

“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

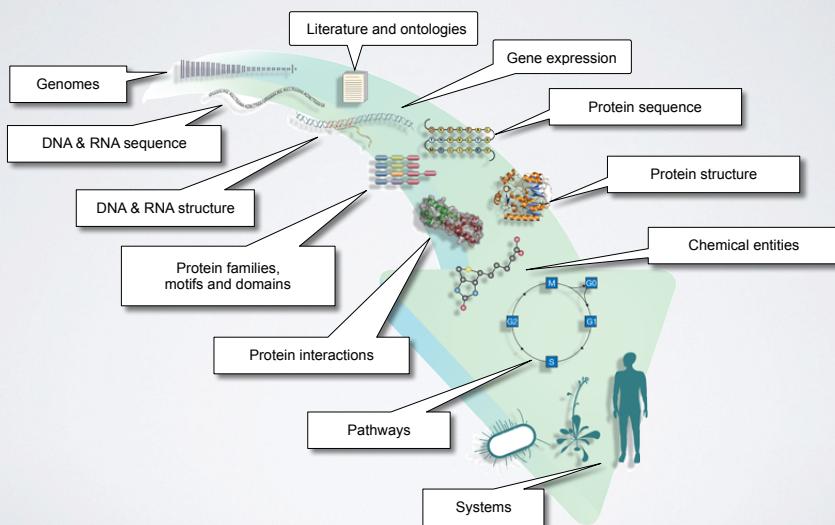
Why should we care?

Why should we care?

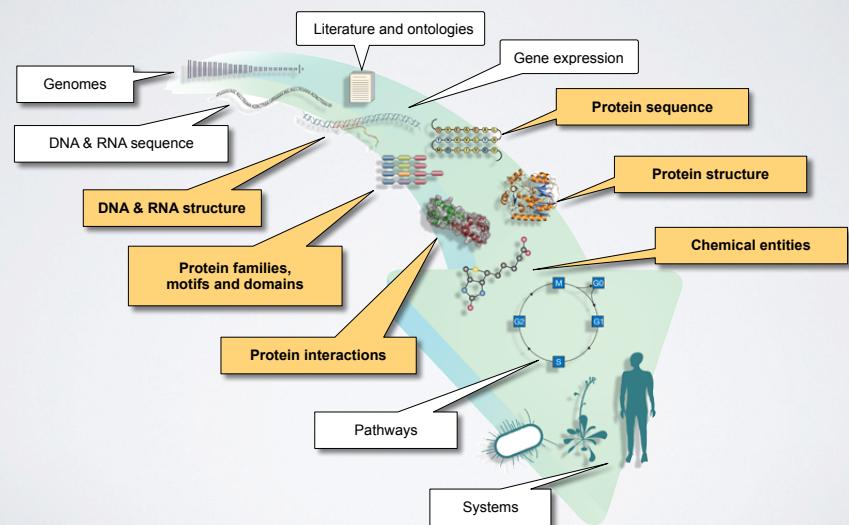
Because biomolecules are “nature’s robots”

