

BGGN 213

Structural Bioinformatics

Lecture 11

Barry Grant
UC San Diego

<http://thegrantlab.org/bggn213>

Download VMD: See class website!

“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

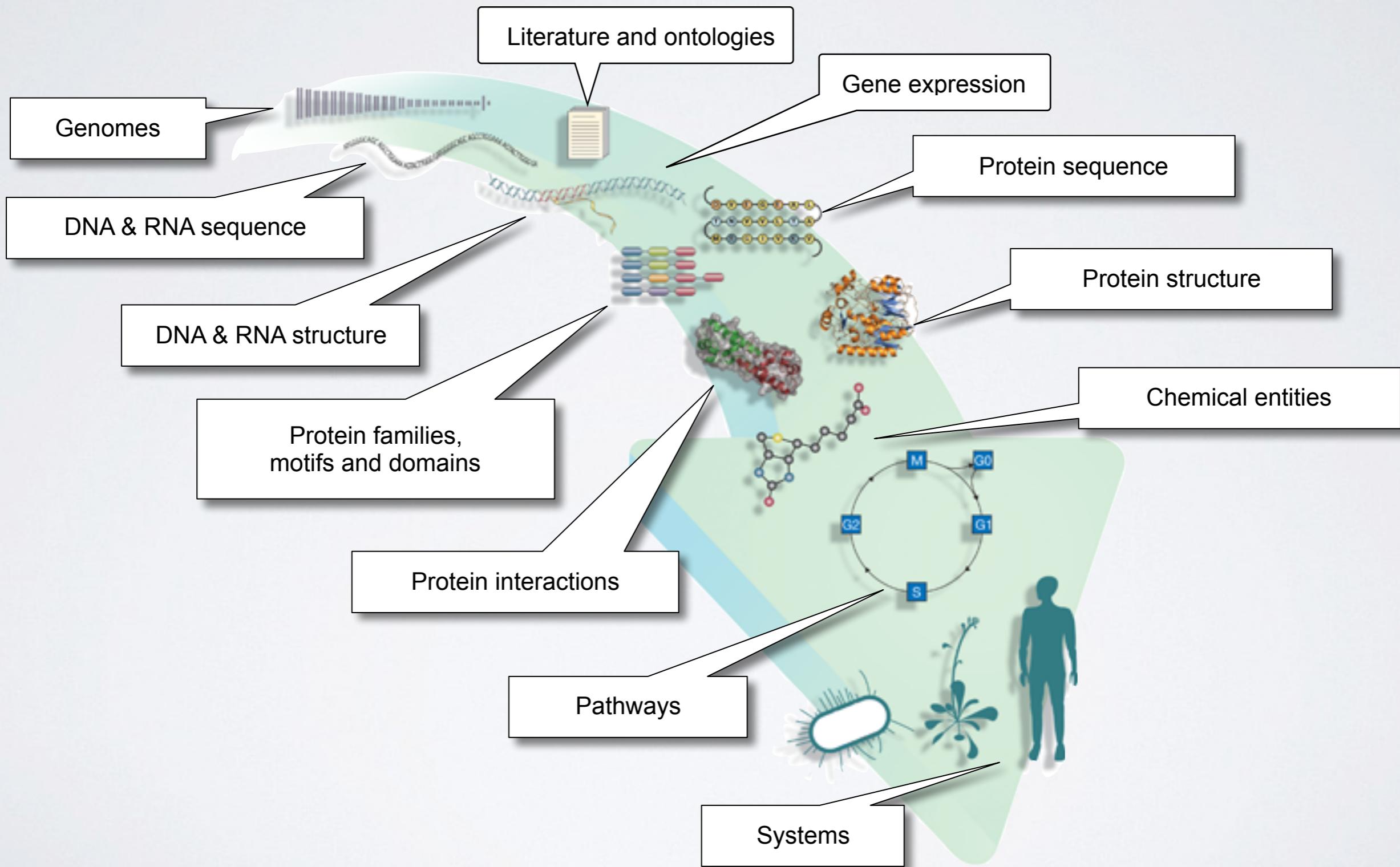
Why should we care?

Why should we care?

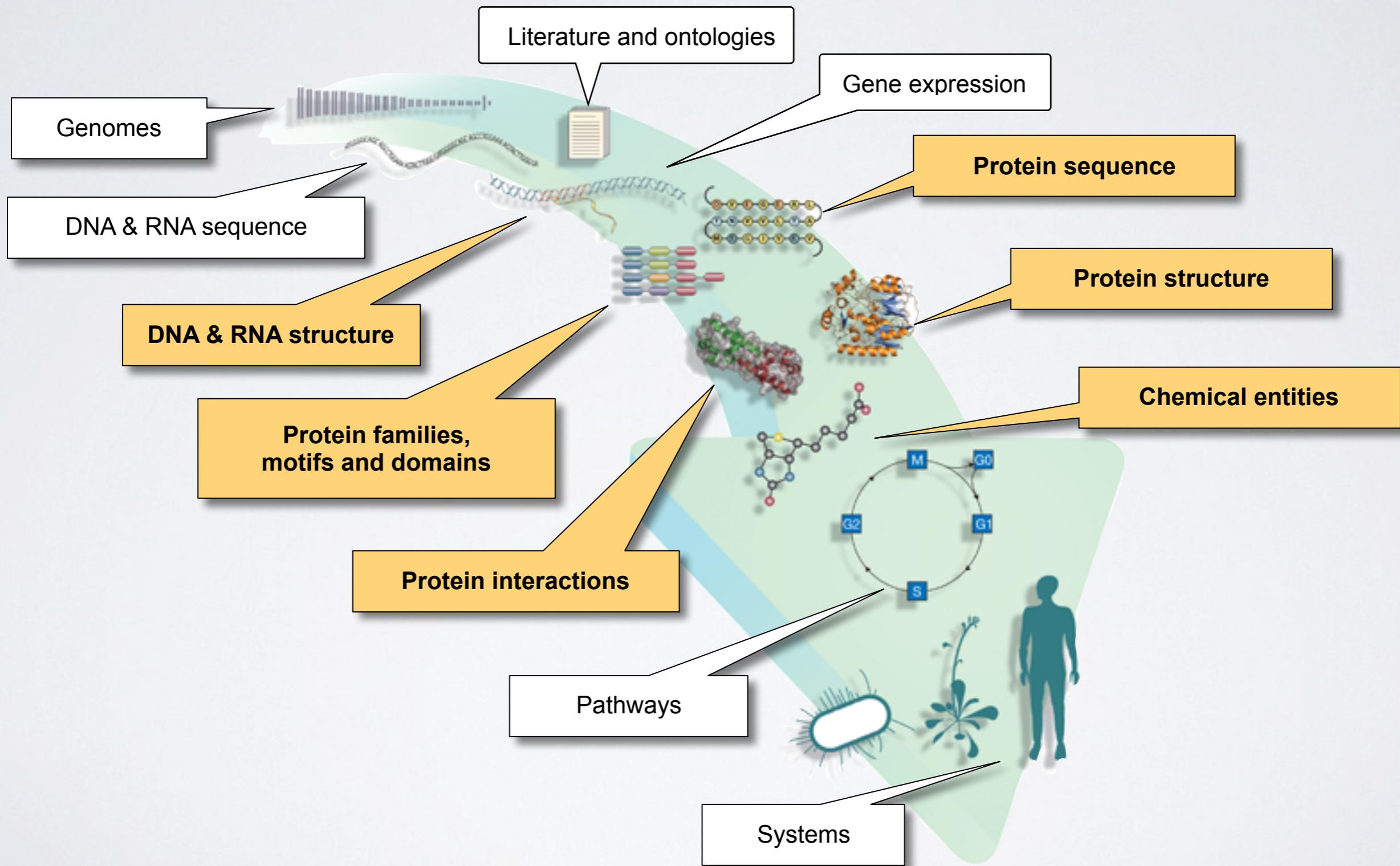
Because biomolecules are “nature’s robots”

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

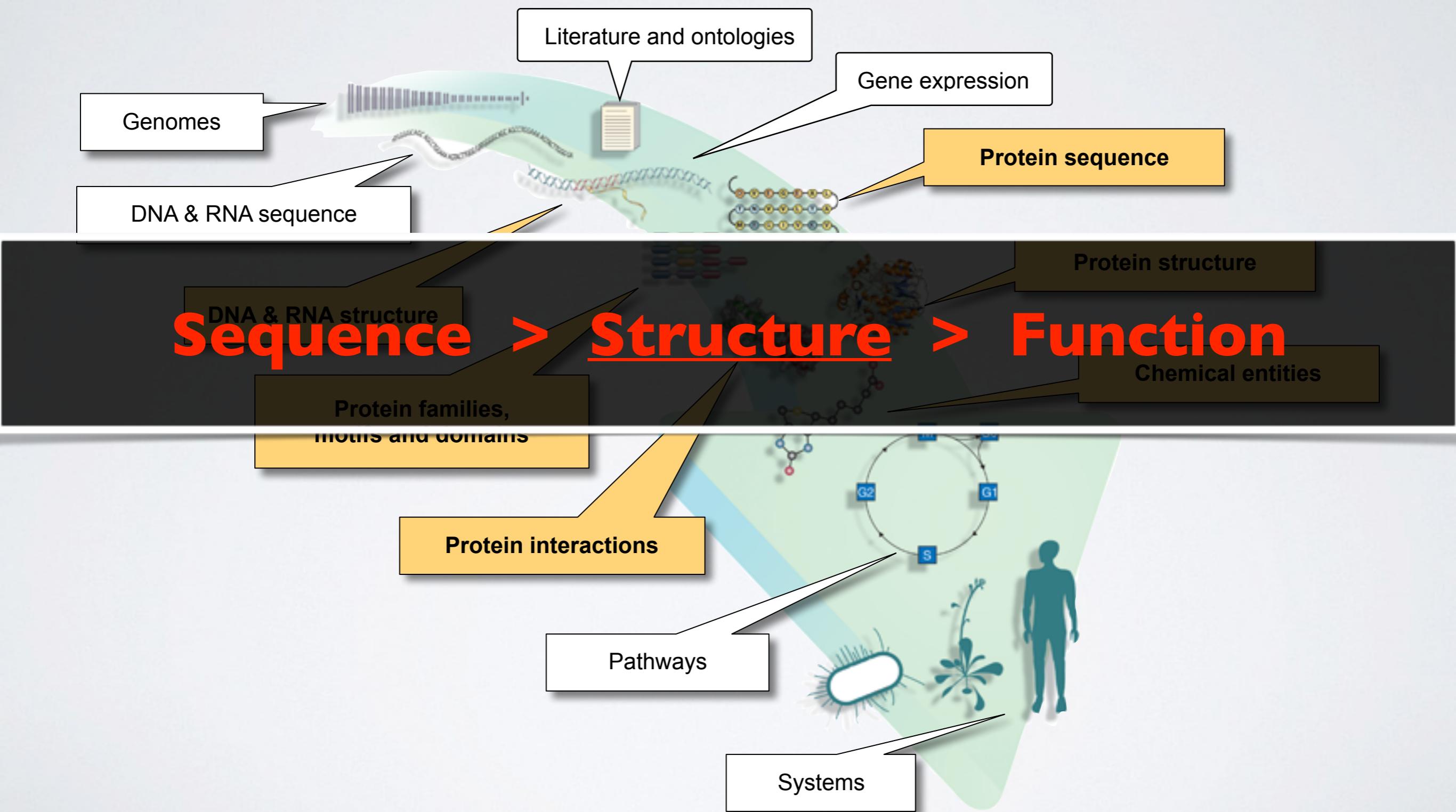
BIOINFORMATICS DATA



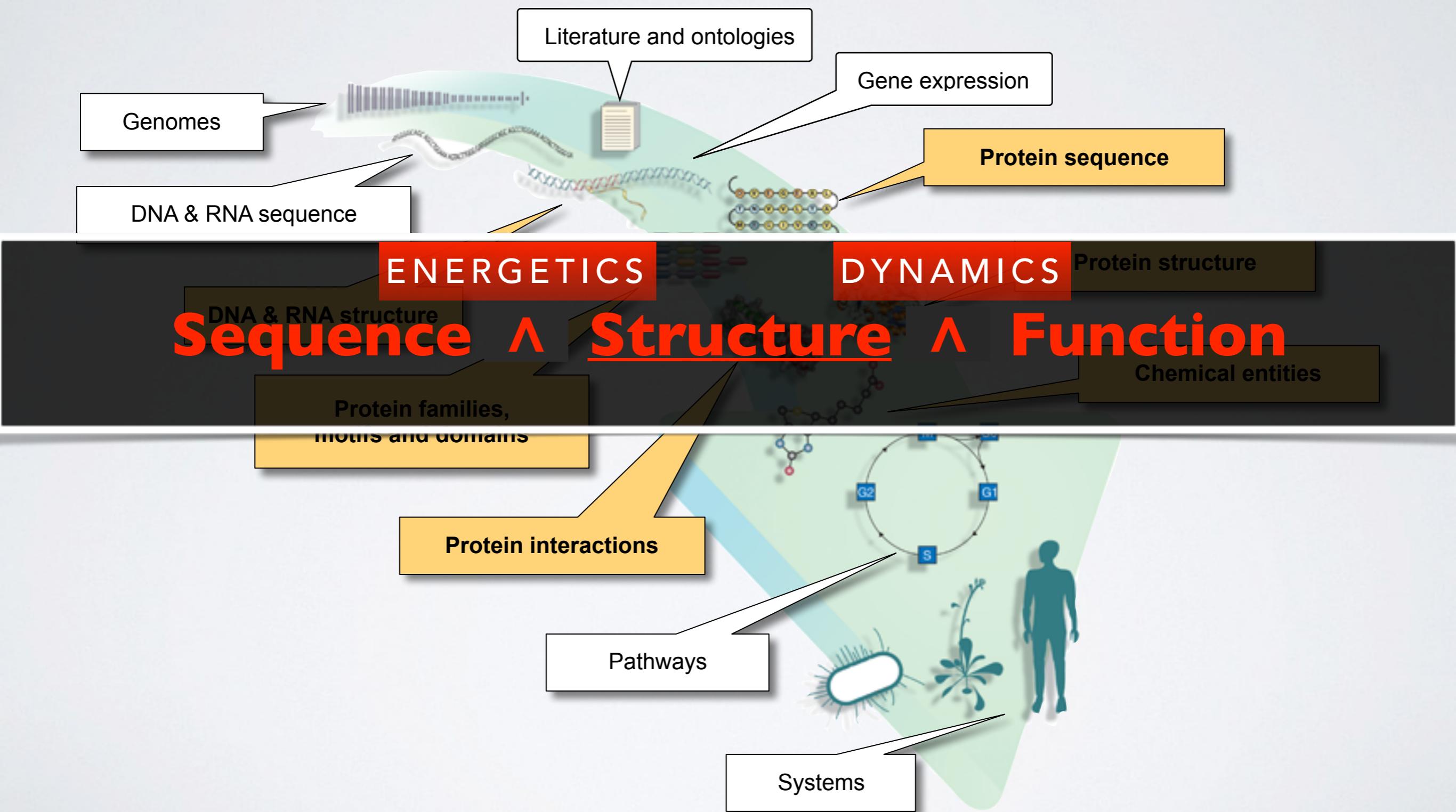
STRUCTURAL DATA IS CENTRAL

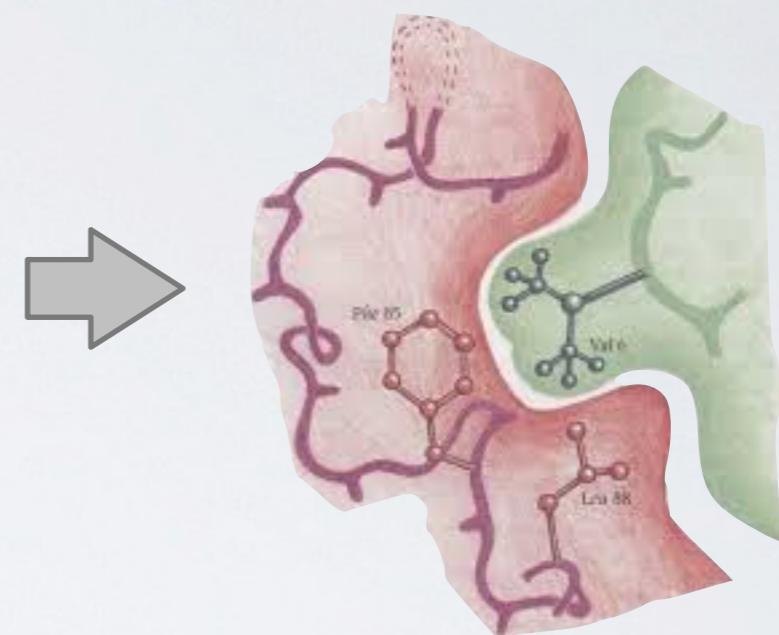
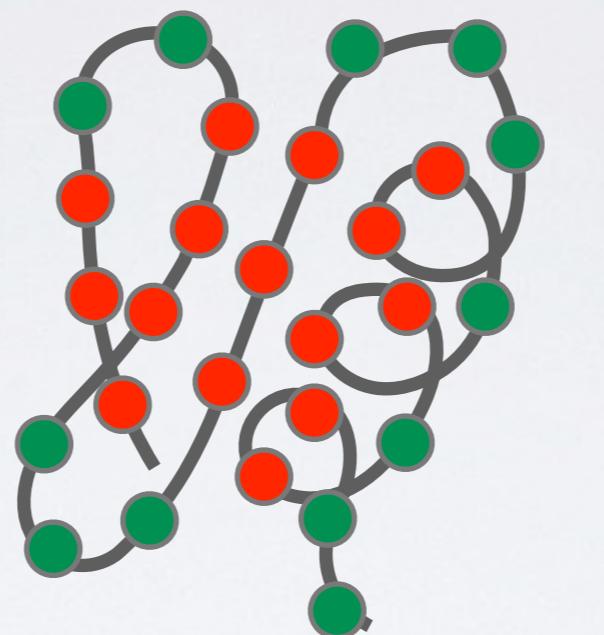
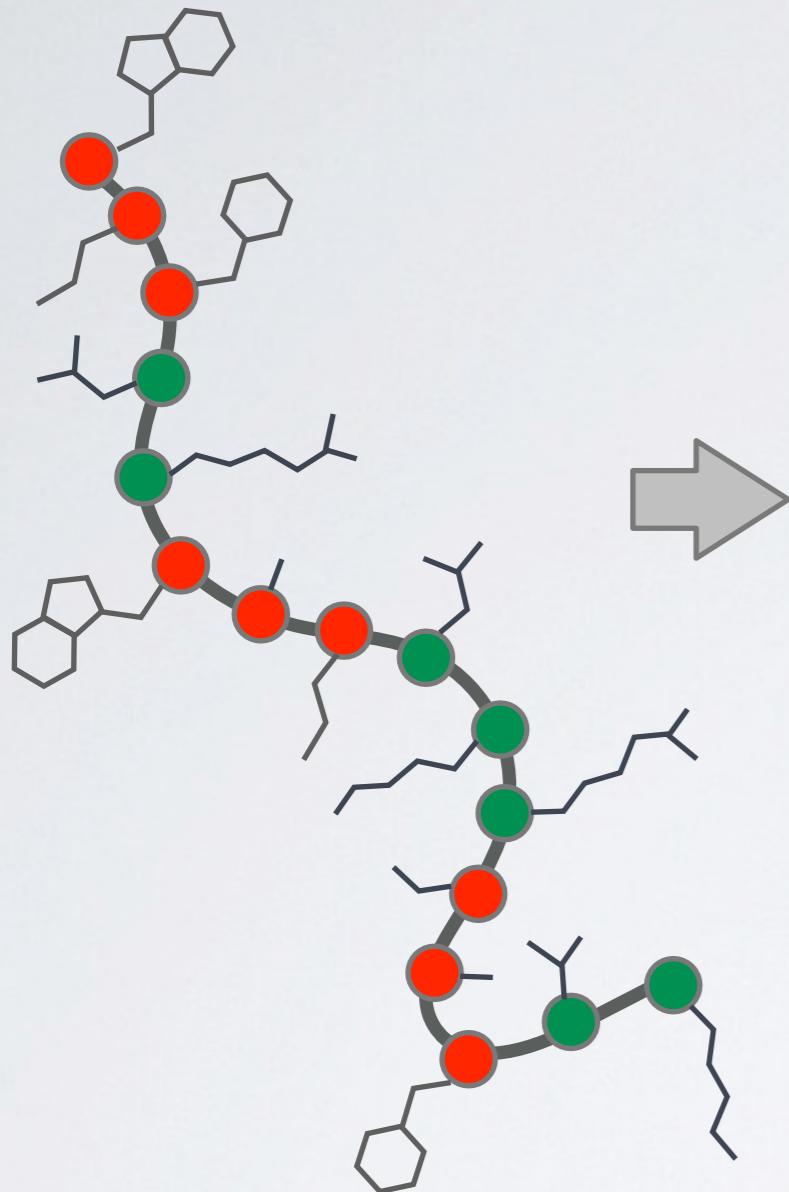


STRUCTURAL DATA IS CENTRAL



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*



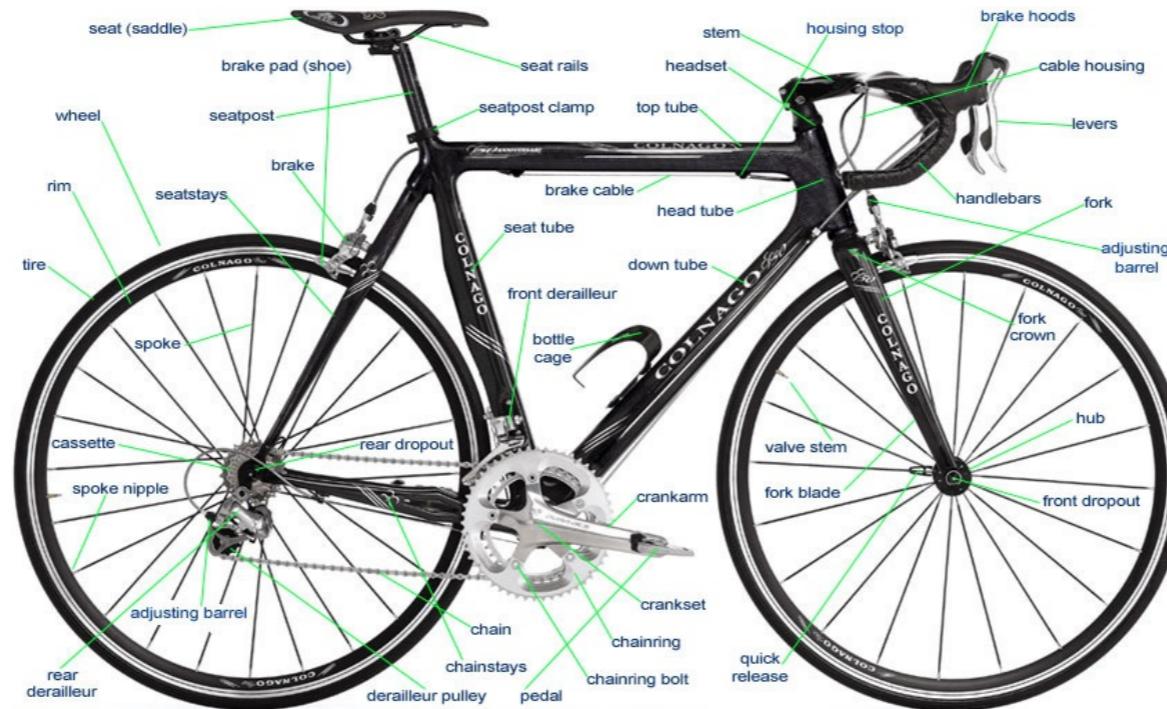
Genomics is a great start

Track Bike – DL 175

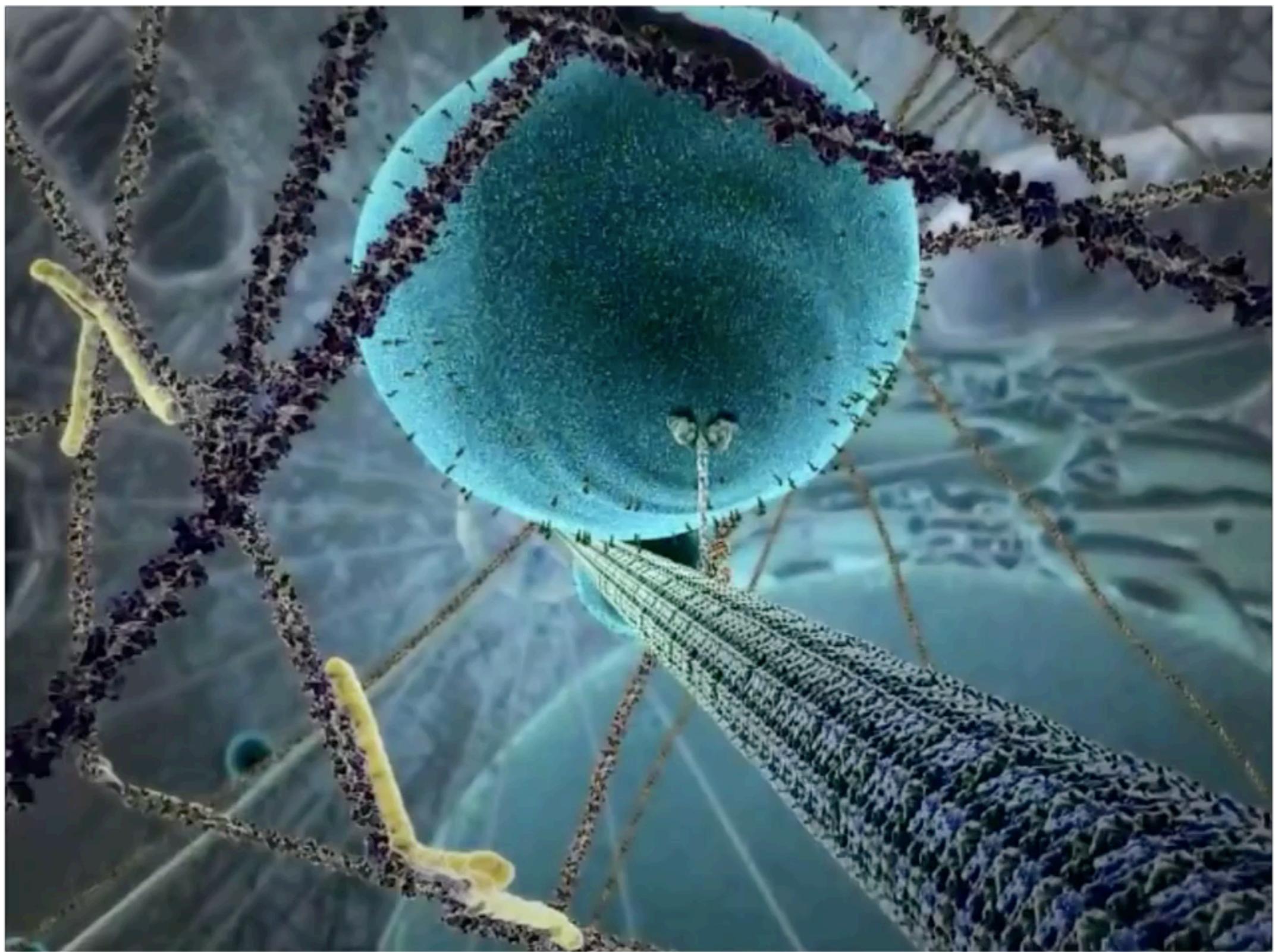
REF. NO.	IBM NO.	DESCRIPTION
1	156011	Track Frame 21", 22", 23", 24", Team Red
2	157040	Fork for 21" Frame
2	157039	Fork for 22" Frame
2	157038	Fork for 23" Frame
2	157037	Fork for 24" Frame
3	191202	Handlebar TTT Competition Track Alloy 15/16"
4		Handlebar Stem, TTT, Specify extension
5	191278	Expander 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

