

Biomolecular structure (including protein structure)

CS/CME/BioE/Biophys/BMI 279

Sept. 17 and 22, 2020

Ron Dror

- If you had trouble enrolling in the course, please try again, or let course staff know
- Please interrupt if you have questions, and especially if you're confused!
- Assignment 1 will be posted by tomorrow
- Tutorial on terminal and Python
 - Tomorrow 3:30-4:20
 - You can also view the recording afterwards
 - Recommended if you haven't used Python or terminal (Mac, Linux) before
 - You can also get help during TA office hours

Outline

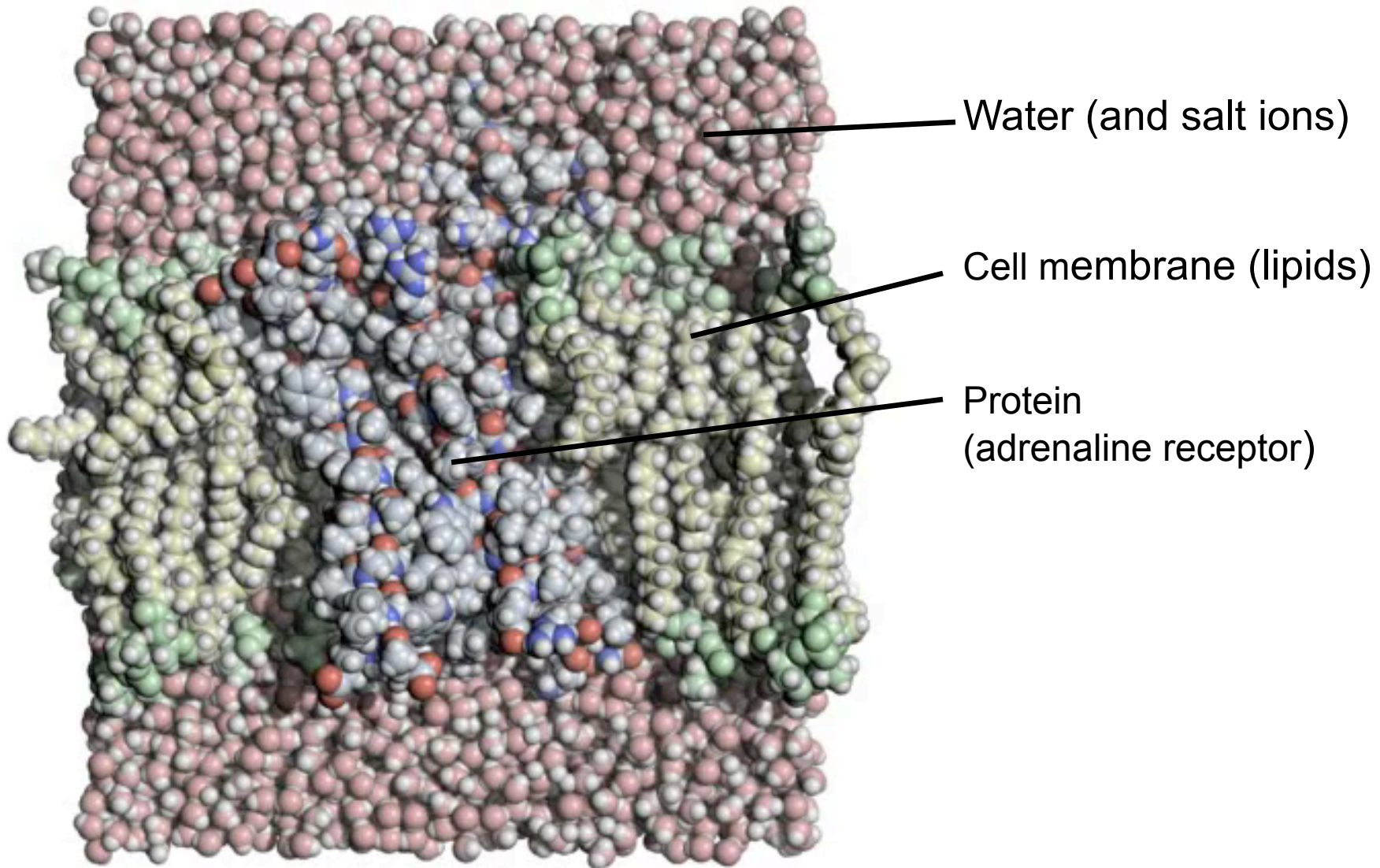
Note: I'll discuss proteins first, as an example.

These concepts apply to other biomolecules as well.

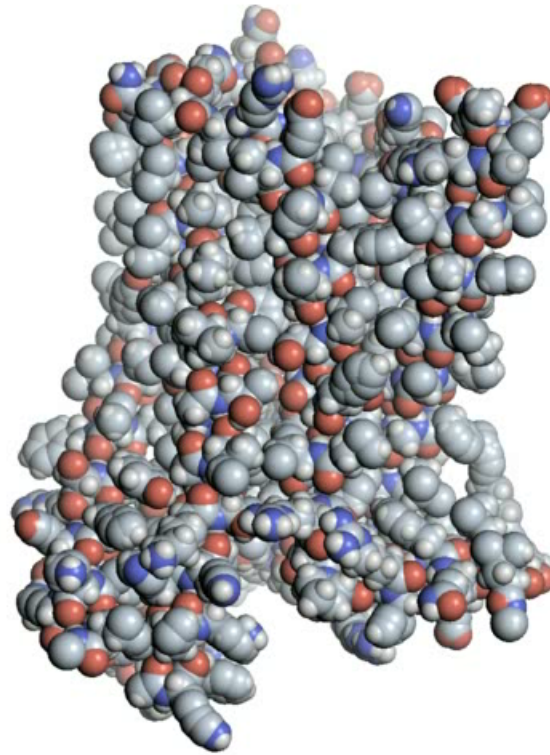
- Visualizing biomolecules (e.g, proteins)
- The Protein Data Bank (PDB)
- Chemical (2D) structure of proteins
- What determines the 3D structure of a protein?
Physics underlying biomolecular structure
 - Basic interactions
 - Complex interactions
- Protein structure: a more detailed view
 - Properties of amino acids
 - Secondary structure
 - Tertiary structure, quaternary structure, and domains
- Structures of other biomolecules

Visualizing biomolecules (e.g., proteins)

Protein surrounded by other molecules (mostly water)

