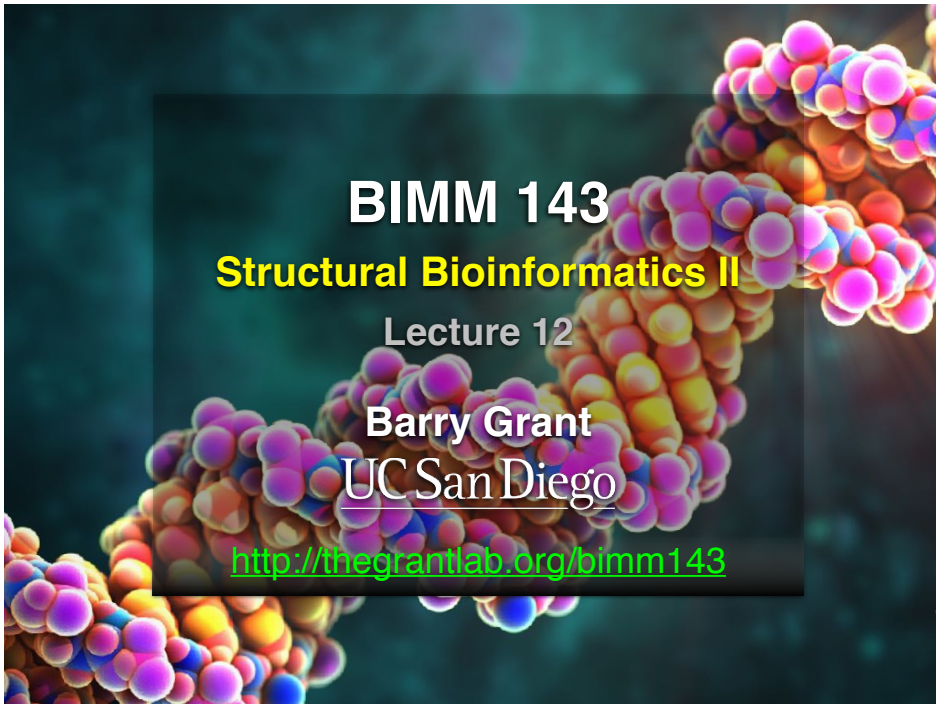
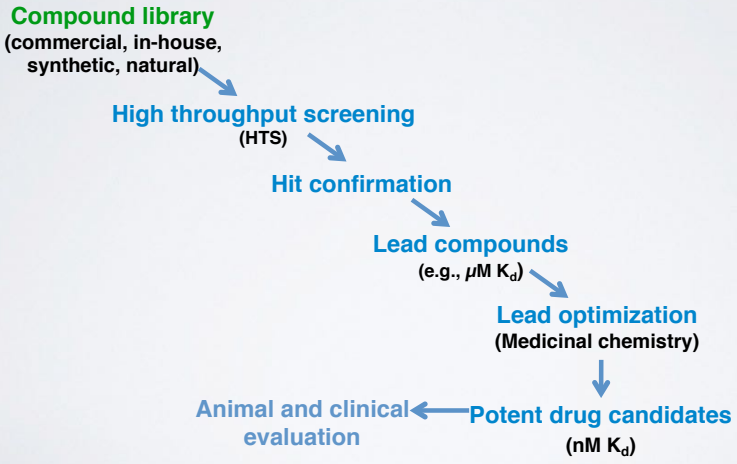



## NEXT UP:

- **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



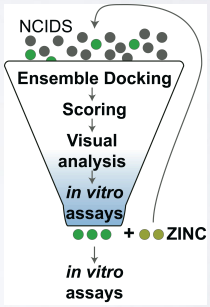

## COMPUTER-AIDED LIGAND DESIGN

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**

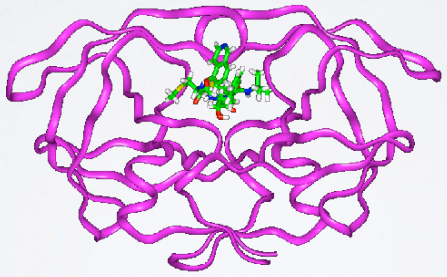
Two main approaches:

(1). **Receptor/Target-Based**

(2). **Ligand/Drug-Based**

# **SCENARIO I**: RECEPTOR-BASED DRUG DISCOVERY

Structure of Targeted Protein Known: Structure-Based Drug Discovery

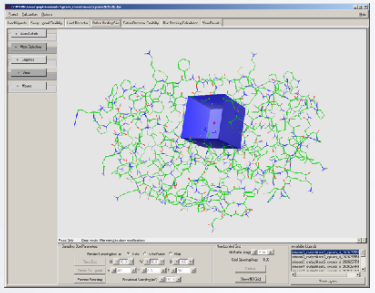


HIV Protease/KNI-272 complex

# PROTEIN-LIGAND DOCKING

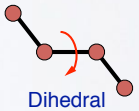
## Structure-Based Ligand Design

**Docking software**
Search for structure of lowest energy

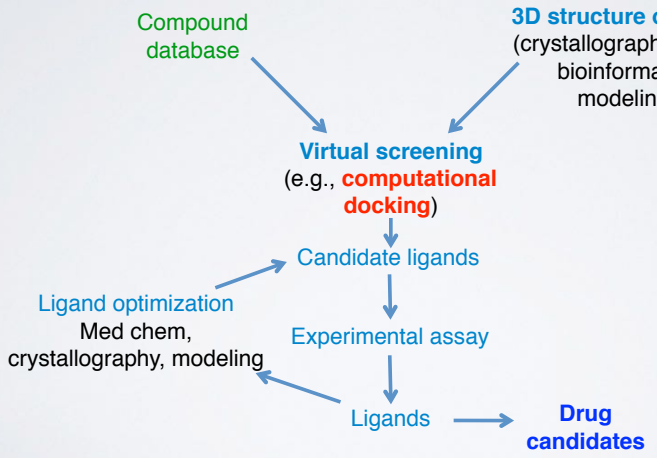


**Potential function**
Energy as function of structure





# STRUCTURE-BASED VIRTUAL SCREENING



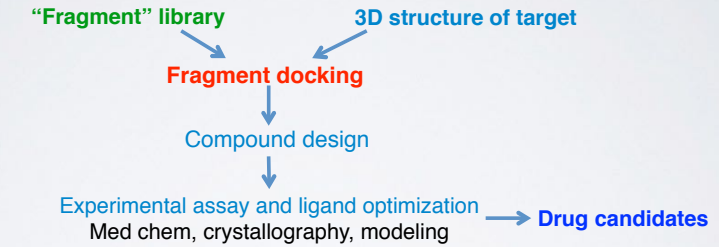

# COMPOUND LIBRARIES

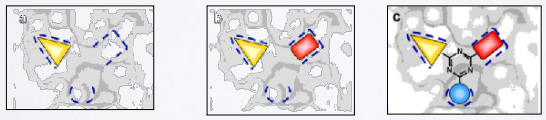
Commercial (in-house pharma)


Government (NIH)


Academia

# FRAGMENTAL STRUCTURE-BASED SCREENING

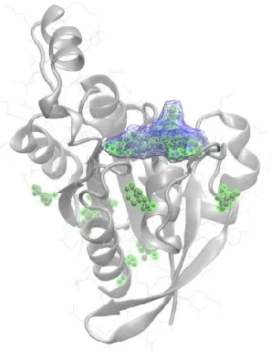


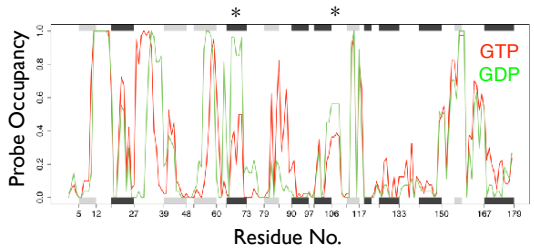



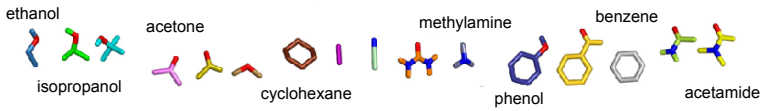
http://www.beilstein-institut.de/bozen2002/proceedings/Jhoti/jhoti.html

## Multiple non active-site pockets identified

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



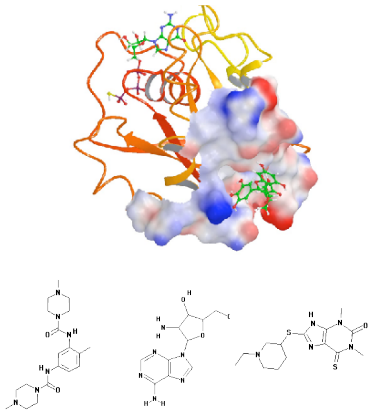


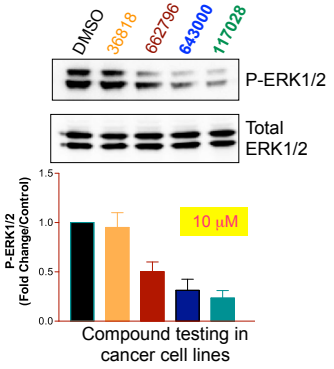
## Ensemble docking & candidate inhibitor testing

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

Ensemble computational docking

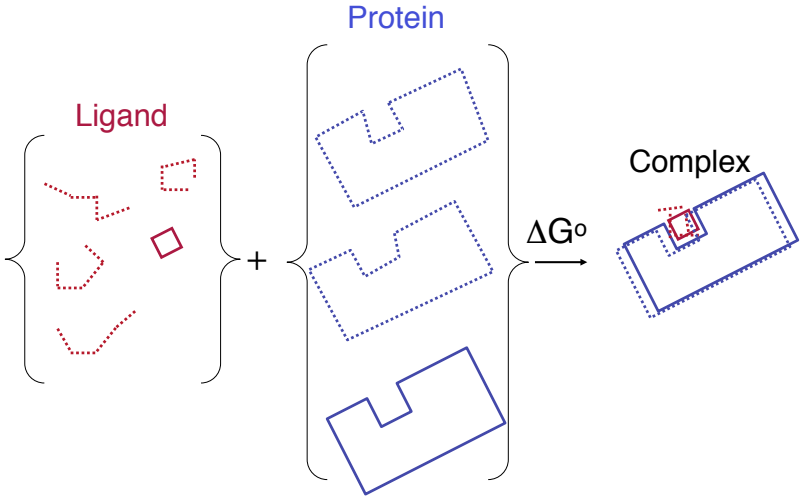


Compound effect on U251 cell line




PLoS One (2011, 2012)

## 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**

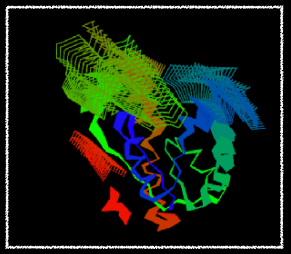
Do it Yourself!

