BGGN 213

Structural Bioinformatics

Lecture 11

Barry Grant

UC San Diego

http://thegrantlab.org/bggn213

"*Bioinformatics is the application of <u>computers</u> to the collection, archiving, organization, and analysis of <u>biological data</u>.*"

… A hybrid of biology and computer science

"Bioinformatics is the application of <u>computers</u> to the collection, archiving, organization, and analysis of <u>biological data</u>."

Bioinformatics is computer aided biology!

"*Bioinformatics is the application of <u>computers</u> to the collection, archiving, organization, and analysis of <u>biological data</u>.*"

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 assembles at the molecular & atomic level
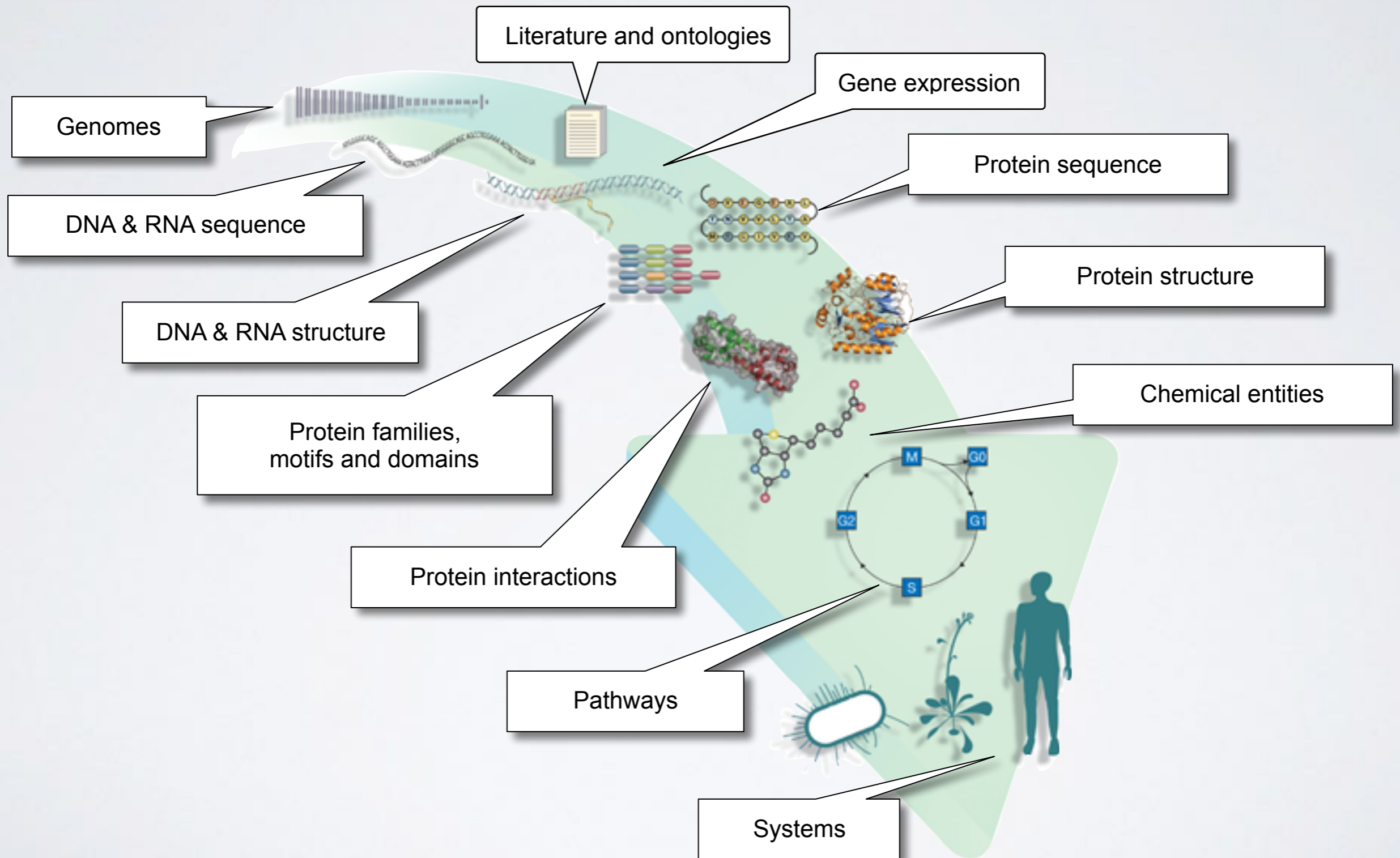
Why should we care?

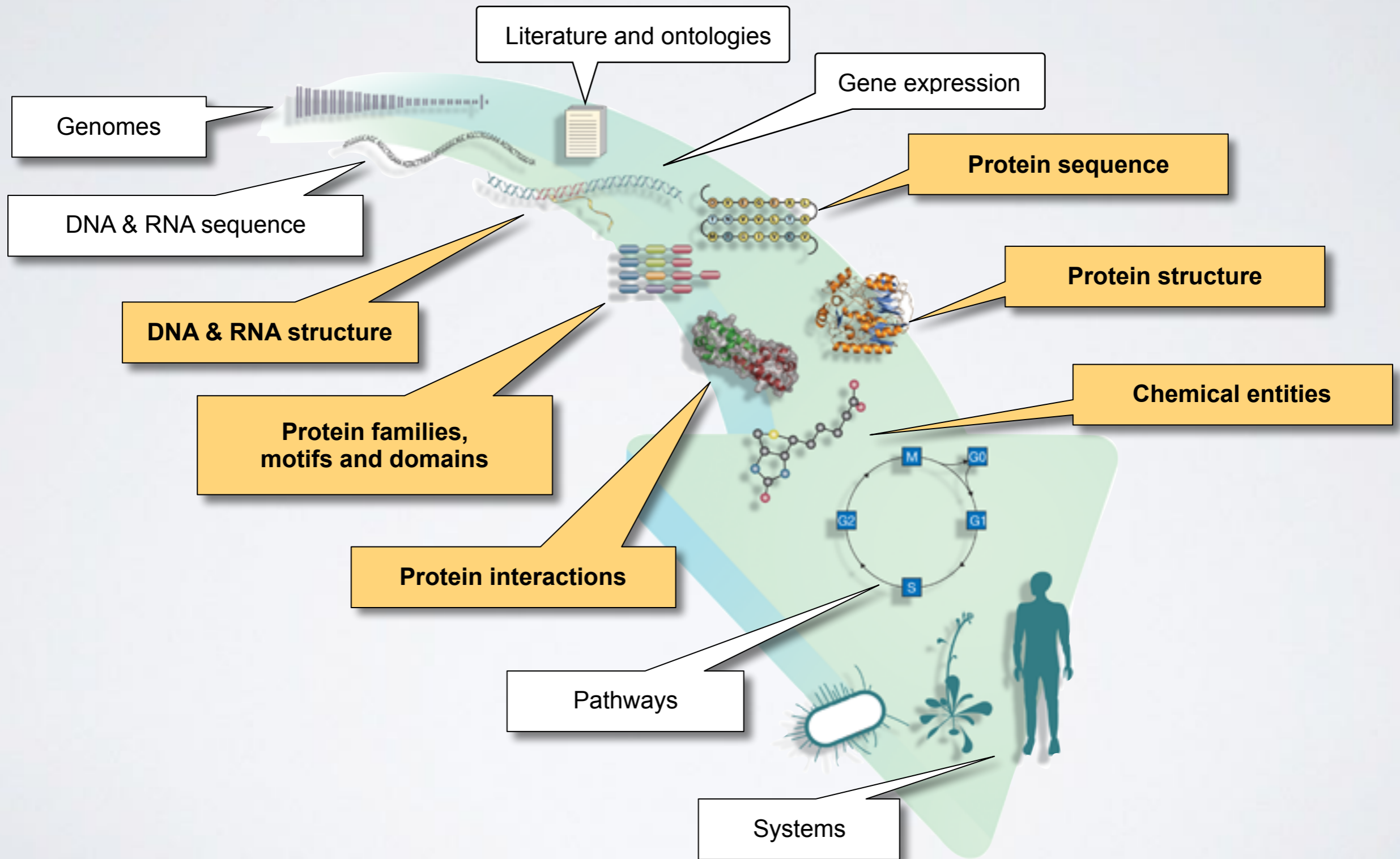# Why should we care?

Because biomolecules are "nature's robots"

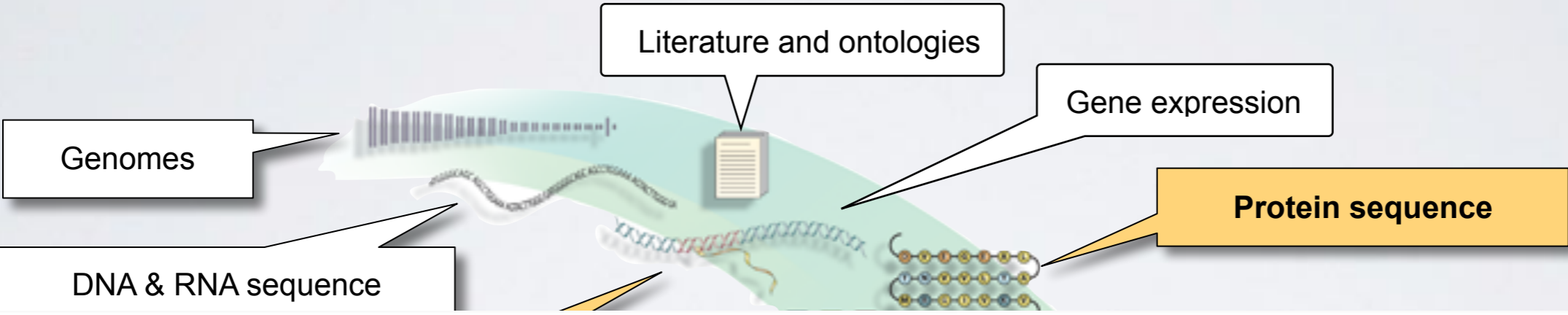… and because it is only by coiling into **specific 3D structures** that they are able to perform their functions

BIOINFORMATICS DATA

Literature and ontologies

Gene expression

Genomes

Protein sequence

DNA & RNA sequence

Protein structure

DNA & RNA structure

Protein families, motifs and domains

Chemical entities

Protein interactions

Pathways

Systems

STRUCTURAL DATA IS CENTRAL

Genomes

Literature and ontologies

Gene expression

DNA & RNA sequence

Protein sequence

DNA & RNA structure

Protein structure

Protein families, motifs and domains

Chemical entities

Protein interactions

Pathways

Systems

STRUCTURAL DATA IS CENTRAL

Genomes

DNA & RNA sequence

Literature and ontologies

Gene expression

Protein sequence

Protein structure

DNA & RNA structure

Sequence > Structure > Function

Chemical entities

Protein families, motifs and domains

Protein interactions

Pathways

Systems

STRUCTURAL DATA IS CENTRAL

Literature and ontologies

Gene expression

Genomes

DNA & RNA sequence

Protein sequence

ENERGETICS

DYNAMICS

Protein structure

DNA & RNA structure

Sequence ∧ Structure ∧ Function

Chemical entities

Protein families, motifs and domains

Protein interactions

Pathways

Systems
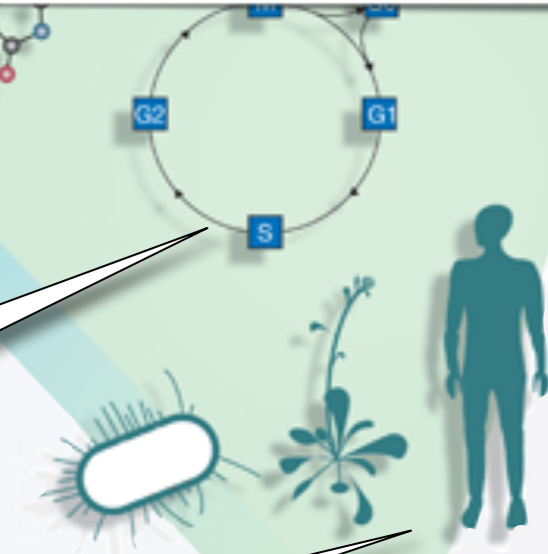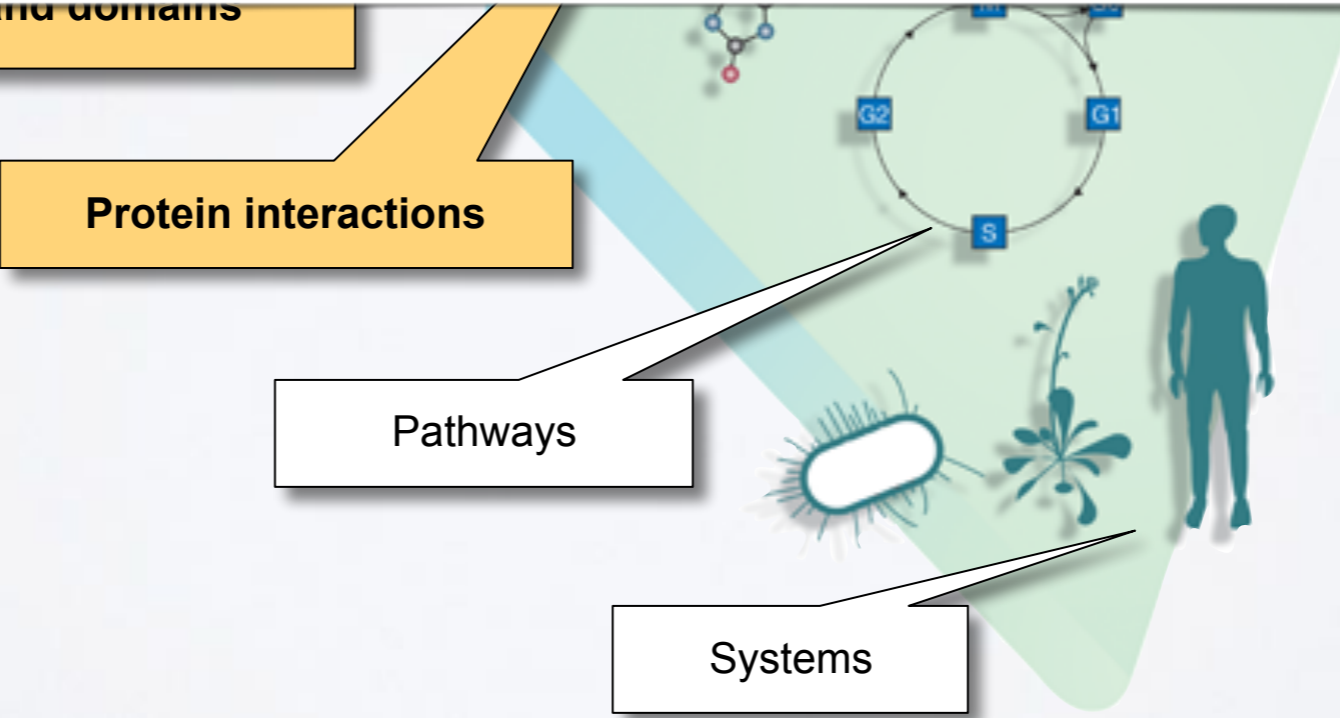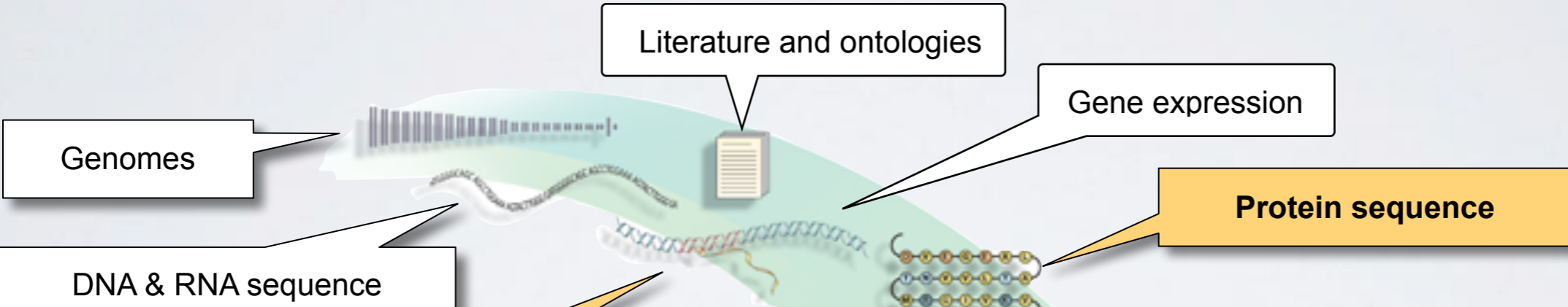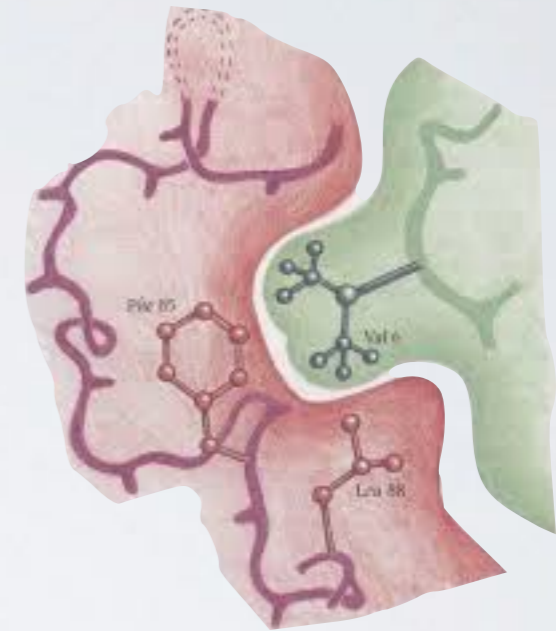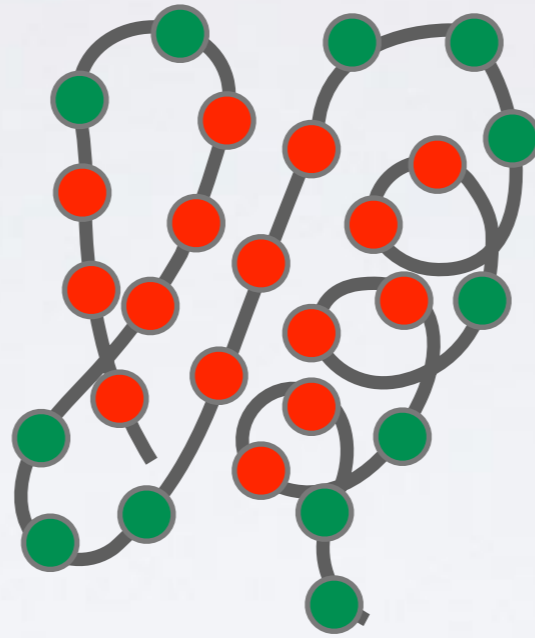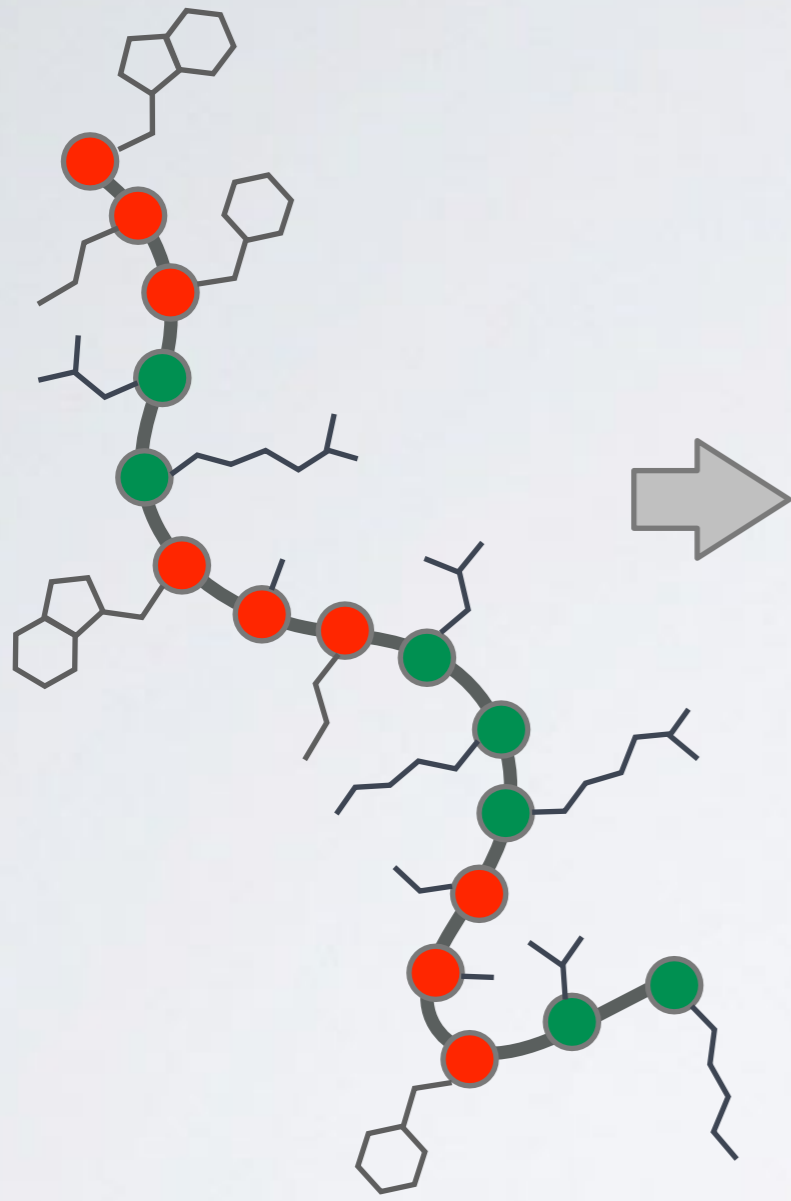
Sequence	Structure	Function
• Unfolded chain of amino acid chain • Highly mobile • Inactive	• Ordered in a precise 3D arrangment • Stable but dynamic	• Active in specific "conformations" • Specific associations & precise reactions

