



BGGN 213

Structural Bioinformatics

Lecture 11

Barry Grant
UC San Diego

<http://thegrantlab.org/bggn213>

“Bioinformatics is the application of computers to the collection, archiving, organization, and analysis of biological data.”

... A hybrid of biology and computer science

“Bioinformatics is the application of computers to the collection, archiving, organization, and analysis of biological data.”

Bioinformatics is computer aided biology!

“Bioinformatics is the application of computers to the collection, archiving, organization, and analysis of biological data.”

Bioinformatics is computer aided biology!

Goal: Data to Knowledge

So what is **structural bioinformatics**?

So what is **structural bioinformatics**?

... computer aided structural biology!

Aims to characterize and interpret biomolecules and their assemblies at the molecular & atomic level

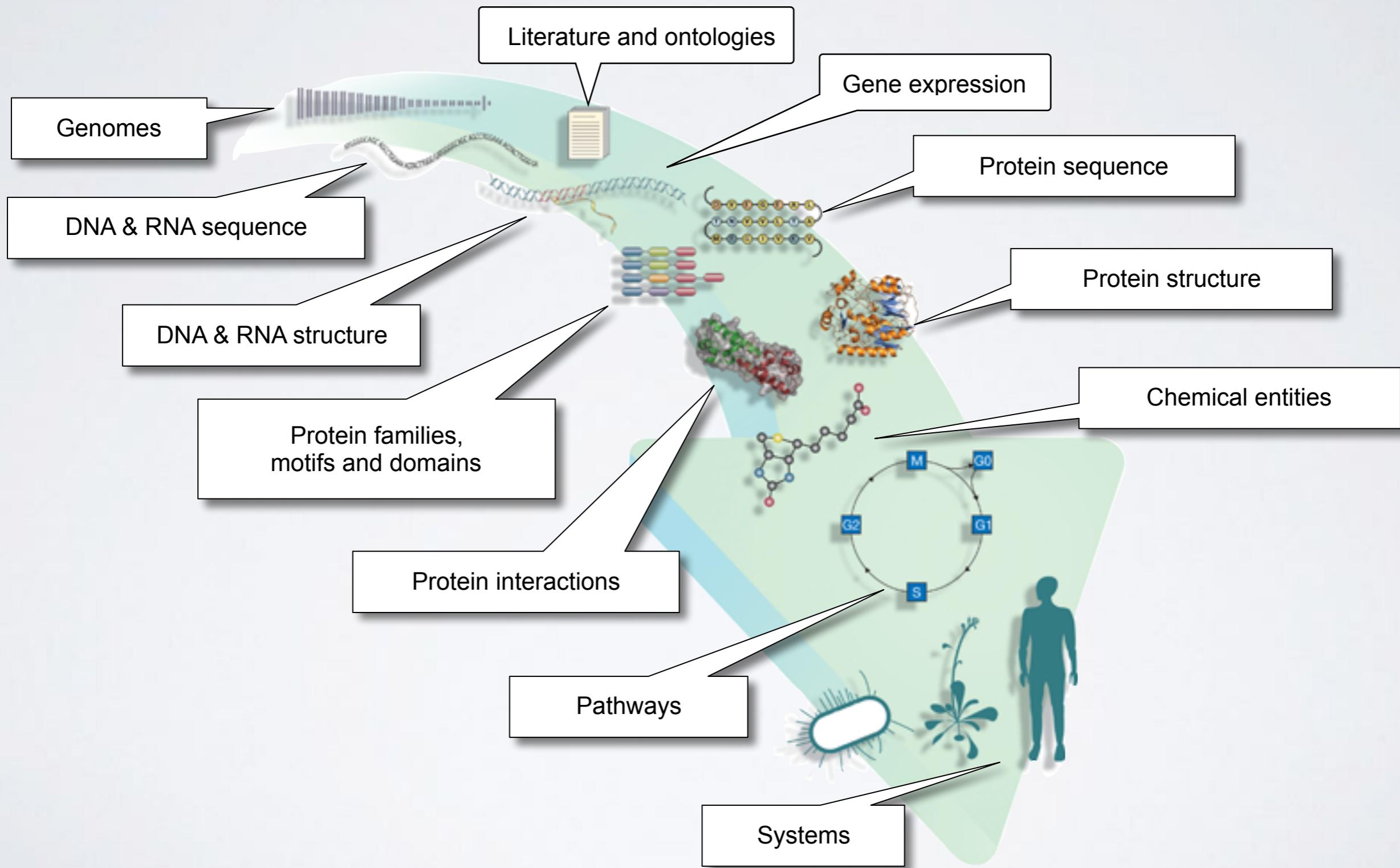
Why should we care?

Why should we care?

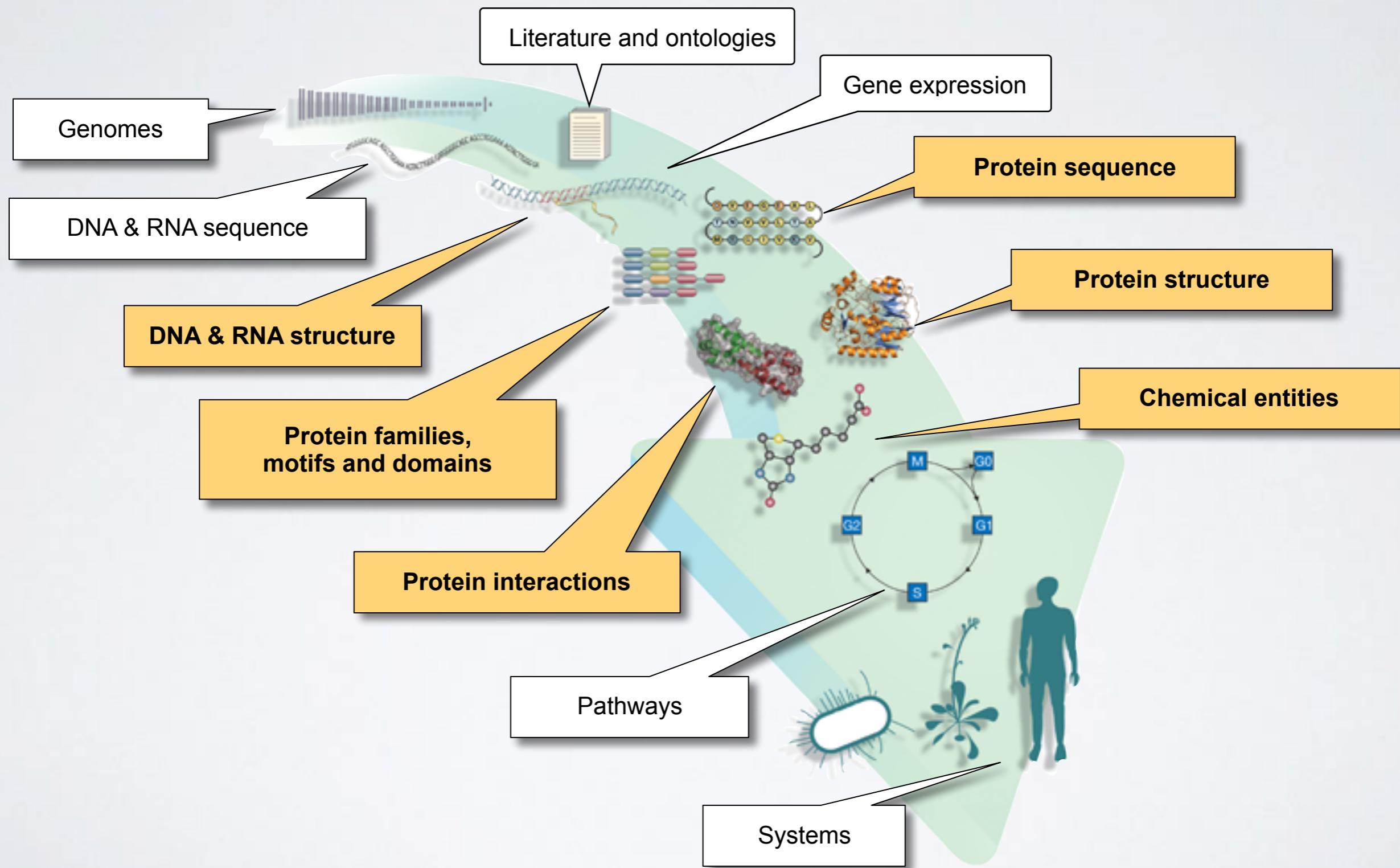
Because biomolecules are “nature’s robots”

... and because it is only by coiling into
specific 3D structures that they are able to
perform their functions

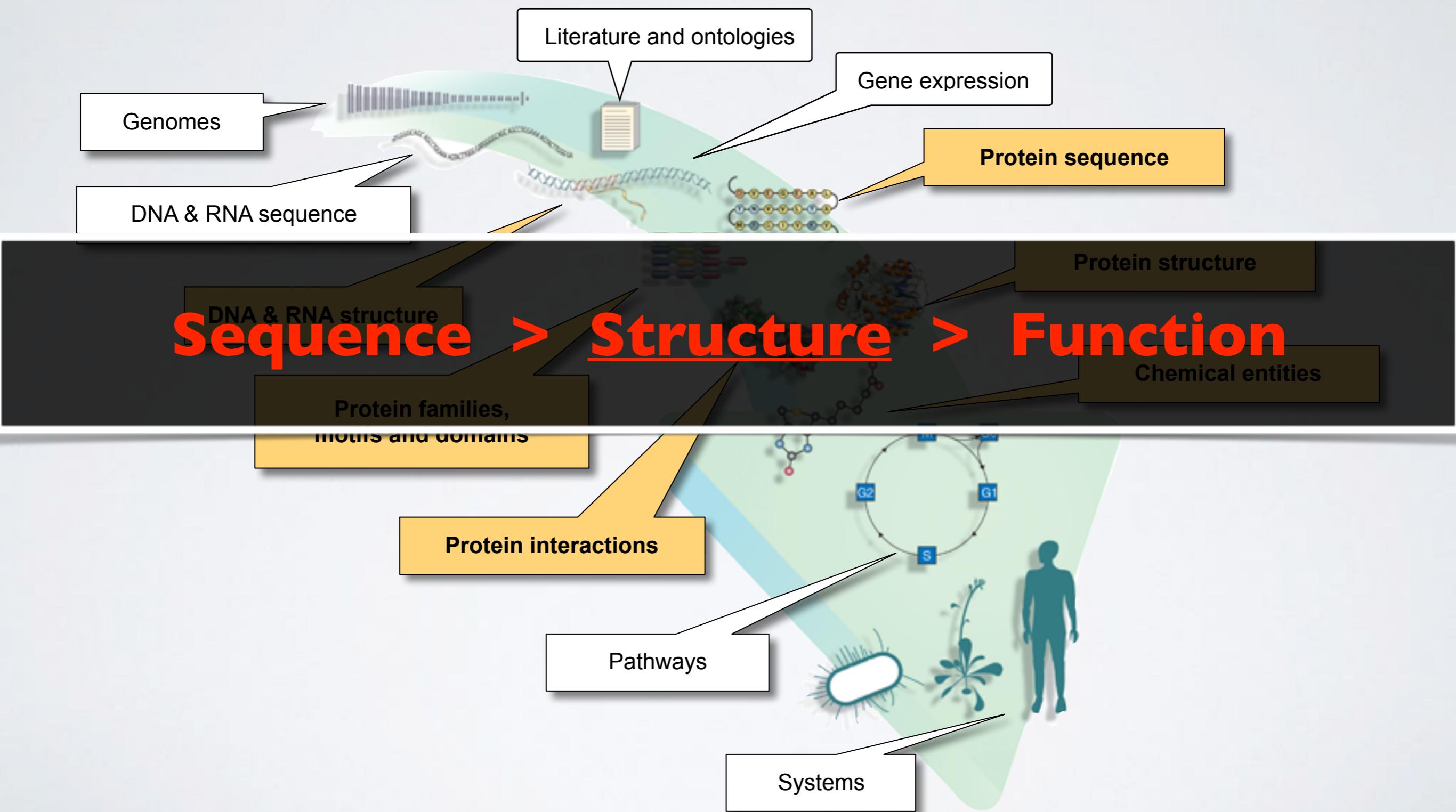
BIOINFORMATICS DATA



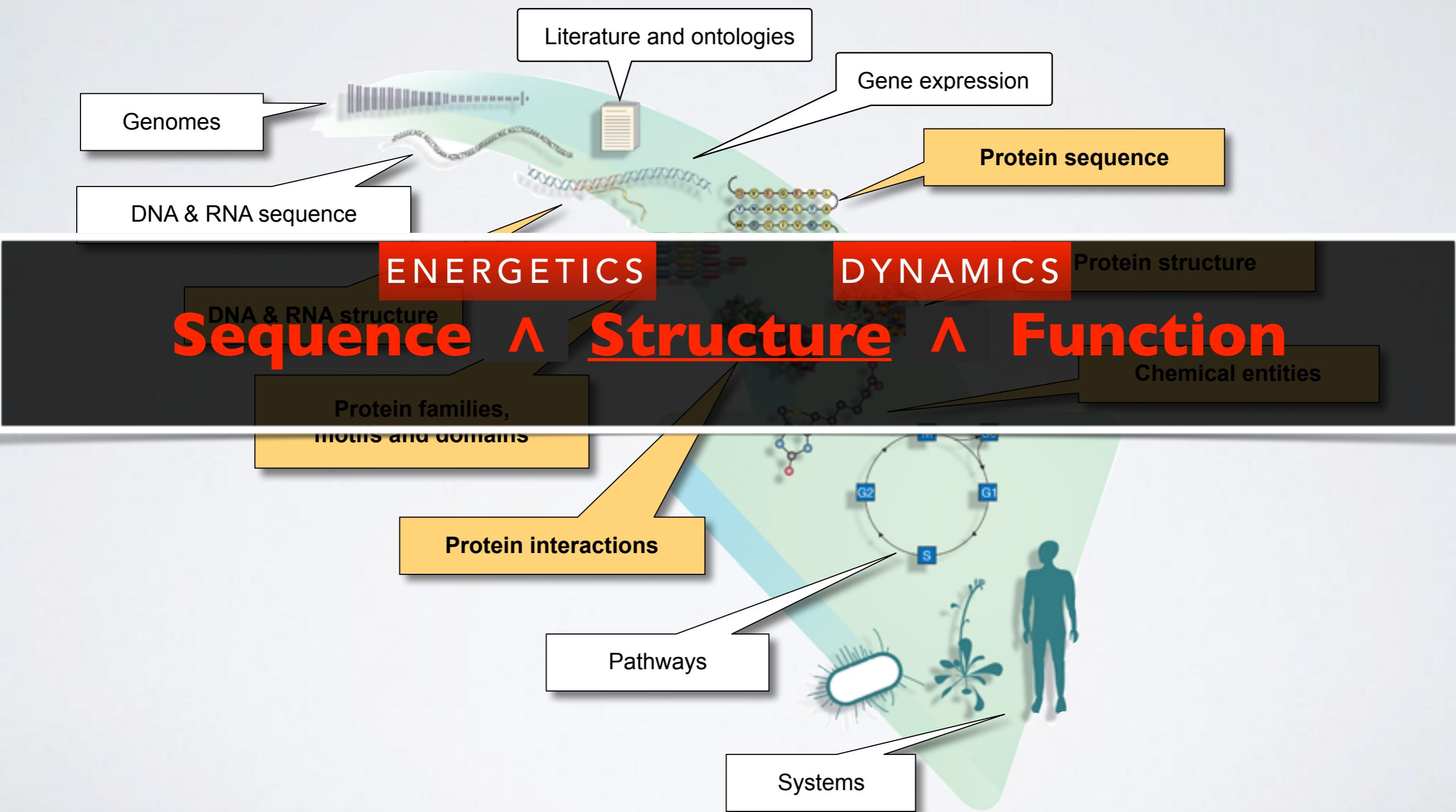
STRUCTURAL DATA IS CENTRAL

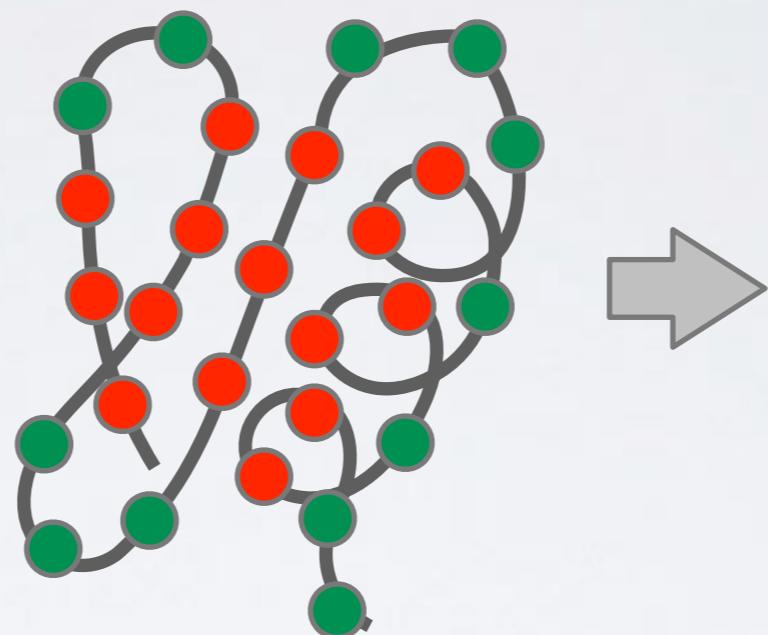
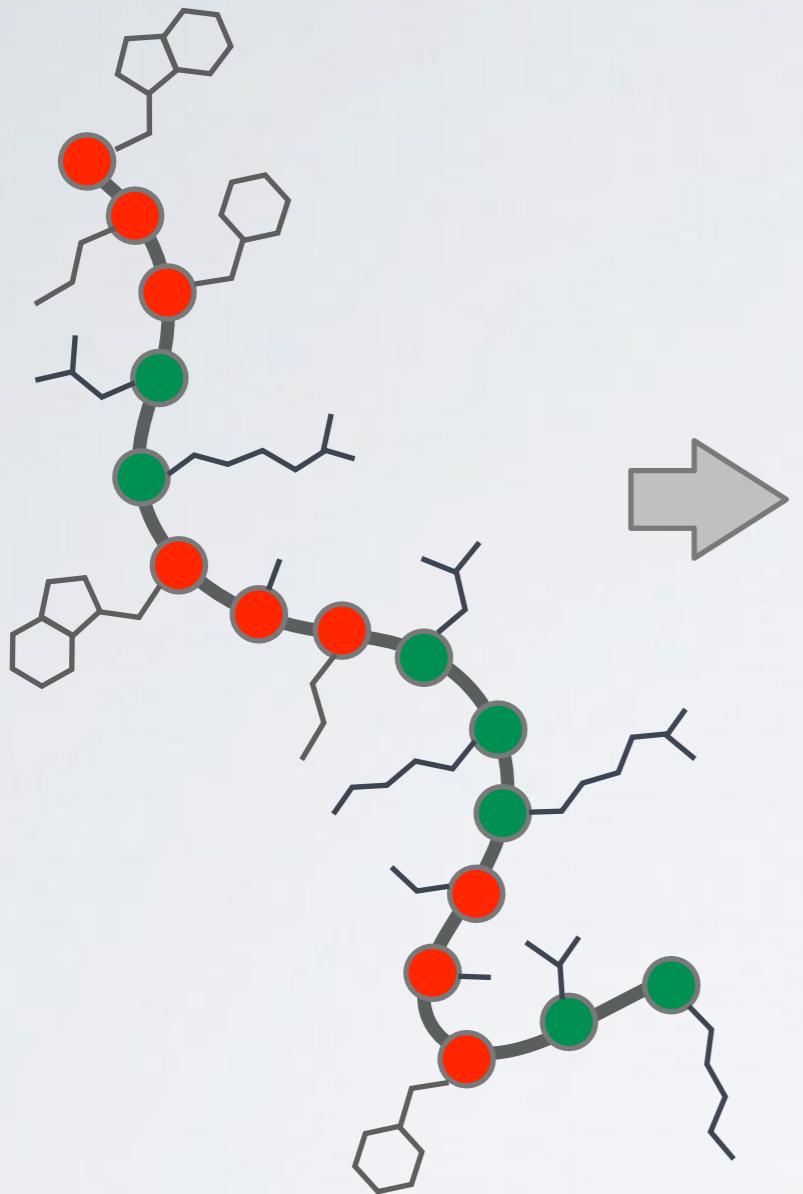


STRUCTURAL DATA IS CENTRAL



STRUCTURAL DATA IS CENTRAL





Sequence

- Unfolded chain of amino acid chain
- Highly mobile
- Inactive

Structure

- Ordered in a precise 3D arrangement
- Stable but dynamic

Function

- Active in specific “conformations”
- Specific associations & precise reactions

In daily life, we use machines
with functional structure and *moving parts*



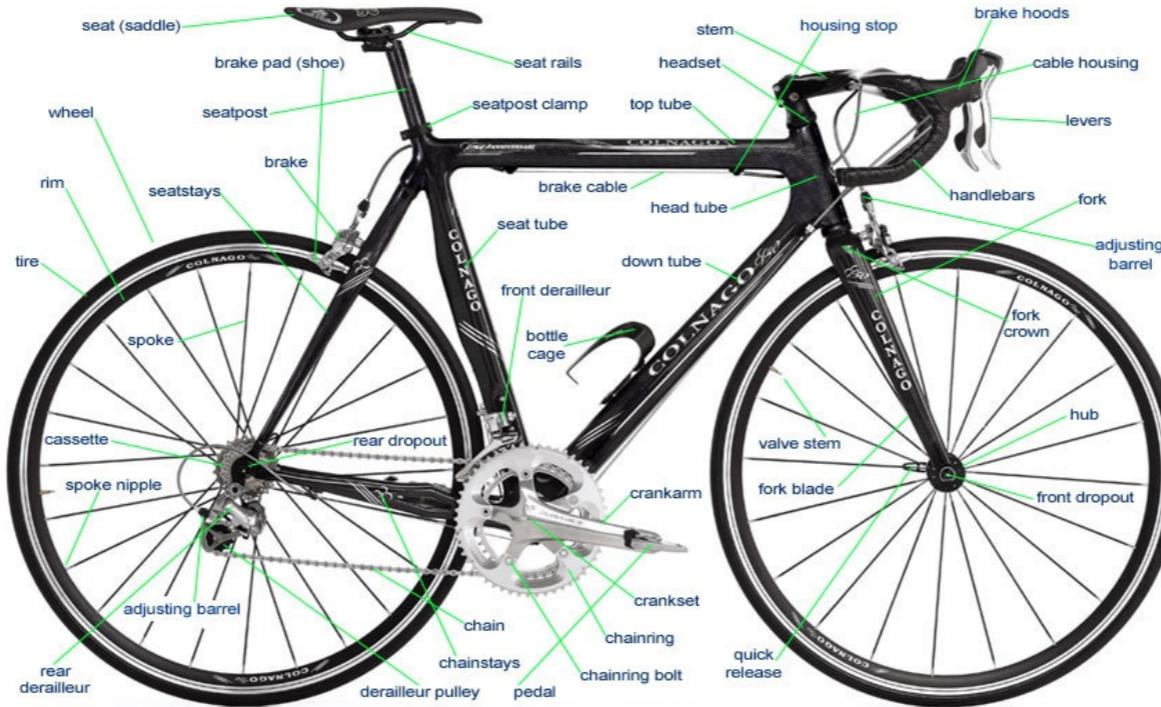
Genomics is a great start

Track Bike – DL 175

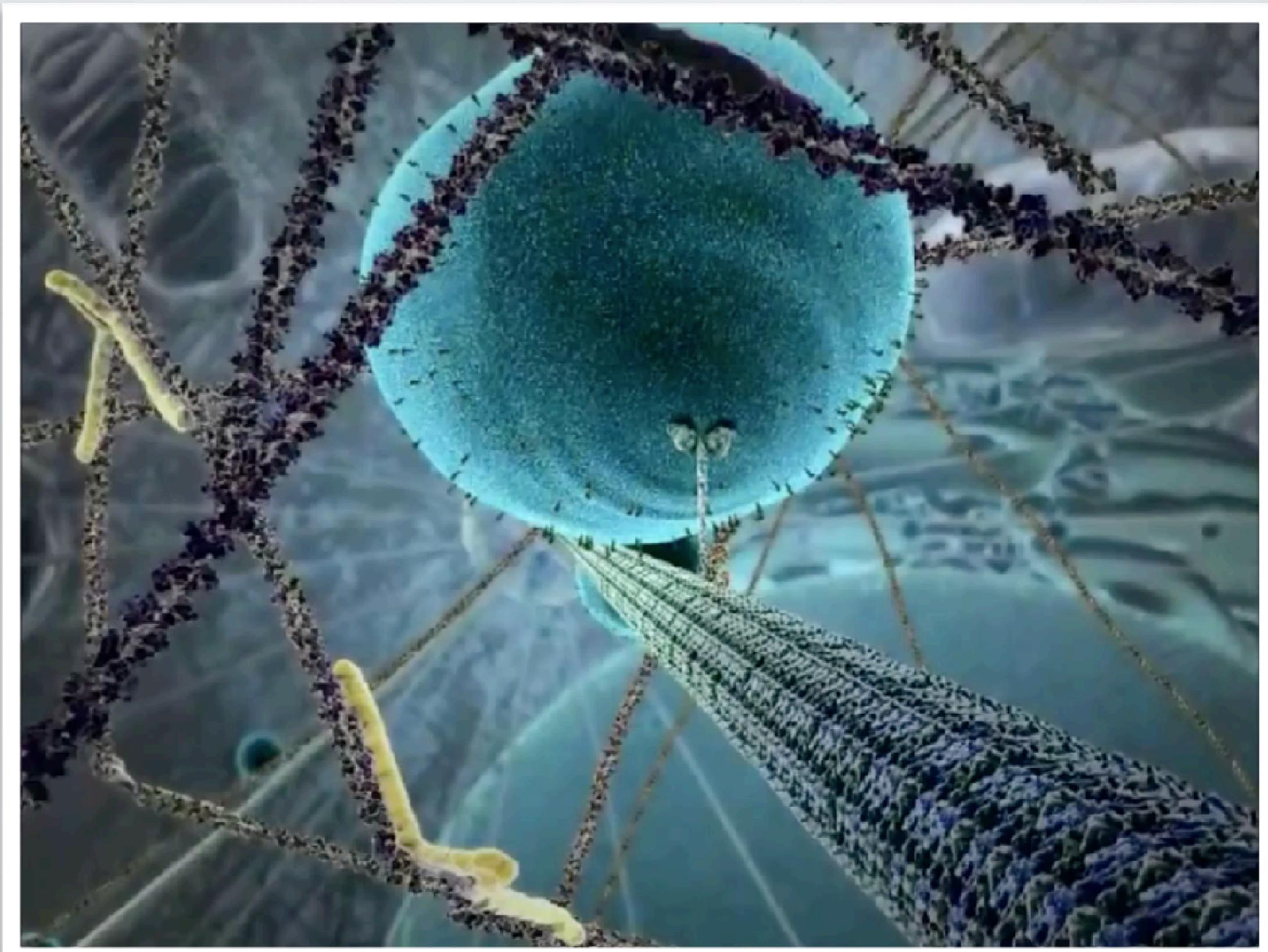
REF. NO.	IBM NO.	DESCRIPTION
1	156011	Track Frame 21", 22", 23", 24", Team Red
2	157040	Fork for 21" Frame
2	157039	Fork for 22" Frame
2	157038	Fork for 23" Frame
2	157037	Fork for 24" Frame
3	191202	Handlebar TTT Competition Track Alloy 15/16"
4		Handlebar Stem, TTT, Specify extension
5	191278	Expander Bolt
6	191272	Clamp Bolt
7	145841	Headset Complete 1 x 24 BSC
8	145842	Ball Bearings
9	190420	175 Raleigh Pistard Seta Tubular Prestavalve 27"
10	190233	Rim, 27" AVA Competition (36H) Alloy Prestavalve
11	145973	Hub, Large Flange Campagnolo Pista Track Alloy (pairs)
12	190014	Spokes, 11 5/8"
13	145837	Sleeve
14	145636	Ball Bearings
15	145170	Bottom Bracket Axle
16	145838	Cone for Sleeve
17	146473	L.H. Adjustable Cup
18	145833	Lockring
19	145239	Straps for Toe Clips
20	145834	Fixing Bolt
21	145835	Fixing Washer
22	145822	Dustcap
23	145823	R.H. and L.H. Crankset with Chainwheel
24	146472	Fixed Cup
25	145235	Toe Clips, Christophe, Chrome (Medium)
26	145684	Pedals, Extra Light, Pairs
27	123021	Chain
28	145980	Seat Post
29		Seat Post Bolt and Nut
30	167002	Saddle, Brooks
31	145933	Track Sprocket, Specify 12, 13, 14, 15, or 16 T.