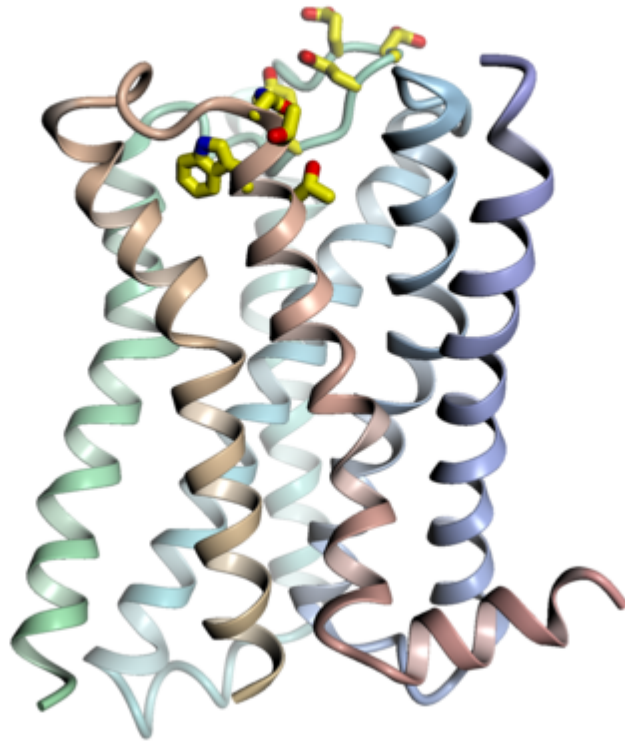


Protein only



Adrenaline receptor

Protein only — simplified representation



Adrenaline receptor

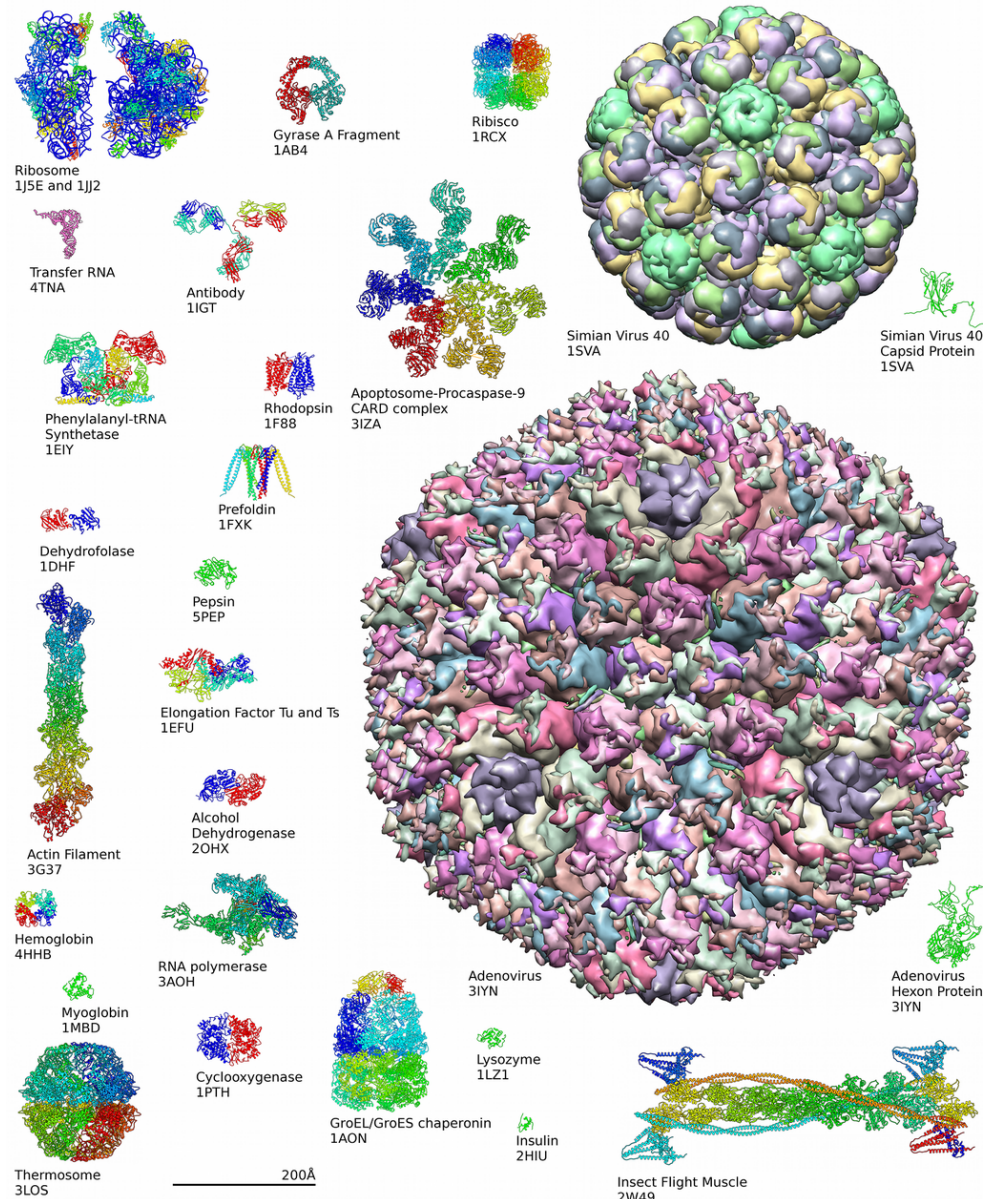
Key take-aways from these visualizations

- Protein is a long chain of amino acids.
- Protein and surrounding atoms fill space (close-packed).
- There are a lot of atoms. Simplified visual representations help you figure out what's going on.

The Protein Data Bank (PDB)

The Protein Data Bank (PDB)

- Examples of structures from PDB.



https://upload.wikimedia.org/wikipedia/commons/thumb/2/24/Protein_structure_examples.png/1024px-Protein_structure_examples.png

(Axel Griewel)

**You're not responsible for these;
they're just examples.**

The Protein Data Bank (PDB)

← → ↻ 🔒 rcsb.org ☆ 🌐 ⚙️ 📄 🌐

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

RCSB PDB PROTEIN DATA BANK **168889** Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Enter search term(s) 🔍

Advanced Search | Browse Annotations

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDatabank EMDataResource Official Data Resource for IUCr NDB NUCLEIC ACID DATABASE Worldwide Protein Data Bank Foundation

📘 🐦 📺 🔄

- Welcome
- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology

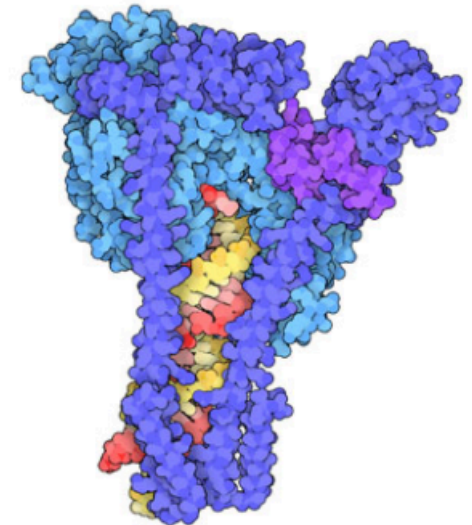
This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

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.



September Molecule of the Month



SARS-CoV-2 RNA-dependent RNA Polymerase

Structure Summary **3D View** Annotations Experiment Sequence

6YYT

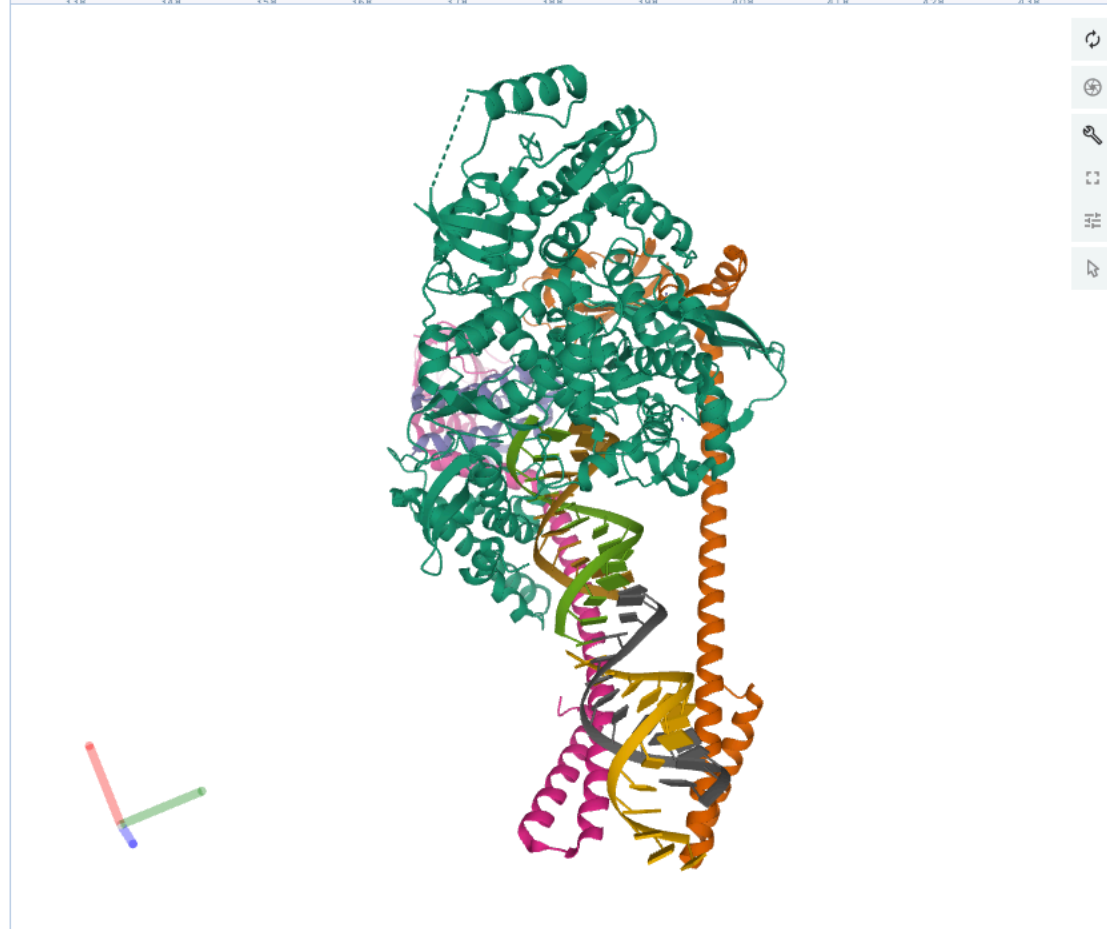
Structure of replicating SARS-CoV-2 polymerase

