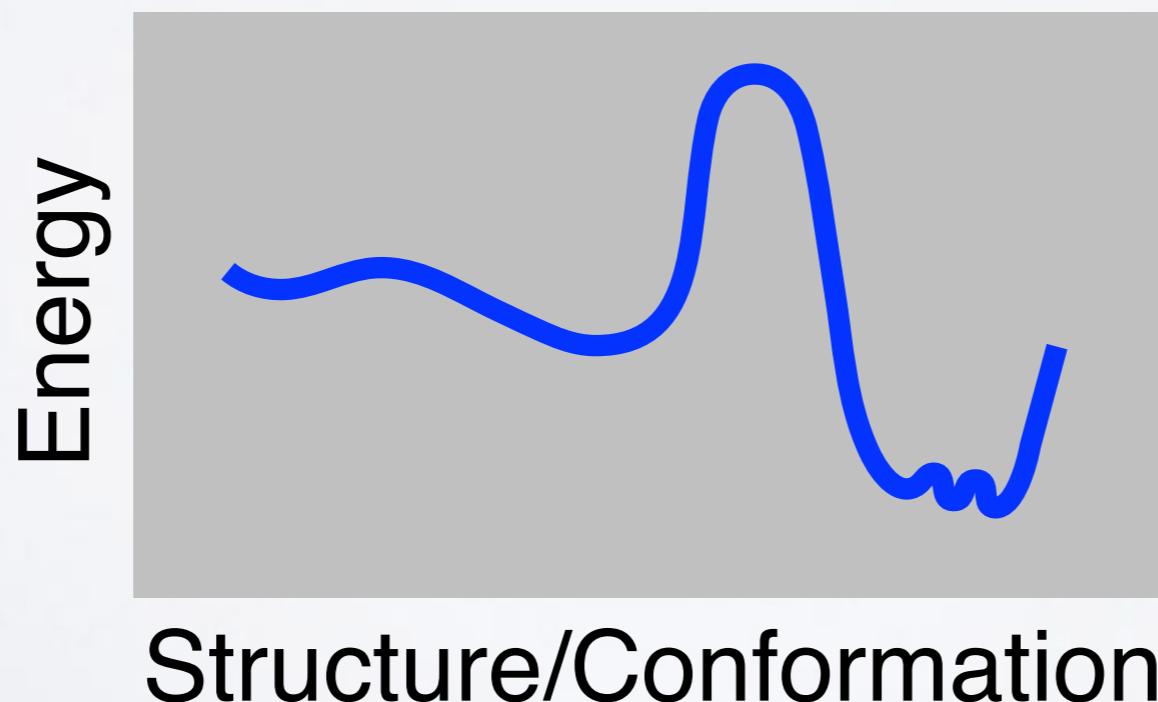


KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

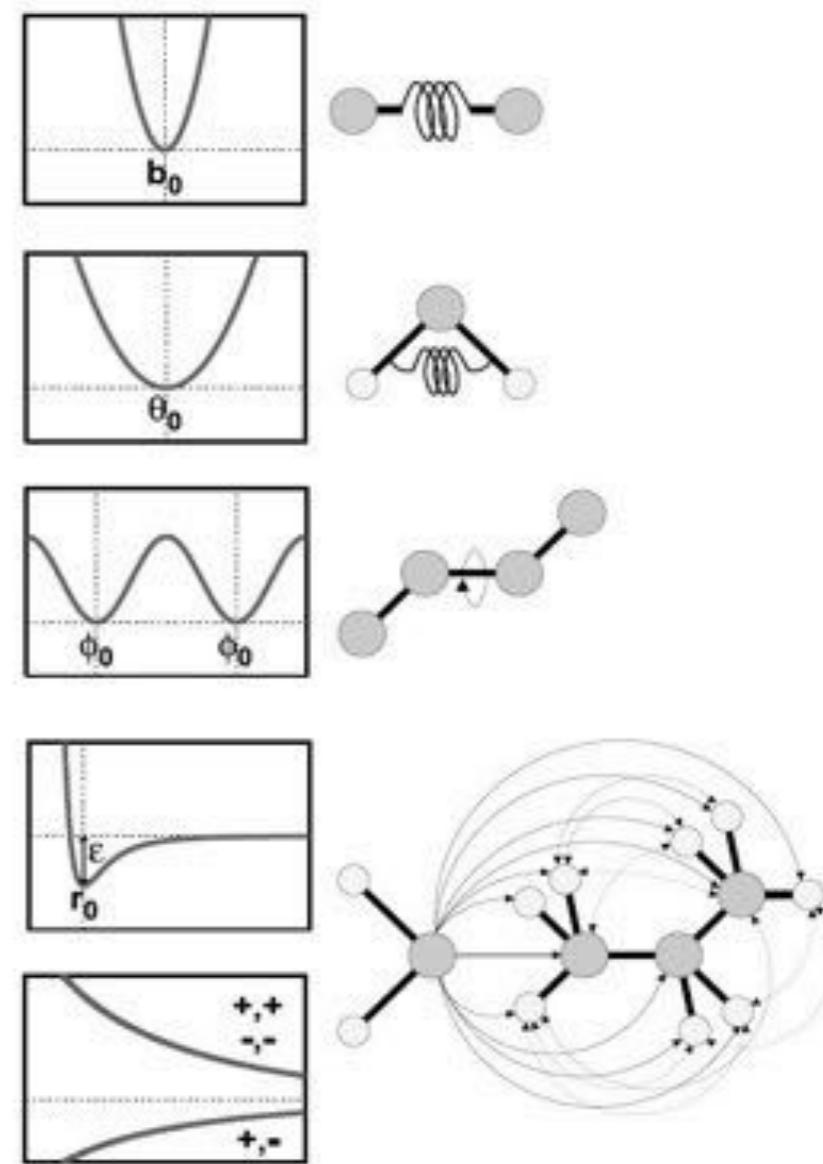
(1). **Physics-Based**

(2). **Knowledge-Based**



PHYSICS-BASED POTENTIALS ENERGY TERMS FROM PHYSICAL THEORY

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond} (r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle} (\theta_i - \theta_0)^2}_{U_{angle}} + \\ \underbrace{\sum_{dihedrals} k_i^{dihed} [1 + \cos(n_i \phi_i + \delta_i)]}_{U_{dihedral}} + \\ \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$