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

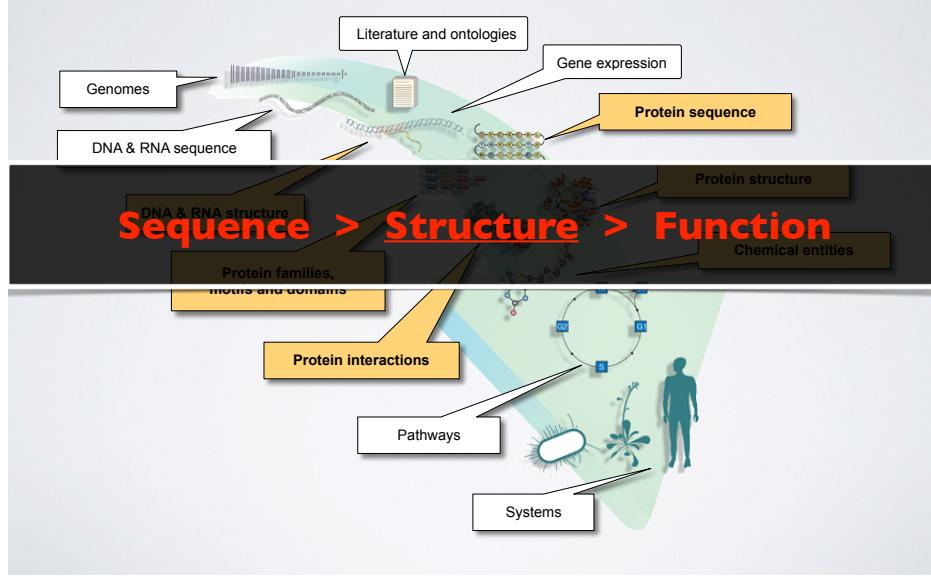
BIOINFORMATICS DATA



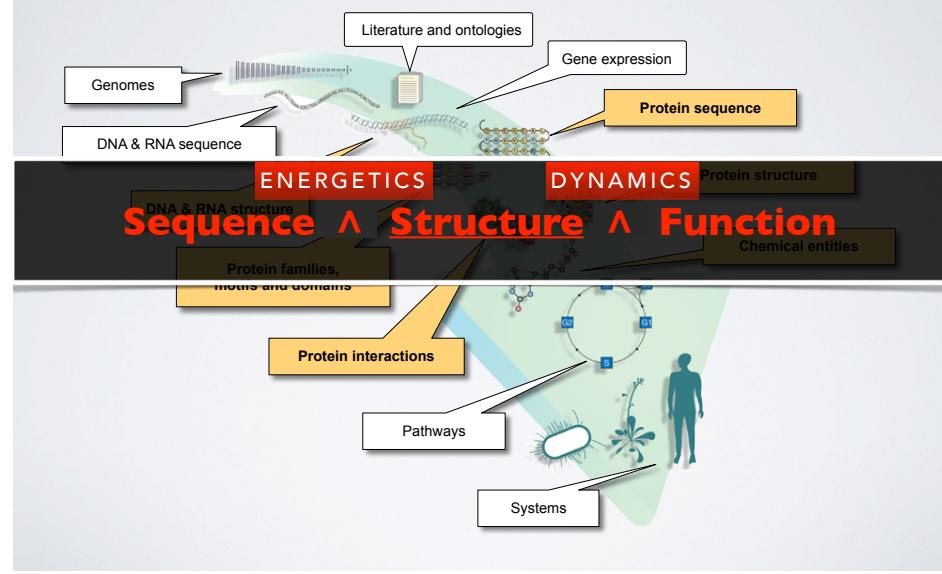
STRUCTURAL DATA IS CENTRAL

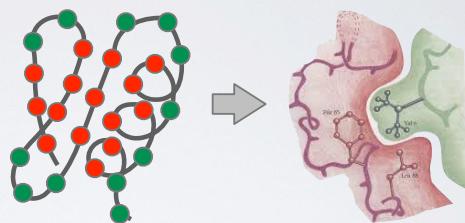
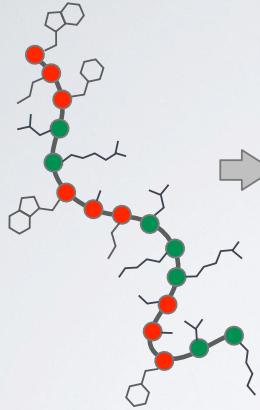


STRUCTURAL DATA IS CENTRAL



STRUCTURAL DATA IS CENTRAL





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



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

Genomics is a great start

Track Bike – DL 175

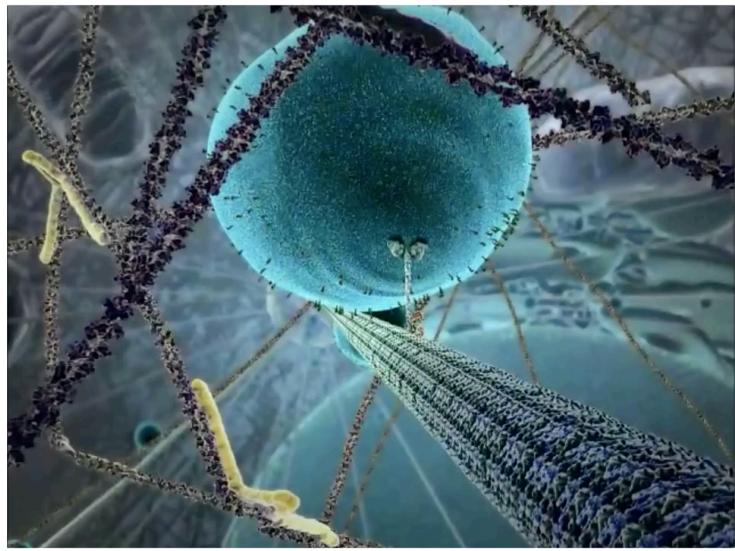
REF. NO.	IBM NO.	DESCRIPTION
1	156011	Track Frame 21", 22", 23", 24", Team Red
2	157040	Fork for 21" Frame
2	157039	Fork for 22" Frame
2	157038	Fork for 23" Frame
2	157037	Fork for 24" Frame
3	191202	Handlebar TTT Competition Track Alloy 15/16" Headset Stem, TTT, Specify extension
4	145278	Examiner Bolt
5	191272	Clamp Bolt
6	145841	Headset Complete 1 x 24 BSC
8	145842	Ball Bearings
9	190420	175 Raleigh Pistard Seta Tubular Prestavalue 27"
10	190233	Rim, 27" AVA Competition (36H) Alloy Prestavalue
11	145973	Hub, Large Flange Campagnolo Pista Track Alloy (pairs)
12	145974	Sprocket, 11 5/8"
13	145937	Sleeve
14	145636	Ball Bearings
15	145170	Bottom Bracket Axle
16	145836	Cone for Sleeve
17	146473	L.H. Adjustable Cup
18	145833	Lockring
19	145834	Straps or Toe Clips
20	145834	Fixed Belt
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	167002	Seat Post Bolt and Nut
30	145933	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