Display Files Download Files

Help

Sequence of 6YYT | Struct... 1: nsp12 A

```
SNASADAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAFLKTNCCRFQEKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFKFRIDGD
118 128 138 148 158 168 178 188 198 208 218
MVPHISRQRLTKYTMADLVYALRHFDEGNCDLKEILVTYNCCDDYFNKKDWYDFVENPDILRVYANLGERVVRQALLKTVQFCDAMRNAGIVGLTLDNQDLNGNWDYDFGD
228 238 248 258 268 278 288 298 308 318 328
FIQTTPGSGVPPVDSYYSLMLPILTLRALTAESHVDTLTKPYIKWDLKDYFTEERLKLFDRYFKYWDQTYHPNCVNCDDRCILHCANFNVLFTVFPPTSFGLVVRKI
338 348 358 368 378 388 398 408 418 428 438
```



Structure

6YYT | Structure of replicating SAR... [📄](#)

Type	Assembly
Asm Id	1: Author And Softwar...

Nothing Focused [🔍](#)

Measurements

Components

6YYT

Preset	+ Add	🔍	🗑️	⋮
Polymer	Cartoon	👁️	🗑️	⋮
Ion	Ball & Stick	👁️	🗑️	⋮

Density

Assembly Symmetry

The Protein Data Bank (PDB)

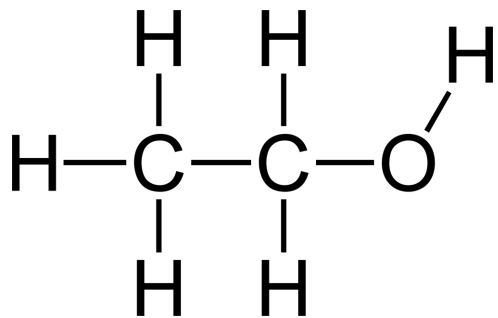
- <http://www.rcsb.org/pdb/home/home.do>
- A collection of (almost) all published experimental structures of biomacromolecules (e.g., proteins, DNA, RNA)
- Each identified by 4-character code (e.g., 1rcx)
- Currently ~169,000 structures. 80% of those are determined by x-ray crystallography.
- Browse it and look at some structures. Options:
 - 3D view in applet on PDB web pages
 - PyMol: fetch 1rcx

Chemical (two-dimensional) structure of proteins

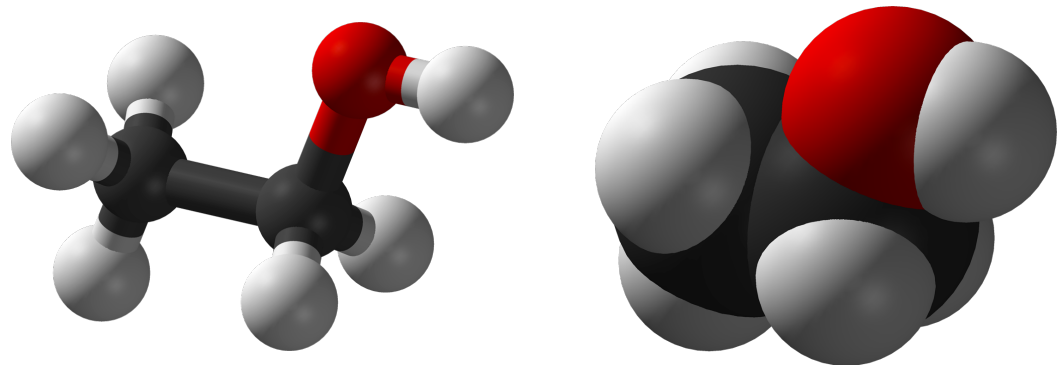
Chemical (two-dimensional) structure vs. three-dimensional structure

- Chemical (two-dimensional) structure shows *covalent bonds* between atoms. Essentially a graph.
- Three-dimensional structure shows relative positions of atoms.

2D structure



3D structure



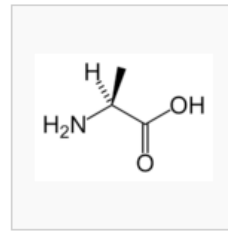
images from <https://en.wikipedia.org/wiki/Ethanol>

Proteins are built from amino acids

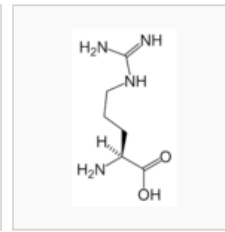
- 20 “standard” amino acids
- Each has three-letter and one-letter abbreviations (e.g., Threonine = Thr = T; Tryptophan = Trp = W)

The “side chain” is different in each amino acid.

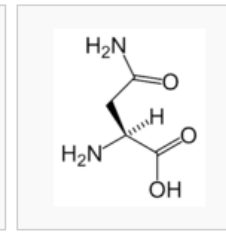
All amino acids have this part in common.



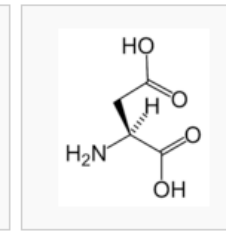
L-Alanine
(Ala / A)



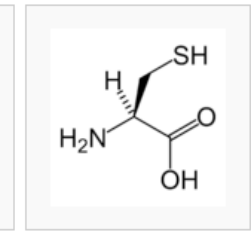
L-Arginine
(Arg / R)



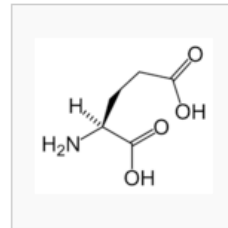
L-Asparagine
(Asn / N)



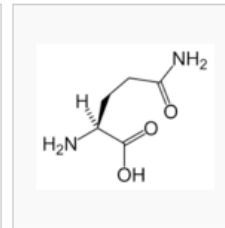
L-Aspartic acid
(Asp / D)



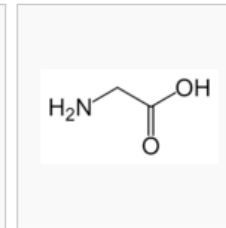
L-Cysteine
(Cys / C)



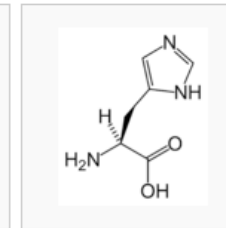
L-Glutamic acid
(Glu / E)



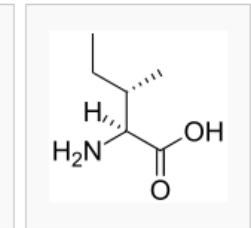
L-Glutamine
(Gln / Q)