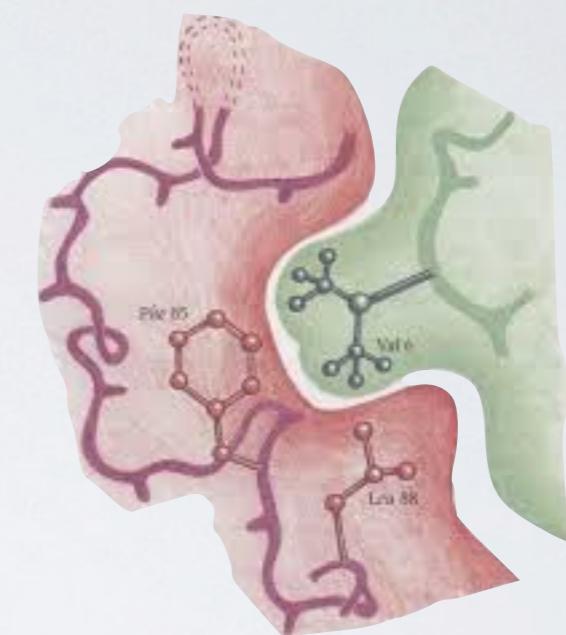
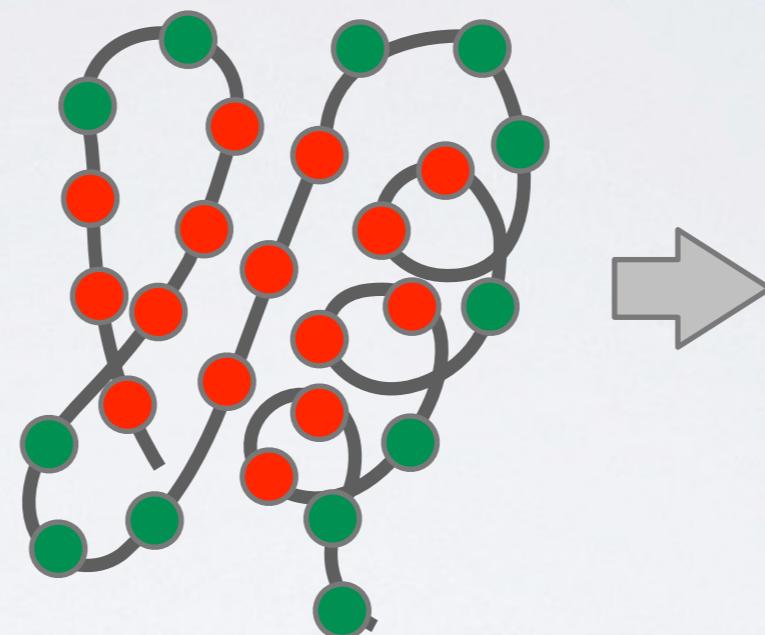
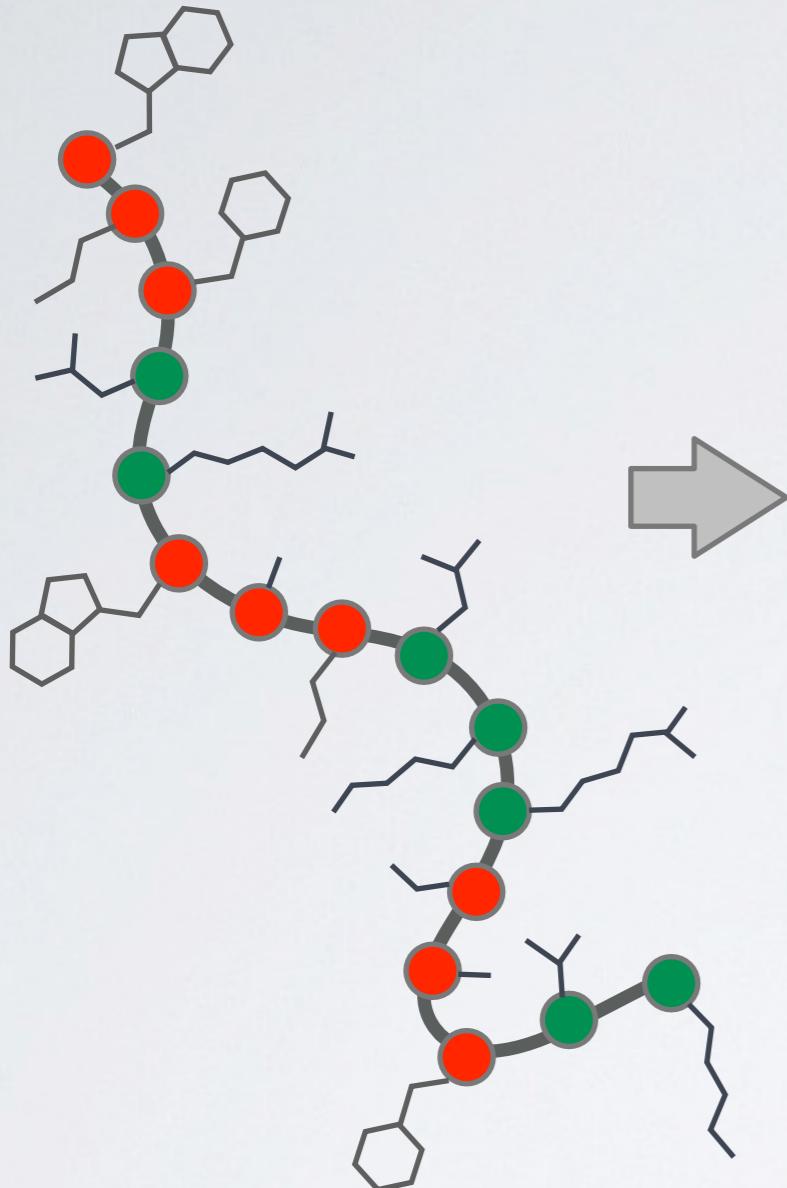
... but not the end



- 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

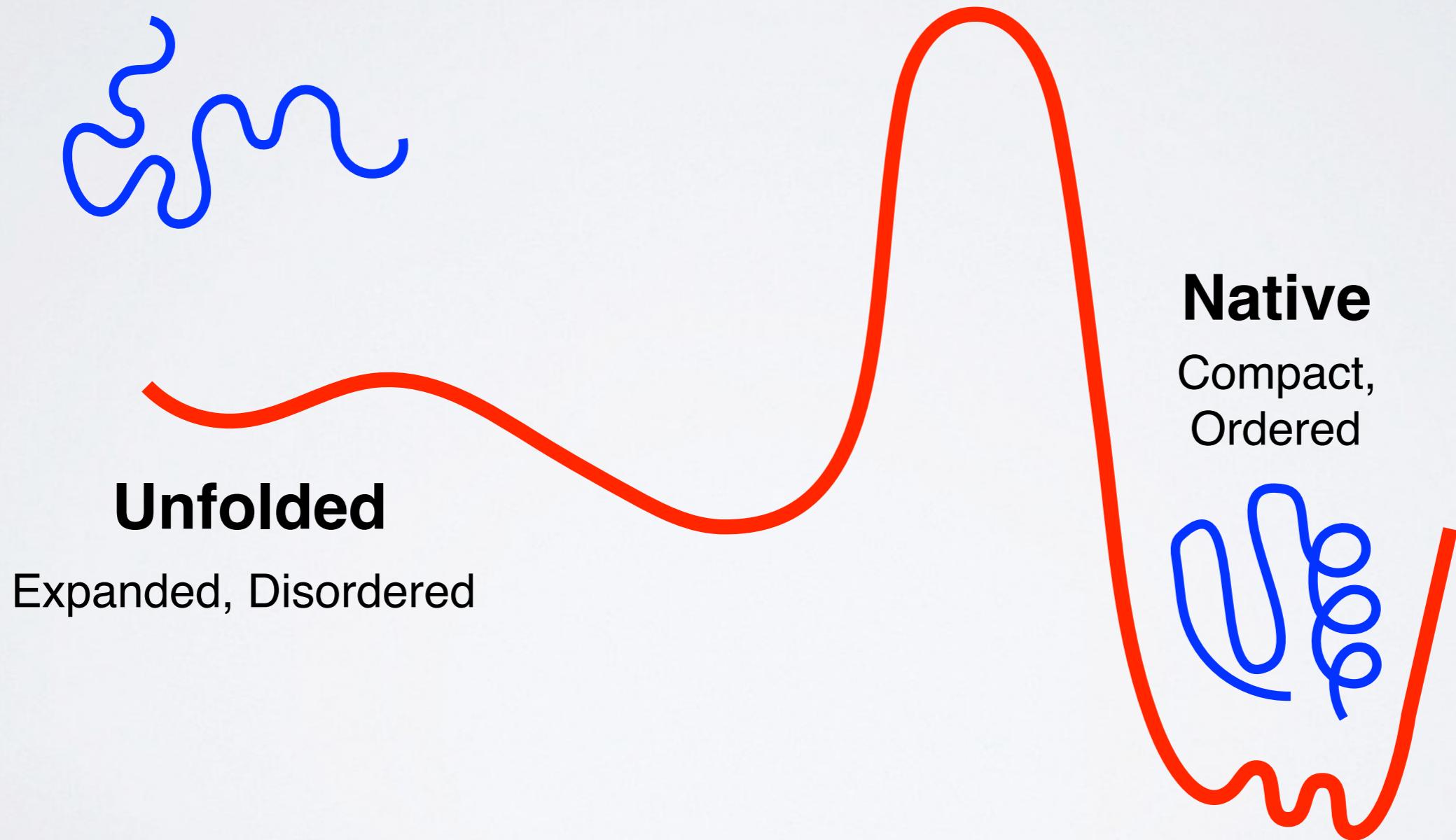
Structure

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

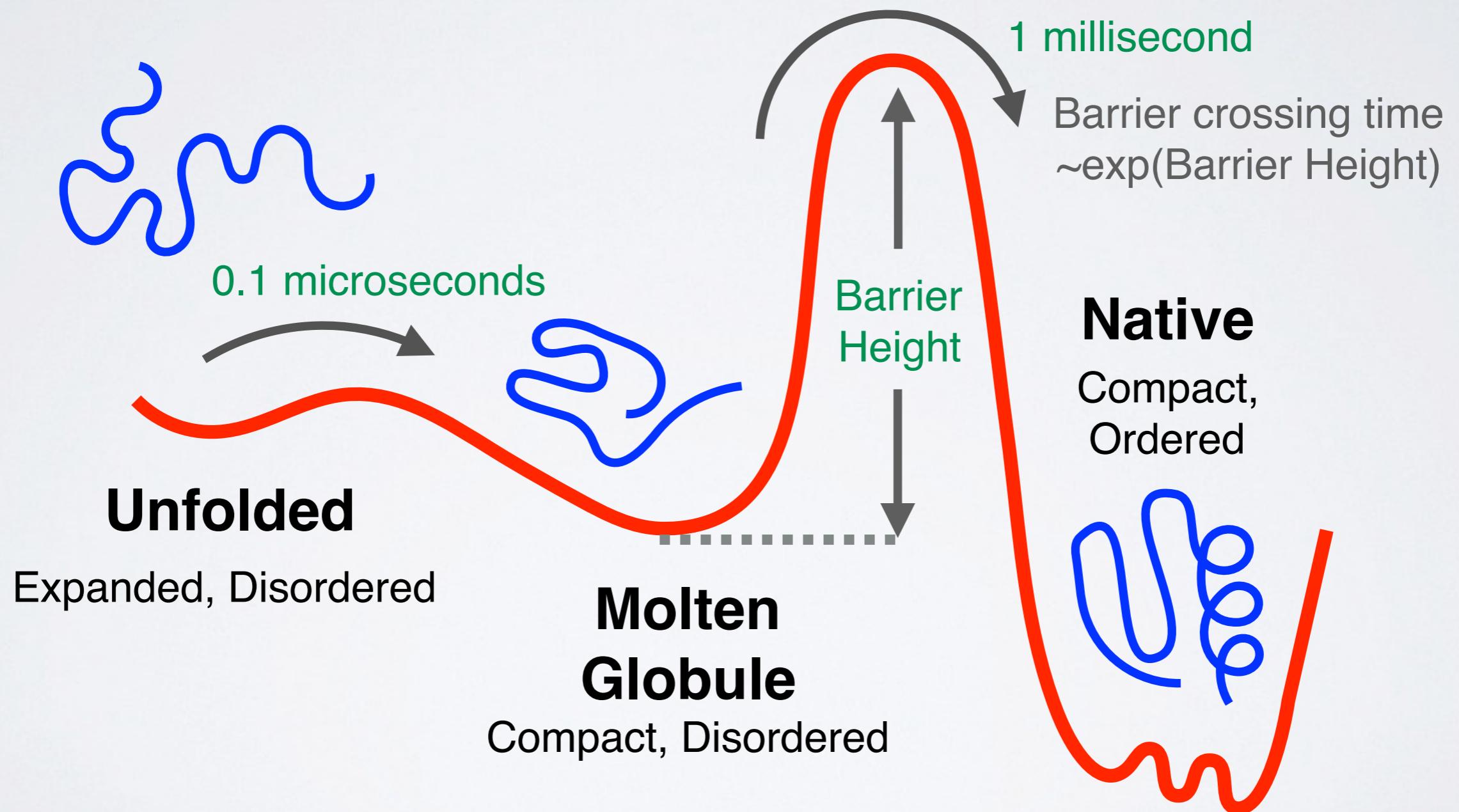
Function

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

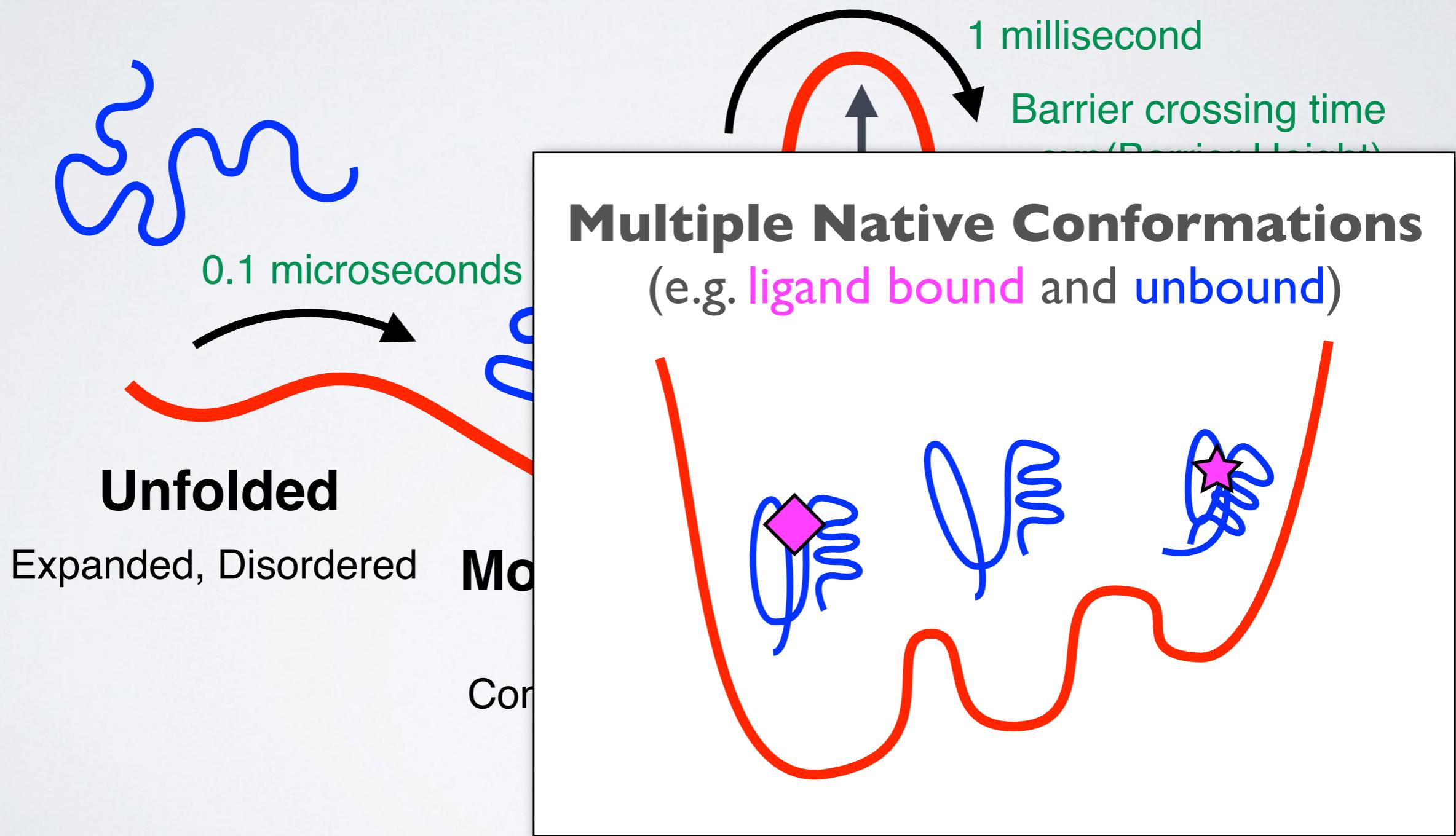
KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: ENERGY LANDSCAPE



Today's Menu

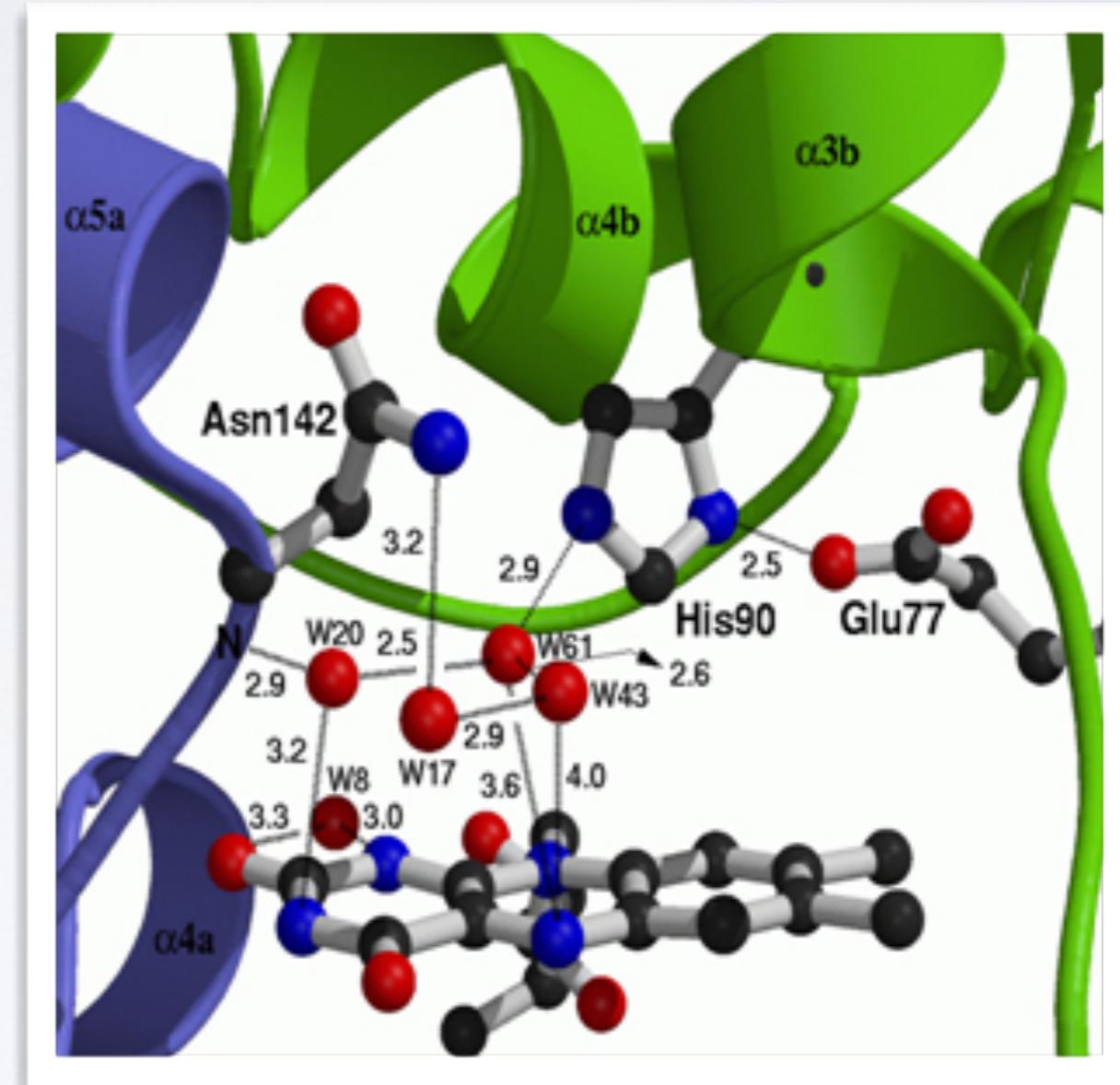
- **Overview of structural bioinformatics**
 - Motivations, goals and challenges
- **Fundamentals of protein structure**
 - Structure composition, form and forces
- **Representing, interpreting & modeling protein structure**
 - Visualizing & interpreting protein structures
 - Analyzing protein structures
 - 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