- But a parts list is not enough to understand how a bicycle works

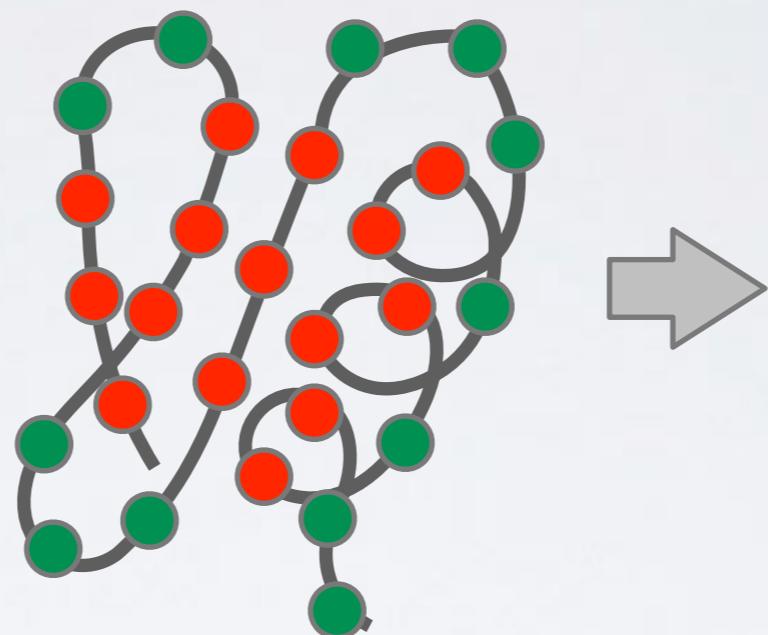
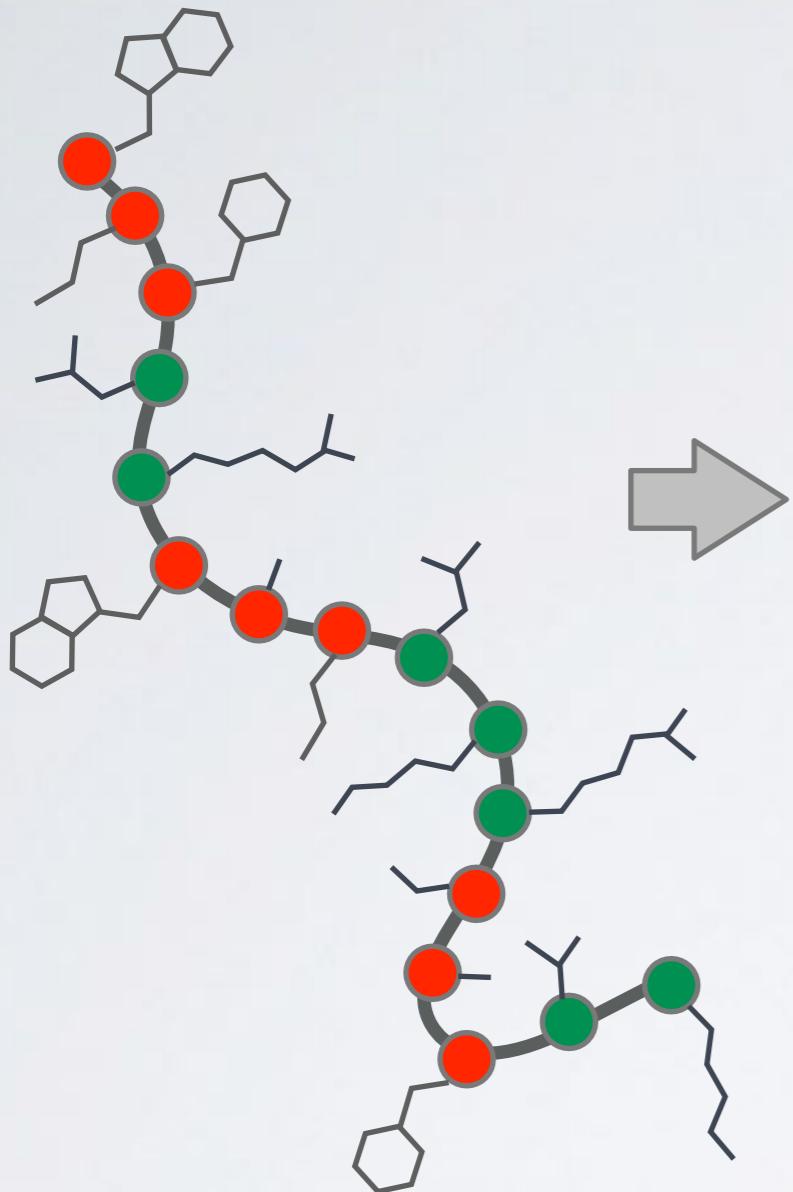
... but not the end



- We want the full spatiotemporal picture, and an ability to control it
- Broad applications, including drug design, medical diagnostics, chemical manufacturing, and energy



Extracted from The Inner Life of a Cell by Cellular Visions and Harvard
[YouTube link: <https://www.youtube.com/watch?v=y-uuk4Pr2i8>]



Sequence

- Unfolded chain of amino acid chain
- Highly mobile
- Inactive

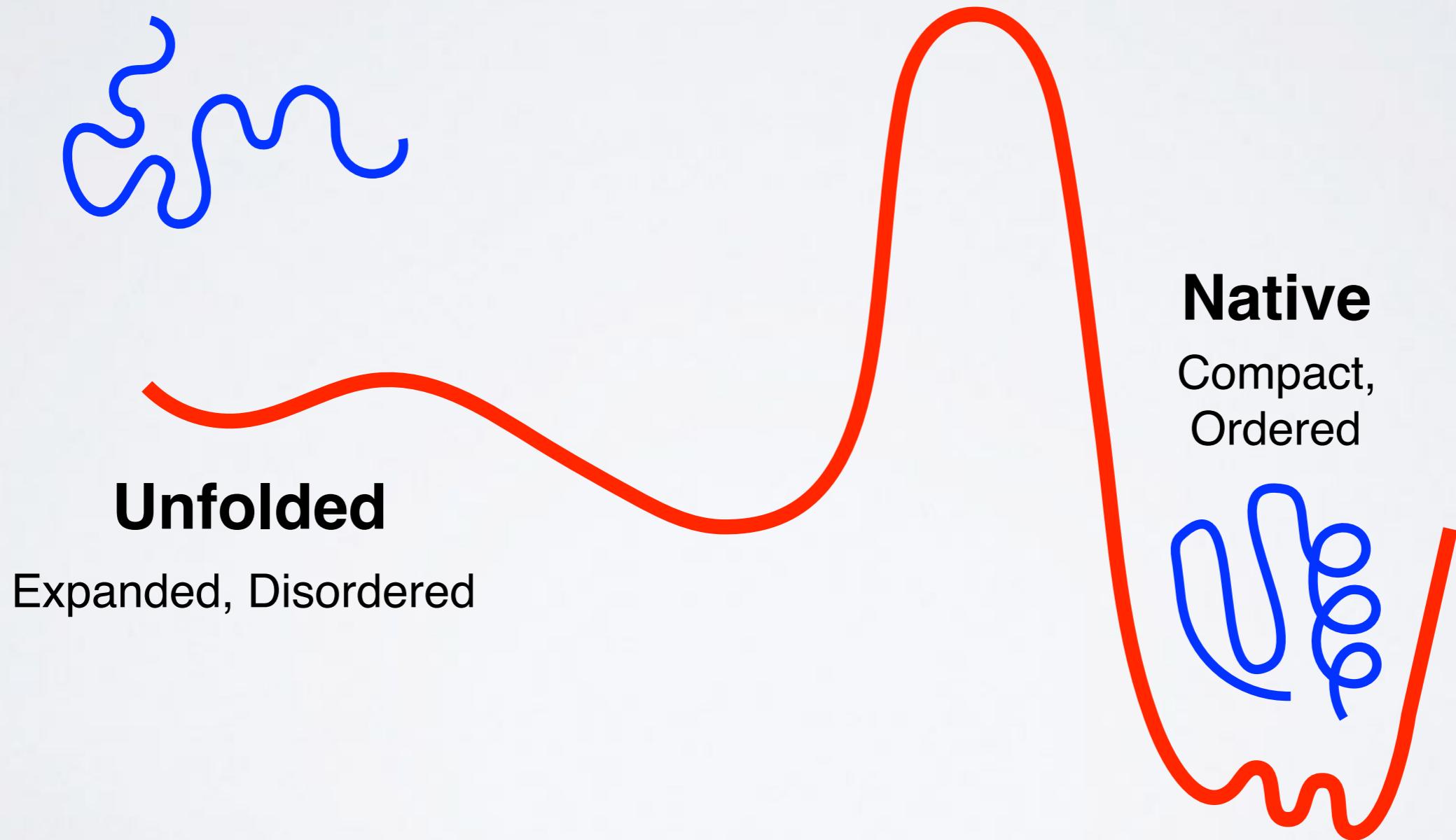
Structure

- Ordered in a precise 3D arrangement
- Stable but dynamic

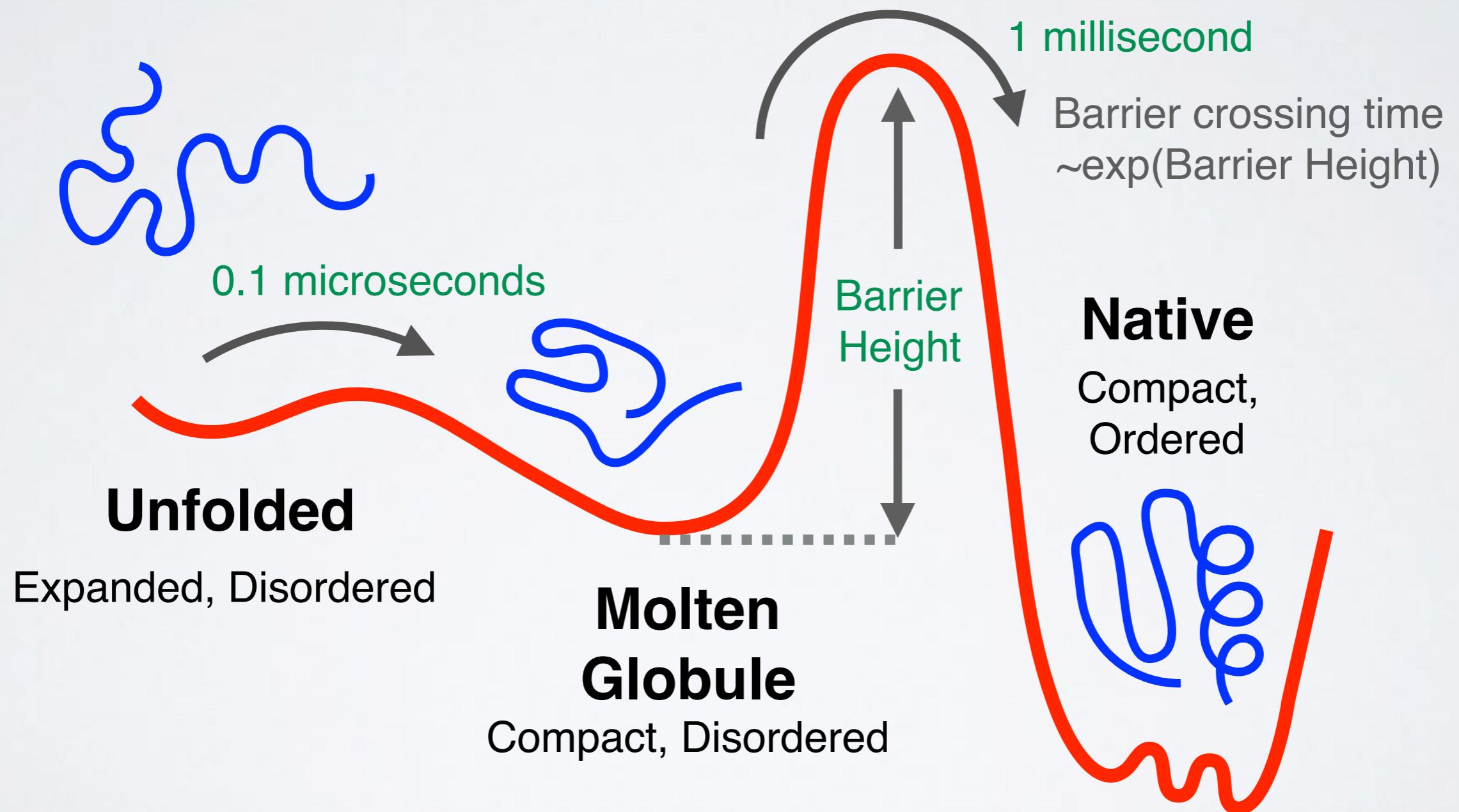
Function

- Active in specific “conformations”
- Specific associations & precise reactions

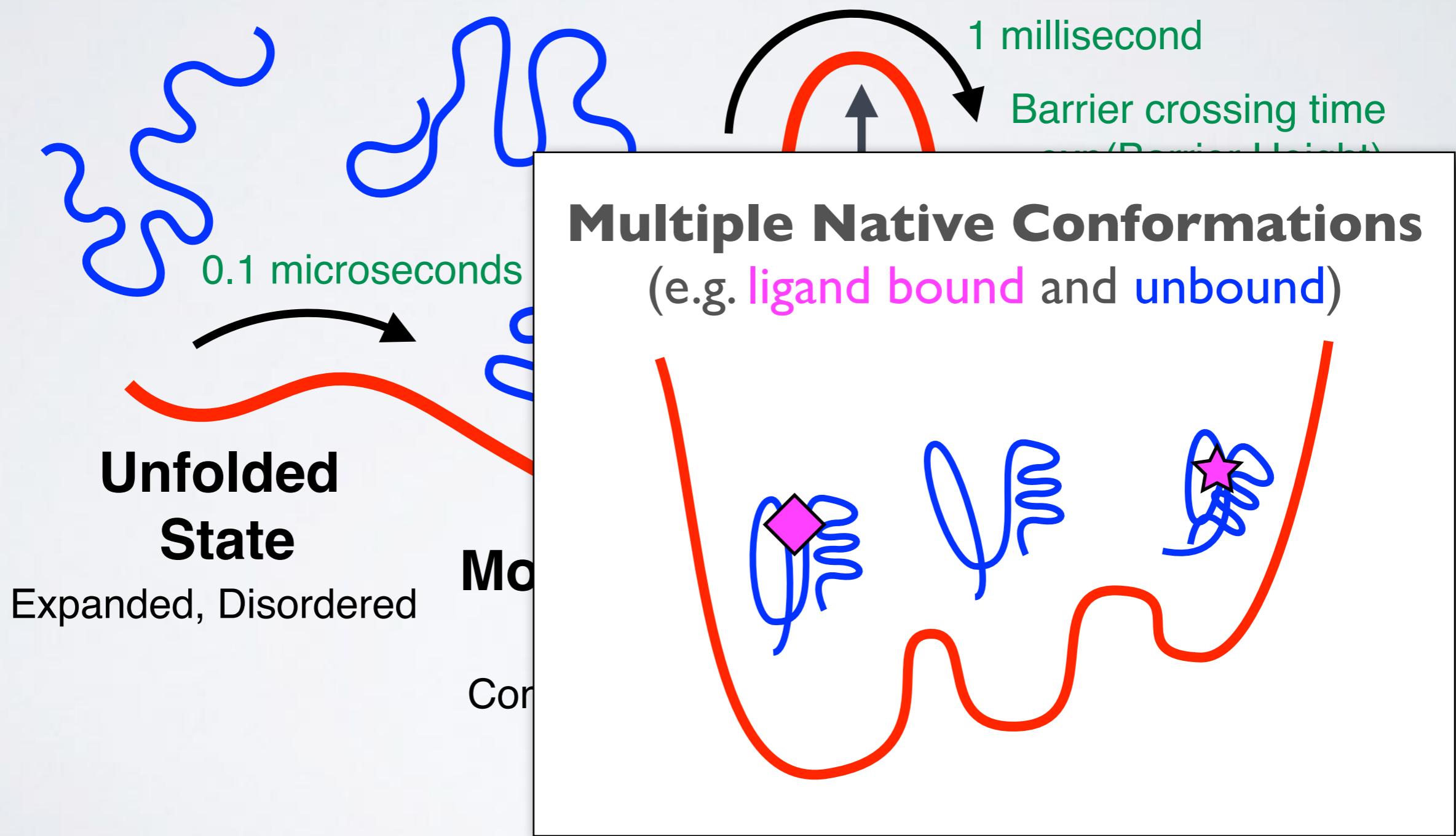
KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: ENERGY LANDSCAPE



OUTLINE:

- ▶ **Overview of structural bioinformatics**
 - Major motivations, goals and challenges
- ▶ **Fundamentals of protein structure**
 - Composition, form, forces and dynamics
- ▶ **Representing and interpreting protein structure**
 - Modeling energy as a function of structure
- ▶ **Example application areas**
 - Predicting functional dynamics & drug discovery

OUTLINE:

- ▶ **Overview of structural bioinformatics**

- Major motivations, goals and challenges

- ▶ **Fundamentals of protein structure**

- Composition, form, forces and dynamics

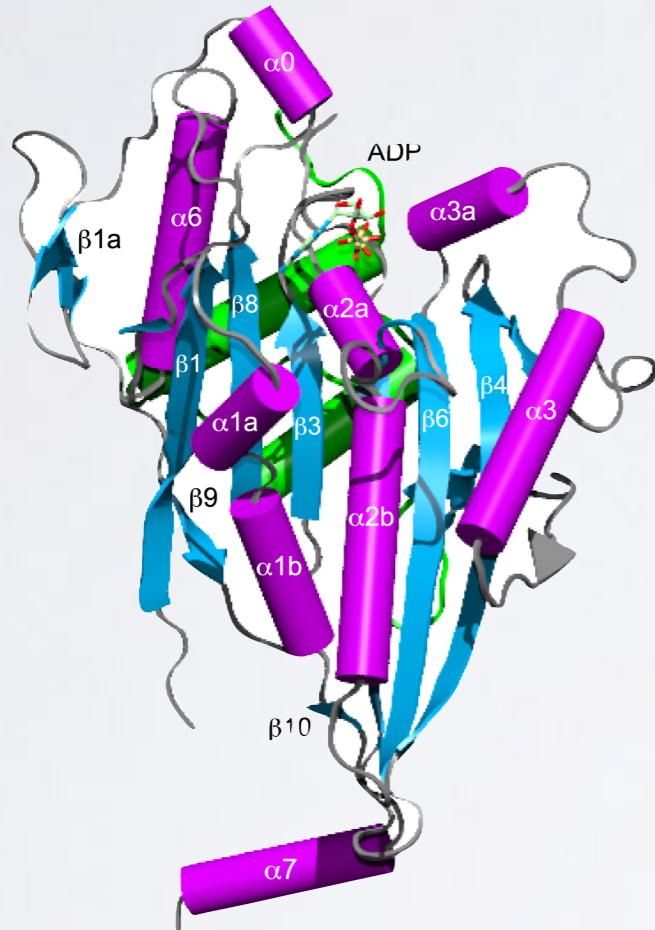
- ▶ **Representing and interpreting protein structure**

- Modeling energy as a function of structure

- ▶ **Example application areas**

- Predicting functional dynamics & drug discovery

TRADITIONAL FOCUS **PROTEIN**, **DNA** AND **SMALL MOLECULE** DATA SETS WITH **MOLECULAR STRUCTURE**



Protein
(PDB)

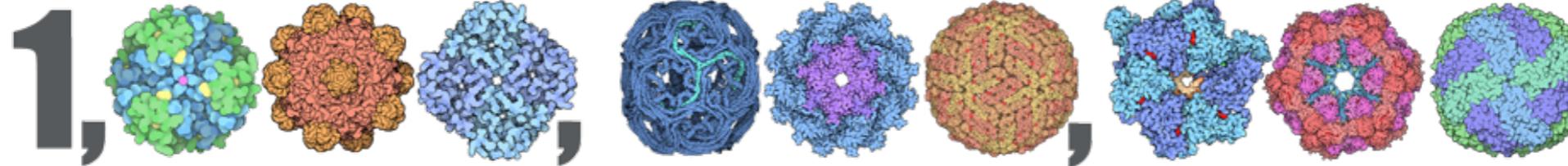


DNA
(NDB)

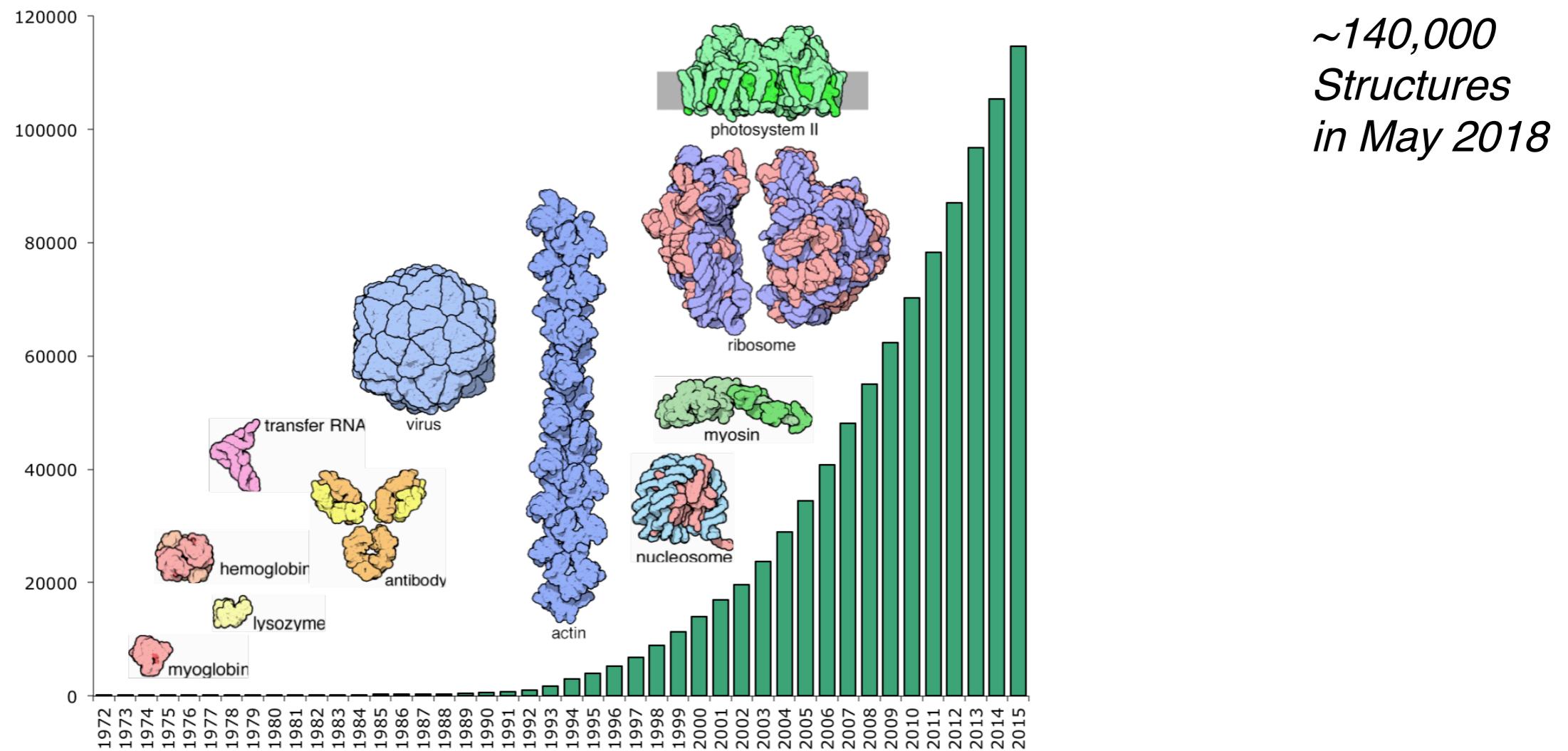


Small Molecules
(CCDB)

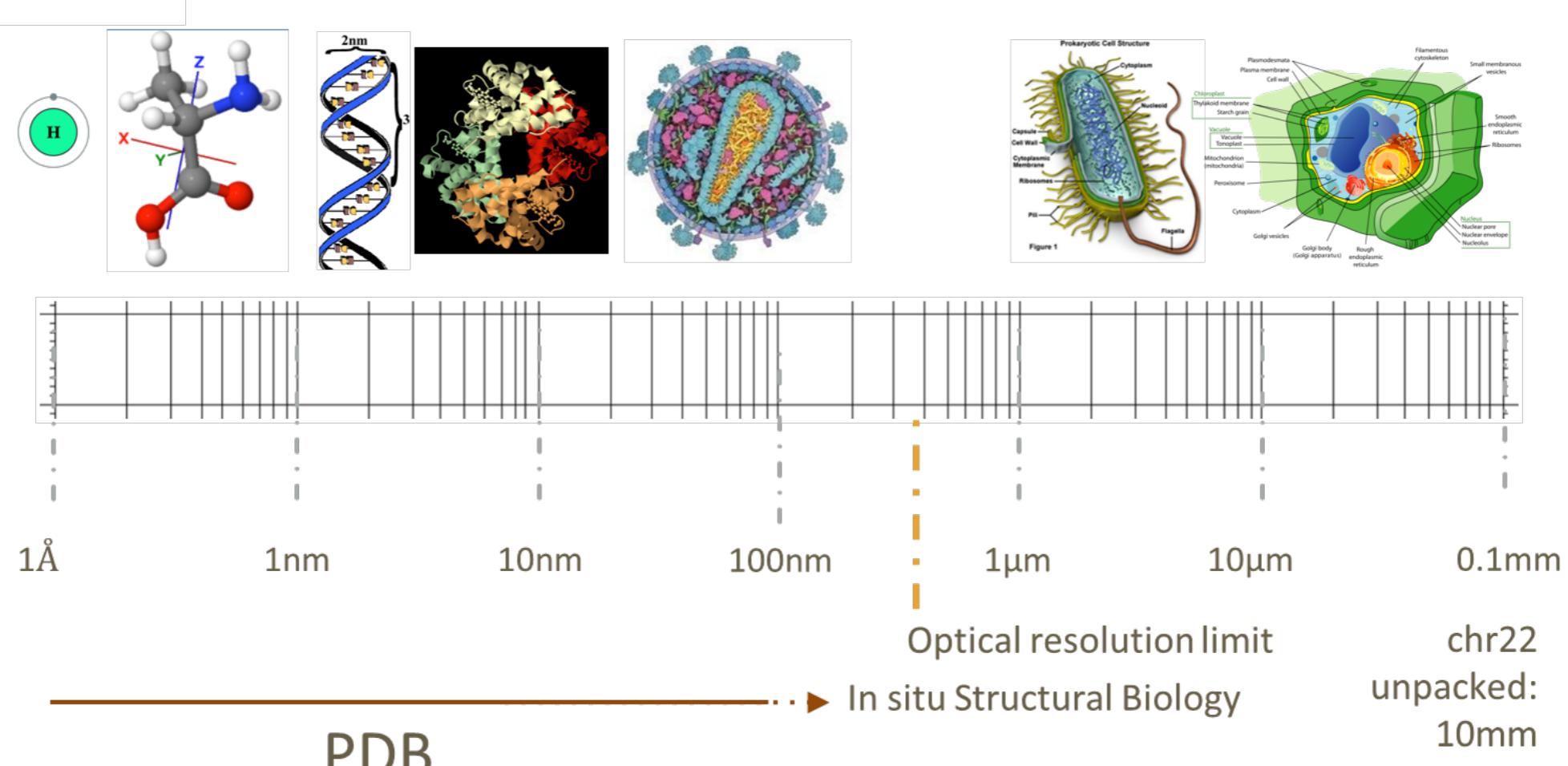
PDB – A Billion Atom Archive



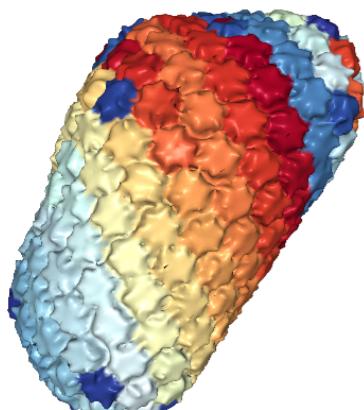
> 1 billion atoms in the asymmetric units



Growing Structure Size and Complexity

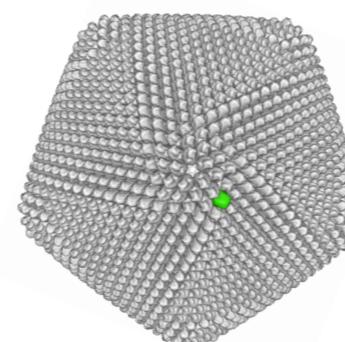


Largest asymmetric structure in PDB



HIV-1 capsid: PDB ID 3J3Q
~2.4M unique atoms

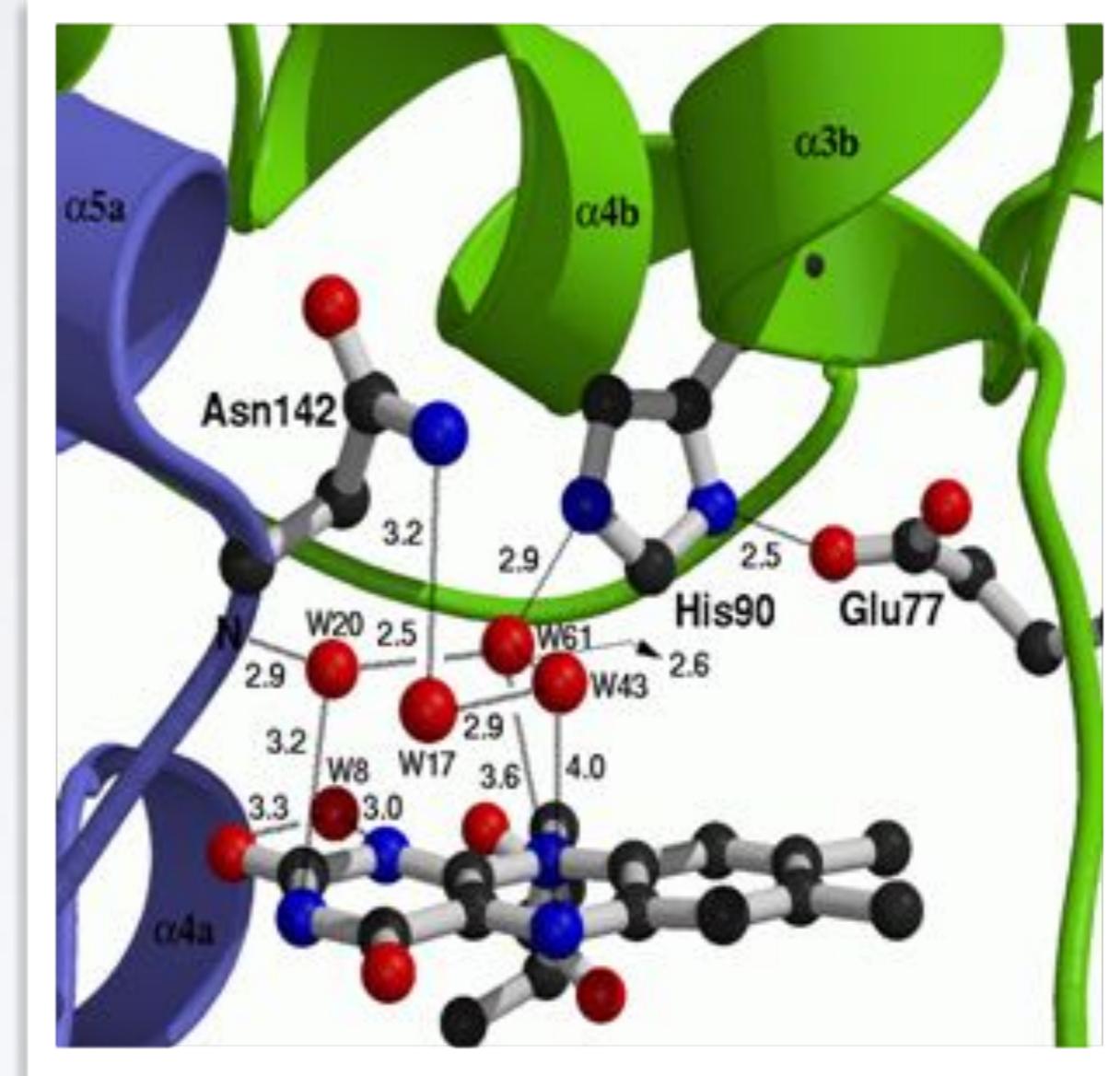
Largest symmetric structure in PDB



Faustovirus major capsid: PDB ID 5J7V
~40M overall atoms

Motivation 1: Detailed understanding of molecular interactions

Provides an invaluable structural context for conservation and mechanistic analysis leading to functional insight.