U_{bond} = oscillations about the equilibrium bond length

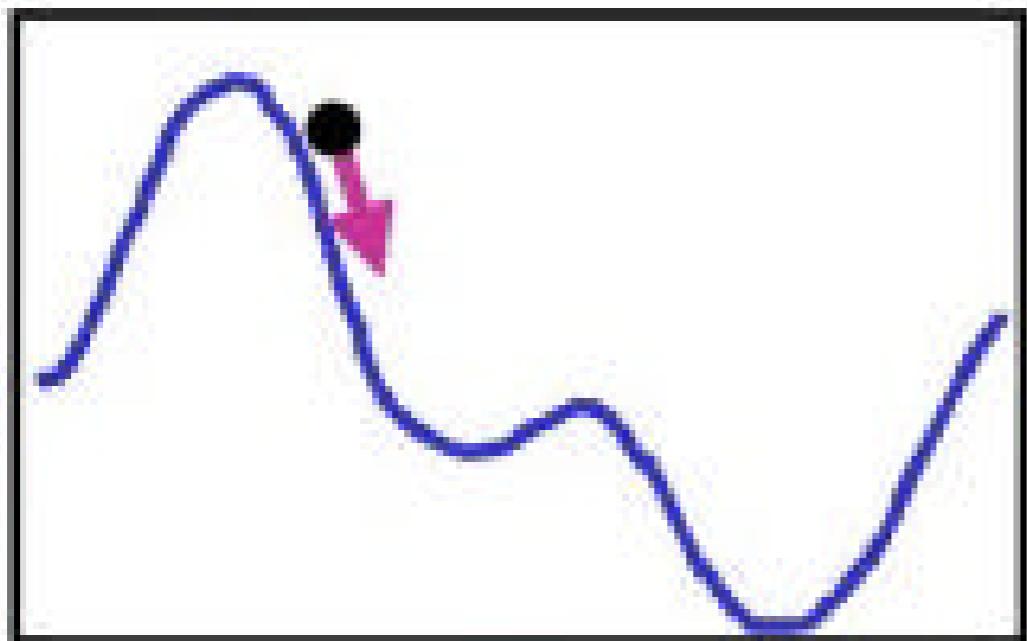
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle

$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond

$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

TOTAL POTENTIAL ENERGY

Energy, U



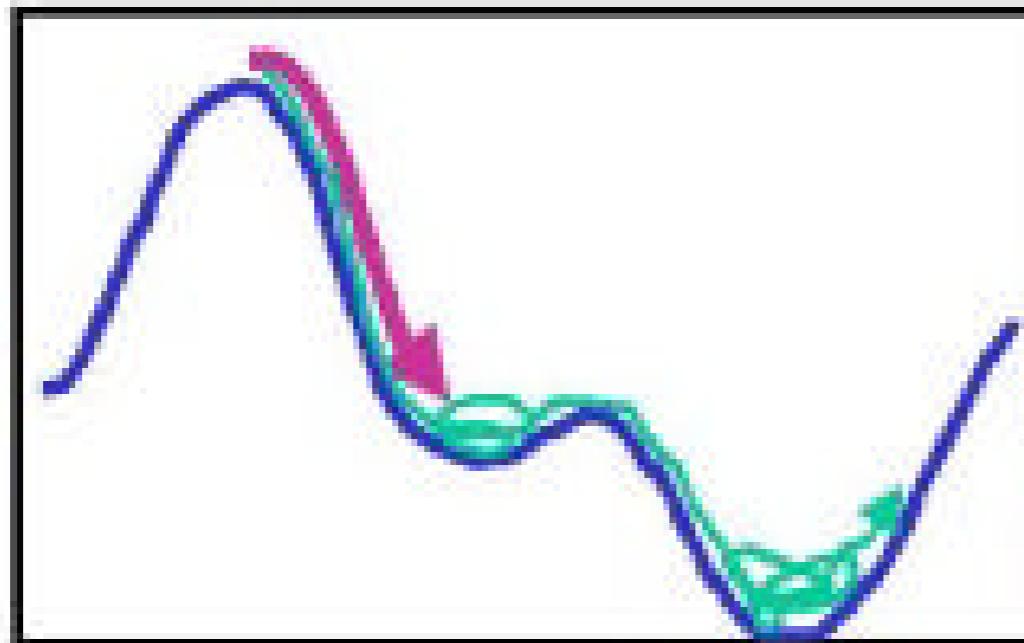
$$F(x) = -dU/dx$$

- The total potential energy or enthalpy fully defines the system, U .
- The forces are the gradients of the energy.

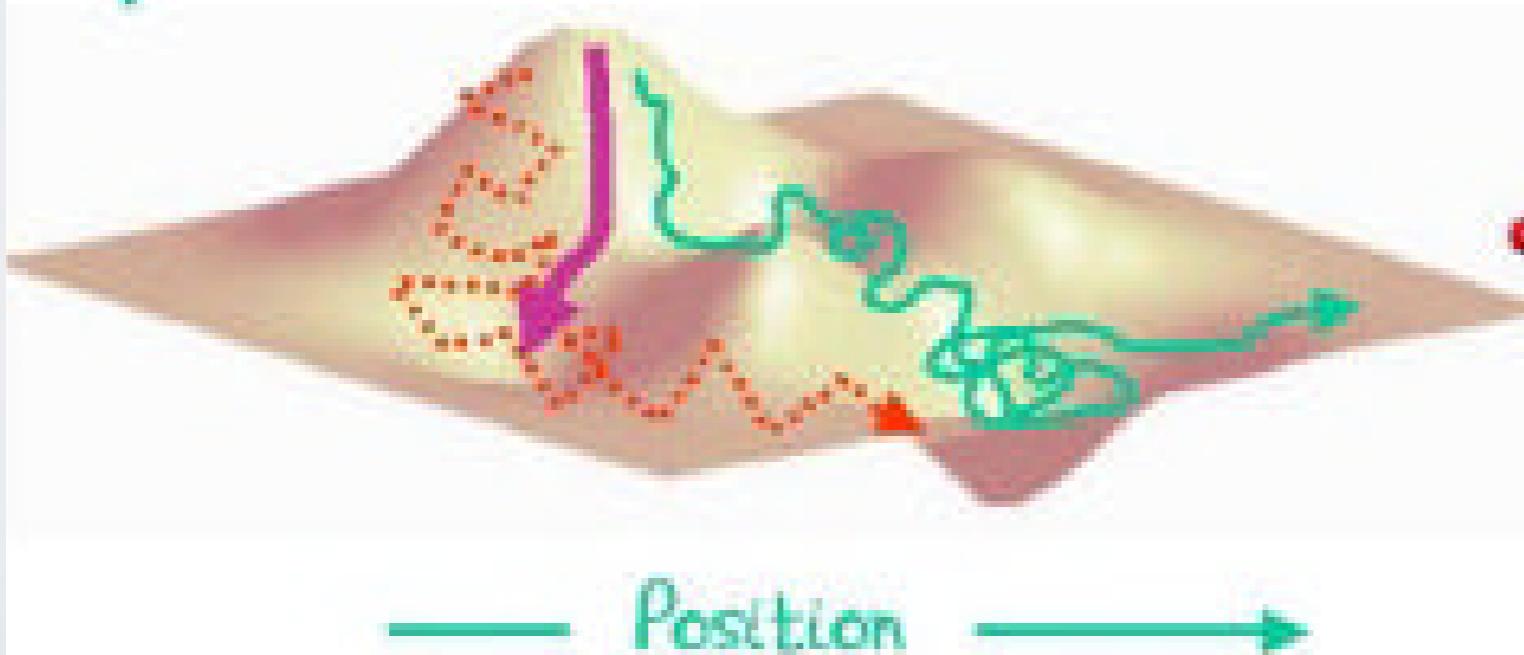
• The energy is a sum of independent terms for:
Bond, Bond angles,
Torsion angles and non-
bonded atom pairs.