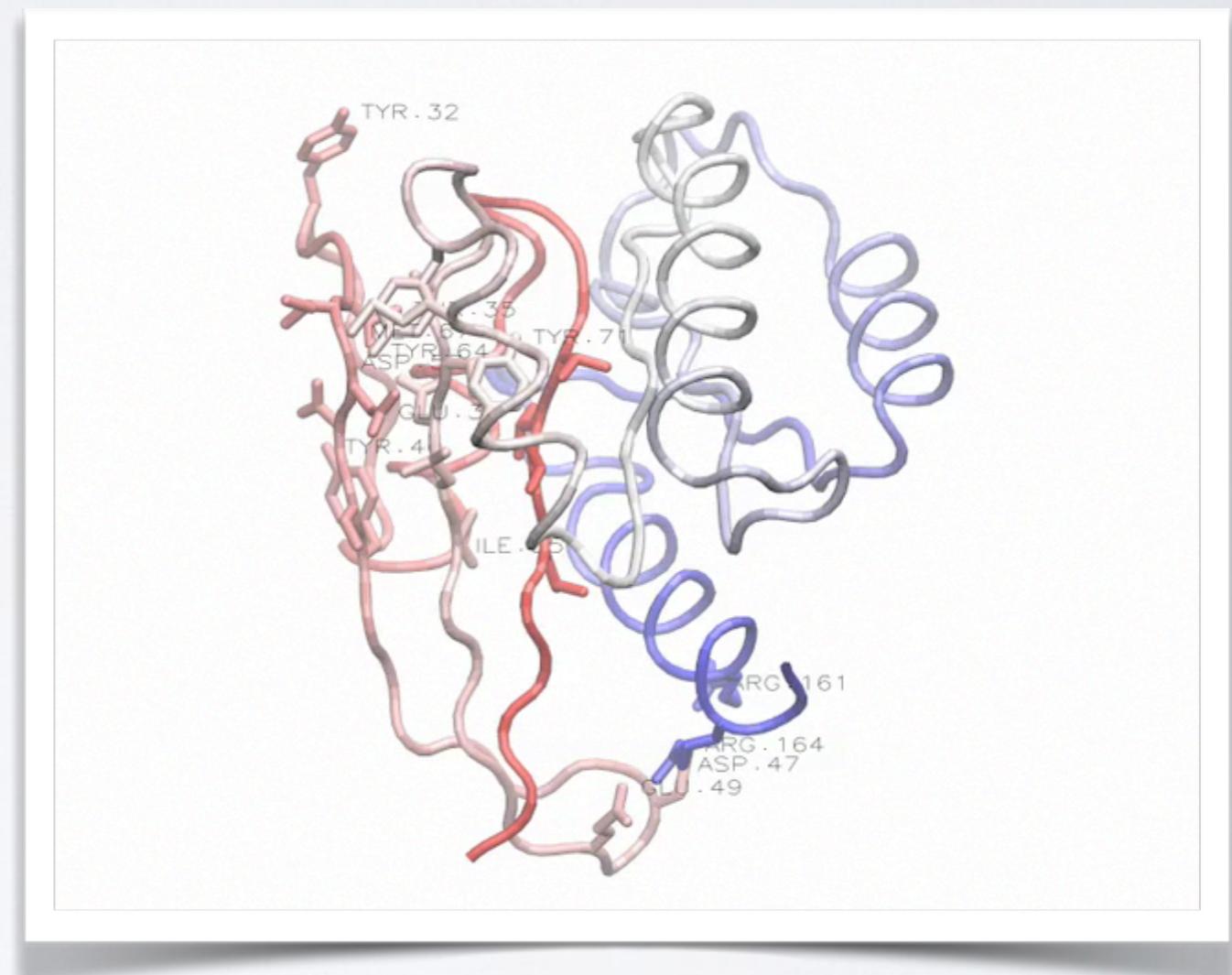
Motivation 1: Detailed understanding of molecular interactions

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



Motivation 1: Detailed understanding of molecular interactions

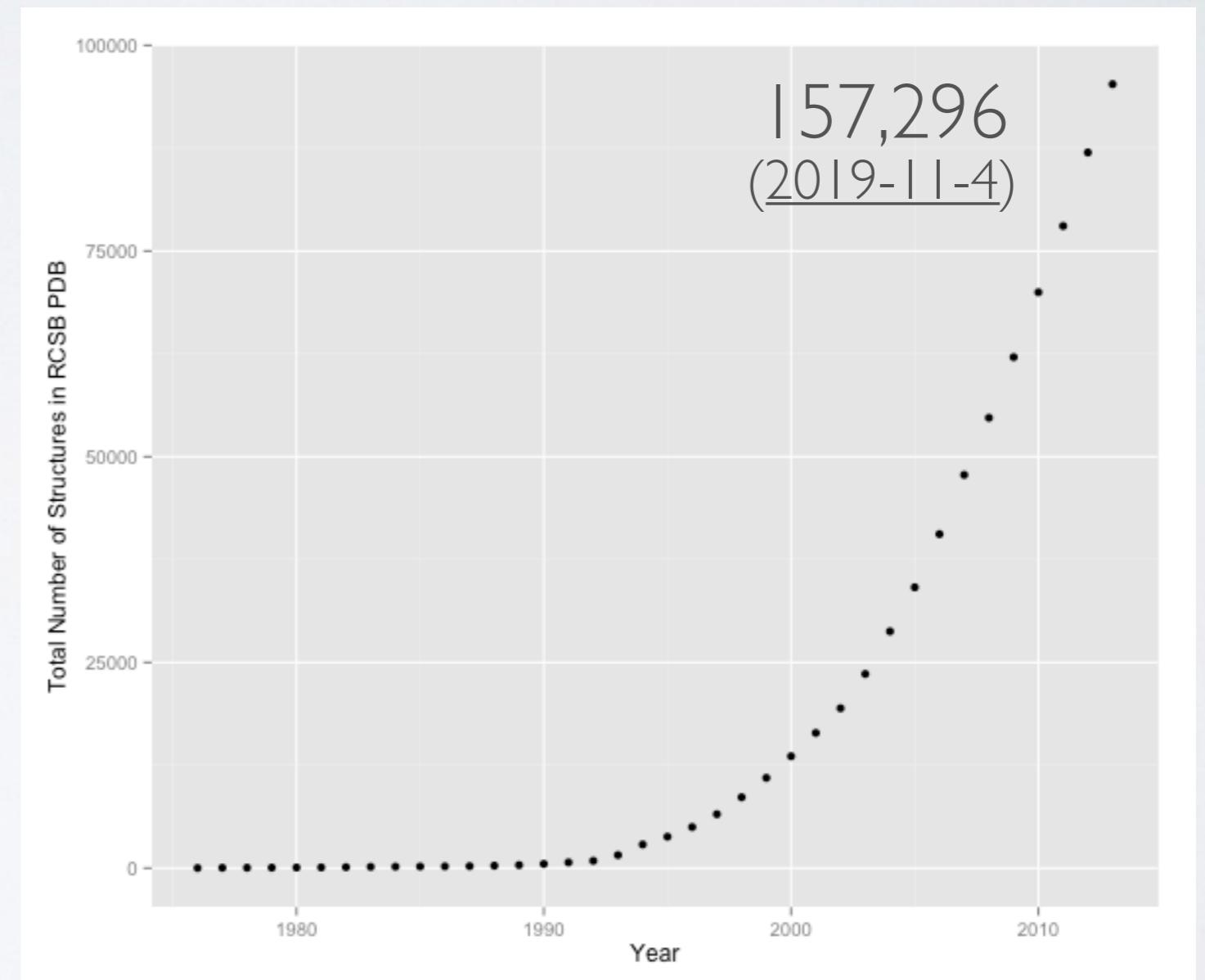
Computational modeling can provide detailed insight into functional interactions, their regulation and potential consequences of perturbation.



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: <https://www.rcsb.org/stats/>

Motivation 2: Lots of structural data is becoming available

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

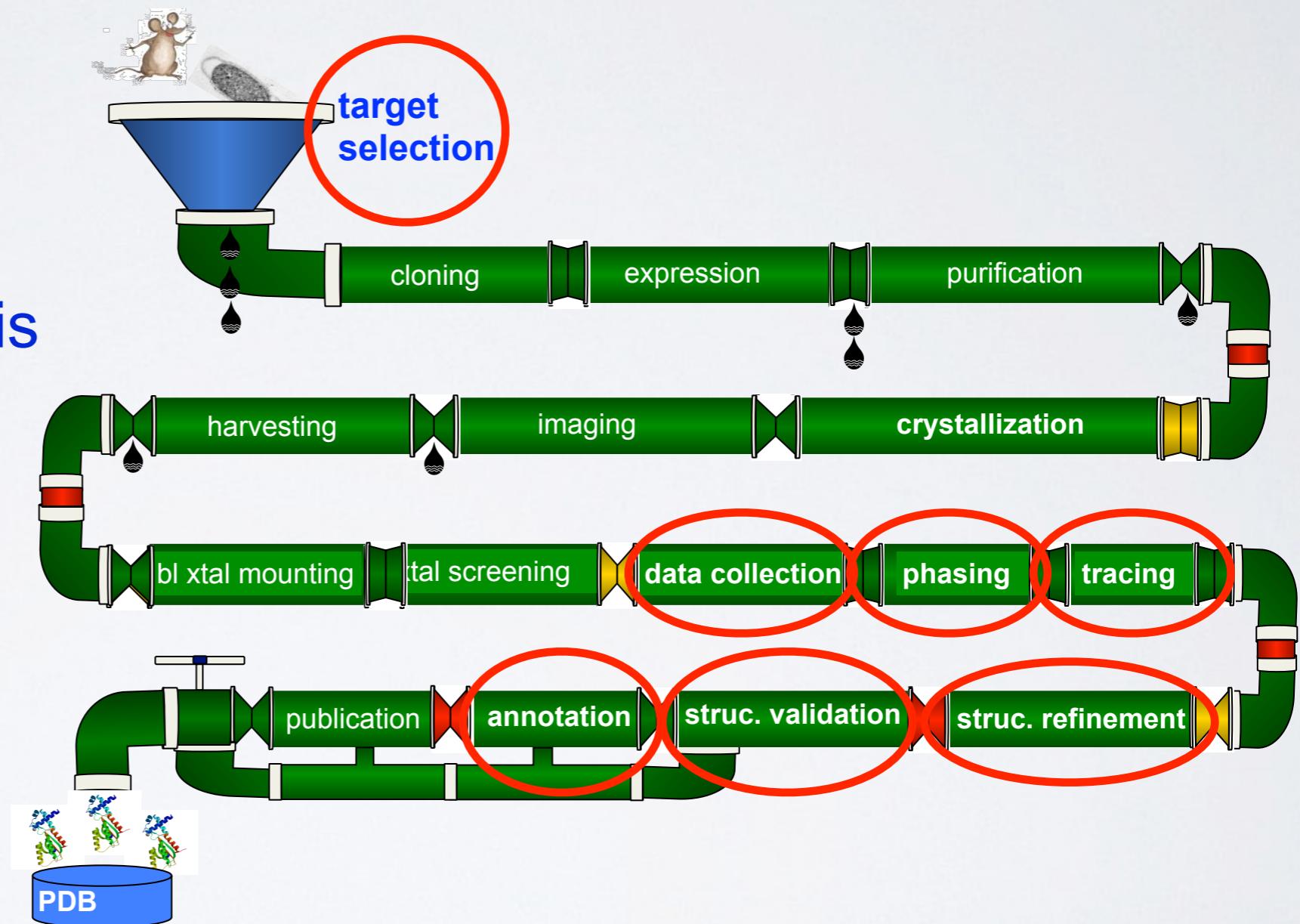
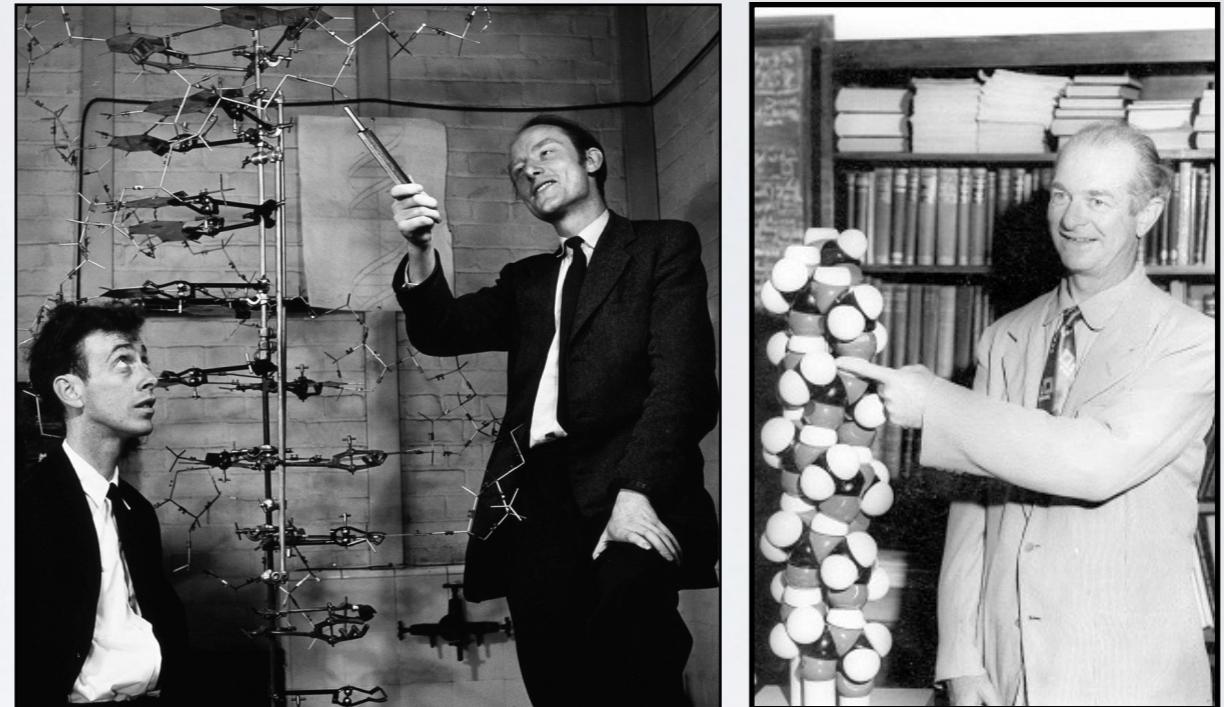


Image Credit: "Structure determination assembly line" Adam Godzik

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:

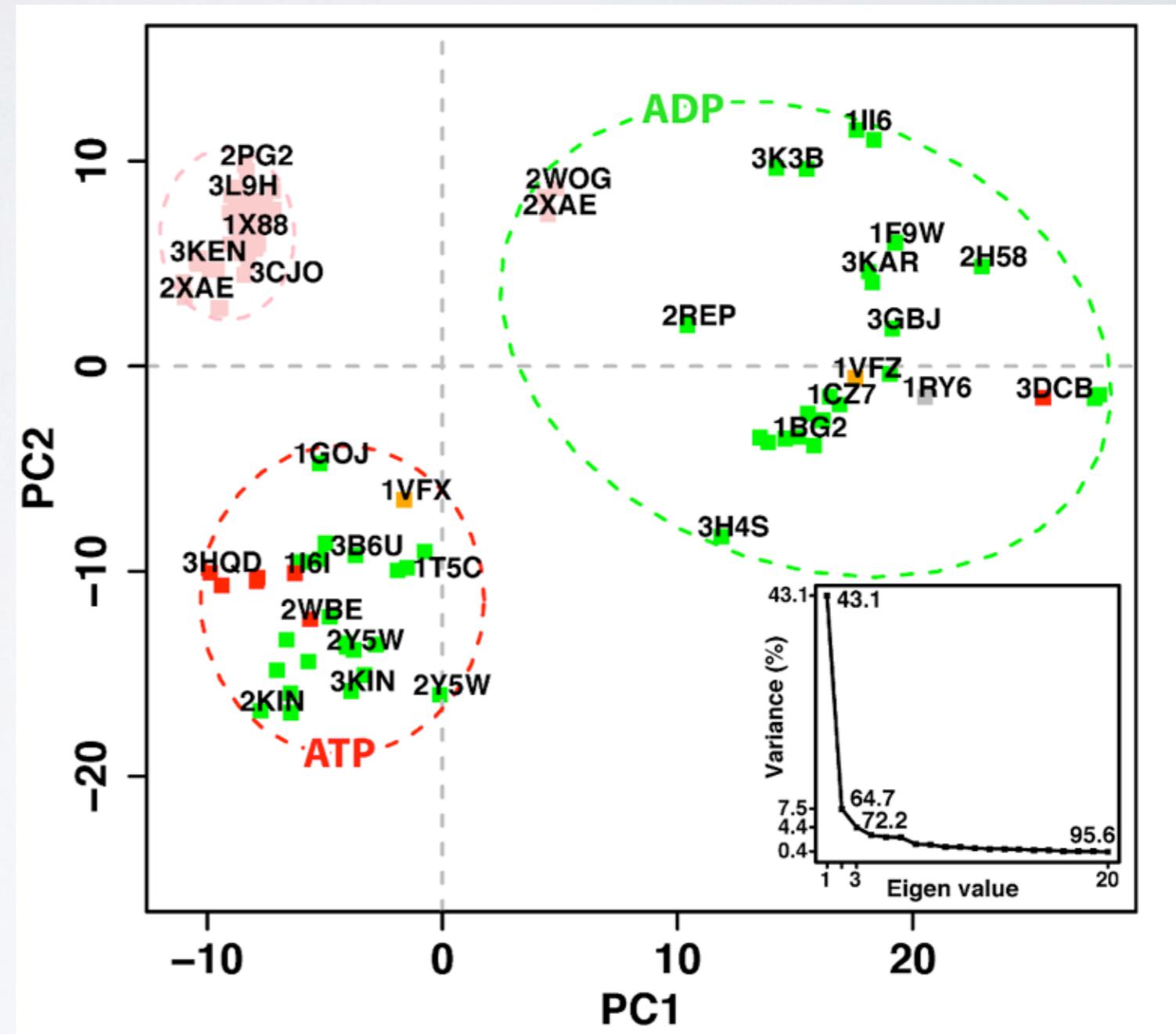
- Visualization
- Analysis
- Comparison
- Prediction
- Design



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

Goals:

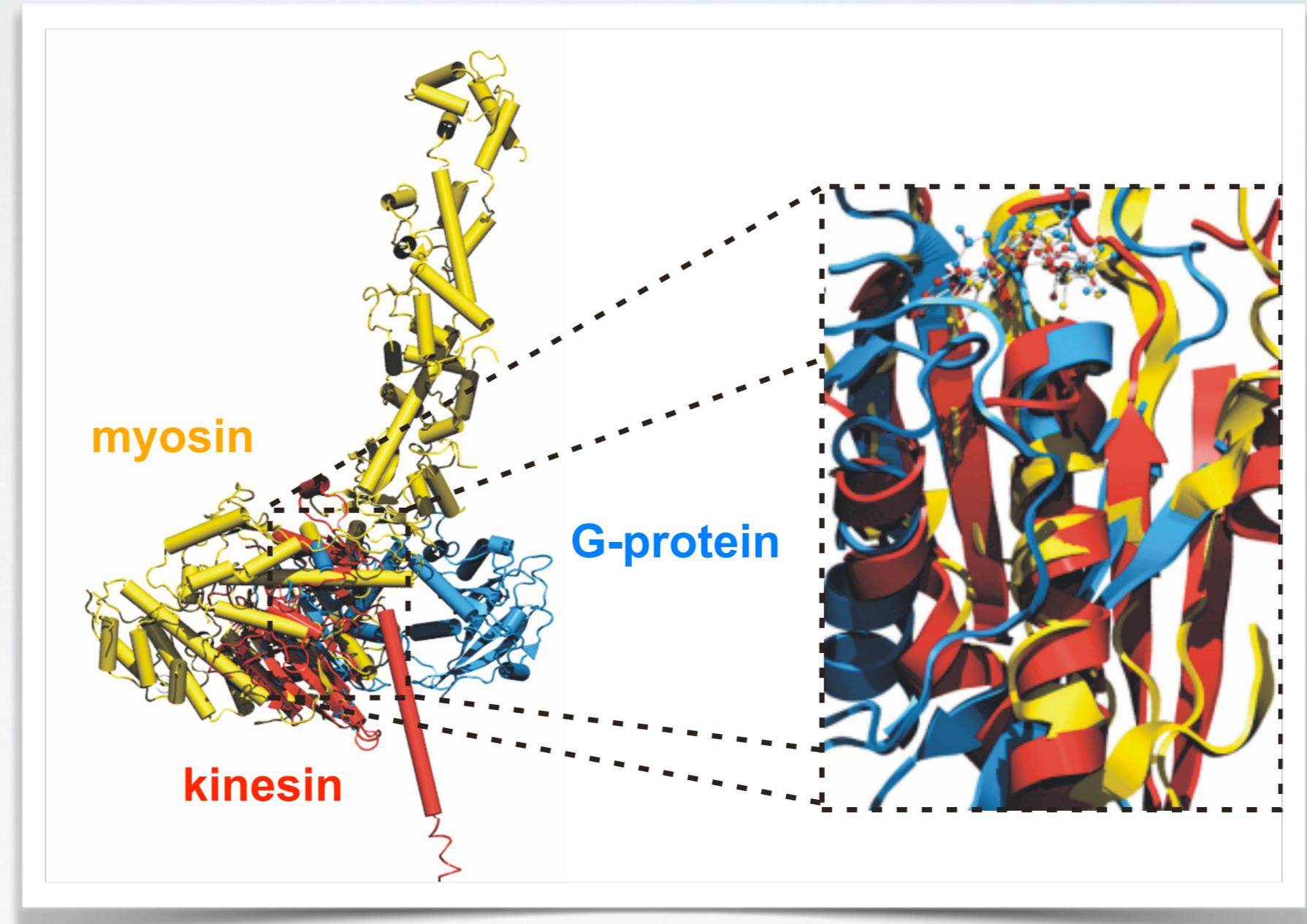
- Visualization
- Analysis
- Comparison
- Prediction
- Design



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

Goals:

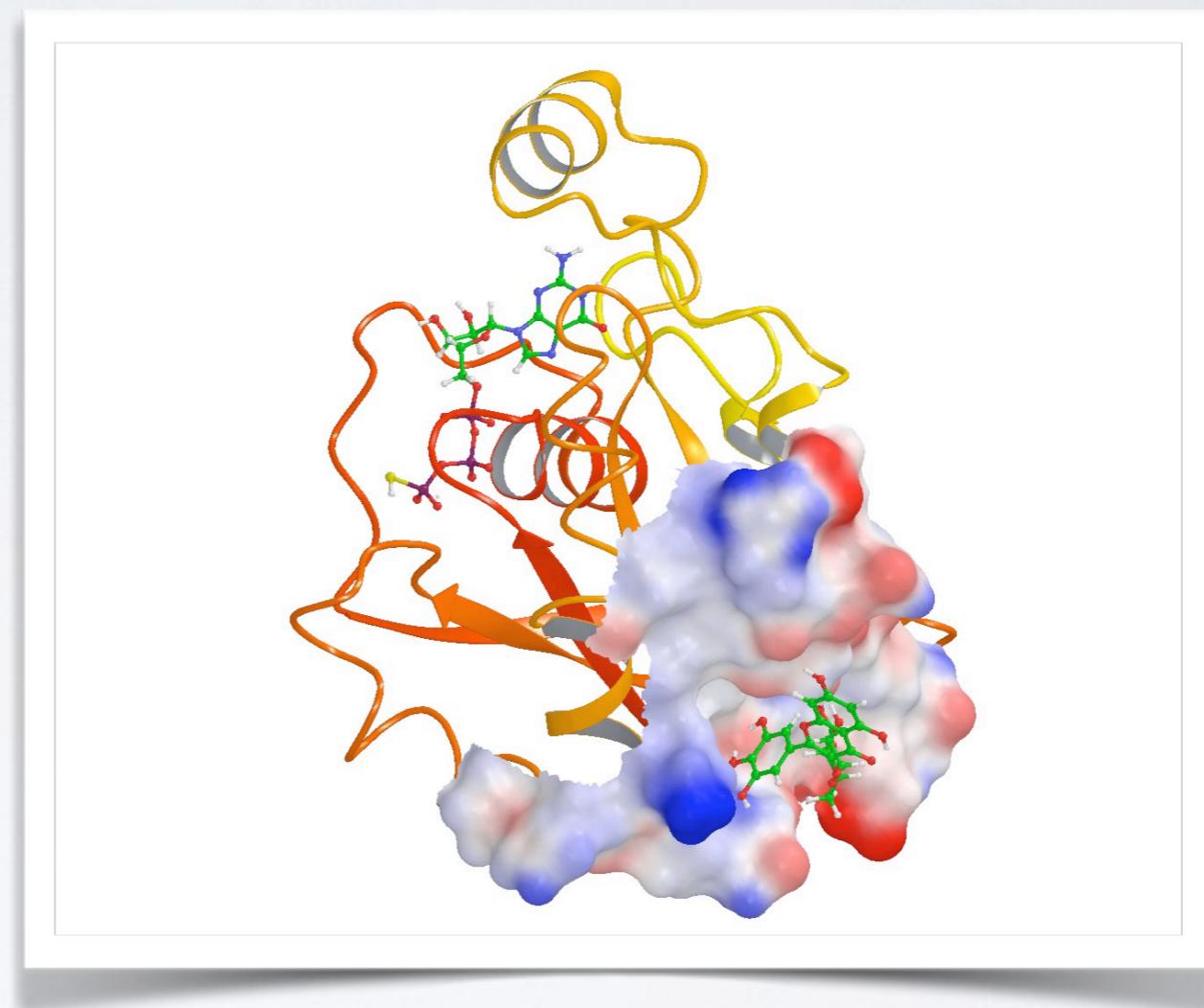
- Visualization
- Analysis
- Comparison
- Prediction
- Design