Motivation 1: Detailed understanding of molecular interactions

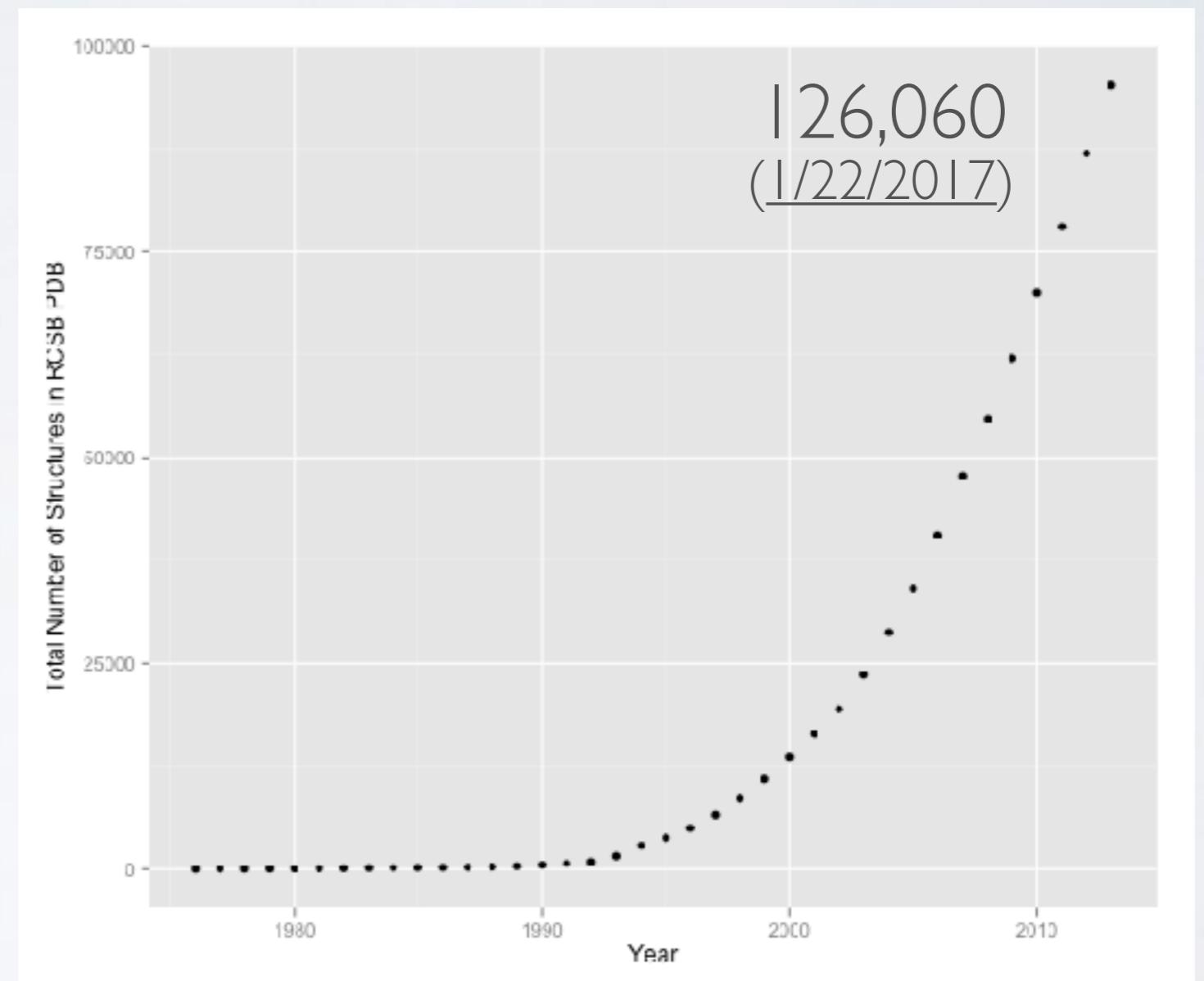
Computational modeling can provide detailed insight into functional interactions, their regulation and potential consequences of perturbation.



Grant et al. PLoS. Comp. Biol. (2010)

Motivation 2: Lots of structural data is becoming available

Structural Genomics has
contributed to driving
down the cost and time
required for structural
determination



Data from: <http://www.rcsb.org/pdb/statistics/>

Motivation 2: Lots of structural data is becoming available

Structural Genomics has
contributed to driving
down the cost and time
required for structural
determination

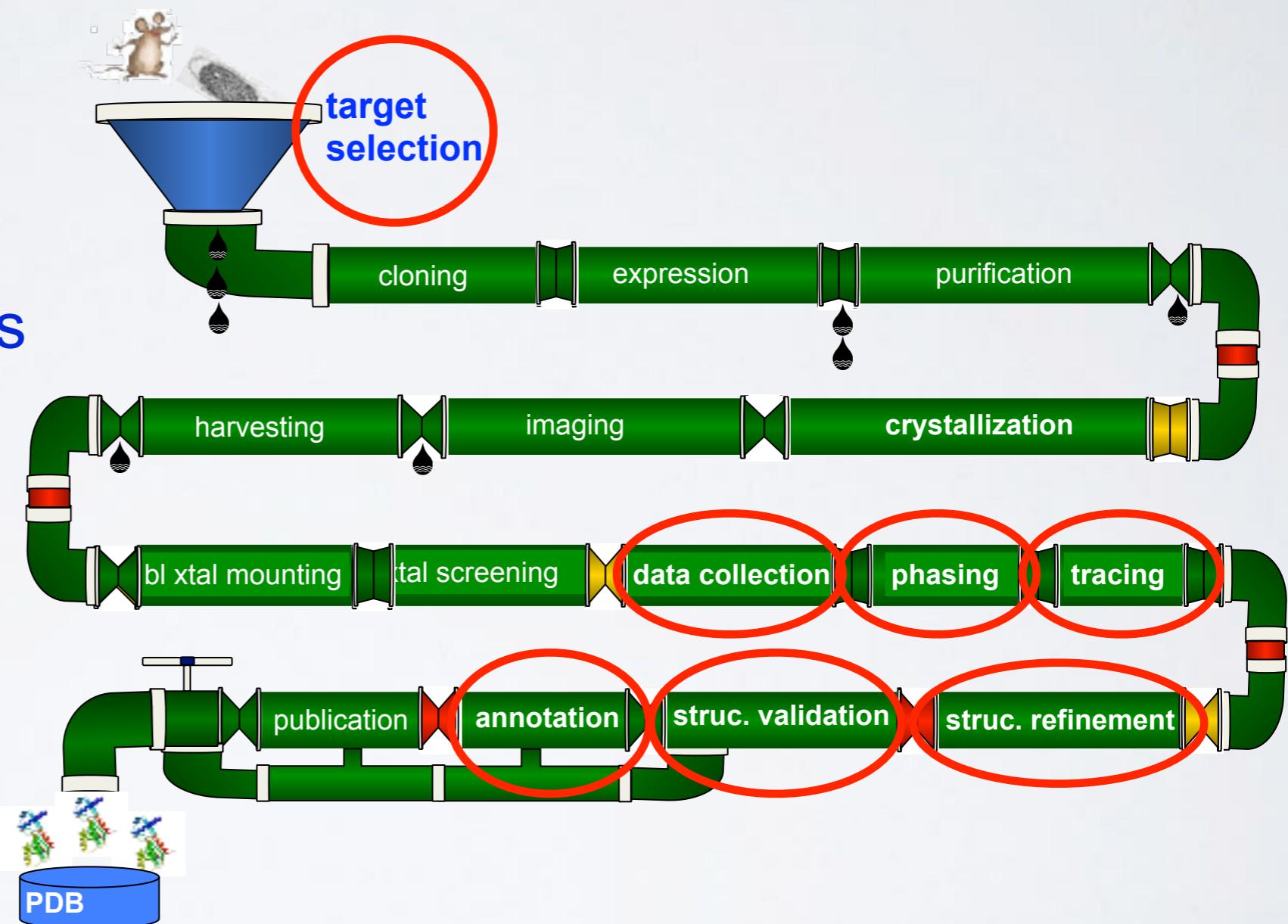
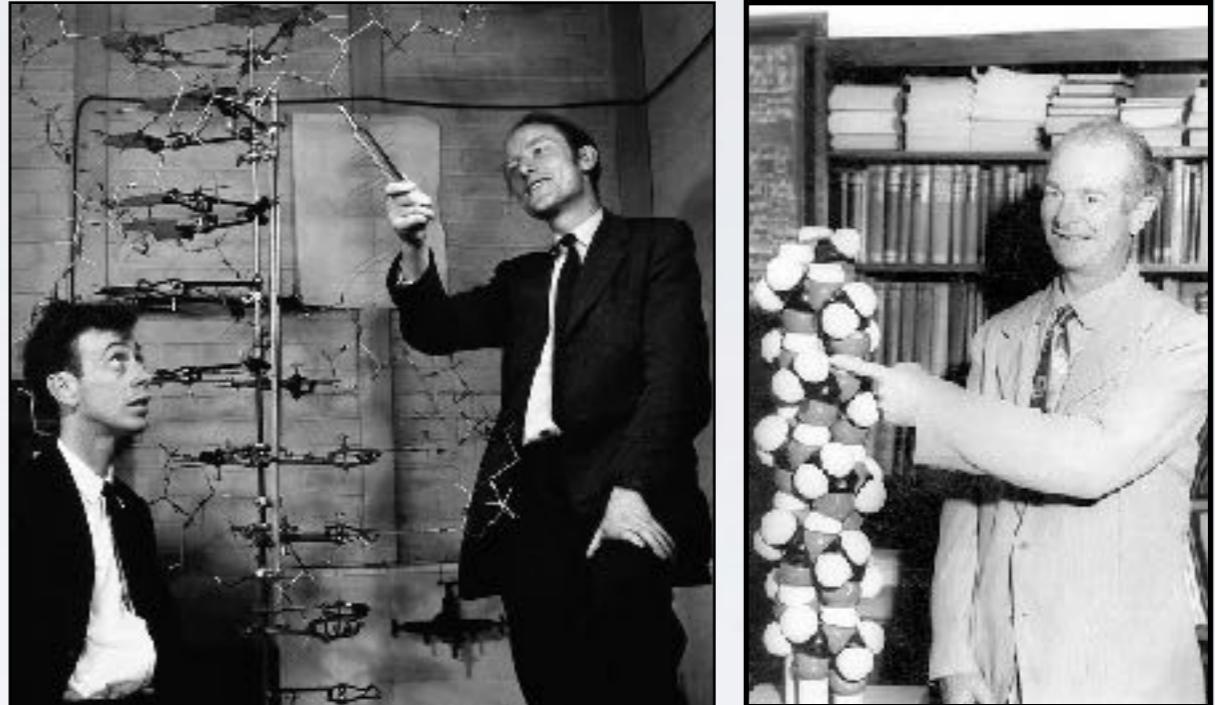


Image Credit: "Structure determination assembly line" Adam Godzik

Motivation 3:
Theoretical and
computational predictions
have been, and continue
to be, enormously
valuable and influential!



SUMMARY OF KEY **MOTIVATIONS**

Sequence > Structure > Function

- Structure determines function, so understanding structure helps our understanding of function

Structure is more conserved than sequence

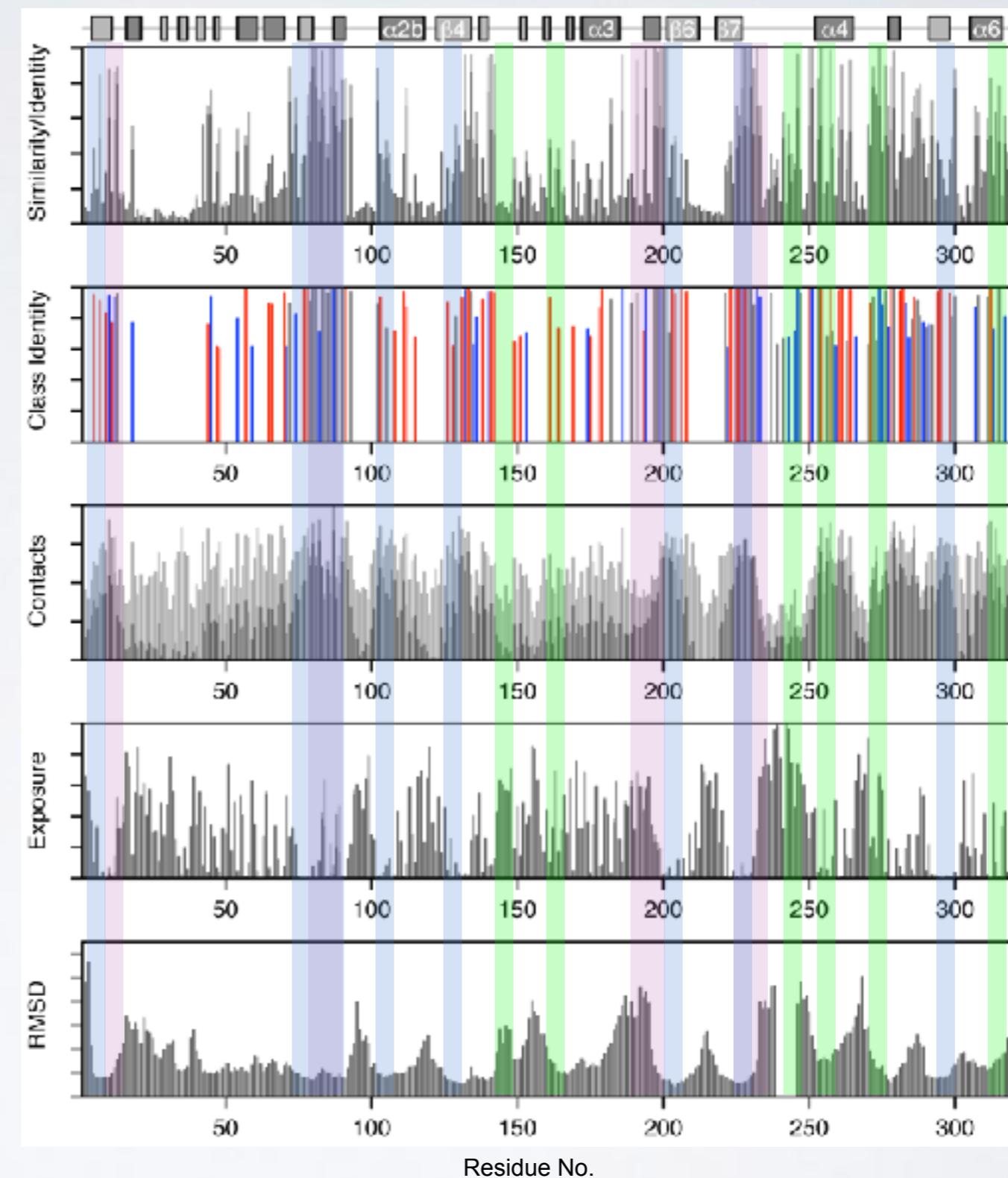
- Structure allows identification of more distant evolutionary relationships

Structure is encoded in sequence

- Understanding the determinants of structure allows design and manipulation of proteins for industrial and medical advantage

Goals:

- Analysis
- Visualization
- Comparison
- Prediction
- Design



Goals:

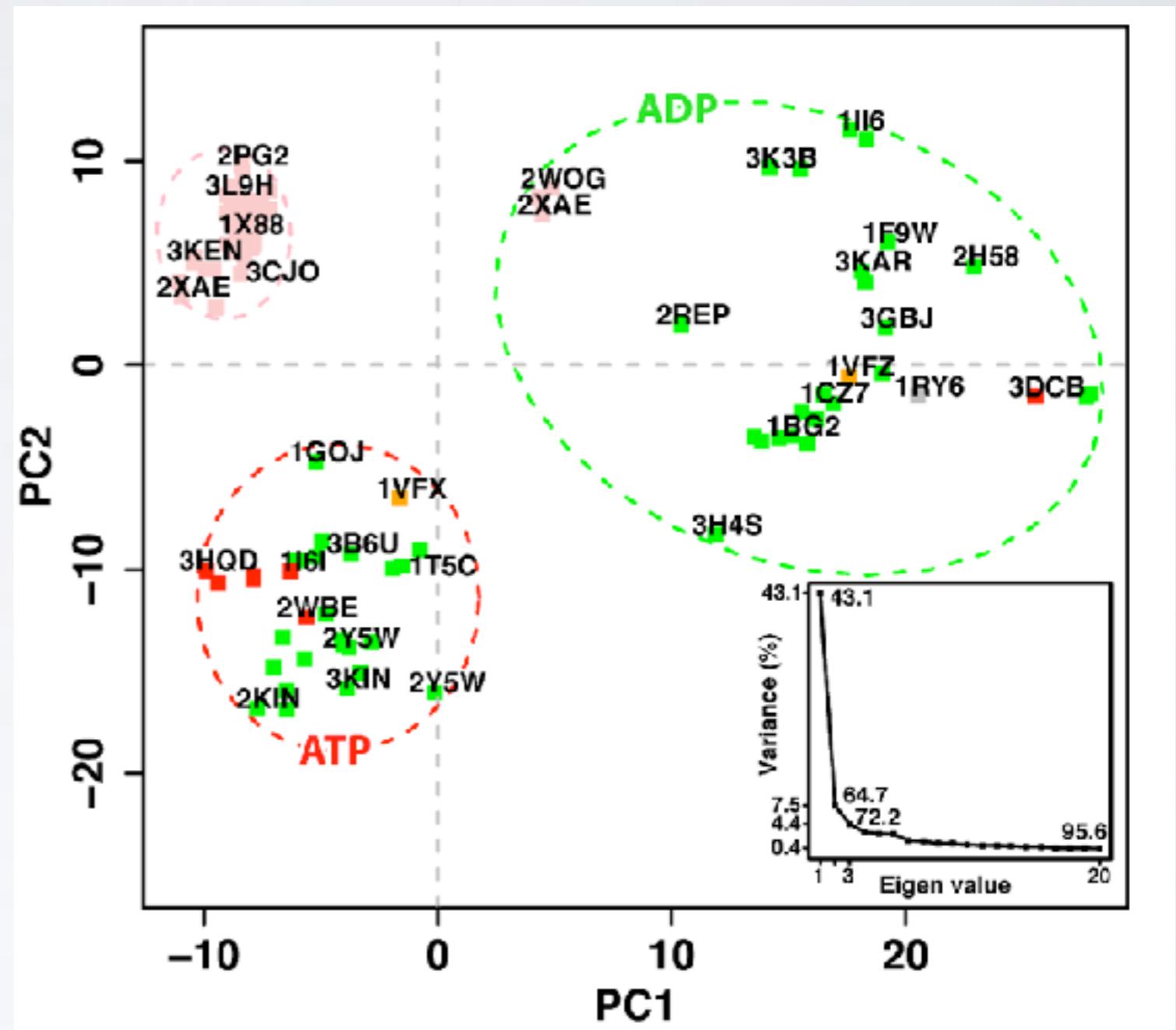
- Analysis
- Visualization
- Comparison
- Prediction
- Design



Scarabelli and Grant. PLoS. Comp. Biol. (2013)

Goals:

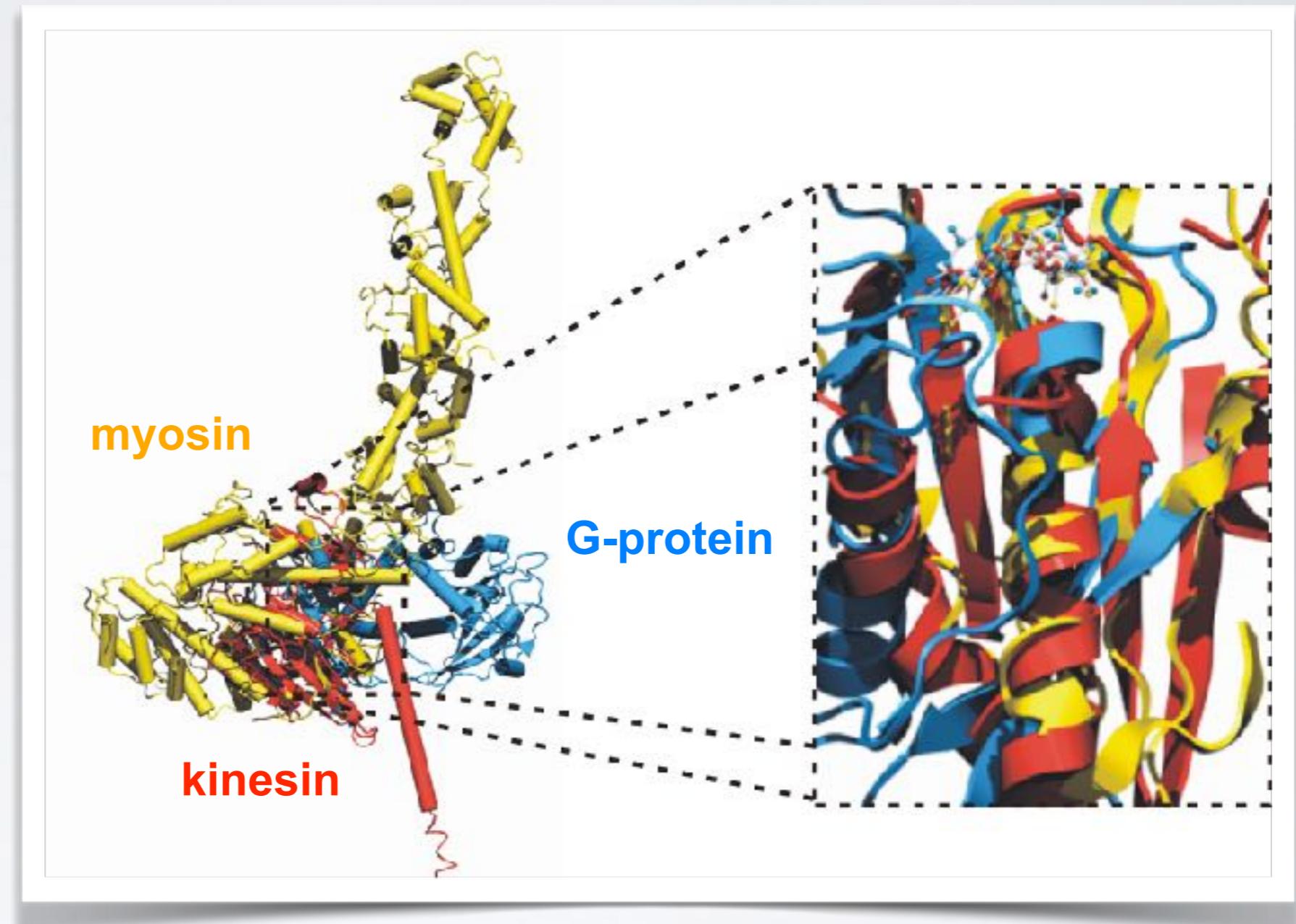
- Analysis
- Visualization
- Comparison
- Prediction
- Design



Scarabelli and Grant. PLoS. Comp. Biol. (2013)

Goals:

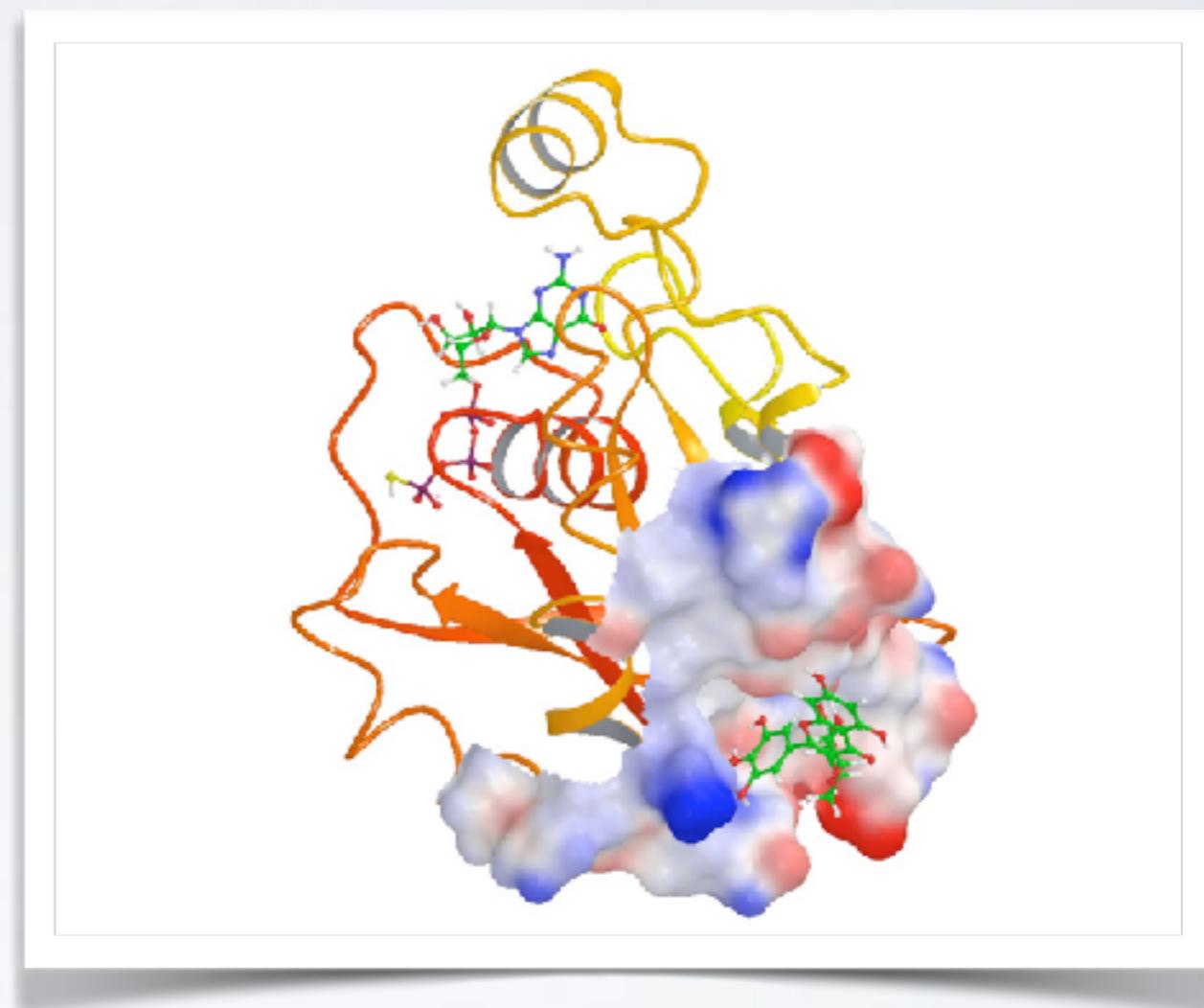
- Analysis
- Visualization
- Comparison
- Prediction
- Design



Grant et al. unpublished

Goals:

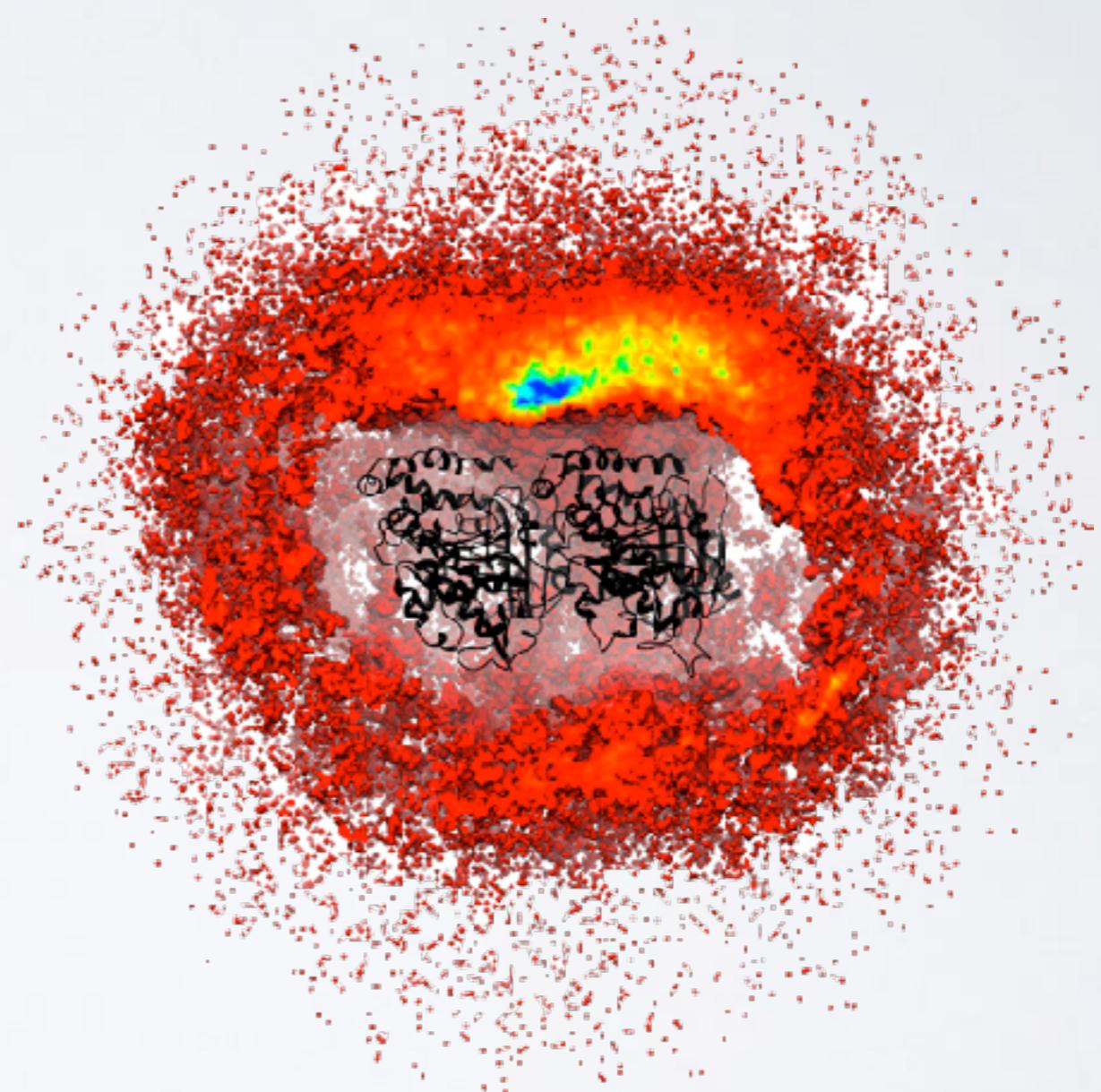
- Analysis
- Visualization
- Comparison
- Prediction
- Design



Grant et al. PLoS One (2011, 2012)

Goals:

- Analysis
- Visualization
- Comparison
- Prediction
- Design



Grant et al. PLoS Biology (2011)

MAJOR RESEARCH AREAS AND CHALLENGES

Include but are not limited to:

- Protein classification
- Structure prediction from sequence
- Binding site detection
- Binding prediction and drug design
- Modeling molecular motions
- Predicting physical properties (stability, binding affinities)
- Design of structure and function
- etc...