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
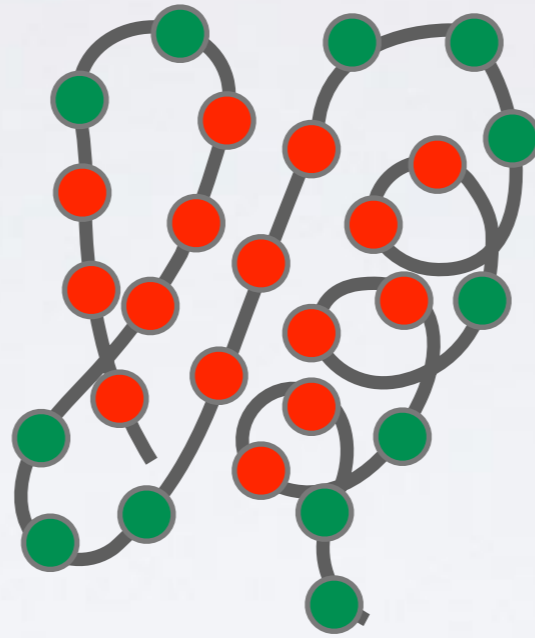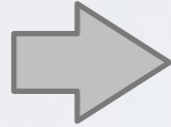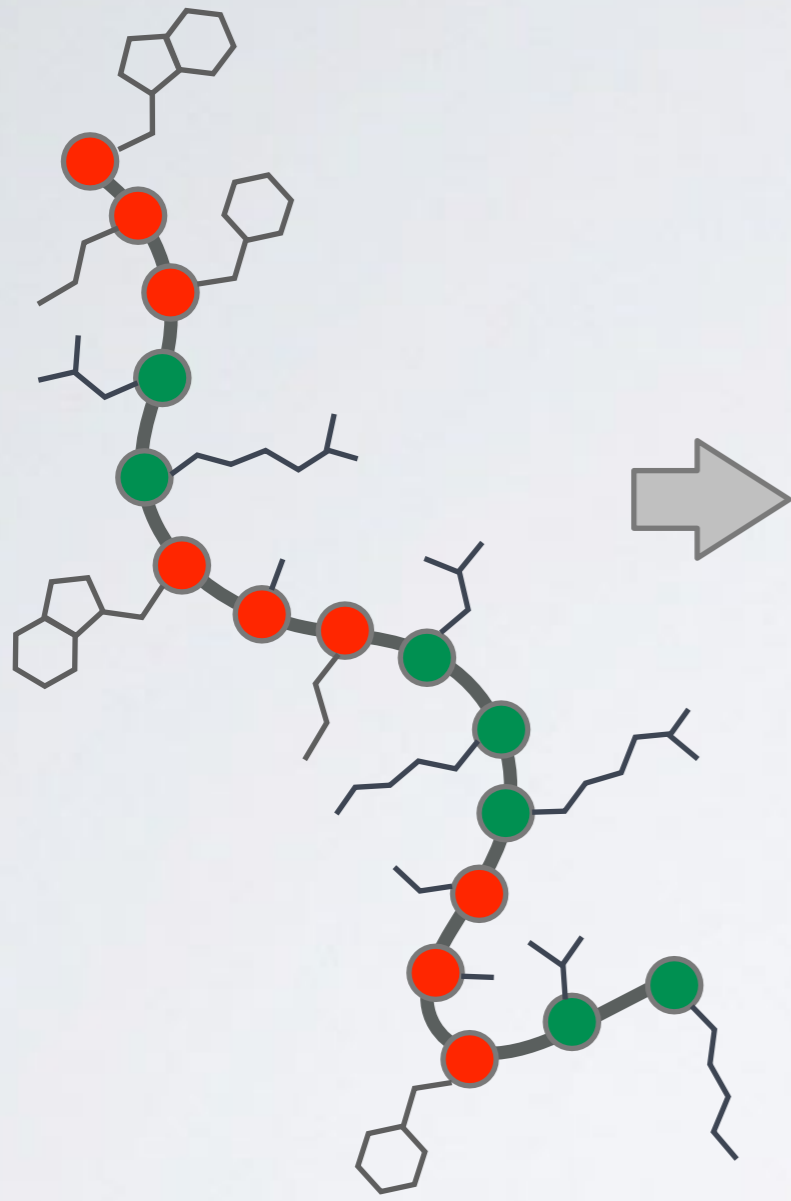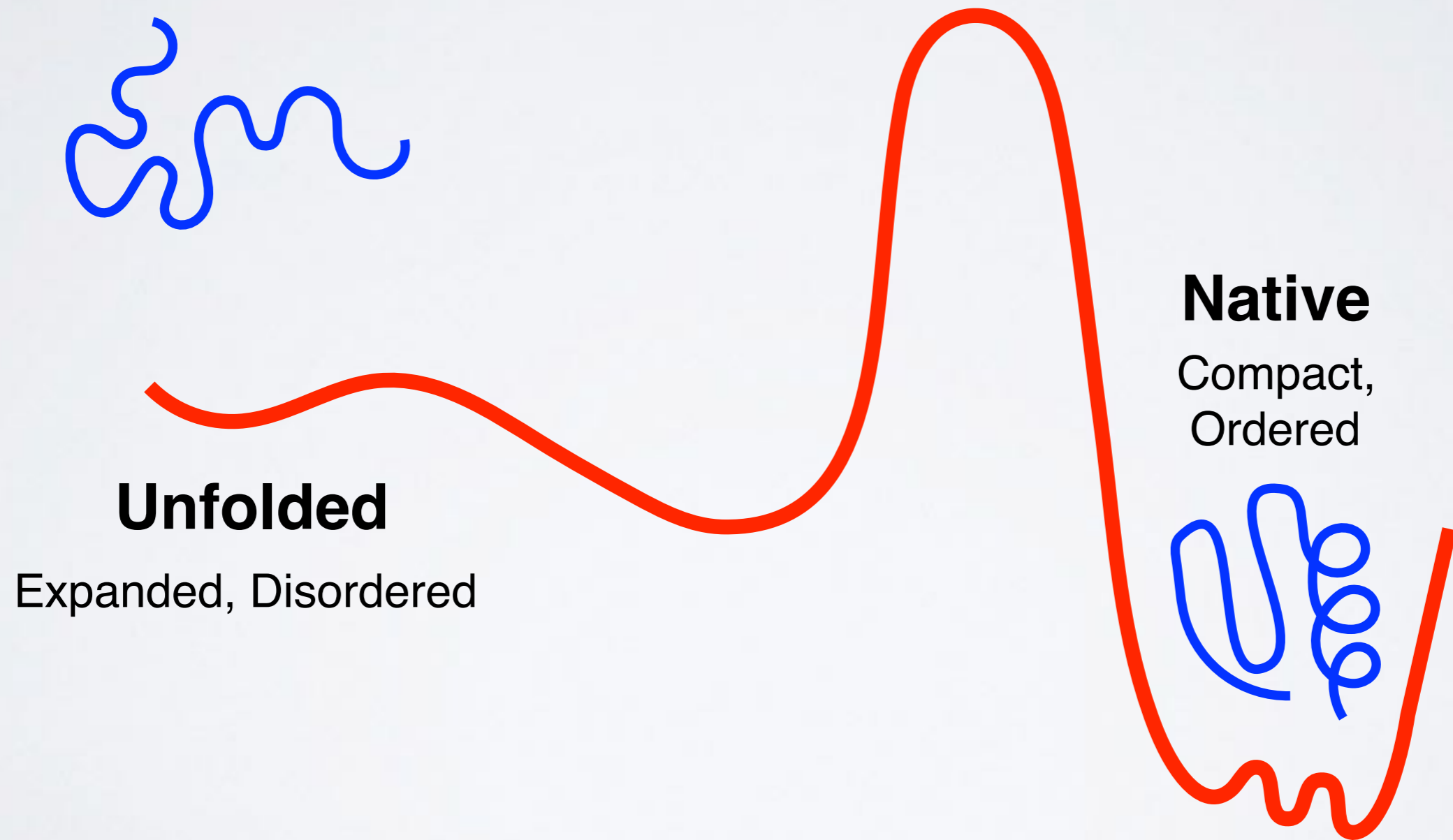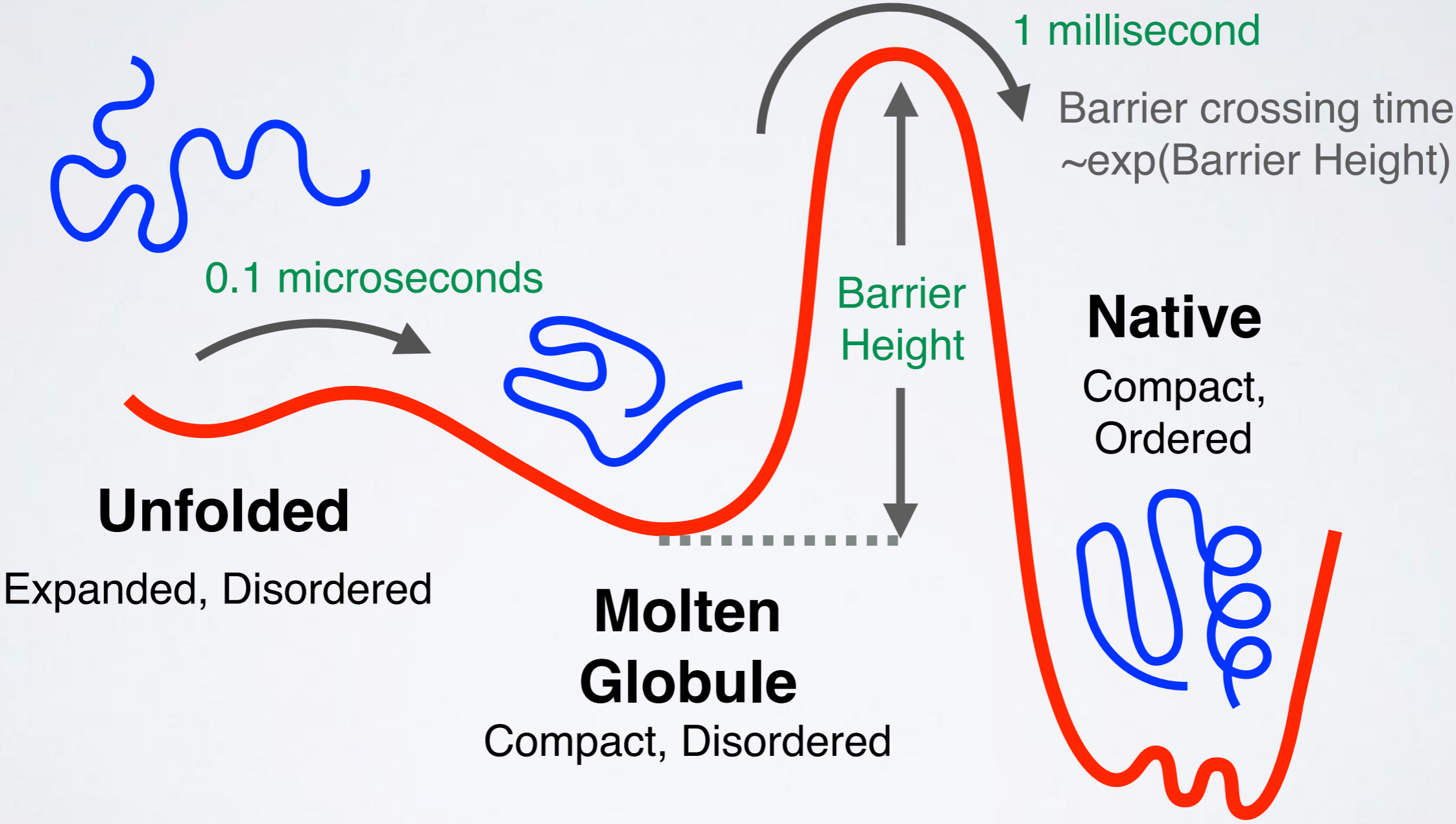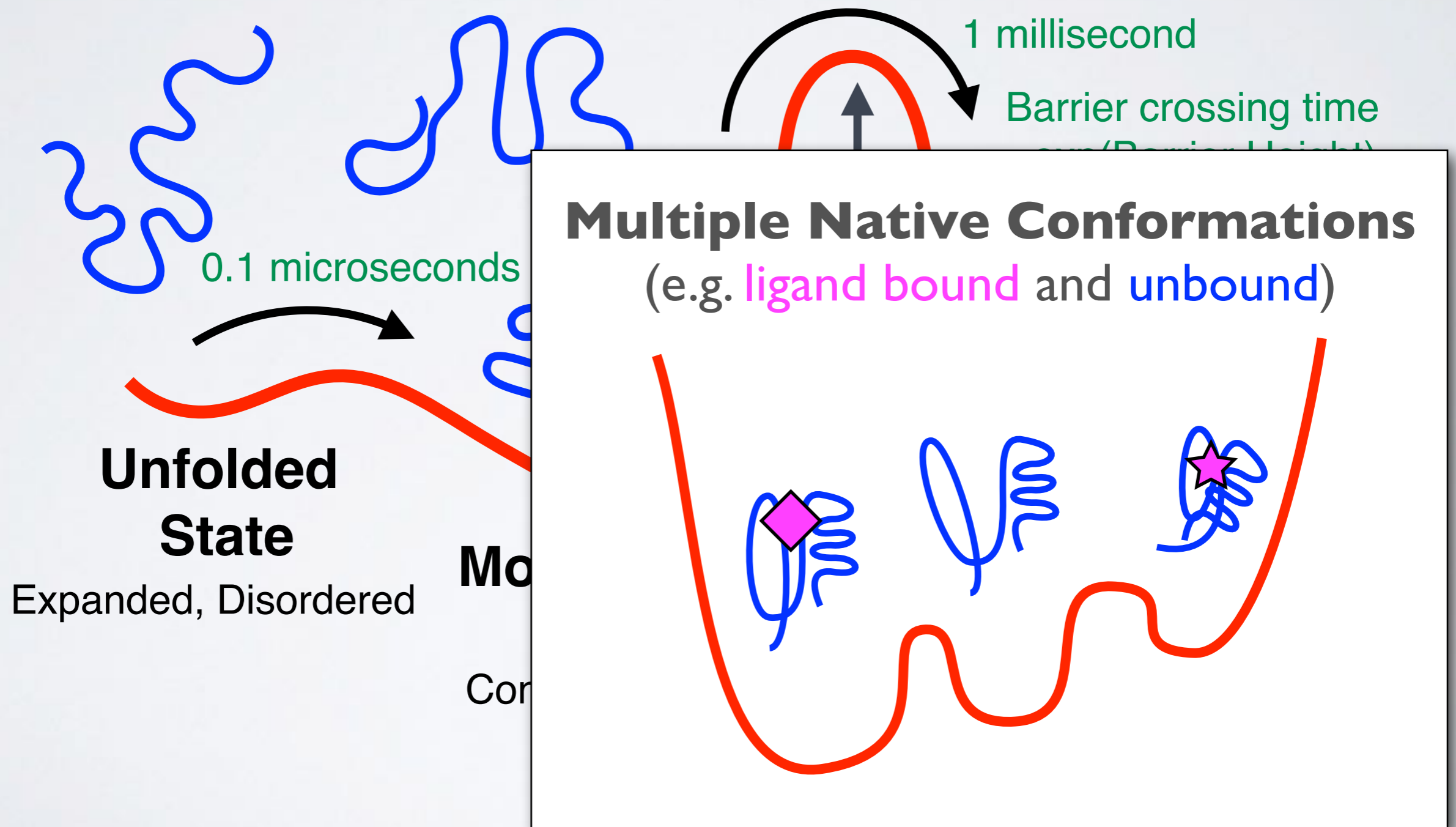
… but not the end
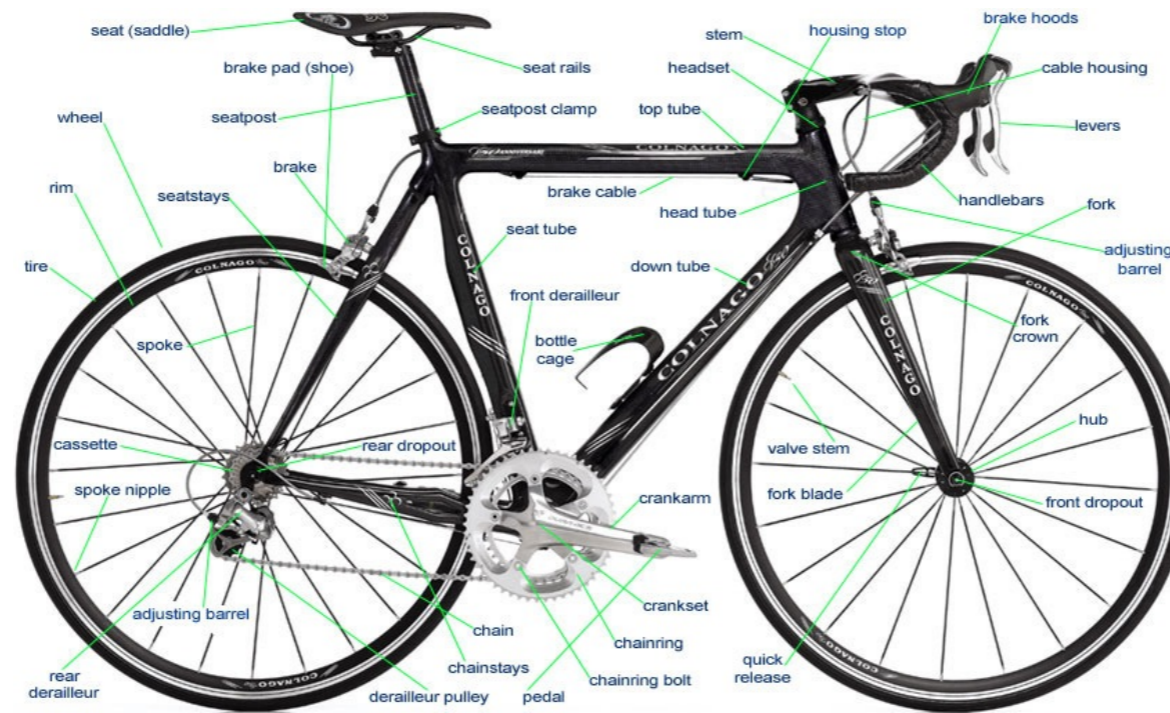
- We want the full spatiotemporal picture, and an ability to control it
- Broad applications, including drug design, medical diagnostics, chemical manufacturing, and energy
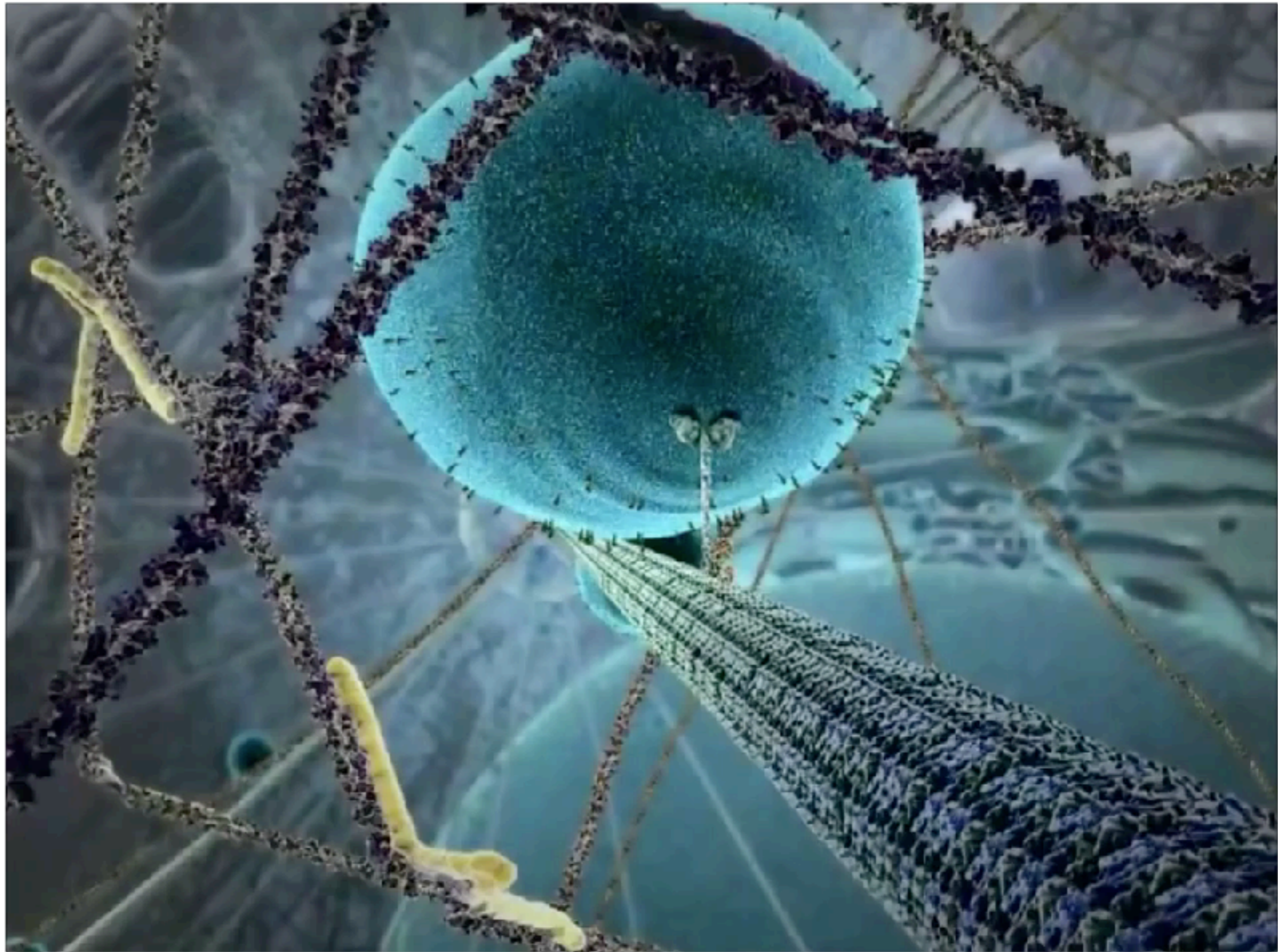
Extracted from The Inner Life of a Cell by Cellular Visions and Harvard
[YouTube link: https://www.youtube.com/watch?v=y-uuk4Pr2i8]
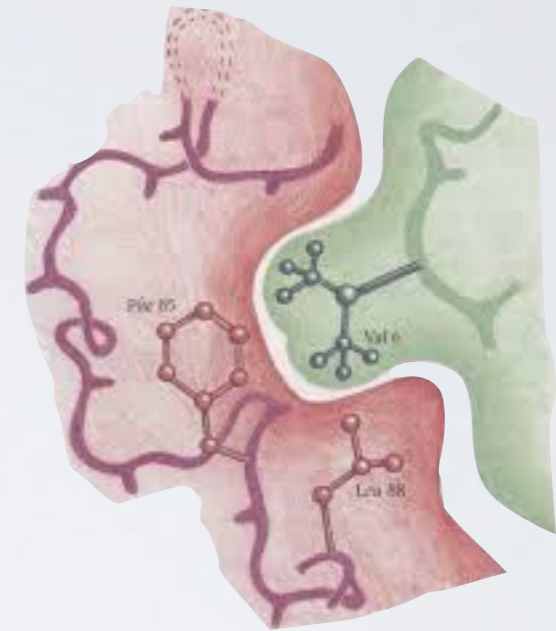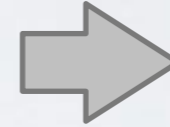
Sequence	**Structure**	**Function**
• Unfolded chain of amino acid chain • Highly mobile • Inactive	• Ordered in a precise 3D arrangment • Stable but dynamic	• Active in specific "conformations" • Specific associations & precise reactions

KEY CONCEPT: ENERGY LANDSCAPE

Unfolded

Expanded, Disordered

Native

Compact,
Ordered

KEY CONCEPT: ENERGY LANDSCAPE

0.1 microseconds

1 millisecond

Barrier crossing time
~exp(Barrier Height)

Barrier Height

Unfolded

Expanded, Disordered

Molten Globule

Compact, Disordered

Native

Compact, Ordered

KEY CONCEPT: **ENERGY LANDSCAPE**

1 millisecond

Barrier crossing time

0.1 microseconds

Unfolded State

Expanded, Disordered

Mo

Co

Multiple Native Conformations
(e.g. ligand bound and unbound)

Today's Menu

- **Overview of structural bioinformatics**
  - Motivations, goals and challenges

- **Fundamentals of protein structure**
  - Structure composition, form and forces

- **Representing, interpreting & modeling protein structure**
  - Visualizing & interpreting protein structures
  - Analyzing protein structures
  - Modeling energy as a function of structure

Today's Menu

- **Overview of structural bioinformatics**
  - Motivations, goals and challenges

- **Fundamentals of protein structure**
  - Structure composition, form and forces

- **Representing, interpreting & modeling protein structure**
  - Visualizing & interpreting protein structures
  - Analyzing protein structures
  - Modeling energy as a function of structure

TRADITIONAL FOCUS **PROTEIN**, **DNA** AND **SMALL MOLECULE** DATA SETS WITH **MOLECULAR STRUCTURE**
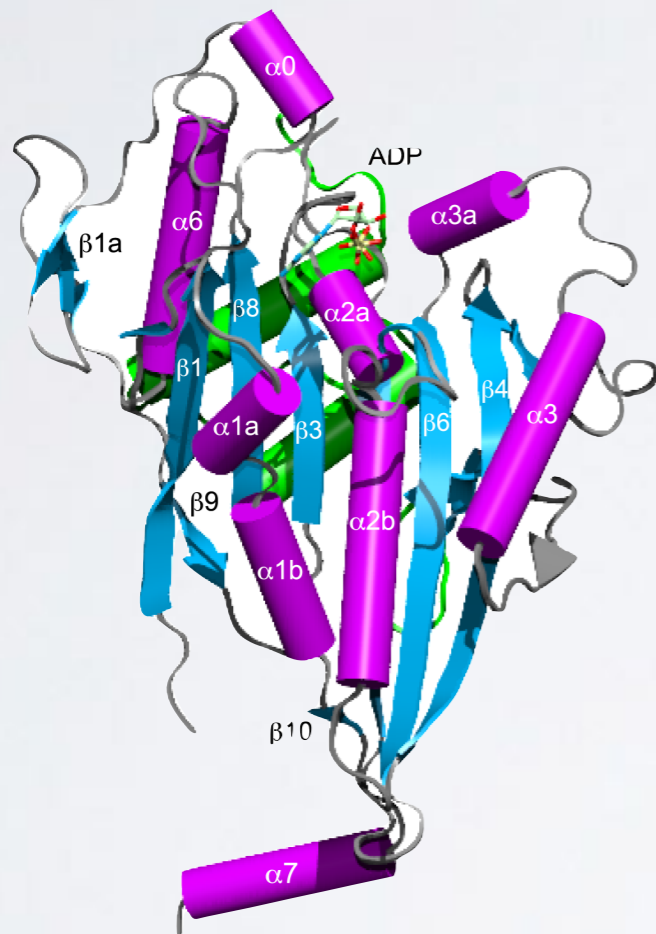
Protein
(PDB)

DNA
(NDB)
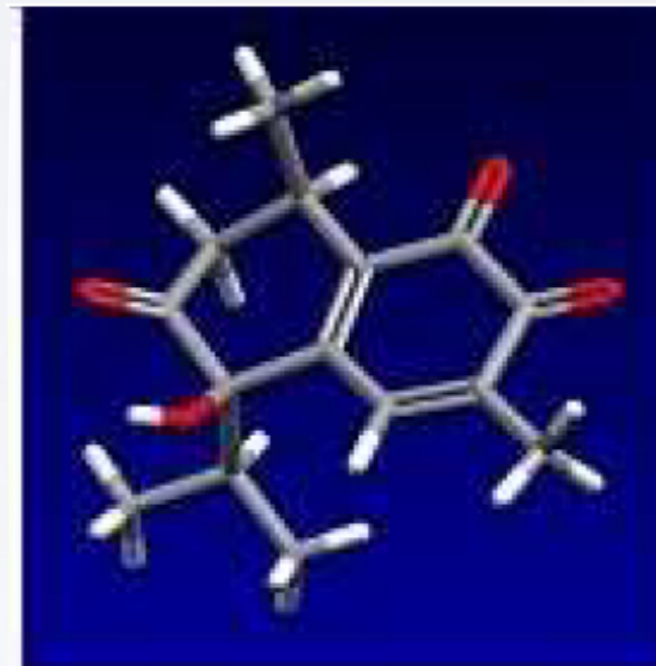
Small Molecules
(CCDB)

PDB – A Billion Atom Archive
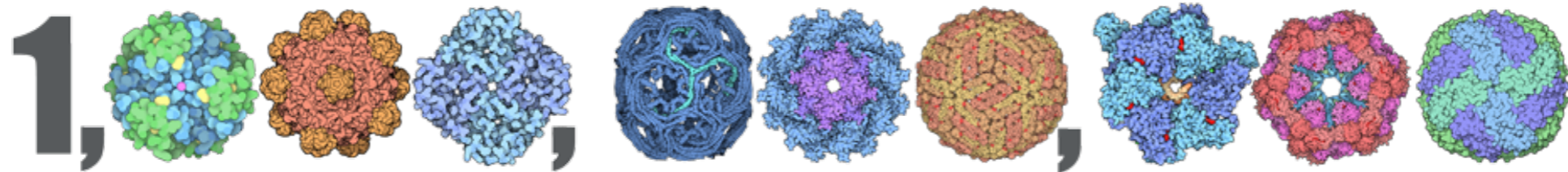
> 1 billion atoms in the asymmetric units
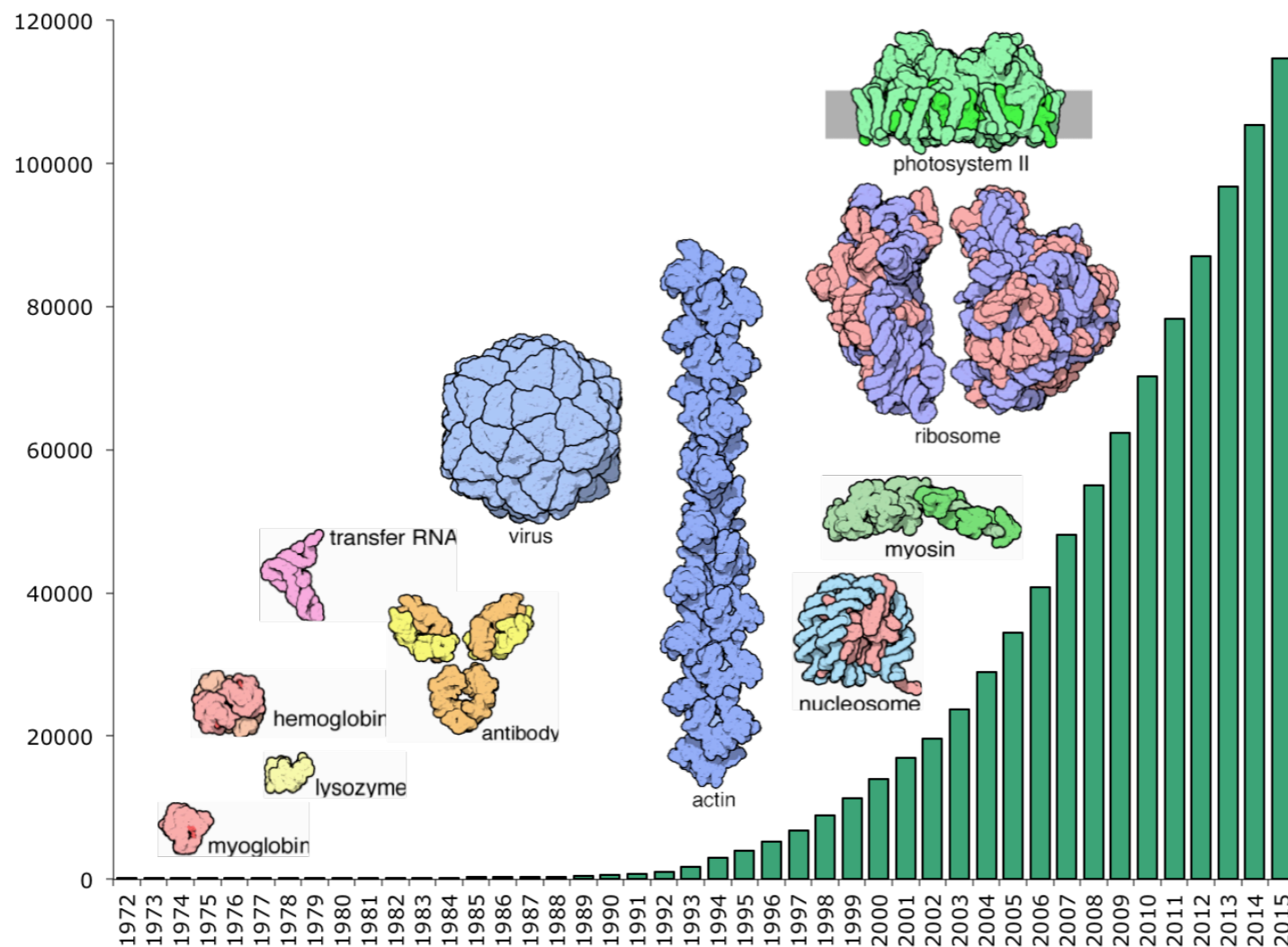
~140,000 Structures in May 2018

Growing Structure Size and Complexity

1Å   1nm   10nm   100nm   1µm   10µm   0.1mm

Optical resolution limit

In situ Structural Biology

PDB

chr22 unpacked: 10mm
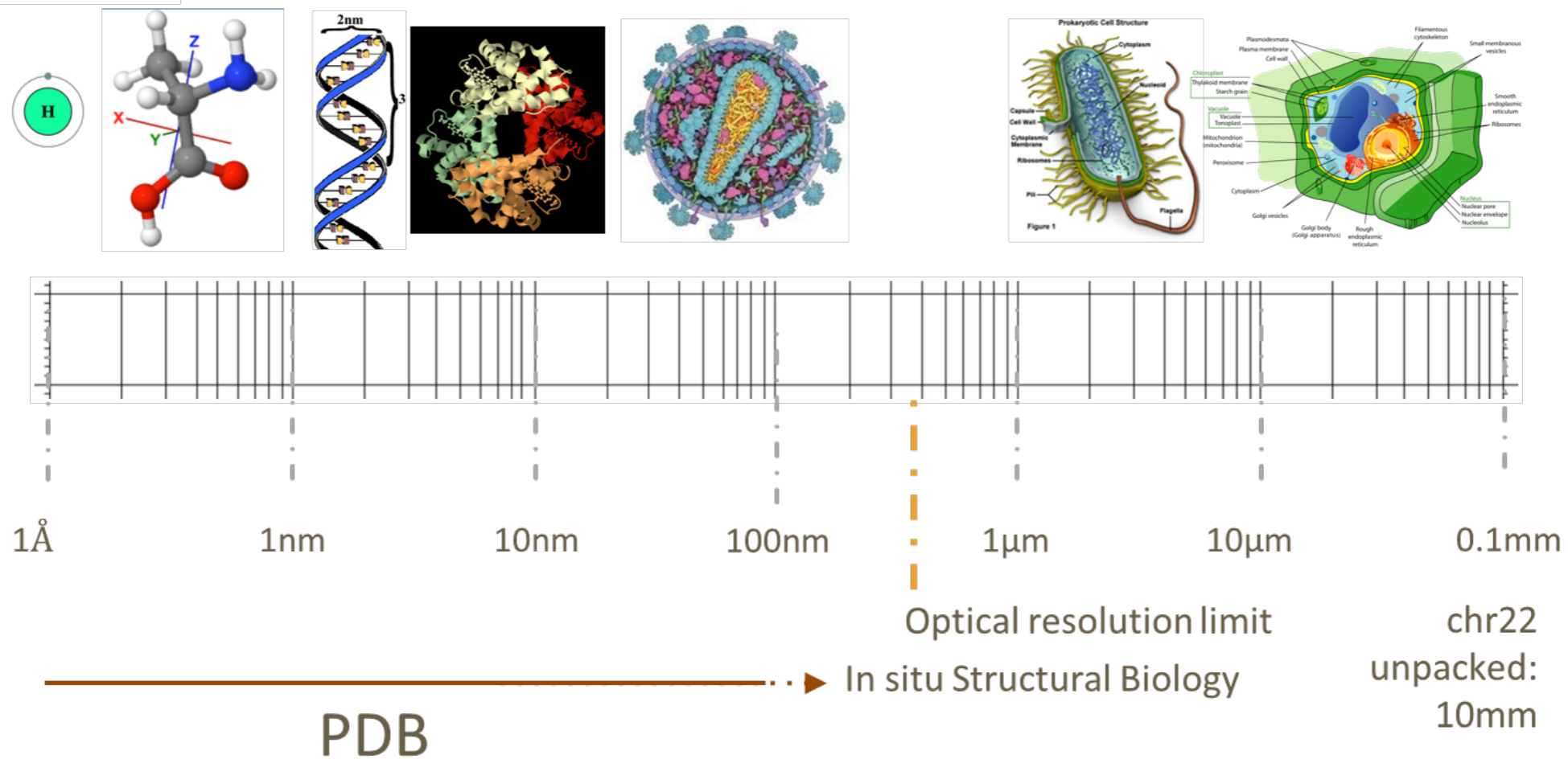
Largest asymmetric structure in PDB

Largest symmetric structure in PDB
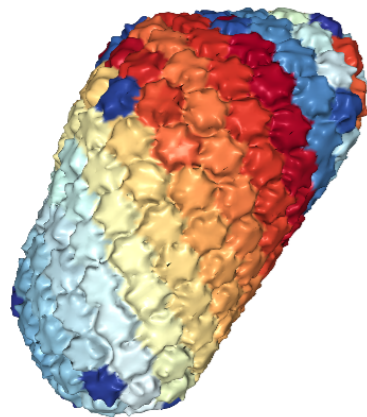
HIV-1 capsid: PDB ID 3J3Q
~2.4M unique atoms
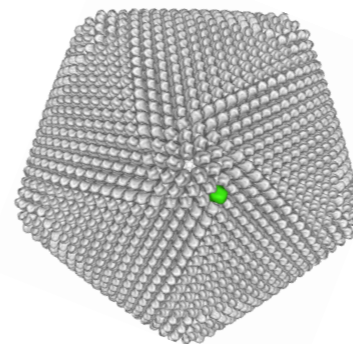
Faustovirus major capsid: PDB ID 5J7V
~40M overall atoms

Slide Credit: Peter Rose

Motivation 1:
Detailed understanding of molecular interactions

Provides an invaluable structural context for conservation and mechanistic analysis leading to functional insight.
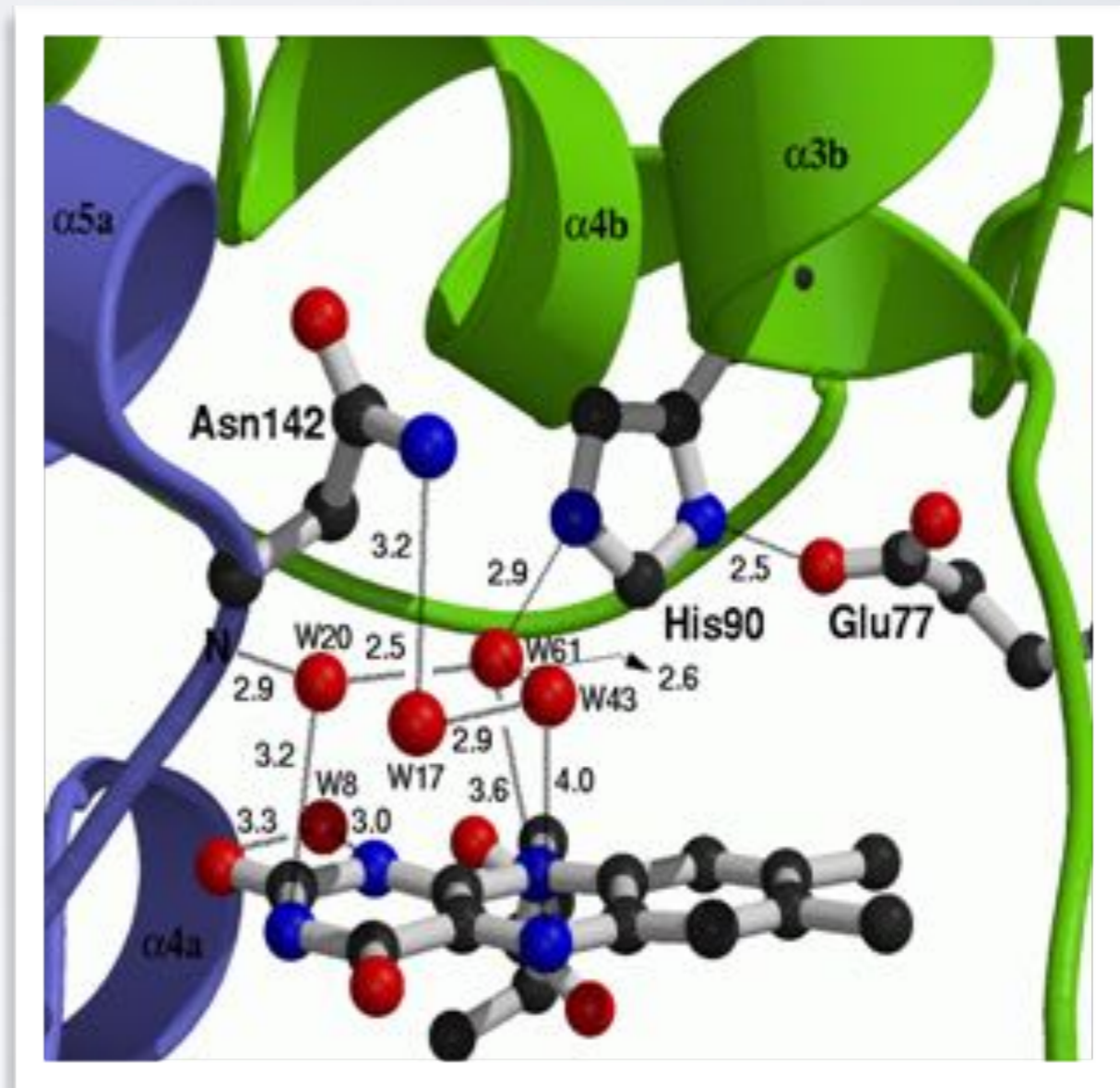
Motivation 1:
Detailed understanding of
molecular interactions