Grant et al. unpublished

Goals:

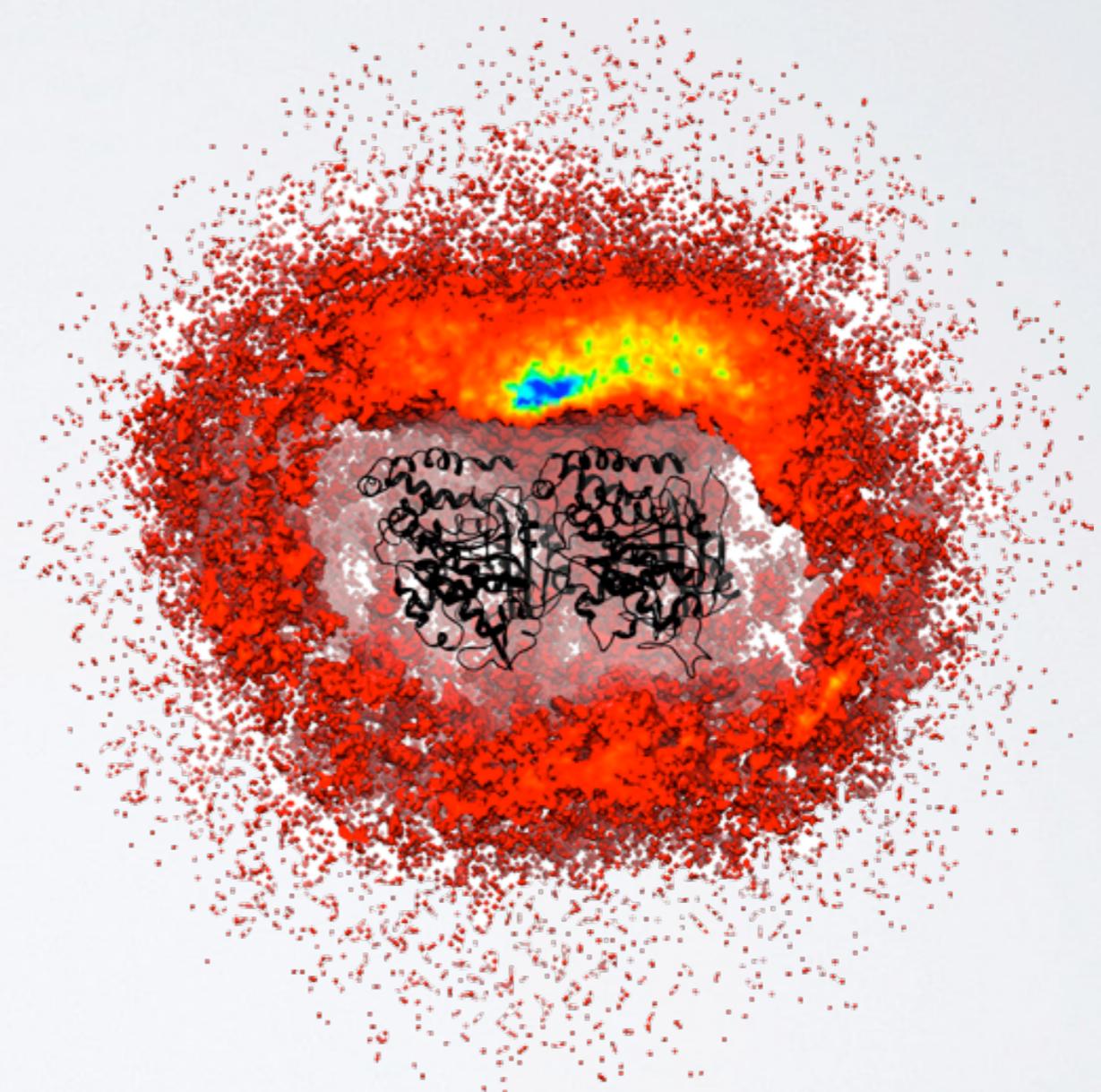
- Visualization
- Analysis
- Comparison
- Prediction
- Design



Grant et al. PLoS One (2011, 2012)

Goals:

- Visualization
- Analysis
- 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...

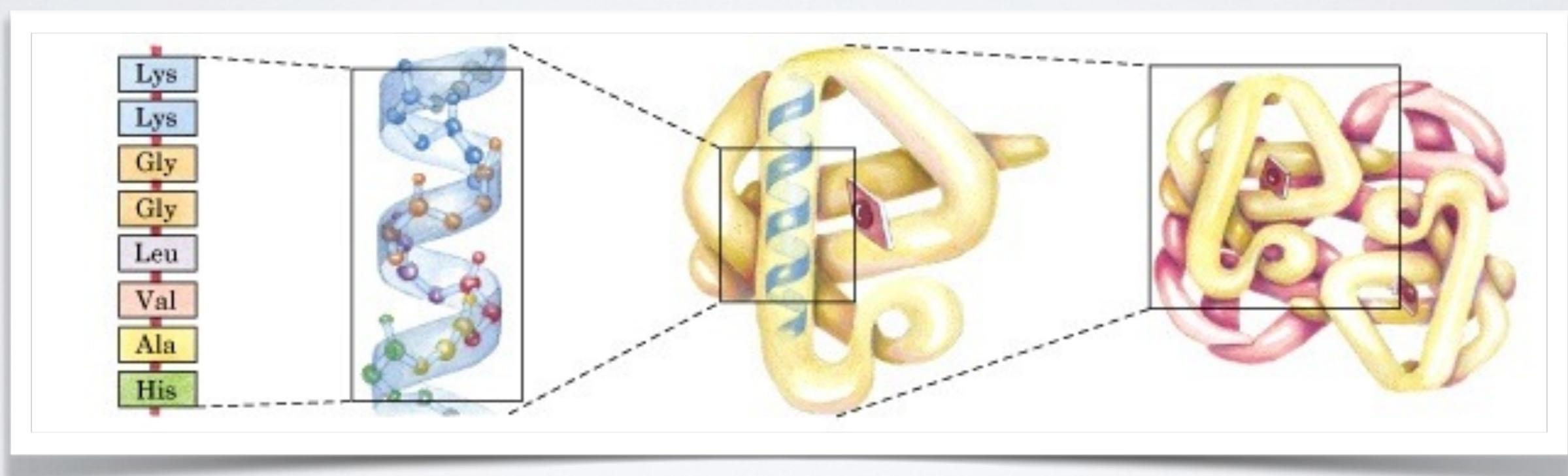
With applications to Biology, Medicine, Agriculture and Industry

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

RECAP: AMINO ACID NOMENCLATURE

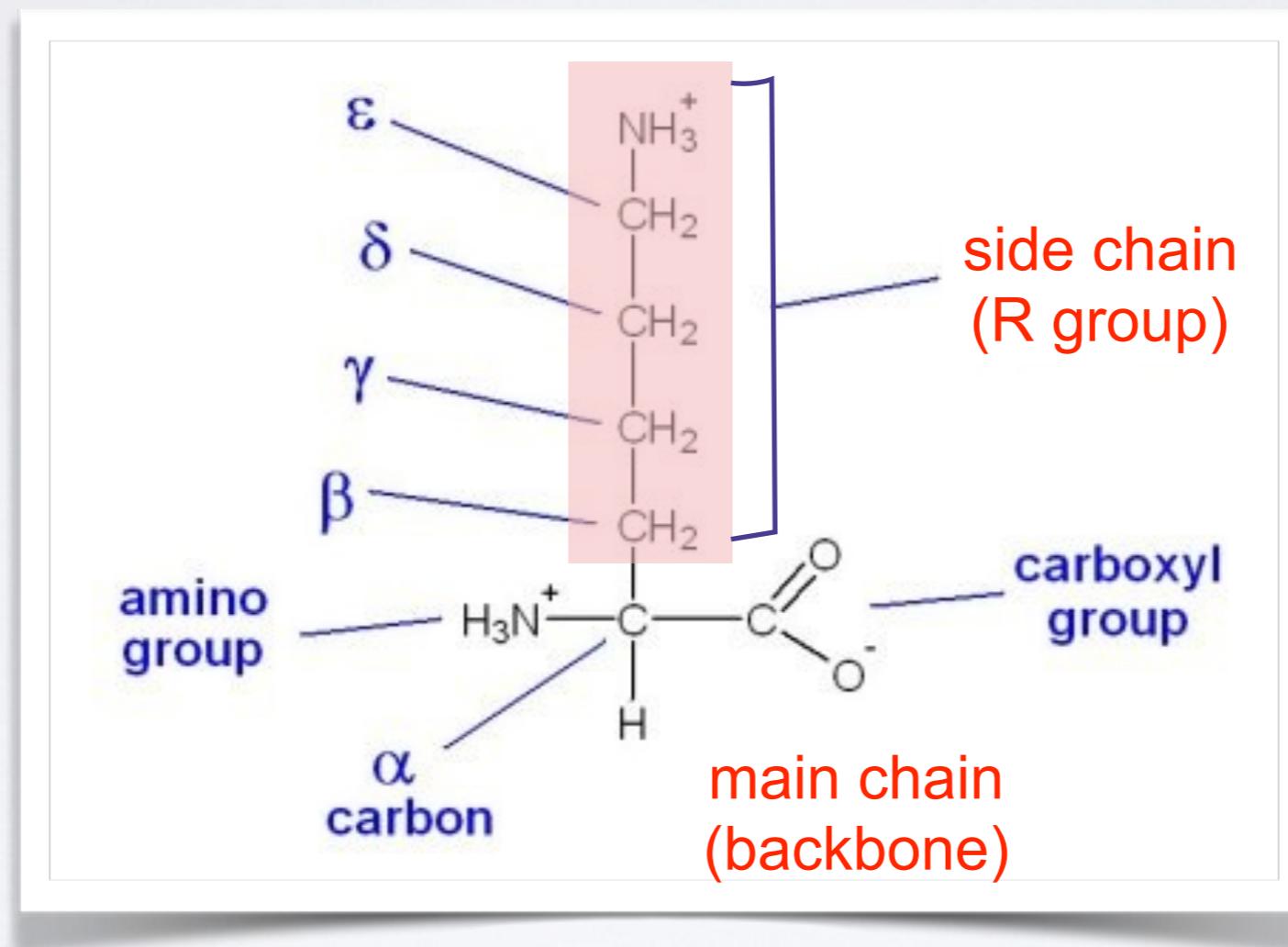
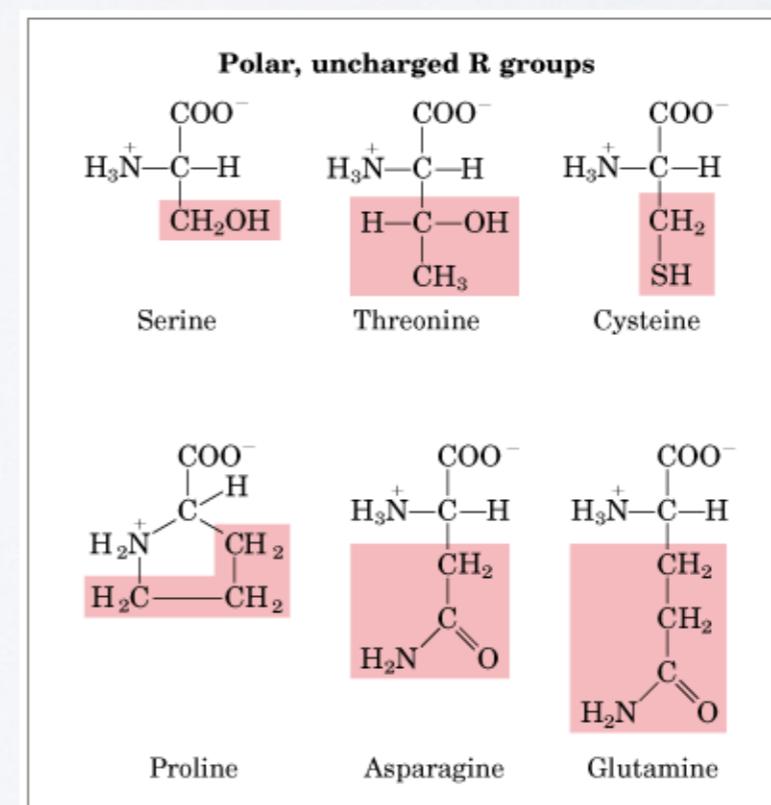
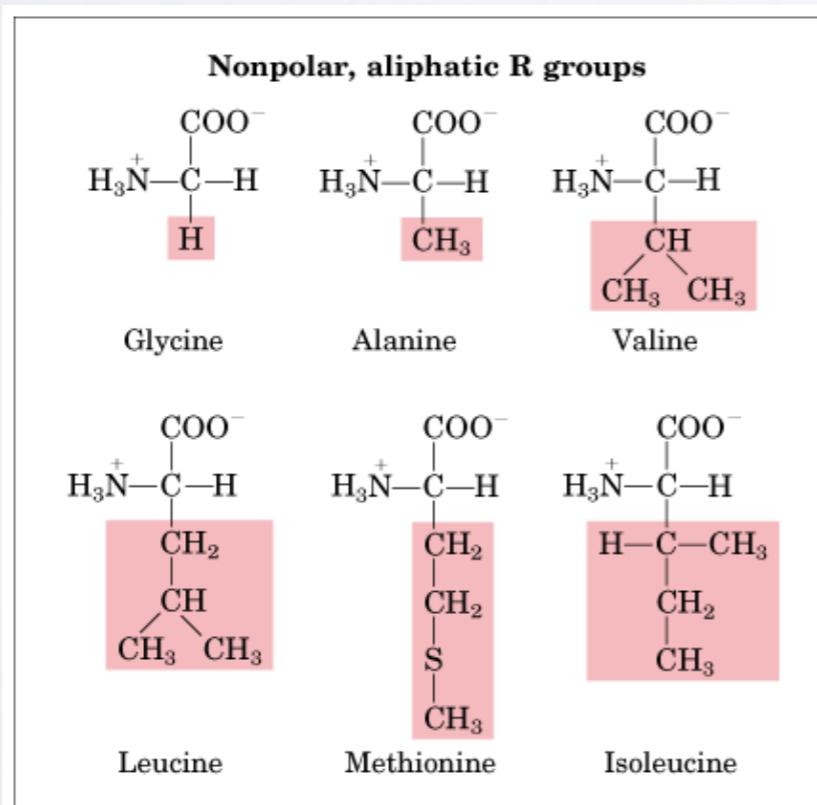
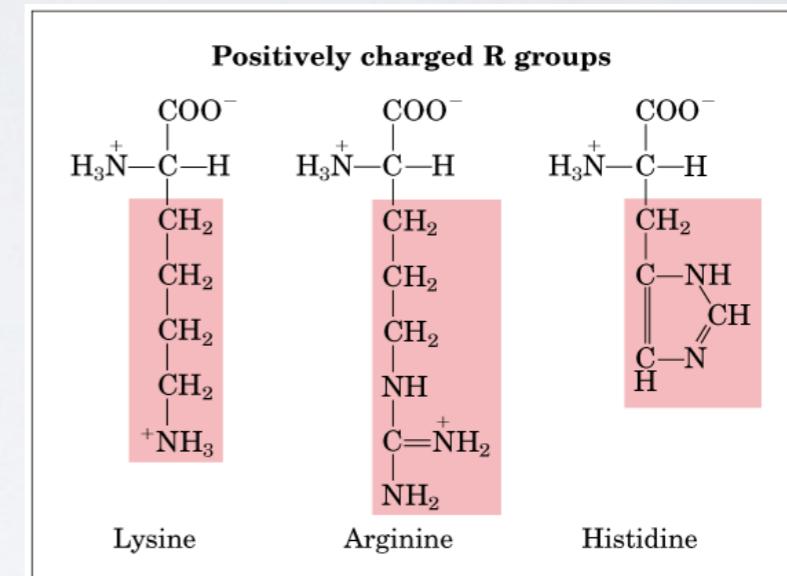
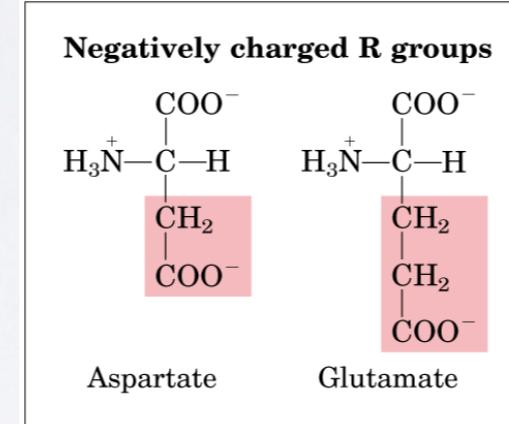
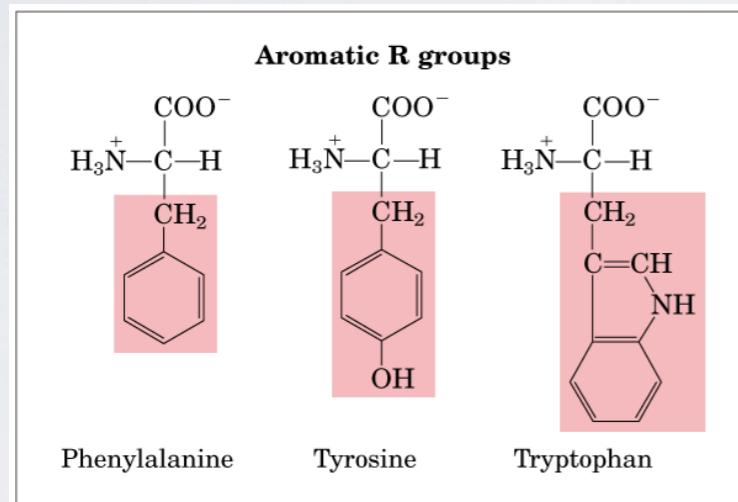
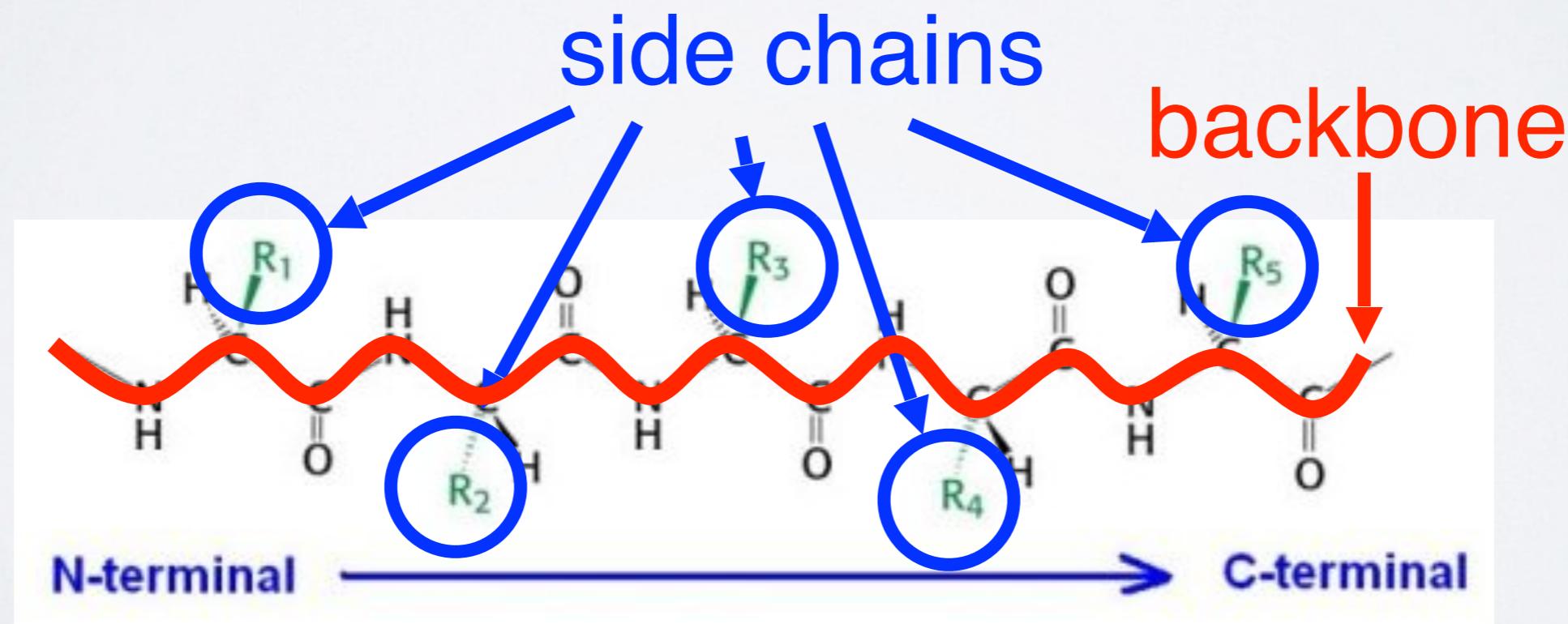
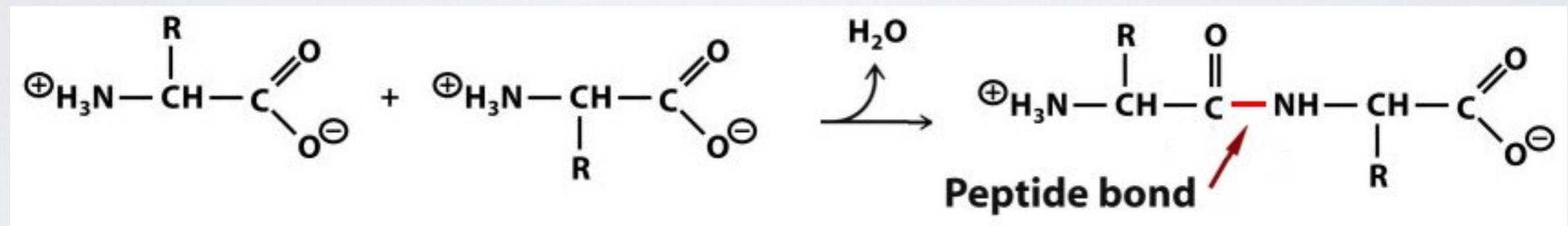