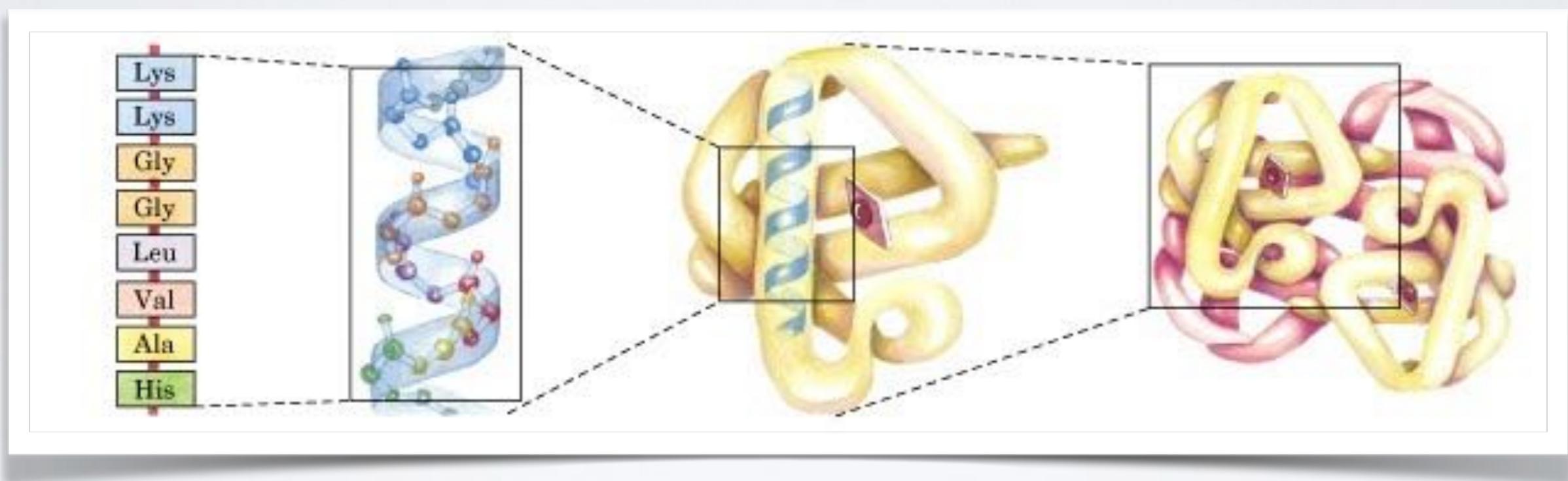
With applications to Biology, Medicine, Agriculture and Industry

NEXT UP:

- ▶ **Overview of structural bioinformatics**
 - Major motivations, goals and challenges
- ▶ **Fundamentals of protein structure**
 - Composition, form, forces and dynamics
- ▶ **Representing and interpreting protein structure**
 - Modeling energy as a function of structure
- ▶ **Example application areas**
 - Predicting functional dynamics & drug discovery

HIERARCHICAL STRUCTURE OF PROTEINS

Primary > Secondary > Tertiary > Quaternary



amino acid
residues

Alpha
helix

Polypeptide
chain

Assembled
subunits

RECAP: AMINO ACID NOMENCLATURE

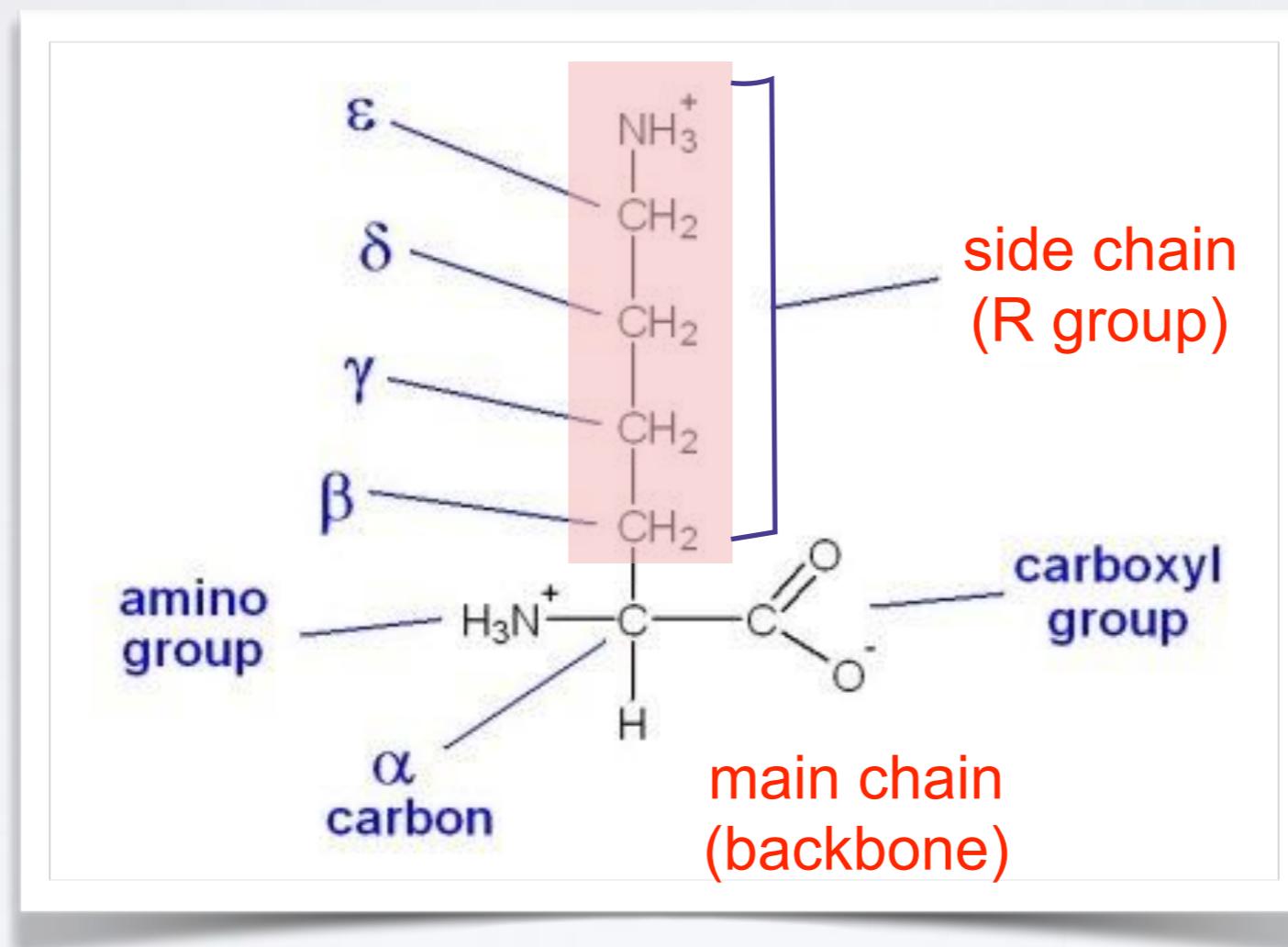
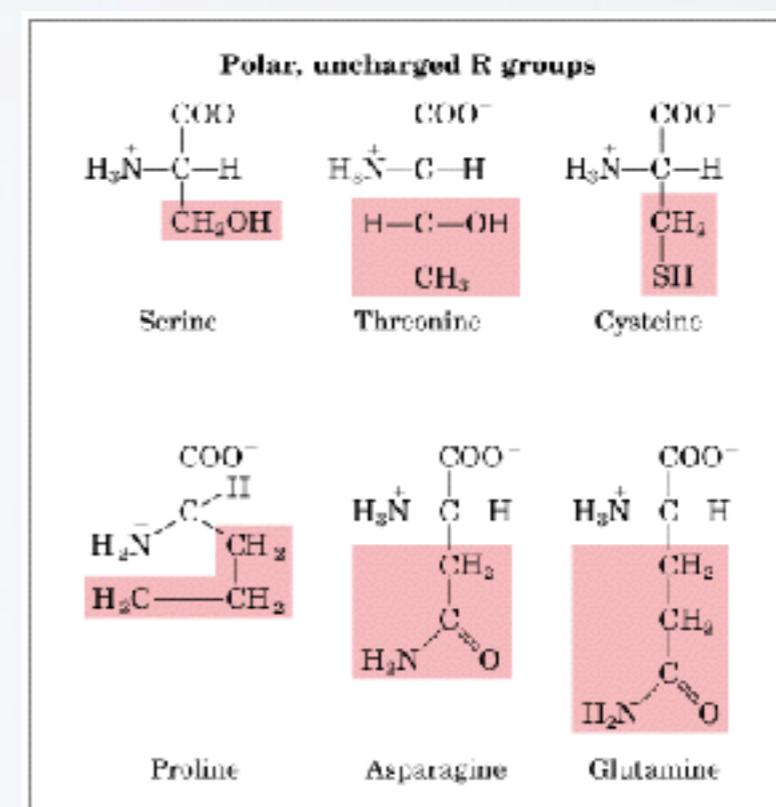
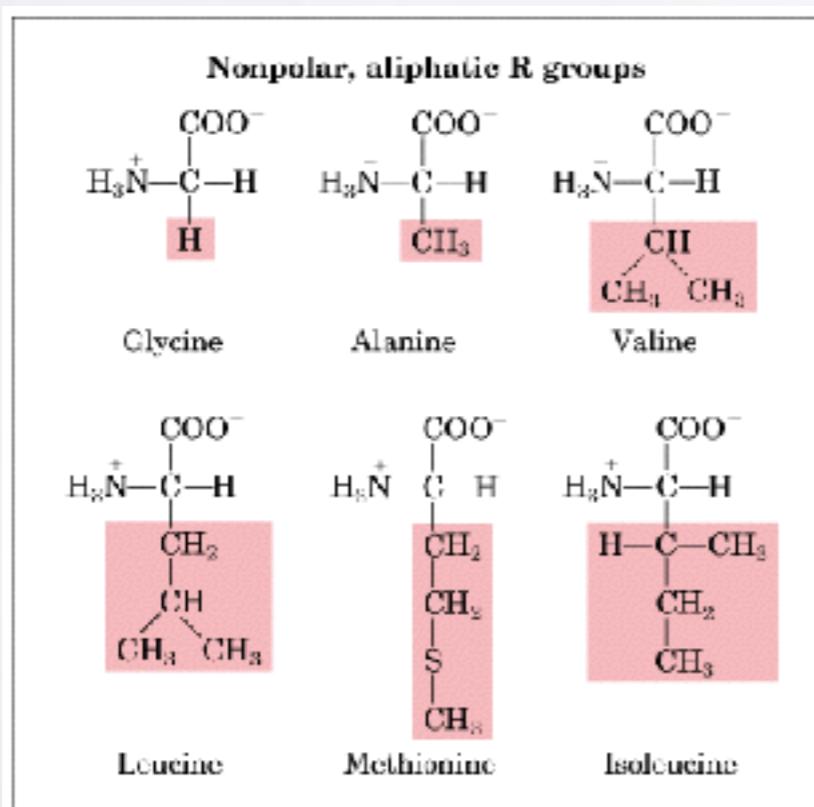
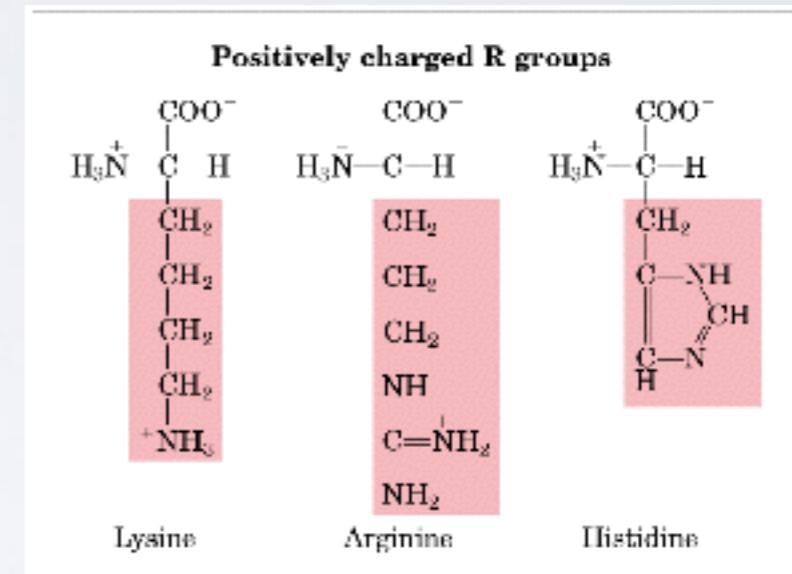
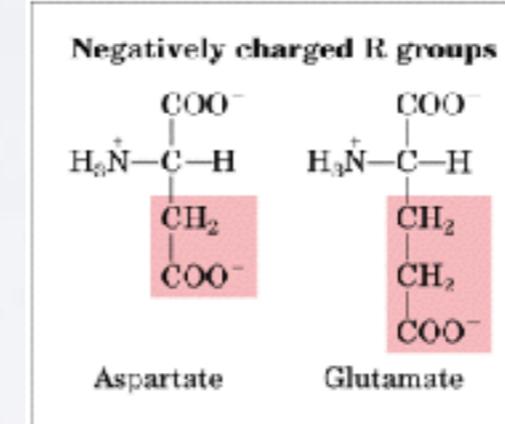
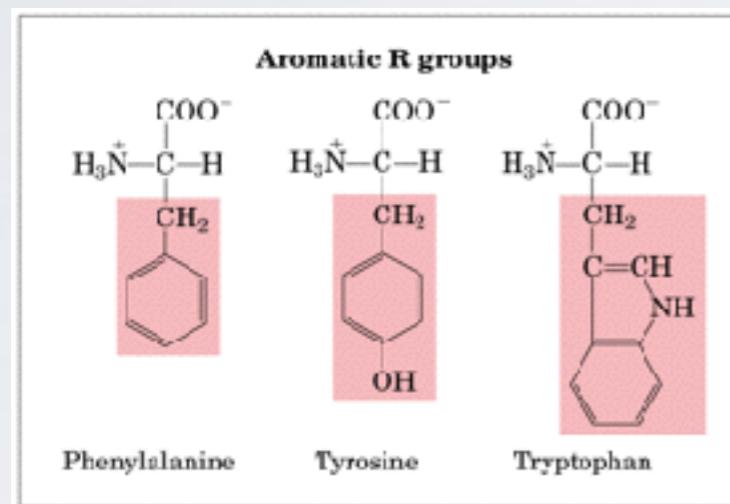
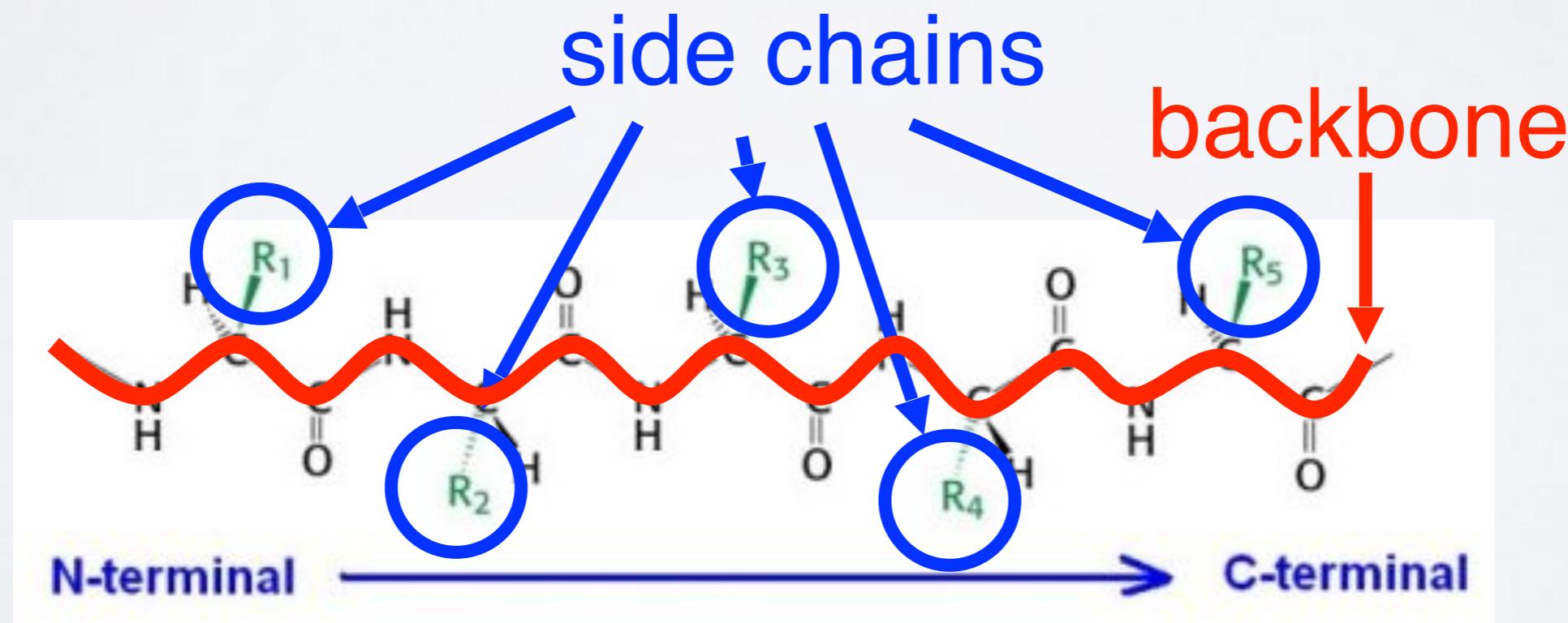
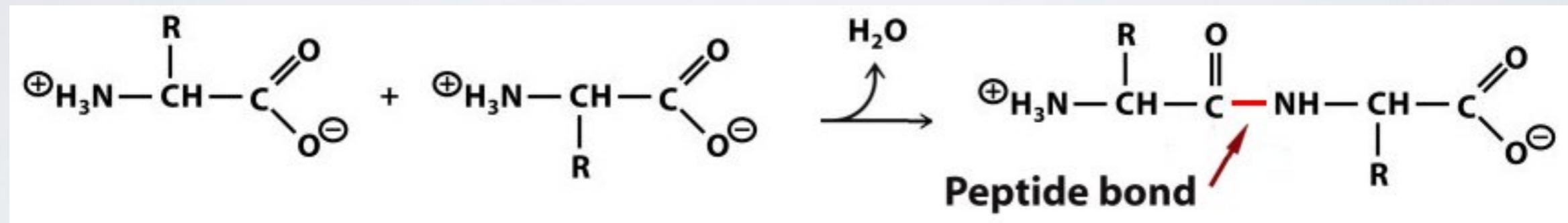


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

AMINO ACIDS CAN BE GROUPED BY THE PHYSIOCHEMICAL PROPERTIES



AMINO ACIDS POLYMERIZE THROUGH PEPTIDE BOND FORMATION



PEPTIDES CAN ADOPT DIFFERENT CONFORMATIONS BY VARYING THEIR PHI & PSI BACKBONE TORSIONS

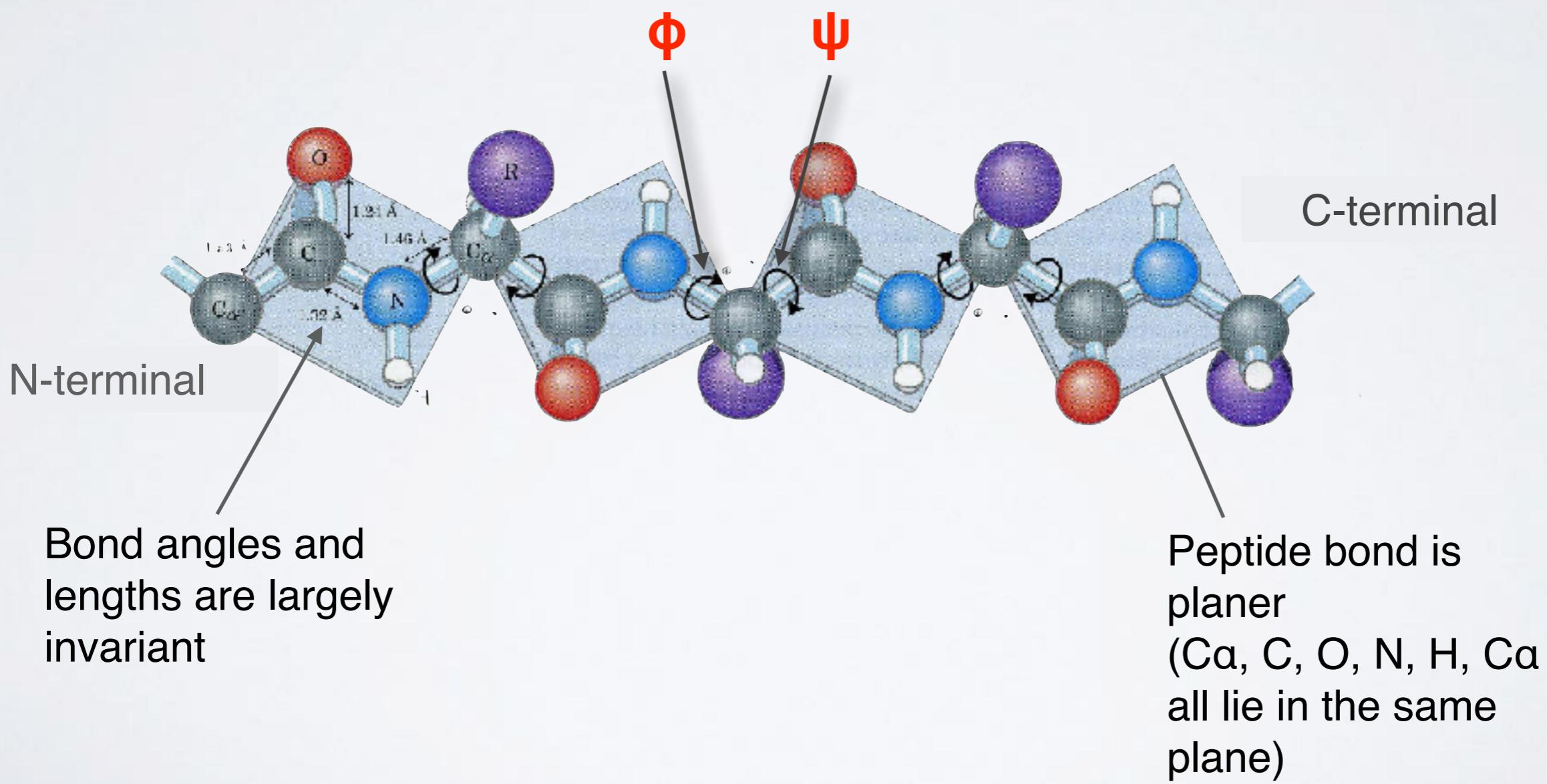
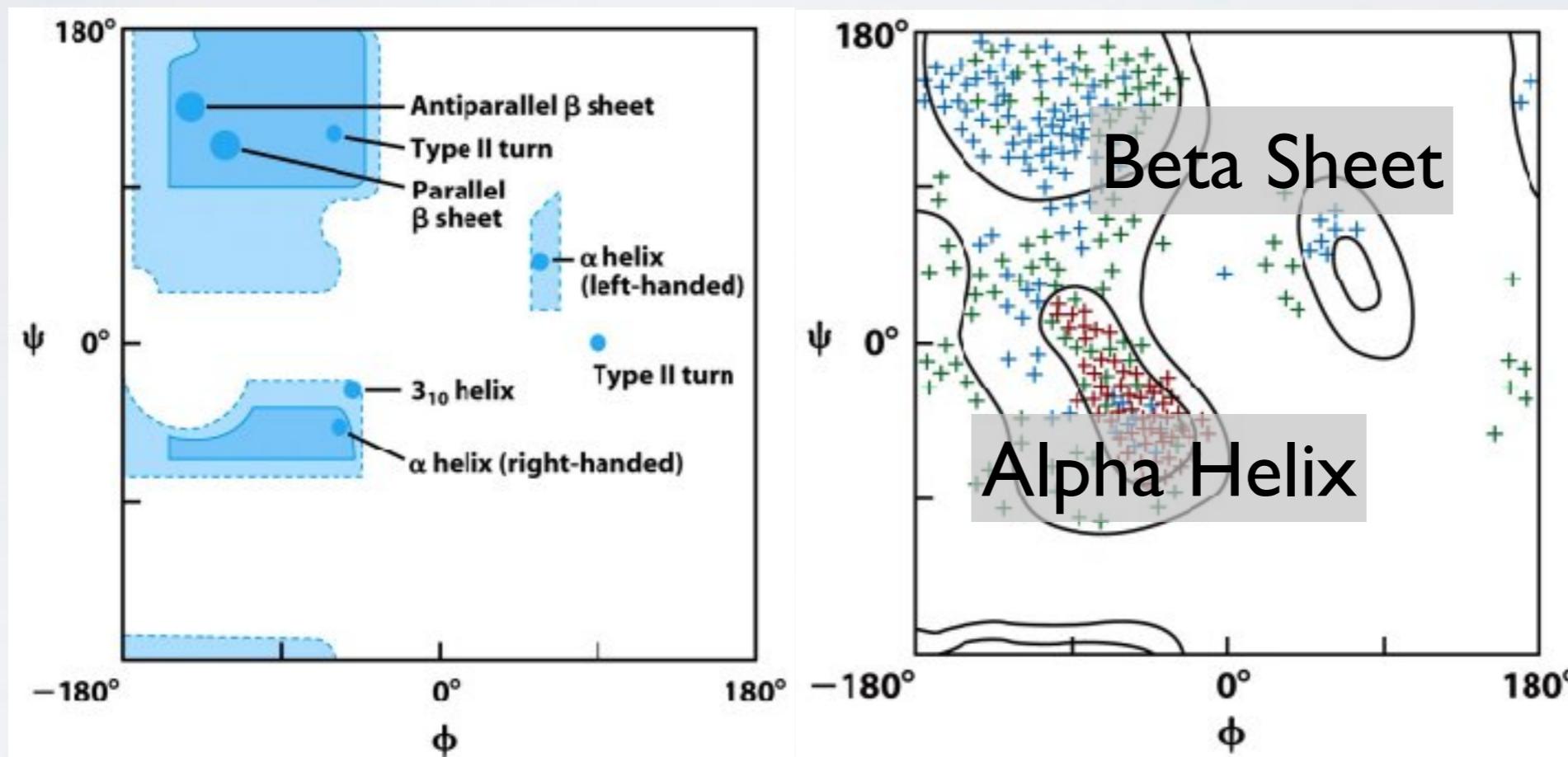


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

PHI vs PSI PLOTS ARE KNOWN AS RAMACHANDRAN DIAGRAMS

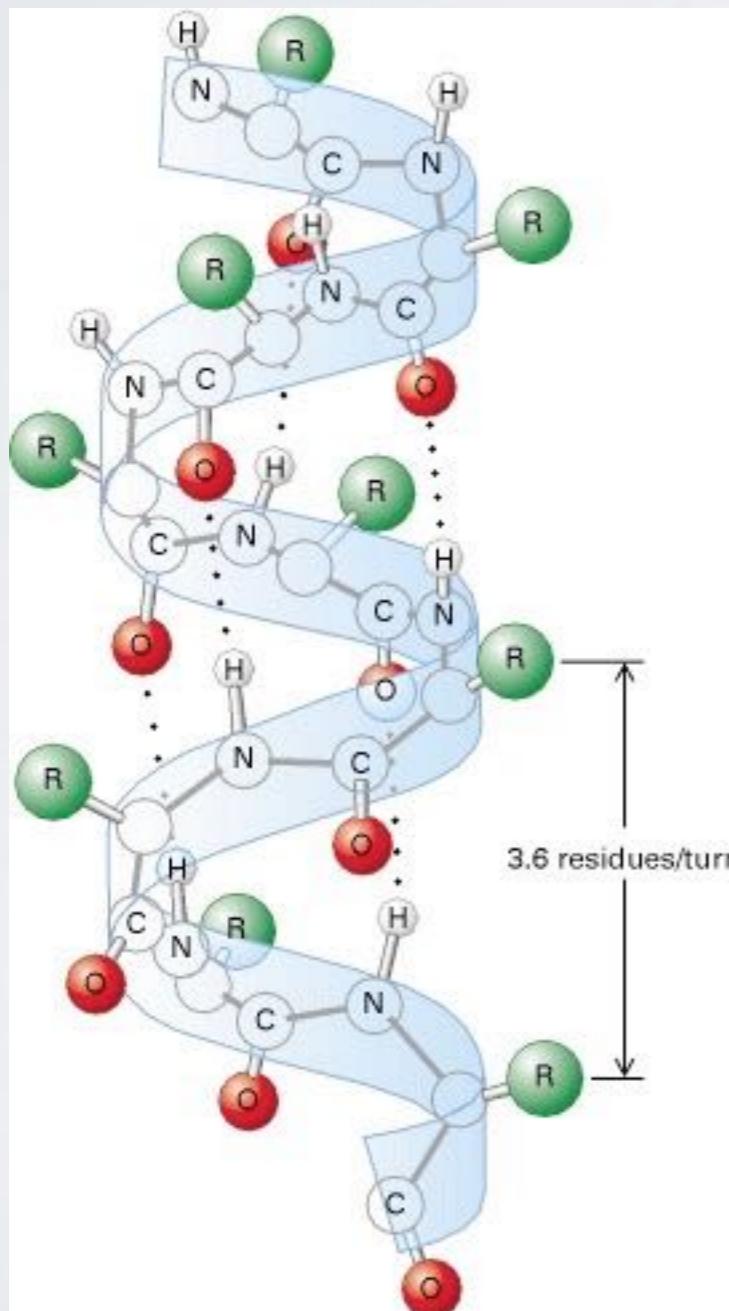


- Steric hindrance dictates torsion angle preference
- Ramachandran plot show preferred regions of ϕ and ψ dihedral angles which correspond to major forms of **secondary structure**

Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

MAJOR SECONDARY STRUCTURE TYPES

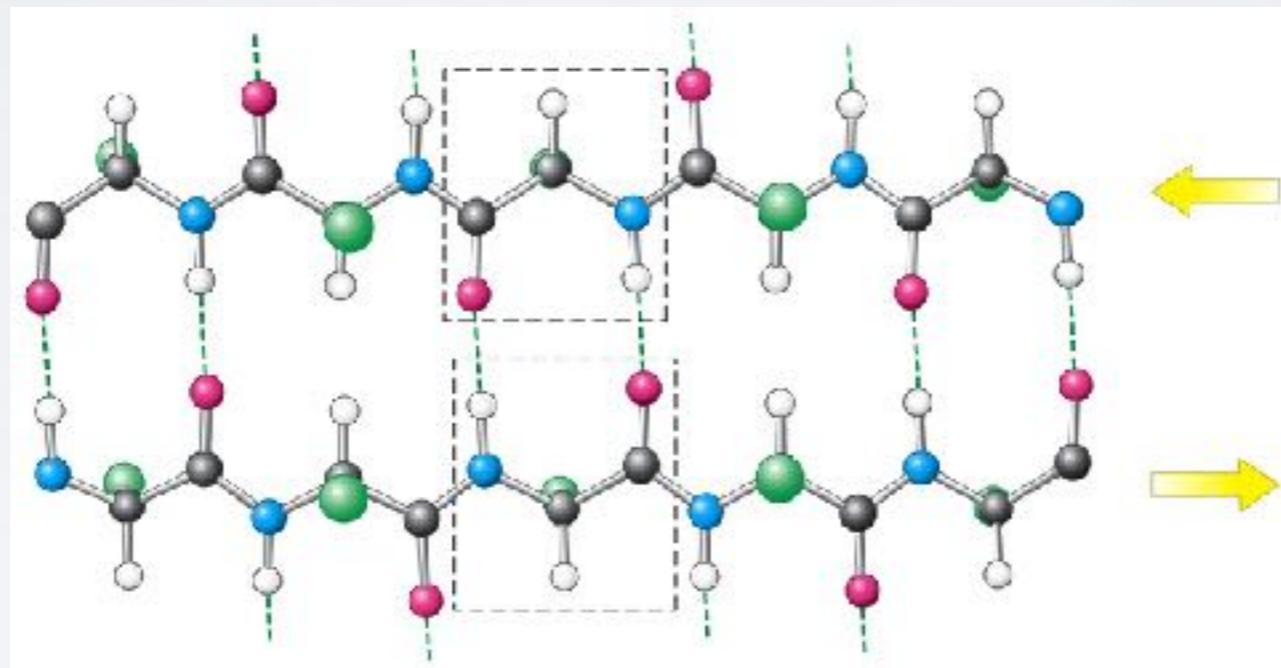
ALPHA HELIX & BETA SHEET



α -helix

- Most common form has 3.6 residues per turn (number of residues in one full rotation)
- Hydrogen bonds (dashed lines) between residue i and $i+4$ stabilize the structure
- The side chains (in green) protrude outward
- 3_{10} -helix and π -helix forms are less common

MAJOR SECONDARY STRUCTURE TYPES ALPHA HELIX & **BETA SHEET**

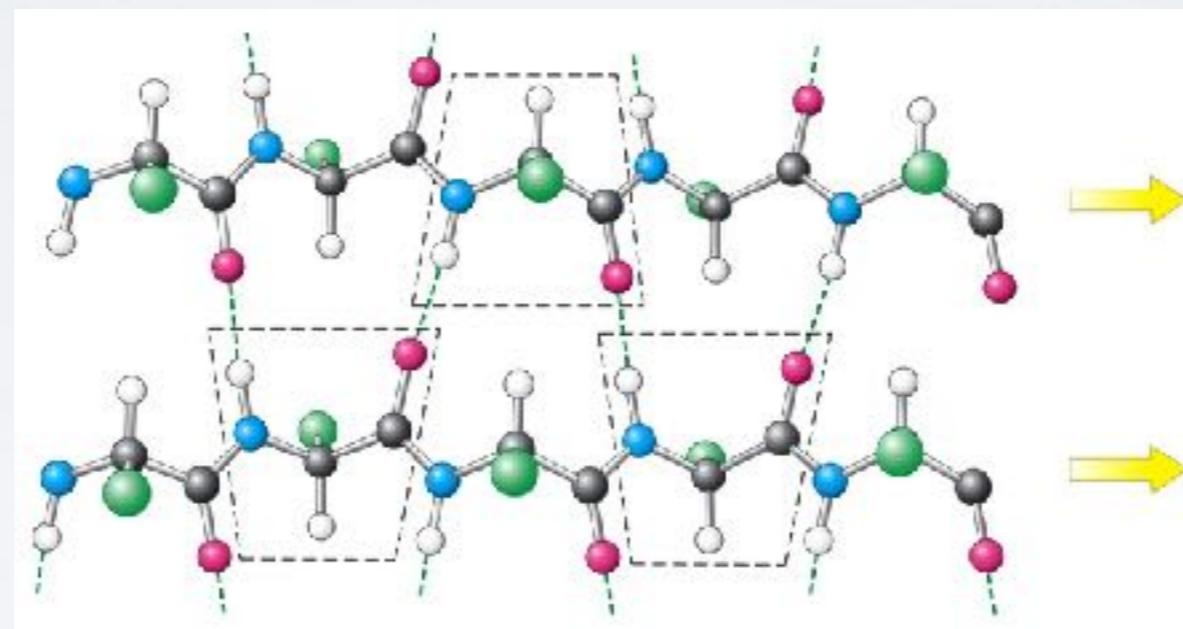


In antiparallel β -sheets

- Adjacent β -strands run in opposite directions
- Hydrogen bonds (dashed lines) between NH and CO stabilize the structure
- The side chains (in green) are above and below the sheet

Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

MAJOR SECONDARY STRUCTURE TYPES ALPHA HELIX & **BETA SHEET**

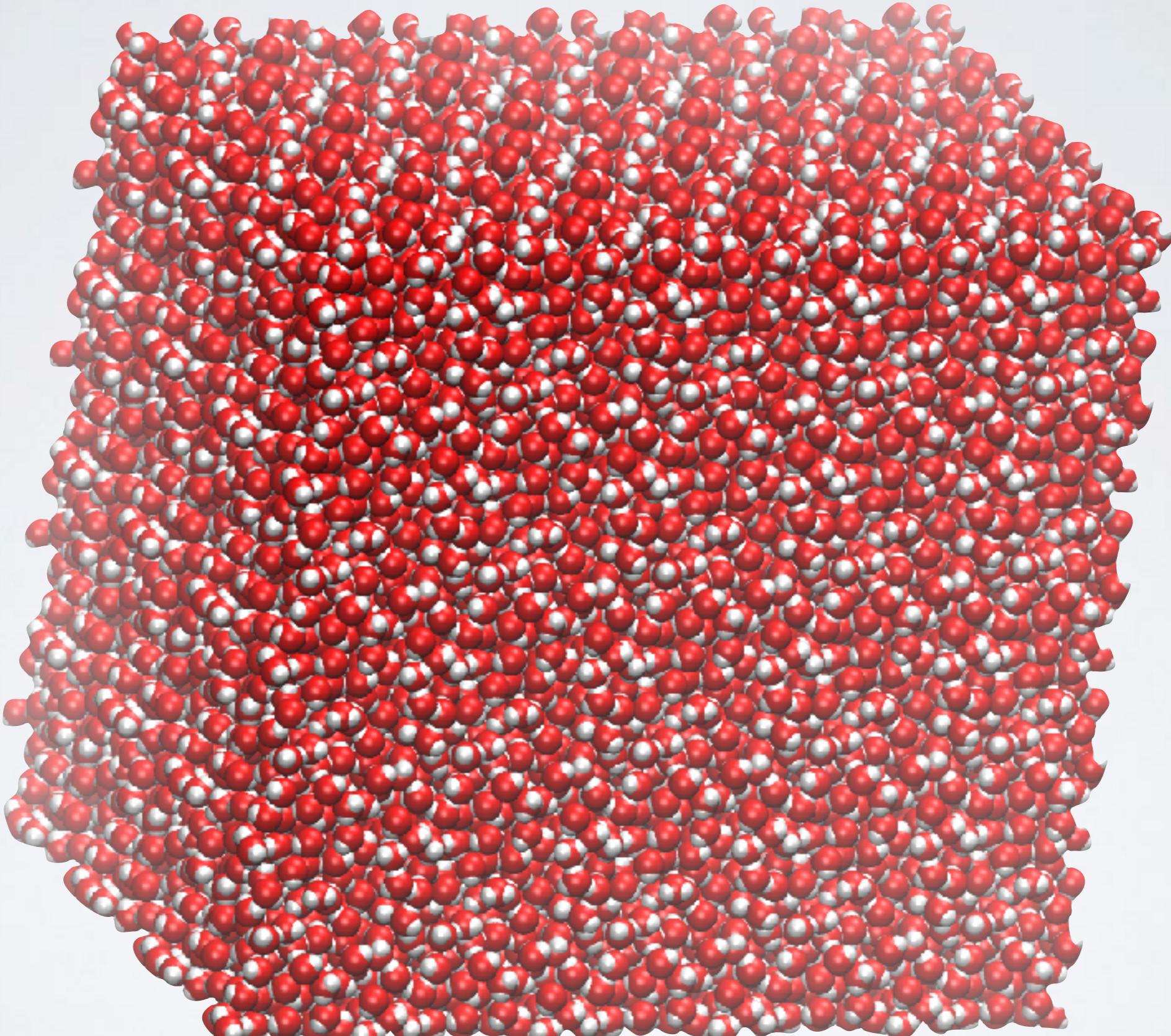


In parallel β -sheets

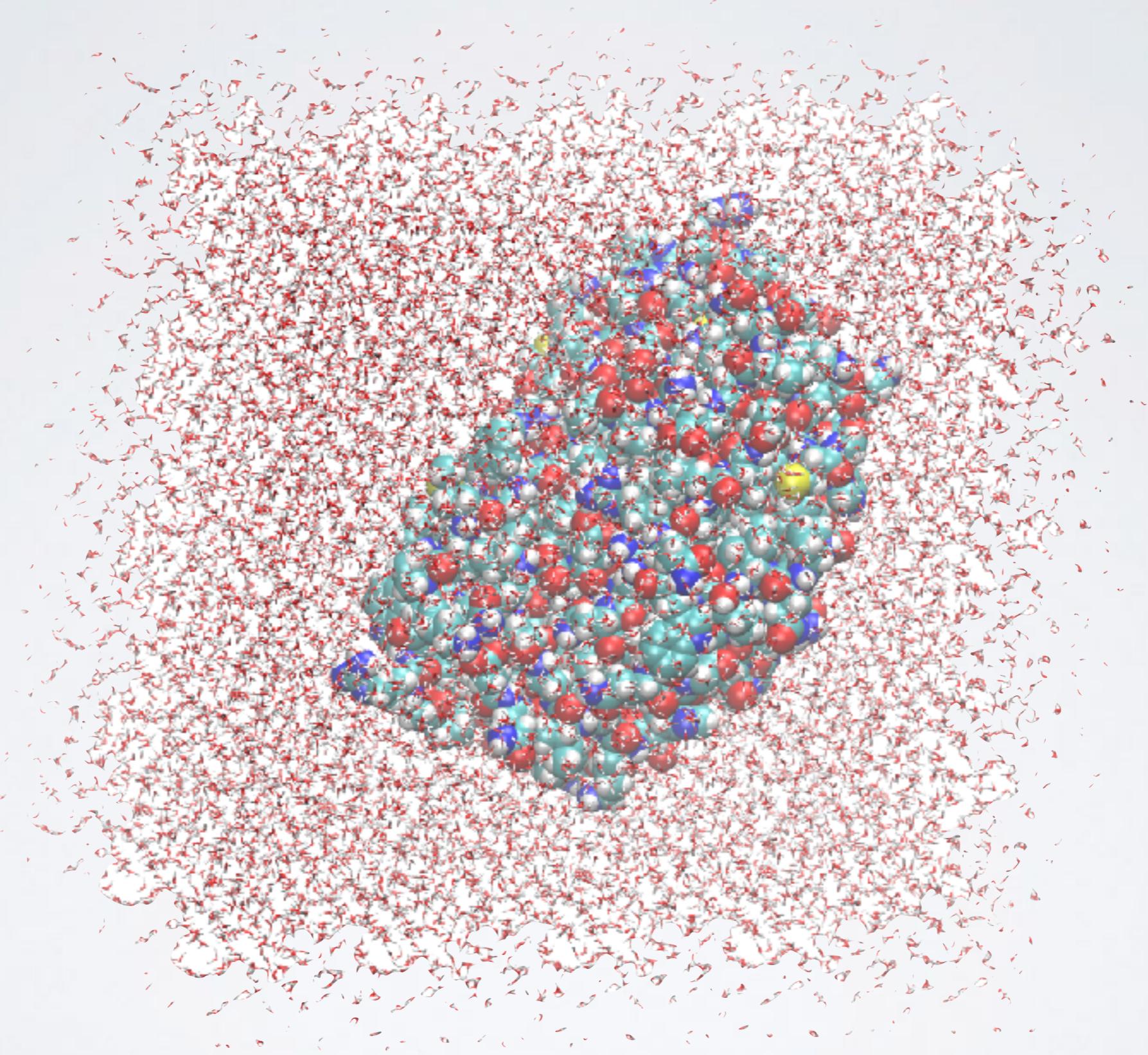
- Adjacent β -strands run in same direction
- Hydrogen bonds (dashed lines) between NH and CO stabilize the structure
- The side chains (in green) are above and below the sheet

Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

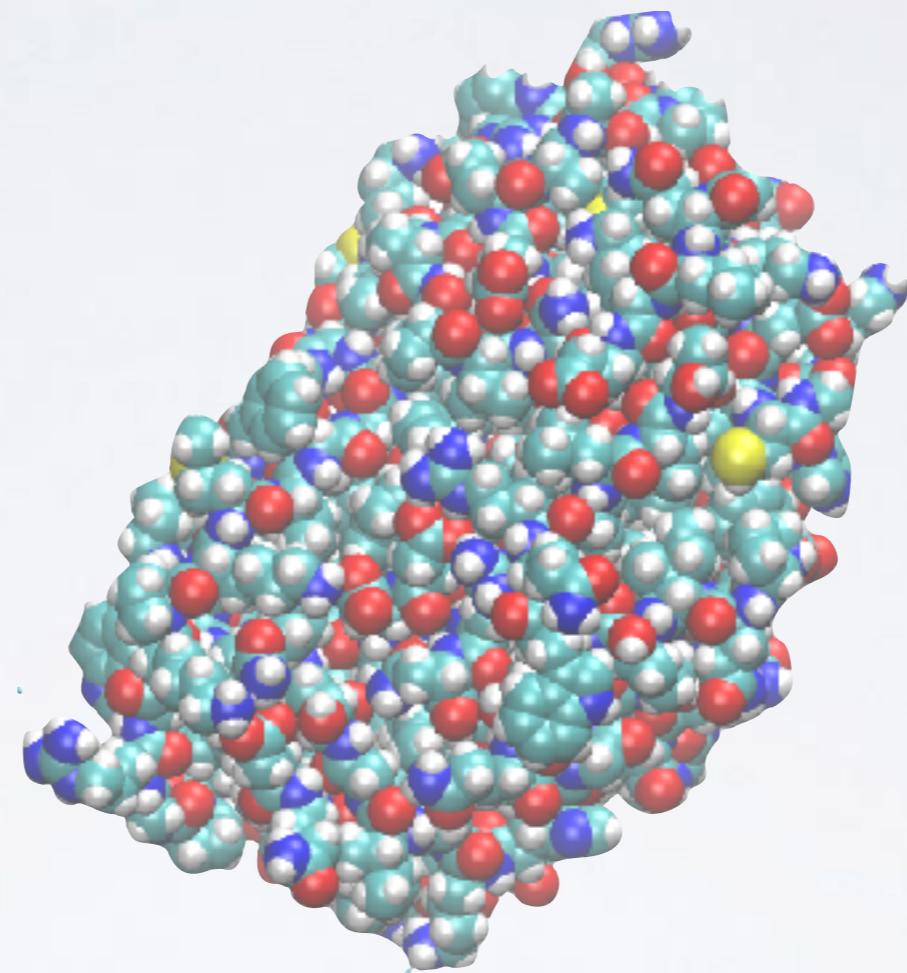
What Does a Protein Look like?



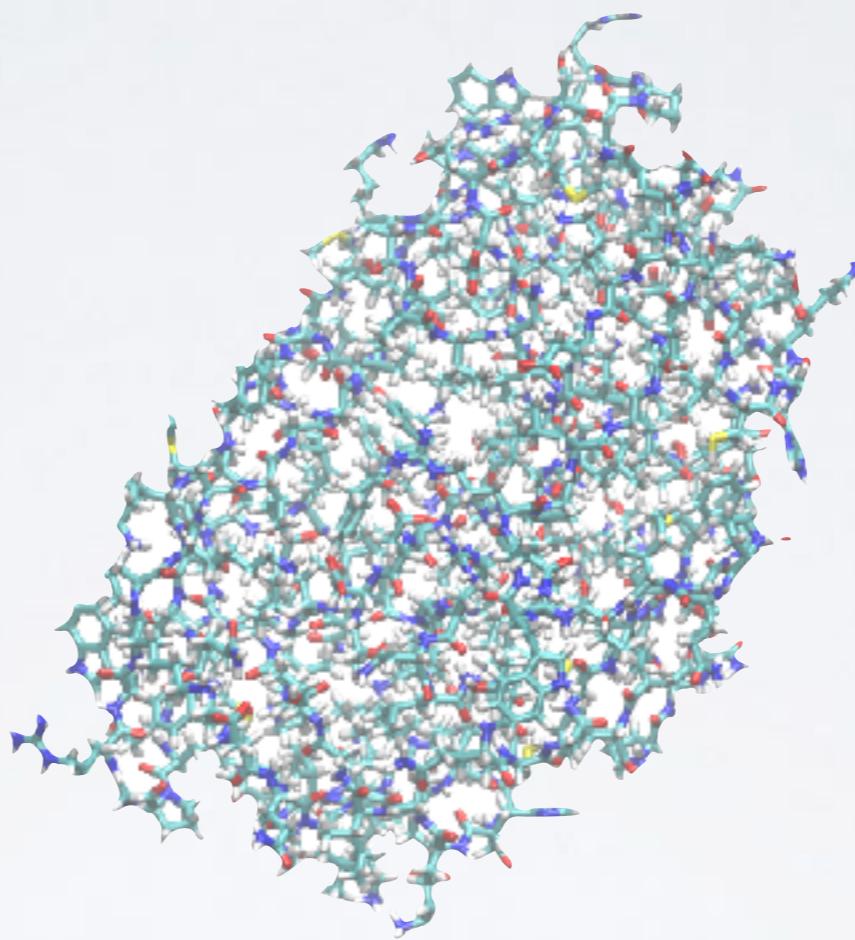
- Proteins are stable (and hidden) in water



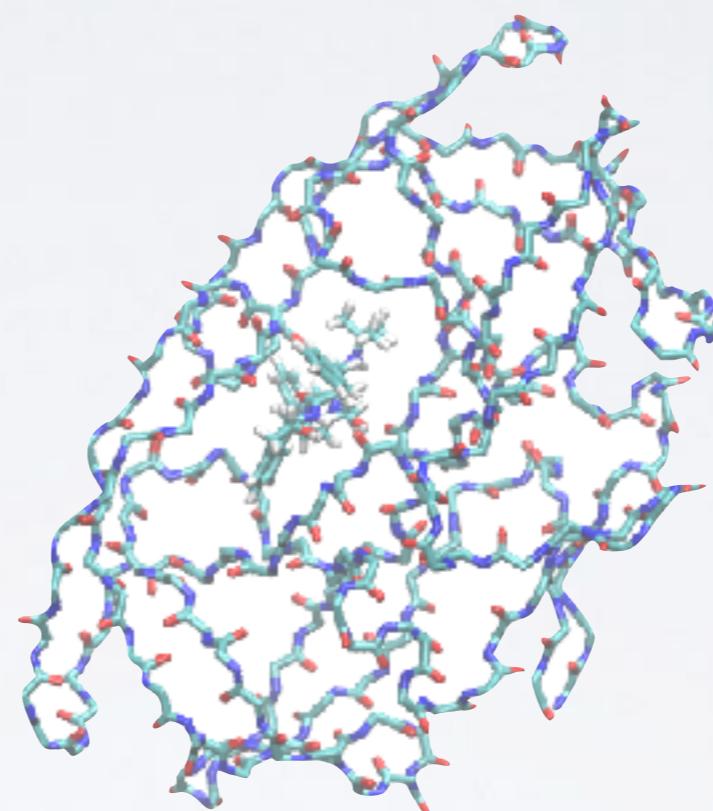
- Proteins closely interact with water



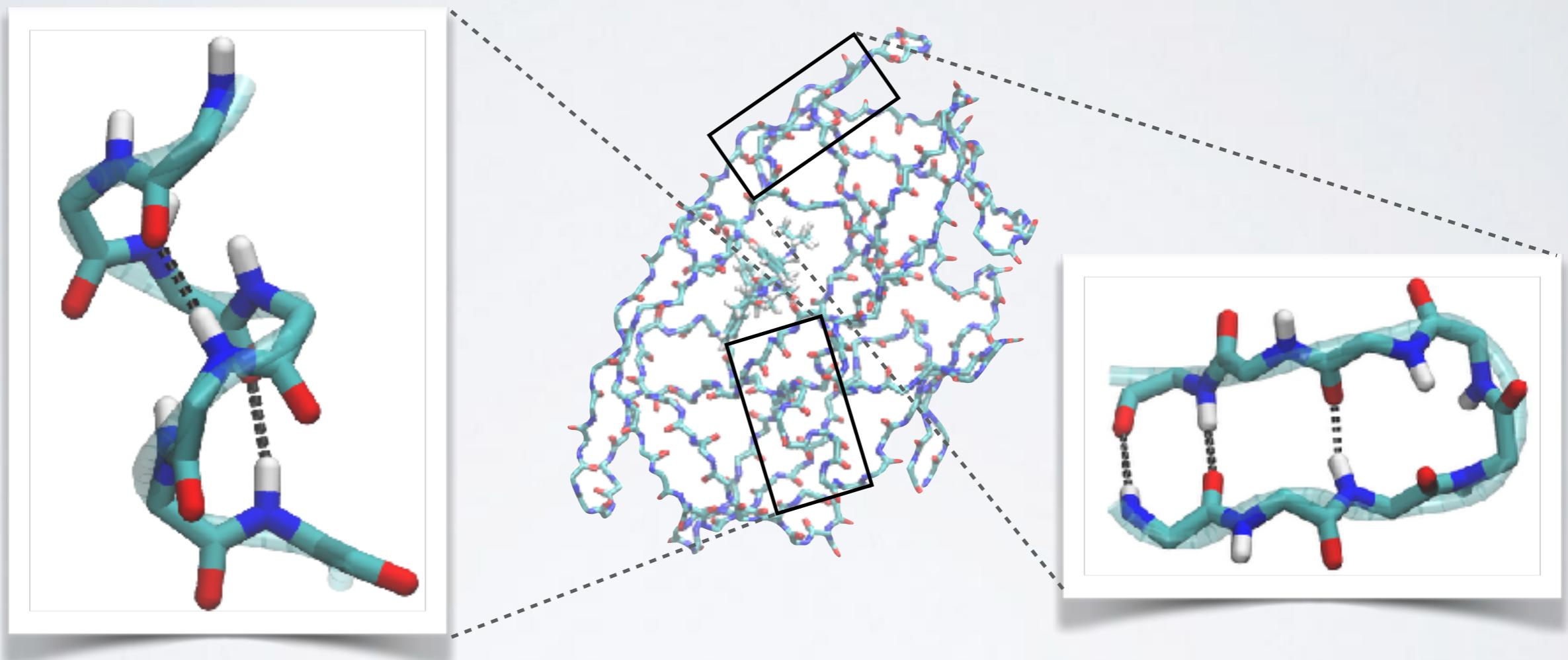
- Proteins are close packed solid but flexible objects (globular)



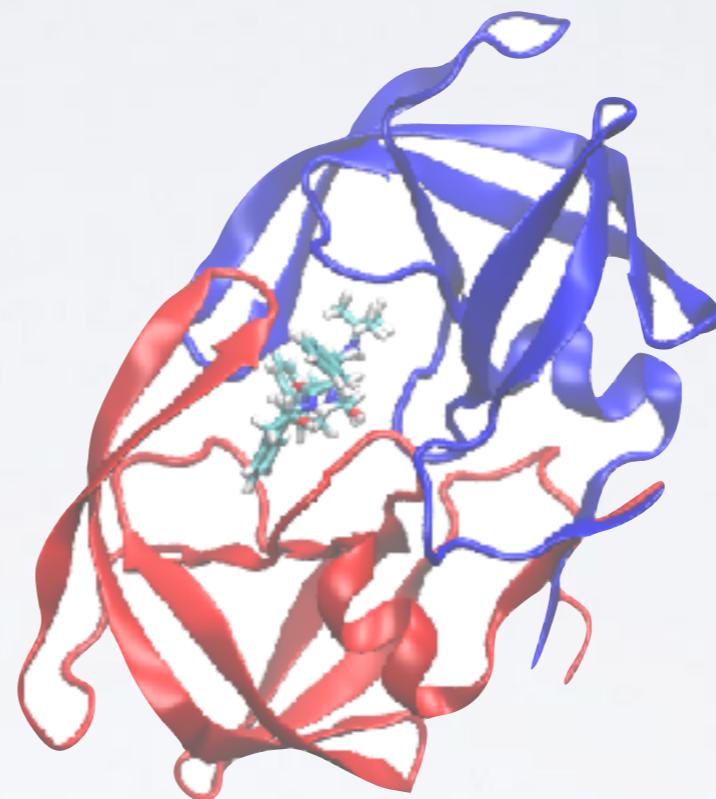
- Due to their large size and complexity it is often hard to see what's important in the structure



- Backbone or main-chain representation can help trace chain topology

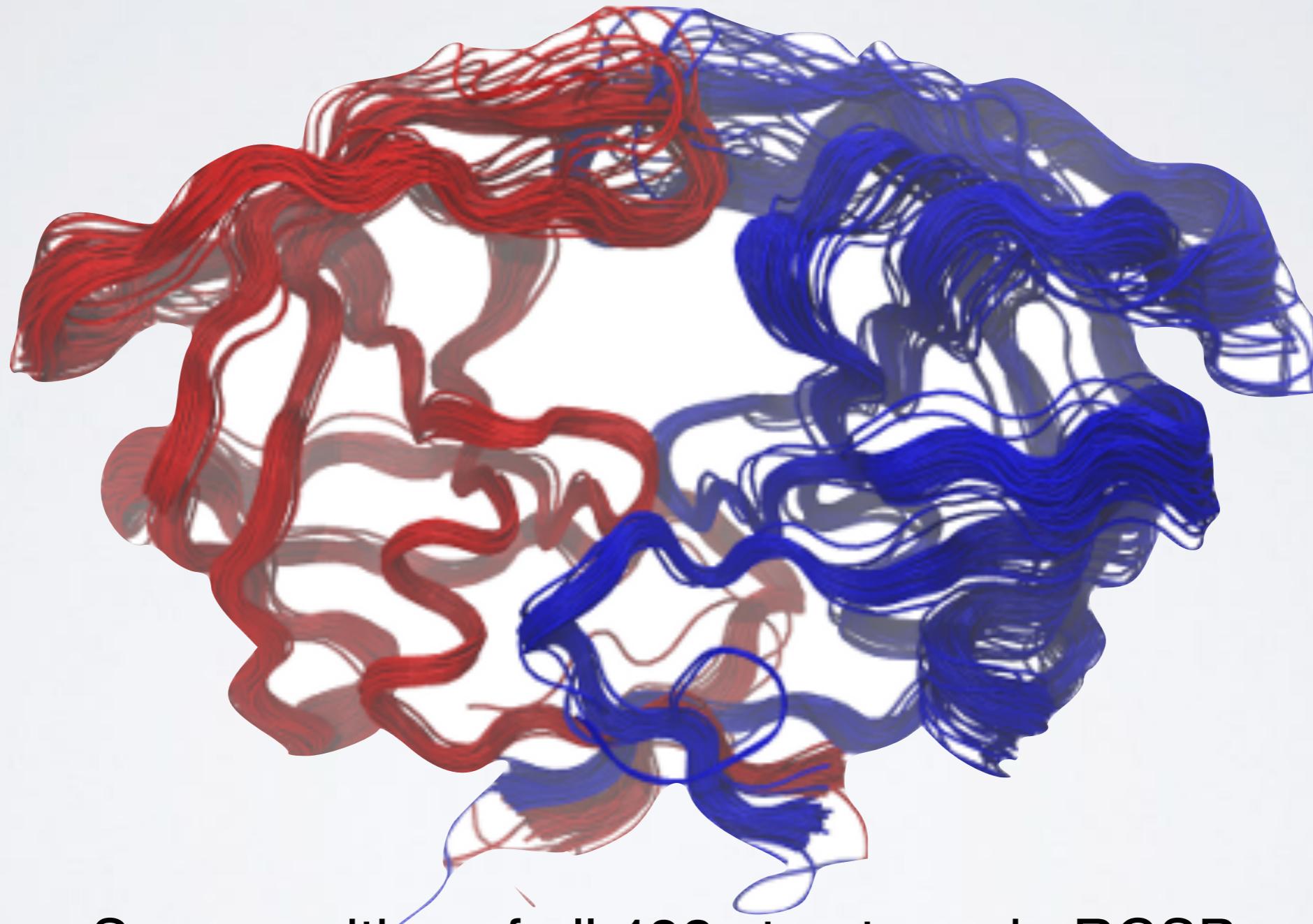


- Backbone or main-chain representation can help trace chain topology & reveal secondary structure