MOVING OVER THE ENERGY SURFACE

Energy, E



- Energy Minimization drops into local minimum.



- Molecular Dynamics uses thermal energy to move smoothly over surface.

- Monte Carlo Moves are random. Accept with probability $\exp(-\Delta U/kT)$.

Slide Credit: Michael Levitt

PHYSICS-ORIENTED APPROACHES

Weaknesses

Fully physical detail becomes computationally intractable

Approximations are unavoidable

(Quantum effects approximated classically, water may be treated crudely)

Parameterization still required

Strengths

Interpretable, provides guides to design

Broadly applicable, in principle at least

Clear pathways to improving accuracy

Status

Useful, widely adopted but far from perfect

Multiple groups working on fewer, better approxs

Force fields, quantum
entropy, water effects

Moore's law: hardware improving

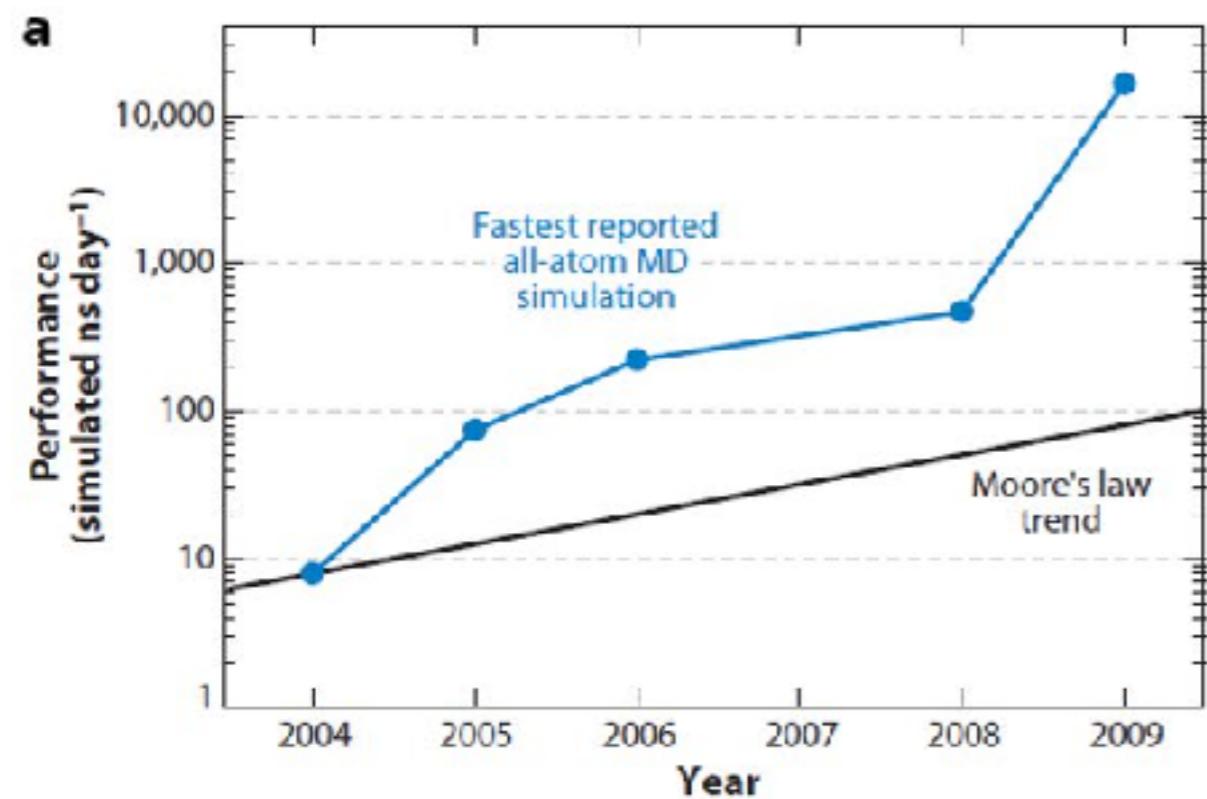
HOW COMPUTERS HAVE CHANGED

DATE	COST	SPEED	MEMORY	SIZE
1967	\$10M	0.1 MHz	1 MB	WALL
2013	\$14,000	1 GHz	10 GB	LAPTOP
CHANGE	10,000	10,000	10,000	10,000

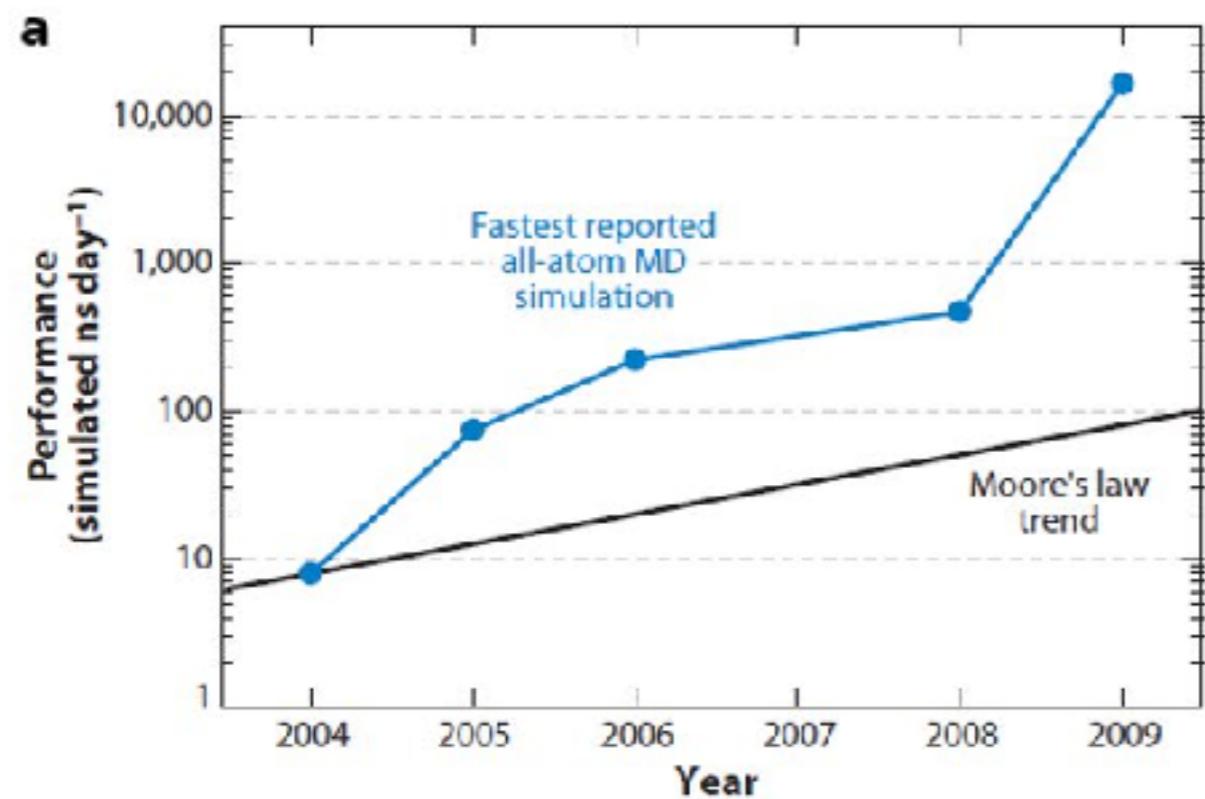
If cars were like computers then a new Vehc
would cost \$3, would have a top speed of
1,000,000 Km/hr, would carry 50,000
adults and would park in a shadow.



SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER



SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER

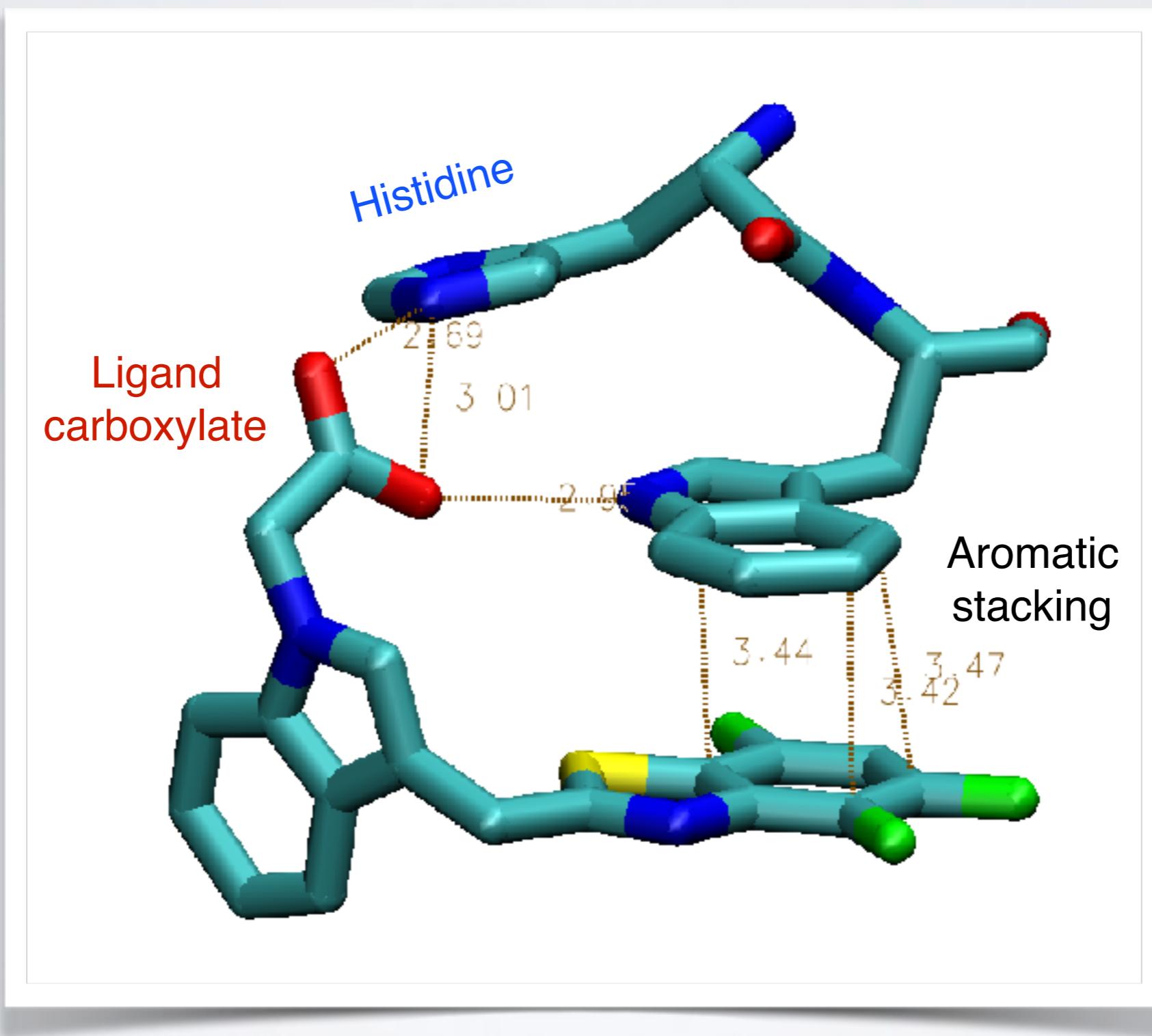


KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY AS A FUNCTION OF ITS **STRUCTURE****

Two main approaches:

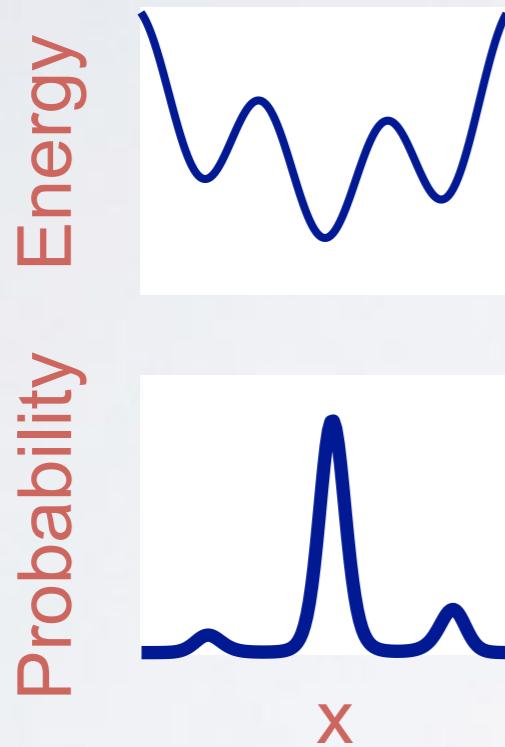
- (1). Physics-Based
- (2). Knowledge-Based

KNOWLEDGE-BASED DOCKING POTENTIALS



ENERGY DETERMINES **PROBABILITY** (STABILITY)