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

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

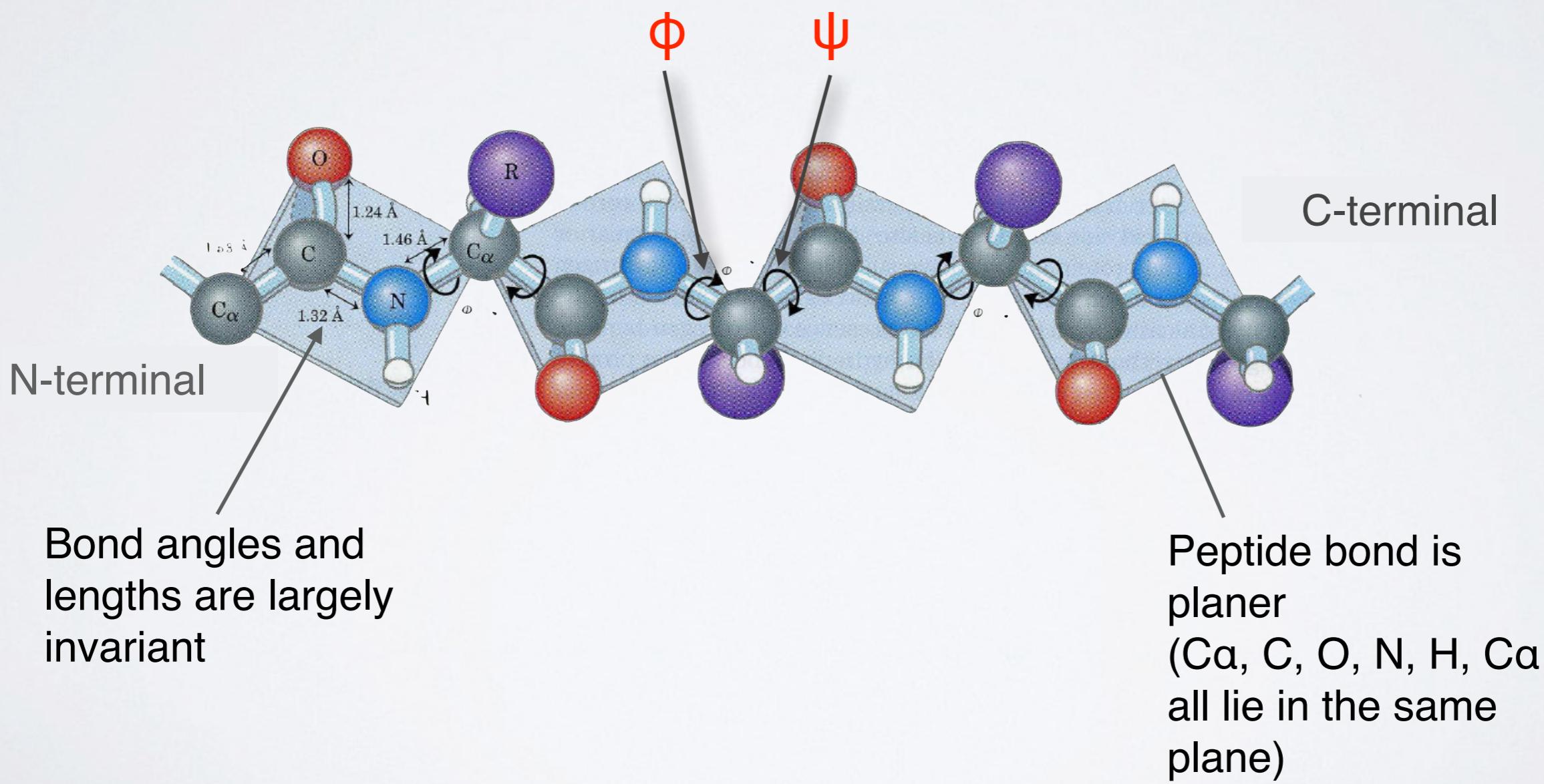
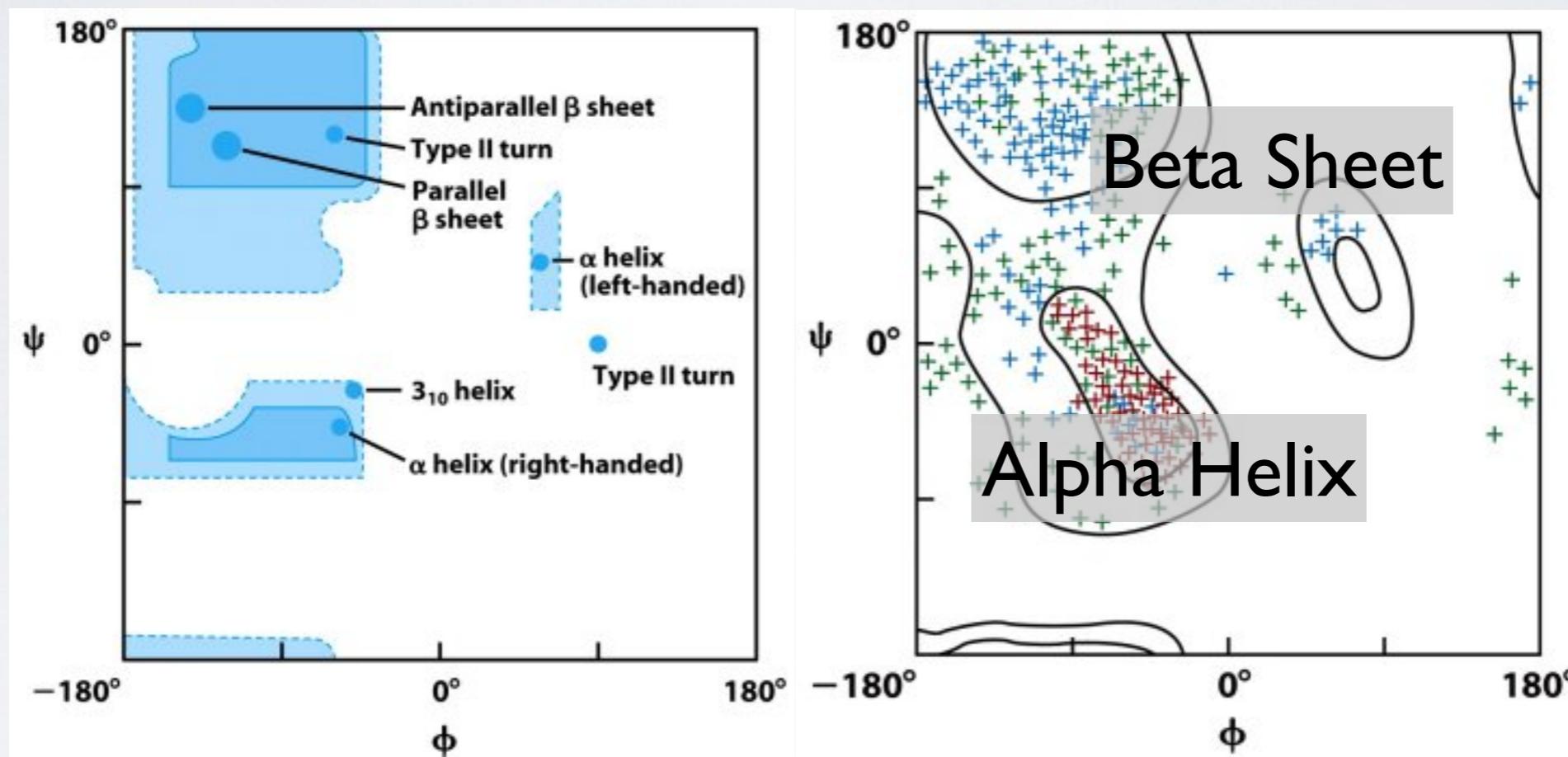


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

PHI vs PSI PLOTS ARE KNOWN AS RAMACHANDRAN DIAGRAMS

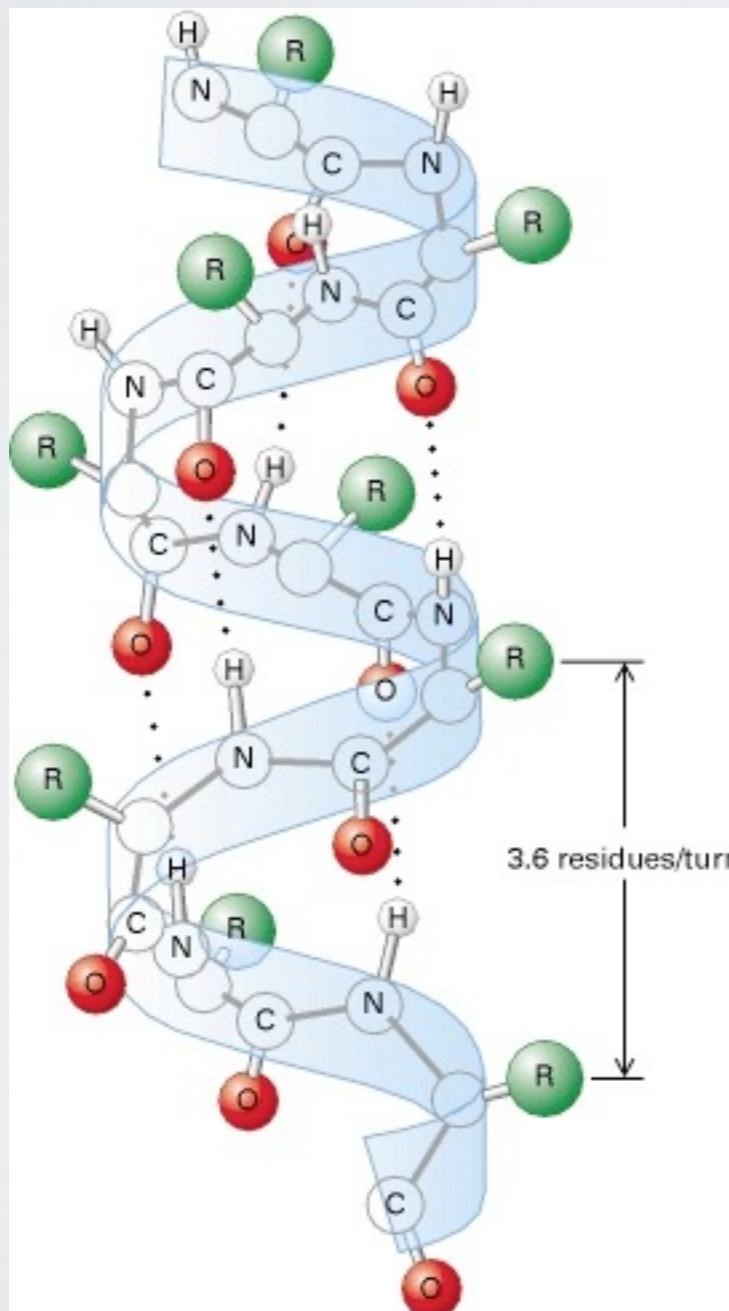


- Steric hindrance dictates torsion angle preference
- Ramachandran plot show preferred regions of ϕ and ψ dihedral angles which correspond to major forms of secondary structure

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

MAJOR SECONDARY STRUCTURE TYPES

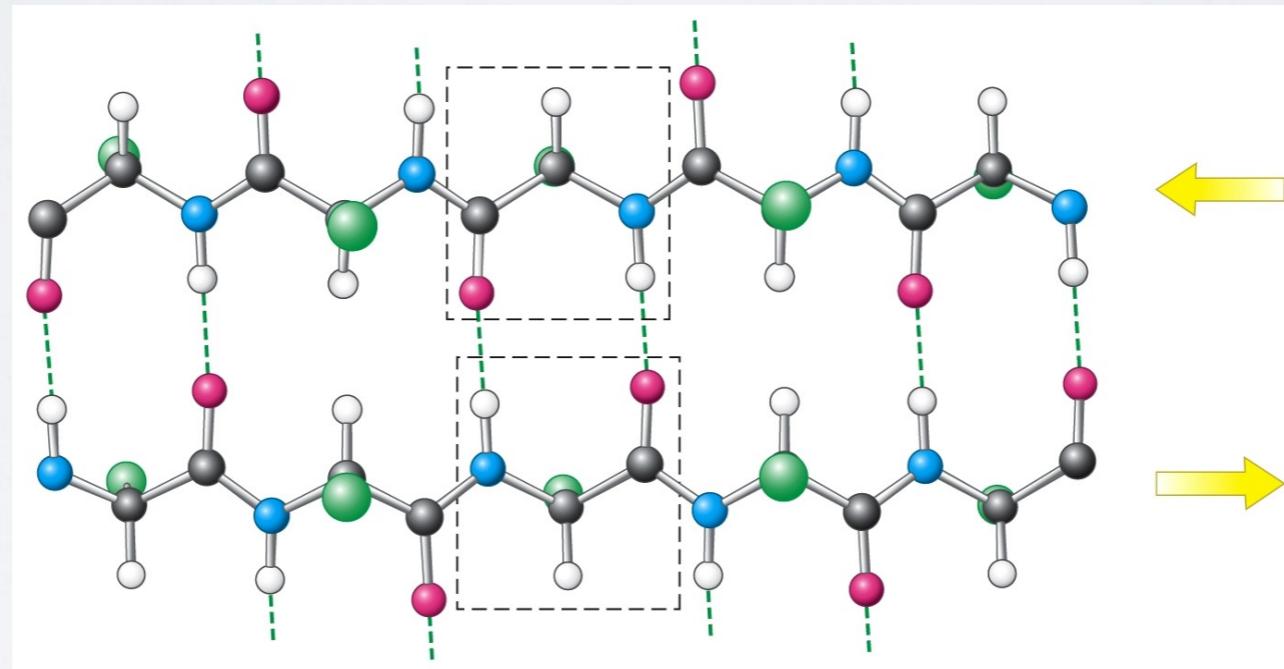
ALPHA HELIX & BETA SHEET



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**

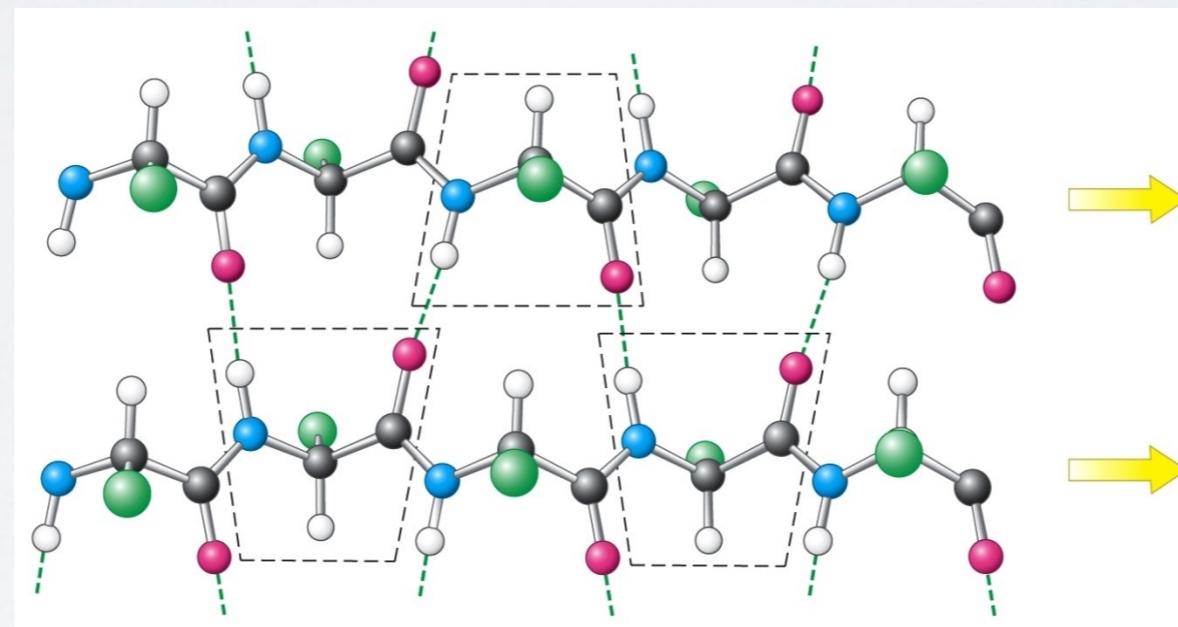


In antiparallel β -sheets

- 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/>

Protein Data Bank (PDB) is the main repository for Biomolecular structure data

<http://www.rcsb.org>

The screenshot shows the RCSB PDB website. At the top, there's a navigation bar with links for Home, Gmail, Gcal, GitHub, BIMM143, BGGN213, Atmosphere, BIMM194, Blink, News, and More. Below the navigation is a search bar with the placeholder "Search by PDB ID, author, macromolecule, sequence, or ligands" and a "Go" button. To the left, the RCSB PDB logo is displayed with the text "153085 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education". Below the logo are links for PDB-101, wwPDB, EMDataResource, NDB, and the Worldwide Protein Data Bank Foundation. On the right, there are social media icons for Facebook, Twitter, YouTube, and Google+. A sidebar on the left contains links for Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area features a section titled "A Structural View of Biology" with text about the resource being powered by the Protein Data Bank archive and its applications in biomedicine and agriculture. It also mentions the High School Antibiotic Resistance Video Challenge, which is highlighted with a purple banner showing "2019 WINNERS". Another section on the right is titled "June Molecule of the Month" and shows a 3D molecular model of MDM2 and Cancer.

RCSB PDB Deposit ▾ Search ▾ Visualize ▾ Analyze ▾ Download ▾ Learn ▾ More ▾ MyPDB

153085 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

PDB-101 Worldwide Protein Data Bank EMDataResource NDB Worldwide Protein Data Bank Foundation

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

High School Antibiotic Resistance Video Challenge

2019 WINNERS

June Molecule of the Month

MDM2 and Cancer

Contact Us

You can search by text (e.g. "ABL kinase"), PDB code (e.g. "1iep") or sequence

<http://www.rcsb.org>

The screenshot shows the RCSB PDB website interface. At the top, a red arrow points from the text "PDB code (e.g. '1iep') or sequence" to the search bar containing the text "1iep". The search bar is highlighted with a red box. Below the search bar, the page title "1iep" is also highlighted with a red box. The main content area displays the crystal structure of the C-ABL kinase domain in complex with STI-571, shown as a ribbon model. To the left of the structure is a "3D View" button. To the right, there is a "Display Files" dropdown and a "Download Files" button. The page title "1IEP" is prominently displayed at the top of the content area. Below it, the subtitle "CRYSTAL STRUCTURE OF THE C-ABL KINASE DOMAIN IN COMPLEX WITH STI-571." is shown. The "Experimental Data Snapshot" section lists the method as X-RAY DIFFRACTION, resolution as 2.1 Å, R-Value Free as 0.262, and R-Value Work as 0.231. The "wwPDB Validation" section includes a table with validation metrics and percentile ranks:

Metric	Percentile Ranks	Value
Rfree	0.264	15
Clashscore	1.3%	2.9%
Ramachandran outliers	15.7%	RSRZ outliers

You can get a **3D View** of and read details about the experiment and molecule

<http://www.rcsb.org>

The screenshot shows the RCSB PDB website interface. At the top, there is a navigation bar with links for Home, Gmail, Gcal, GitHub, BIMM143, BGGN213, Atmosphere, BIMM194, Blink, News, and More. Below the navigation bar is a secondary menu with tabs for Structure Summary, 3D View, Annotations, Sequence, Sequence Similarity, Structure Similarity, and Experiment. The '3D View' tab is highlighted with a red box and a red arrow points to it from the text above. The main content area displays the title '1IEP CRYSTAL STRUCTURE OF THE C-ABL KINASE DOMAIN IN COMPLEX WITH STI-571.' Below the title is a note: 'Note: Use your mouse to drag, rotate, and zoom in and out of the structure. Mouse-over to identify atoms and bonds. [Mouse controls documentation.](#)' To the right of the title is a panel with buttons for 'Display Files' and 'Download Files'. Further down on the right is a panel titled 'Structure View Documentation' with various settings for viewing the structure, including Assembly (Bioassembly 1), Model (Model 1), Symmetry (None), Style (Cartoon), Color (Rainbow), Ligand (Ball & Stick), Quality (Automatic), Water (unchecked), Ions (checked), Hydrogens (checked), and Clashes (unchecked). A 'Default Structure View' button is also present. A vertical 'Contact Us' button is located on the far right edge of the page.

You can display or download PDB format files for a particular entry

<http://www.rcsb.org>

The screenshot shows the RCSB PDB website interface. At the top, there is a navigation bar with links for Home, Gmail, Gcal, GitHub, BIMM143, BGGN213, Atmosphere, BIMM194, Blink, News, and More. Below the navigation bar is a secondary menu with tabs for Structure Summary, 3D View, Annotations, Sequence, Sequence Similarity, Structure Similarity, and Experiment. The '3D View' tab is currently selected. The main content area displays the title '1IEP CRYSTAL STRUCTURE OF THE C-ABL KINASE DOMAIN IN COMPLEX WITH STI-571.' Below the title is a note: 'Note: Use your mouse to drag, rotate, and zoom in and out of the structure. Mouse-over to identify atoms and bonds. [Mouse controls documentation.](#)' To the right of the title is a panel titled 'Structure View Documentation' containing various settings for viewing the structure, such as Assembly (Bioassembly 1), Model (Model 1), Symmetry (None), Style (Cartoon), Color (Rainbow), Ligand (Ball & Stick), Quality (Automatic), Water (unchecked), Ions (checked), Hydrogens (checked), and Clashes (unchecked). A red arrow points from the text 'You can display or download PDB format files for a particular entry' to the 'Download Files' button in this panel. The bottom right corner of the page has a 'Contact Us' link.

Side-Note: PDB File Format

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

		Amino Acid		Sequence/Residue Number	Coordinates			(etc.)	
	Element		Chain		X	Y	Z		
ATOM	1	N	MET	A	1	19.353	41.547	-3.887	...
ATOM	2	CA	MET	A	1	20.513	40.939	-4.592	...
ATOM	3	C	MET	A	1	20.150	39.658	-5.355	...
ATOM	4	O	MET	A	1	19.053	39.551	-5.903	...
ATOM	5	CB	MET	A	1	21.642	40.678	-3.592	...
ATOM	6	CG	MET	A	1	21.233	39.903	-2.360	...
ATOM	7	SD	MET	A	1	22.533	39.928	-1.113	...
ATOM	8	CE	MET	A	1	23.771	38.881	-1.885	...
ATOM	9	N	ASP	A	2	21.068	38.694	-5.390	...
ATOM	10	CA	ASP	A	2	20.856	37.440	-6.117	...
ATOM	11	C	ASP	A	2	20.124	36.371	-5.299	...
ATOM	12	O	ASP	A	2	20.680	35.818	-4.351	...

Element position within amino acid

Side-Note: PDB File Format

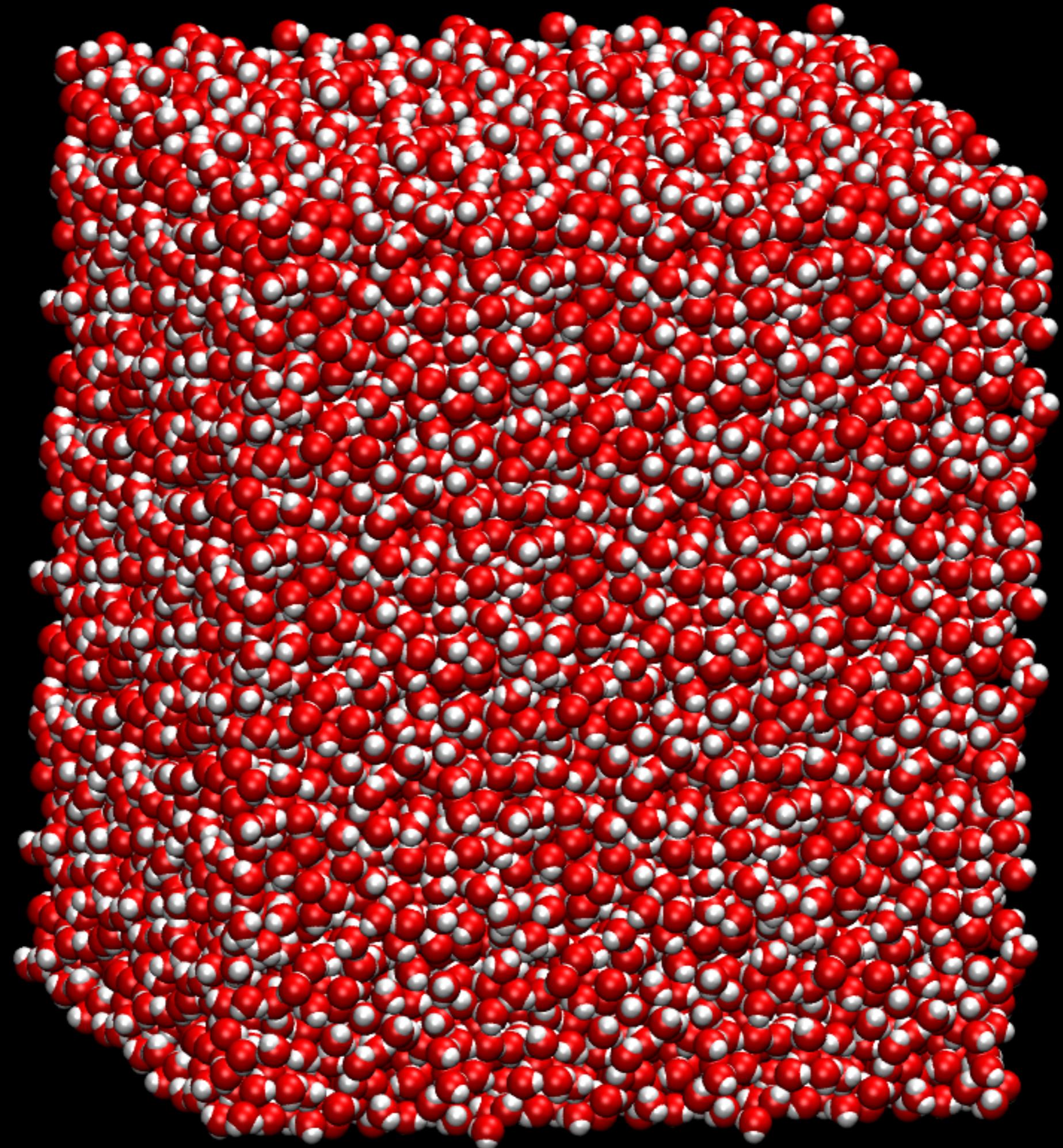
- PDB files contains atomic **coordinates** and associated information.