# 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**:

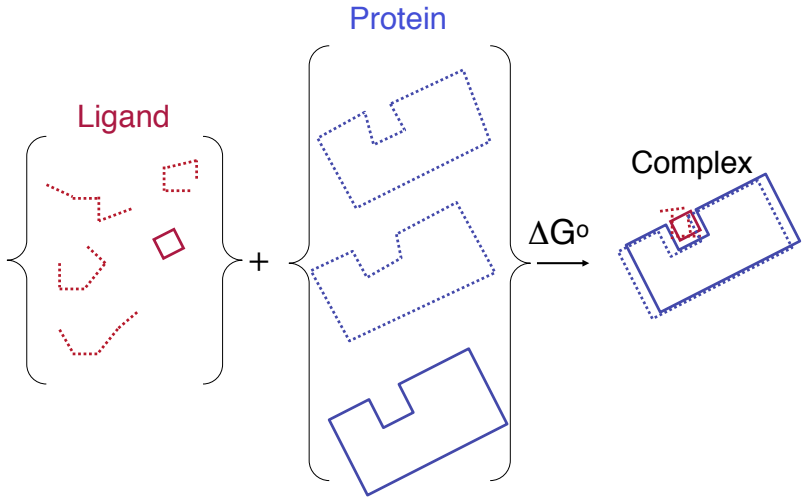
```
> download.file("https://tinyurl.com/bio3d-mac", "bio3d.tar.gz")
> install.packages("bio3d.tar.gz", repos = NULL)
```

- 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:

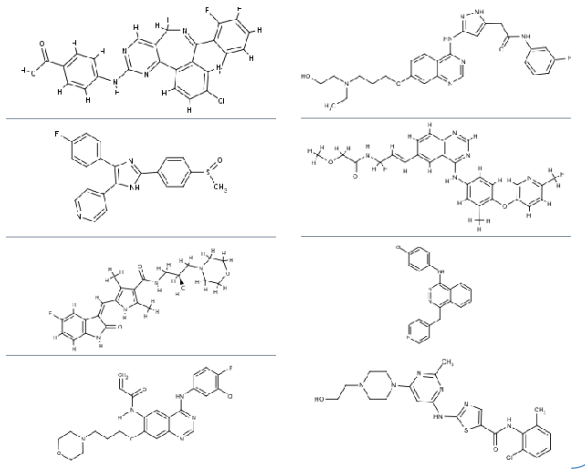
(1). **Receptor/Target-Based**

(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

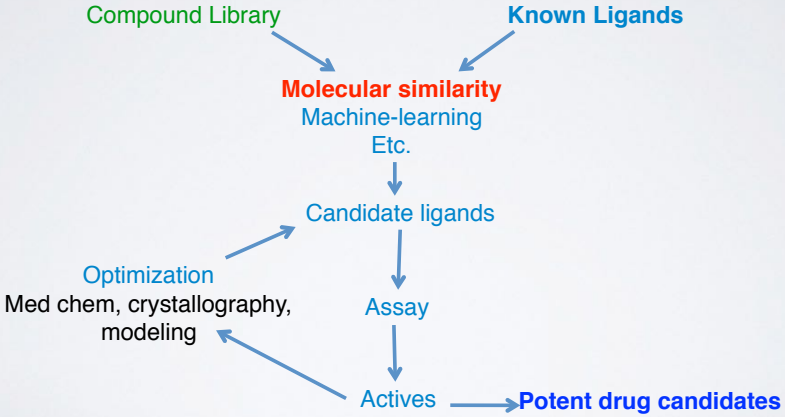
## Why Look for Another Ligand if You Already Have Some?

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

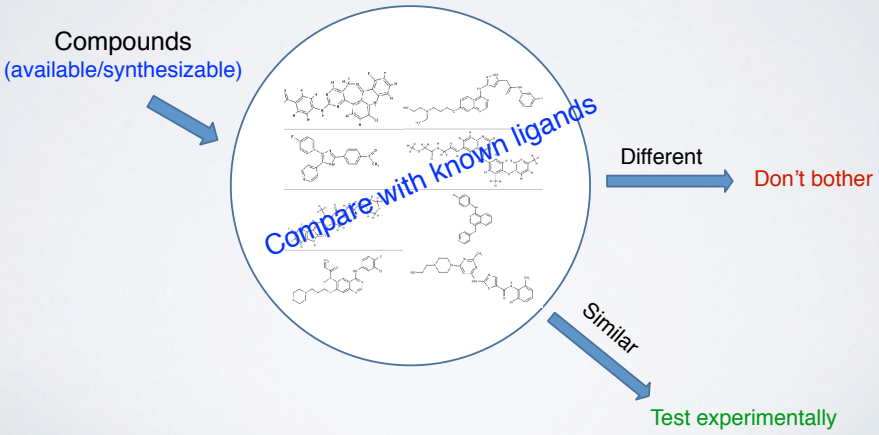
A company wants to work around another company's chemical patents