Basic idea: Use probability as a proxy for energy



Boltzmann:

$$p(r) \propto e^{-E(r)/RT}$$

Inverse Boltzmann:

$$E(r) = -RT \ln[p(r)]$$

Example: ligand carboxylate O to protein histidine N

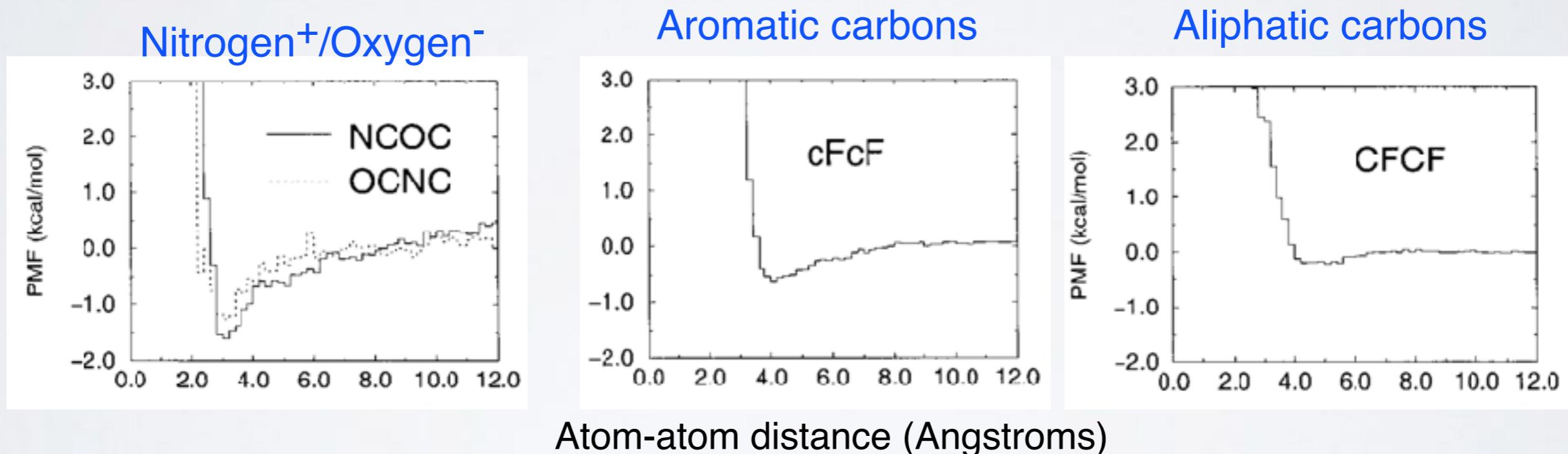
Find all protein-ligand structures in the PDB with a ligand carboxylate O

1. For each structure, histogram the distances from O to every histidine N
2. Sum the histograms over all structures to obtain $p(r_{O-N})$
3. Compute $E(r_{O-N})$ from $p(r_{O-N})$

KNOWLEDGE-BASED DOCKING POTENTIALS

“PMF”, Muegge & Martin, J. Med. Chem. (1999) 42:791

A few types of atom pairs, out of several hundred total



$$E_{prot-lig} = E_{vdw} + \sum_{pairs(ij)} E_{type(ij)}(r_{ij})$$

KNOWLEDGE-BASED POTENTIALS

Weaknesses

Accuracy limited by availability of data

Strengths

Relatively easy to implement

Computationally fast

Status

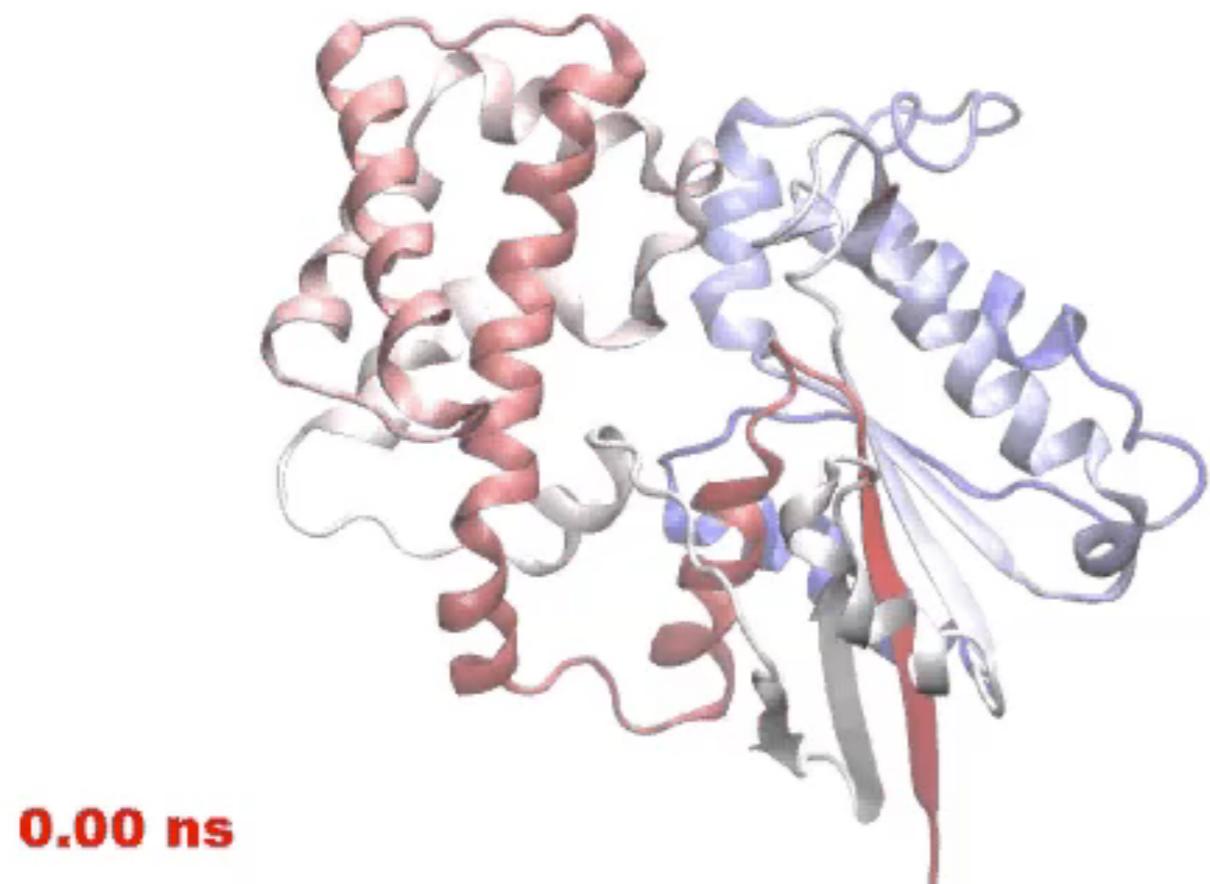
Useful, far from perfect

May be at point of diminishing returns

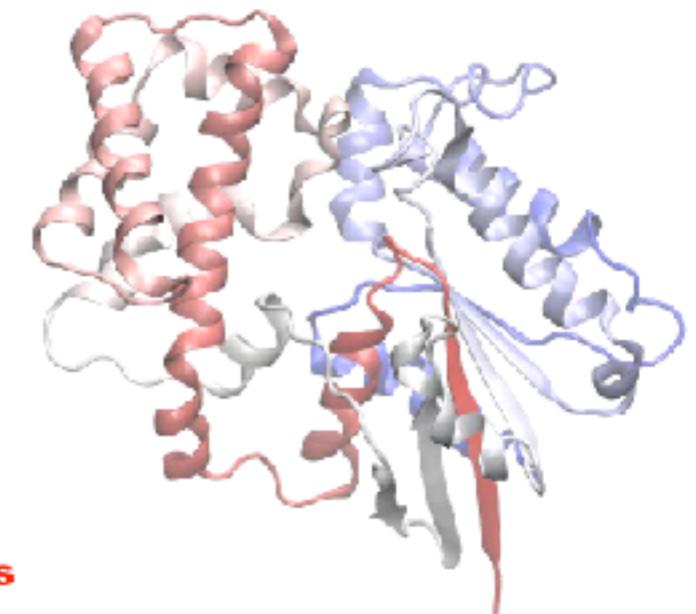
(not always clear how to make improvements)