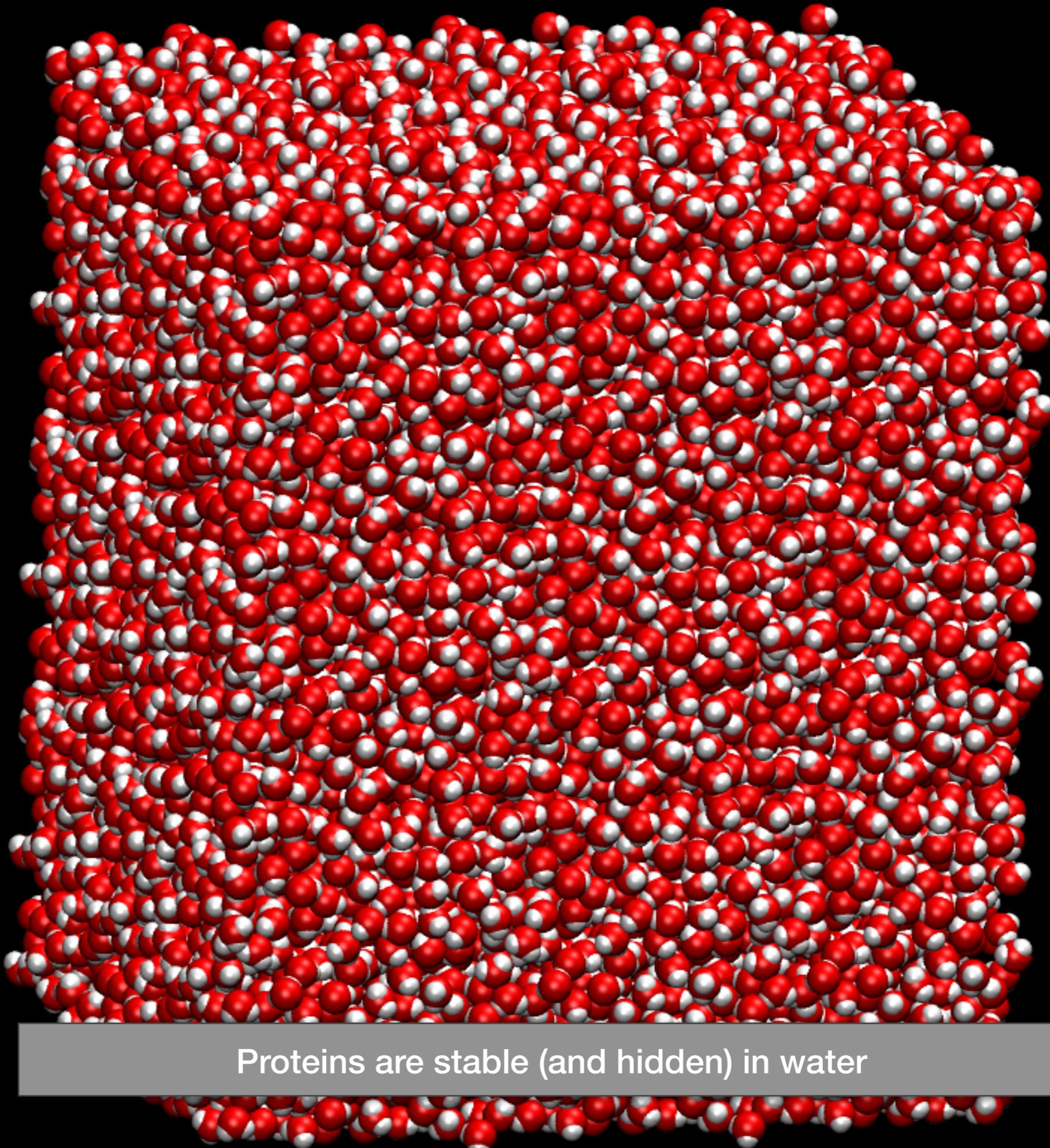
Element						
ATOM	1	N	MET	A		
ATOM	2	CA	MET	A		
ATOM	3	C	MET	A		
ATOM	4	O	MET	A		
ATOM	5	CB	MET	A		
ATOM	6	CG	MET	A		
ATOM	7	SD	MET	A		
ATOM	8	CE	MET	A		
ATOM	9	N	ASP	A		
ATOM	10	CA	ASP	A		
ATOM	11	C	ASP	A		
ATOM	12	O	ASP	A		

The diagram illustrates the chemical structure of a Metionine (Met) amino acid. It features a central alpha carbon atom (labeled 'α carbon') bonded to a hydrogen atom (H), a methyl group (CH₃), a carboxyl group (-C(=O)O⁻), and an amino group (-NH₃⁺). The methyl group is further substituted with four methylene groups (CH₂) and an amino group (-NH₃⁺). The labels 'ε', 'δ', 'γ', and 'β' point to the carbons of the side chain, which is labeled 'side chain (R group)'. The 'amino group' and 'carboxyl group' are also labeled.

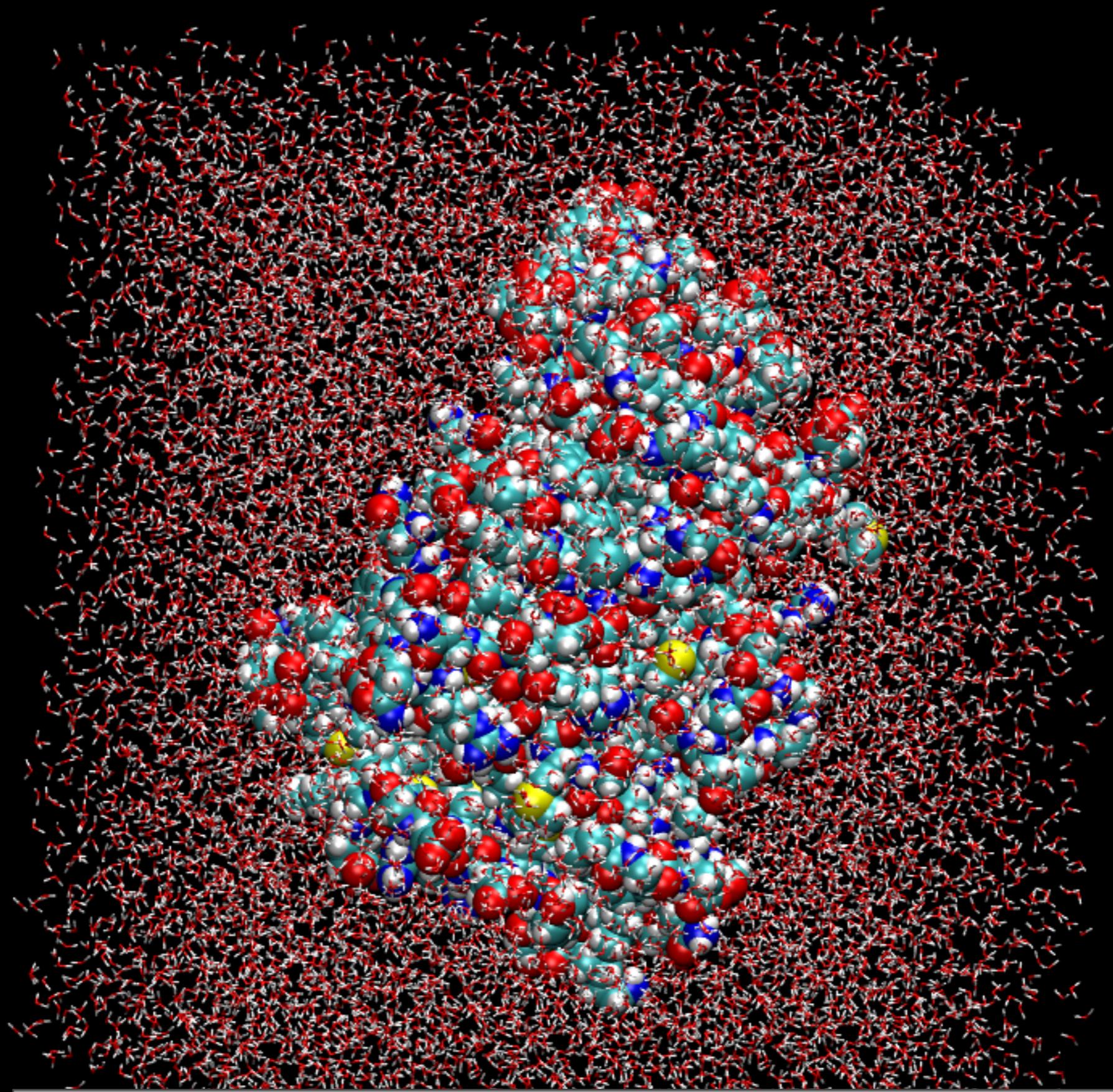
Element position within amino acid

What Does a Protein Look like?

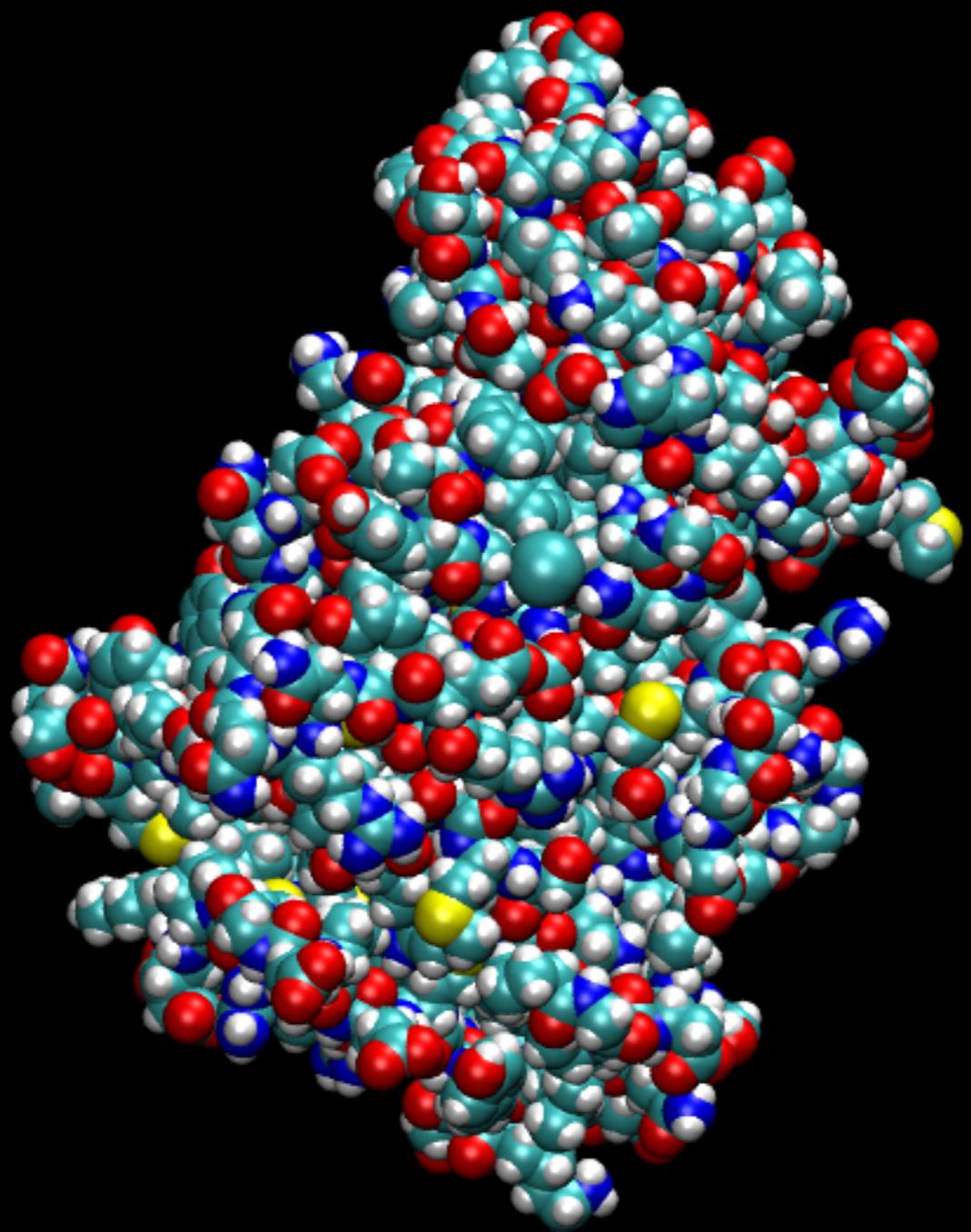




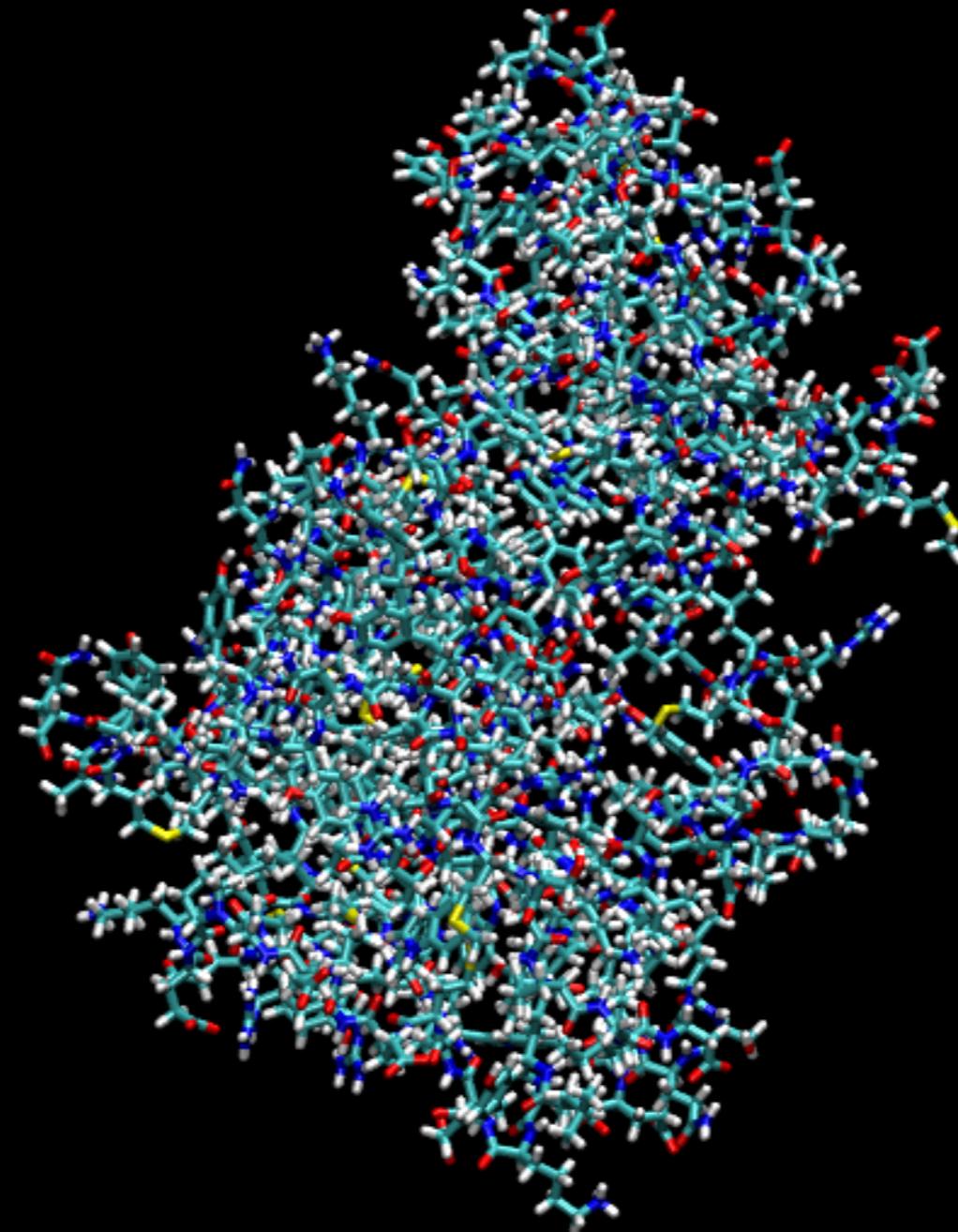
Proteins are stable (and hidden) in water



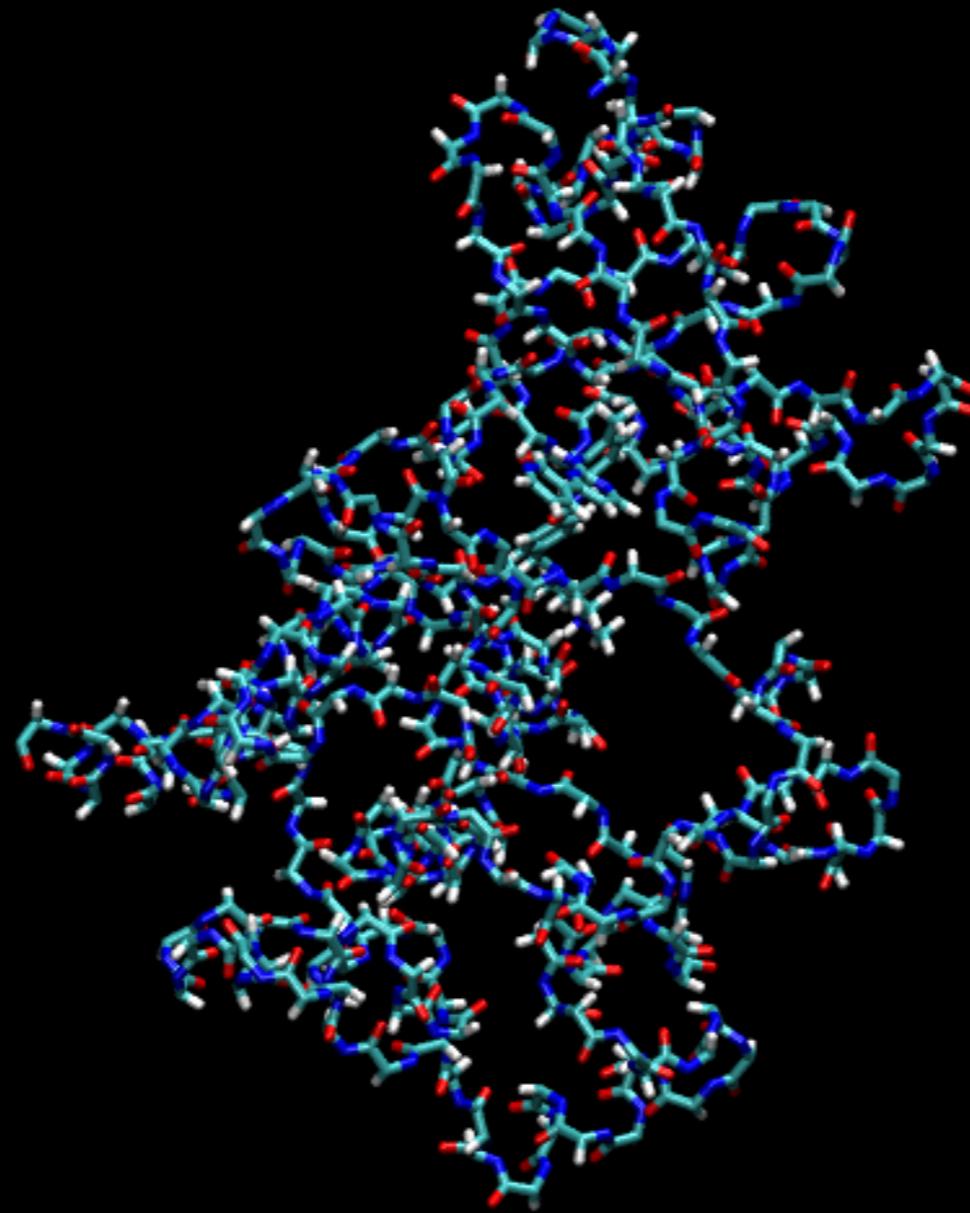
Proteins closely interact with water



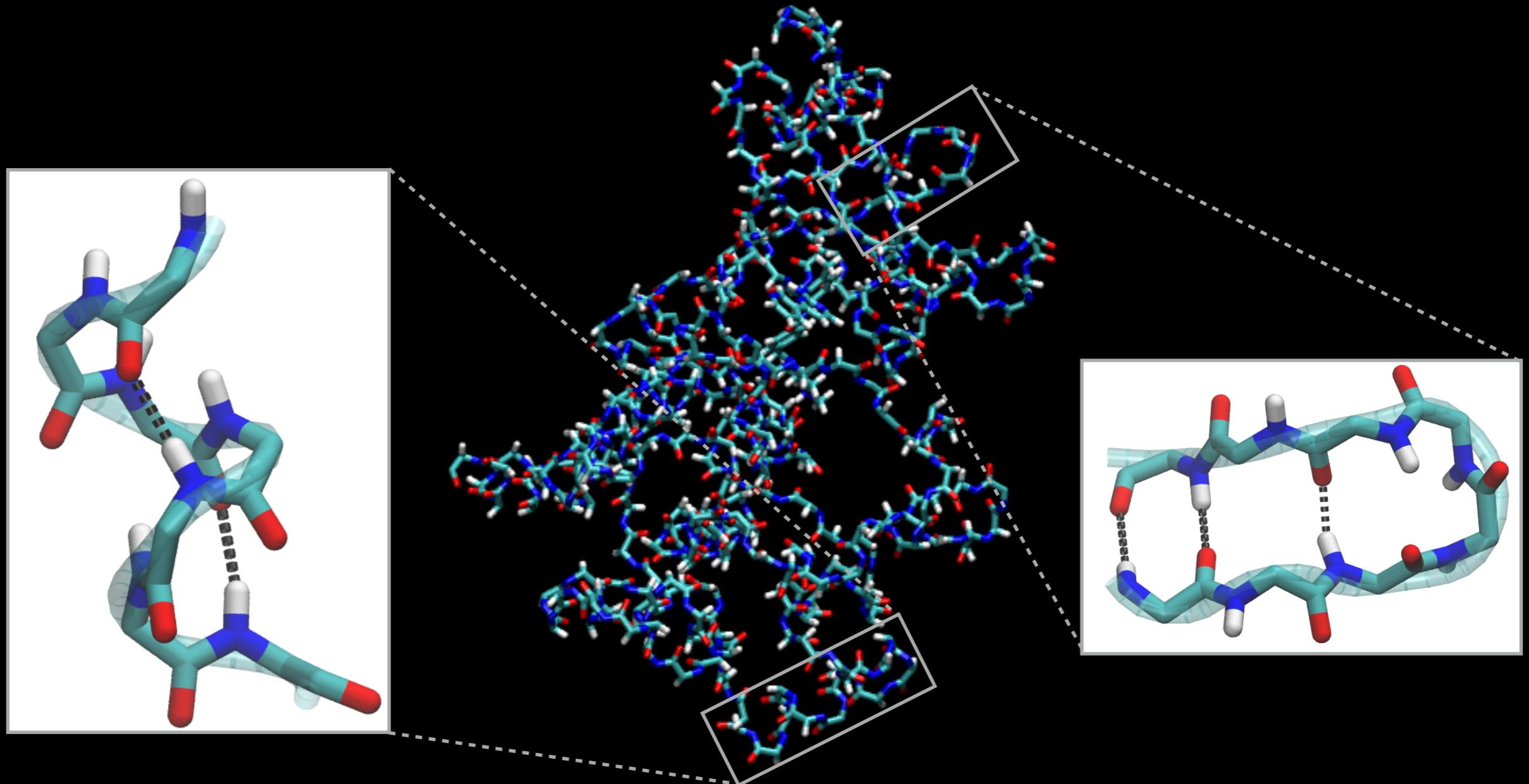
Proteins are close packed solid but flexible objects (globular)



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



Tube or trace representation is one of the simplest views



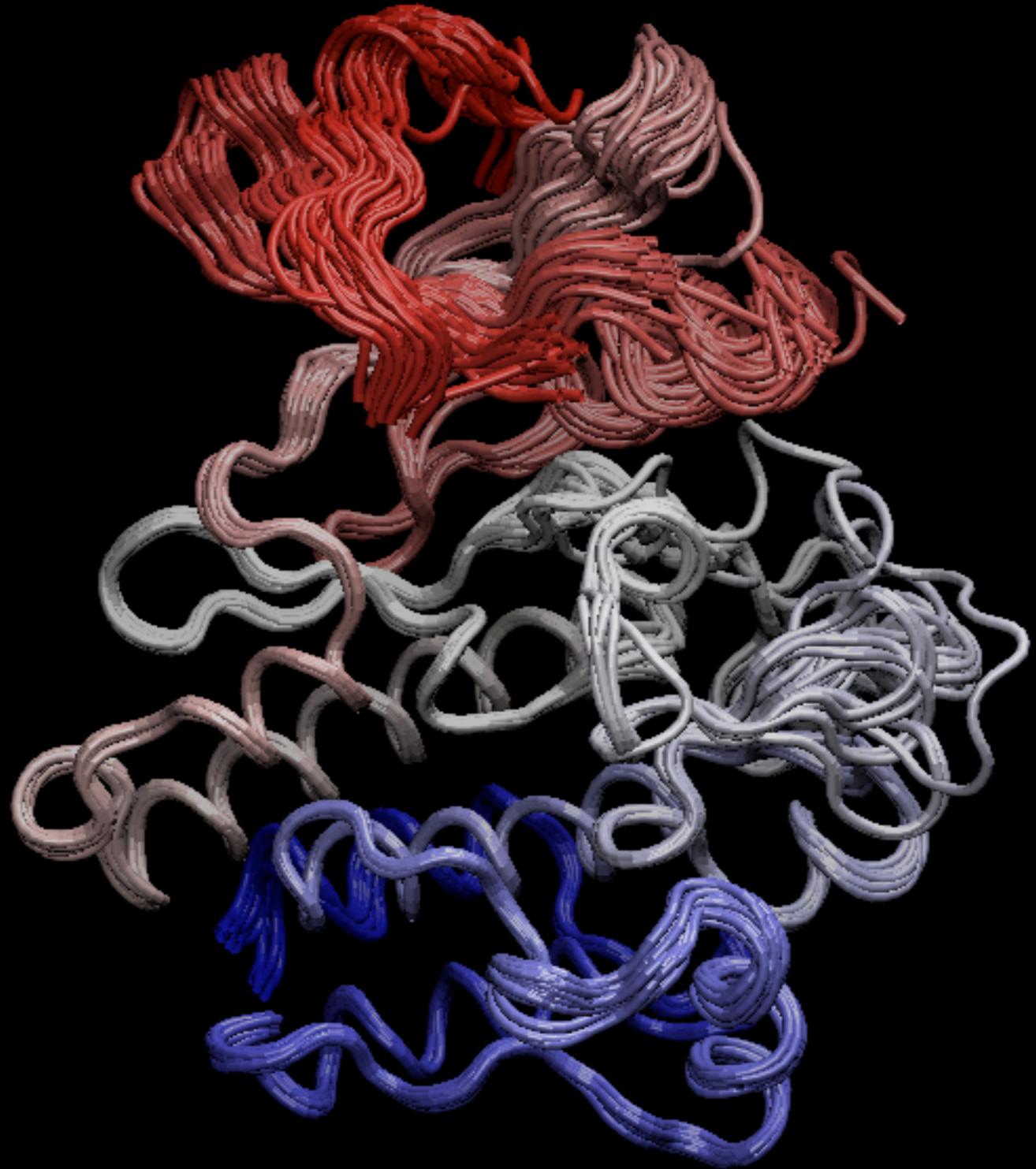
Tube with added colors to highlight secondary structure