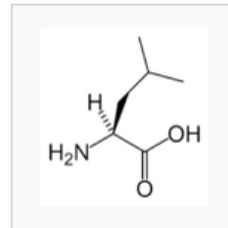
Glycine
(Gly / G)



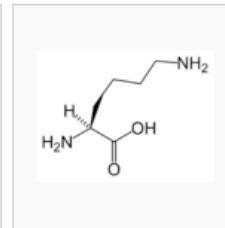
L-Histidine
(His / H)



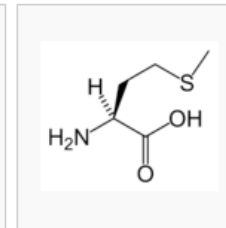
L-Isoleucine
(Ile / I)



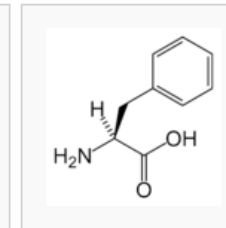
L-Leucine
(Leu / L)



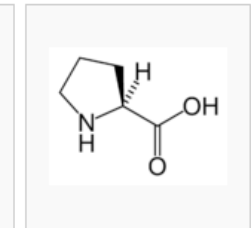
L-Lysine
(Lys / K)



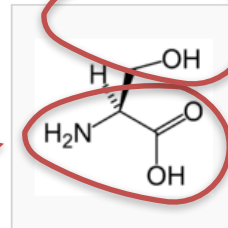
L-Methionine
(Met / M)



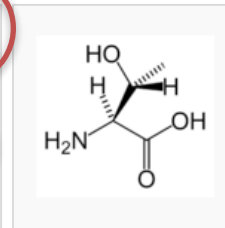
L-Phenylalanine
(Phe / F)



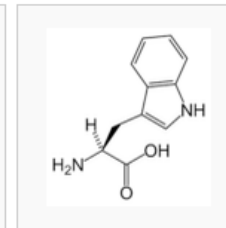
L-Proline
(Pro / P)



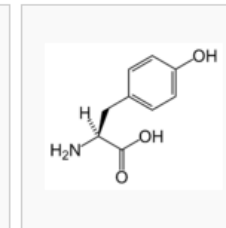
L-Serine
(Ser / S)



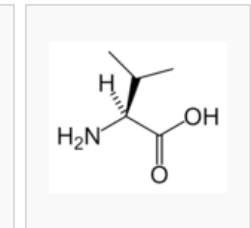
L-Threonine
(Thr / T)



L-Tryptophan
(Trp / W)



L-Tyrosine
(Tyr / Y)

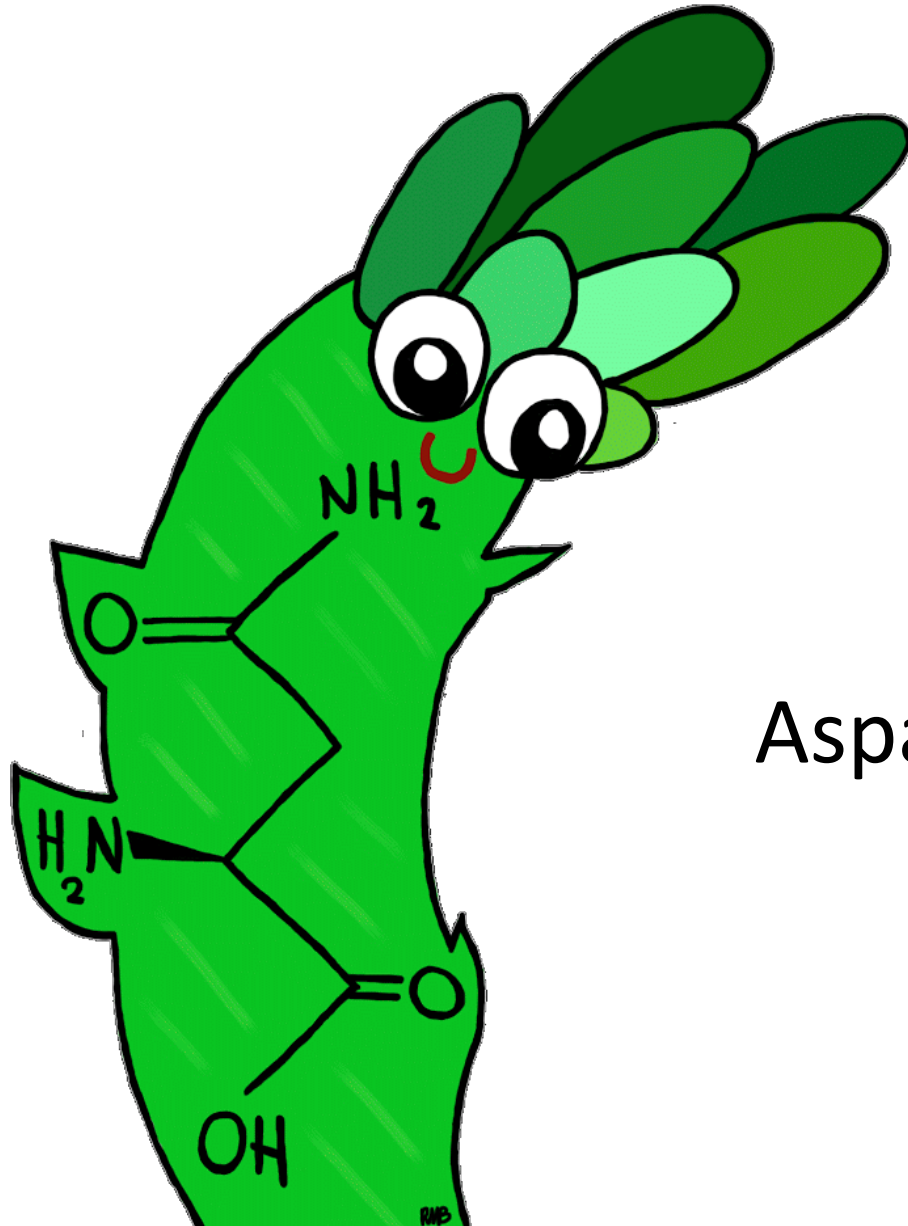


L-Valine
(Val / V)



THAT'S RIGHT, FOUREYES!
YOU'RE **NOTHING**
WITHOUT ME! WHILE
I'M AN ESSENTIAL PART
OF ANY PROTEIN, EVEN
YOURS, YOU'RE STILL
A SO-SO PROFESSOR
WITH **NO CHANCE**
OF TENURE! **HAHAHA**