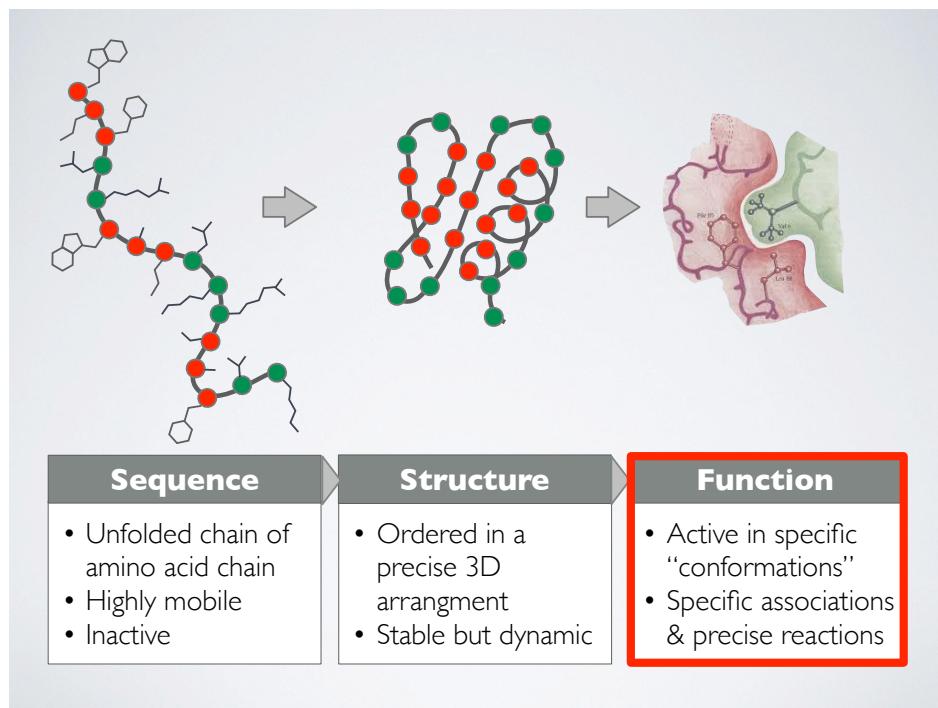
... but not the end



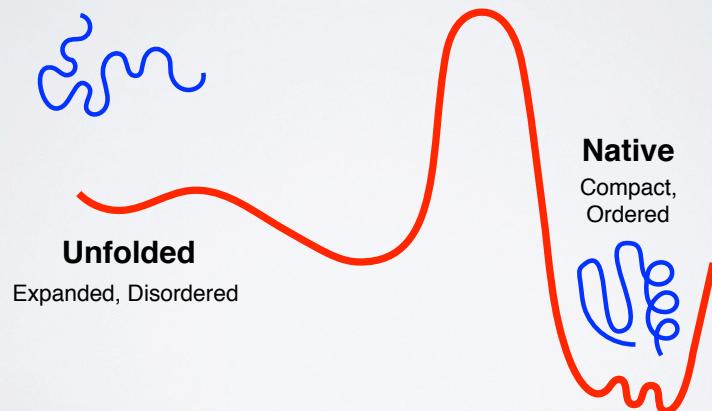
- We want the full spatiotemporal picture, and an ability to control it
- Broad applications, including drug design, medical diagnostics, chemical manufacturing, and energy



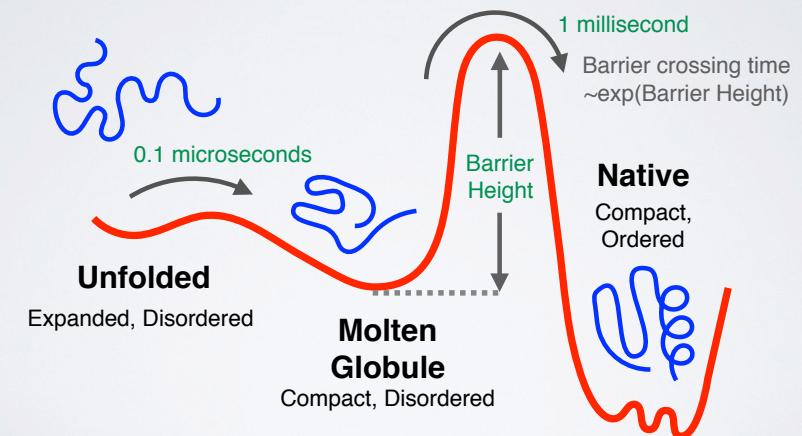
Extracted from The Inner Life of a Cell by Cellular Visions and Harvard
[YouTube link: <https://www.youtube.com/watch?v=y-uuk4Pr2i8>]



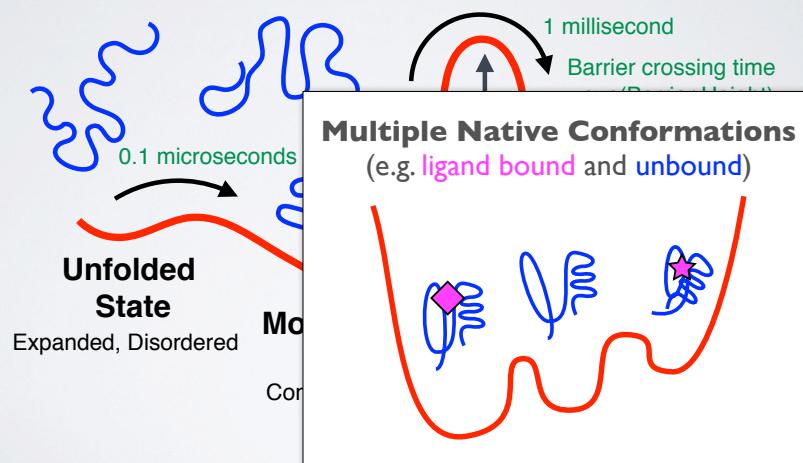
KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: ENERGY LANDSCAPE



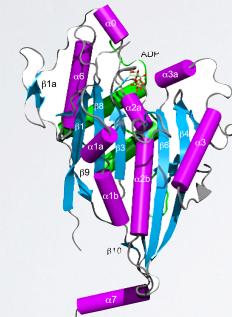
Today's Menu

- 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

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



Protein
(PDB)

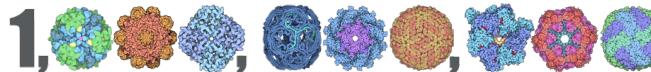


DNA
(NDB)

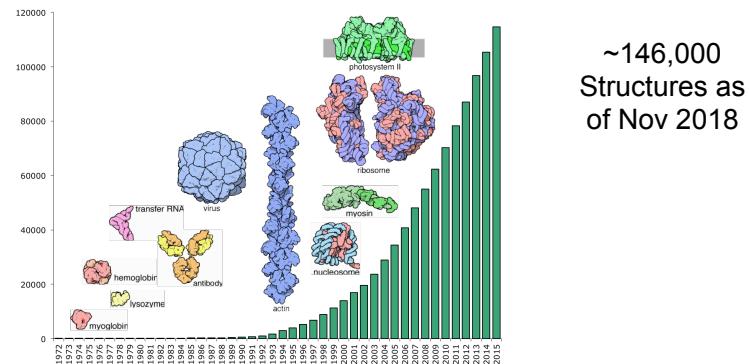


Small Molecules
(CCDB)