Computational modeling can
provide detailed insight into
functional interactions, their
regulation and potential
consequences of perturbation.
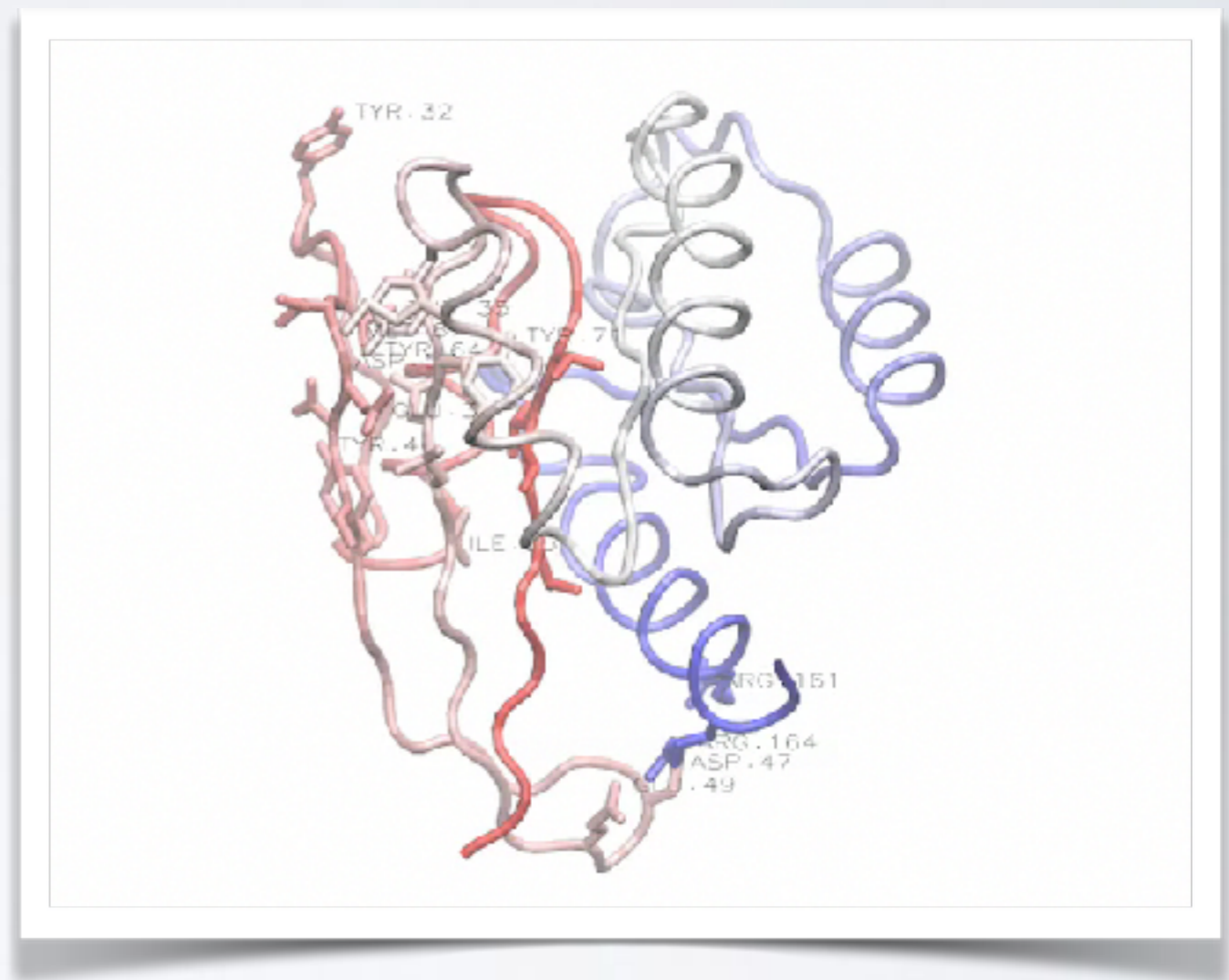
Grant *et al.* PLoS. Comp. Biol. (2010)

Motivation 2:
Lots of structural data is becoming available

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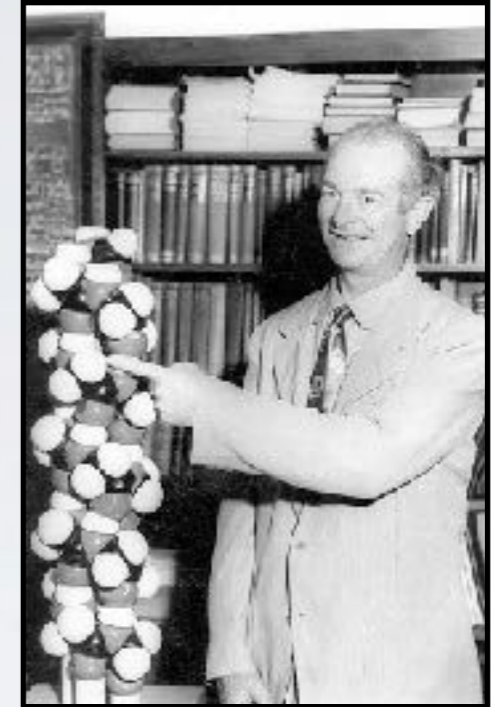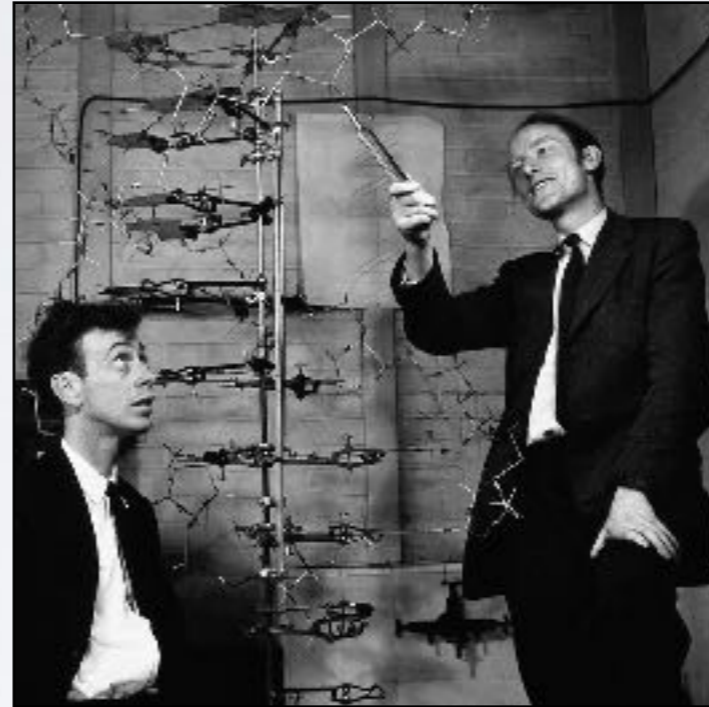

126,060
(1/22/2017)

Data from: http://www.rcsb.org/pdb/statistics/
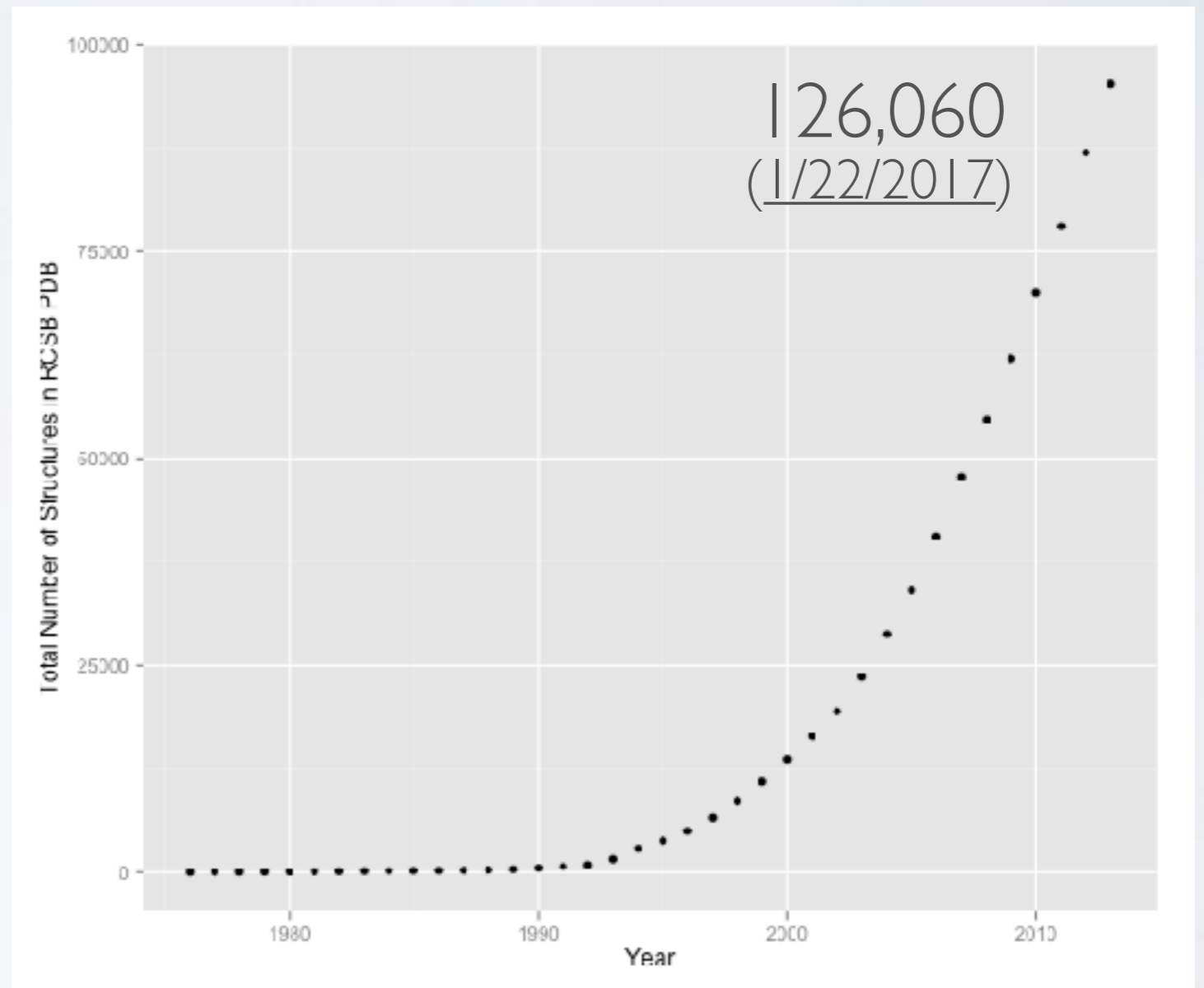
Motivation 2:
Lots of structural data is becoming available

Structural Genomics has contributed to driving down the cost and time required for structural determination
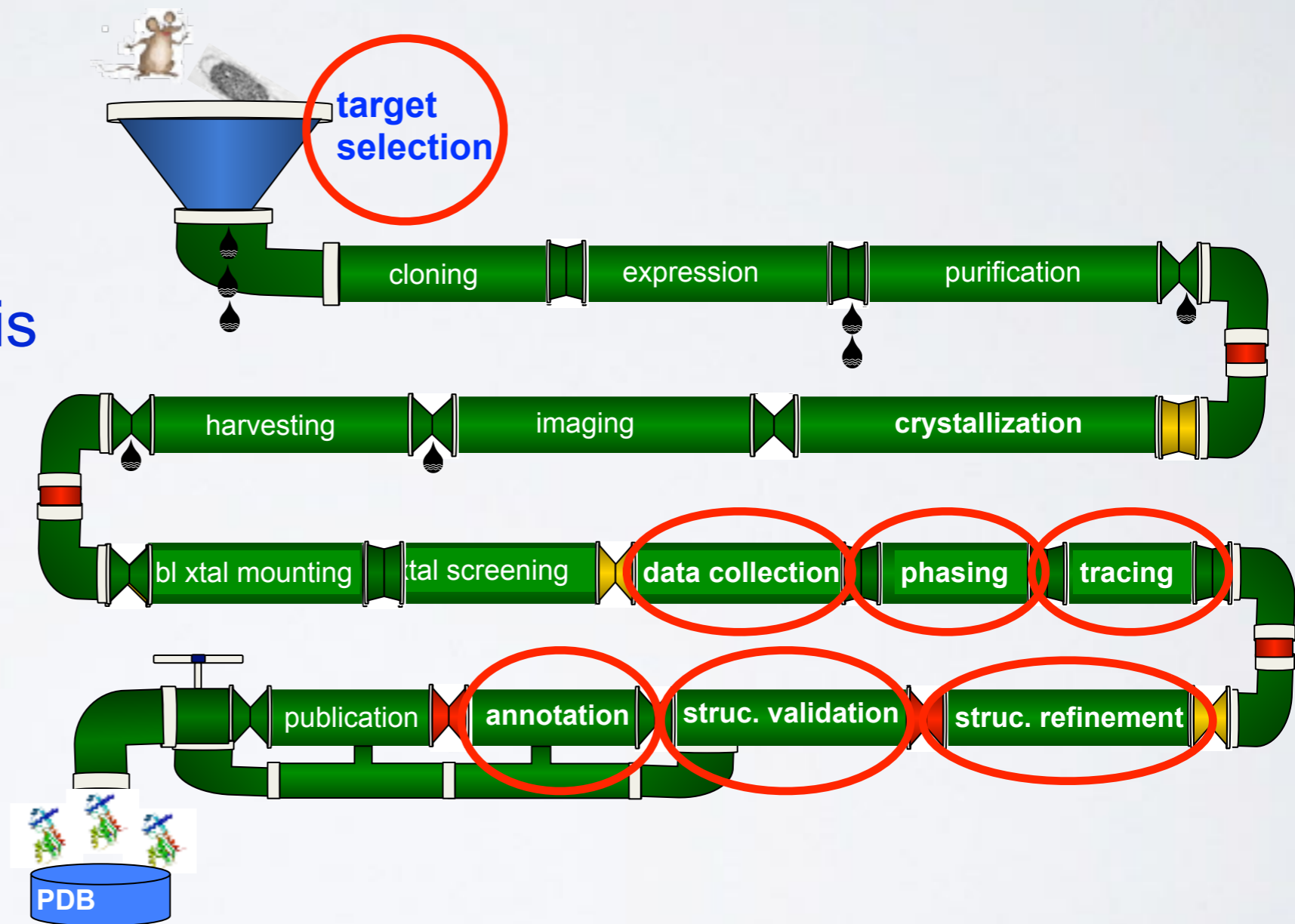
Image Credit: "Structure determination assembly line" Adam Godzik

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
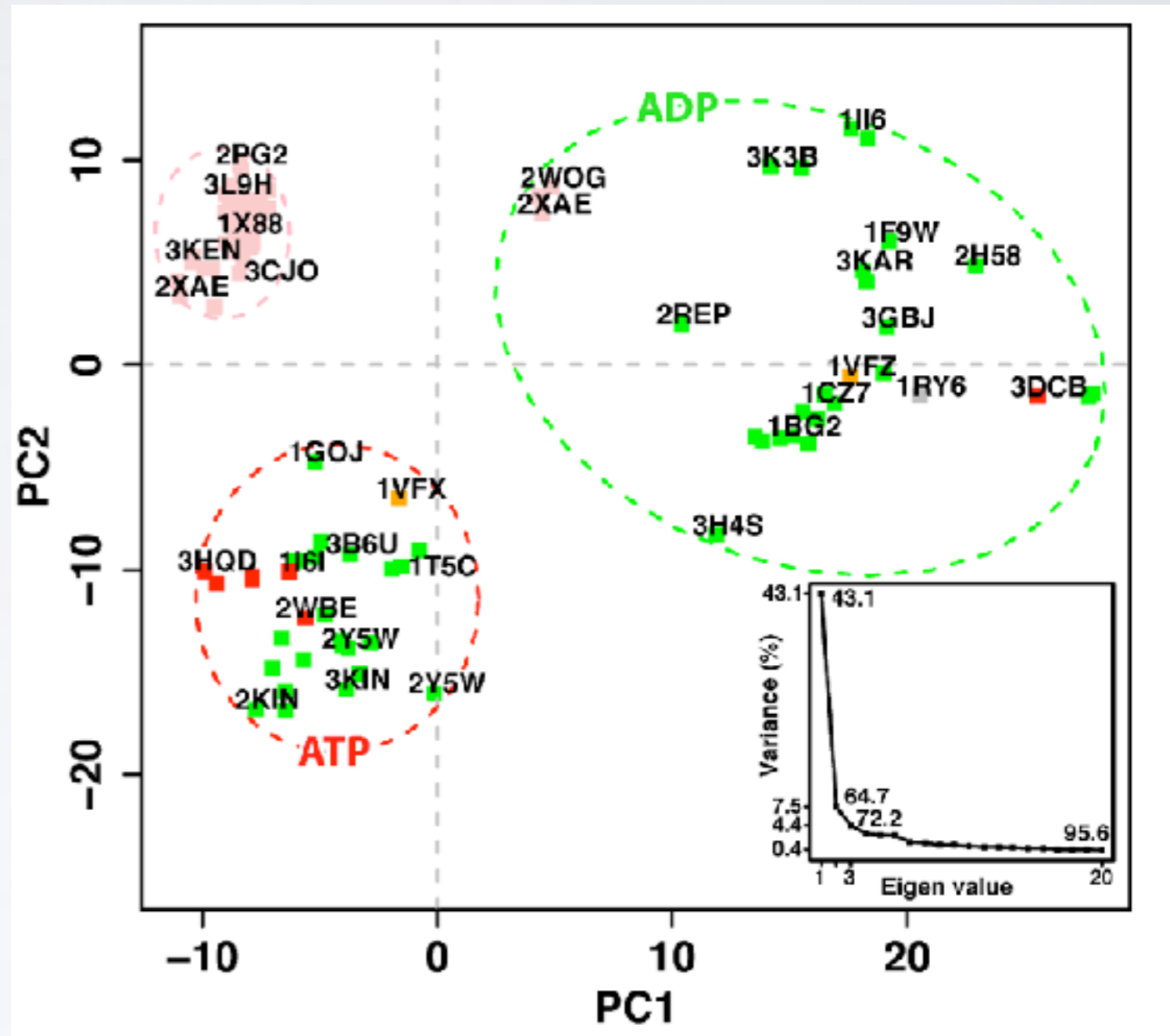
Goals:

- Analysis
- Visualization
- Comparison
- Prediction
- Design
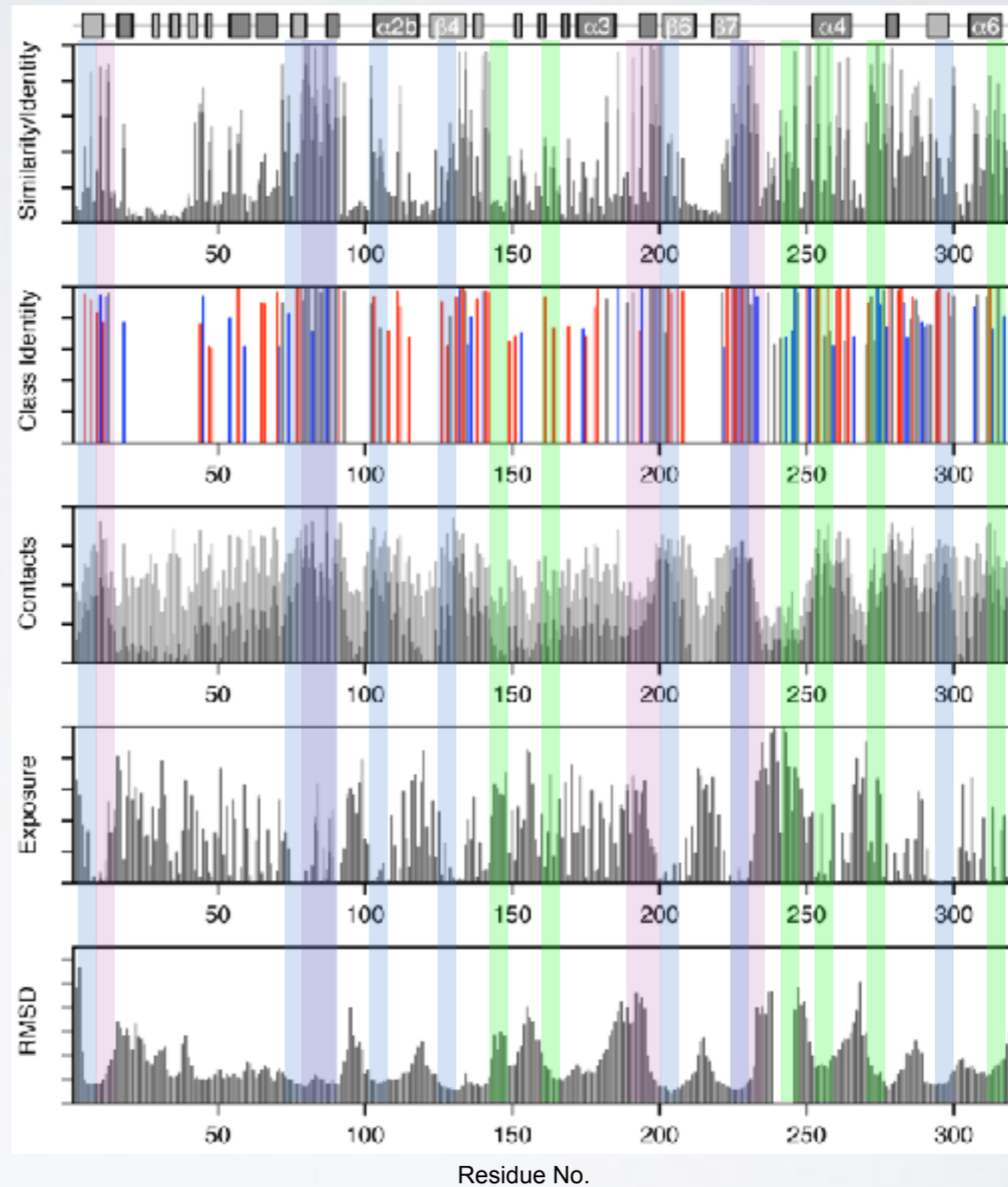
Grant *et al.* JMB. (2007)

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

myosin

G-protein

kinesin

Grant *et al.* unpublished
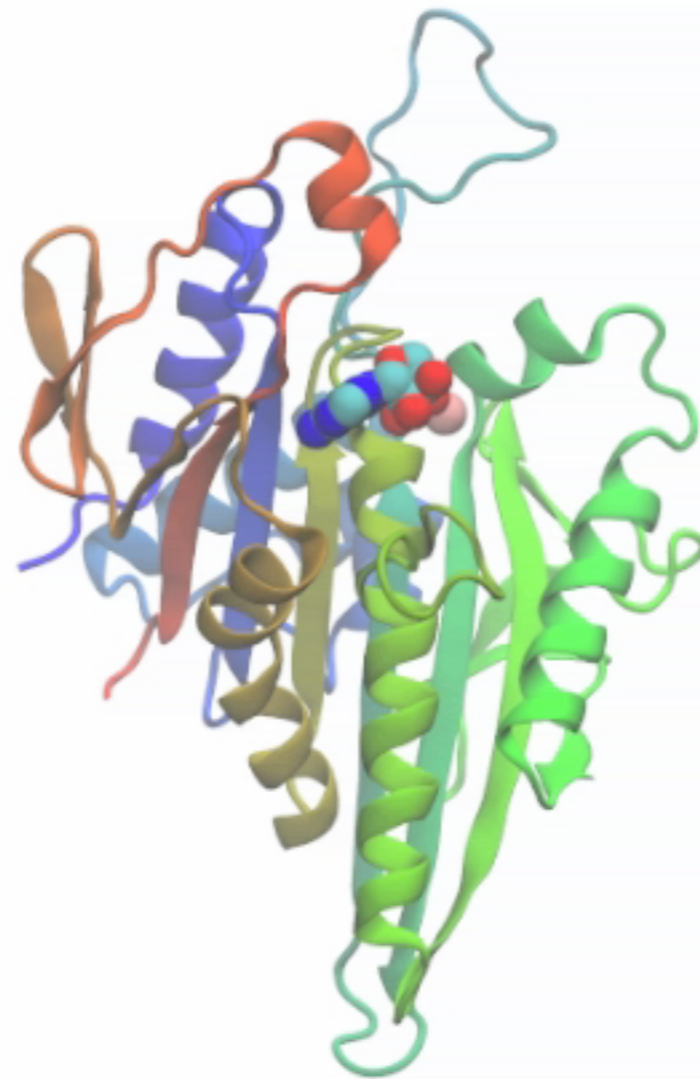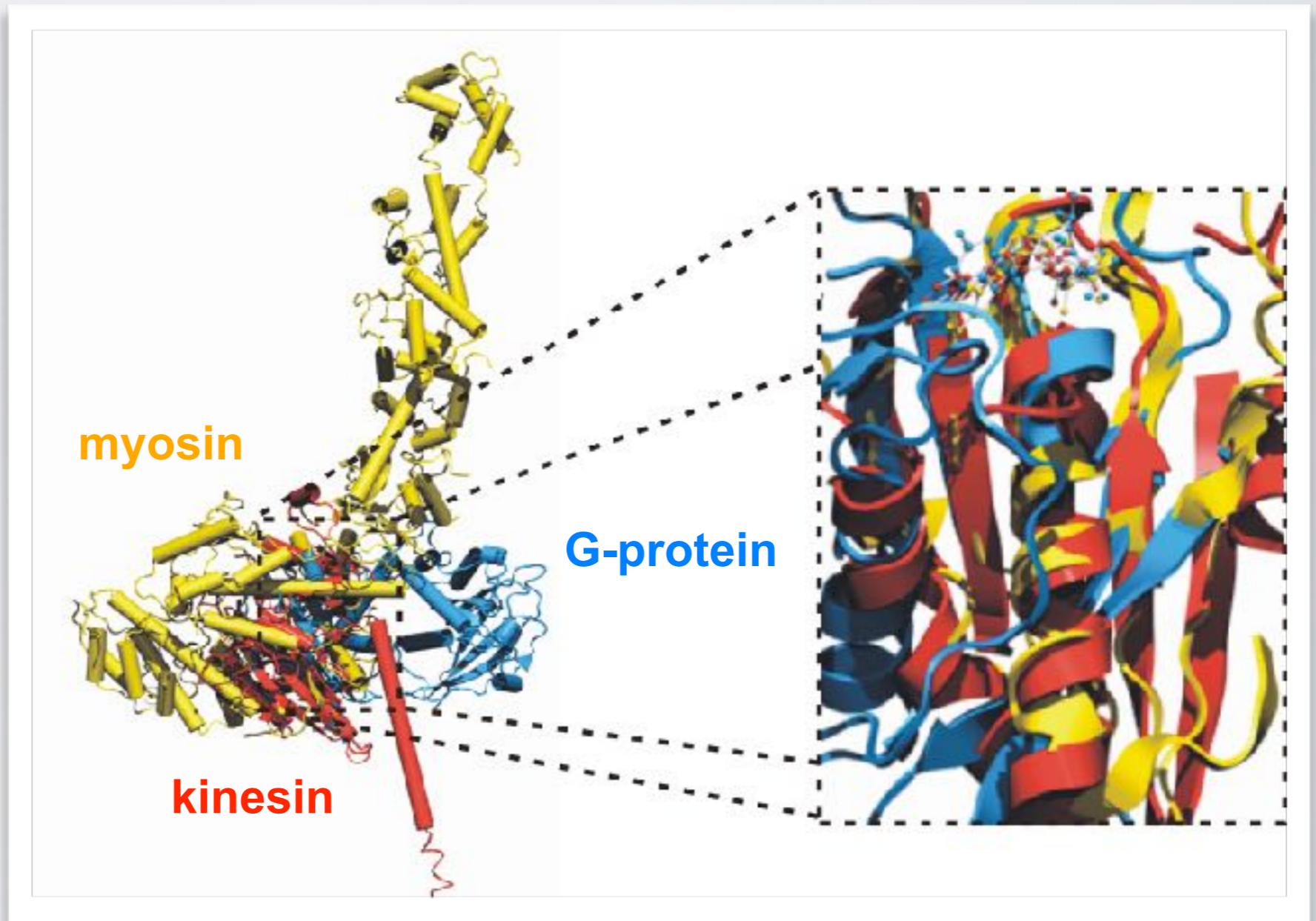
Goals:
- Analysis
- Visualization
- Comparison
- Prediction
- Design
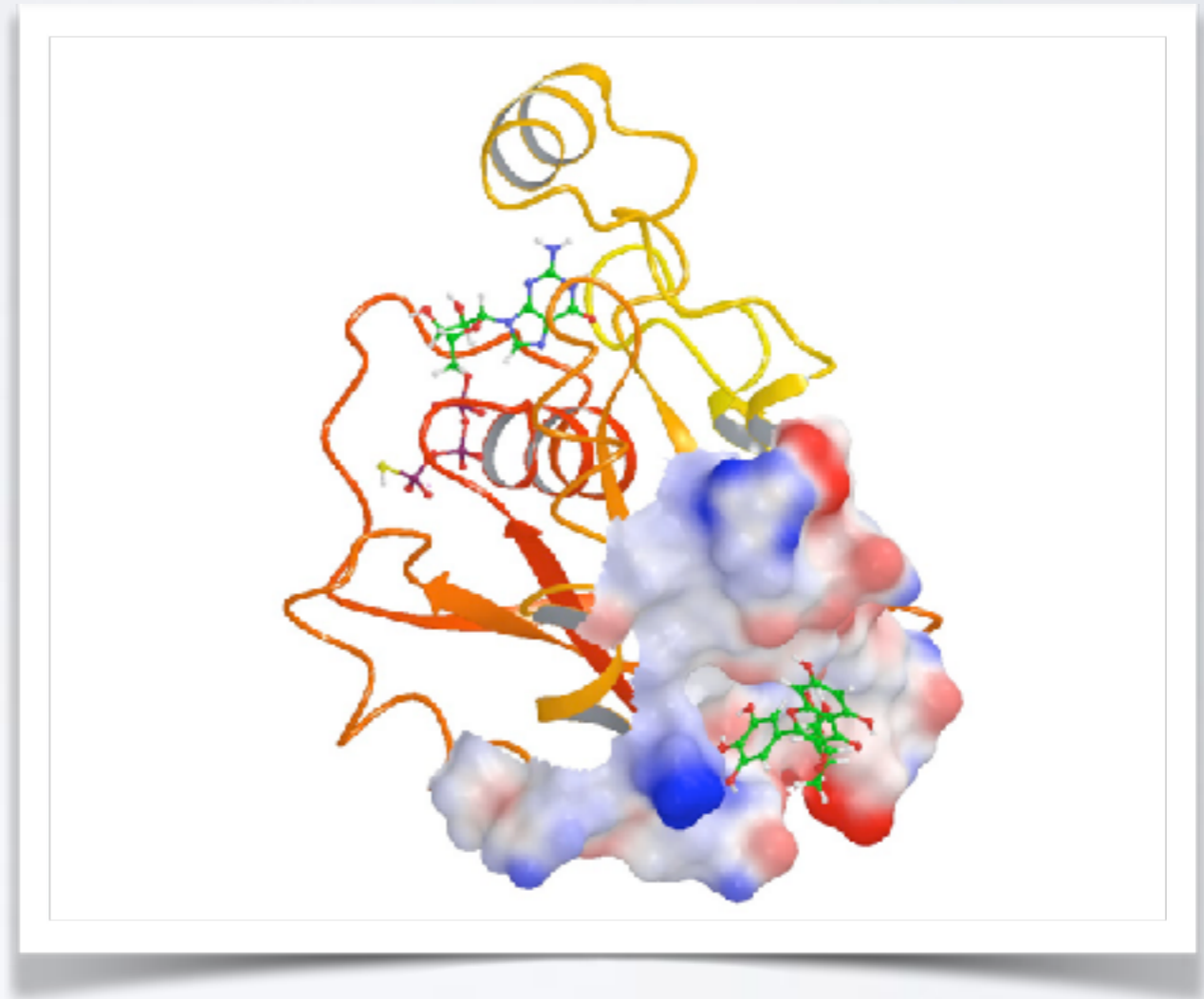
Grant *et al.* PLoS One (2011, 2012)