MD Prediction of Functional Motions

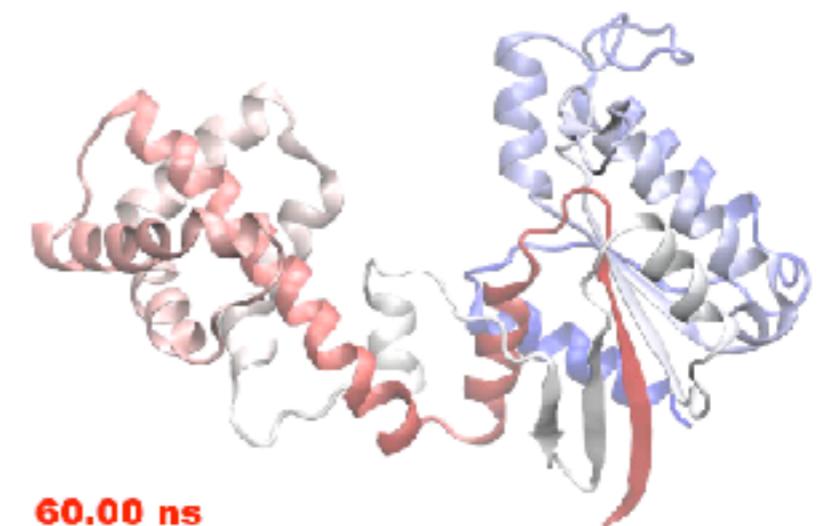
Accelerated MD simulation of
nucleotide-free transducin alpha subunit



“close”



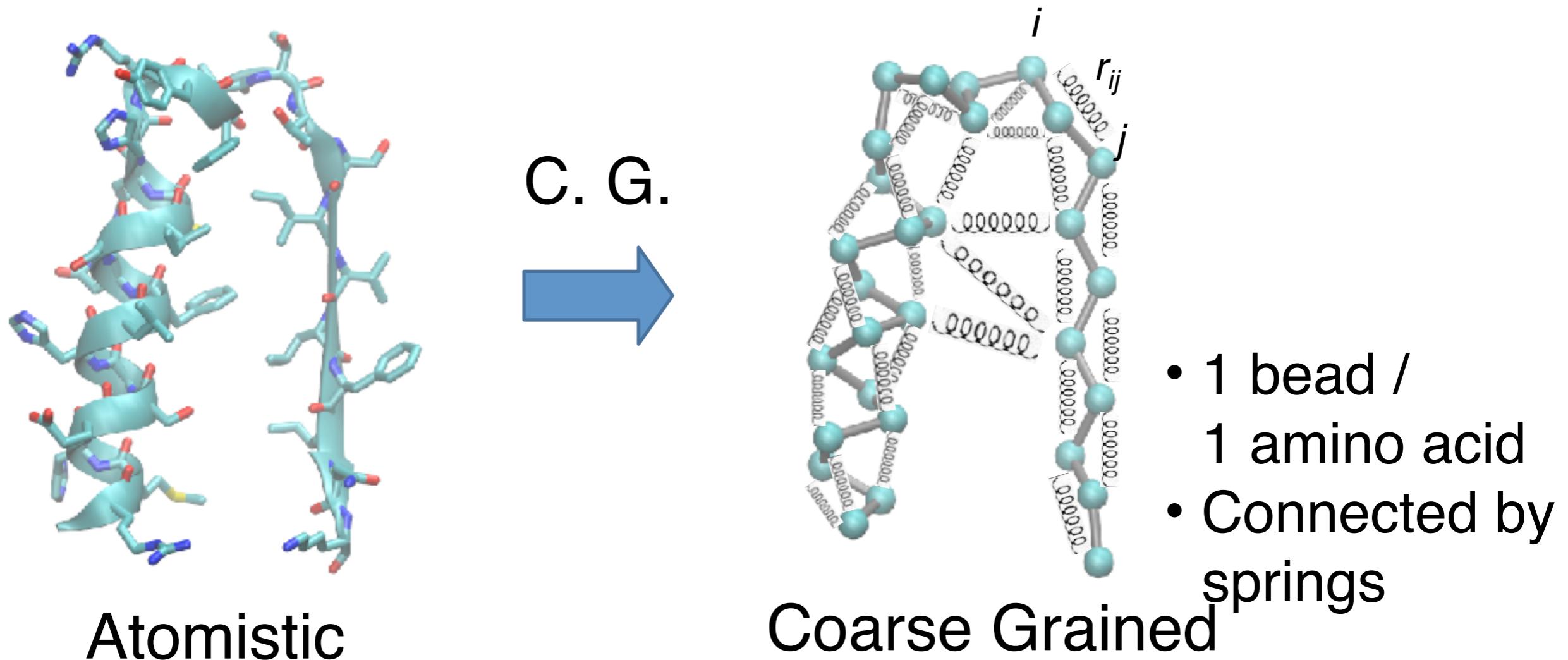
“open”