PDB – A Billion Atom Archive



> 1 billion atoms in the asymmetric units



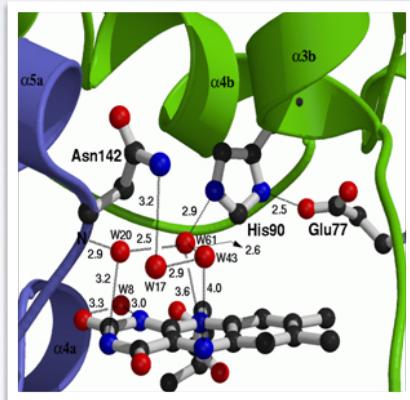
SDSC SAN DIEGO SUPERCOMPUTER CENTER

Slide Credit: Peter Rose

UC San Diego

Motivation 1: Detailed understanding of molecular interactions

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

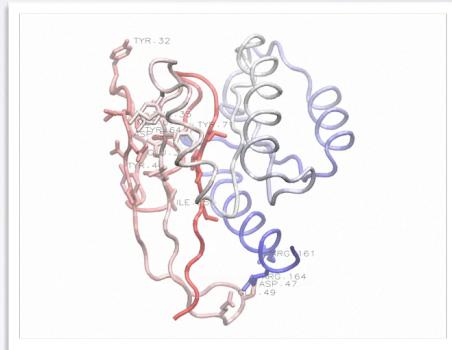


Motivation 2: Lots of structural data is becoming available

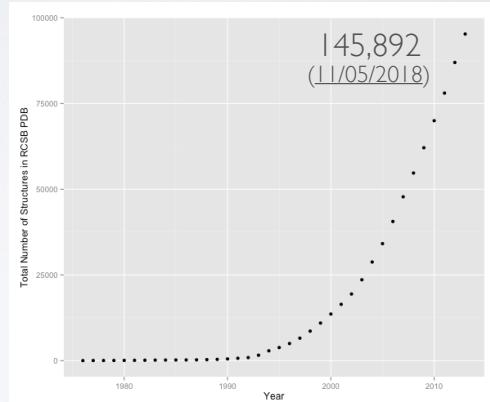
Structural Genomics has contributed to driving down the cost and time required for structural determination

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)



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

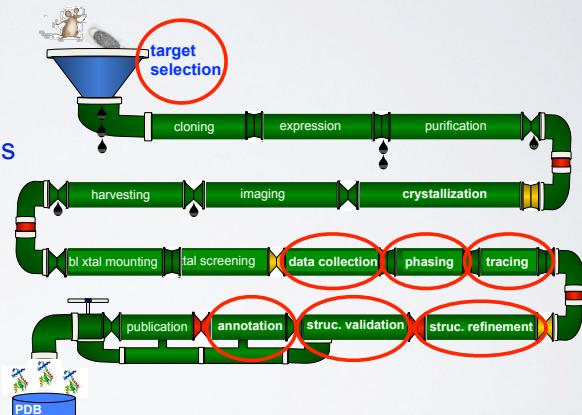
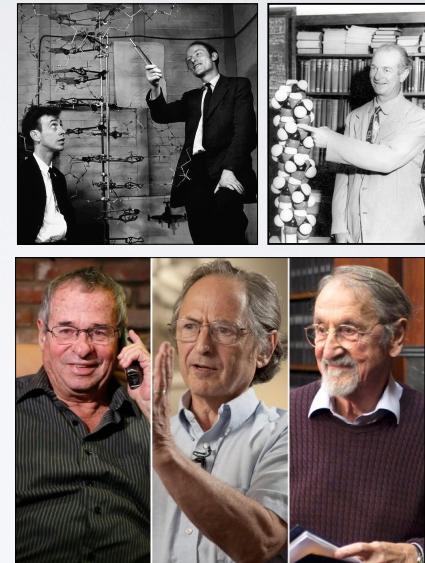


Image Credit: "Structure determination assembly line" Adam Godzik

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

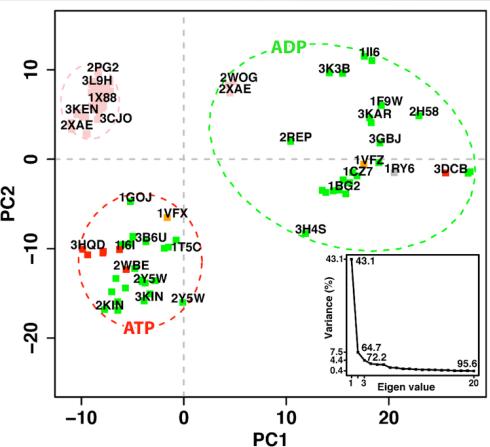
Goals:

- Visualization
- Analysis
- Comparison
- Prediction
- Design



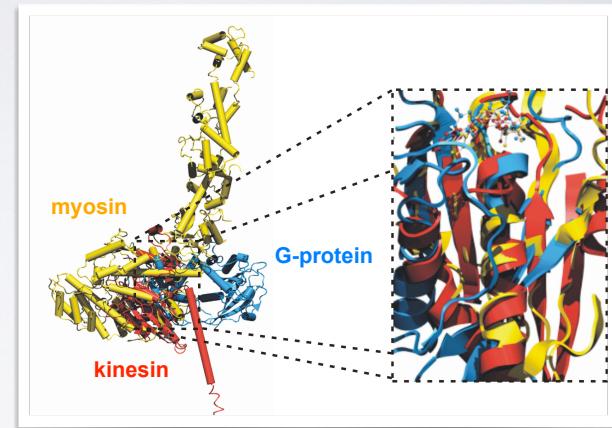
Scarabelli and Grant. PLoS. Comp. Biol. (2013)