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

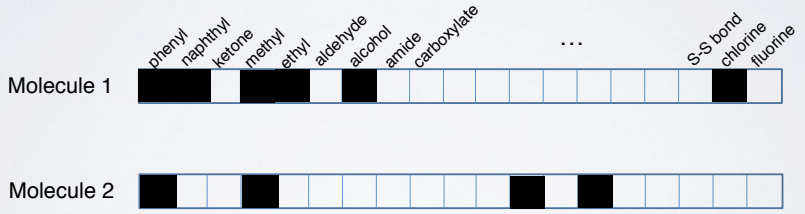
## LIGAND-BASED VIRTUAL SCREENING



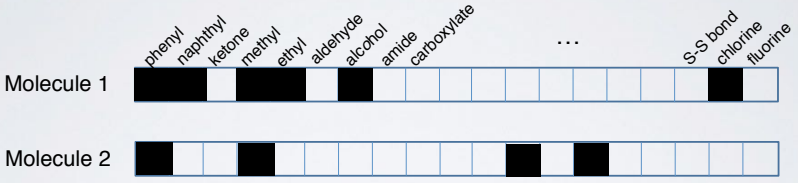
# CHEMICAL SIMILARITY
# LIGAND-BASED DRUG-DISCOVERY




# CHEMICAL FINGERPRINTS
## BINARY STRUCTURE KEYS

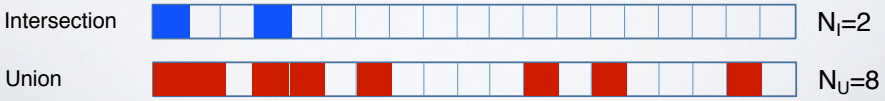



# CHEMICAL SIMILARITY FROM FINGERPRINTS



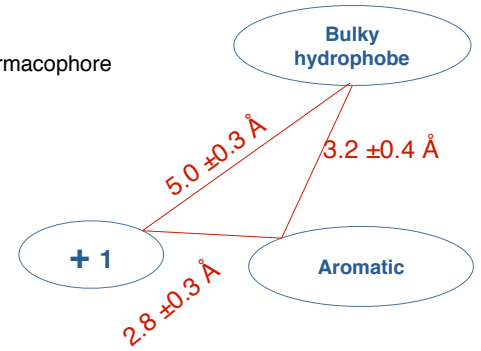

Tanimoto Similarity
(or Jaccard Index), T

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




# Pharmacophore Models
Φάρμακο (drug) + Φορά (carry)

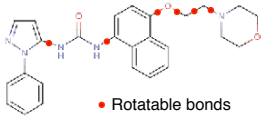
A 3-point pharmacophore



# 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 clogP)



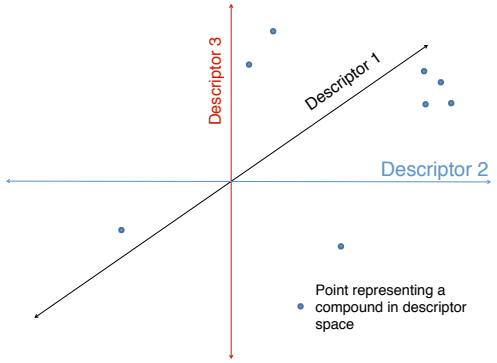

Topological
- branching index
- measures of linearity vs interconnectedness

Etc. etc.

# A High-Dimensional "Chemical Space"

## Each compound is at a point in an n-dimensional space

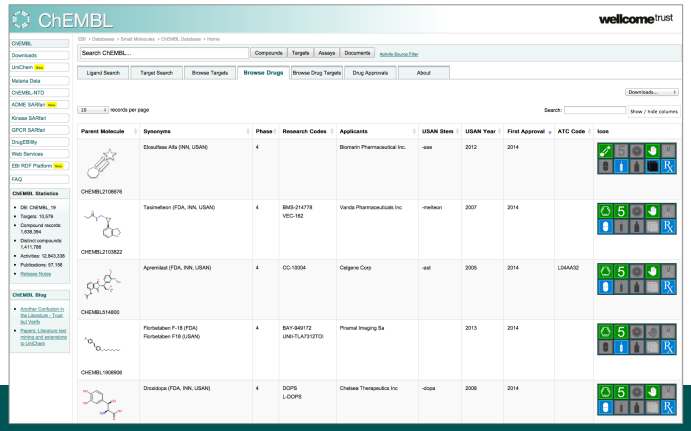
Compounds with similar properties are near each other




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

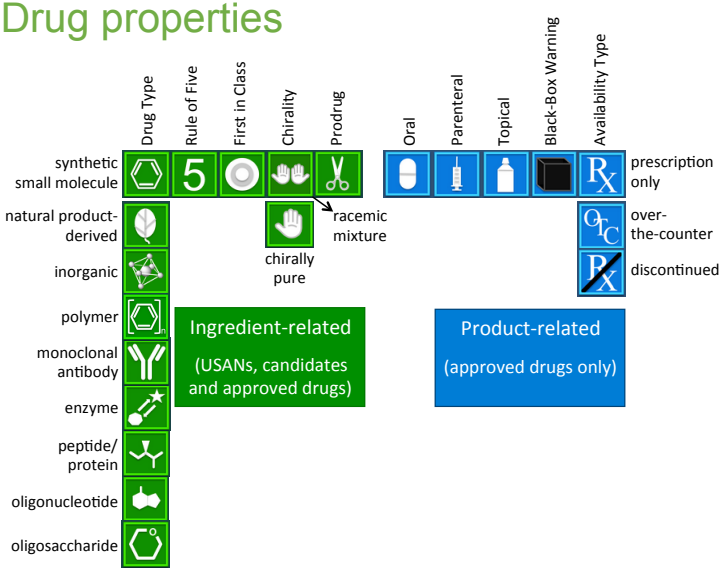
# Approved drugs and clinical candidates