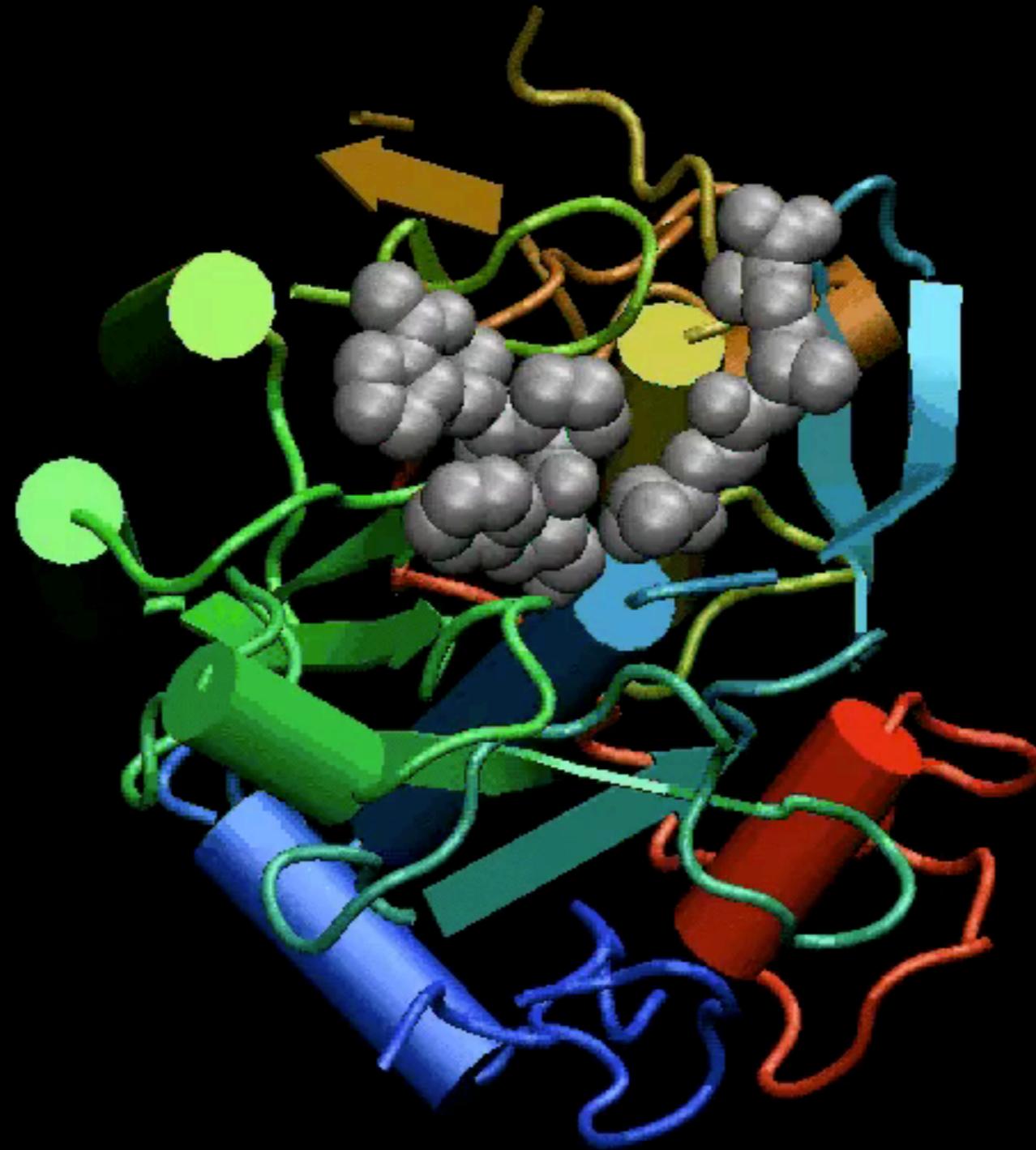
Yao and Grant, Biophys J. (2013)

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



Proteinase K

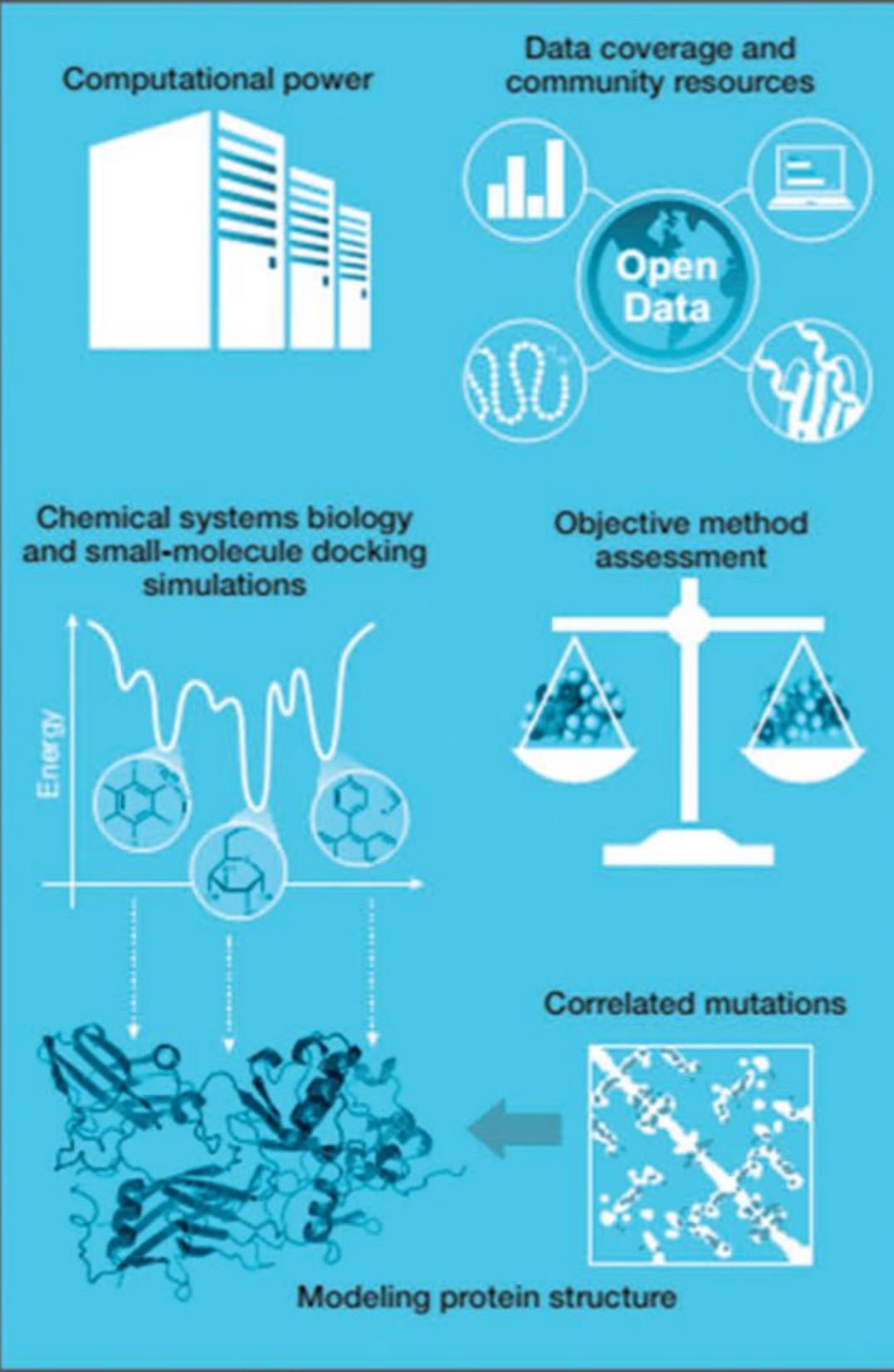
Do it Yourself!

Hand-on time!