- Goals:
- Visualization
 - Analysis
 - Comparison
 - Prediction
 - Design



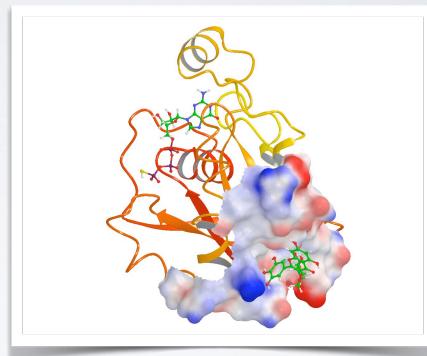
Scarabelli and Grant. PLoS. Comp. Biol. (2013)

- 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

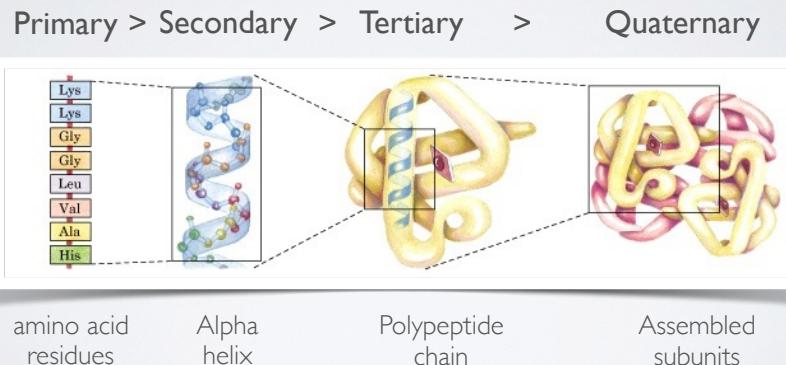


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

RECAP: AMINO ACID NOMENCLATURE

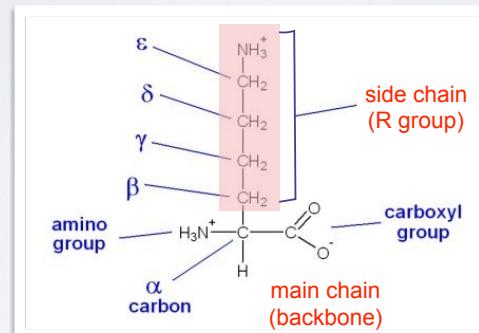


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

AMINO ACIDS CAN BE GROUPED BY THE PHYSIOCHEMICAL PROPERTIES

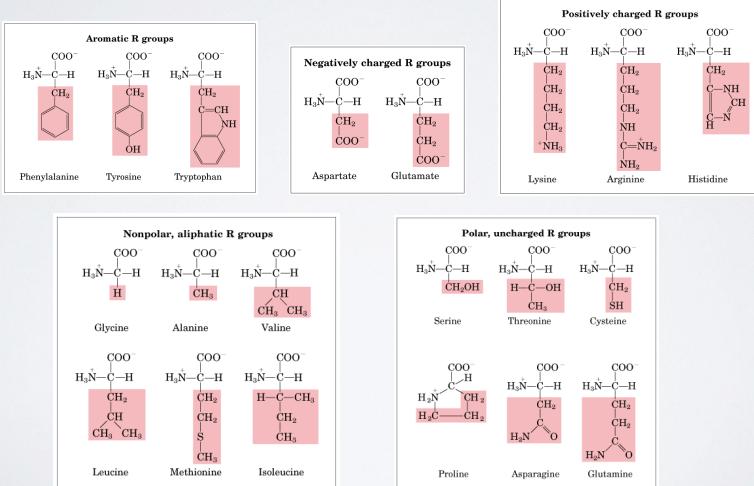


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

AMINO ACIDS POLYMERIZE THROUGH PEPTIDE BOND FORMATION

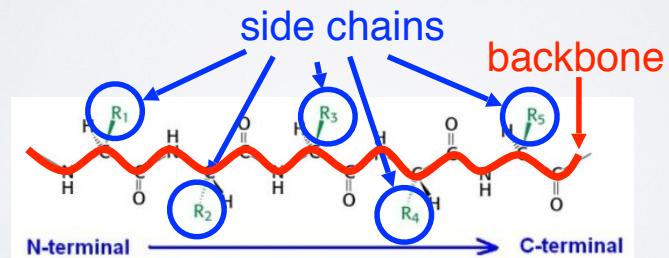
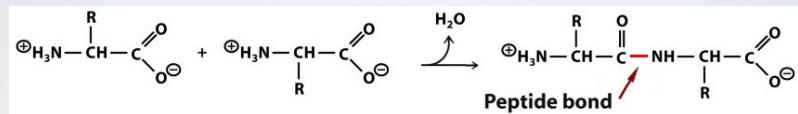


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

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

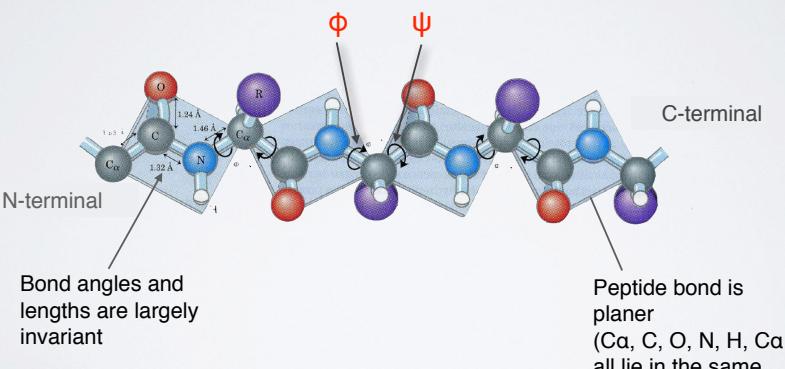
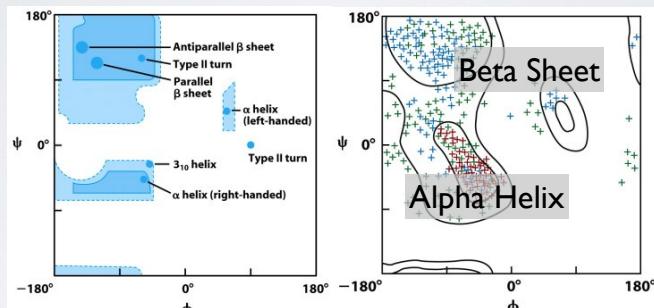