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






# Drug properties



# LIPINSKI'S RULE OF FIVE

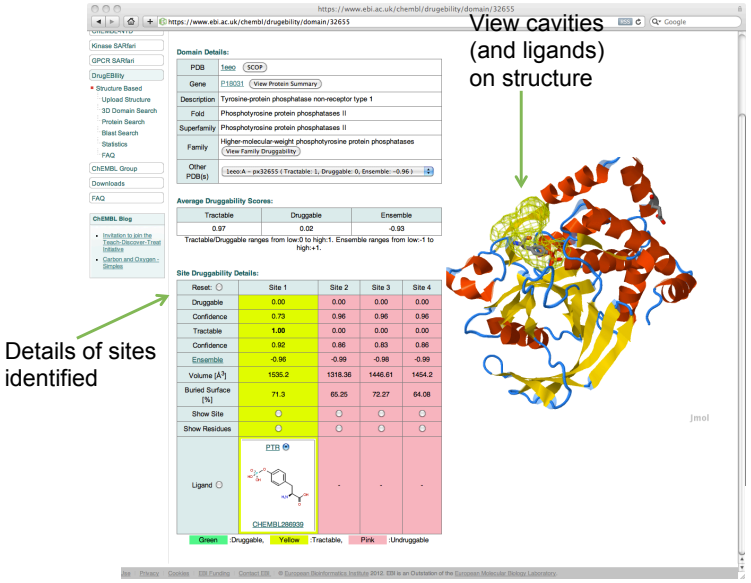
Lipinski's rule of five states that, in general, an orally active drug has no more than one violation of the following criteria:

- 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?

## Druggability prediction



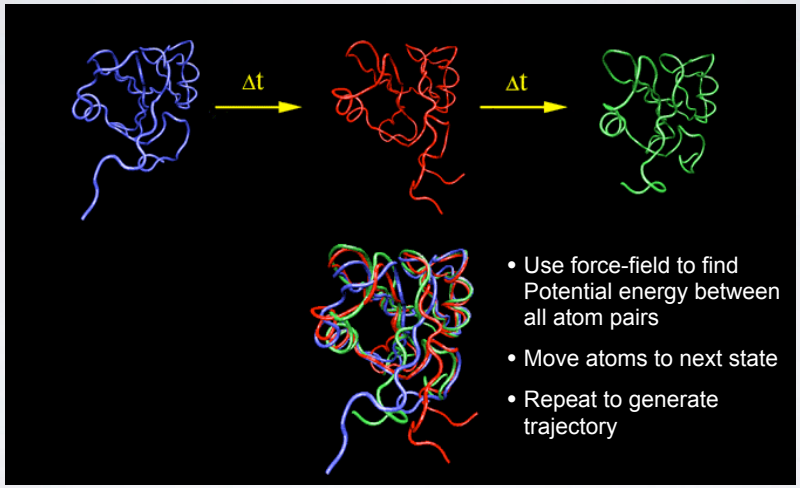

## NEXT UP:

- **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**

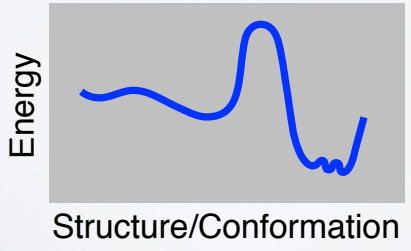
(2). **Knowledge-Based**


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

Two main approaches:

(1). **Physics-Based**

(2). **Knowledge-Based**

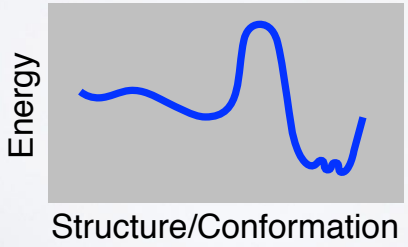



**KEY CONCEPT**: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

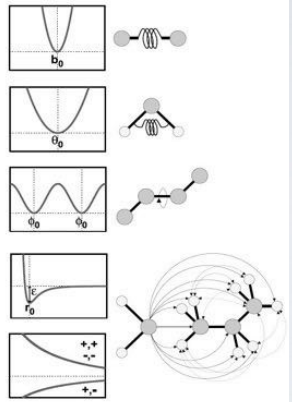
(1). **Physics-Based**

(2). **Knowledge-Based**



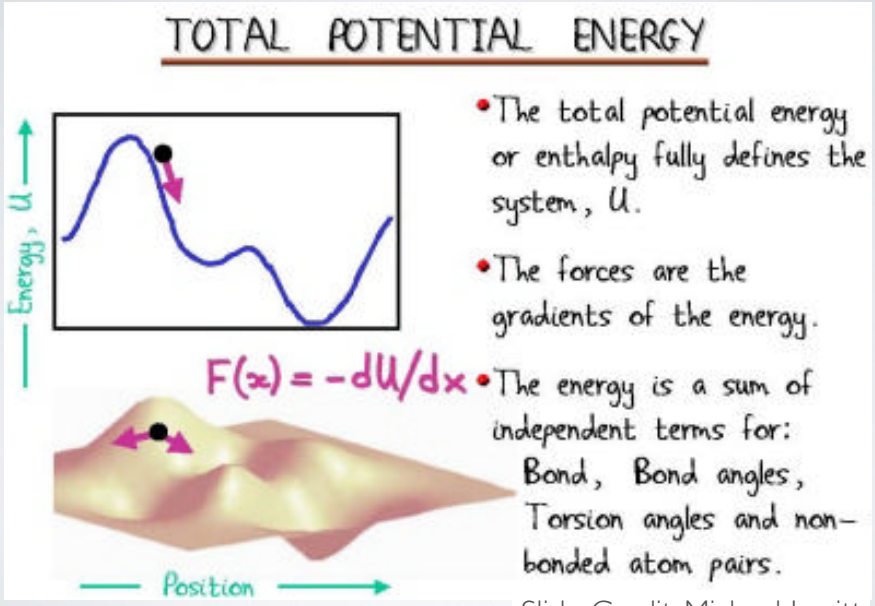

## **PHYSICS-BASED** POTENTIALS
## ENERGY TERMS FROM PHYSICAL THEORY

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

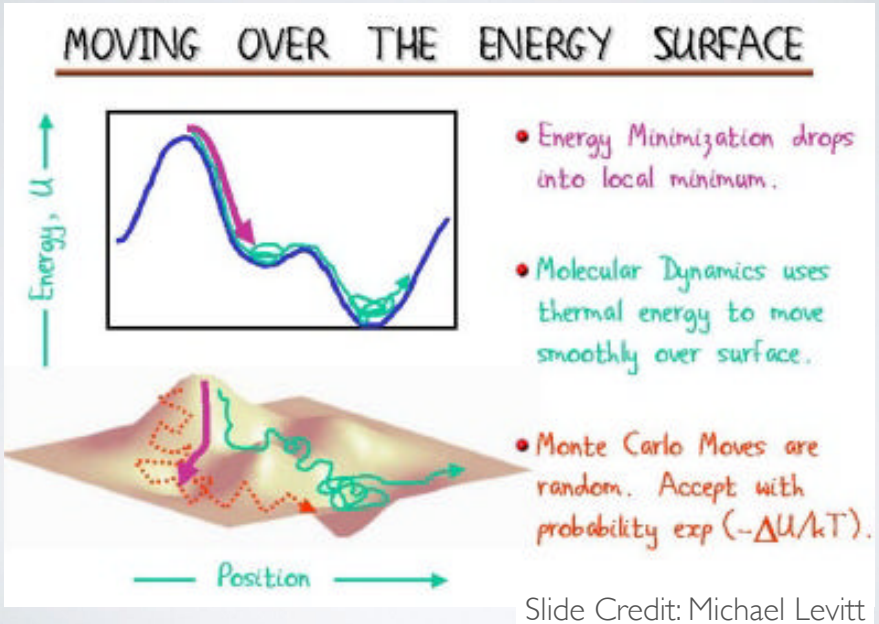


$U_{bond}$ = oscillations about the equilibrium bond length
$U_{angle}$ = oscillations of 3 atoms about an equilibrium bond angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

CHARMM P.E. function, see: http://www.charmm.org/




Slide Credit: Michael Levitt




Slide Credit: Michael Levitt

# PHYSICS-ORIENTED APPROACHES

## Weaknesses

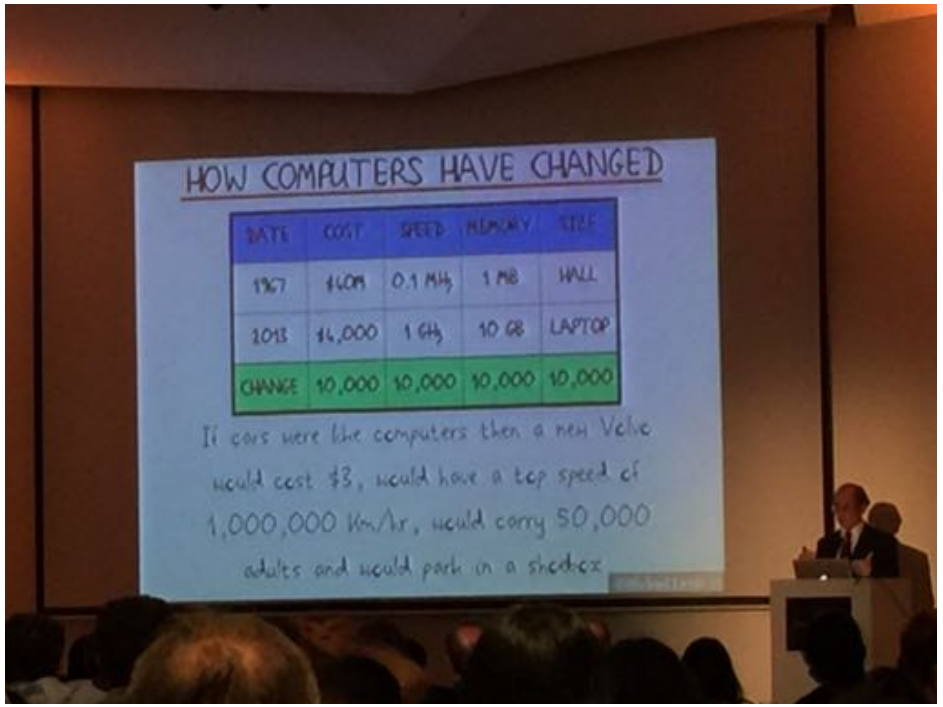
- Fully physical detail becomes computationally intractable
- Approximations are unavoidable
  - (Quantum effects approximated classically, water may be treated crudely)
- Parameterization still required