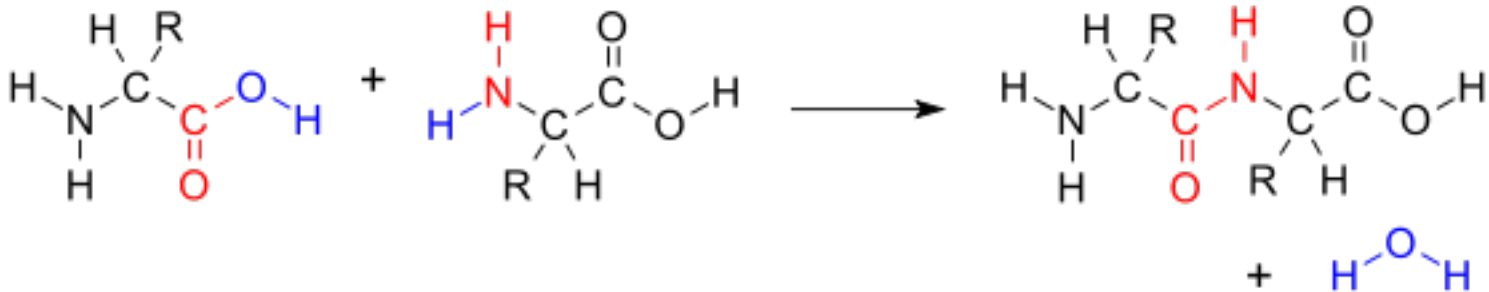
a mean o' acid



Asparagine

Proteins are chains of amino acids

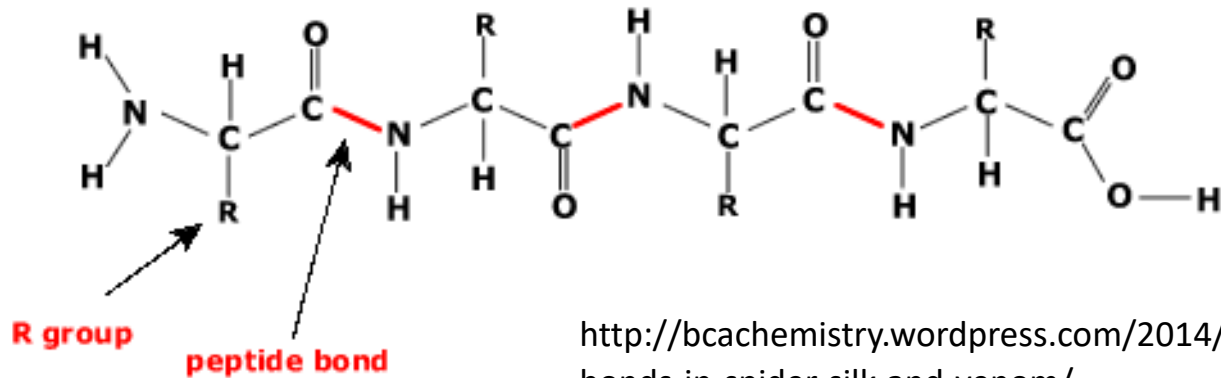
- Amino acids link together through a chemical reaction (“condensation”)



http://en.wikipedia.org/wiki/Condensation_reaction

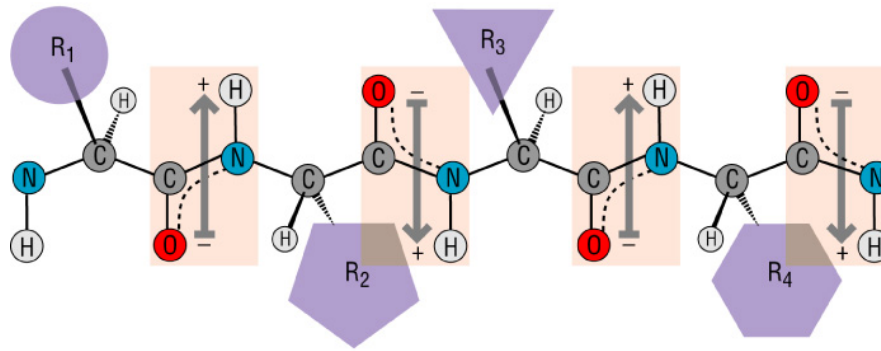
- Elements of the chain are called “amino acid residues” or just “**residues**” (important term!)
- The bonds linking these residues are “peptide bonds.” The chains are also called “polypeptides”

Proteins have uniform backbones with differing side chains



<http://bcachemistry.wordpress.com/2014/05/28/chemical-bonds-in-spider-silk-and-venom/>

From **Protein Structure and Function** by Gregory A Petsko and Dagmar Ringe



© 1999–2004 New Science Press

What determines the 3D structure of a protein?
Physics underlying biomolecular structure

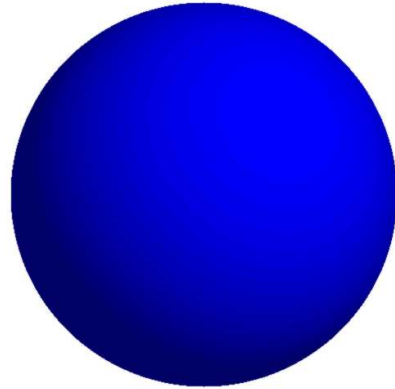
Why do proteins have well-defined structure?

- The sequence of amino acids in a protein (usually) suffices to determine its structure.
- A chain of amino acids (usually) “folds” spontaneously into the protein’s preferred structure, known as the “native structure”
- Why?
 - Intuitively: some amino acids prefer to be inside, some prefer to be outside, some pairs prefer to be near one another, etc.
 - To understand this better, examine forces acting between atoms

What determines the 3D structure of a protein?
Physics underlying biomolecular structure