Image from: <http://www.ncbi.nlm.nih.gov/books/NBK21581/>

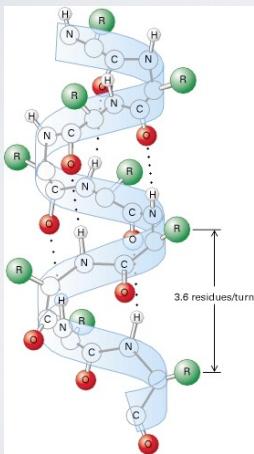
PHI vs PSI PLOTS ARE KNOWN AS RAMACHANDRAN DIAGRAMS



- 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

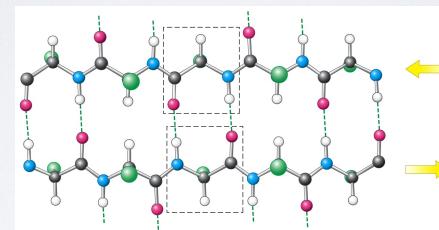


α -helix

- 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
- β_{10} -helix and π -helix forms are less common

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

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

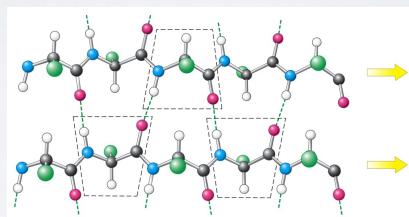


In antiparallel β -sheets

- 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**

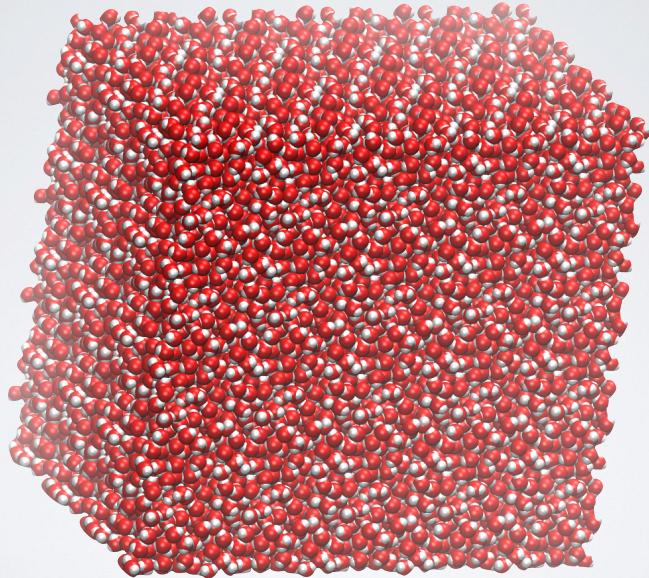


In parallel β -sheets

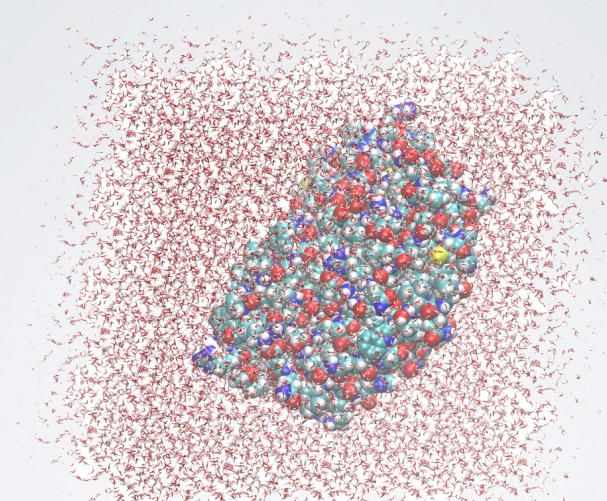
- Adjacent β -strands run in same direction
- Hydrogen bonds (dashed lines) between NH and CO stabilize the structure
- The side chains (in green) are above and below the sheet

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

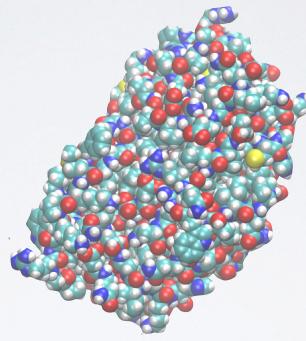
What Does a Protein Look like?