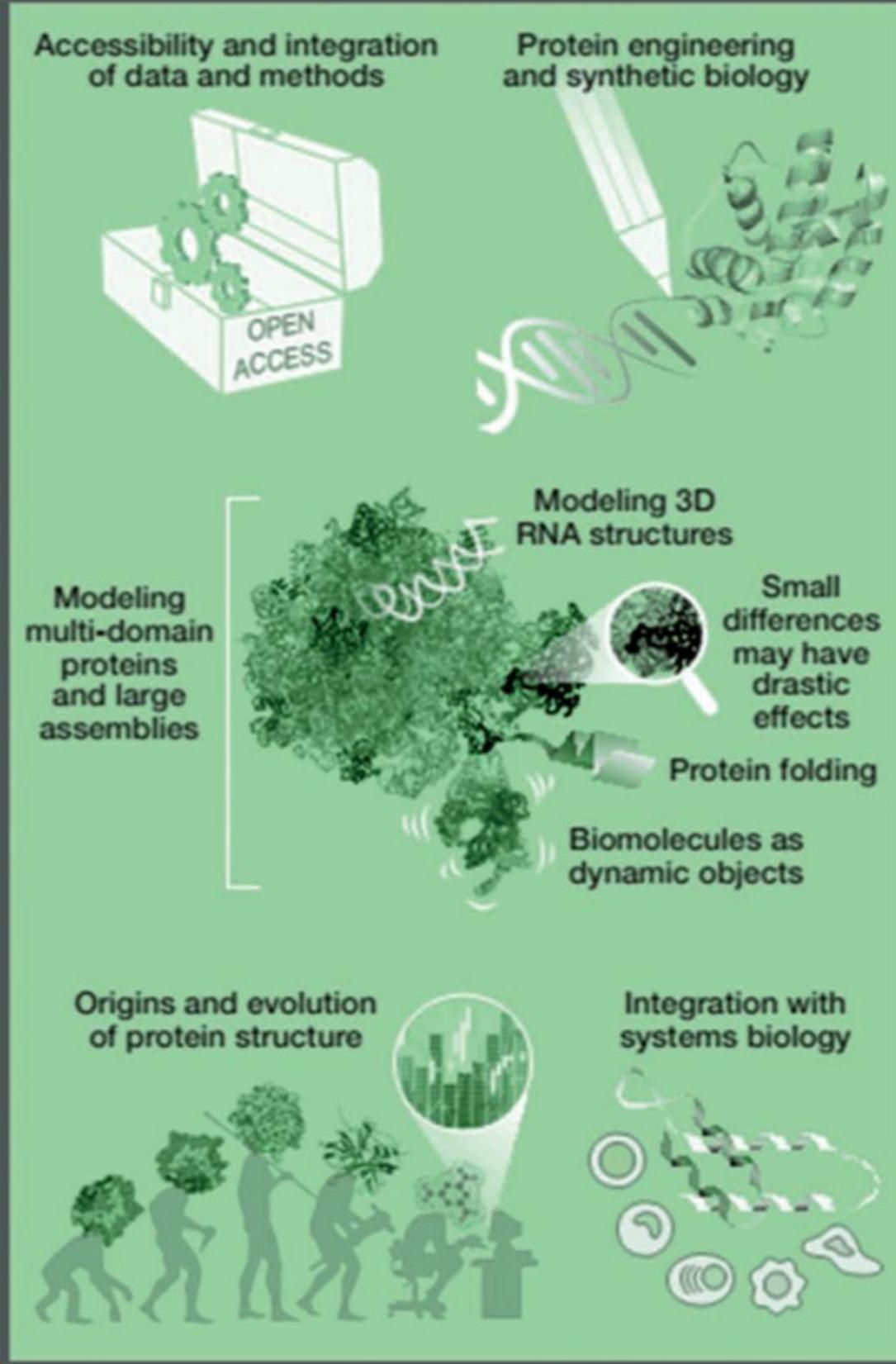
https://bioboot.github.io/bimm143_S18/lectures/#12

Focus on **section 3 & 4** exploring **PCA** and **NMA apps**

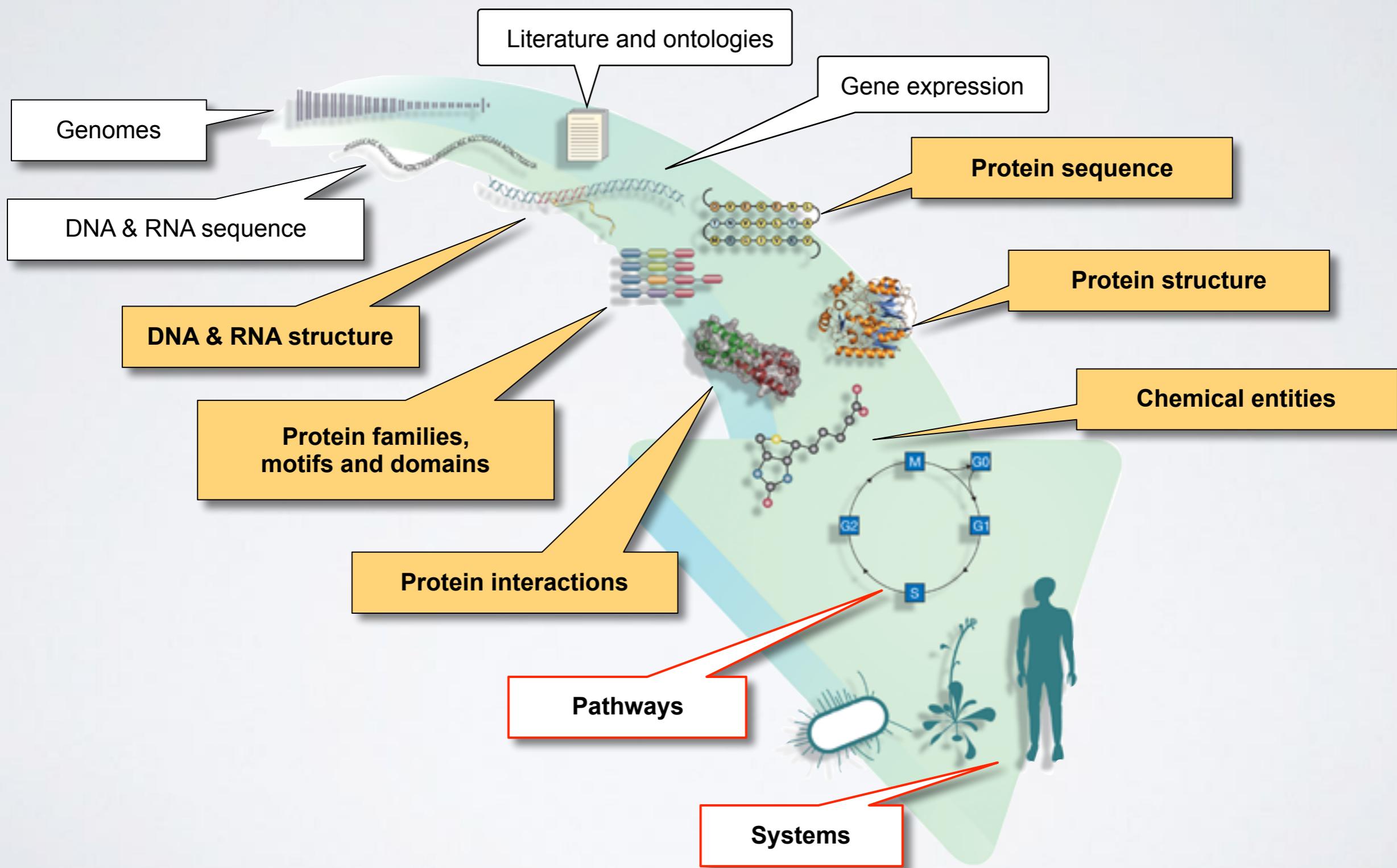
ACHIEVEMENTS



CHALLENGES



INFORMING SYSTEMS BIOLOGY?



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