## Strengths

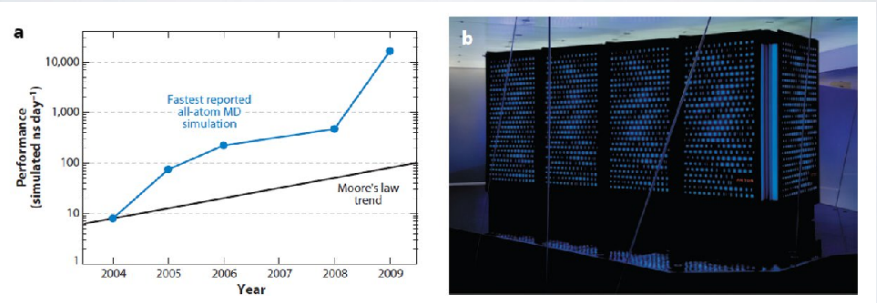
- Interpretable, provides guides to design
- Broadly applicable, in principle at least
- Clear pathways to improving accuracy

## 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

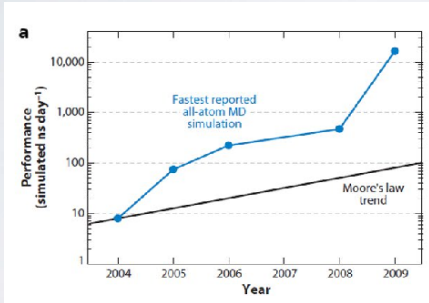



# SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER





# SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER

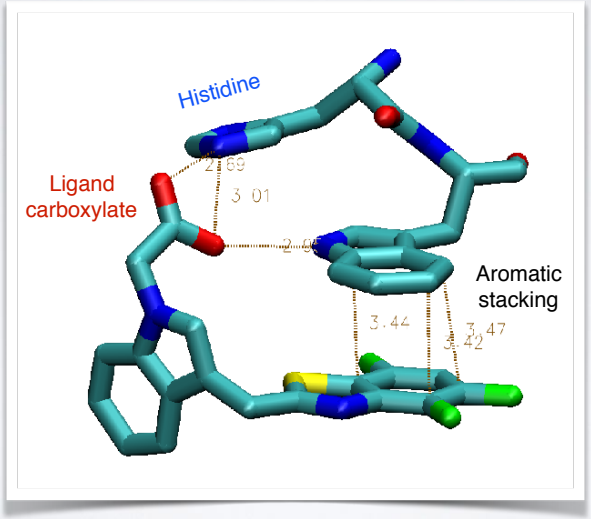

## **KEY CONCEPT**: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

(1). **Physics-Based**

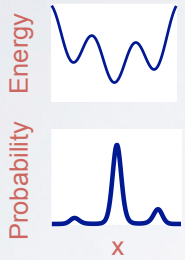
(2). **Knowledge-Based**

## 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:

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

Example: ligand carboxylate O to protein histidine N

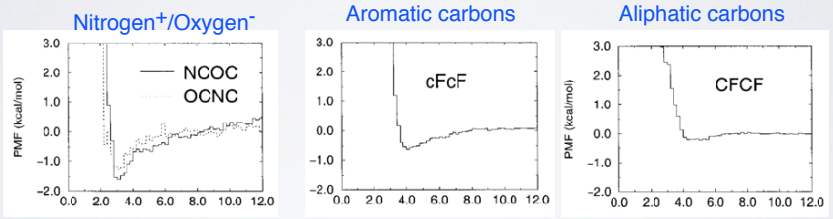
Find all protein-ligand structures in the PDB with a ligand carboxylate O

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

## 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

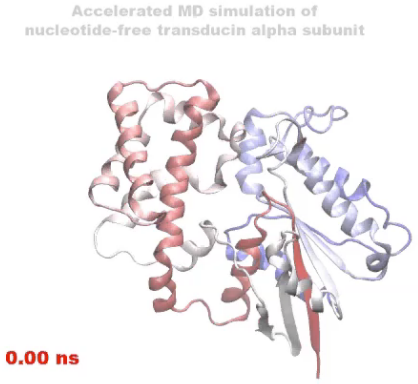
### Strengths

Relatively easy to implement
Computationally fast

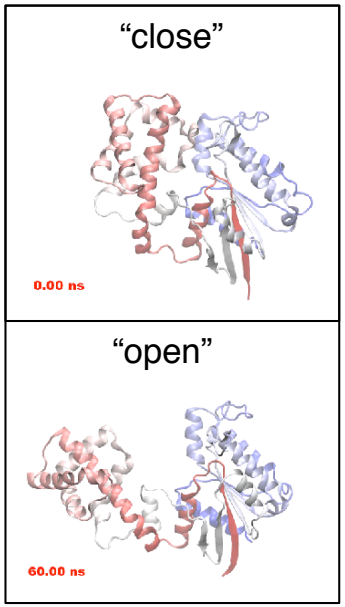
### Status

Useful, far from perfect
May be at point of diminishing returns
(not always clear how to make improvements)

## MD Prediction of Functional Motions

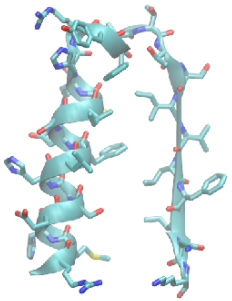



Yao and Grant, Biophys J. (2013)



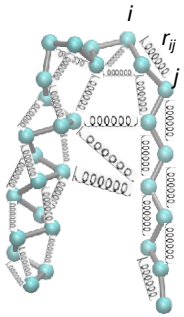

## COARSE GRAINING: **NORMAL MODE ANALYSIS** (NMA)

- 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.