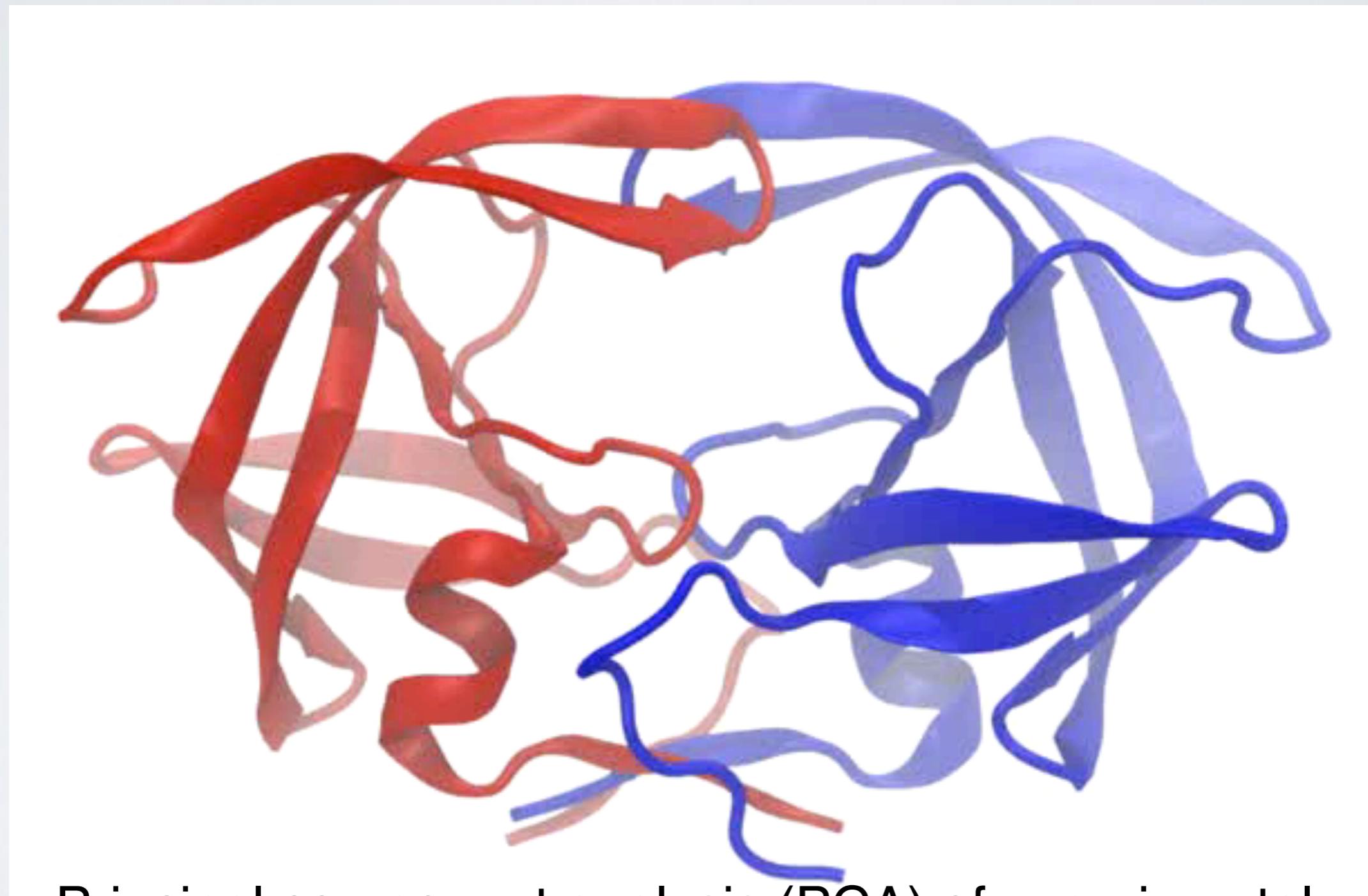
- Simplified secondary structure representations are commonly used to communicate structural details
- Now we can clearly see 2^o, 3^o and 4^o structure
- Coiled chain of connected secondary structures

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY



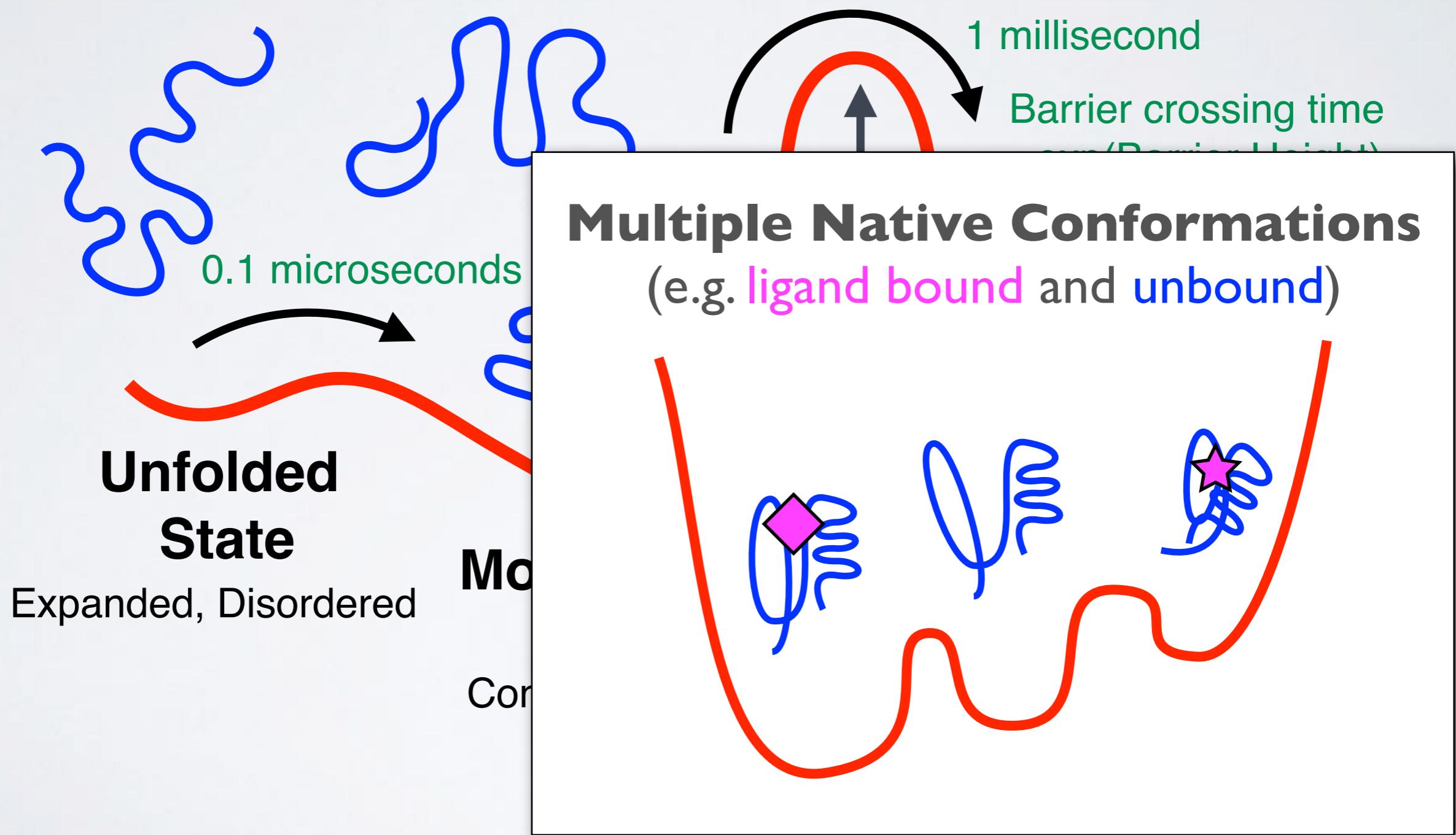
Superposition of all 482 structures in RCSB
PDB (23/09/2015)

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY



Principal component analysis (PCA) of experimental structures

KEY CONCEPT: ENERGY LANDSCAPE



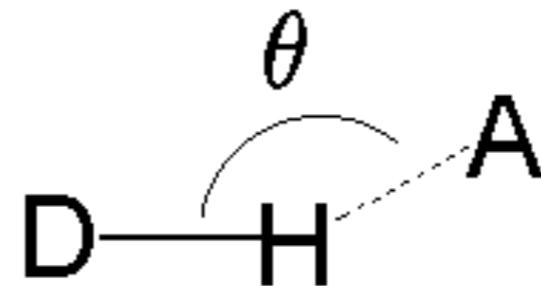
Key forces affecting structure:

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

Hydrogen-
bond donor Hydrogen-
bond acceptor



← d →

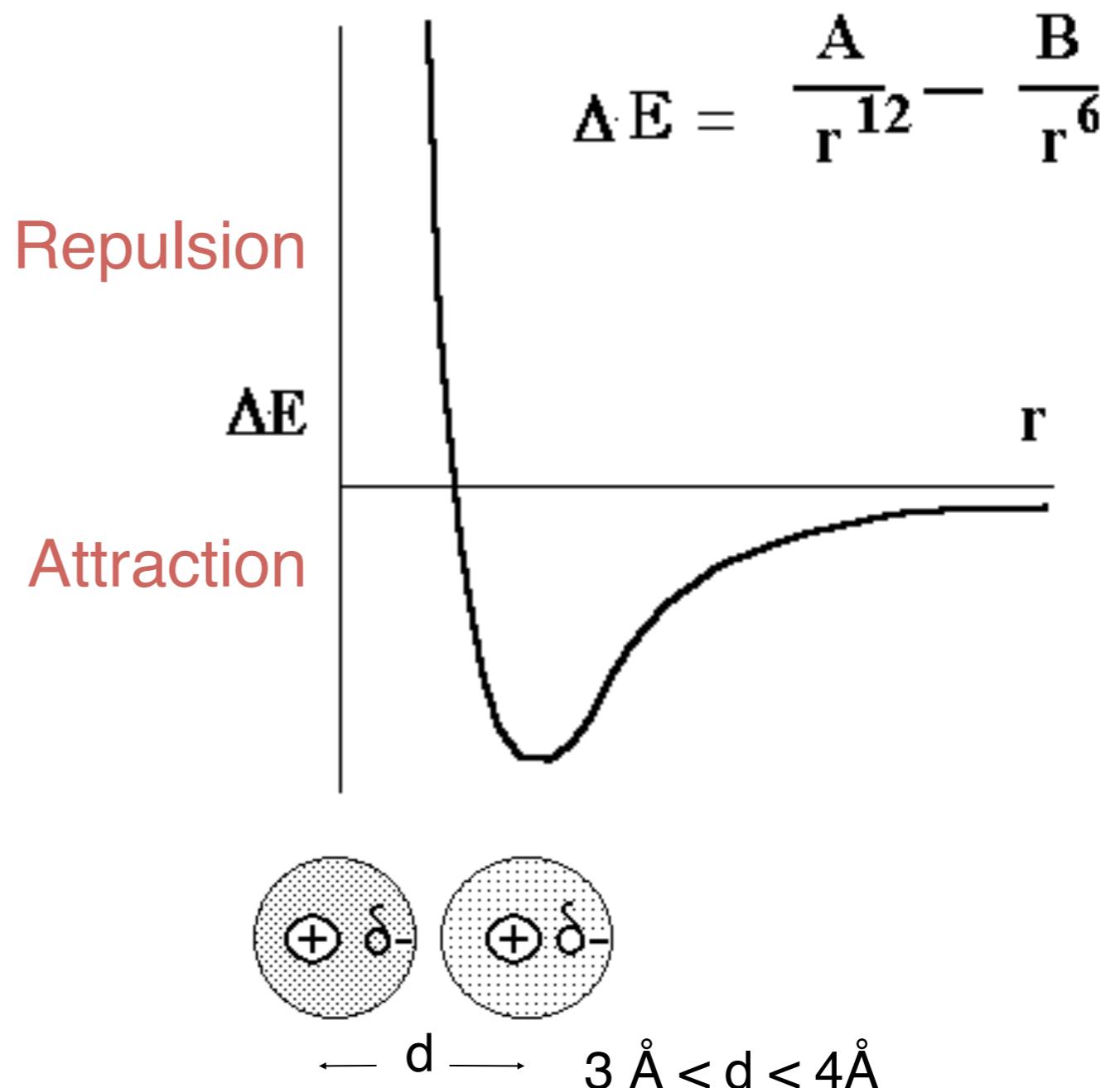


$2.6 \text{ \AA} < d < 3.1 \text{ \AA}$

$150^\circ < \theta < 180^\circ$

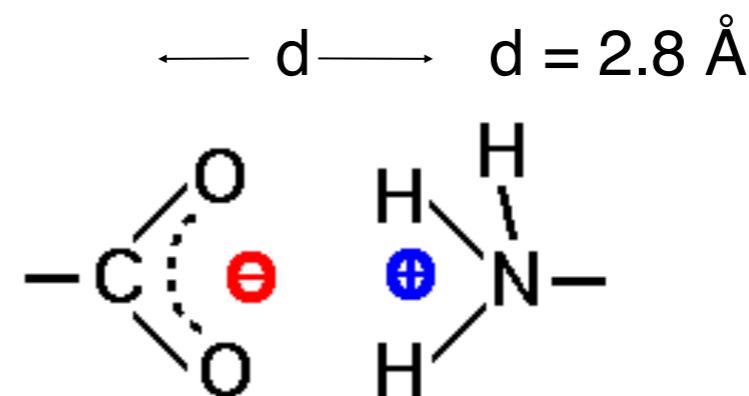
Key forces affecting structure:

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



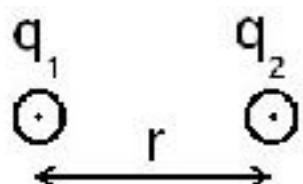
Key forces affecting structure:

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



carboxyl group and amino group

(some time called IONIC BONDs or SALT BRIDGEs)



Coulomb's law

$$E = \frac{K q_1 q_2}{D r}$$

E = Energy

k = constant

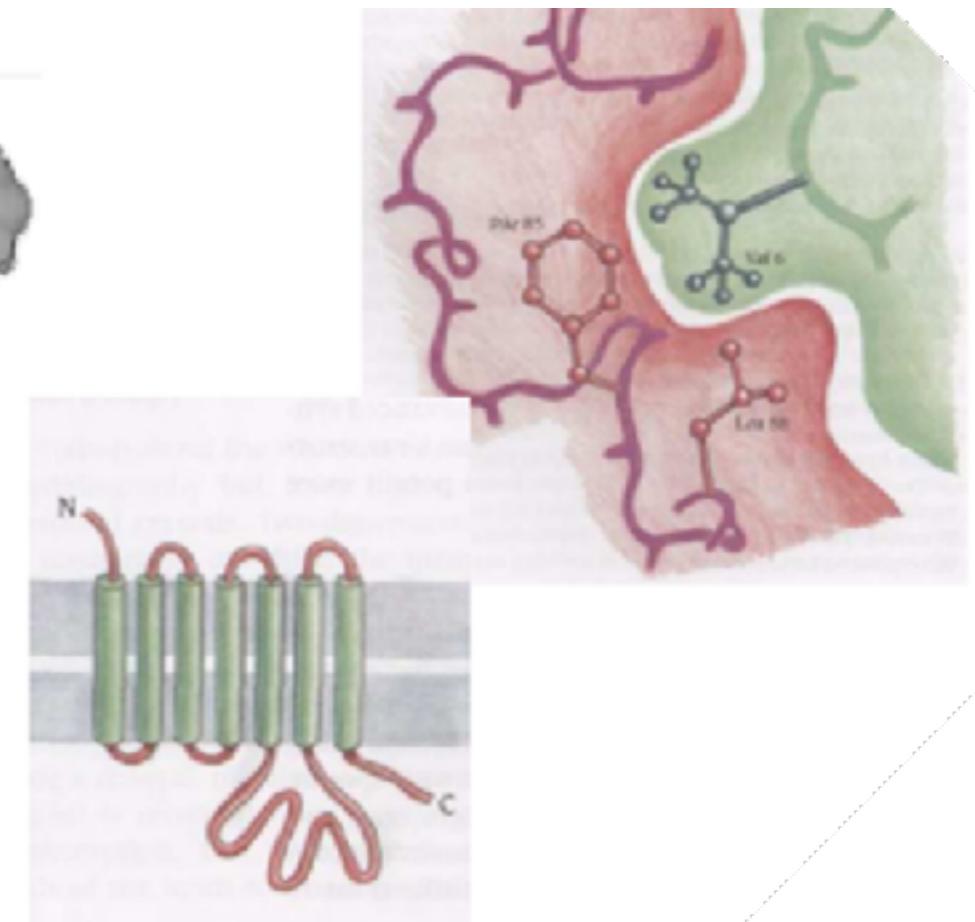
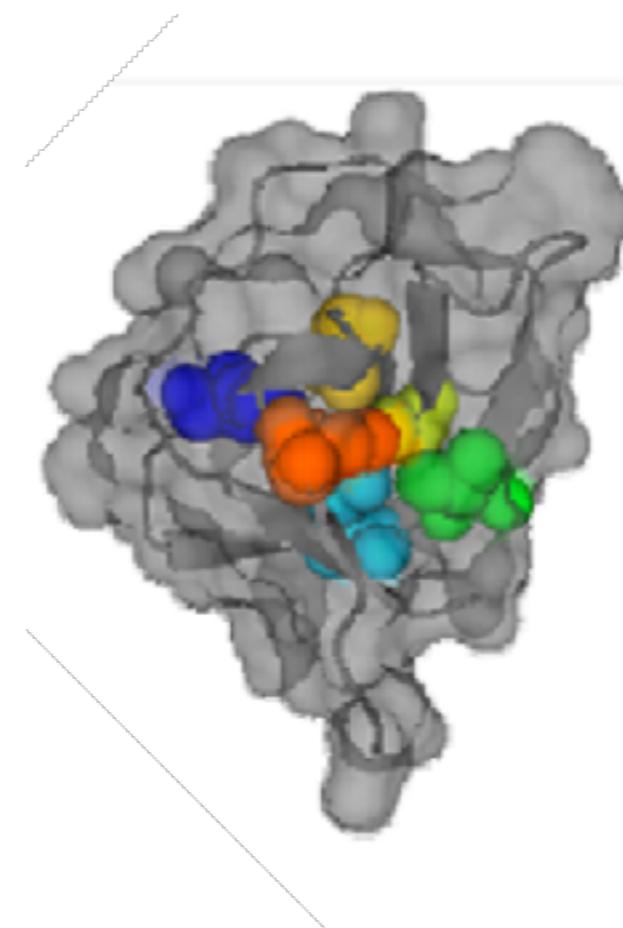
D = Dielectric constant (vacuum = 1; H₂O = 80)

q₁ & q₂ = electronic charges (Coulombs)

r = distance (Å)

Key forces affecting structure:

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



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

Do it Yourself!

Hand-on time!

https://bioboot.github.io/bimm143_S18/lectures/#11

Focus on **section 1 to 3** only please!

SIDE-NOTE: PDB FILE FORMAT

				Chain name			
				Sequence Number			
				-----Coordinates-----			
		Amino Acid	Element	X	Y	Z	(etc.)
ATOM	1	N	ASP L	1	4.060	7.307	5.186
ATOM	2	CA	ASP L	1	4.042	7.776	6.553
ATOM	3	C	ASP L	1	2.668	8.426	6.644
ATOM	4	O	ASP L	1	1.987	8.438	5.606
ATOM	5	CB	ASP L	1	5.090	8.827	6.797
ATOM	6	CG	ASP L	1	6.338	8.761	5.929
ATOM	7	OD1	ASP L	1	6.576	9.758	5.241
ATOM	8	OD2	ASP L	1	7.065	7.759	5.948

\\
Element position within amino acid

- **PDB files** contains atomic coordinates and associated information.

SIDE-NOTE: PDB FILE FORMAT

Amino Acid
Element
Chain
Seq

ATOM	1	N	ASP	L	1
ATOM	2	CA	ASP	L	1
ATOM	3	C	ASP	L	1
ATOM	4	O	ASP	L	1
ATOM	5	CB	ASP	L	1
ATOM	6	CG	ASP	L	1
ATOM	7	OD1	ASP	L	1
ATOM	8	OD2	ASP	L	1

\\
Element position within amino acid

Diagram illustrating the structure of an alpha-amino acid. The structure shows the alpha-carbon (C-alpha) bonded to an amino group (H_3N^+), a side chain (R-group), and a carboxyl group ($\text{C}(=\text{O})\text{O}^-$). The side chain is highlighted in pink and labeled with Greek letters ϵ , δ , γ , β , and α from the amino group to the carbonyl group respectively.

- **PDB files** contains atomic coordinates and associated information.

NEXT UP:

- ▶ **Overview of structural bioinformatics**
 - Major motivations, goals and challenges
- ▶ **Fundamentals of protein structure**
 - Composition and dynamics
- ▶ **Representing protein structure**
 - Modeling energy as a function of structure
- ▶ **Example application areas**
 - Predicting functional dynamics & drug discovery

Change from here on!!!

Just do docking???

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

Two main approaches:

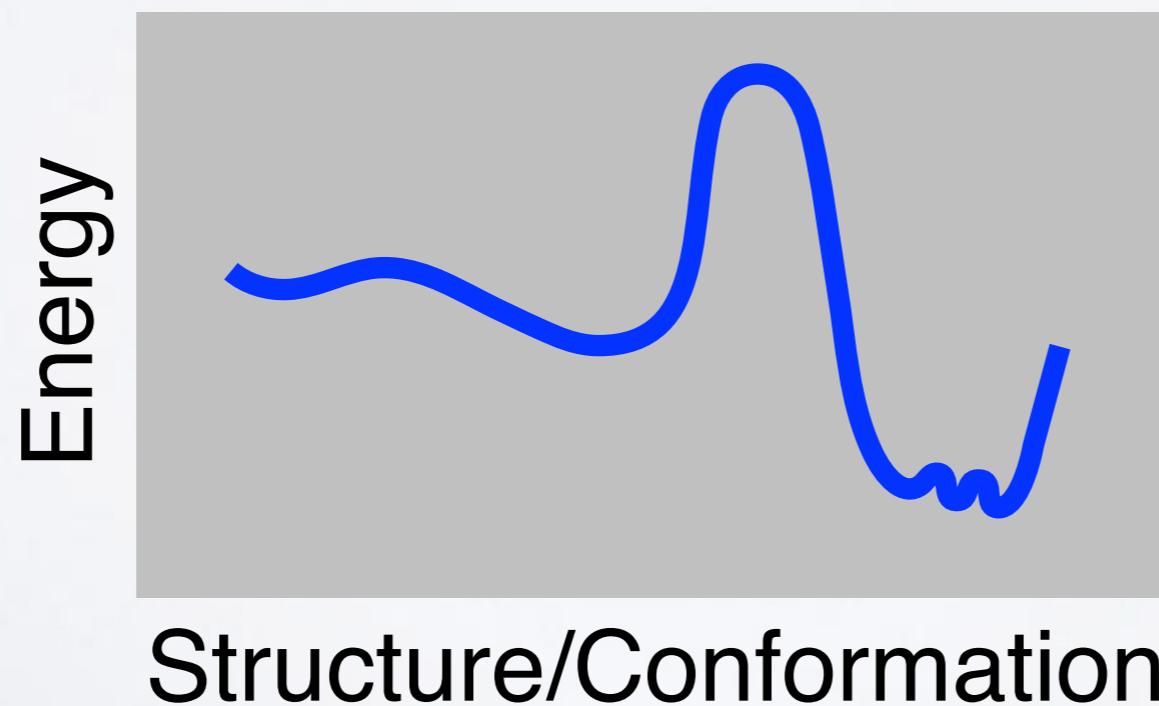
- (1). Physics-Based
- (2). Knowledge-Based

KEY CONCEPT: POTENTIAL FUNCTIONS

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

Two main approaches:

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

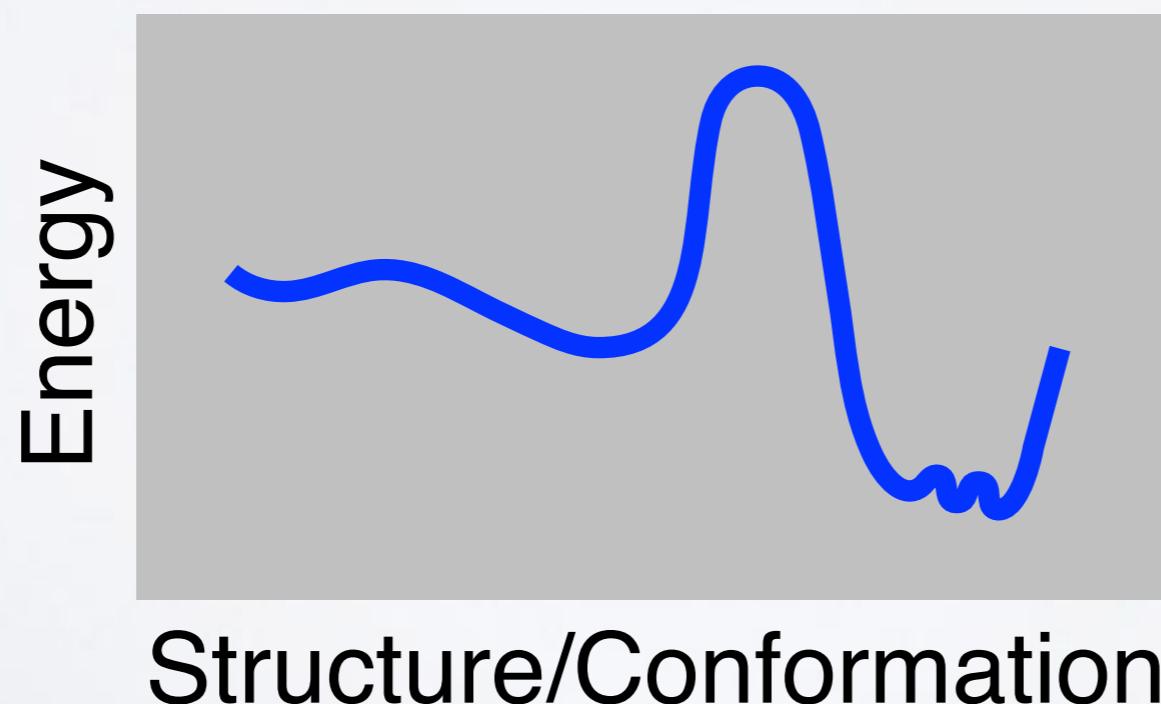


KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

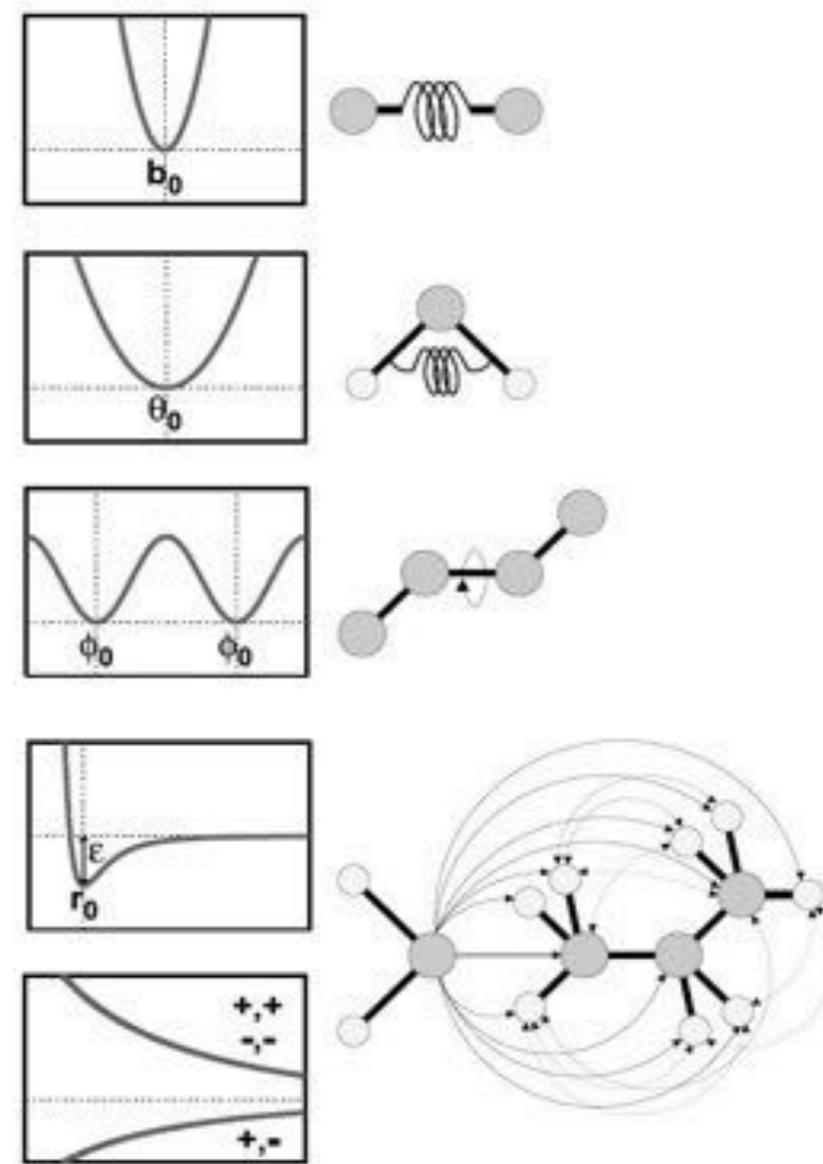
(1). **Physics-Based**

(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

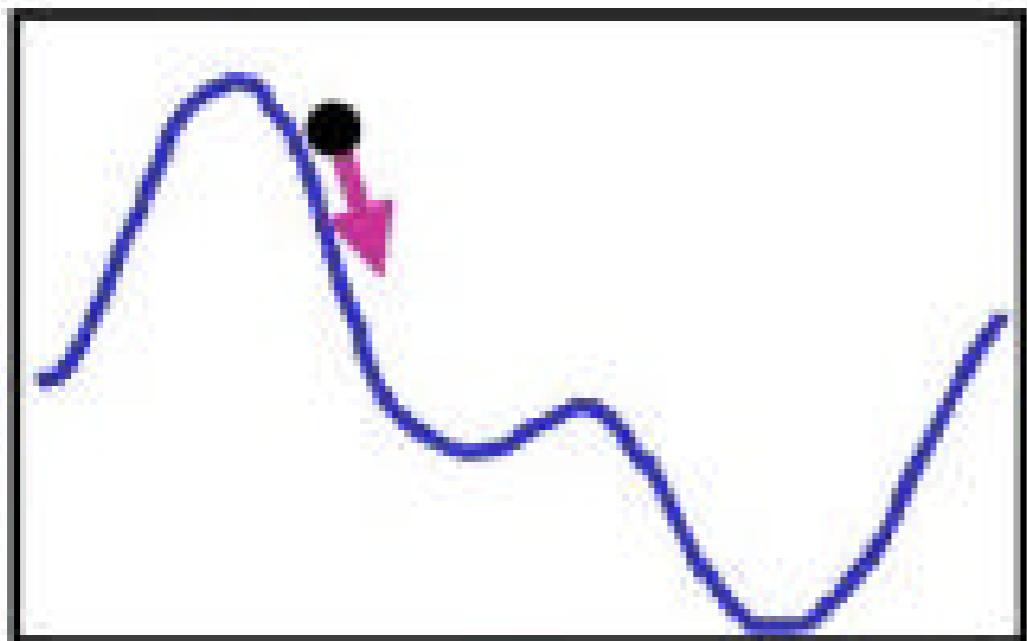
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle

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

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

TOTAL POTENTIAL ENERGY

Energy, U



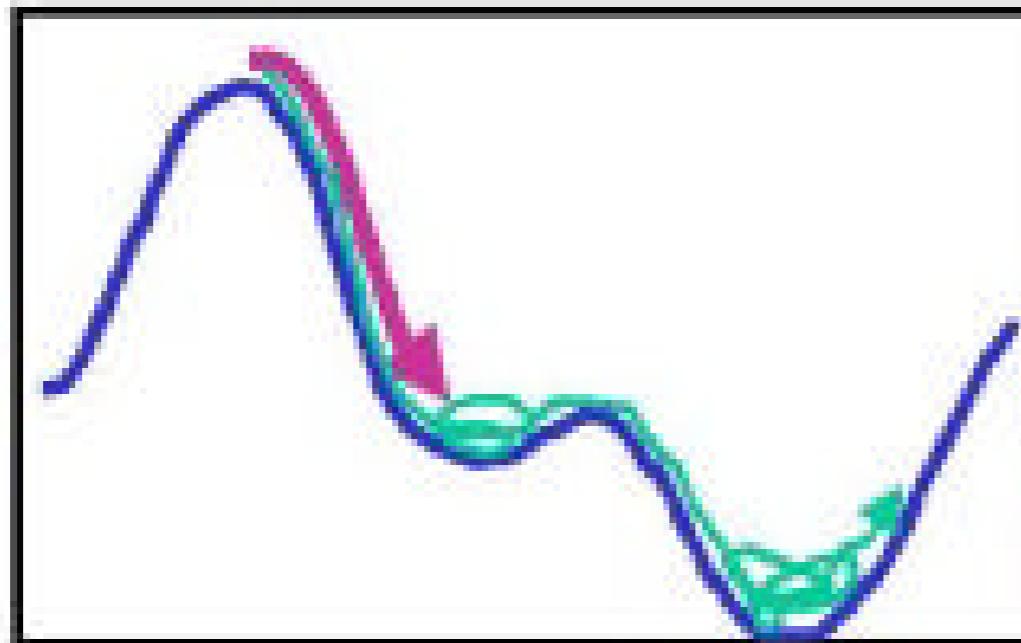
$$F(x) = -dU/dx$$

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

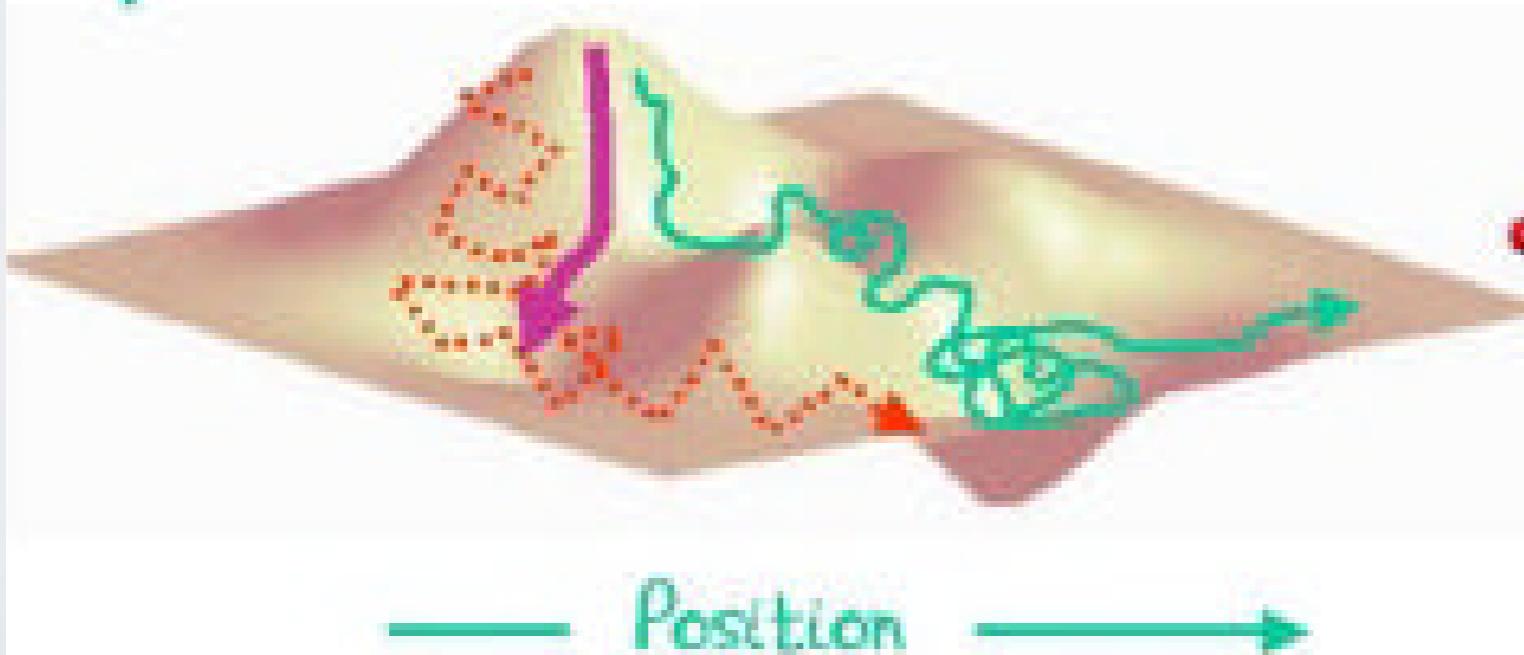
• The energy is a sum of independent terms for:
Bond, Bond angles,
Torsion angles and non-
bonded atom pairs.

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

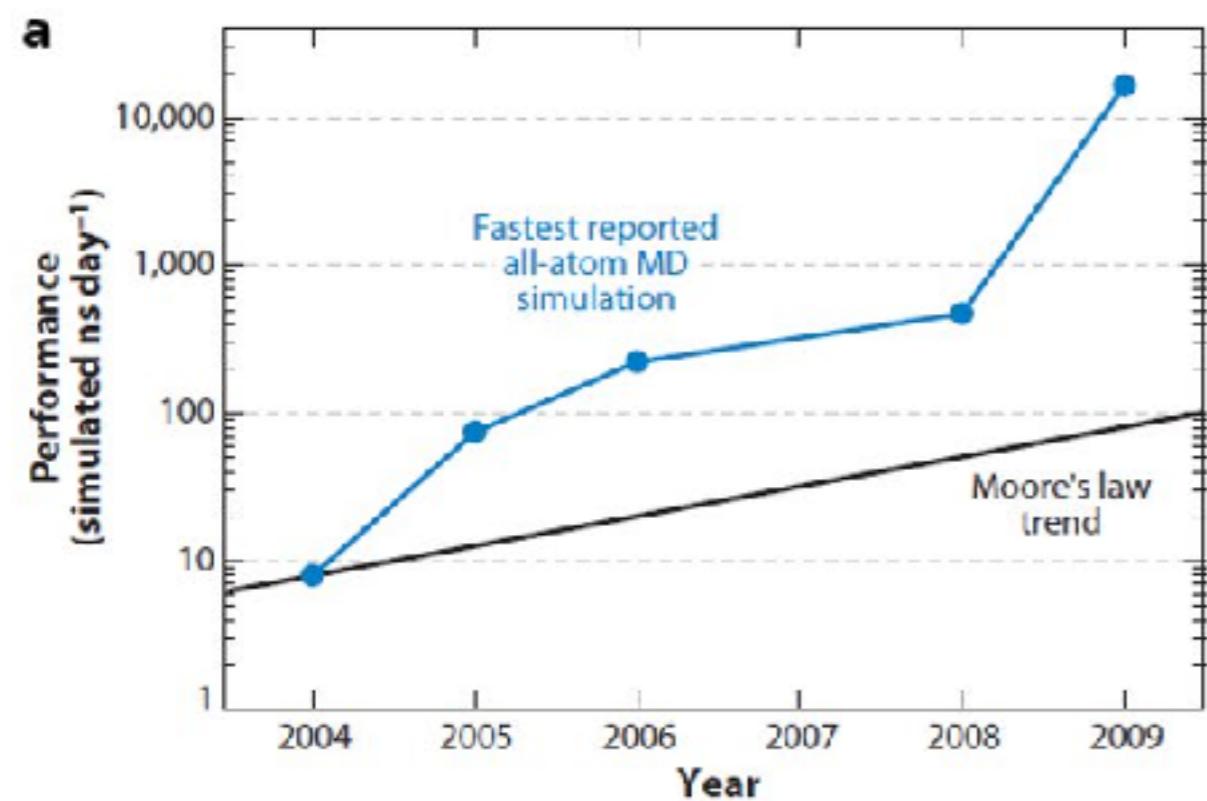
HOW COMPUTERS HAVE CHANGED

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

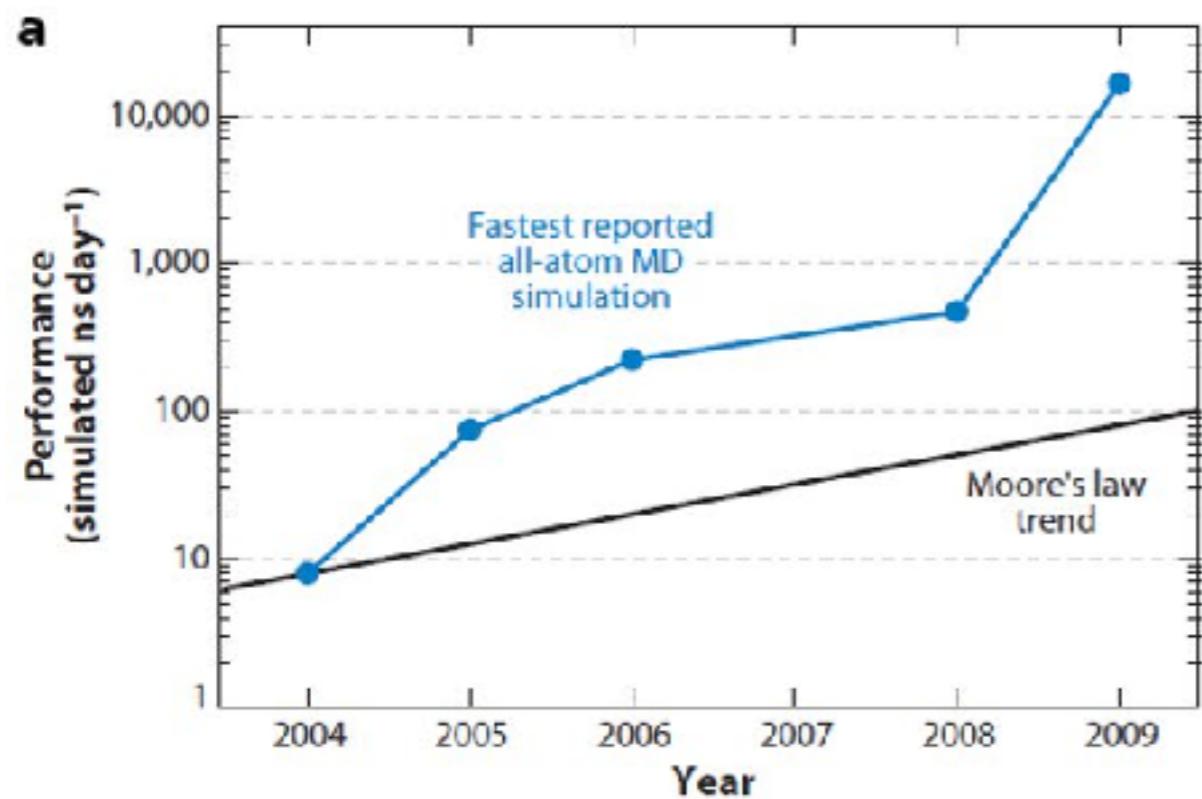
If cars were like computers then a new Vehc
would cost \$3, would have a top speed of
1,000,000 Km/hr, would carry 50,000
adults and would park in a shadow.



SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER



SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER

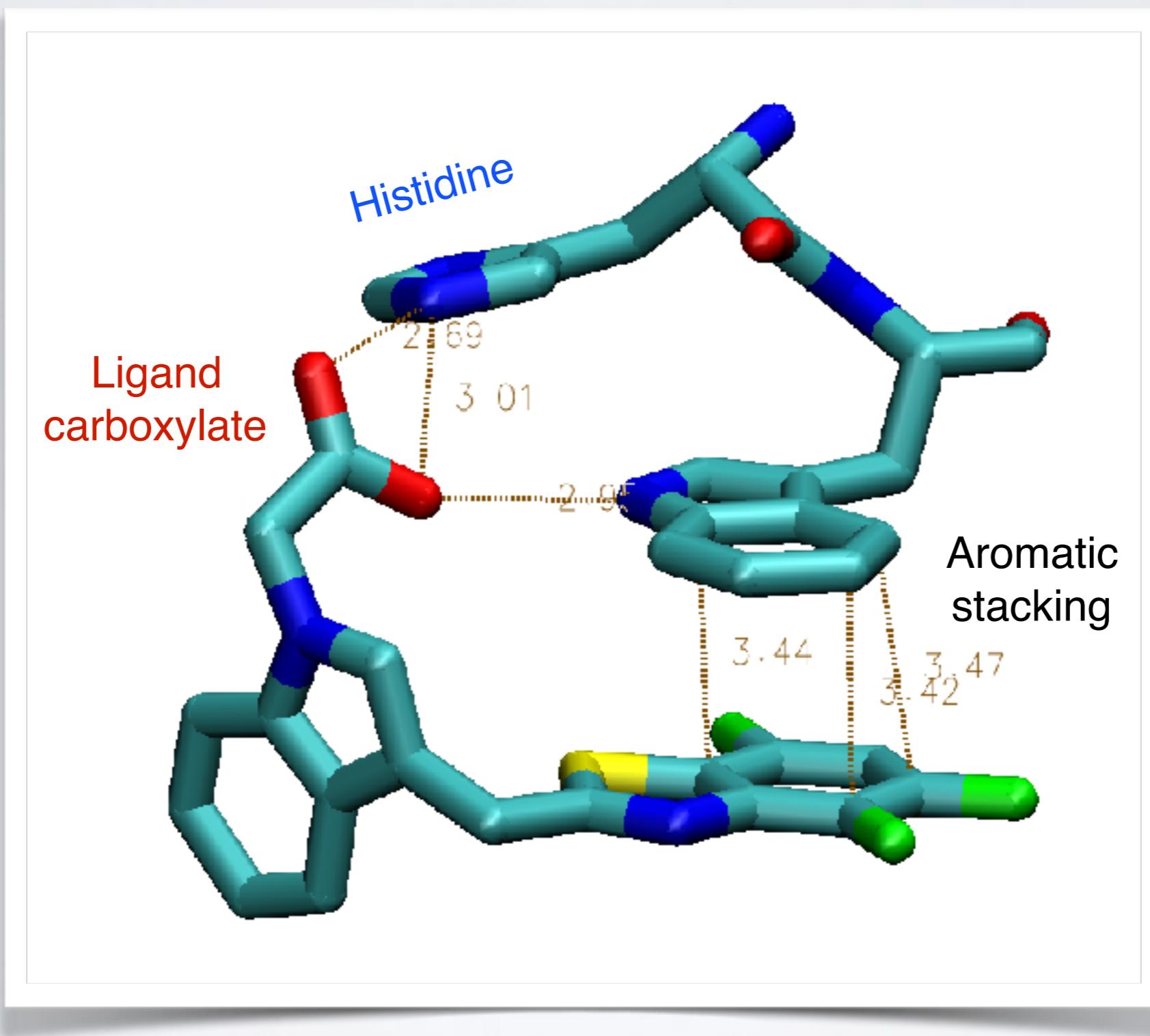


KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

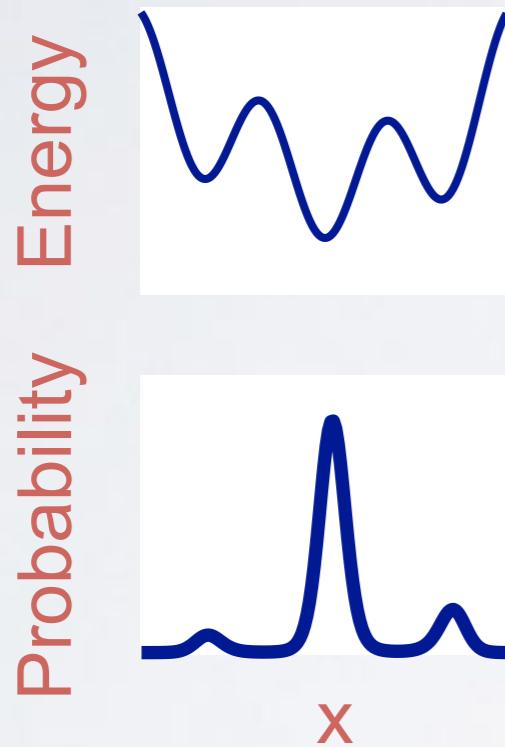
- (1). Physics-Based
- (2). Knowledge-Based

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

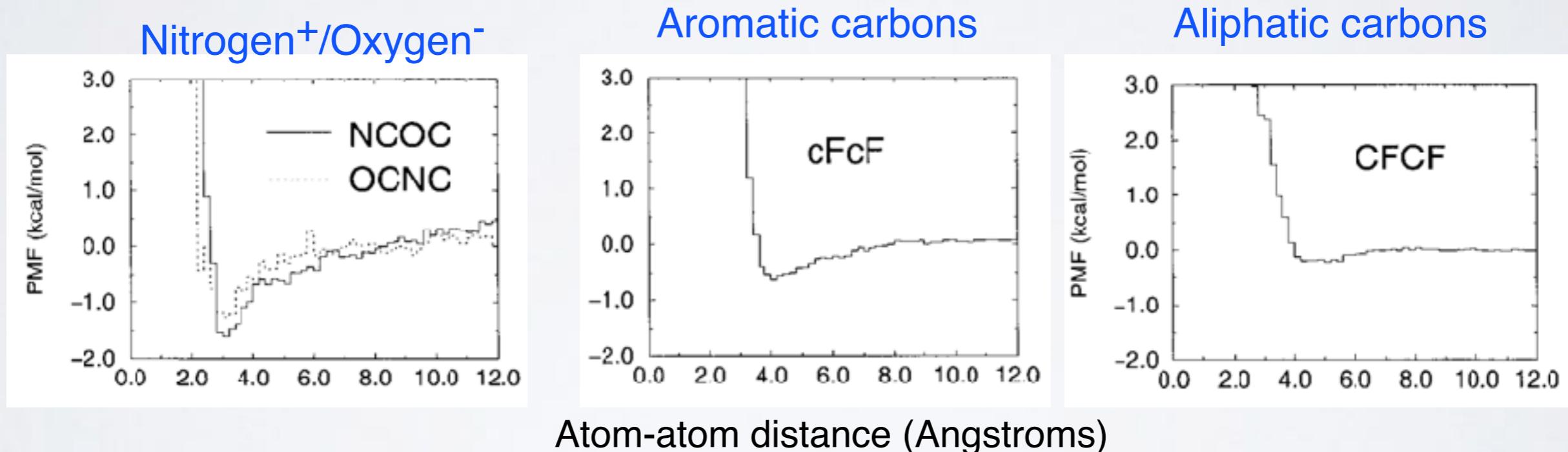
Find all protein-ligand structures in the PDB with a ligand carboxylate O

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)

Do it Yourself!

Hand-on time!

https://bioboot.github.io/bimm143_S18/lectures/#11

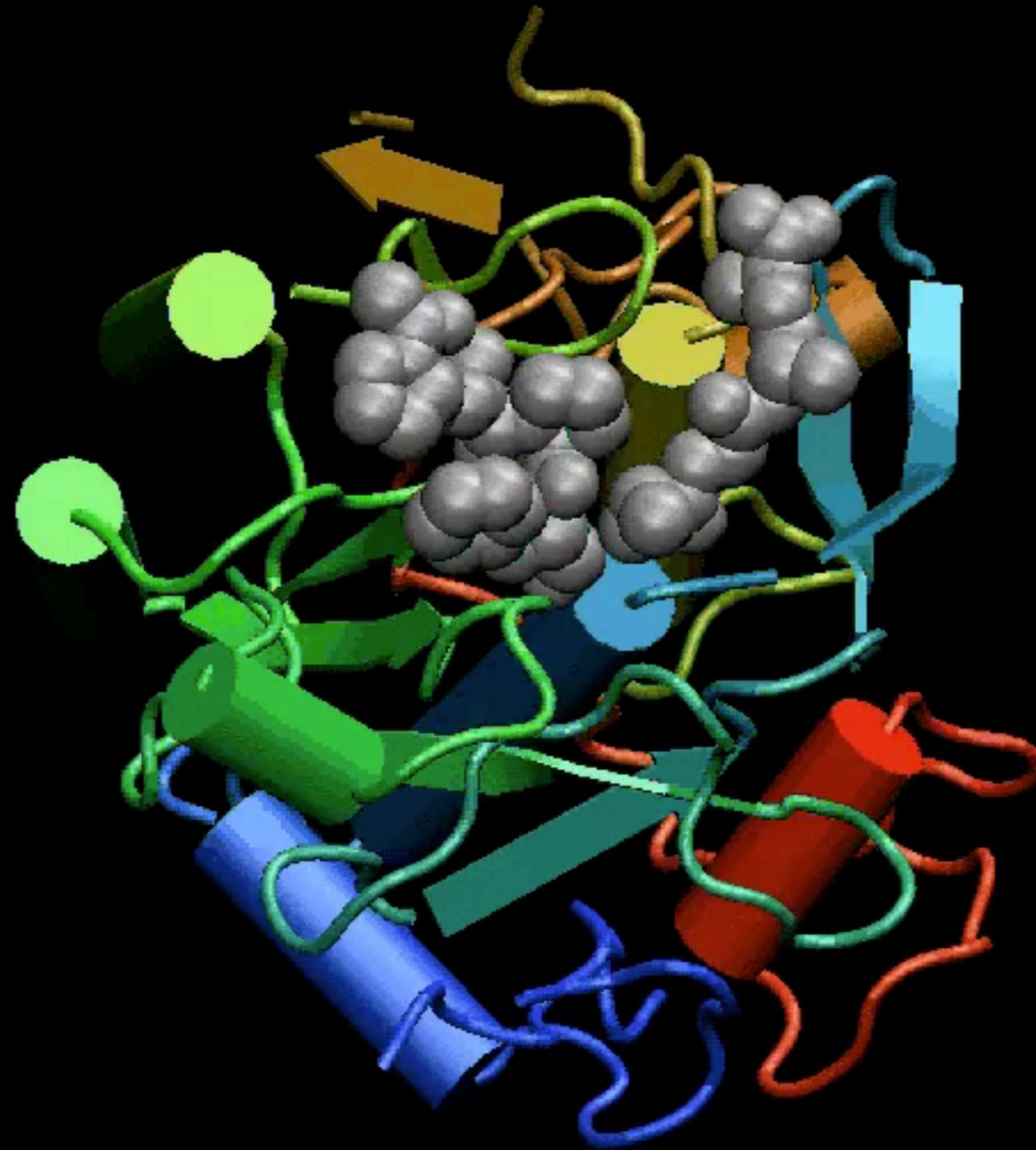
Focus on **section 4 & 5**

NEXT UP:

- ▶ **Overview of structural bioinformatics**
 - Major motivations, goals and challenges
- ▶ **Fundamentals of protein structure**
 - Composition, form, forces and dynamics
- ▶ **Representing and interpreting protein structure**
 - Modeling energy as a function of structure

- ▶ **Example application areas**
 - Predicting functional dynamics & drug discovery

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/#11

Focus on **section 6** to **7**

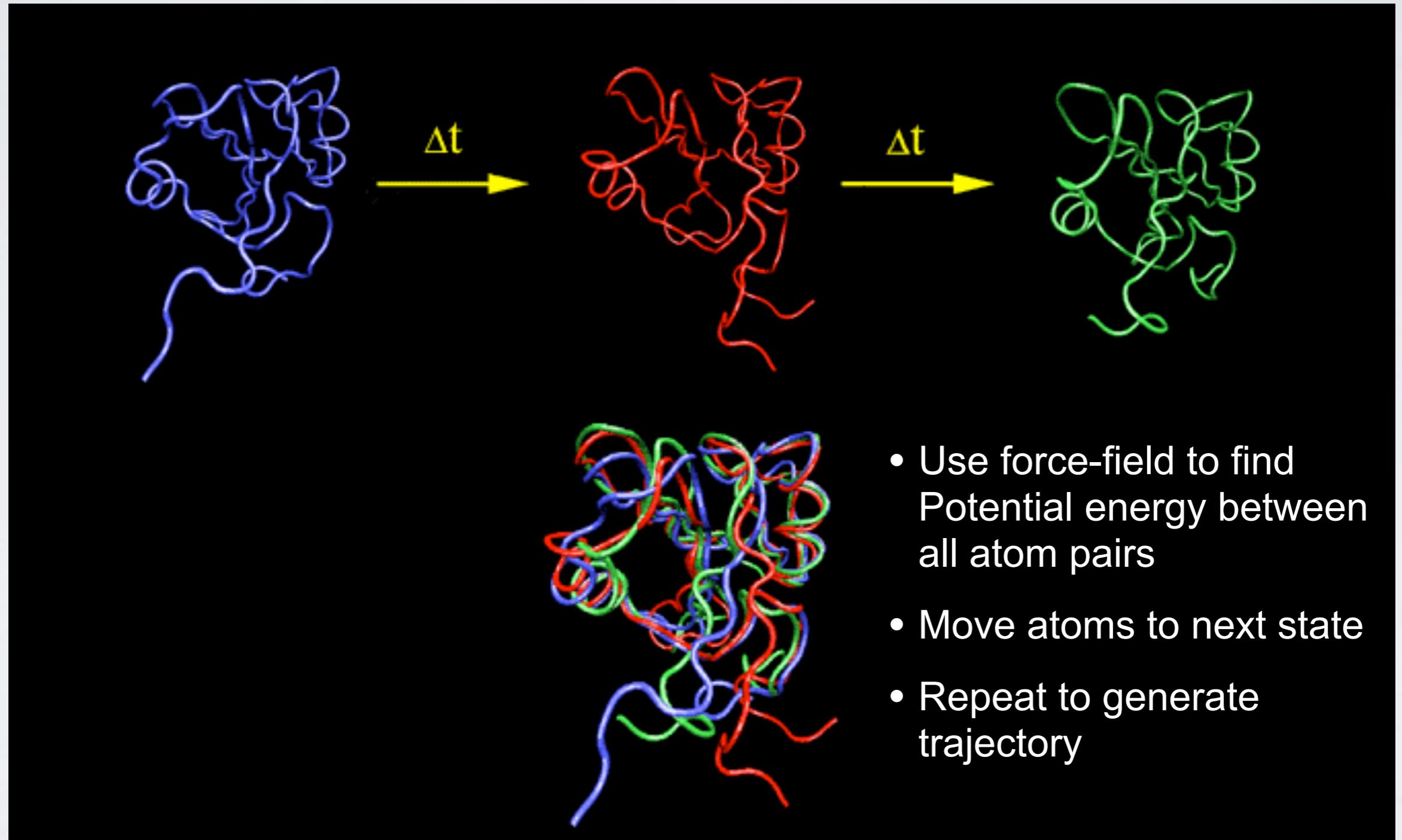
Optional:
Stop here for Today!