Simplified "cartoon" secondary structure representations are commonly used to communicate structural details



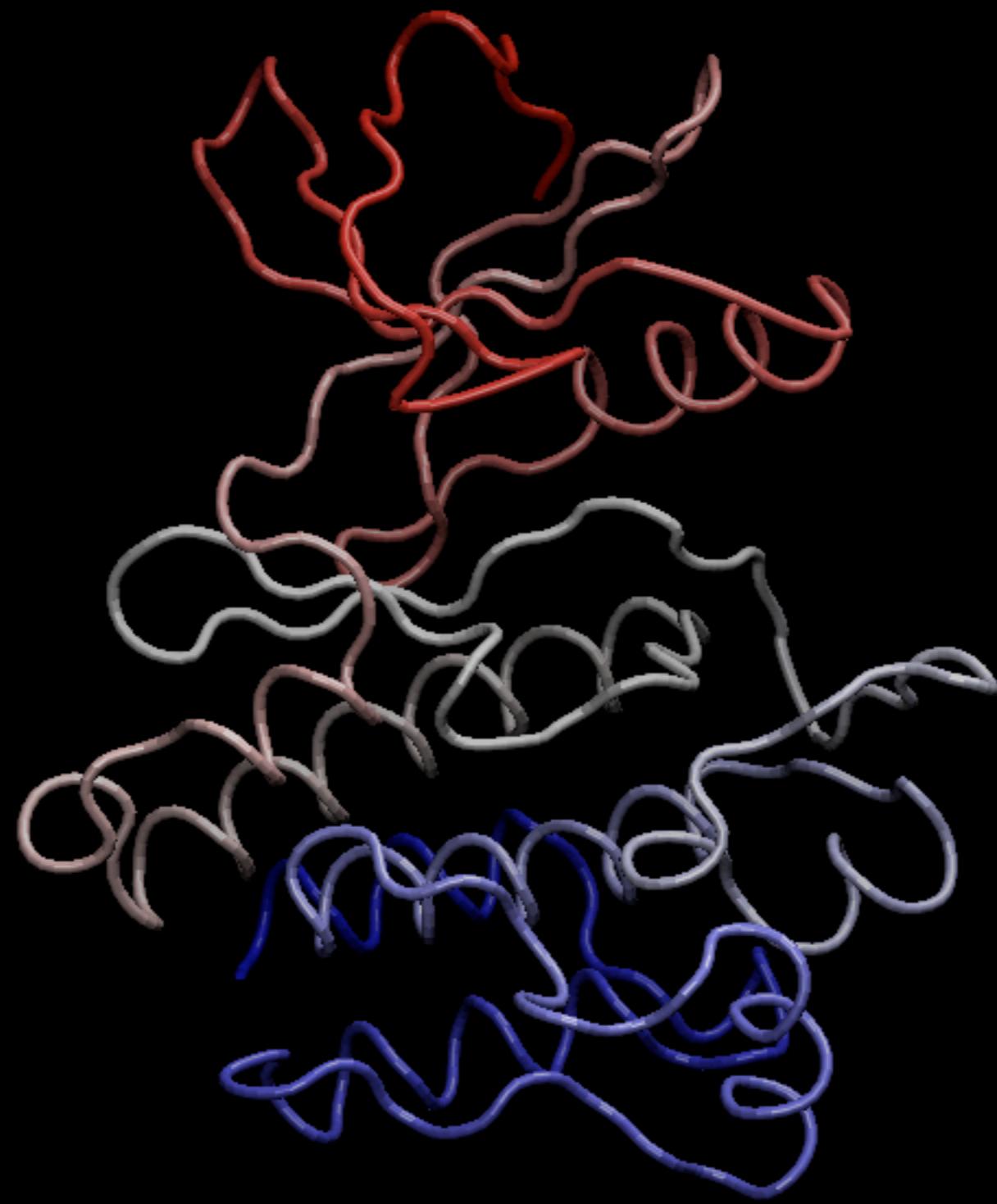
Viewing in 3D is often essential for interpretation.
Now we can clearly see 2° and 3° structure - the
coiled chain of connected secondary structures



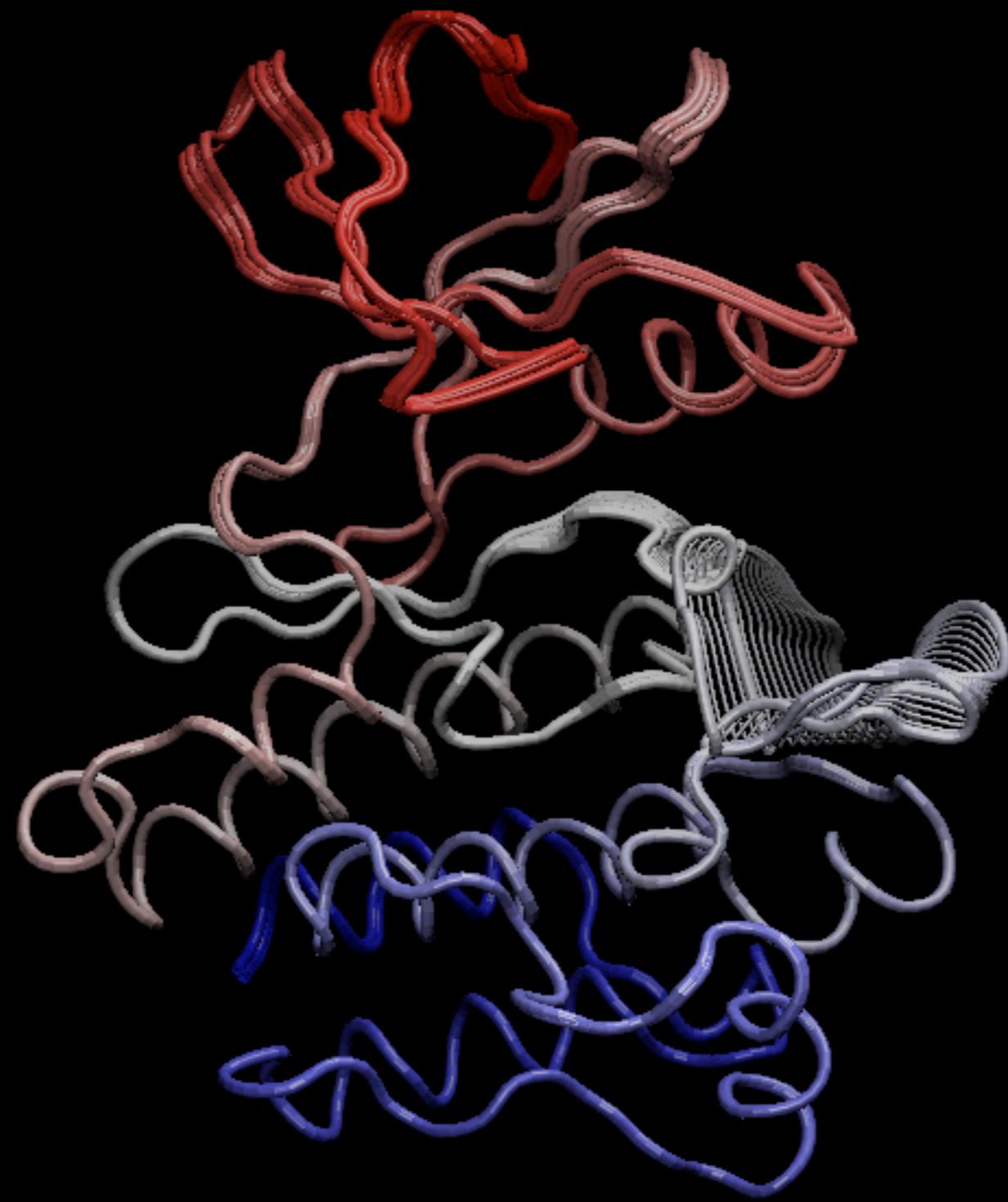
Viewing multiple superposed structures solved under different conditions can highlight flexible regions



Viewing multiple superposed structures solved under different conditions can highlight distinct conformations



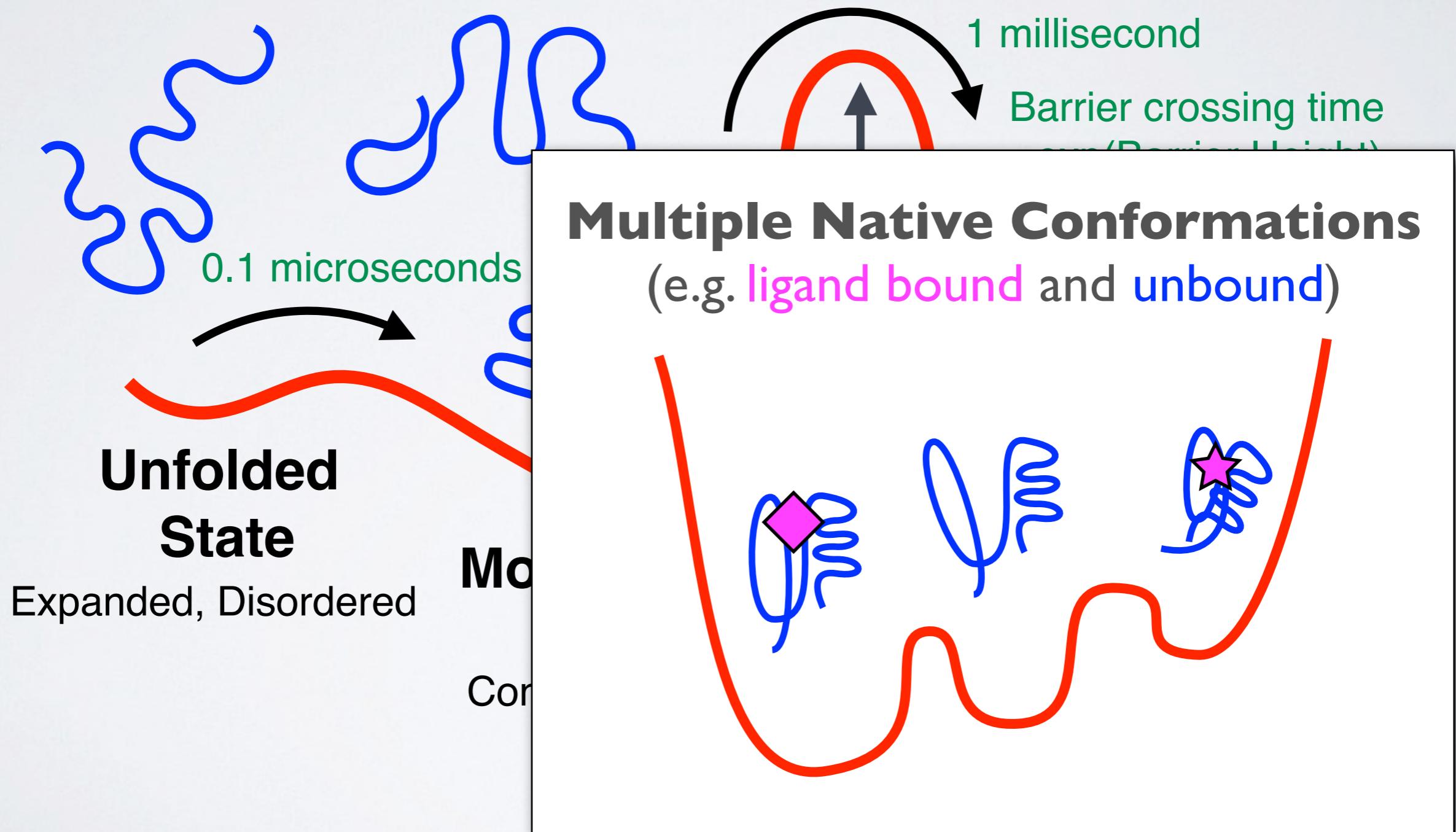
Analyzing these multiple structures can reveal **functional motions**
- i.e. displacements that are essential for regulating function



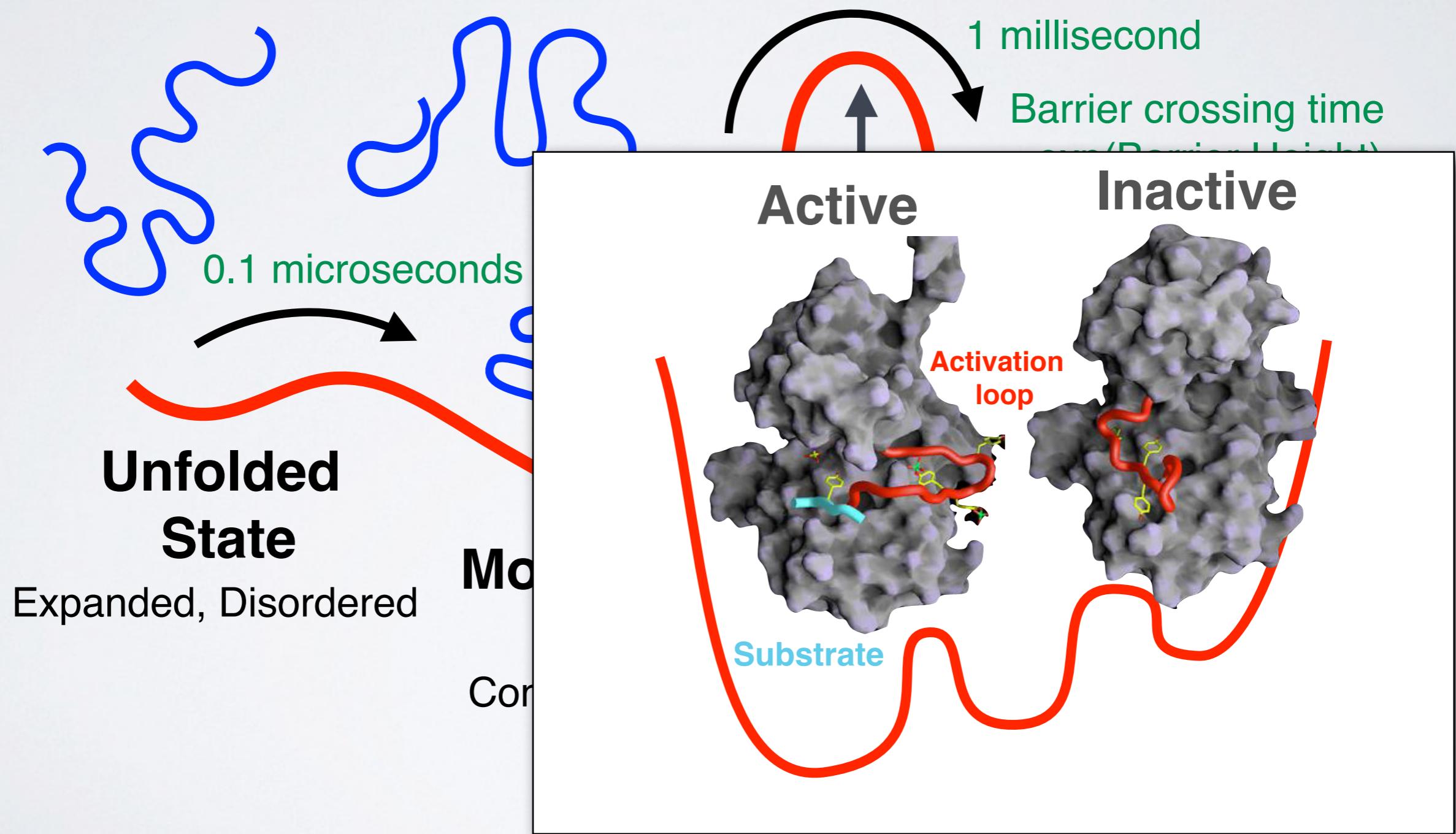
"Activation loop"

Analyzing these multiple structures can reveal functional motions
- i.e. displacements that are essential for regulating function

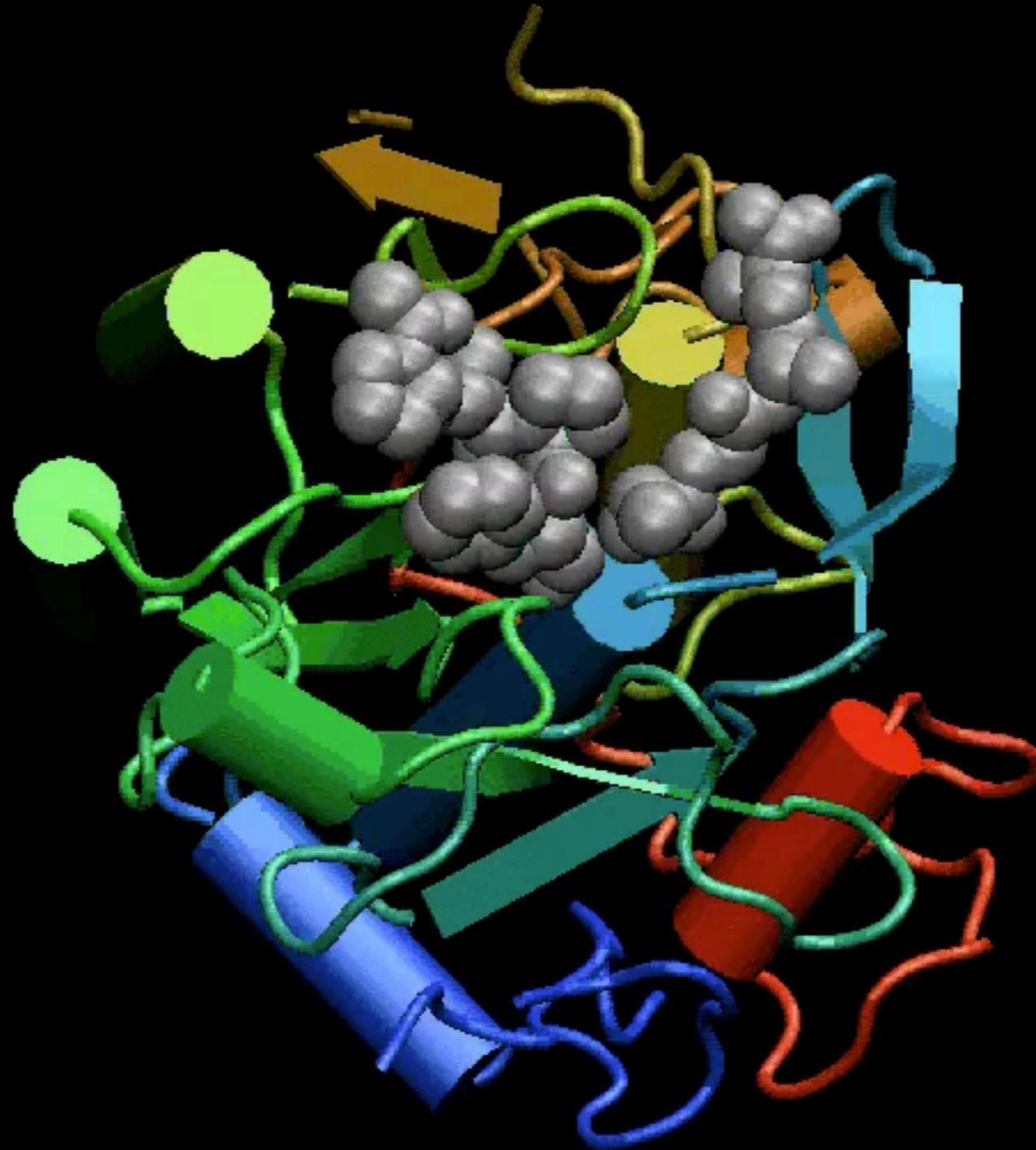
KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: ENERGY LANDSCAPE



Normal Mode Analysis (NMA) models the protein as a network of elastic strings



NMA is a bioinformatics method to predict the intrinsic dynamics of biomolecules

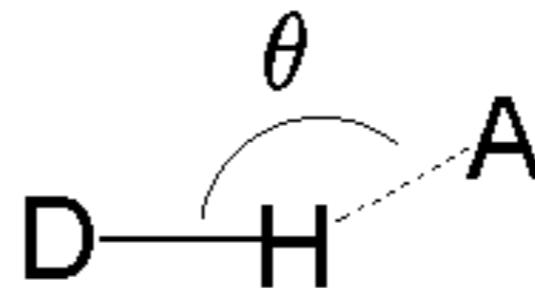
Key forces affecting structure:

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

Hydrogen-bond donor Hydrogen-bond acceptor



← d →

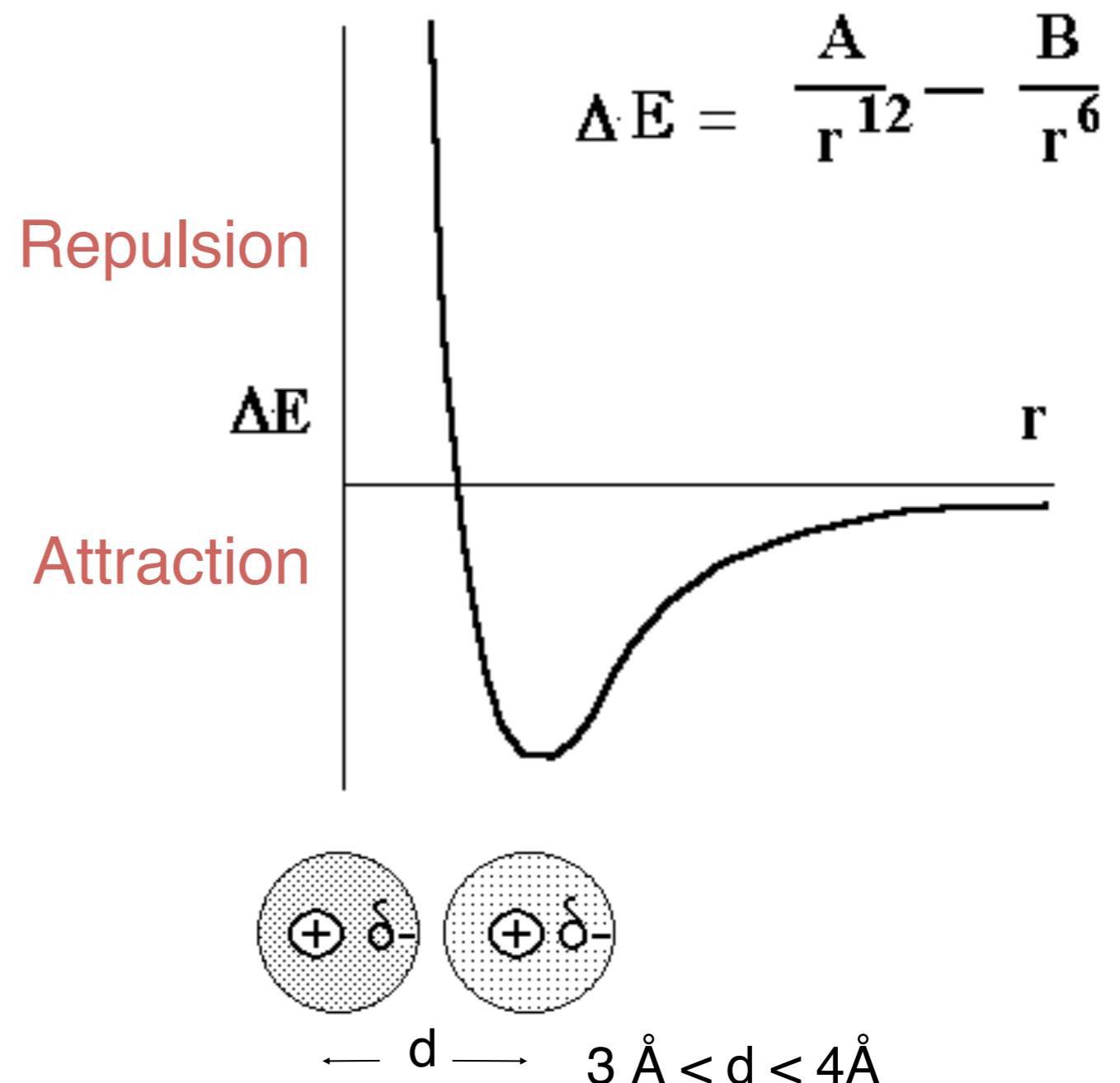


$2.6 \text{ \AA} < d < 3.1 \text{ \AA}$

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

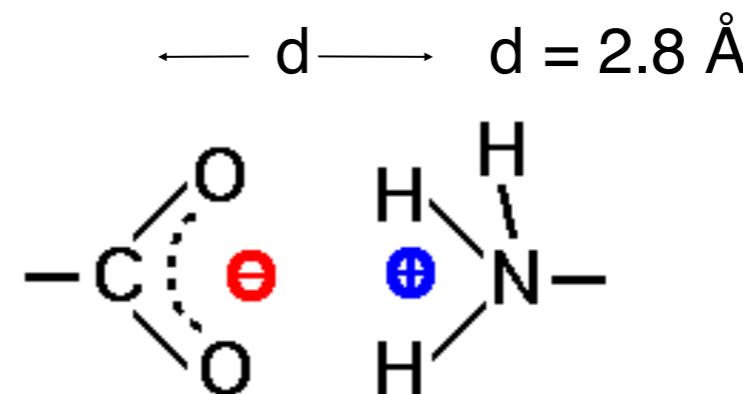
Key forces affecting structure:

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



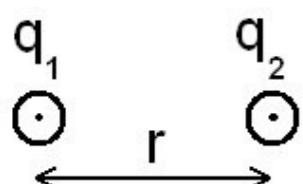
Key forces affecting structure:

- 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

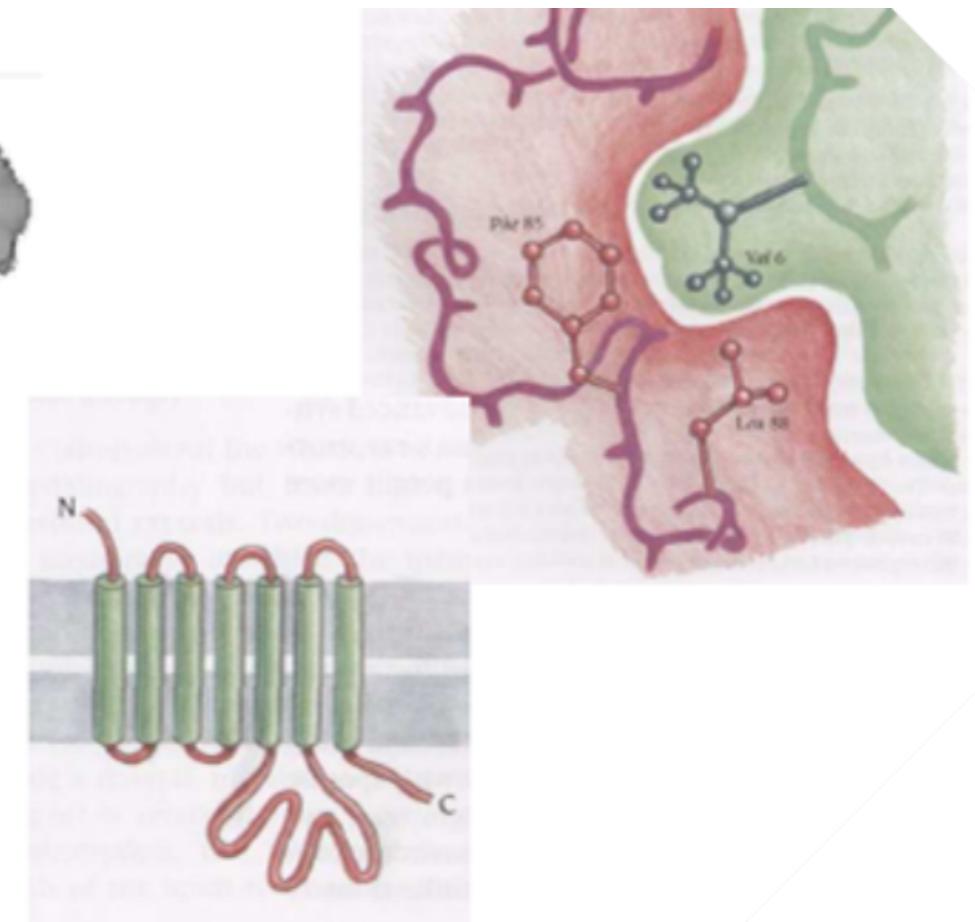
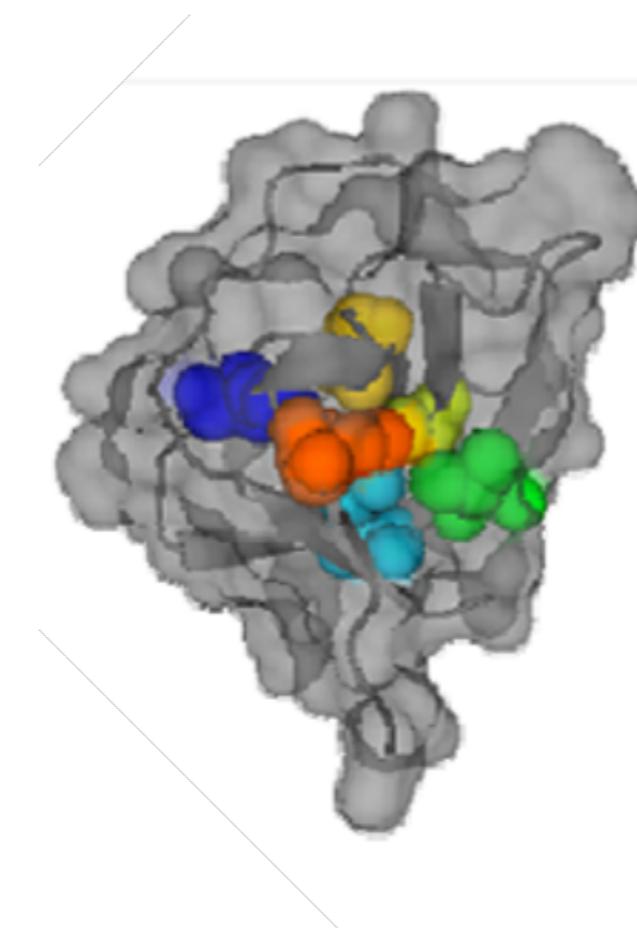
D = Dielectric constant (vacuum = 1; H₂O = 80)

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.

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

Do it Yourself!

Hand-on time!

Focus on **section 1** only please!

[~20 mins]

N.B. Remember to make your new **class11** RStudio project inside your GitHub tracked directory from last day and **UNCHECK** the "Create a Git repository" option...

Side-Note: PDB File Format

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

	Element	Amino Acid	Sequence/Residue Number		Coordinates			(etc.)
			X	Y	Z			
ATOM	1	N	MET A	1	19.353	41.547	-3.887	...
ATOM	2	CA	MET A	1	20.513	40.939	-4.592	...
ATOM	3	C	MET A	1	20.150	39.658	-5.355	...
ATOM	4	O	MET A	1	19.053	39.551	-5.903	...
ATOM	5	CB	MET A	1	21.642	40.678	-3.592	...
ATOM	6	CG	MET A	1	21.233	39.903	-2.360	...
ATOM	7	SD	MET A	1	22.533	39.928	-1.113	...
ATOM	8	CE	MET A	1	23.771	38.881	-1.885	...
ATOM	9	N	ASP A	2	21.068	38.694	-5.390	...
ATOM	10	CA	ASP A	2	20.856	37.440	-6.117	...
ATOM	11	C	ASP A	2	20.124	36.371	-5.299	...
ATOM	12	O	ASP A	2	20.680	35.818	-4.351	...

Element position within amino acid

Side-Note: PDB File Format

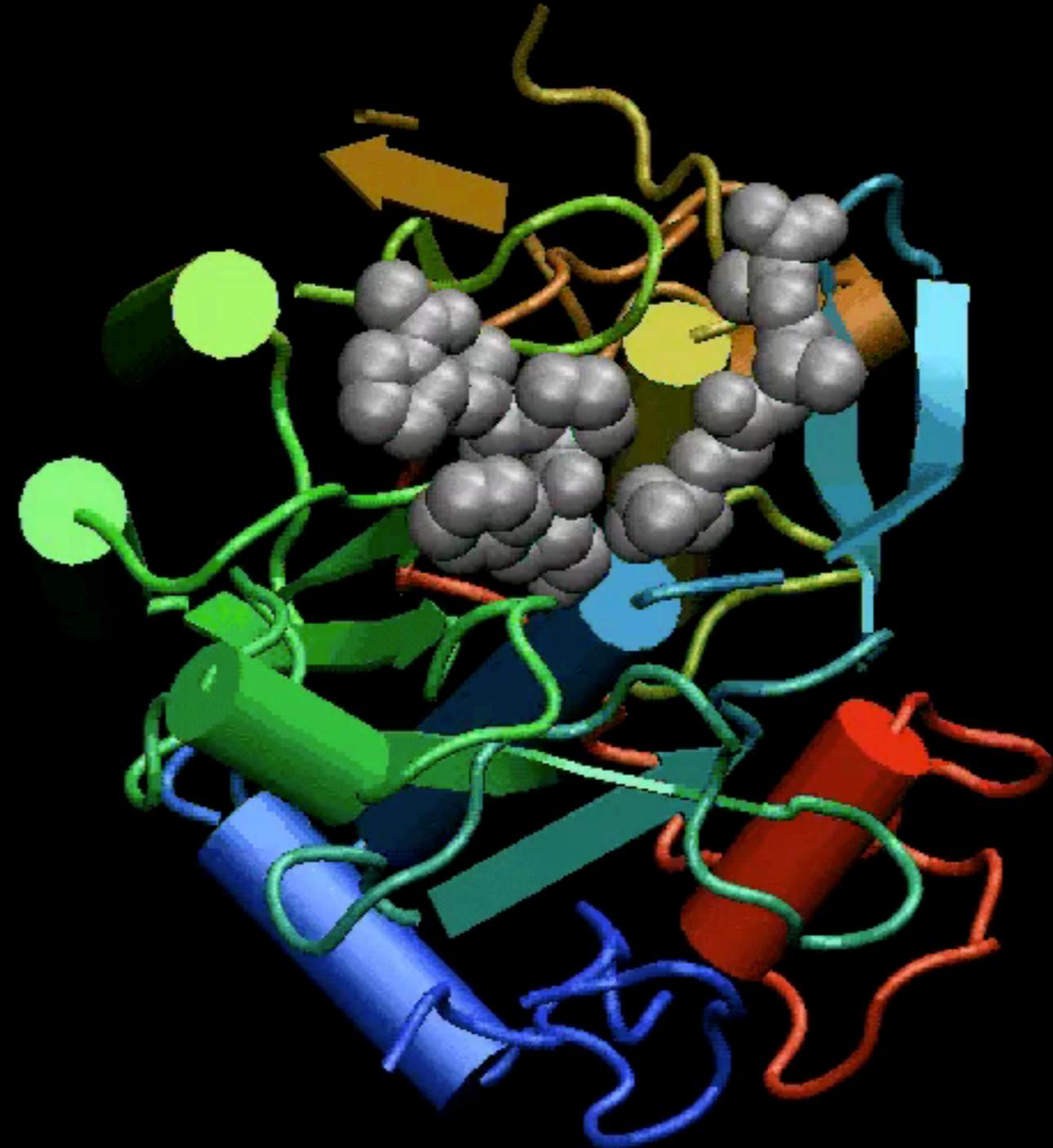
- PDB files contains atomic **coordinates** and associated information.

Element						
ATOM	1	N	MET	A		
ATOM	2	CA	MET	A		
ATOM	3	C	MET	A		
ATOM	4	O	MET	A		
ATOM	5	CB	MET	A		
ATOM	6	CG	MET	A		
ATOM	7	SD	MET	A		
ATOM	8	CE	MET	A		
ATOM	9	N	ASP	A		
ATOM	10	CA	ASP	A		
ATOM	11	C	ASP	A		
ATOM	12	O	ASP	A		

The diagram illustrates the chemical structure of a Metionine (Met) amino acid. It features a central alpha carbon atom (labeled 'α carbon') bonded to a hydrogen atom (H), a methyl group (CH₃), a carboxyl group (-C(=O)O⁻), and an amino group (-NH₃⁺). The methyl group is further substituted with four methylene groups (CH₂) and an amino group (-NH₃⁺). The labels 'ε', 'δ', 'γ', and 'β' are positioned above the respective methylene groups, indicating their positions along the side chain. A pink shaded box highlights the side chain (R group) of the Metionine residue. Labels 'amino group' and 'carboxyl group' point to their respective functional groups.

Element position within amino acid

[Download VMD](#)



Hands-on Time!

Focus on **section 2** of "Lab Sheet" (using VMD)
[~30 mins]

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

Do it Yourself!

Hand-on time!

Focus on **section 3 to 5**

[~60 mins]

Bio3D view()

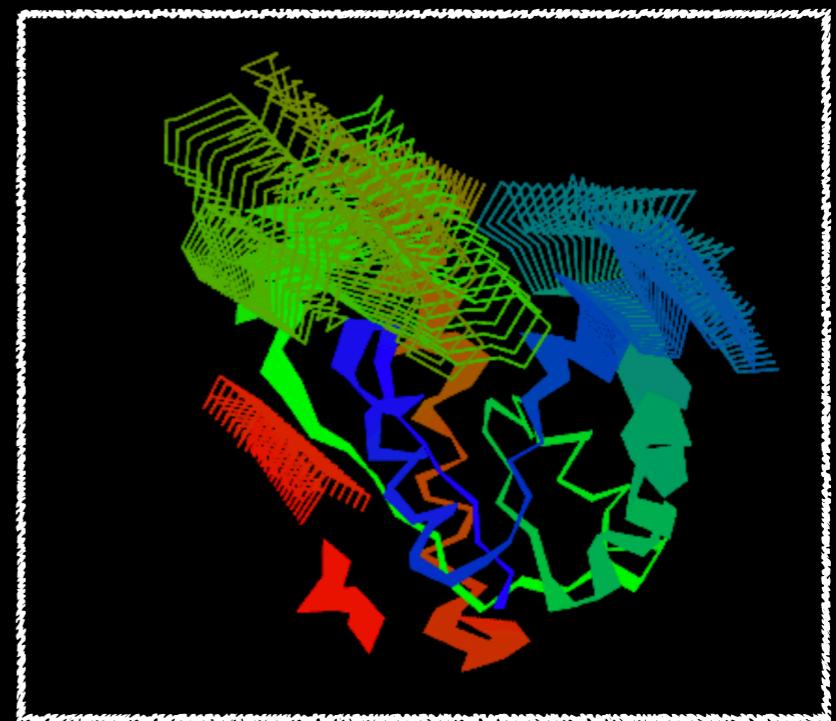
- If you want the 3D viewer in your R markdown you can install the development version of `bio3d.view`

- In your R console:

```
> install.packages("devtools")
• > devtools::install_bitbucket("Grantlab/bio3d-view")
```

- To use in your R session:

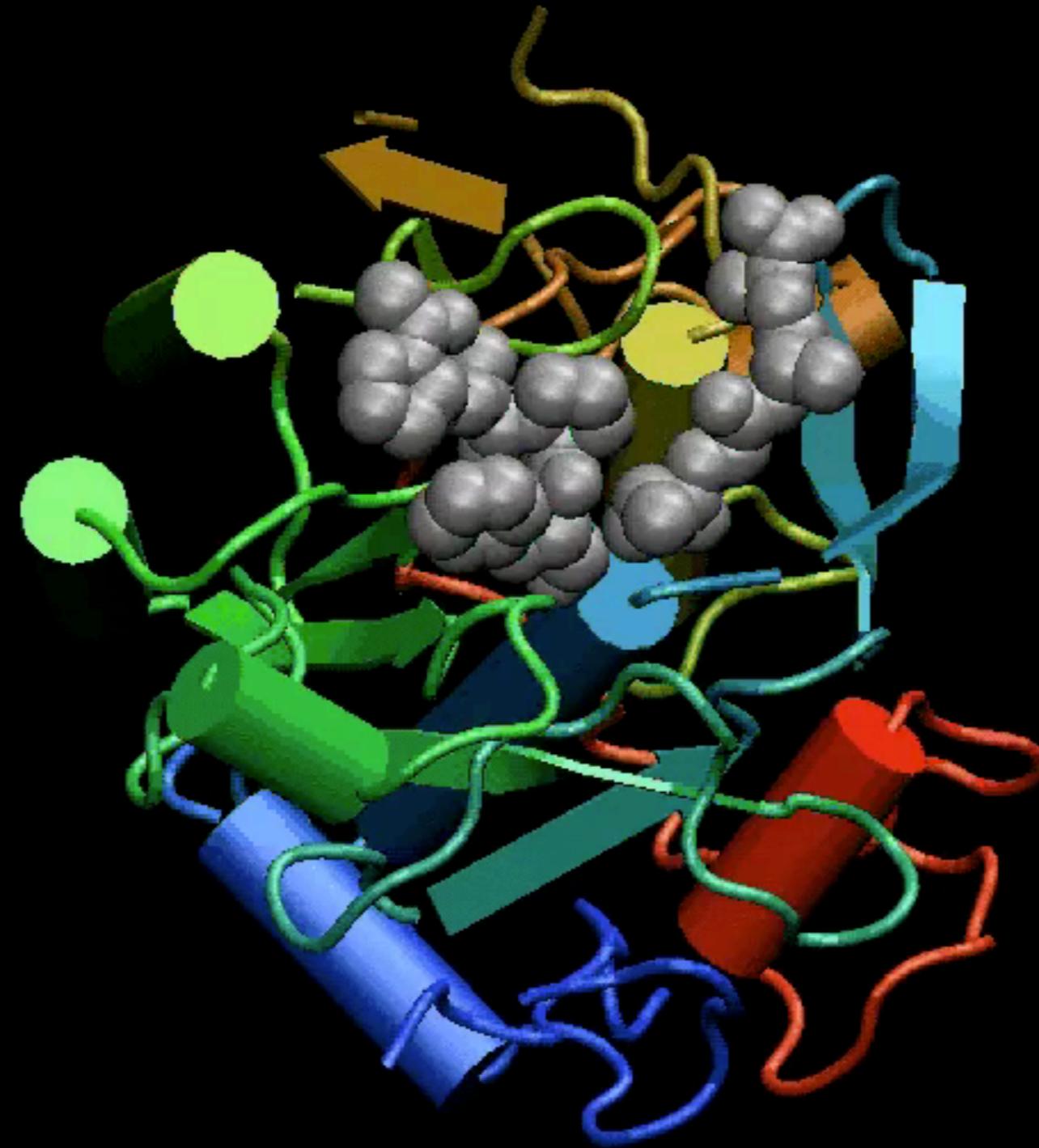
```
> library("bio3d.view")
> pdb <- read.pdb("5p21")
> view(pdb)
> view(pdb, "overview", col="sse")
```



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

NMA models the protein as a network of elastic strings



Proteinase K

NMA in Bio3D

- Normal Mode Analysis (NMA) is a bioinformatics method that can predict the major motions of biomolecules.

```
```{r}
library(bio3d)
library(bio3d.view)
```
```

```
```{r}
pdb <- read.pdb("1hel")
modes <- nma(pdb)
m7 <- mktrj(modes, mode=7, file="mode_7.pdb")

view(m7, col=vec2color(rmsf(m7)))
```
```

Bio3D view()

- If you want the interactive 3D viewer in **Rmd** rendered to output: **html_output** document:

```
```{r}
library(bio3d.view)
library(rgl)
````
```

```
```{r}
modes <- nma(read.pdb("1hel"))
m7 <- mktrj(modes, mode=7, file="mode_7.pdb")

view(m7, col=vec2color(rmsf(m7)))
rglwidget(width=500, height=500)
````
```

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

Two main approaches:

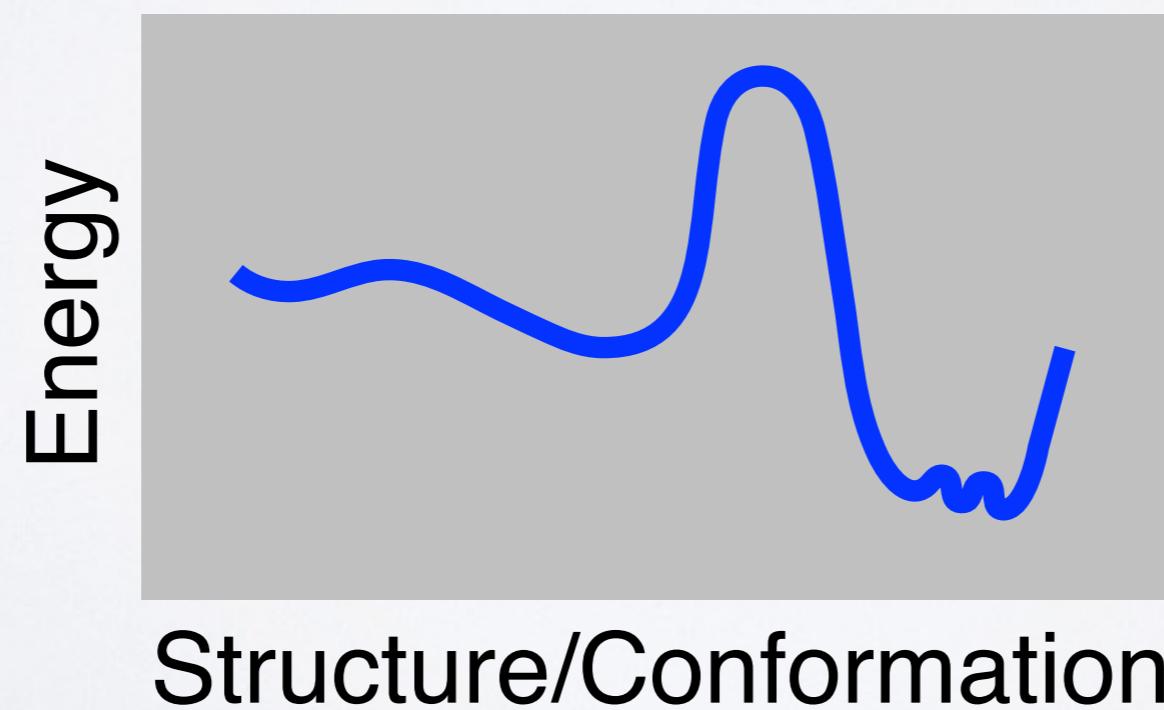
- (1). Physics-Based
- (2). Knowledge-Based

KEY CONCEPT: POTENTIAL FUNCTIONS

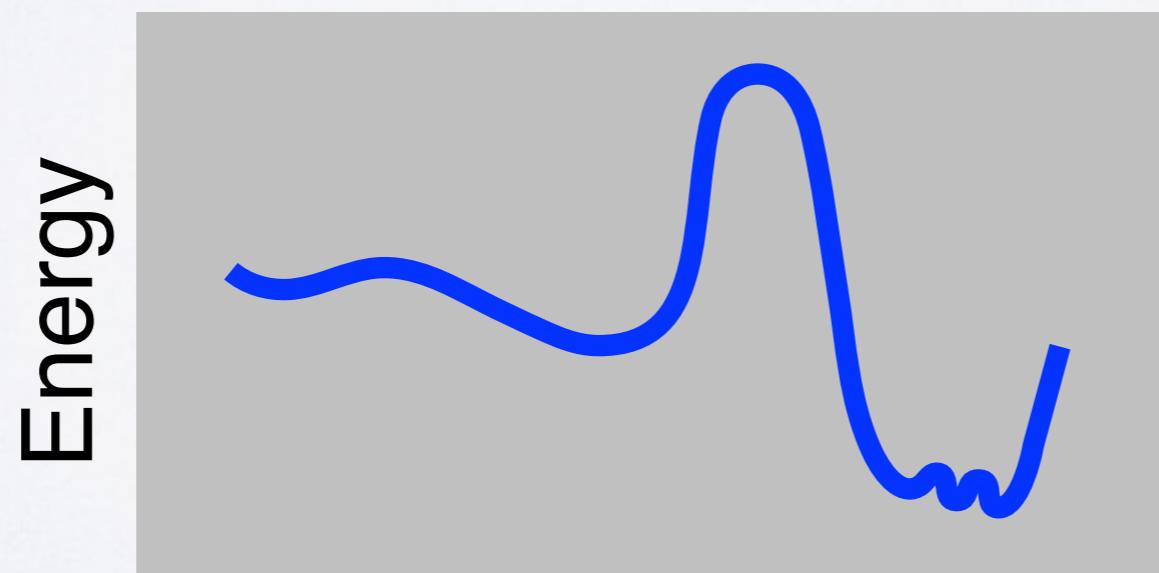
DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION
OF ITS **STRUCTURE**

Two main approaches:

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



This will be the focus of the next class!



Structure/Conformation

SUMMARY

- Structural bioinformatics is computer aided structural biology
- Described major motivations, goals and challenges of structural bioinformatics
- Reviewed the fundamentals of protein structure
- Explored how to use R to perform advanced custom structural bioinformatics analysis!

- Introduced both physics and knowledge based modeling approaches for describing the structure, energetics and dynamics of proteins computationally

[Muddy Point Assessment]