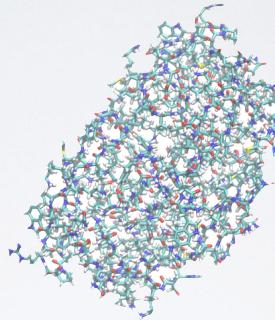
- Proteins are stable (and hidden) in water



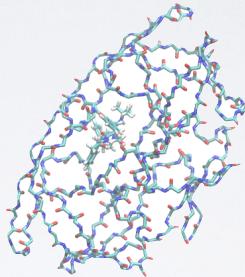
- Proteins closely interact with water



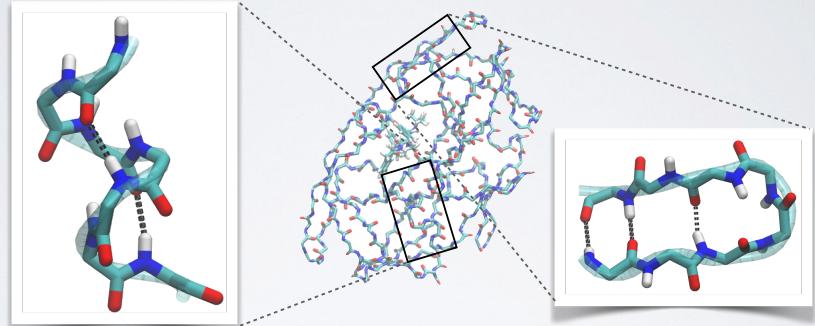
- Proteins are close packed solid but flexible objects (globular)



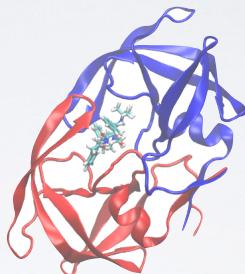
- Due to their large size and complexity it is often hard to see what's important in the structure



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

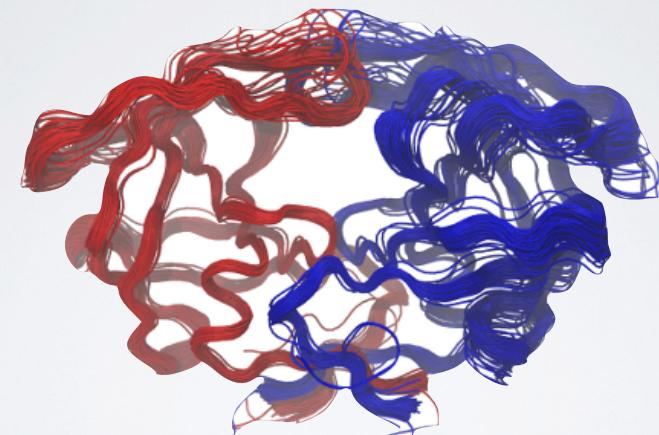


- Backbone or main-chain representation can help trace chain topology & reveal secondary structure



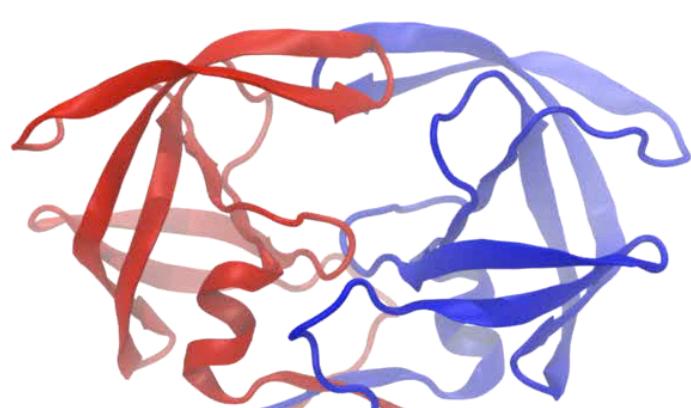
- Simplified secondary structure representations are commonly used to communicate structural details
- Now we can clearly see 2^o, 3^o and 4^o structure
- Coiled chain of connected secondary structures

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY



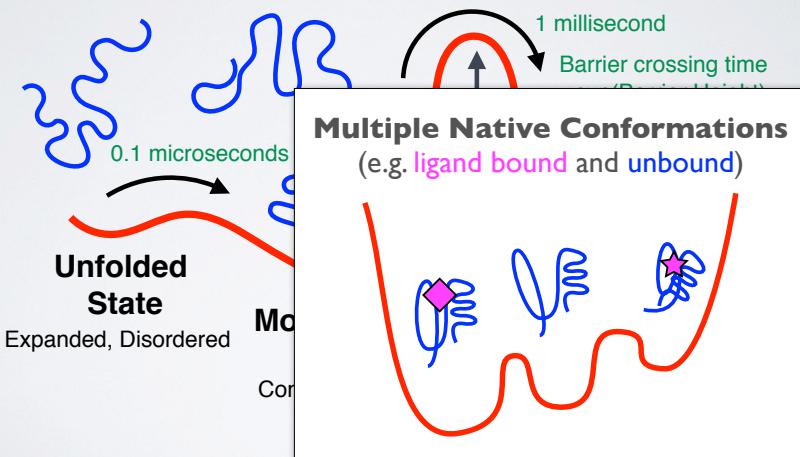
Superposition of all 482 structures in RCSB PDB (23/09/2015)

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY



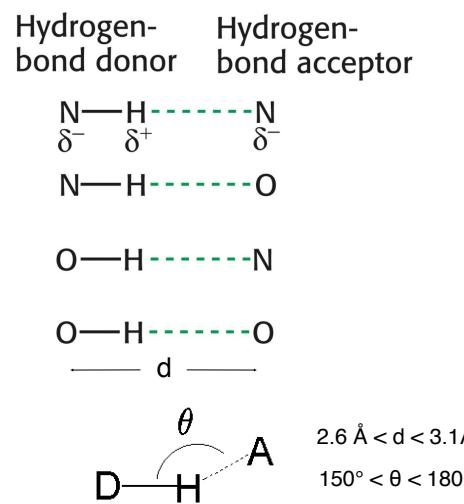
Principal component analysis (PCA) of experimental structures