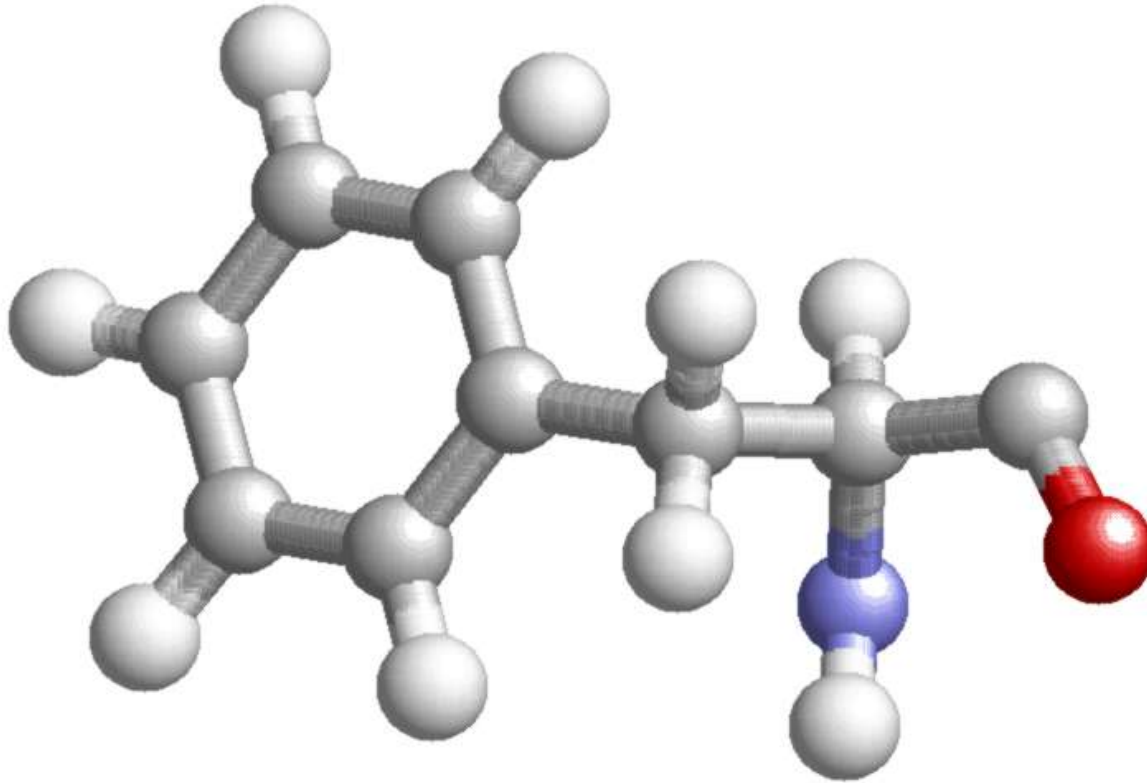
Basic interactions

Geometry of an atom



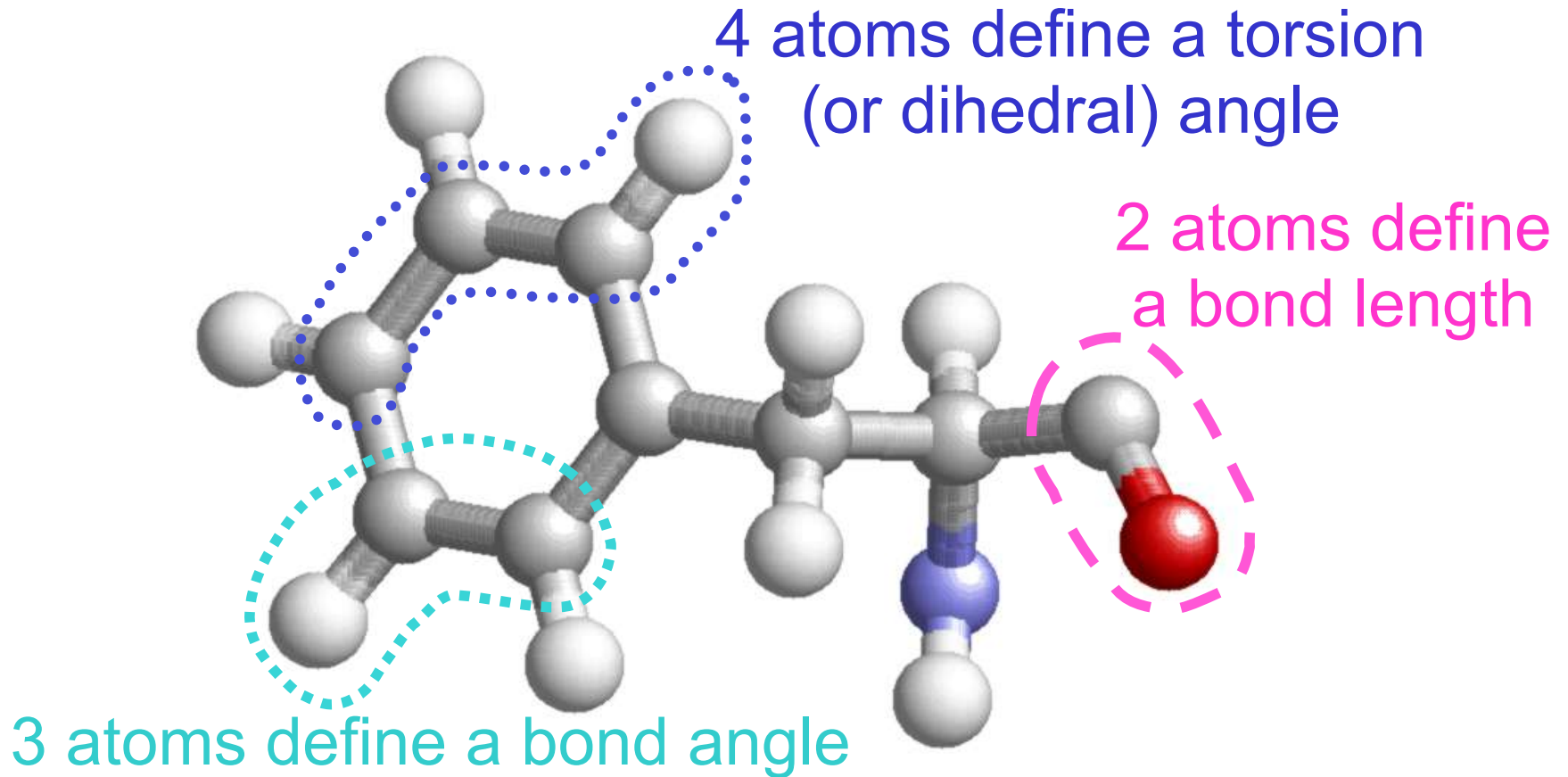
- To a first approximation (which suffices for the purposes of this course), we can think of an atom simply as a sphere.
- It occupies a position in space, specified by the (x, y, z) coordinates of its center, at a given point in time

Geometry of a molecule



- A molecule is a set of atoms connected in a graph
- (x, y, z) coordinates of every atom specify the molecule's geometry

Geometry of a molecule



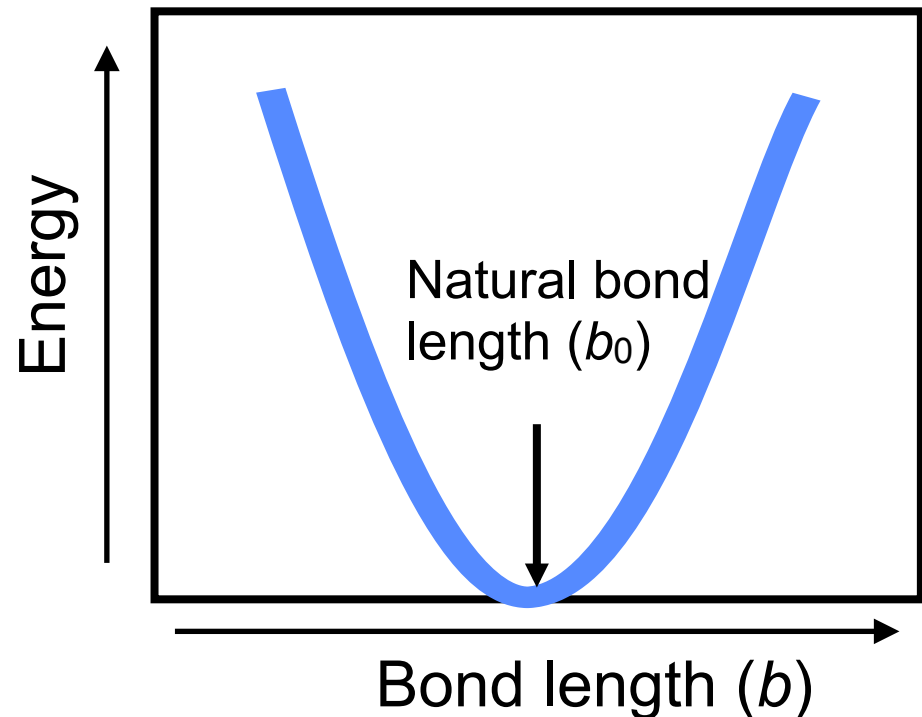
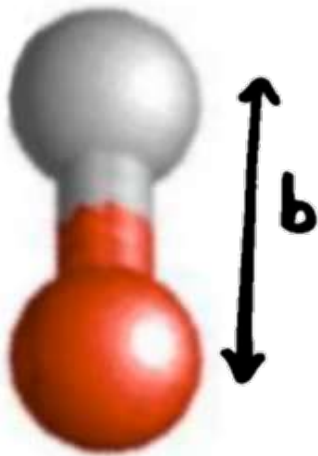
- Alternatively, we can specify the geometry of a molecule using bond lengths, bond angles, and torsion angles

Forces between atoms

- We can approximate the total potential energy of a molecular system as a sum of individual contributions. Terms are additive.
 - Thus force on each atom is also a sum of individual contributions. Remember: force is the derivative of energy.
 - We will ignore quantum effects. Think of atoms as balls and forces as springs.
- Two types of forces:
 - Bonded forces: act between closely connected sets of atoms in the graph of covalent bonds
 - Non-bonded forces: act between all pairs of atoms