Goals:
- Analysis
- Visualization
- Comparison
- Prediction
- Design
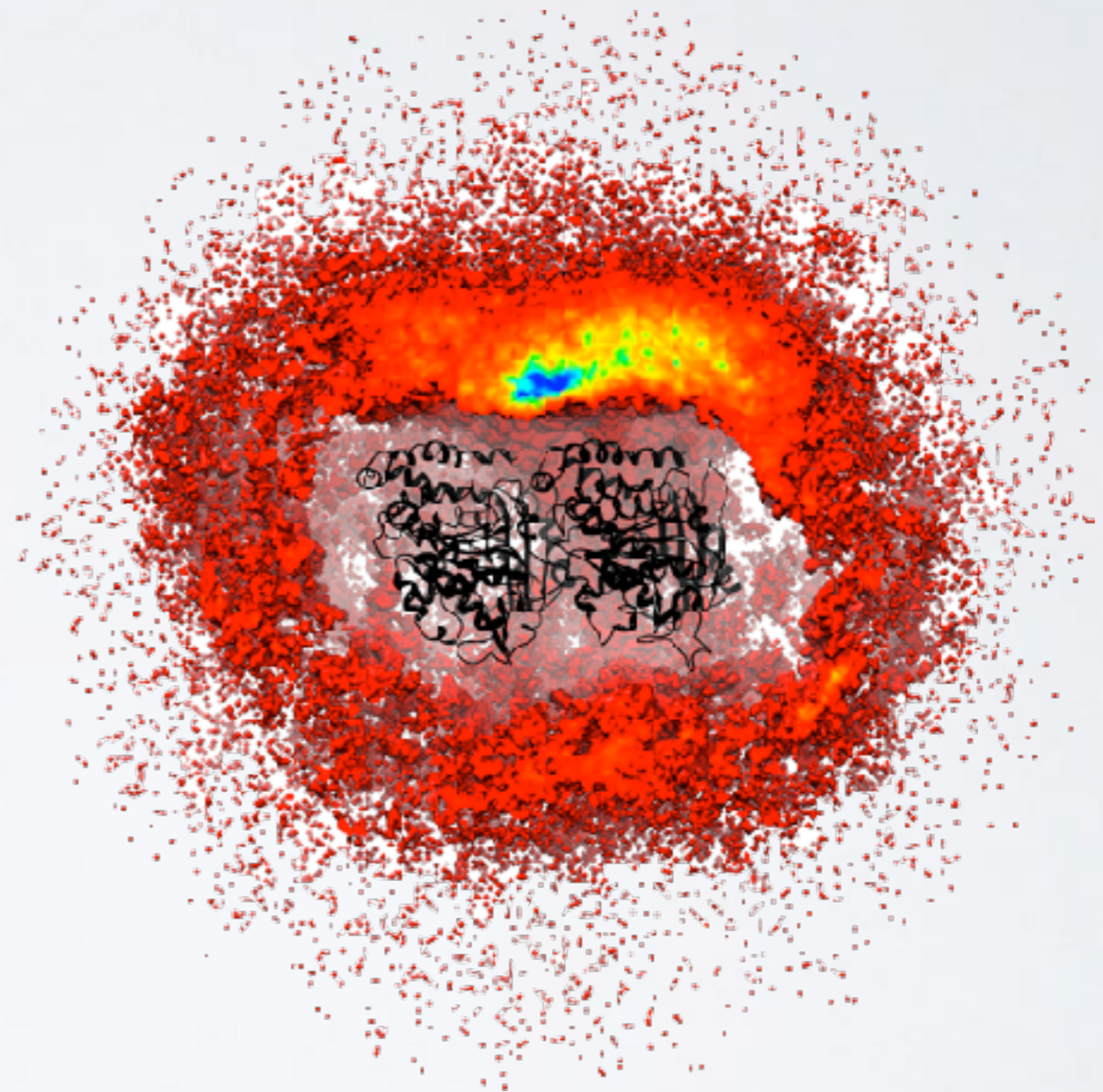
Grant *et al.* PLoS Biology (2011)

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

Today's Menu

- **Overview of structural bioinformatics**
  - Motivations, goals and challenges

- **Fundamentals of protein structure**
  - Structure composition, form and forces

- **Representing, interpreting & modeling protein structure**
  - Visualizing & interpreting protein structures
  - Analyzing protein structures
  - Modeling energy as a function of structure

HIERARCHICAL STRUCTURE OF PROTEINS
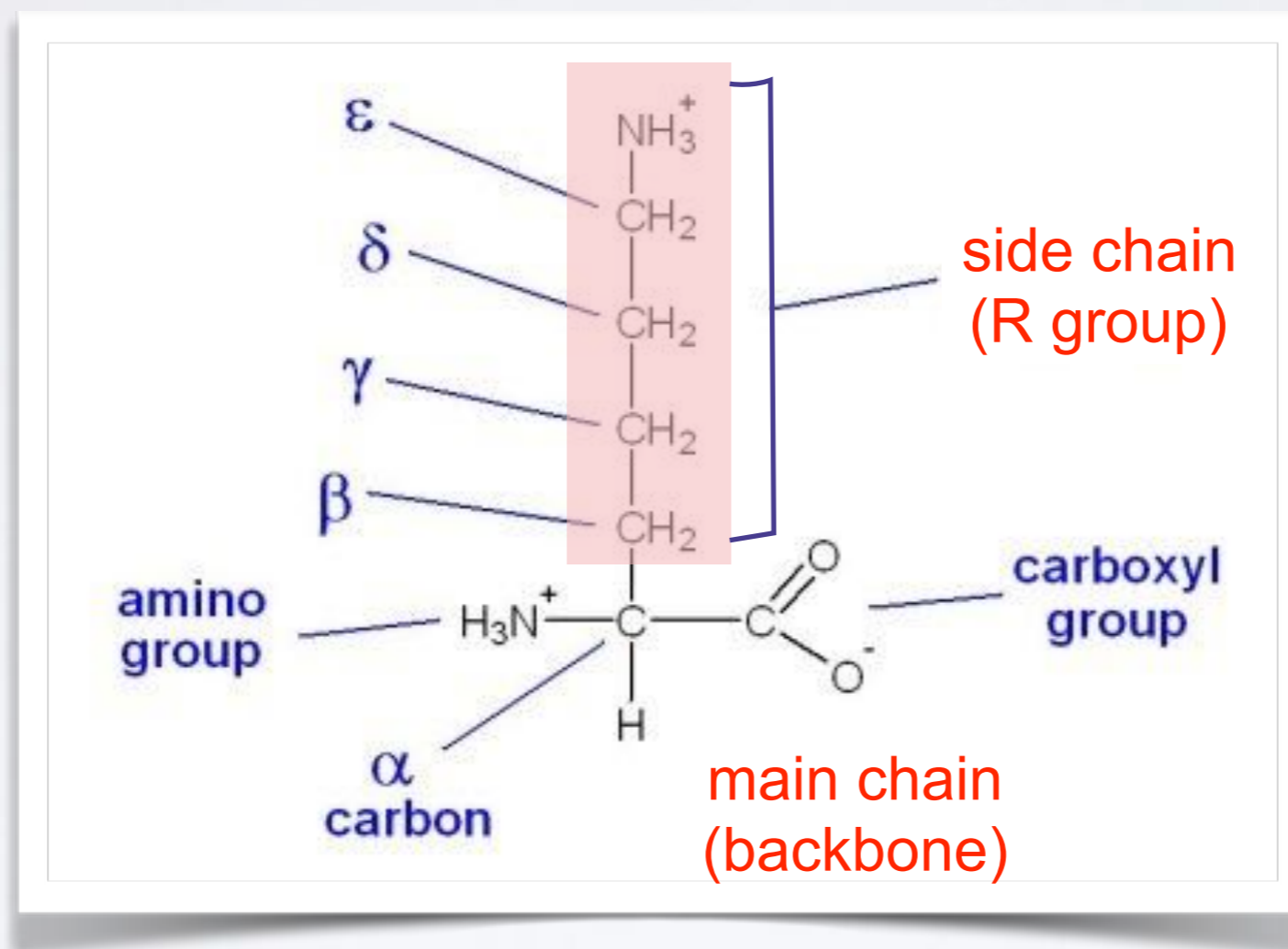
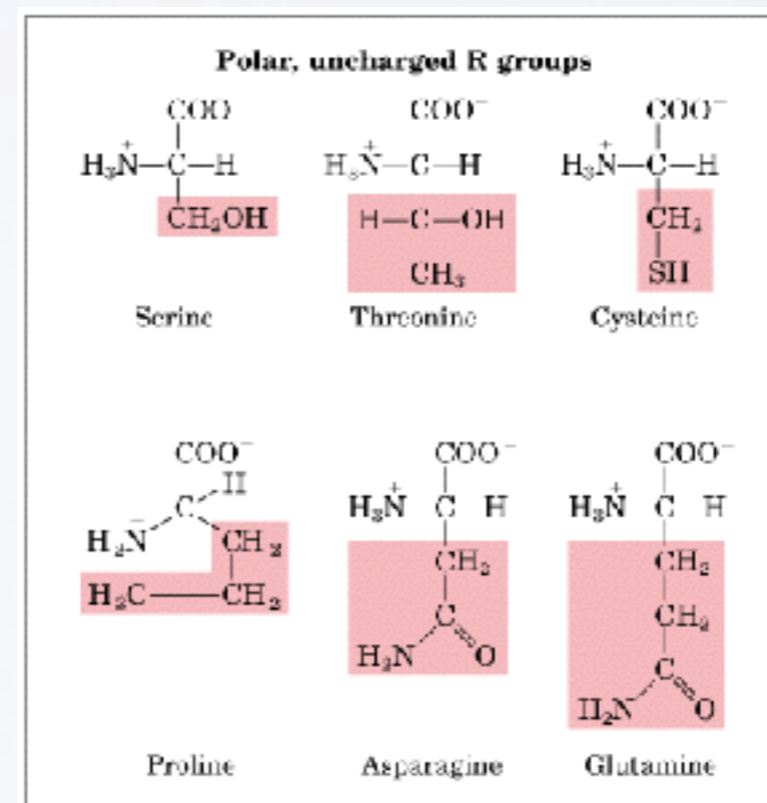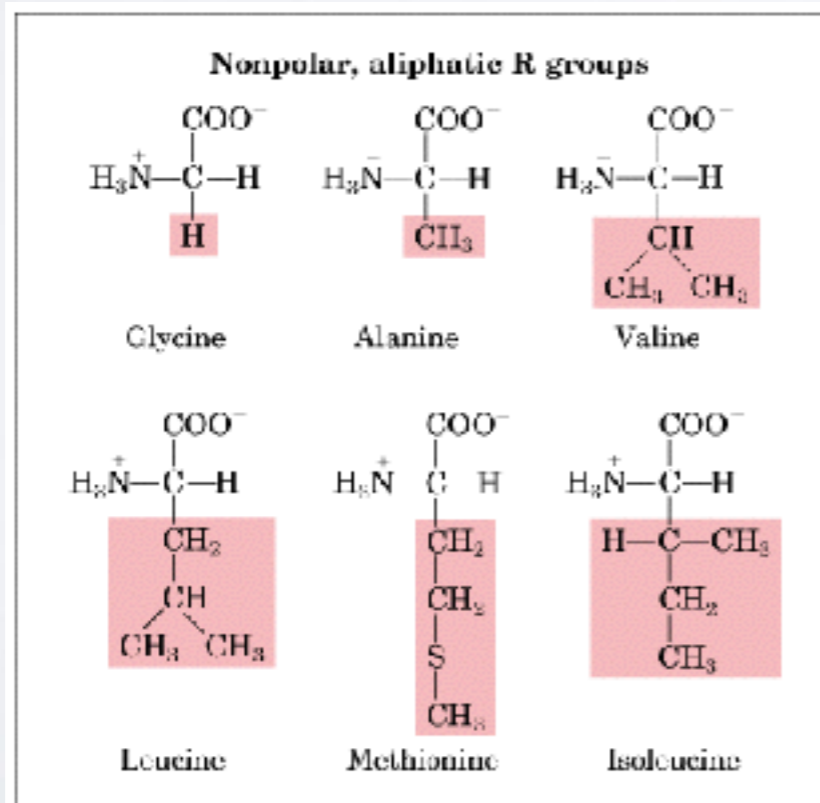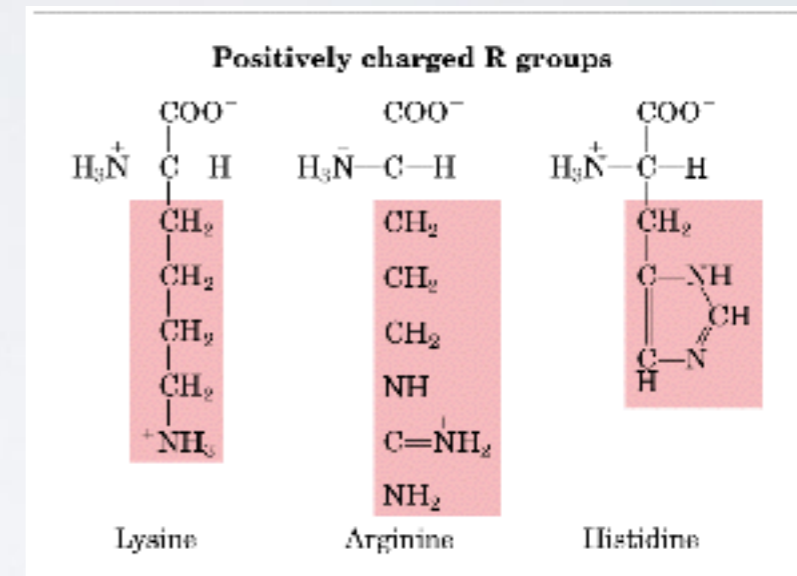**Primary > Secondary > Tertiary > Quaternary**
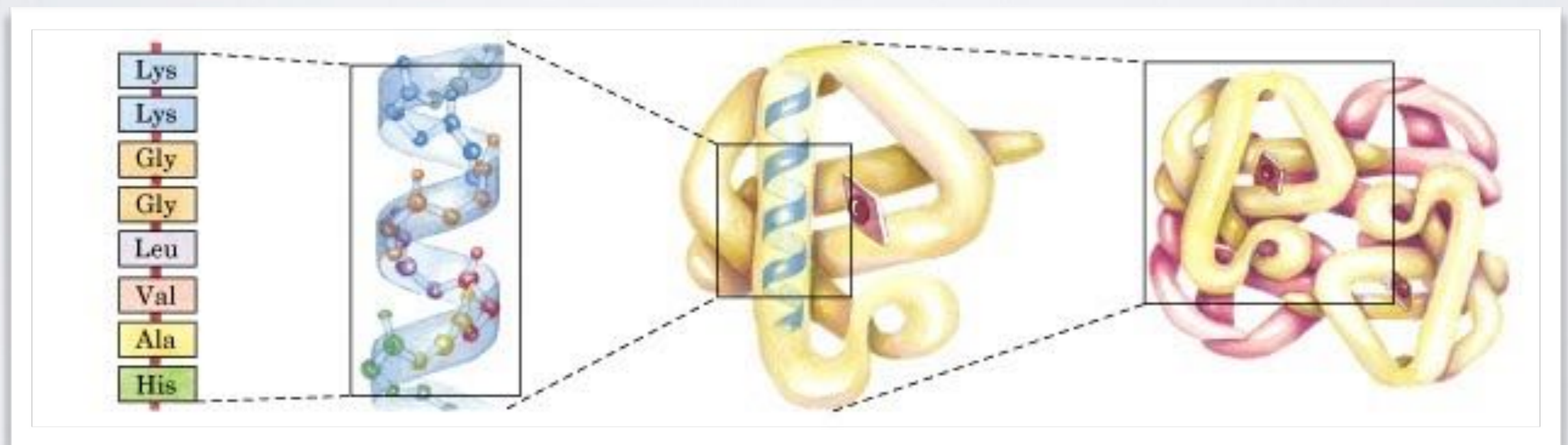
amino acid residues | Alpha helix | Polypeptide chain | Assembled subunits

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

RECAP: AMINO ACID NOMENCLATURE

AMINO ACIDS CAN BE GROUPED BY THE
PHYSIOCHEMICAL PROPERTIES

Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/
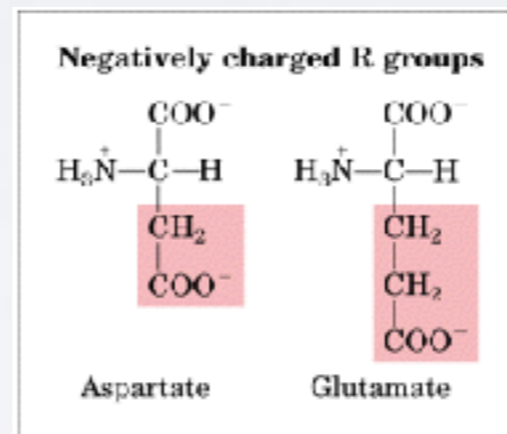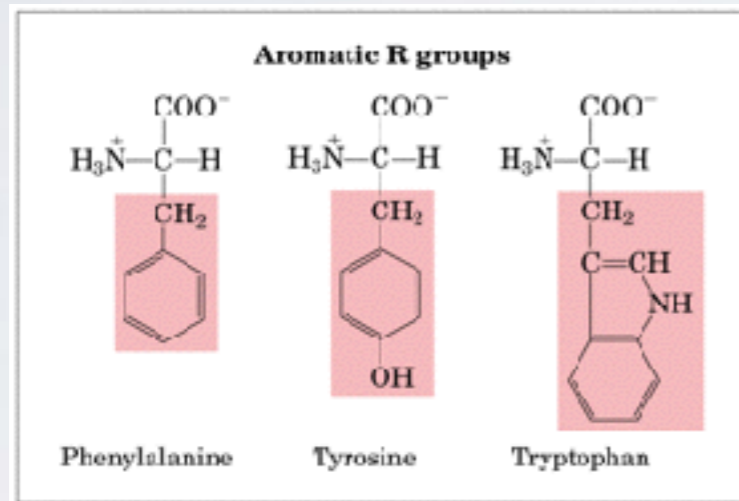
AMINO ACIDS POLYMERIZE THROUGH **PEPTIDE BOND** FORMATION

side chains

backbone

N-terminal ⟶ C-terminal
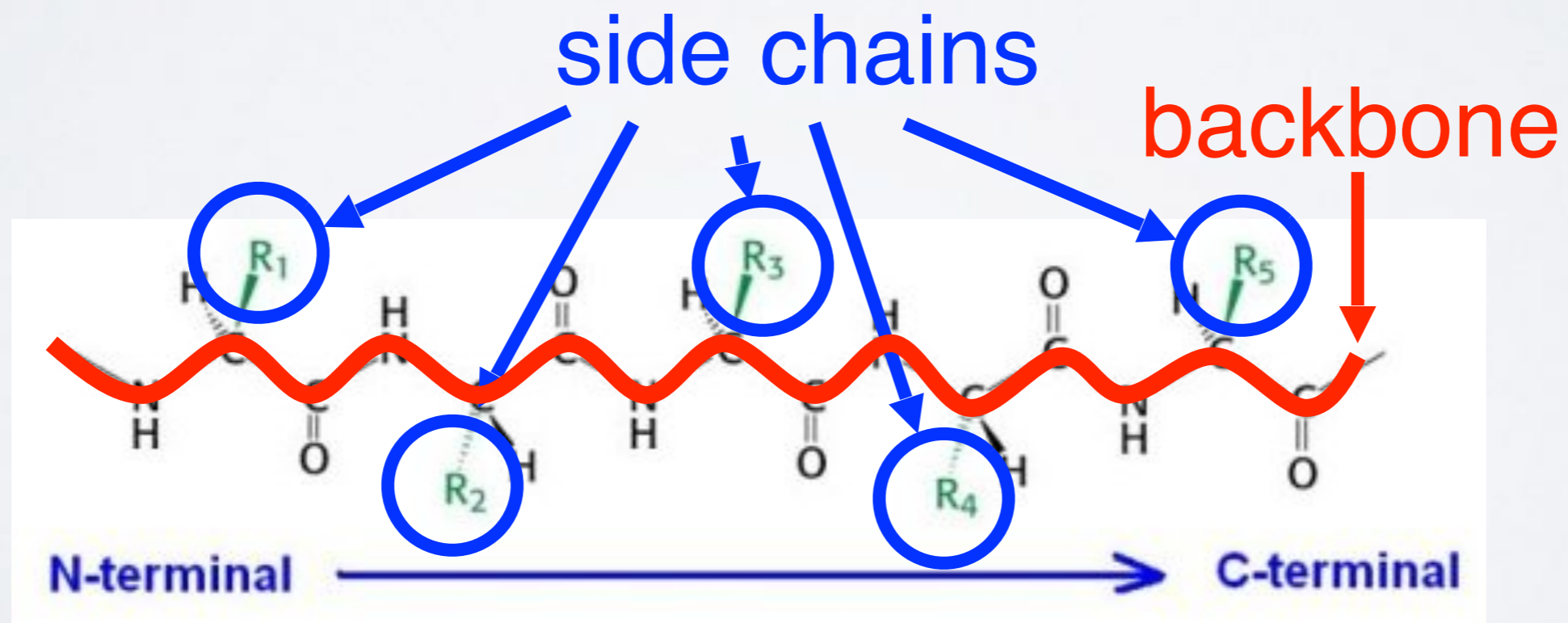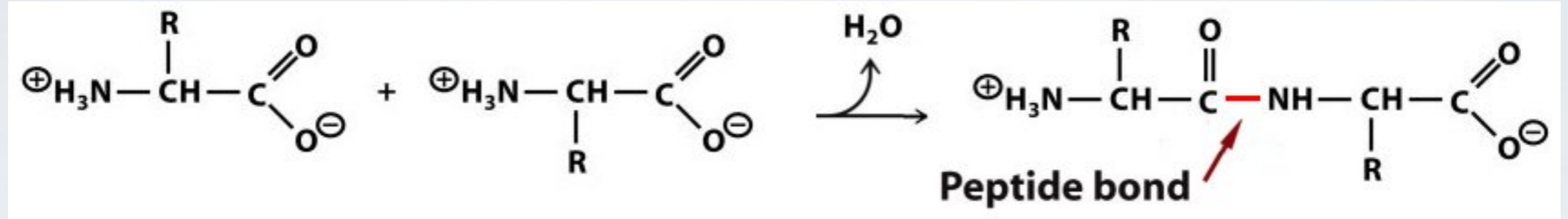
PEPTIDES CAN ADOPT DIFFERENT CONFORMATIONS BY VARYING THEIR
PHI & PSI BACKBONE TORSIONS

ϕ ψ

C-terminal

N-terminal

Bond angles and
lengths are largely
invariant

Peptide bond is
planer
($C\alpha$, C, O, N, H, $C\alpha$
all lie in the same
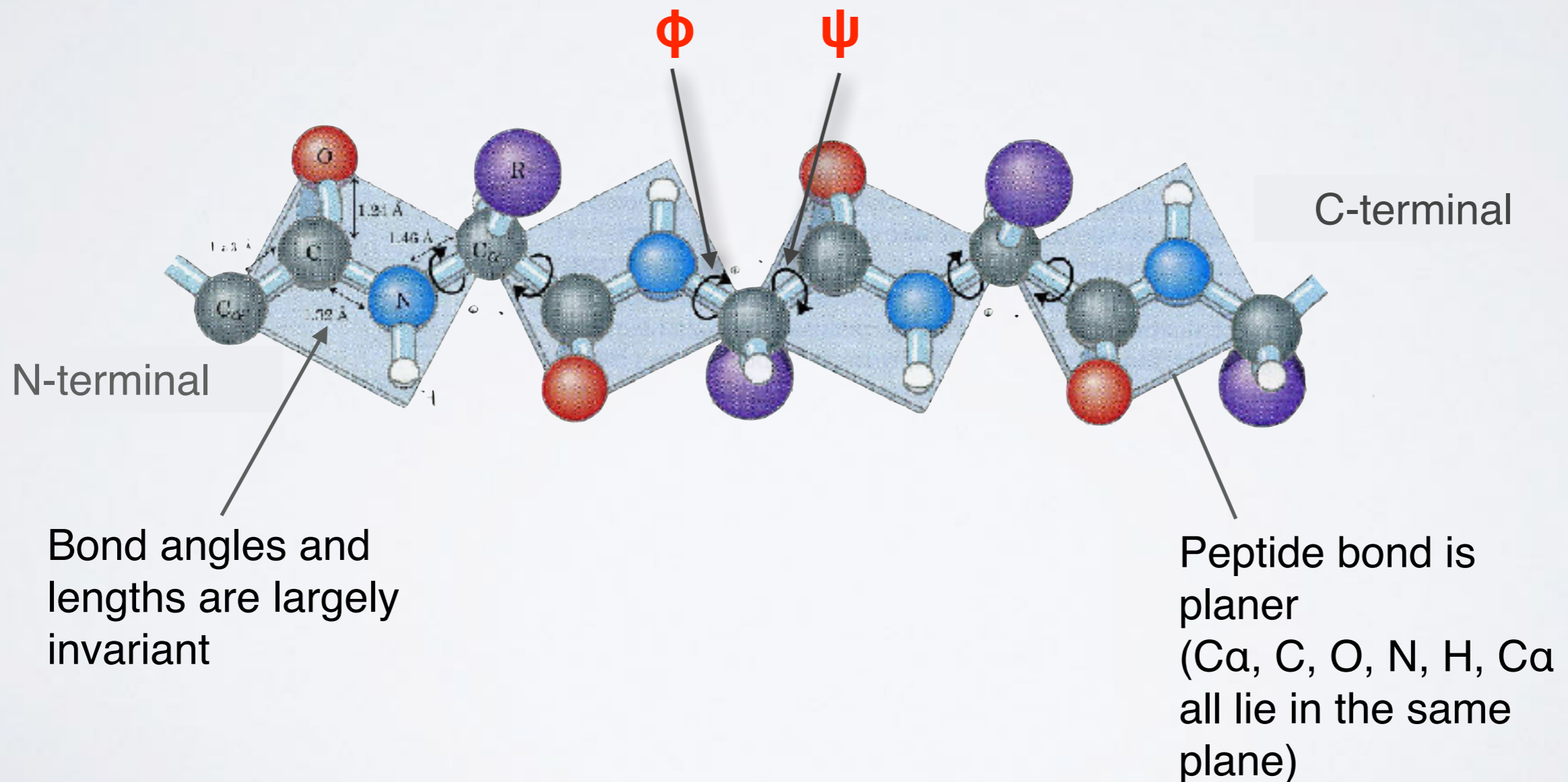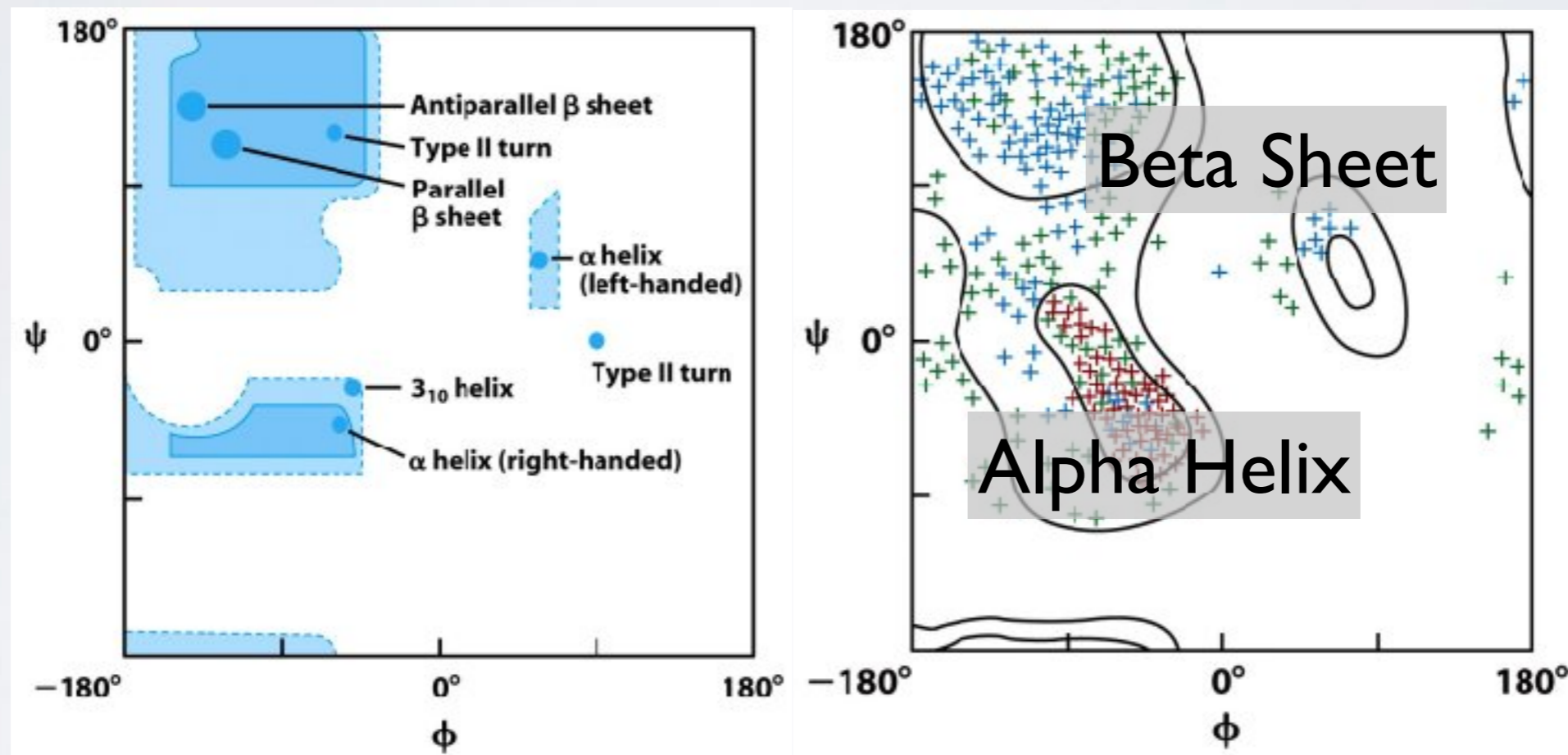plane)