KEY CONCEPT: ENERGY LANDSCAPE



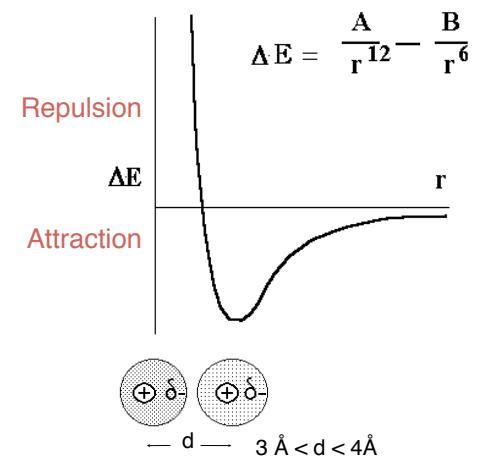
Key forces affecting structure:

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



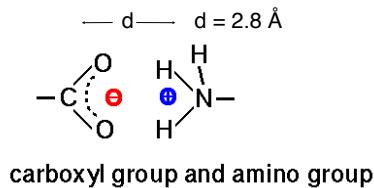
Key forces affecting structure:

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

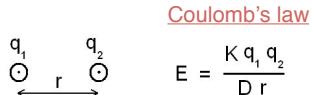


Key forces affecting structure:

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



(some time called IONIC BONDS or SALT BRIDGEs)



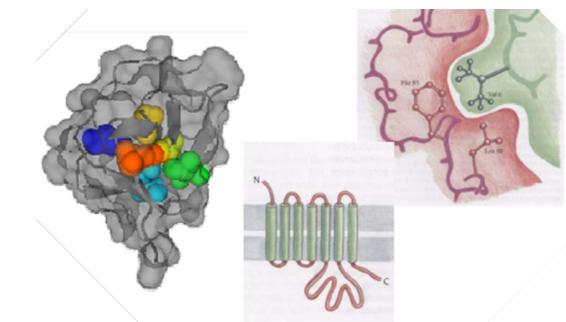
Coulomb's law

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

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!

https://bioboot.github.io/bimm143_S19/lectures/#11

Focus on **section 1** only please!

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

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

\\ Element position within amino acid

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

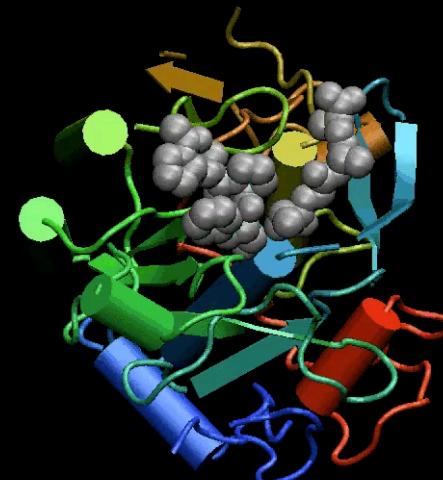
SIDE-NOTE: PDB FILE FORMAT

Element	Amino Acid			Chain Sequence			Element	Side chain (R group)			(etc.)
	1	2	3	4	5	6	7	8	9	10	
ATOM	N	CA	C	O	CB	CG	OD1	OD2	ASP L	ASP L	...
ATOM		ASP L	ASP L	ASP L	ASP L	ASP L	ASP L	ASP L			...
ATOM											...
ATOM											...
ATOM											...
ATOM											...
ATOM											...

\\ Element position within amino acid

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

Download VMD



https://bioboot.github.io/bimm143_W19/lectures/#11

Focus on **section 2** of "Lab Sheet" (using VMD)

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!

https://bioboot.github.io/bimm143_S19/lectures/#11

Focus on **section 3** to **5**

Side Note: Section 4.1

- Download MUSCLE for your OS from:
<https://www.drive5.com/muscle/downloads.htm>
- On **MAC** use your TERMINAL to enter the commands:

```
> tar -xvf ~/Downloads/muscle3.8.31_i86darwin32.tar  
> sudo mv muscle3.8.31_i86darwin32 /usr/local/bin/muscle
```
- On **Windows** use file explorer to:
 - Move the downloaded **muscle3.8.31_i86win32.exe** from your *Downloads* folder to your *Project* folder.
 - Then right click to rename to **muscle.exe**

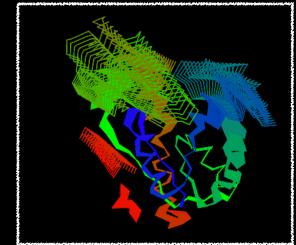
```
> ./muscle.exe -version
```

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:

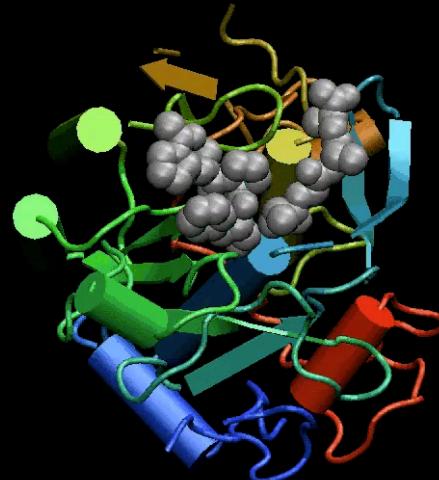
```
> library("bio3d.view")  
> pdb <- read.pdb("5p21")  
> view(pdb)  
> view(pdb, "overview", col="sse")
```



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**

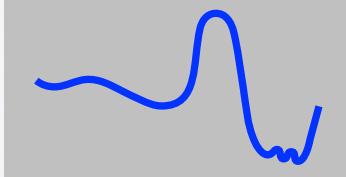
Two main approaches:

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

Two main approaches:

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

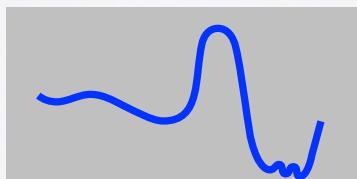
Energy



Structure/Conformation

This will be the focus of the next class!

Energy



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](#)]