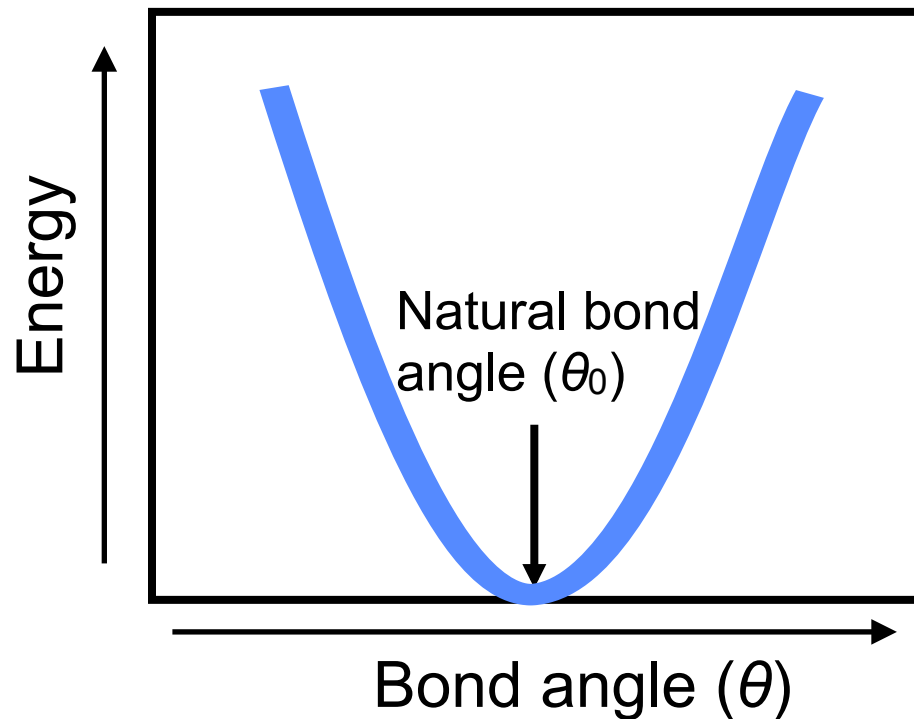
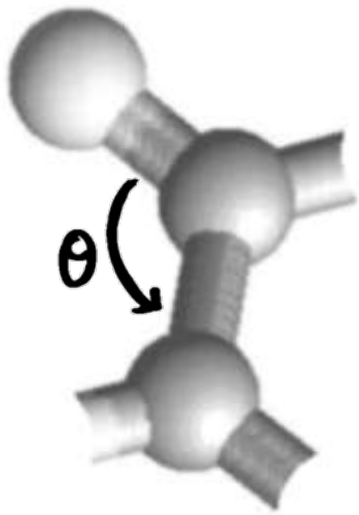
Bond length stretching

- A covalently bonded pair of atoms is effectively connected by a spring with some preferred (natural) length. Stretching or compressing it requires energy.



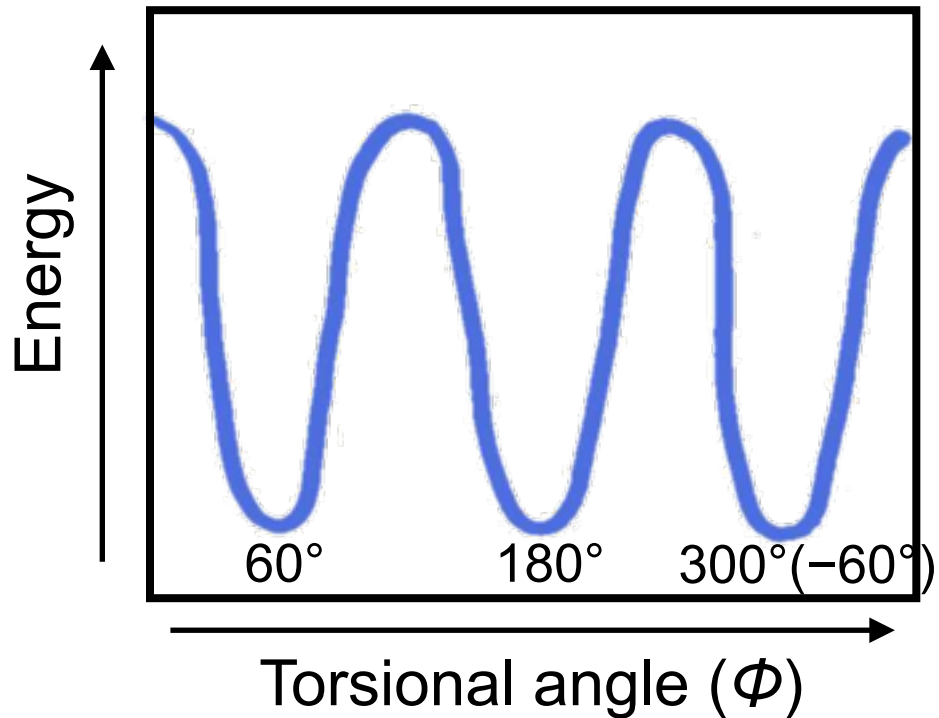
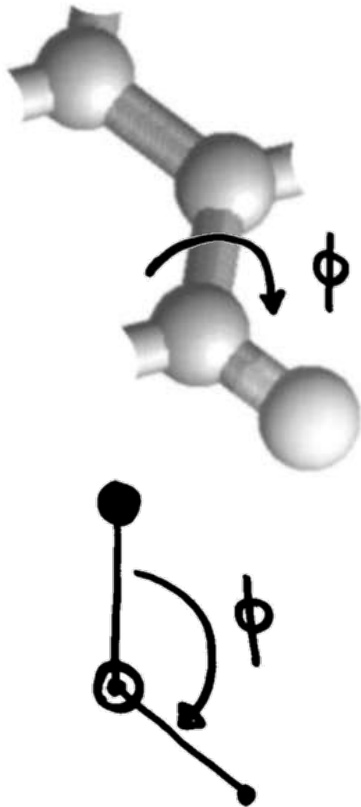
Bond angle bending

- Likewise, each bond angle has some natural value. Increasing or decreasing it requires energy.

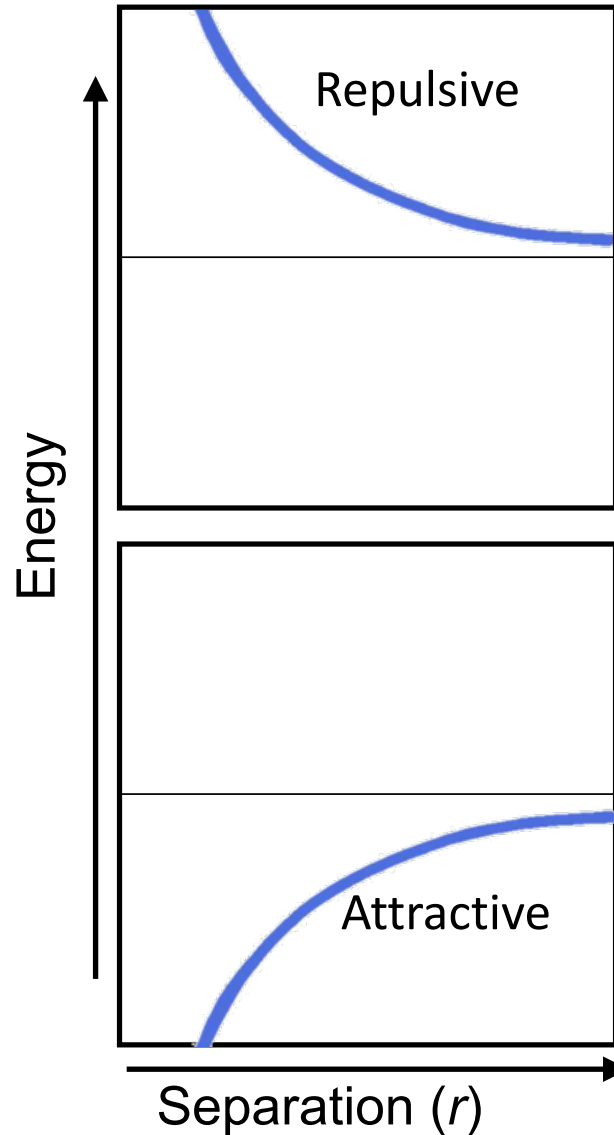
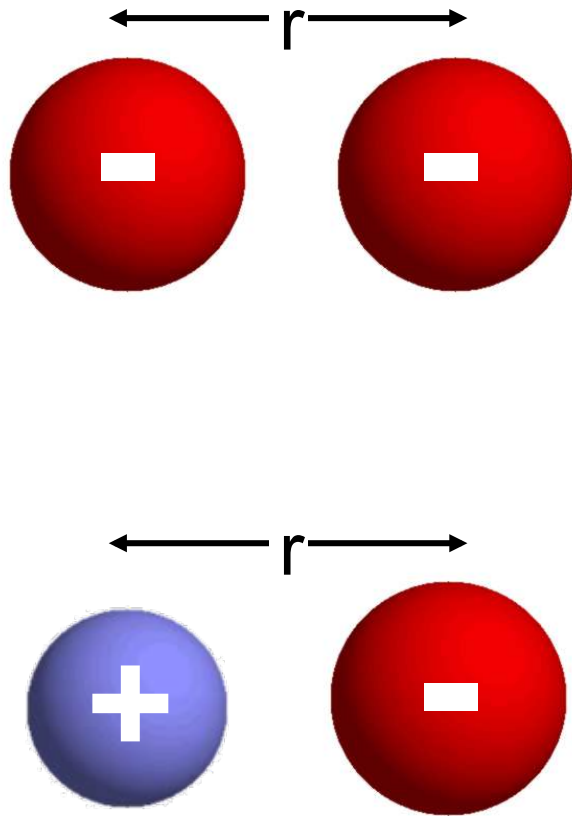


Torsional angle twisting

- Certain values of each torsional angle are preferred over others.