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
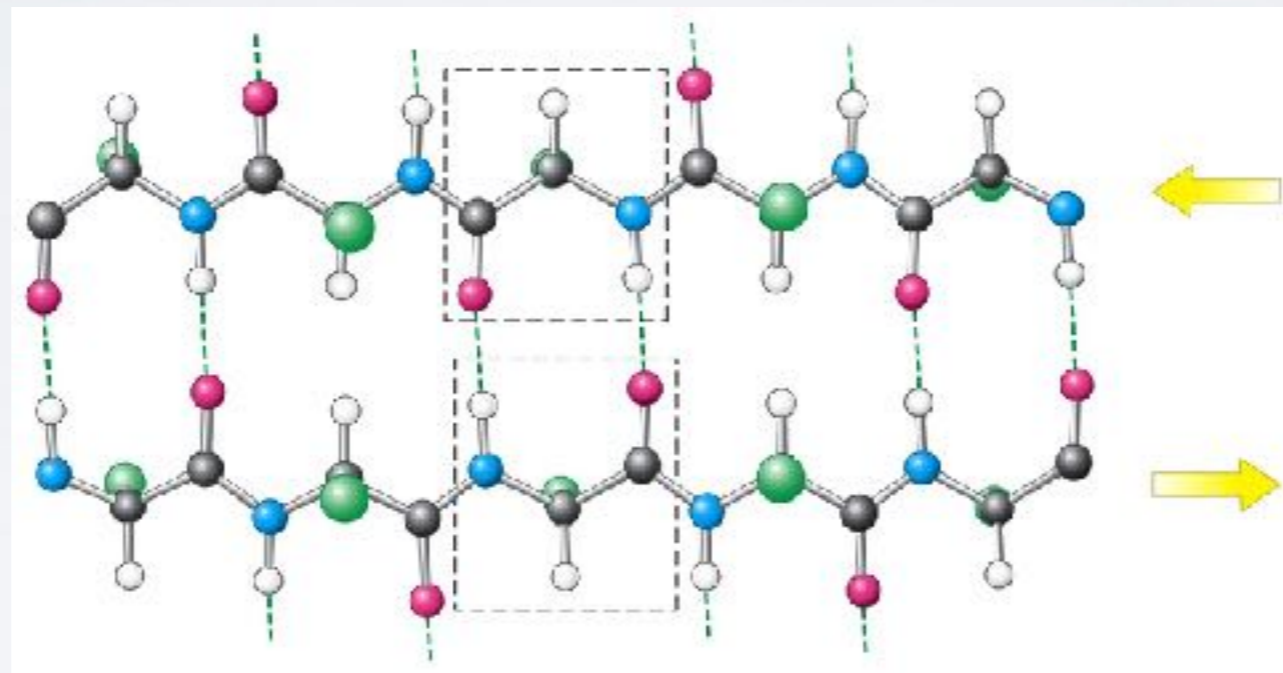## **ALPHA HELIX** & BETA SHEET
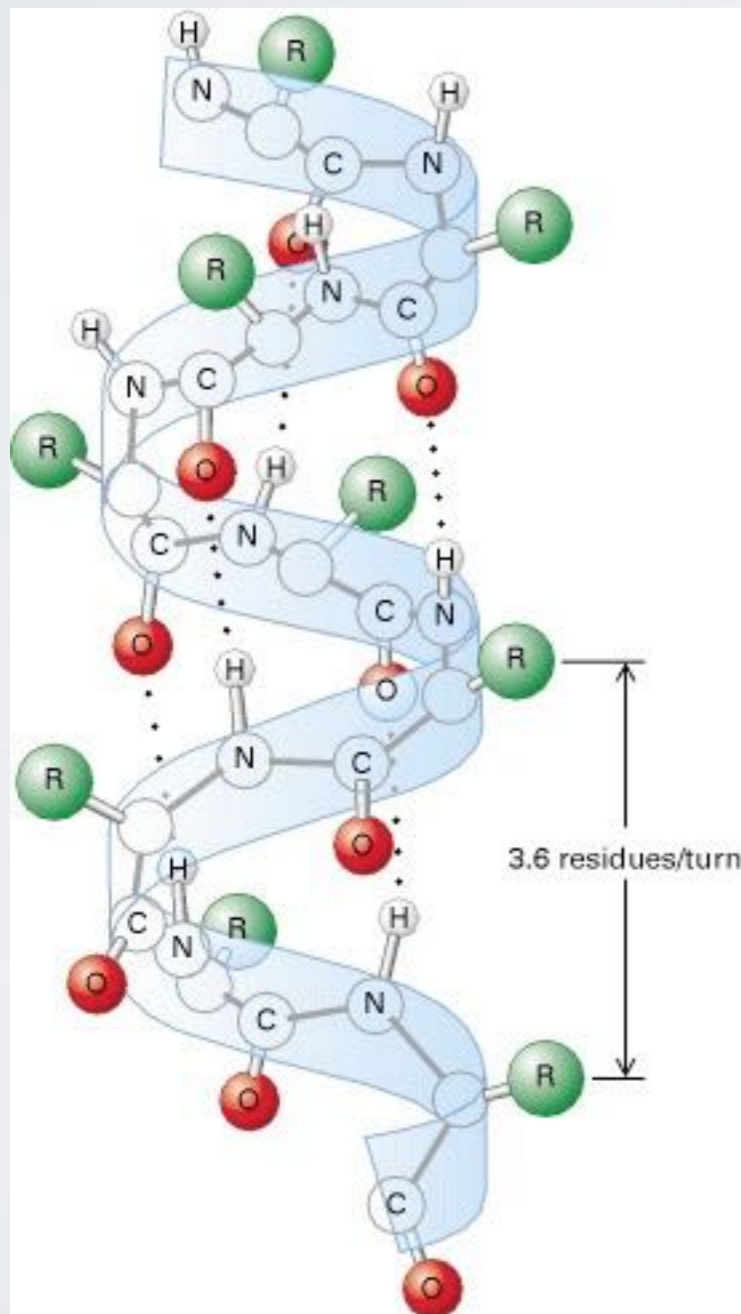
3.6 residues/turn

α-helix

- Most common from has <u>3.6 residues per turn</u> (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**
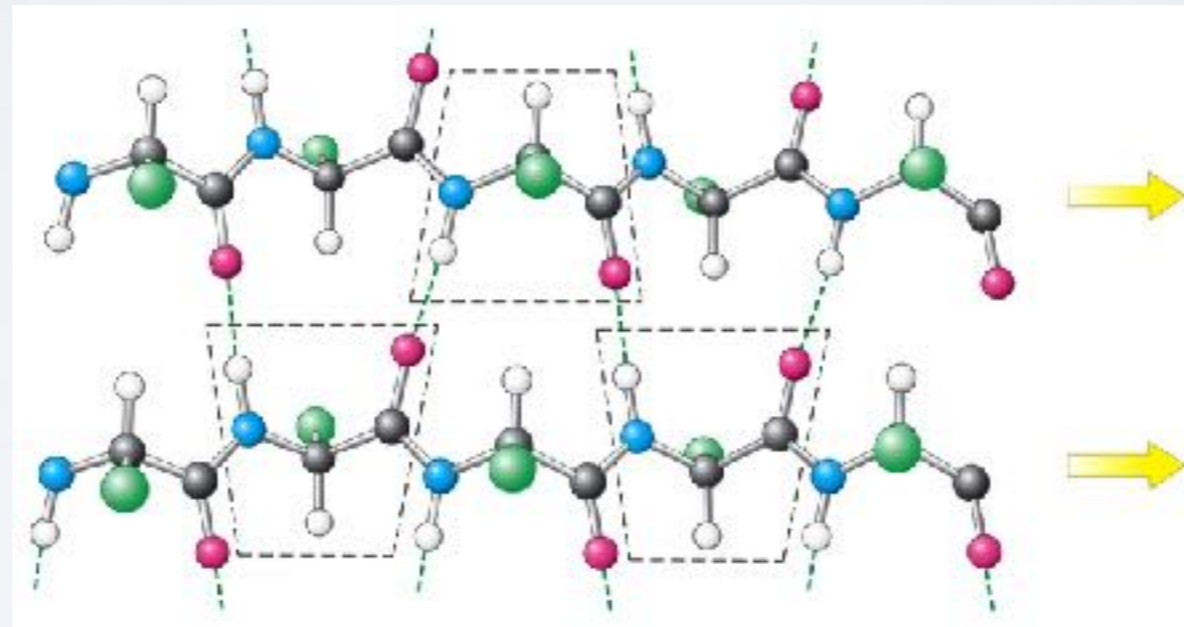

In **<u>antiparallel</u> β-sheets**

- Adjacent β-strands run in <u>opposite</u> 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**
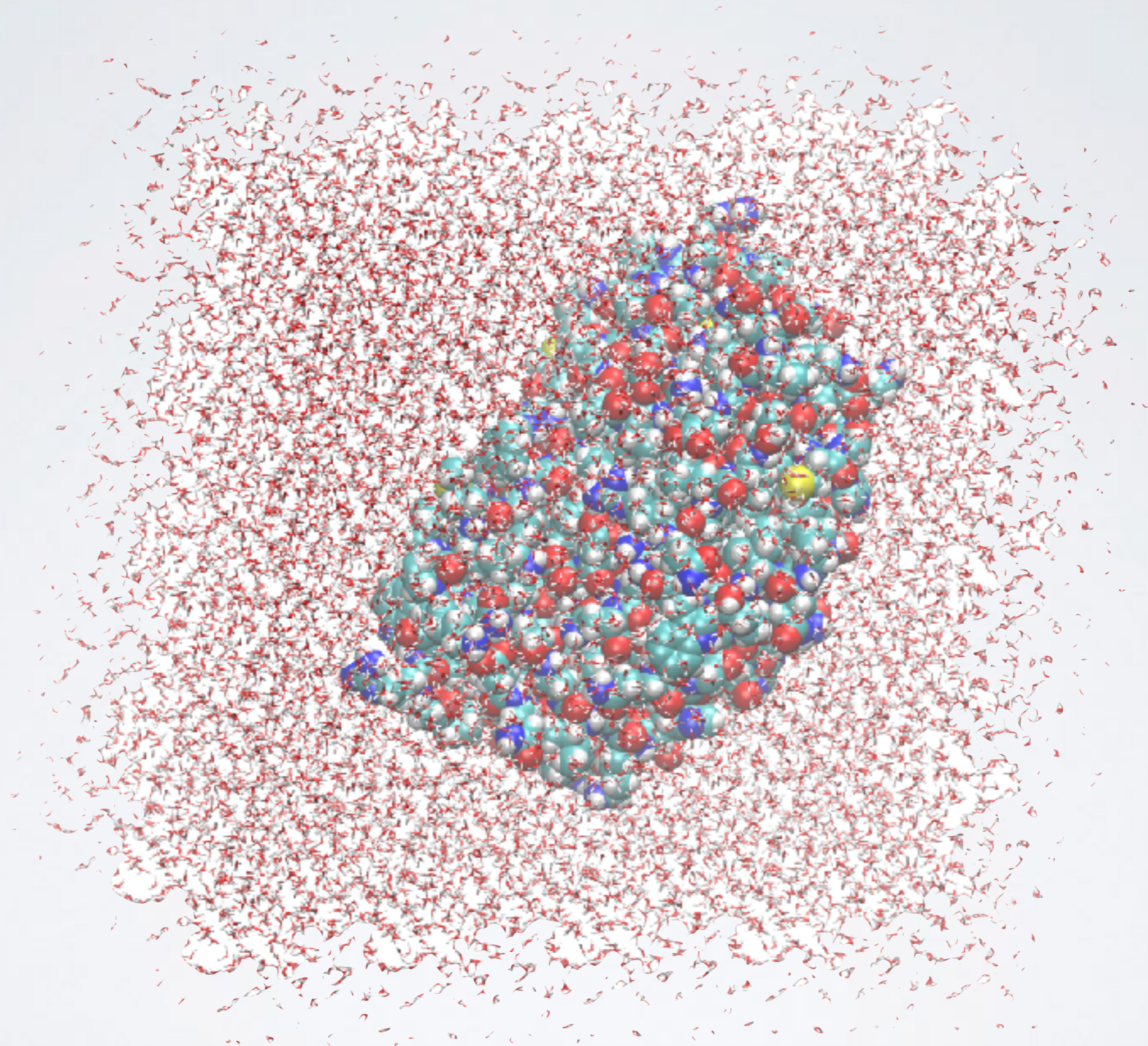
- Adjacent β-strands run in <u>same</u> direction
- Hydrogen bonds (dashed lines) between NH and CO stabilize the structure
- The side chains (in green) are above and below the sheet
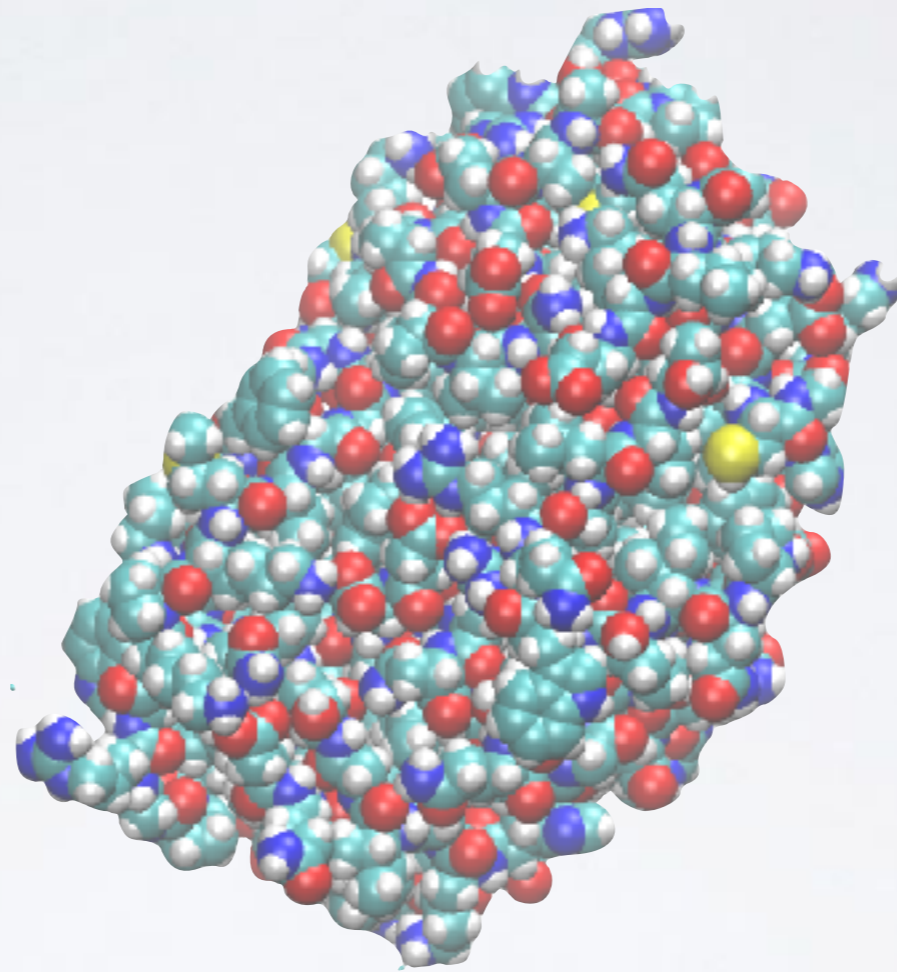
Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/
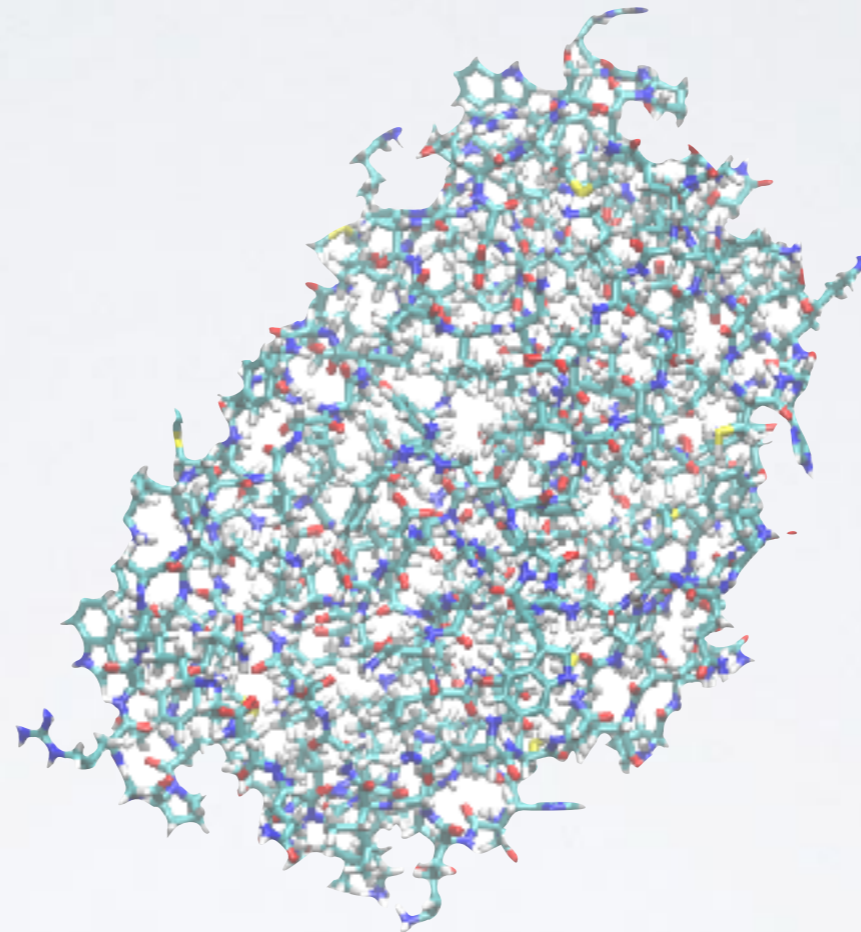
What Does a Protein Look like?
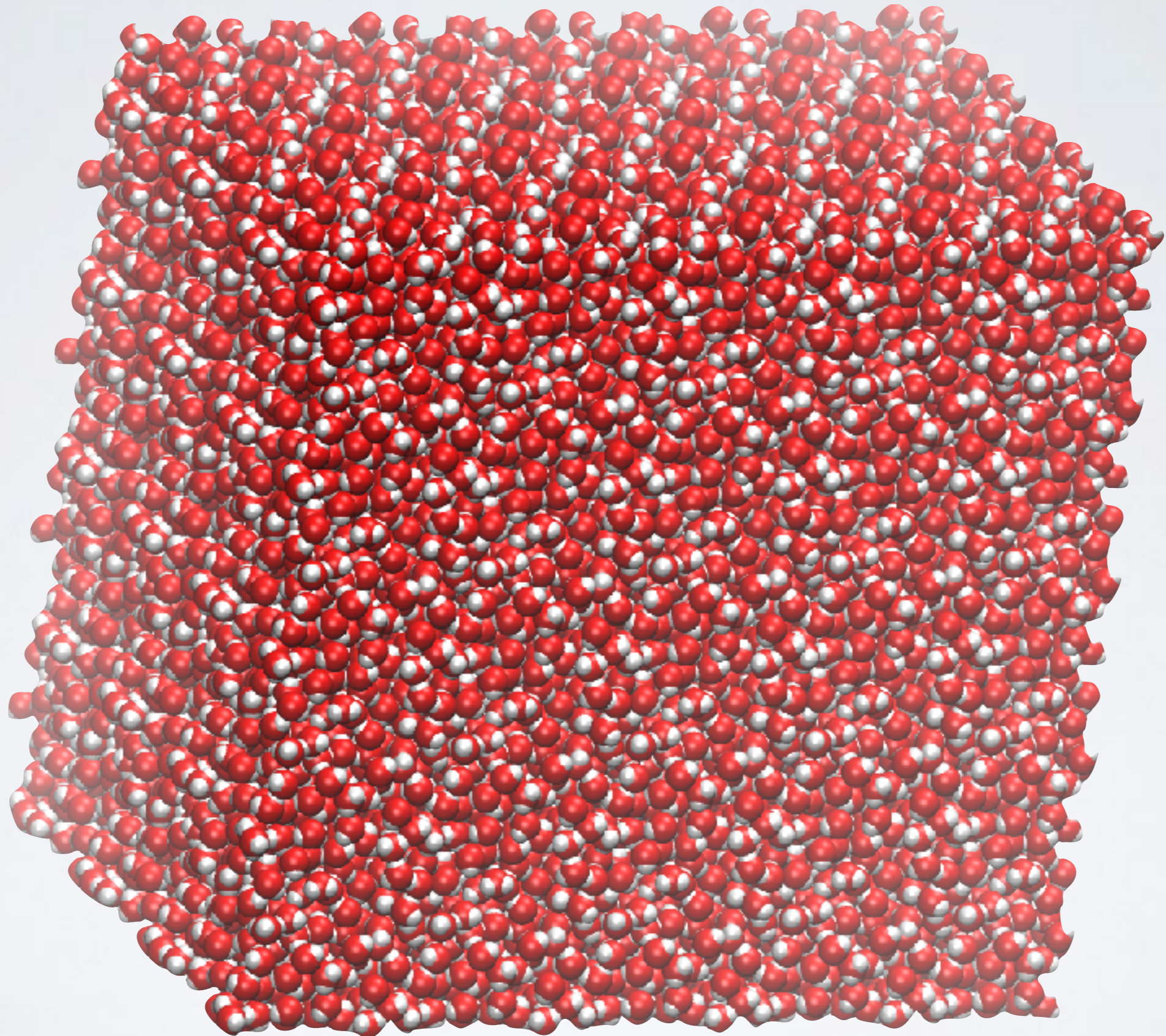
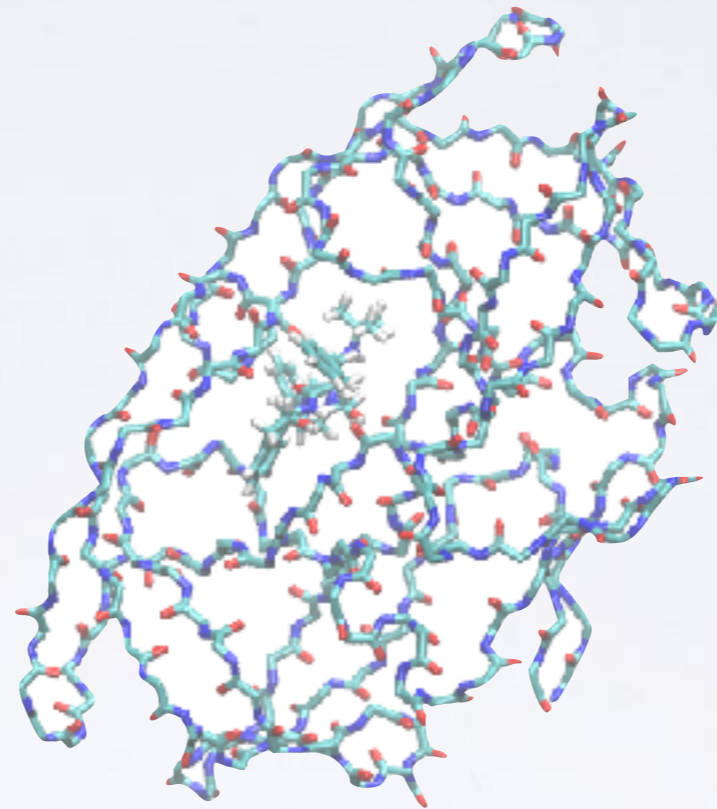- Proteins are stable (and hidden) in water
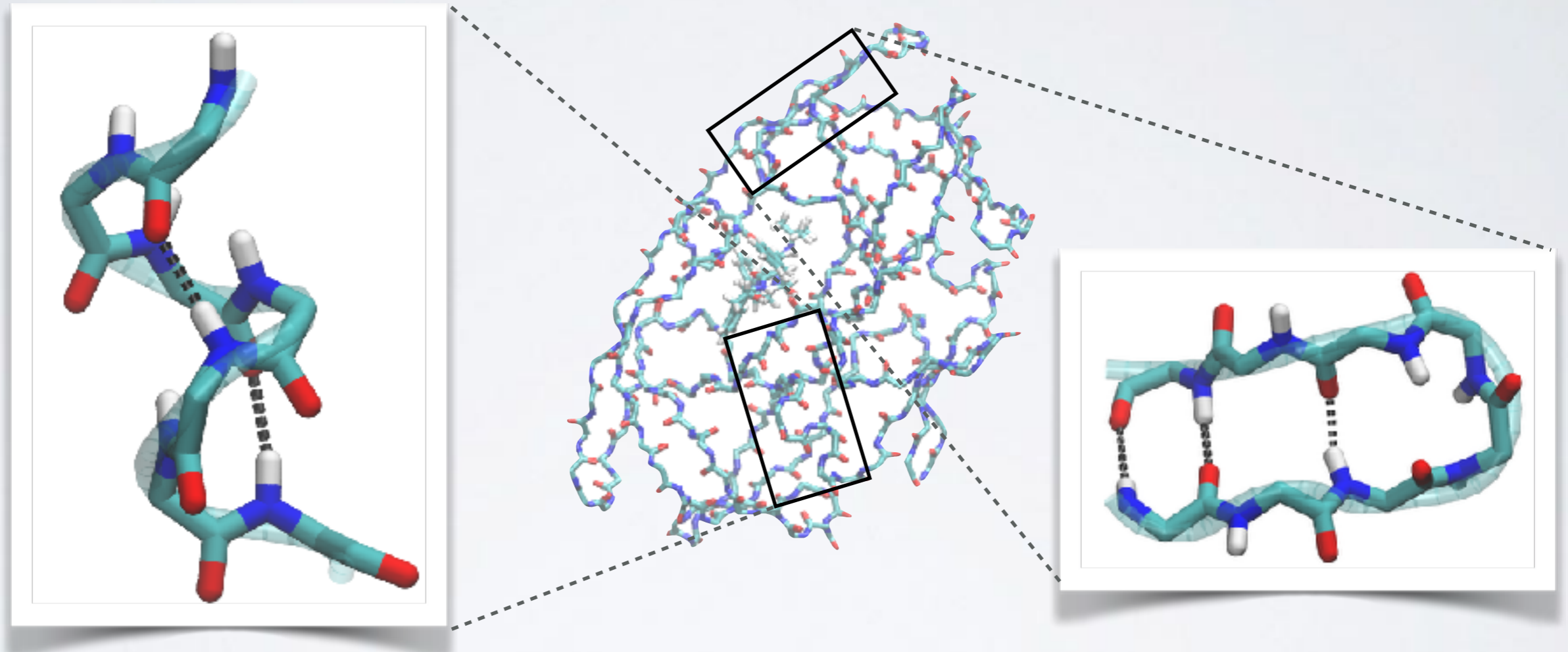
- Proteins closely interact with water

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

- Due to their large size and complexity it is often hard to see whats important in the structure
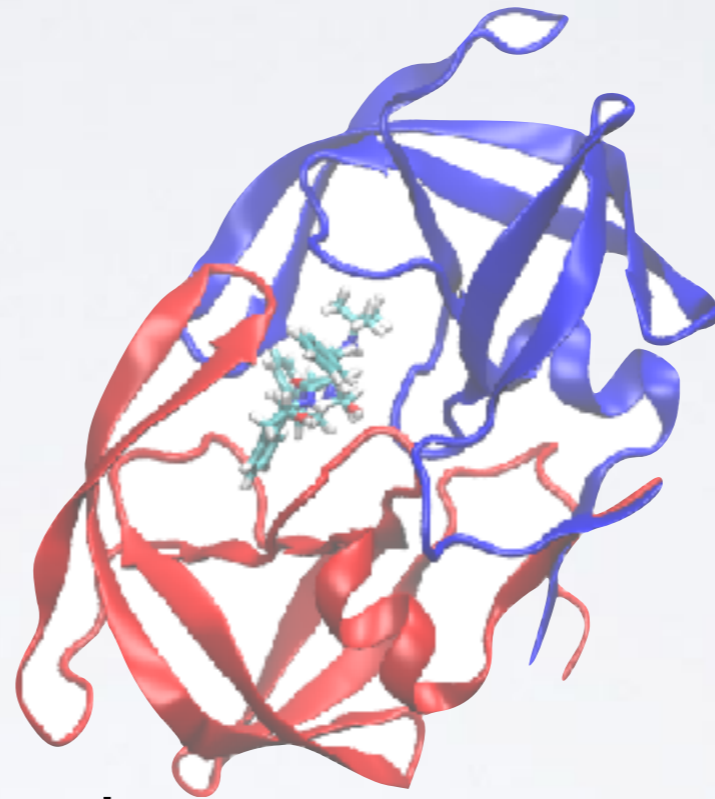
- Backbone or main-chain representation can help trace chain topology
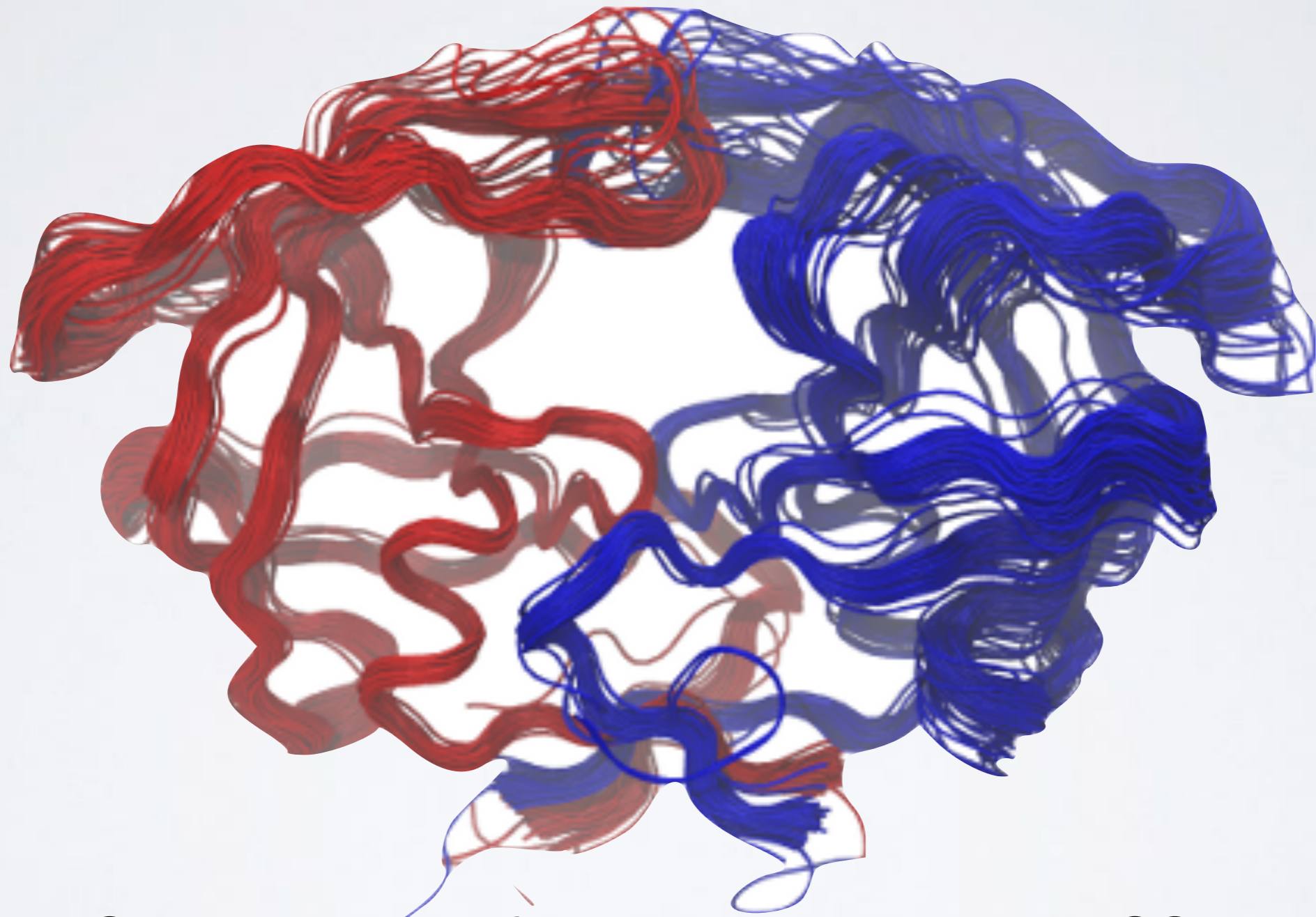
- Backbone or main-chain representation can help trace chain topology & reveal secondary structure
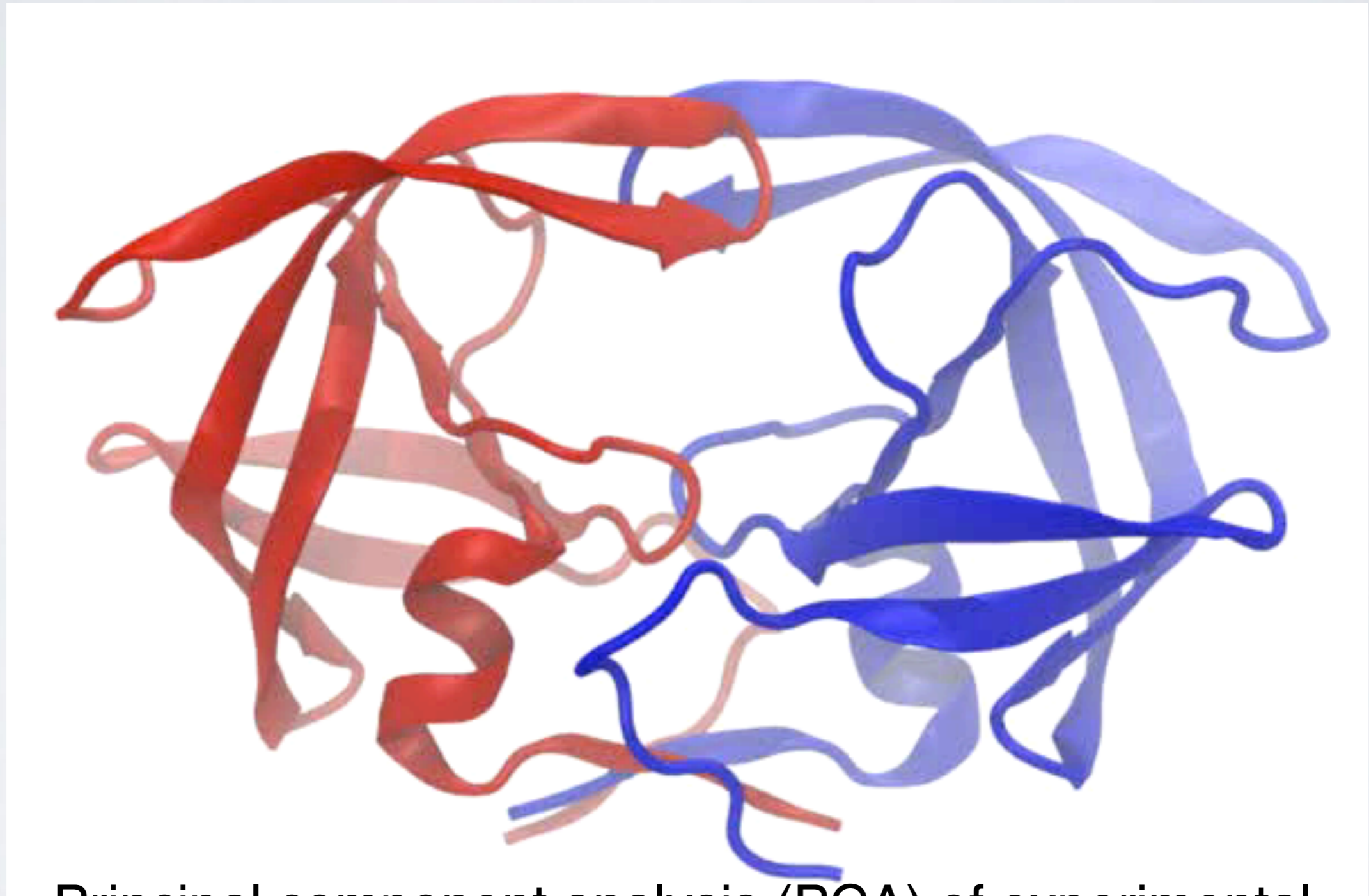
- Simplified secondary structure representations are commonly used to communicate structural details