[Muddy Point Assessment]

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



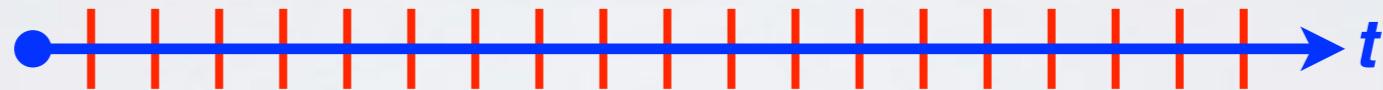
McCammon, Gelin & Karplus, *Nature* (1977)

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

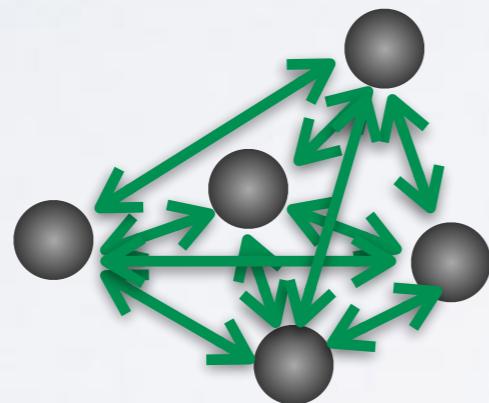
- ▶ Divide **time** into discrete ($\sim 1\text{fs}$) **time steps (Δt)**
(for integrating equations of motion, see below)



- ▶ Divide **time** into discrete ($\sim 1\text{fs}$) **time steps (Δt)**
(for integrating equations of motion, see below)



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



Nucleic motion described classically

$$m_i \frac{d^2}{dt^2} \vec{R}_i = -\vec{\nabla}_i E(\vec{R})$$

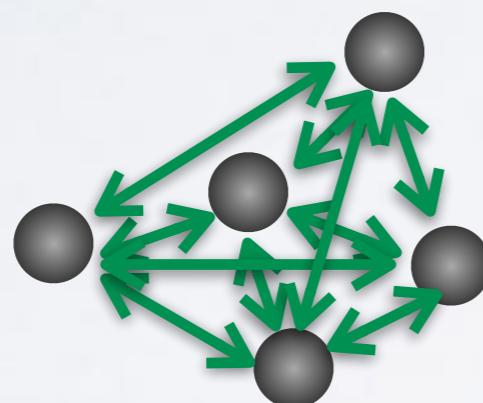
Empirical force field

$$E(\vec{R}) = \sum_{\text{bonded}} E_i(\vec{R}) + \sum_{\text{non-bonded}} E_i(\vec{R})$$

- ▶ Divide **time** into discrete ($\sim 1\text{fs}$) **time steps (Δt)**
(for integrating equations of motion, see below)



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



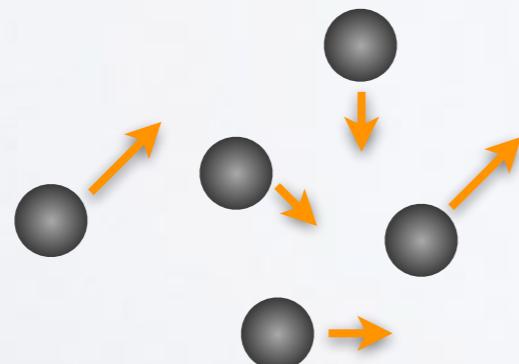
Nucleic motion described classically

$$m_i \frac{d^2}{dt^2} \vec{R}_i = -\vec{\nabla}_i E(\vec{R})$$

Empirical force field

$$E(\vec{R}) = \sum_{\text{bonded}} E_i(\vec{R}) + \sum_{\text{non-bonded}} E_i(\vec{R})$$

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



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

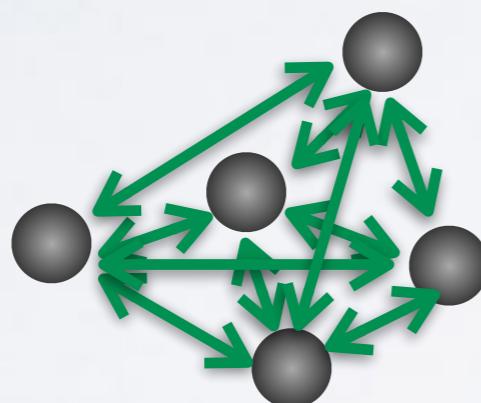
$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \boxed{v(t + \frac{\Delta t}{2})} \Delta t$$

BASIC ANATOMY OF A MD SIMULATION

- ▶ Divide **time** into discrete ($\sim 1\text{fs}$) **time steps** (Δt)
(for integrating equations of motion, see below)



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



Nucleic motion described classically

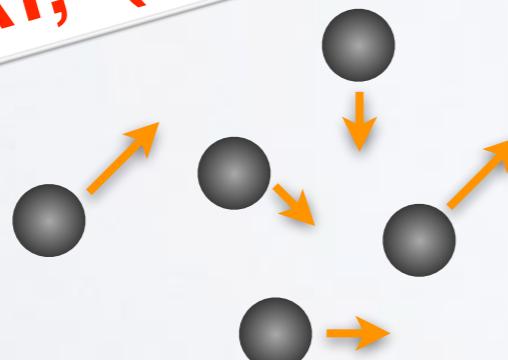
$$m_i \frac{d^2}{dt^2} \vec{R}_i = -\vec{\nabla}_i E(\vec{R})$$

Empirical force function

$$E(\vec{R}) = \sum_{i=1}^N \sum_{j \neq i, \text{non-bonded}} E_i(\vec{R})$$

- ▶ Use the forces to calculate **velocities** and move atoms to new **positions**
(numerically via the “leapfrog” scheme)

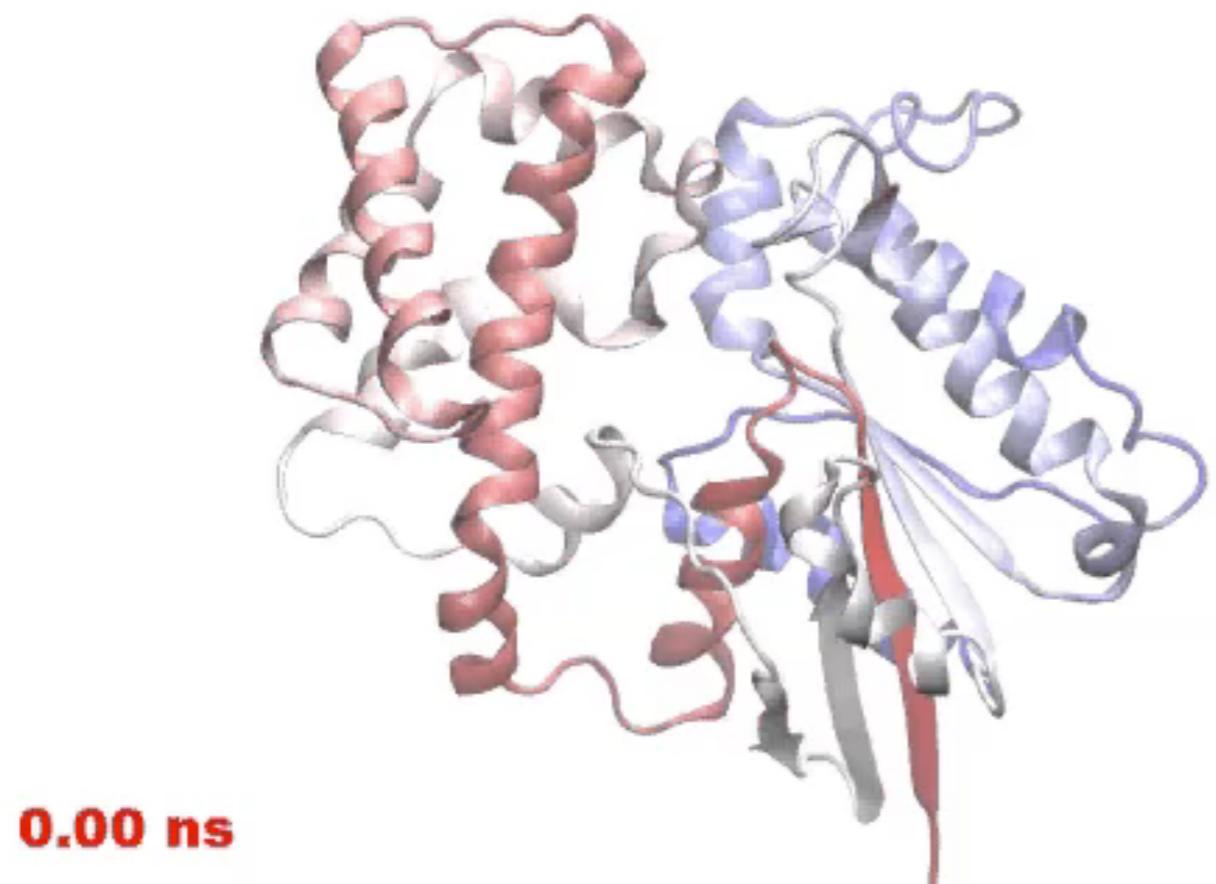
REPEAT, (iterate many, many times... 1ms = 10^{12} time steps)



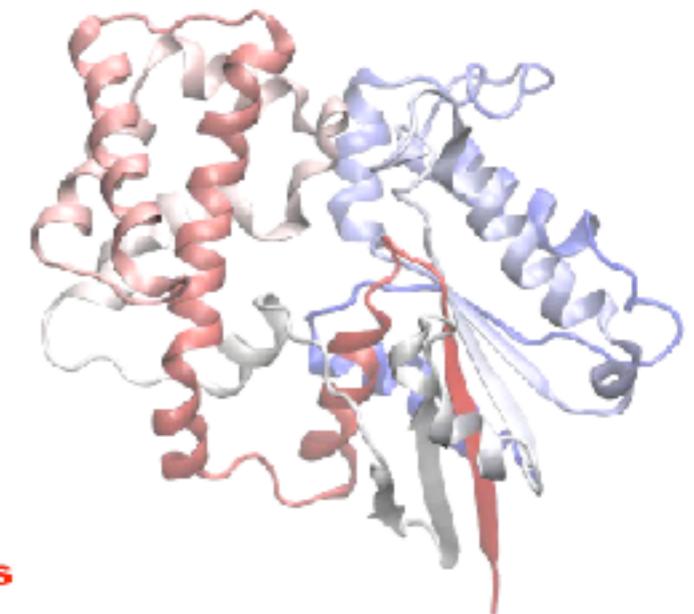
$$\begin{aligned} \mathbf{v}(t + \frac{\Delta t}{2}) &= \mathbf{v}(t - \frac{\Delta t}{2}) + \frac{\mathbf{F}(t)}{m} \Delta t \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t + \frac{\Delta t}{2}) \Delta t \end{aligned}$$

MD Prediction of Functional Motions

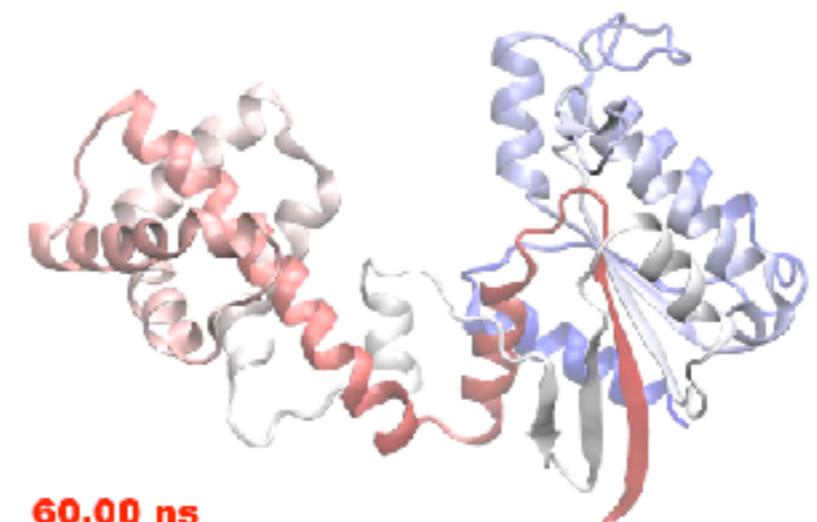
Accelerated MD simulation of
nucleotide-free transducin alpha subunit



“close”

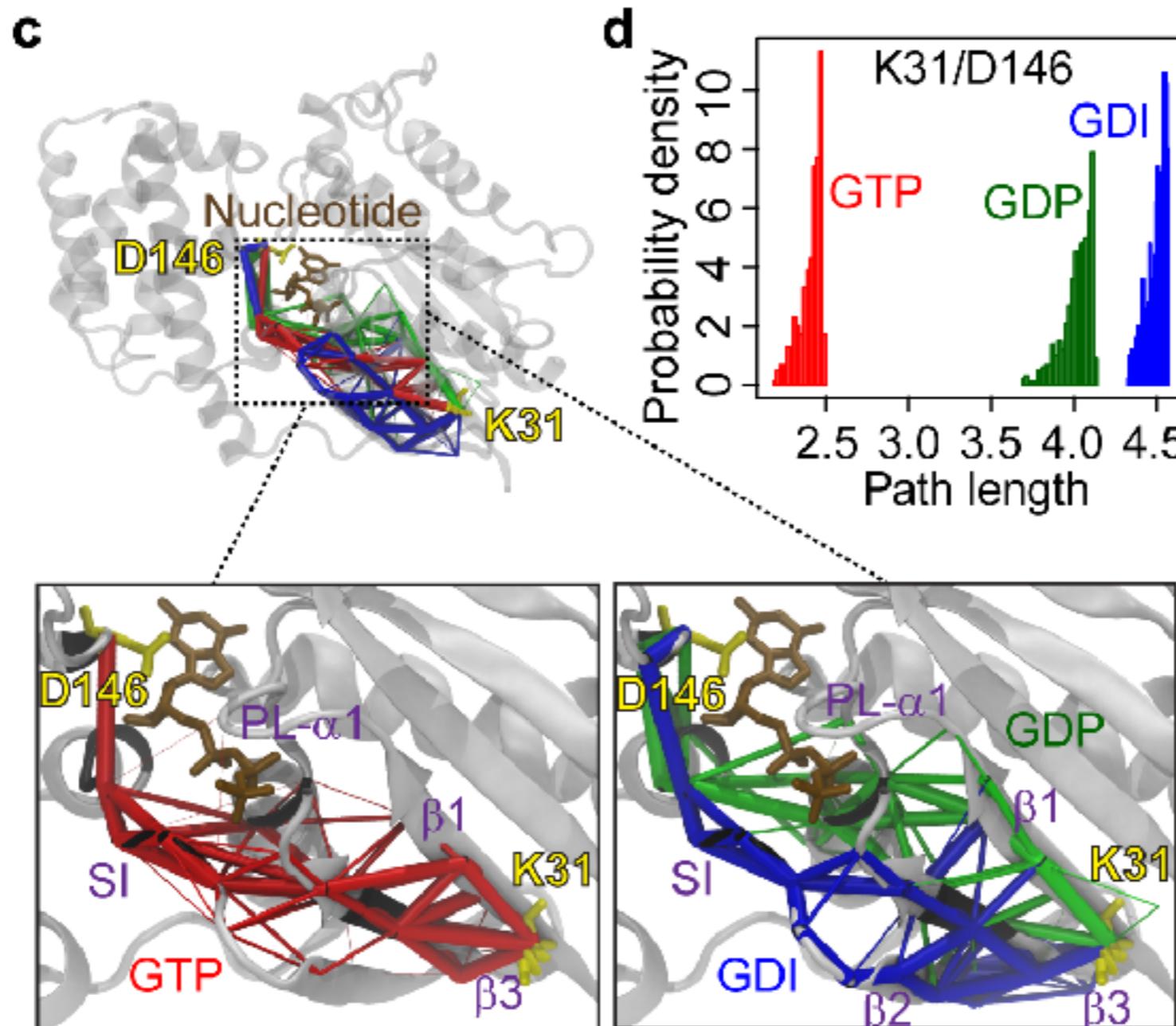


“open”

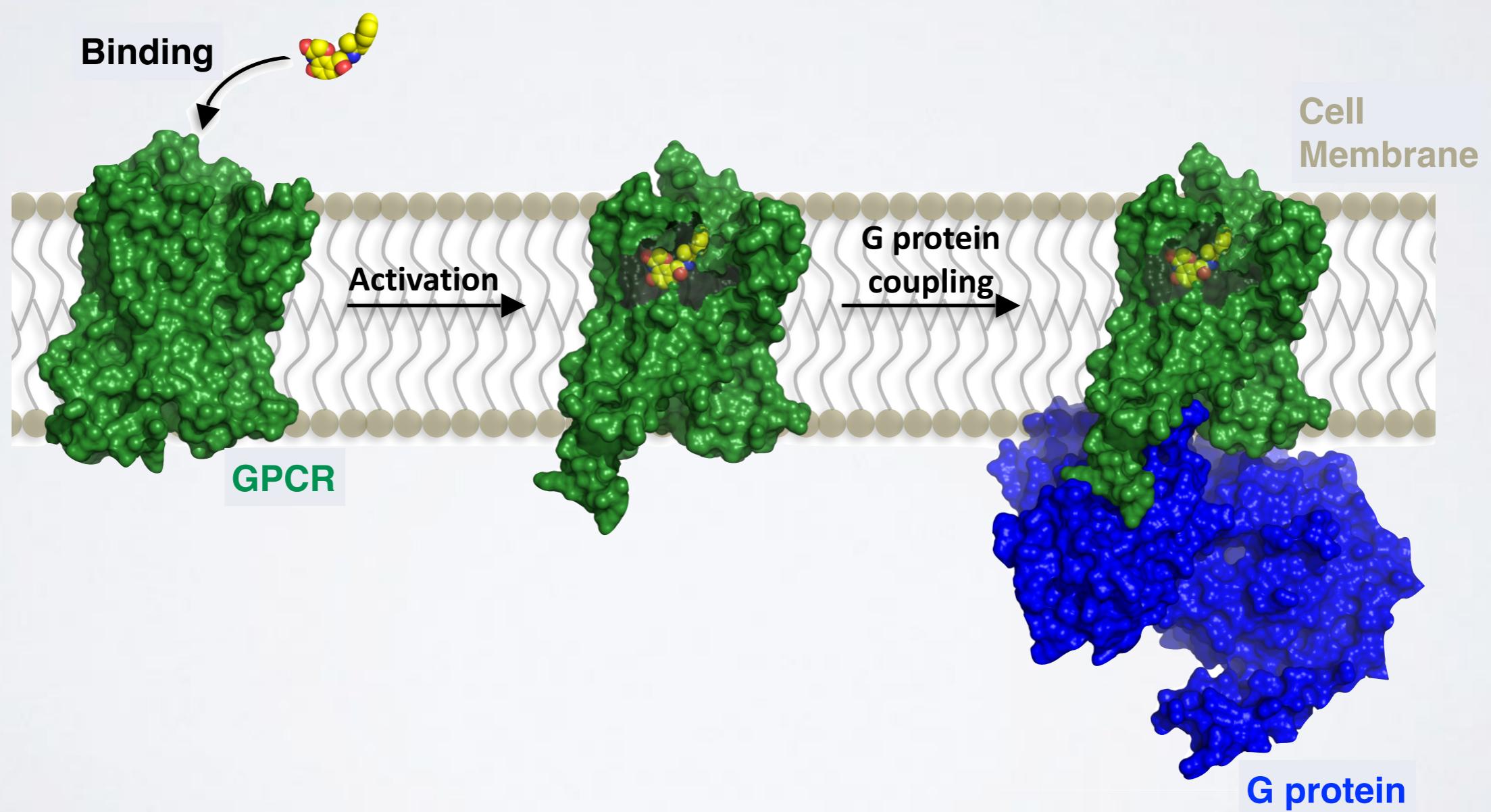


Yao and Grant, Biophys J. (2013)

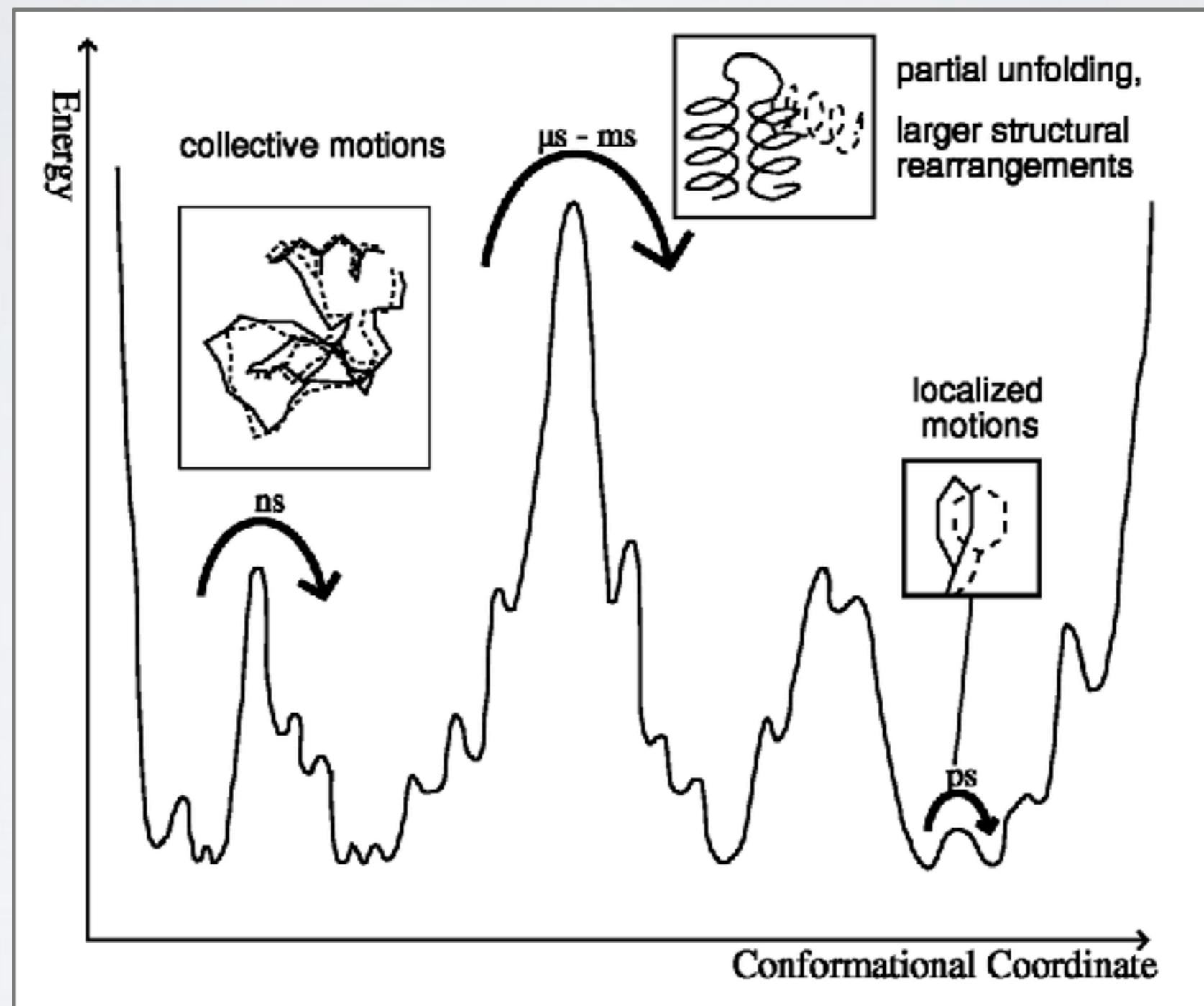
Simulations Identify Key Residues Mediating Dynamic Activation



EXAMPLE APPLICATION OF MOLECULAR SIMULATIONS TO GPCRS



PROTEINS JUMP BETWEEN MANY, HIERARCHICALLY ORDERED “CONFORMATIONAL SUBSTATES”



H. Frauenfelder et al., *Science* **229** (1985) 337

MOLECULAR DYNAMICS IS VERY

Example: F₁-ATPase in water (183,674 atoms) for 1 nanosecond:

- => 10⁶ integration steps
- => 8.4 * 10¹¹ floating point operations/step
[n(n-1)/2 interactions]

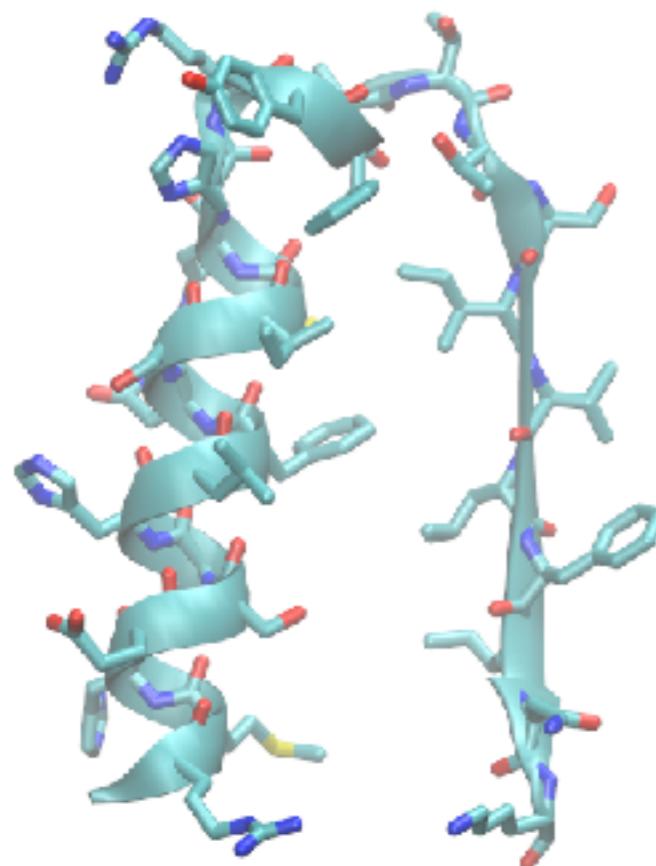
Total: 8.4 * 10¹⁷ flop
(on a 100 Gflop/s cpu: **ca 25 years!**)

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

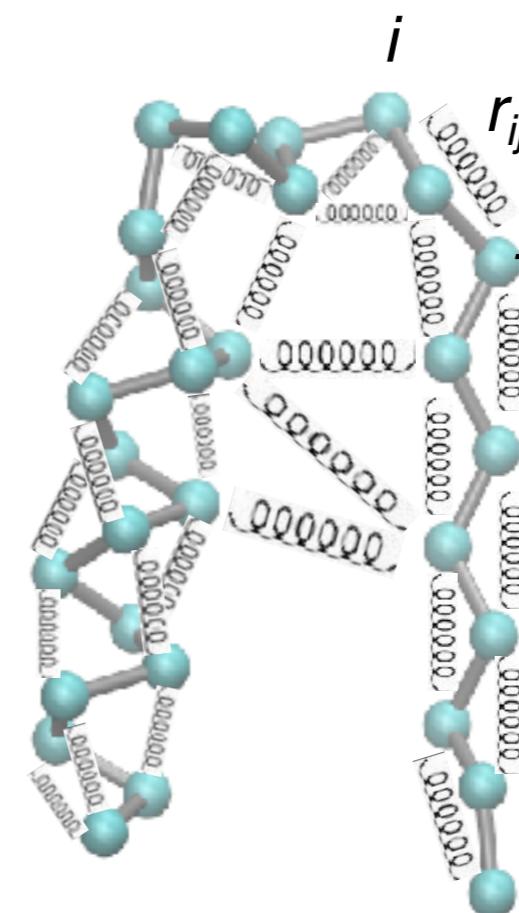
multiple time stepping	ca. 2.5 years
fast multipole methods	ca. 1 year
parallel computers	ca. 5 days
modern GPUs	ca. 1 day
(Anton supercomputer	ca. minutes)

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

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



C. G.

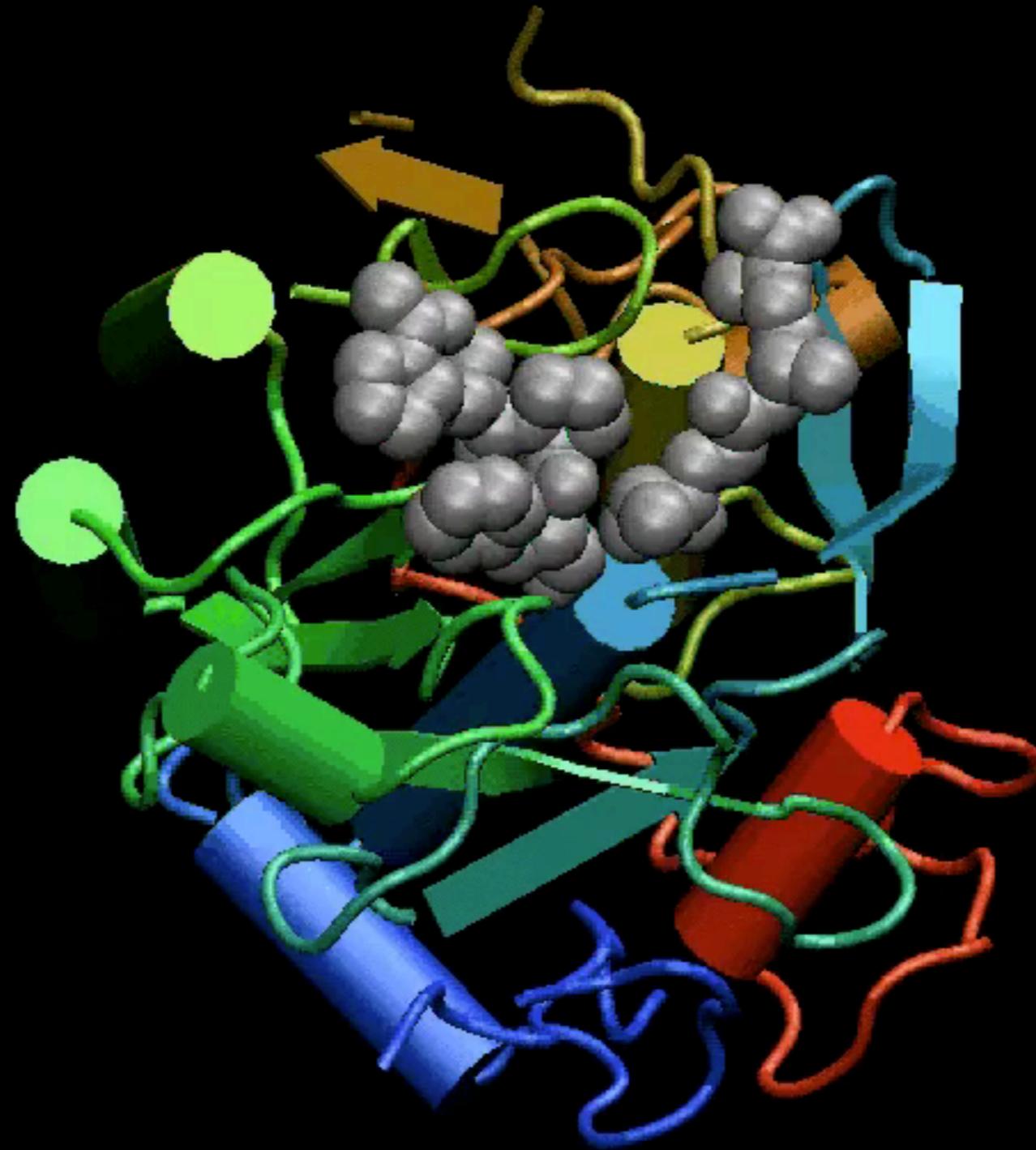


Atomistic

Coarse Grained

- 1 bead /
1 amino acid
- Connected by
springs

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/#11

Focus on **section 6** to **7**

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