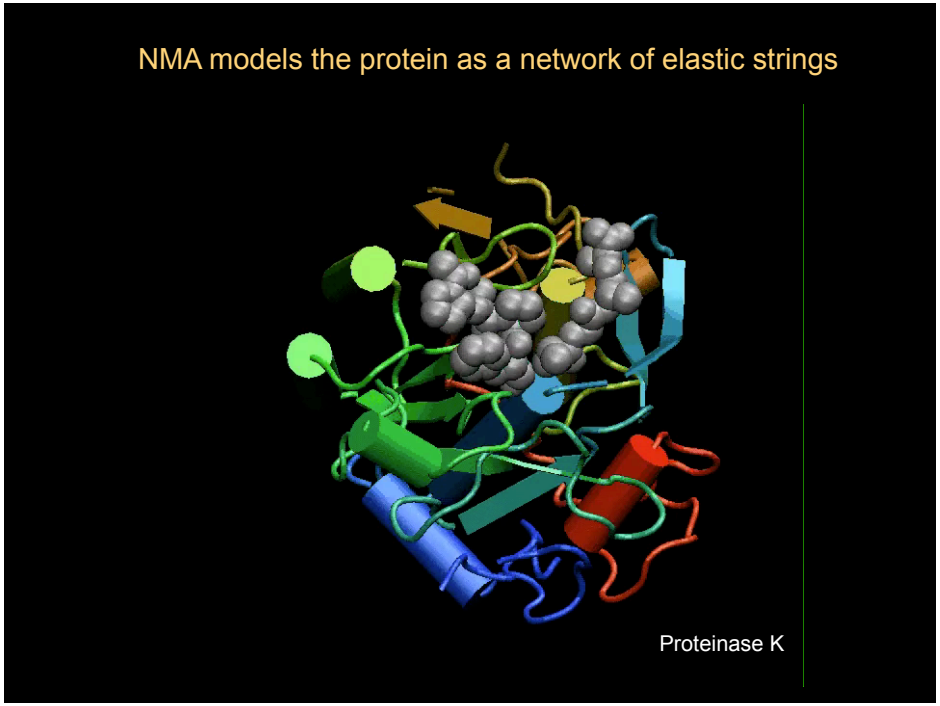
Atomistic

C. G.




- 1 bead / 1 amino acid
- Connected by springs

Coarse Grained

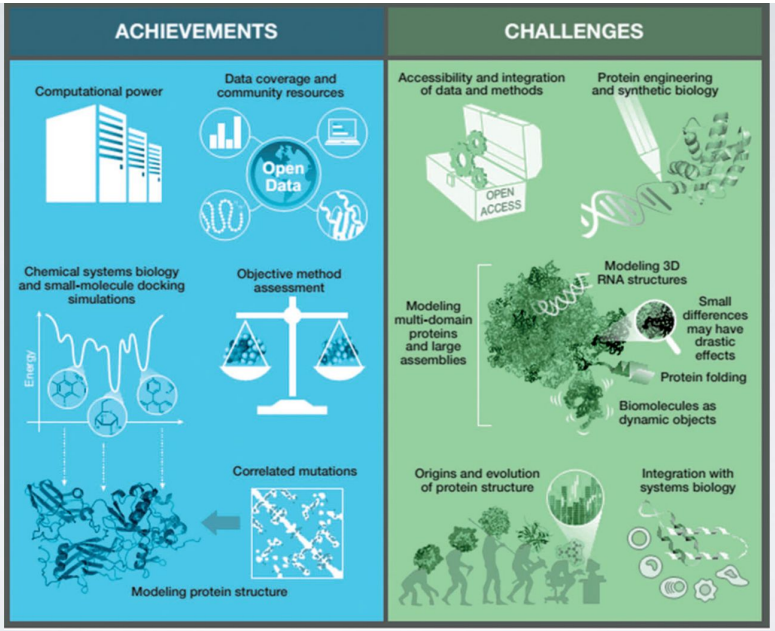


Do it Yourself!

# Hand-on time!

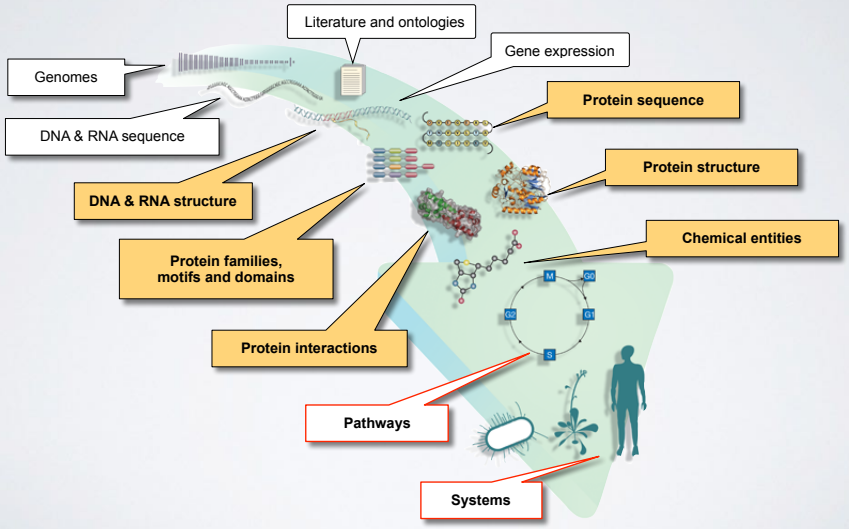
https://bioboot.github.io/bimm143_S18/lectures/#12

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




Ilan Samish et al. Bioinformatics 2015;31:146-150

## 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