- Now we can clearly see 2°, 3° and 4° structure

- Coiled chain of connected secondary structures

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY
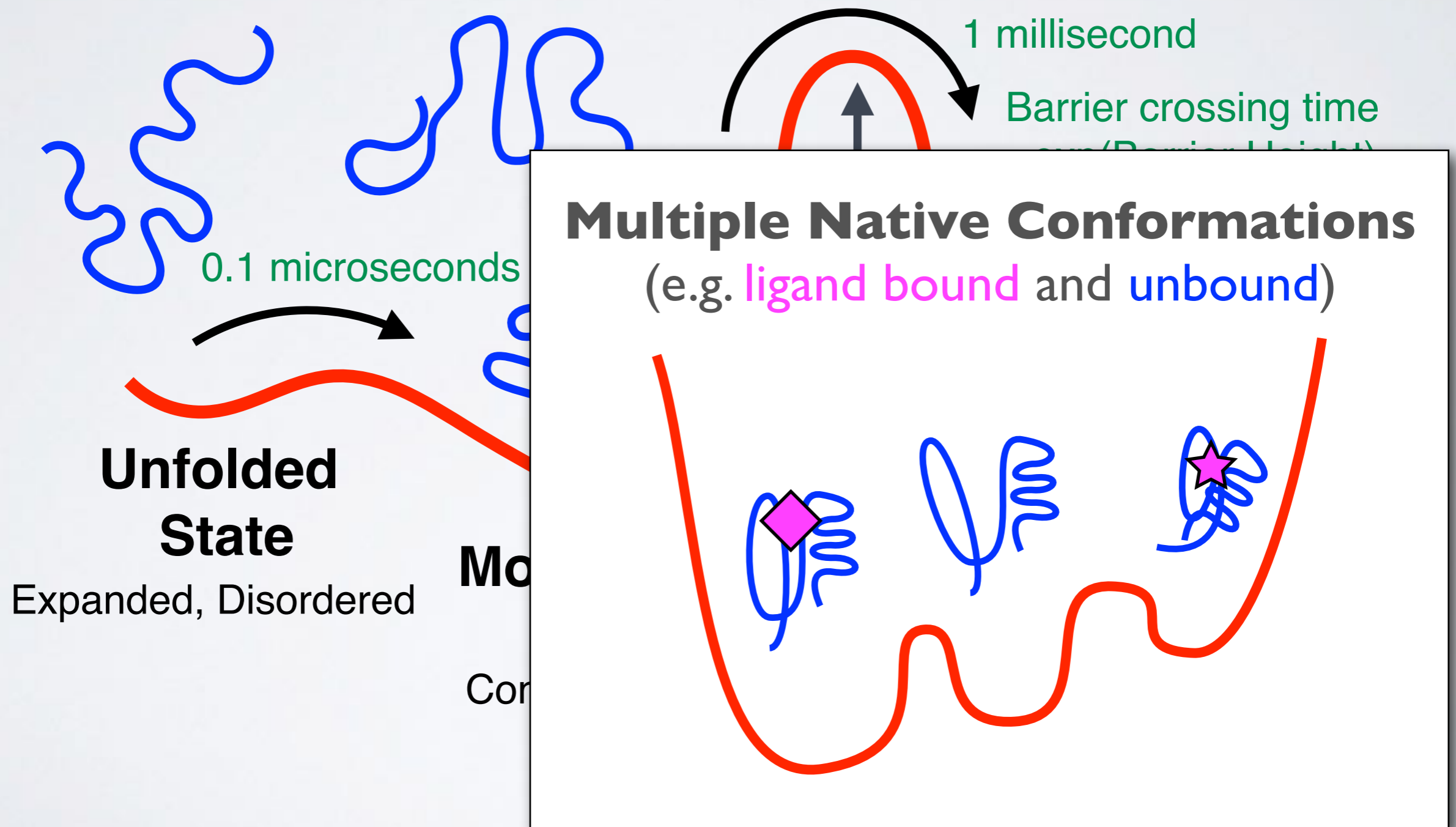

Superposition of all 482 structures in RCSB
PDB (23/09/2015)

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY

<u>P</u>rincipal <u>c</u>omponent <u>a</u>nalysis (PCA) of experimental structures

KEY CONCEPT: **ENERGY LANDSCAPE**

1 millisecond

Barrier crossing time

0.1 microseconds

Unfolded State

Expanded, Disordered

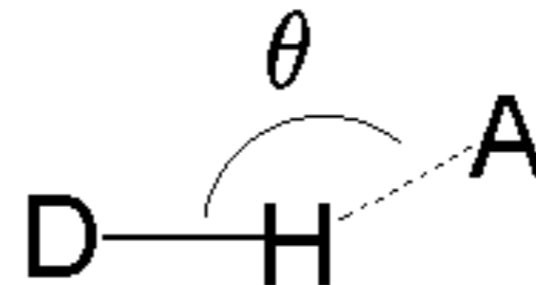**Multiple Native Conformations**
(e.g. ligand bound and unbound)

Mo

Con

Key forces affecting structure:

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

Hydrogen-bond donor   Hydrogen-bond acceptor

$N—H\text{------}N$
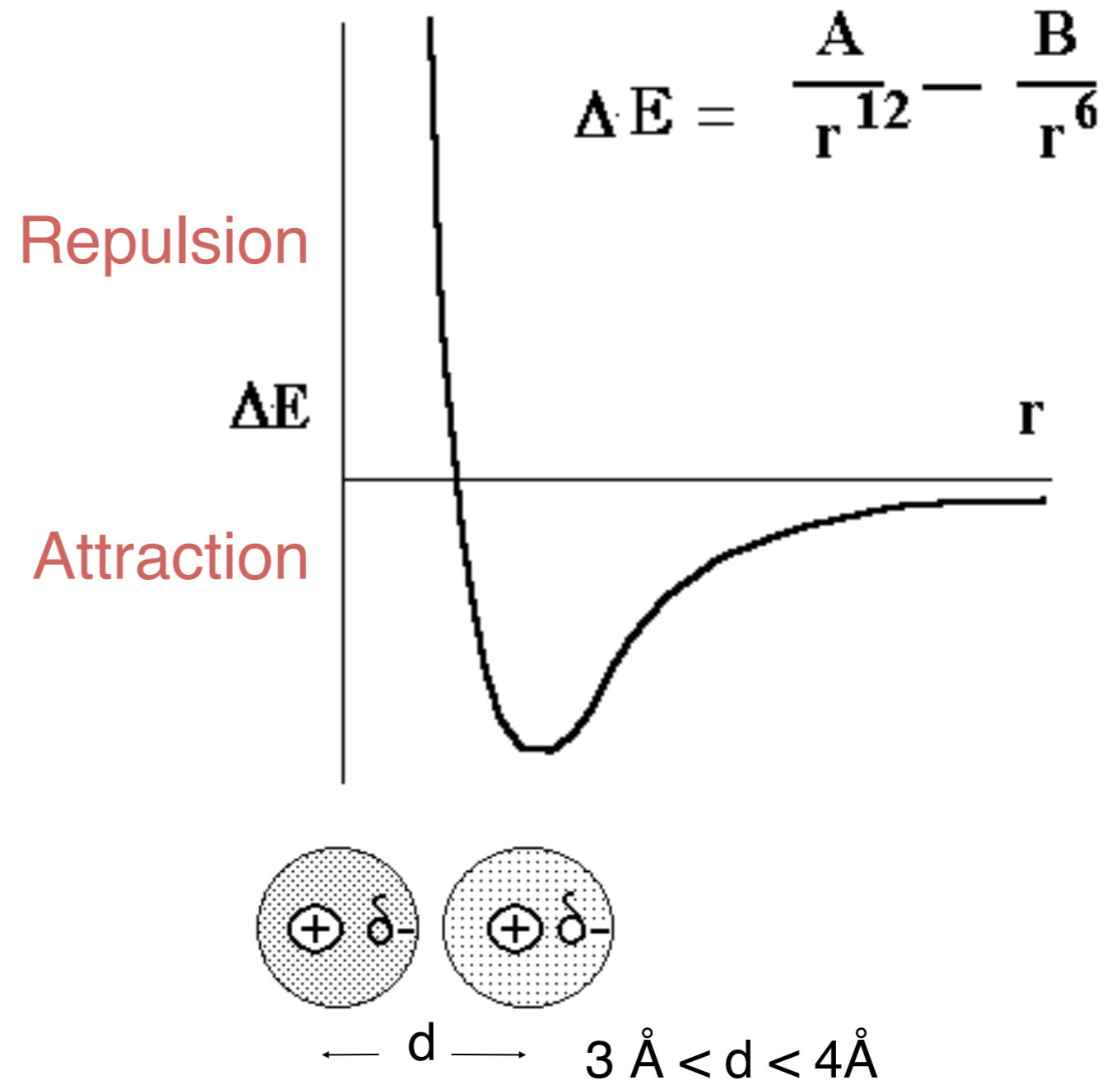$\delta^- \quad \delta^+ \qquad \delta^-$

$N—H\text{------}O$

$O—H\text{------}N$

$O—H\text{------}O$

$\longleftarrow \; d \; \longrightarrow$

θ

$D—H\;\;\;\;\;A$

$2.6\,\text{Å} < d < 3.1\text{Å}$

$150° < \theta < 180°$

Key forces affecting structure:

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

$$\Delta E = \frac{A}{r^{12}} - \frac{B}{r^6}$$

Repulsion

ΔE

Attraction

r

$\oplus \delta- \quad \oplus \delta-$

$\longleftarrow d \longrightarrow$ $3\,\text{Å} < d < 4\text{Å}$

Key forces affecting structure:

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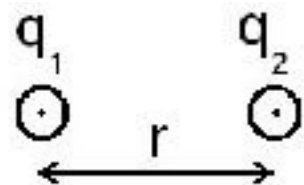- Van der Waals
- Electrostatics
- Hydrophobicity

$d = 2.8\ \text{Å}$
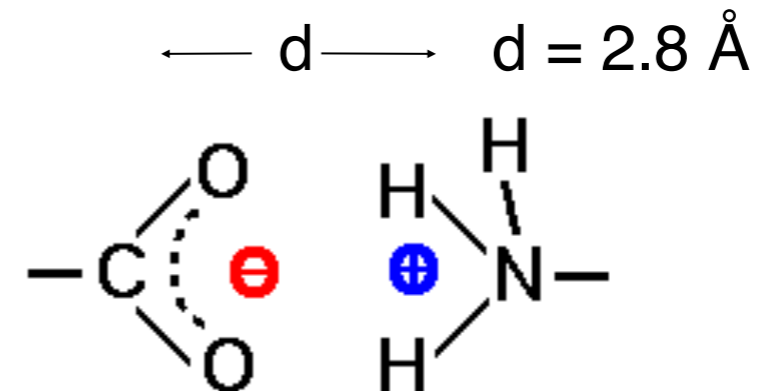
carboxyl group and amino group

(some time called IONIC BONDs or SALT BRIDGEs)

Coulomb's law

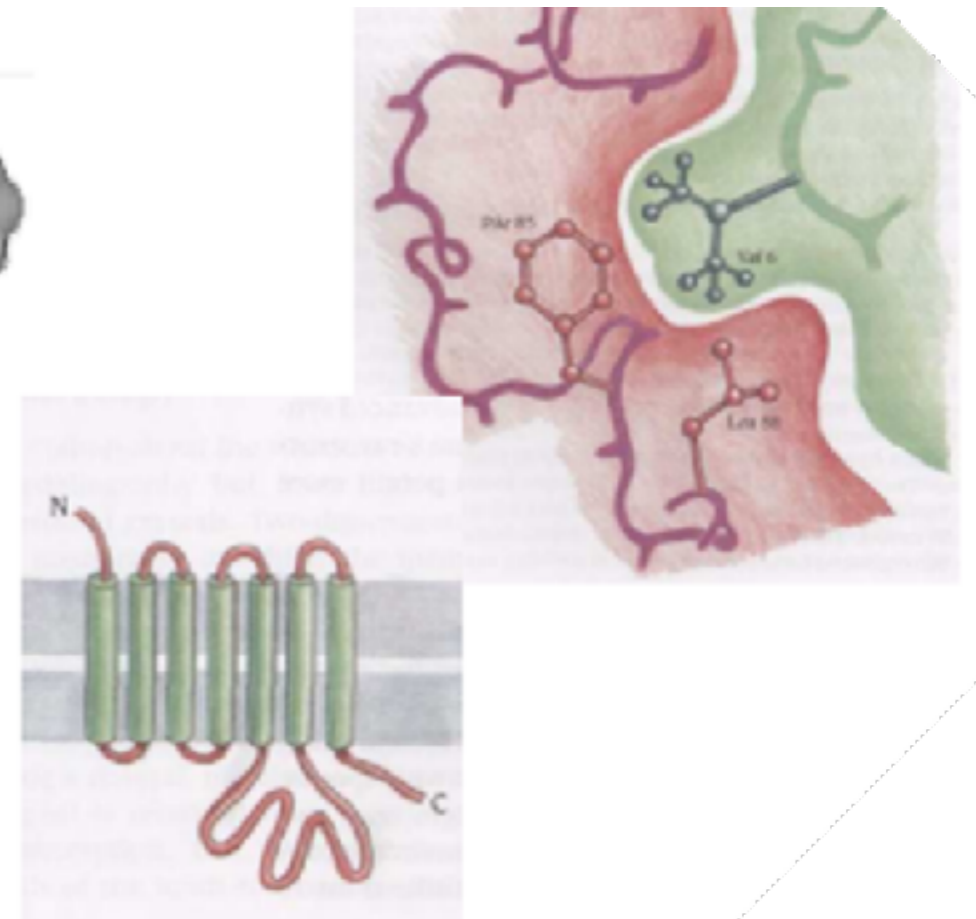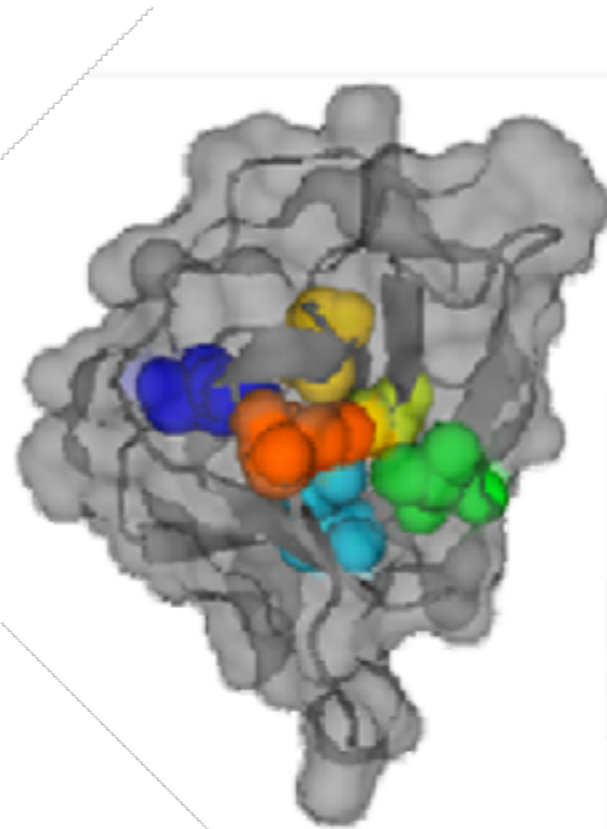$$E = \frac{K\, q_1\, q_2}{D\, r}$$

E = Energy
k = constant
D = Dielectric constant (vacuum = 1; H_2O = 80)
q_1 & q_2 = 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.

Today's Menu

- **Overview of structural bioinformatics**
  - Motivations, goals and challenges

- **Fundamentals of protein structure**
  - Structure composition, form and forces

- **Representing, interpreting & modeling protein structure**
  - Visualizing & interpreting protein structures
  - Analyzing protein structures
  - Modeling energy as a function of structure

Today's Menu

- **Overview of structural bioinformatics**
  - Motivations, goals and challenges

- **Fundamentals of protein structure**
  - Structure composition, form and forces

- **Representing, interpreting & modeling protein structure**
  - Visualizing & interpreting protein structures
  - Analyzing protein structures
  - Modeling energy as a function of structure

Hand-on time!

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

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

SIDE-NOTE: **PDB FILE FORMAT**

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

SIDE-NOTE: **PDB FILE FORMAT**

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

Today's Menu

- **Overview of structural bioinformatics**
  - Motivations, goals and challenges

- **Fundamentals of protein structure**
  - Structure composition, form and forces

- **Representing, interpreting & modeling protein structure**
  - Visualizing and interpreting protein structures
  - Analyzing protein structures
  - Modeling energy as a function of structure

Hand-on time!

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

Focus on **section 4** & **5**

Today's Menu

- **Overview of structural bioinformatics**
  - Motivations, goals and challenges

- **Fundamentals of protein structure**
  - Structure composition, form and forces

- **Representing, interpreting & modeling protein structure**
  - Visualizing and interpreting protein structures
  - Analyzing protein structures
  - Modeling energy as a function of structure

KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**
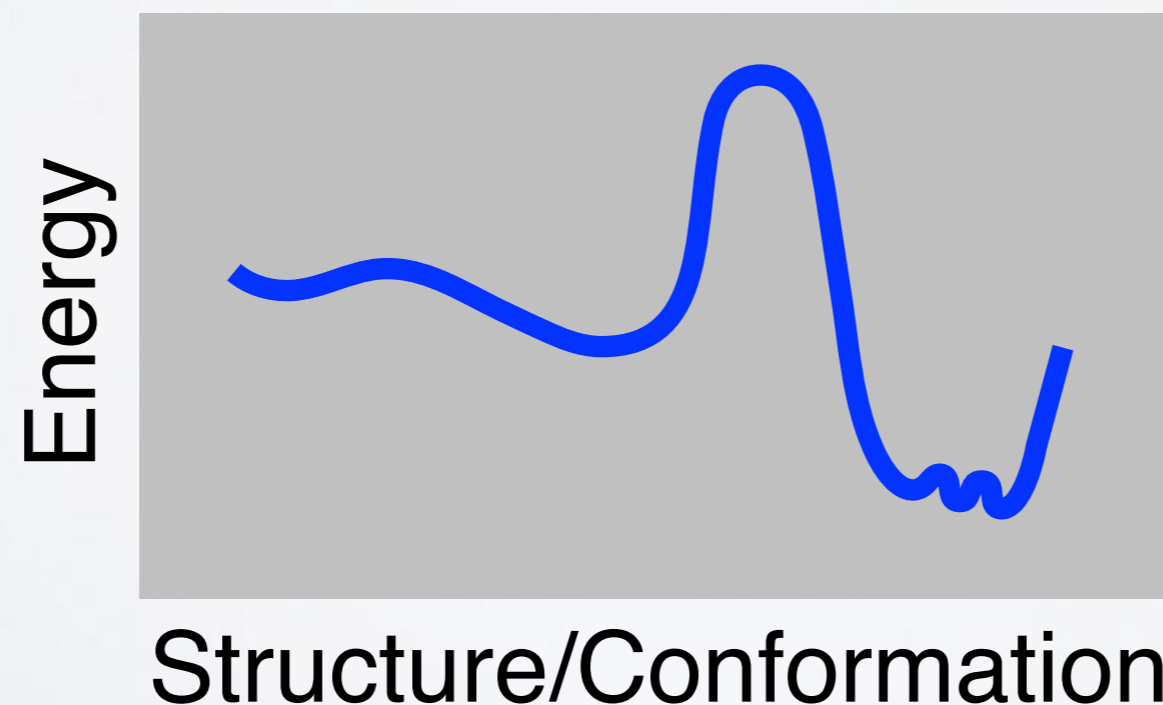
Two main approaches:
(1). **Physics-Based**
(2). **Knowledge-Based**

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

Two main approaches:

(1). **Physics**-Based

(2). **Knowledge**-Based

Energy vs. Structure/Conformation

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

Two main approaches:

(1). **Physics-Based**

(2). **Knowledge-Based**

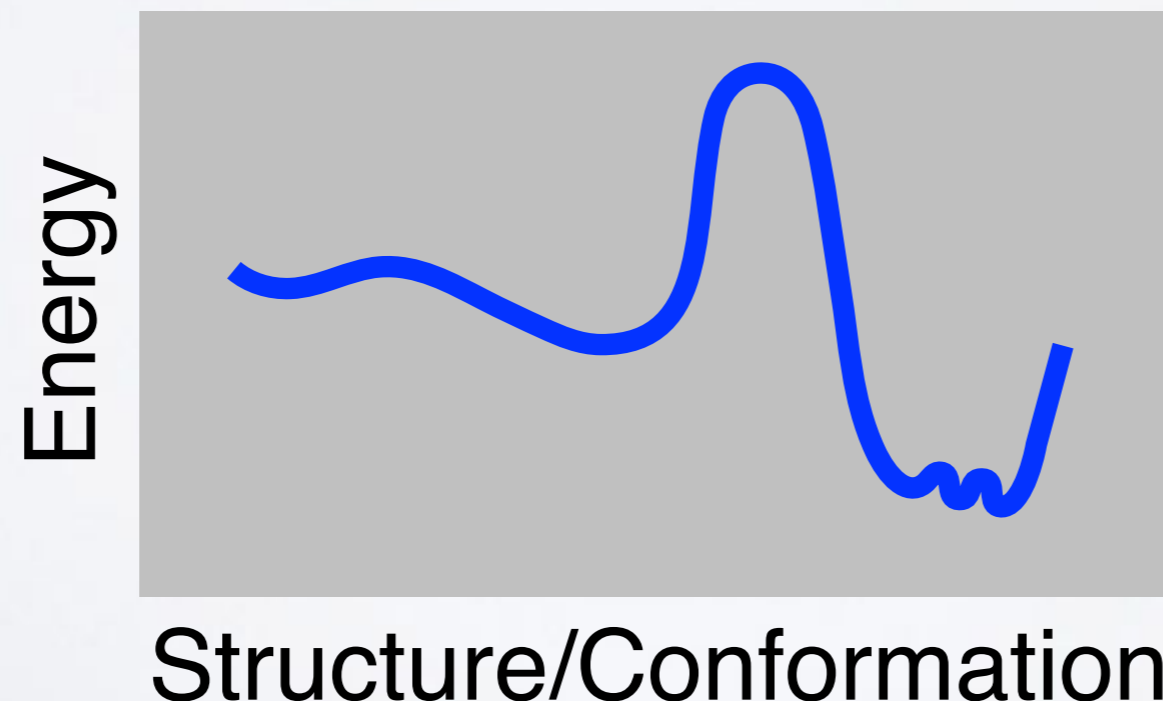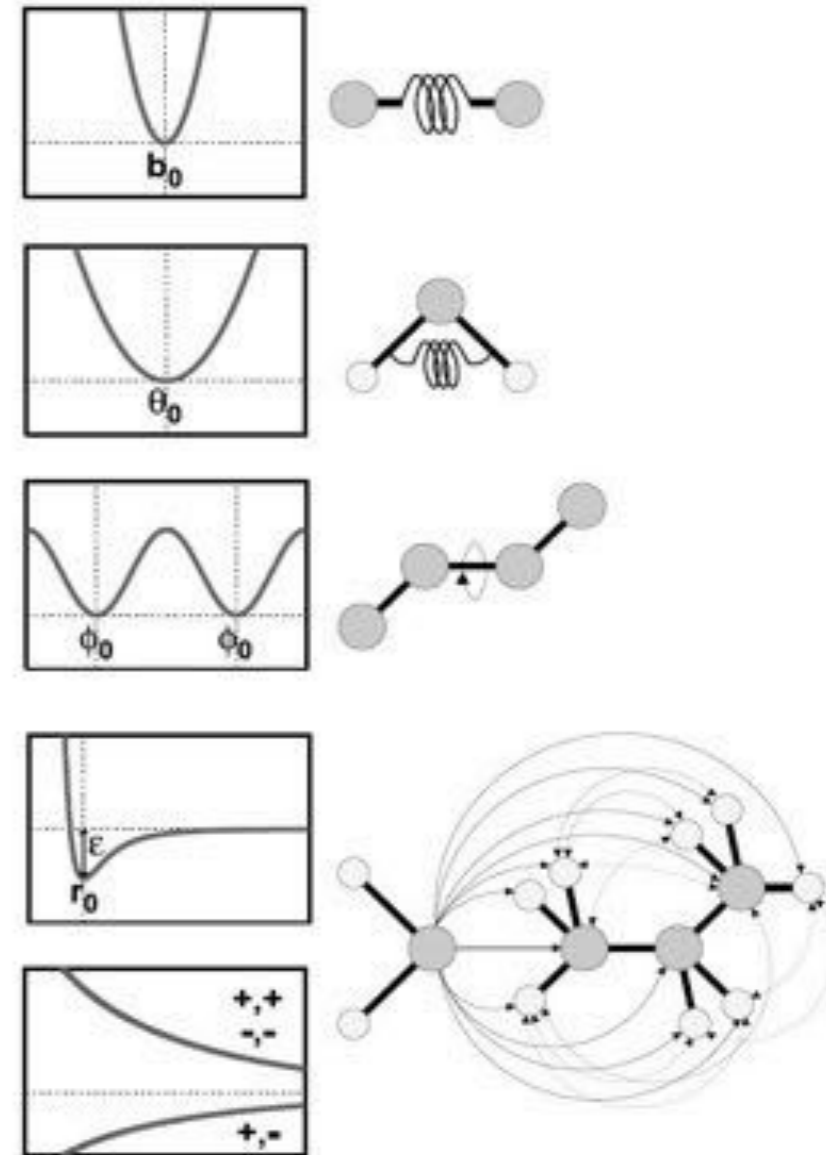# PHYSICS-BASED POTENTIALS
ENERGY TERMS FROM PHYSICAL THEORY

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i \phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$
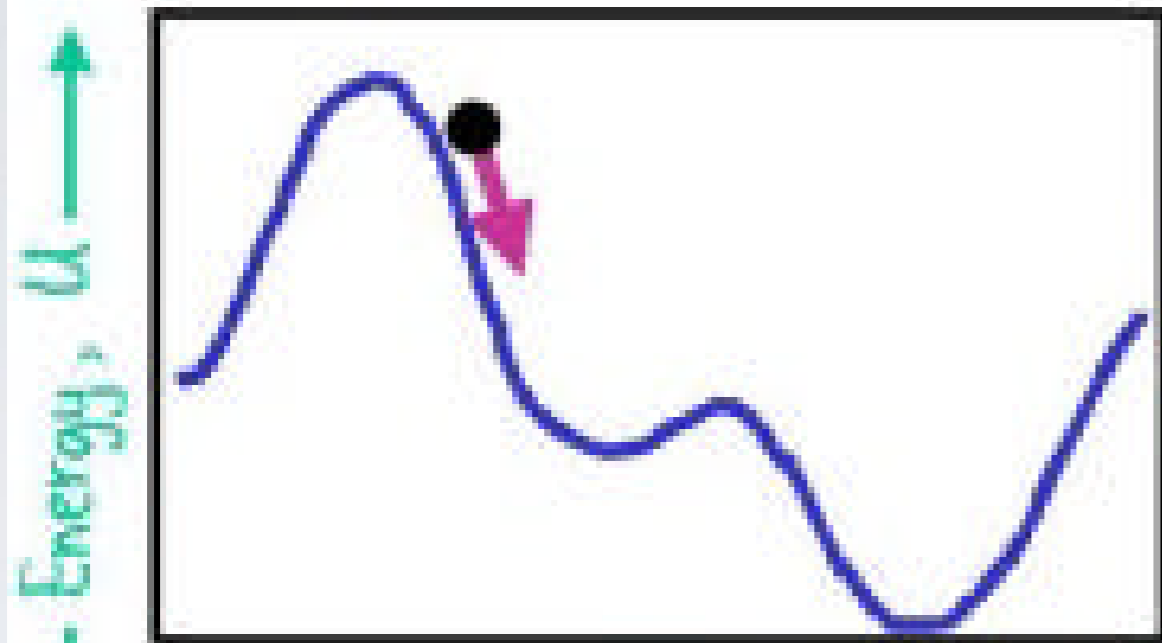
U_{bond} = oscillations about the equilibrium bond length
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)
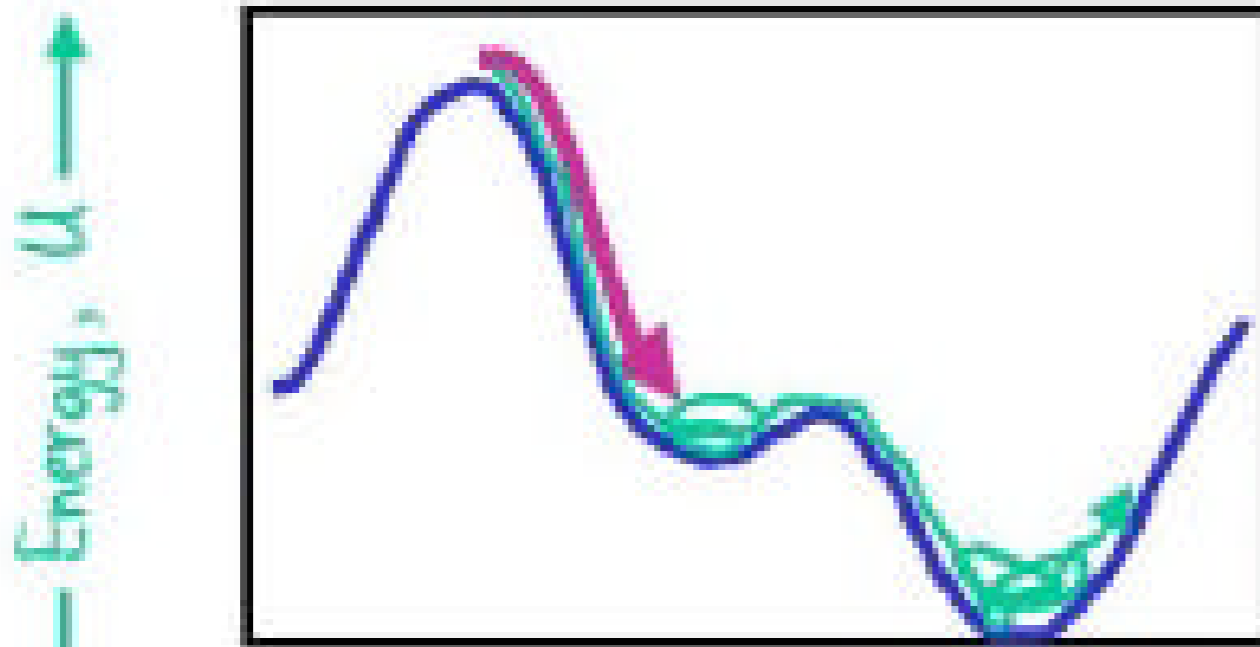
TOTAL POTENTIAL ENERGY

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

- 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.
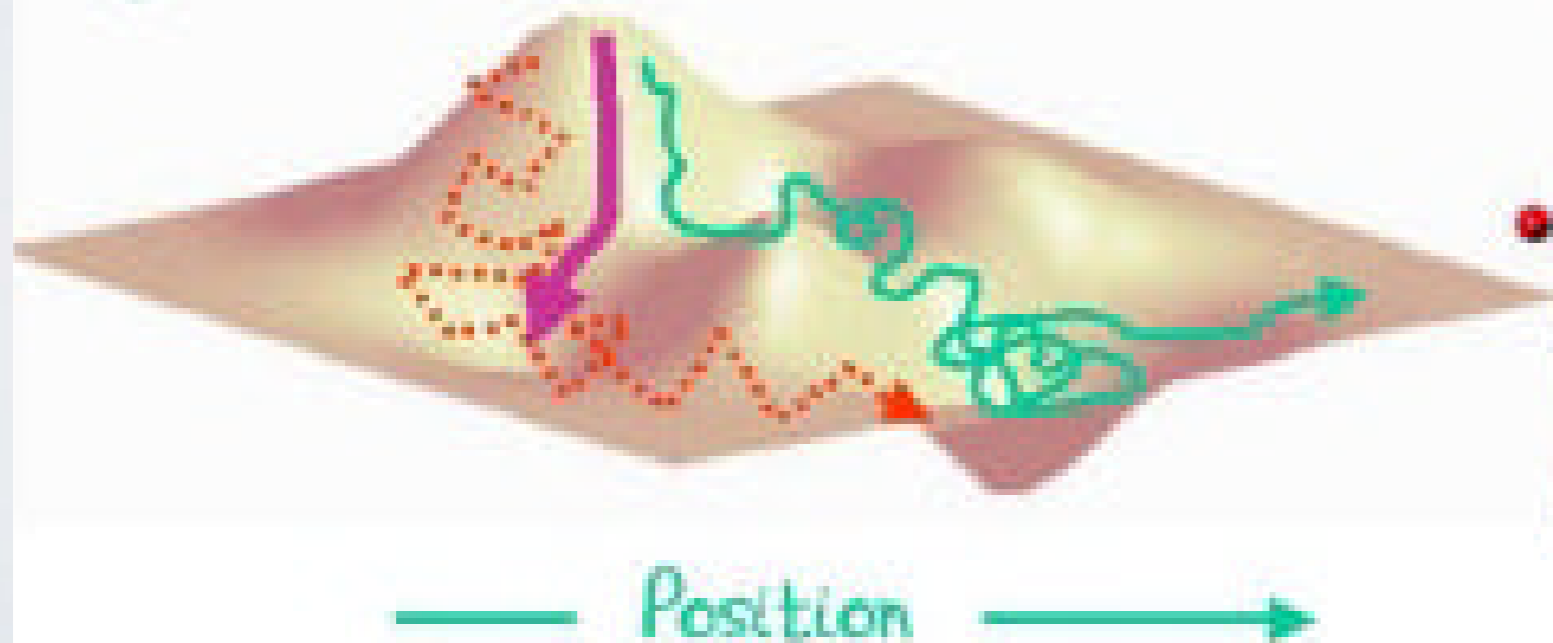
Slide Credit: Michael Levitt

MOVING OVER THE ENERGY SURFACE

- 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
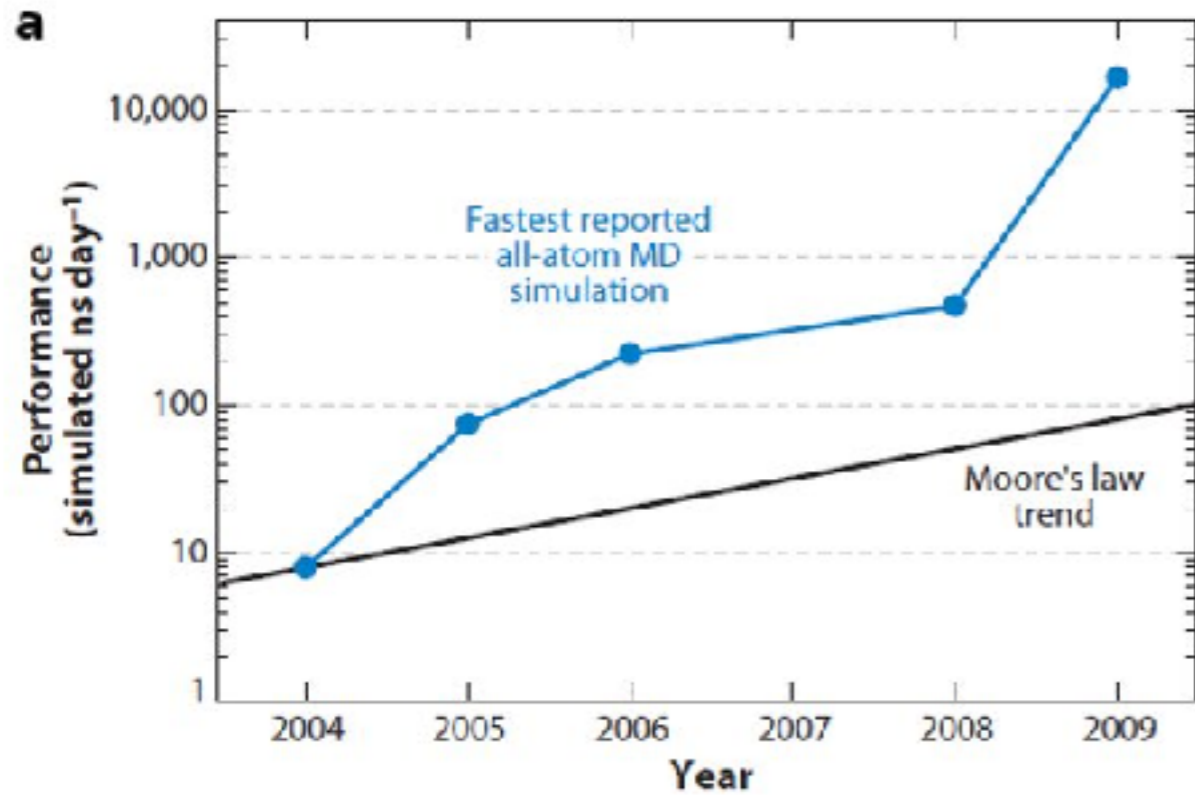
Moore's law: hardware improving

SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER

SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER
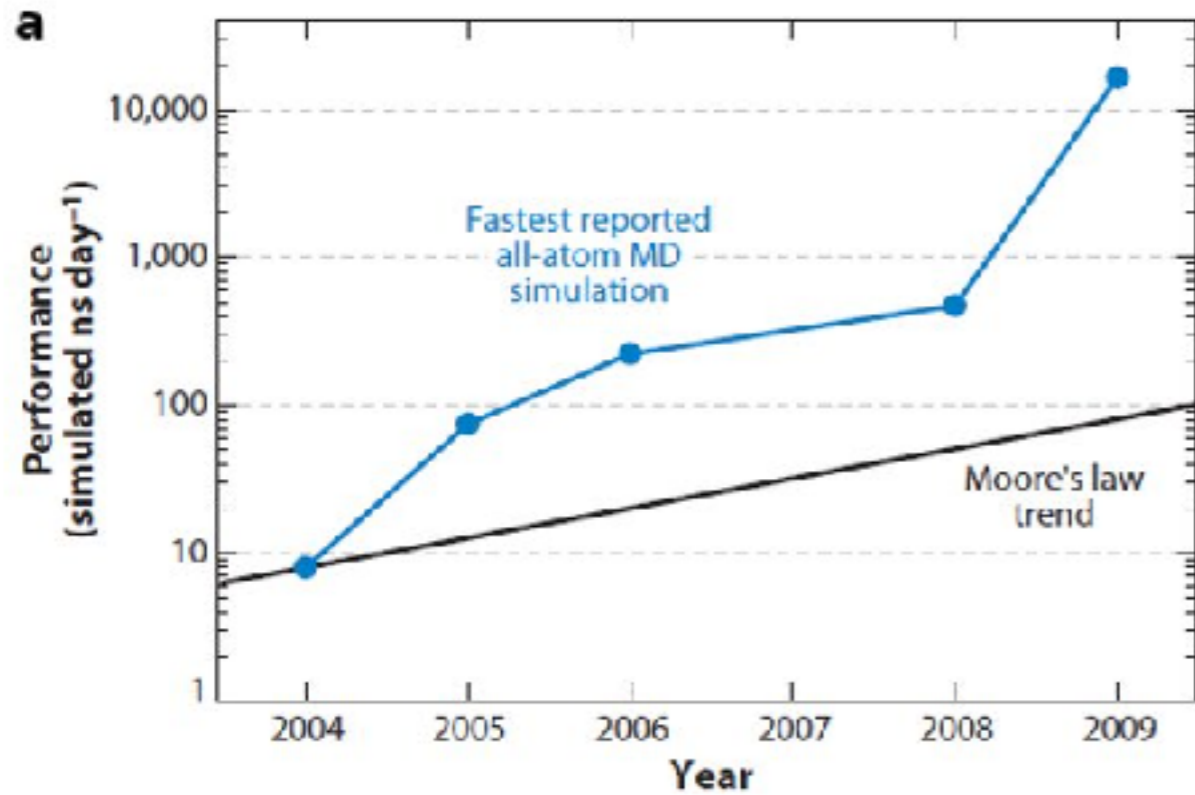
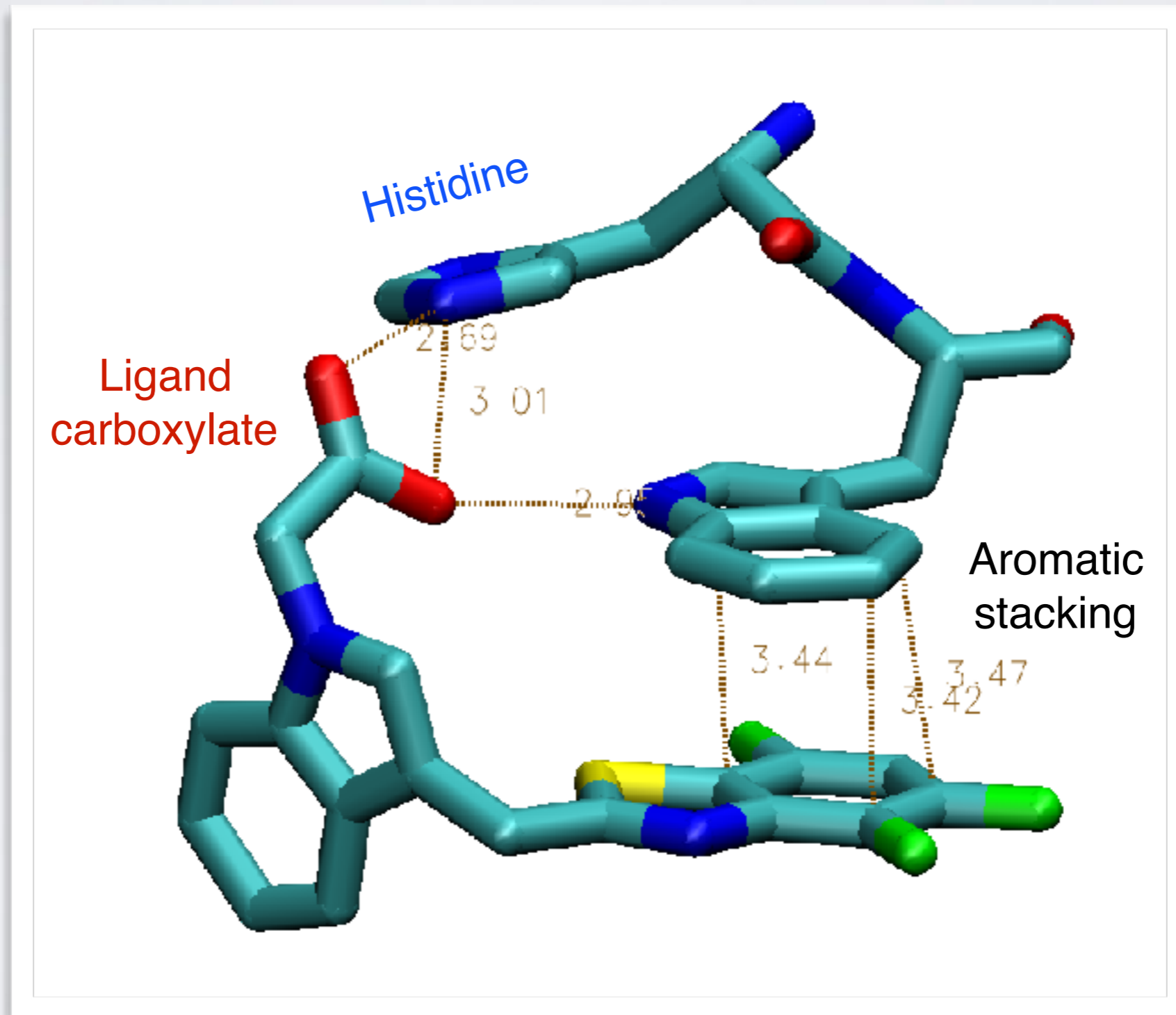# KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**
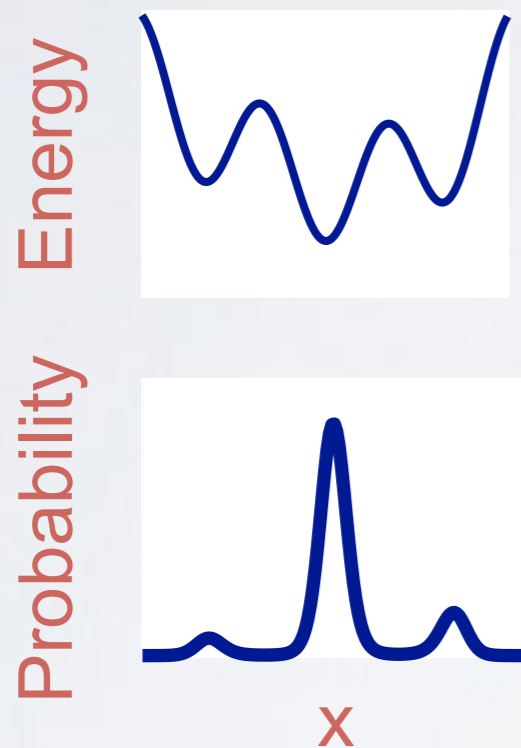
Two main approaches:

(1). **Physics-Based**

(2). **Knowledge-Based**

ENERGY DETERMINES **PROBABILITY** (STABILITY)

Basic idea: Use probability as a proxy for energy

Boltzmann:

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

Inverse Boltzmann:

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

Example: ligand carboxylate O to protein histidine N

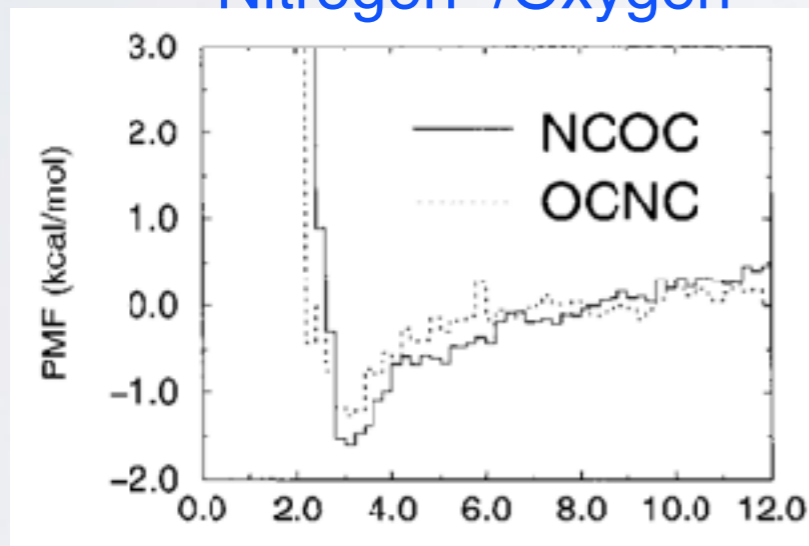Find all protein-ligand structures in the PDB with a ligand carboxylate O
  1. For each structure, histogram the distances from O to every histidine N
  2. Sum the histograms over all structures to obtain $p(r_{O\text{-}N})$
  3. Compute $E(r_{O\text{-}N})$ from $p(r_{O\text{-}N})$

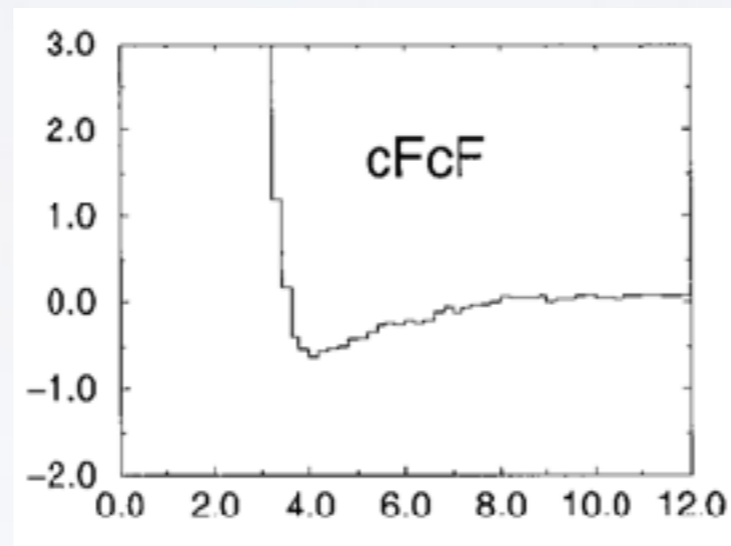# KNOWLEDGE-BASED DOCKING POTENTIALS

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

A few types of atom pairs, out of several hundred total
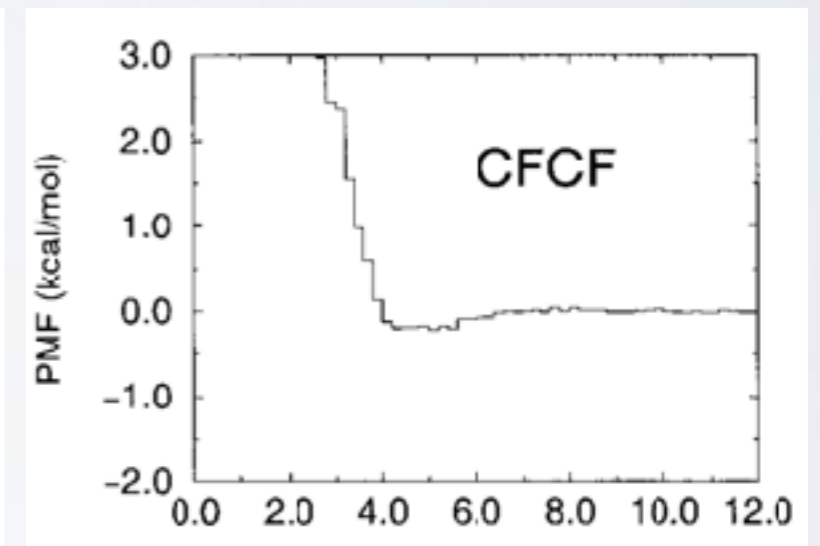
Atom-atom distance (Angstroms)

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

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)
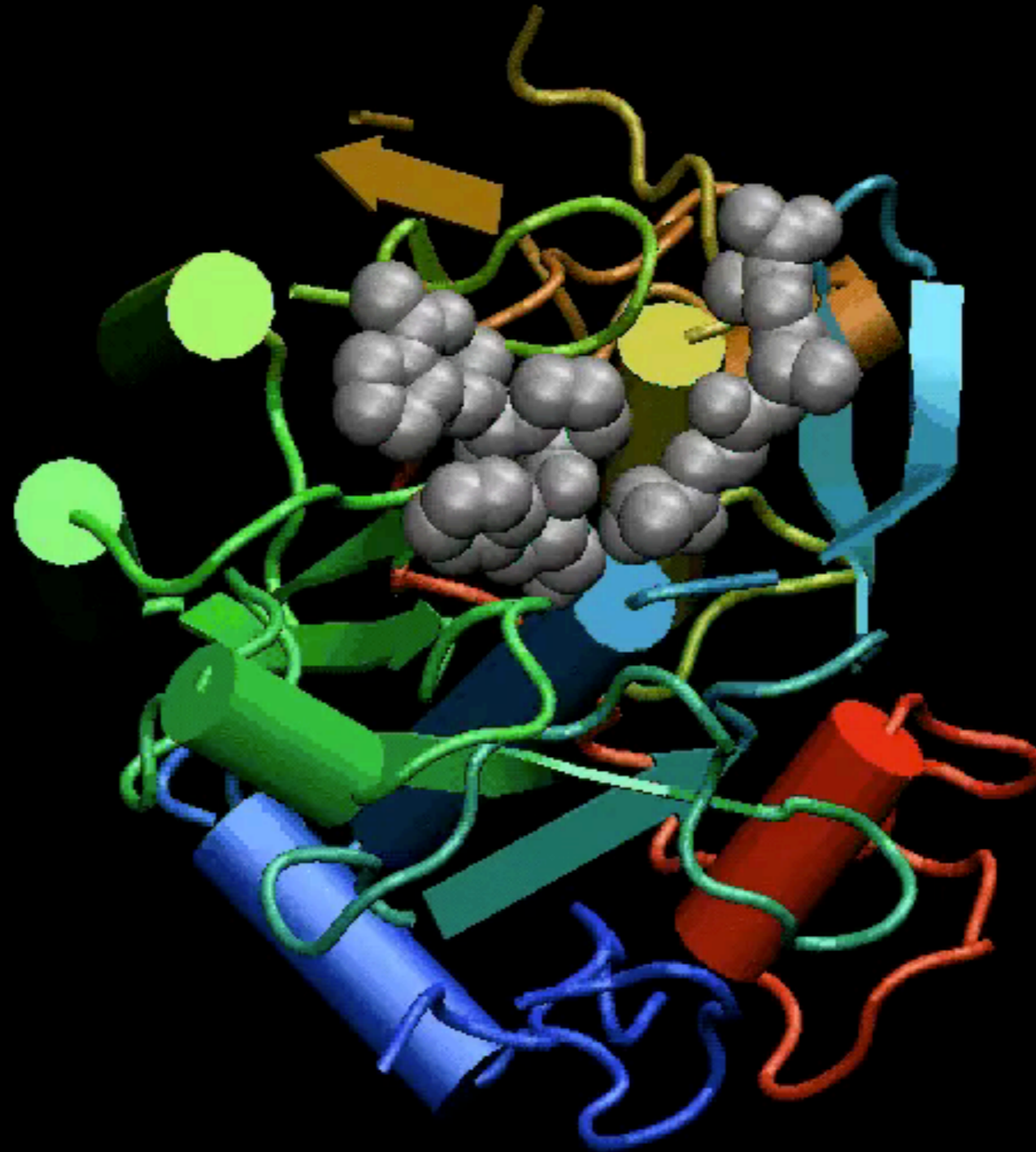
Hand-on time!

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

Focus on **section 6** & **7**

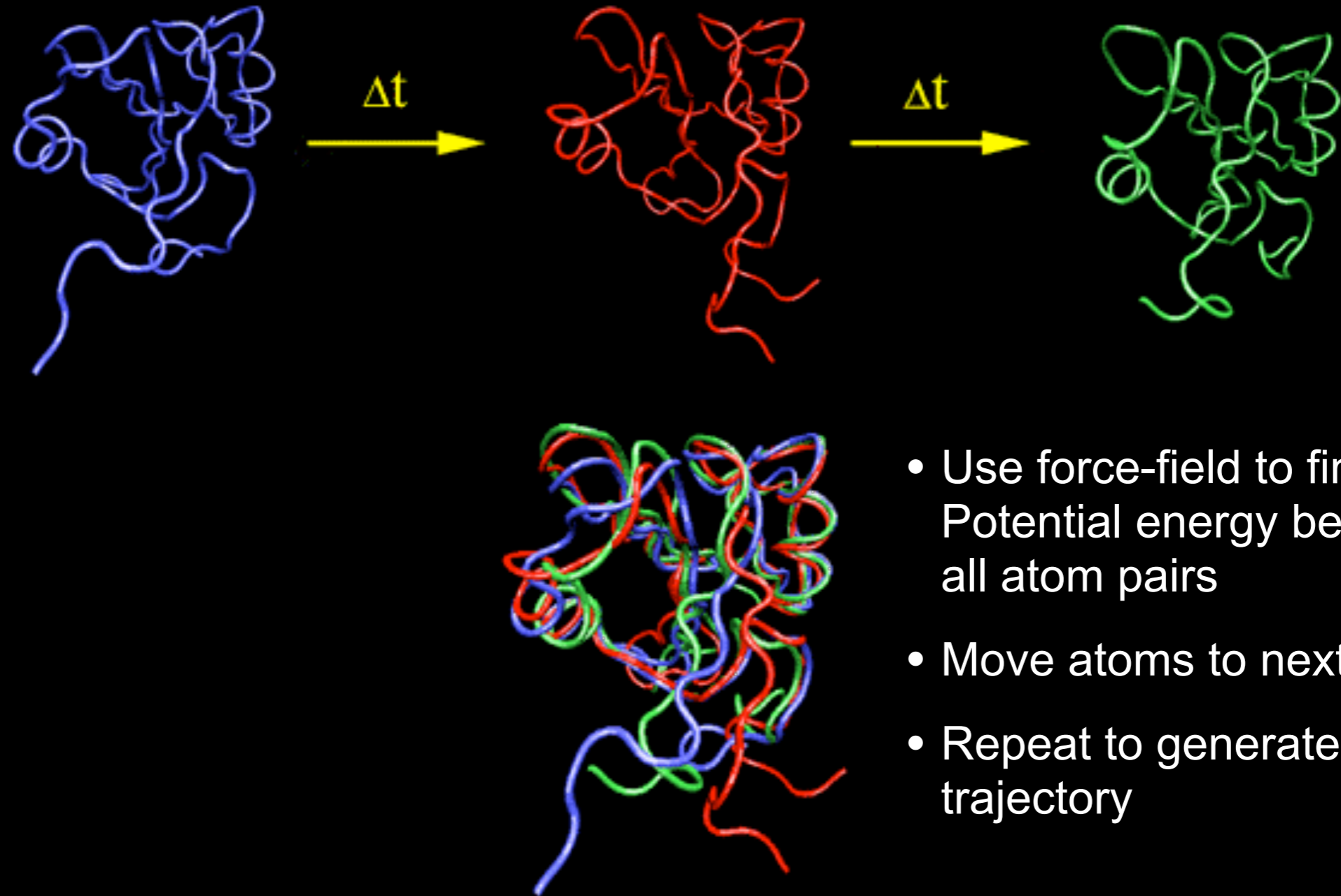# NMA models the protein as a network of elastic strings

Proteinase K

Optional:
Stop here for Today!

[Muddy Point Assessment]

PREDICTING FUNCTIONAL DYNAMICS

- **Proteins are <u>intrinsically flexible</u> 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 <u>mapping of structure to function</u>**
  - **<u>Molecular dynamics</u>** (MD) and **normal mode analysis** (NMA) are two major methods for predicting and characterizing molecular motions and their properties

MOLECULAR DYNAMICS SIMULATION

- Use force-field to find Potential energy between all atom pairs

- Move atoms to next state

- Repeat to generate trajectory

McCammon, Gelin & Karplus, *Nature* (1977)
[See: https://www.youtube.com/watch?v=ui1ZysMFcKk]

▷ Divide **time** into discrete (~1fs) **time steps** (**Δt**)
(for integrating equations of motion, see below)

▷ Divide **time** into discrete (~1fs) **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

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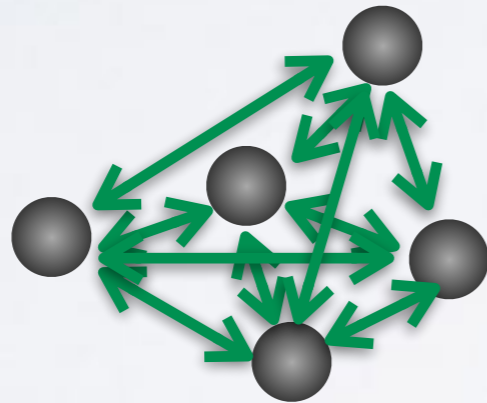
Empirical force field

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

▷ Divide **time** into discrete (~1fs) **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

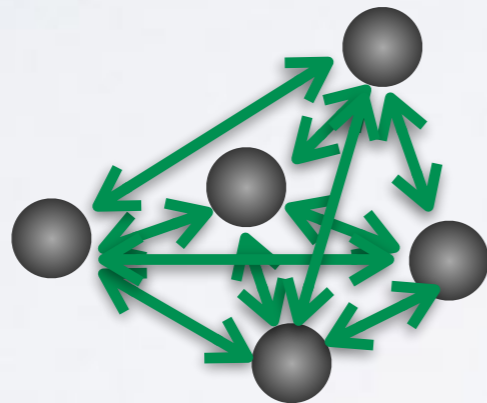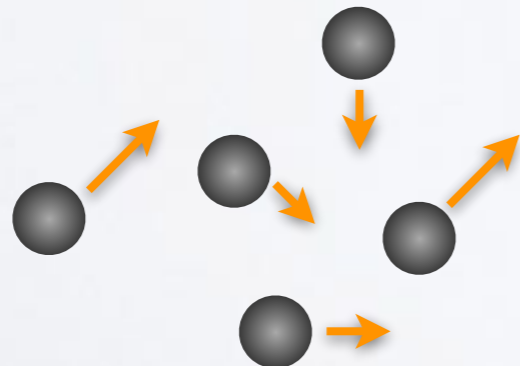$$m_i \frac{d^2}{dt^2}\vec{R}_i = -\vec{\nabla}_i E(\vec{R})$$

Empirical force field

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

▷ Use the forces to calculate **velocities** and move atoms to new **positions**
(by integrating numerically via the "leapfrog" scheme)

$$\boldsymbol{v}\left(t + \frac{\Delta t}{2}\right) = \boldsymbol{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\boldsymbol{F}(t)}{m}\Delta t$$

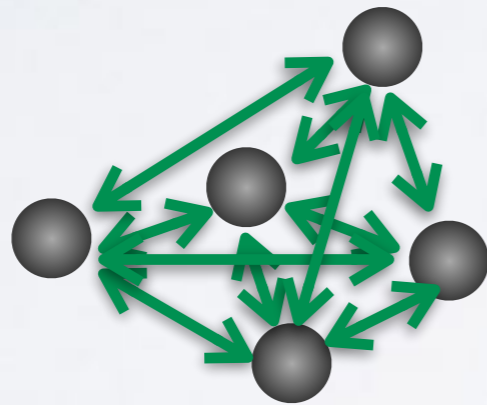$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}\left(t + \frac{\Delta t}{2}\right)\Delta t$$

BASIC ANATOMY OF A MD SIMULATION

▷ Divide **time** into discrete (~1fs) **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)
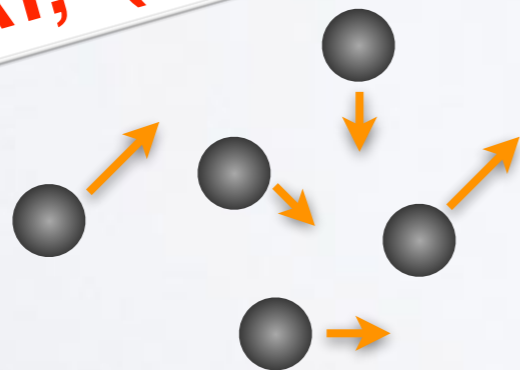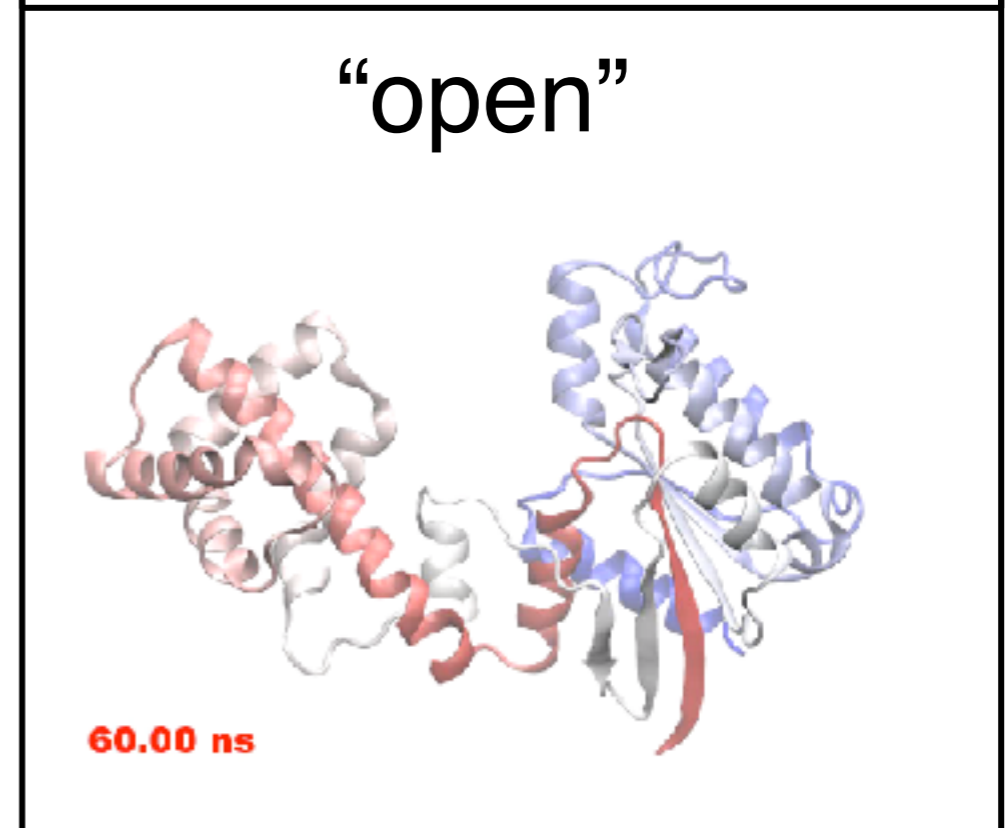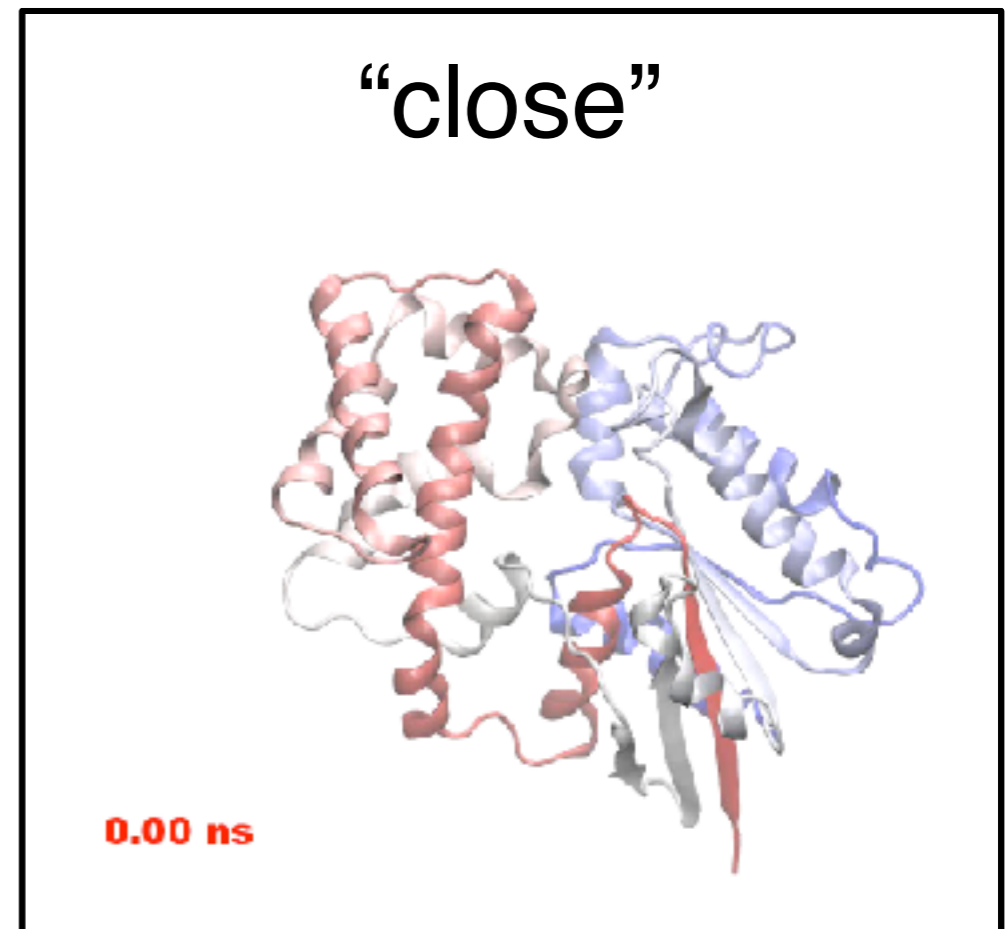
Nucleic motion described classically

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

Empirical force f...

$$E(\vec{R}) = \sum \ldots E_i(R)$$

...bonded

▷ Use the for... ...ate **velocities** and move atoms to new **positions**
(... ...g numerically via the "leapfrog" scheme)

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

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right)\Delta t$$

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

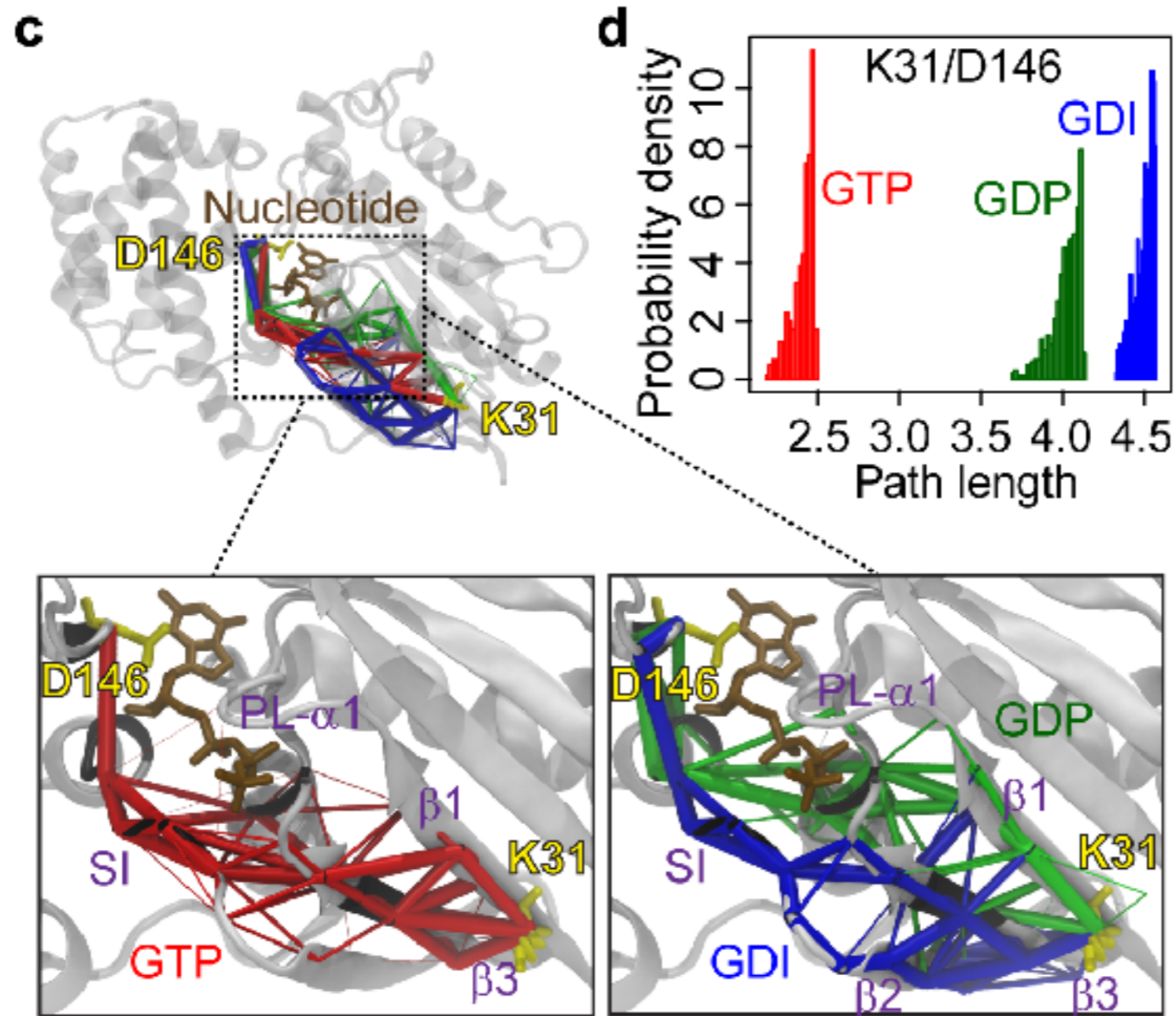# MD Prediction of Functional Motions

Accelerated MD simulation of
nucleotide-free transducin alpha subunit

0.00 ns
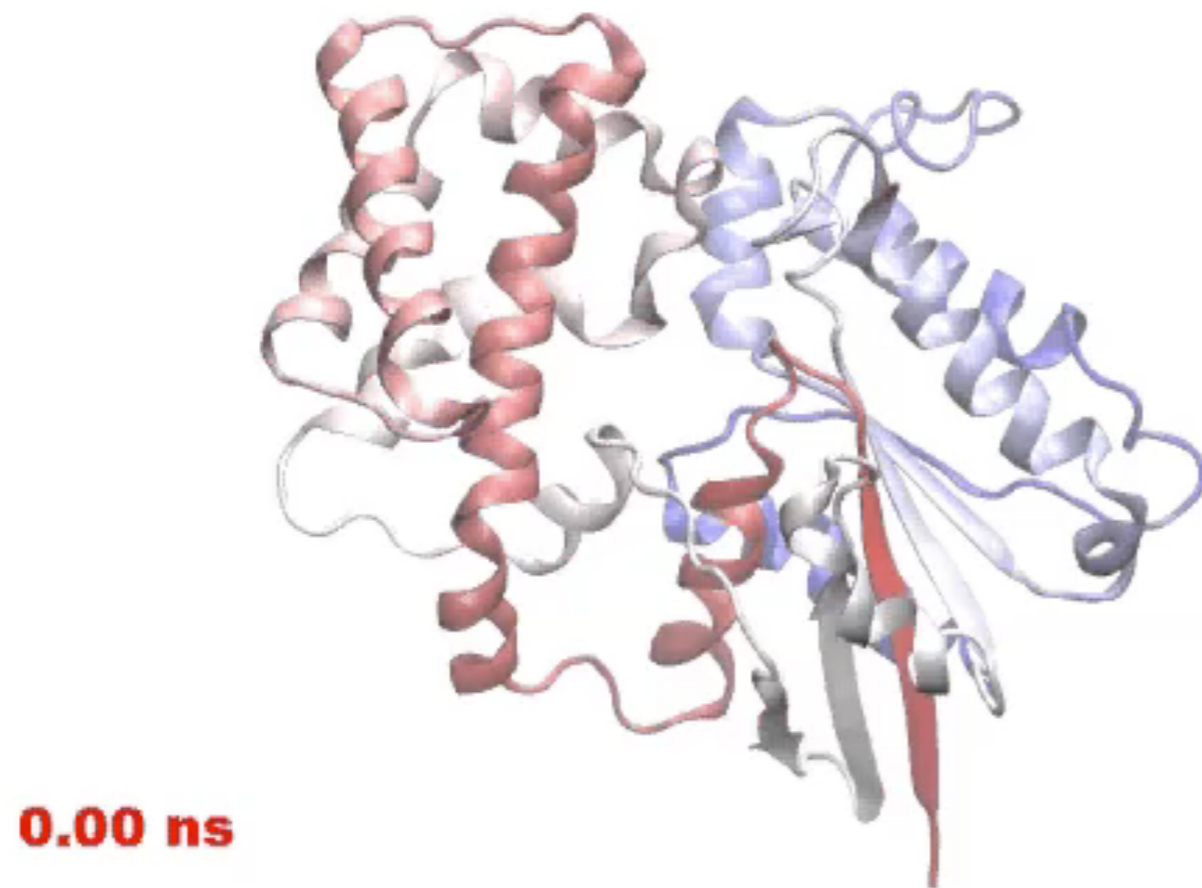
Yao and Grant, Biophys J. (2013)

"close"

0.00 ns

"open"

60.00 ns

Simulations Identify Key Residues Mediating Dynamic Activation

Yao … Grant, Journal of Biological Chemistry (2016)

EXAMPLE APPLICATION OF MOLECULAR SIMULATIONS TO GPCRS

Binding

Cell Membrane

Activation

G protein coupling

GPCR
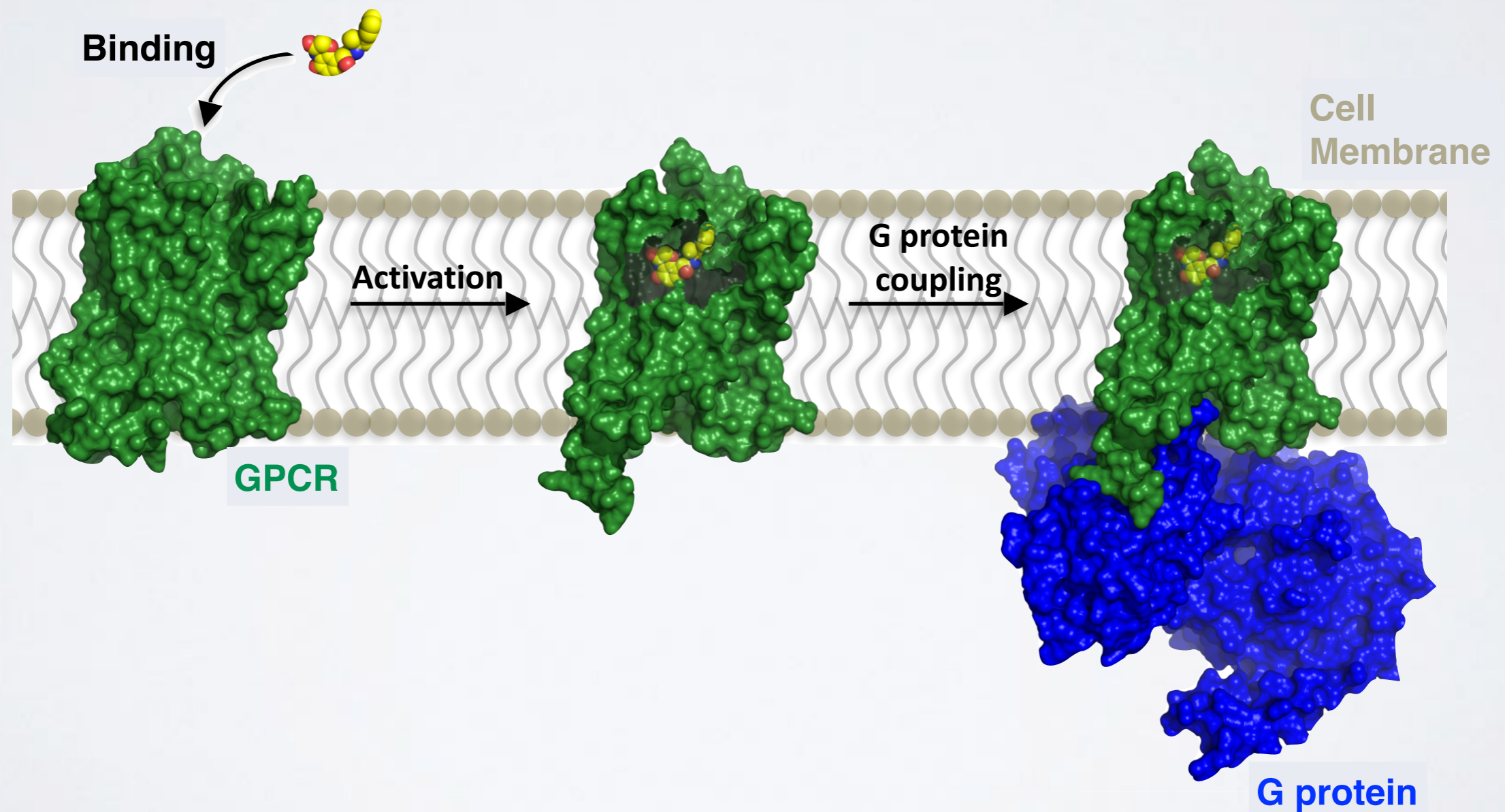
G protein

PROTEINS JUMP BETWEEN MANY, HIERARCHICALLY ORDERED "CONFORMATIONAL SUBSTATES"
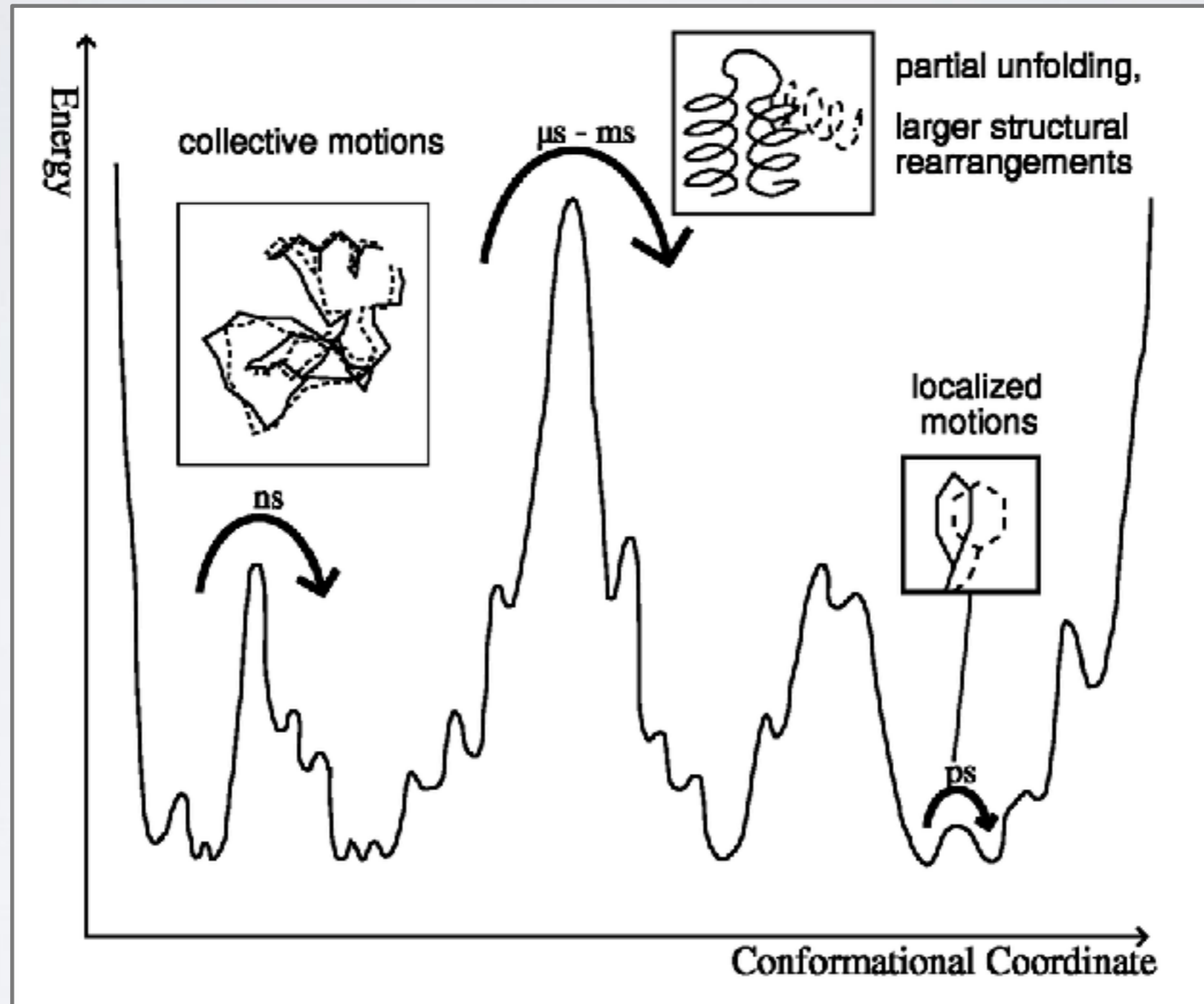
H. Frauenfelder et al., *Science* **229** (1985) 337

MOLECULAR DYNAMICS IS VERY

Example: F_1-ATPase in water (183,674 atoms) for 1 nanosecond:

    => 10^6 integration steps

    => $8.4 * 10^{11}$ floating point operations/step

    [n(n-1)/2 interactions]

      Total: $8.4 * 10^{17}$ 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
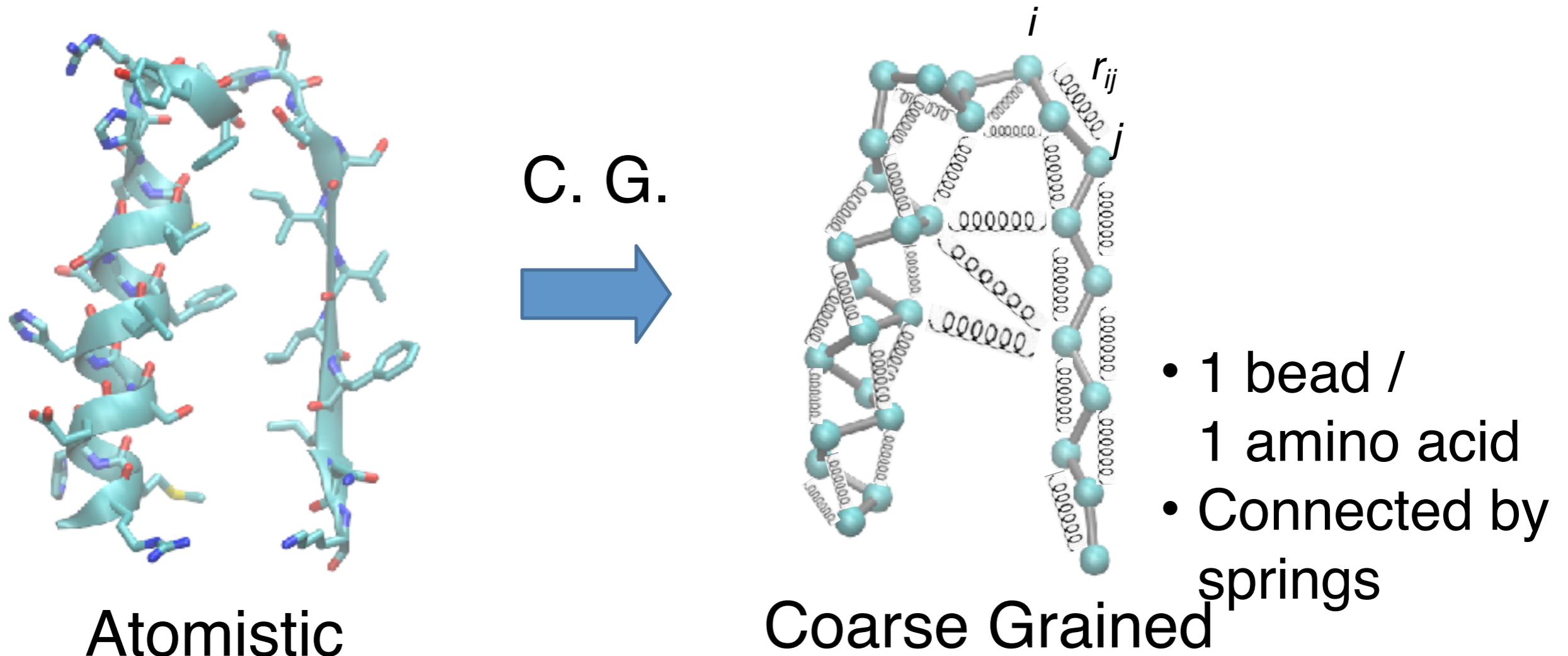
NMA models the protein as a network of elastic strings
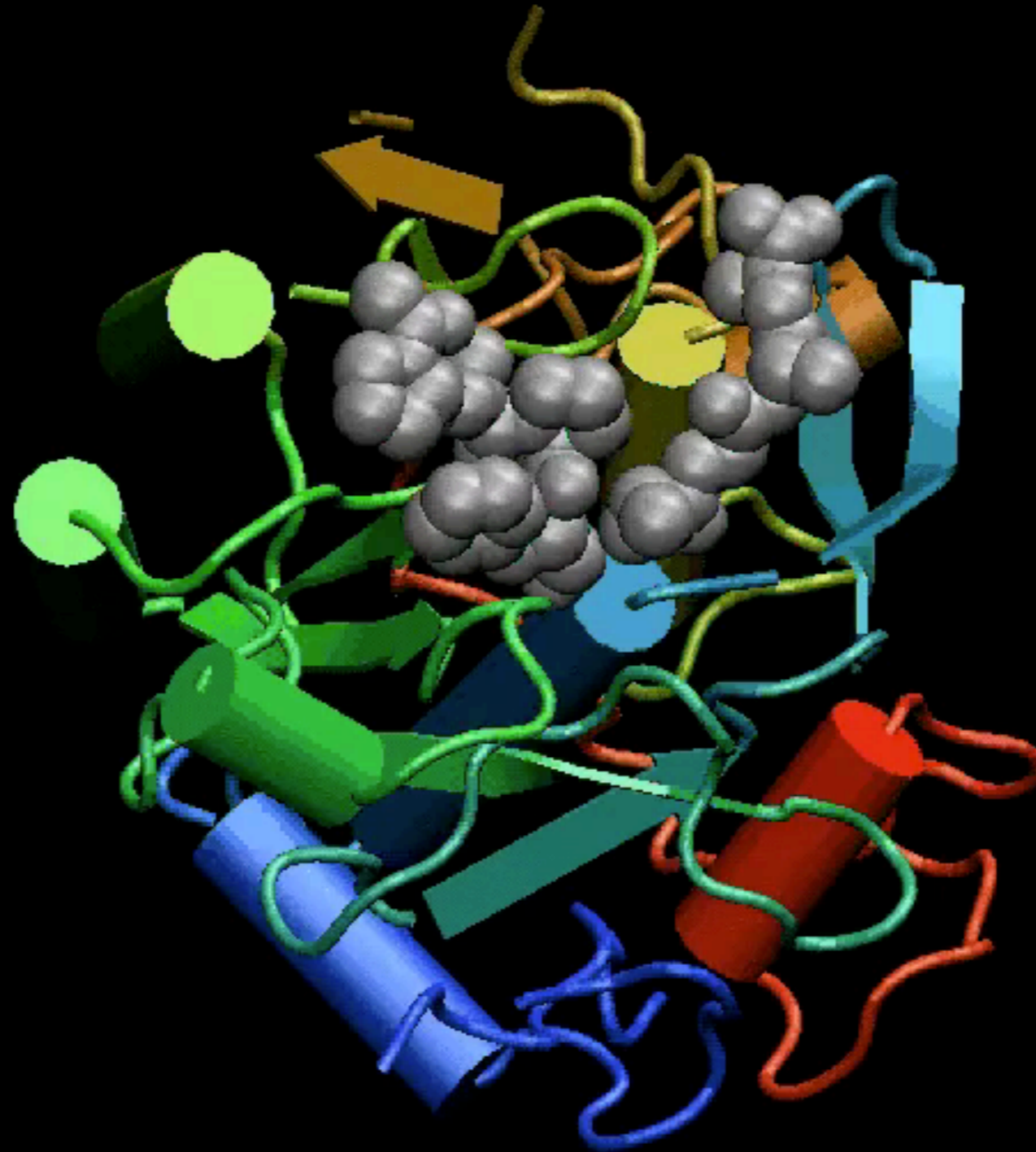
Proteinase K

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