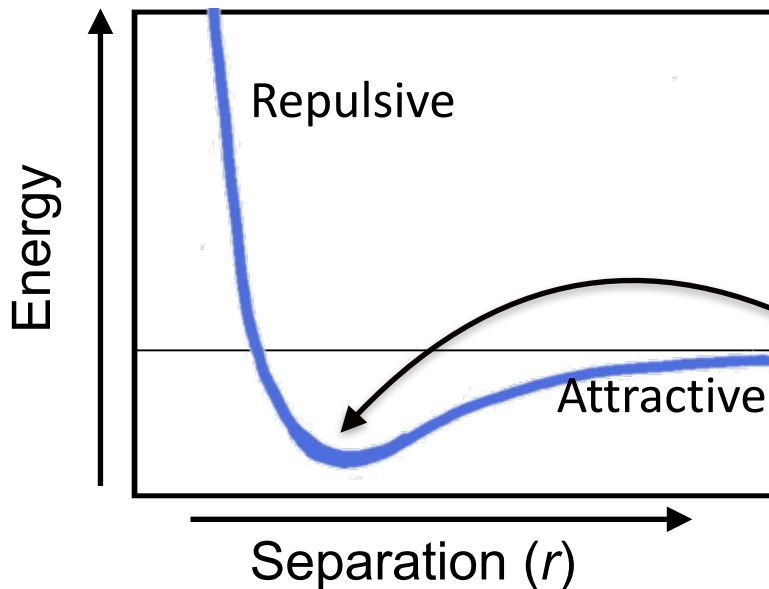
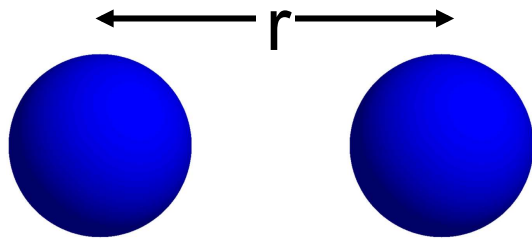


Electrostatic interaction



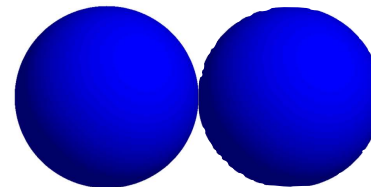
- Like charges repel. Opposite charges attract.
- Electrostatic forces act between all pairs of atoms, including those in different molecules.
- Each atom carries some “partial charge” (may be a fraction of an elementary charge), which depends on which other atoms it’s connected to.

van der Waals interaction



- van der Waals forces act between all pairs of atoms and do not depend on charge.
- When two atoms are too close together, they repel strongly.
- When two atoms are a bit further apart, they attract one another weakly.

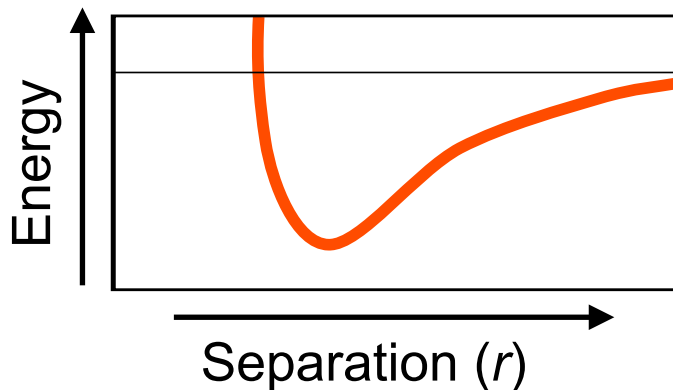
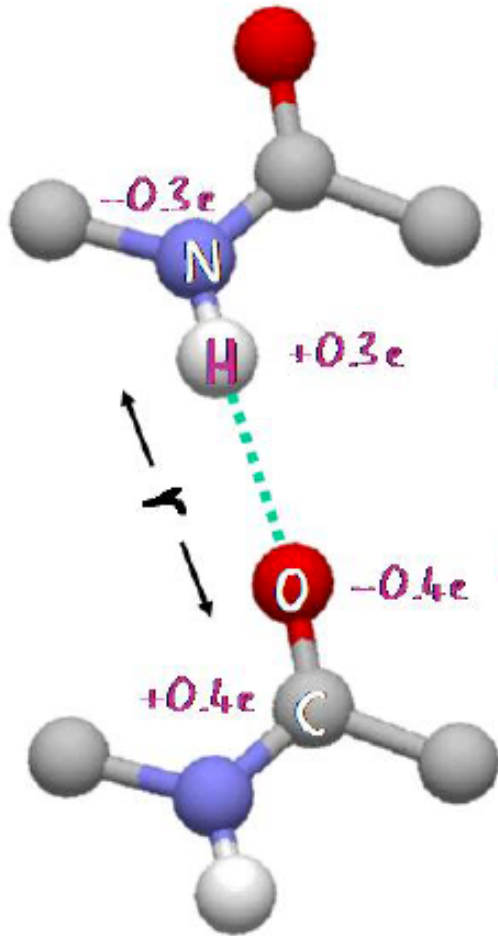
Energy is minimal when atoms are "just touching" one another



What determines the 3D structure of a protein?
Physics underlying biomolecular structure

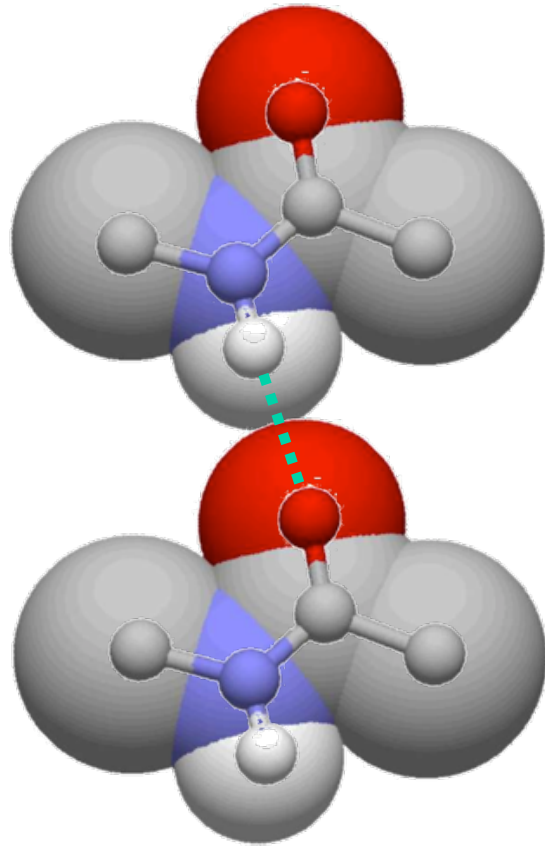
Complex interactions

Hydrogen bonds

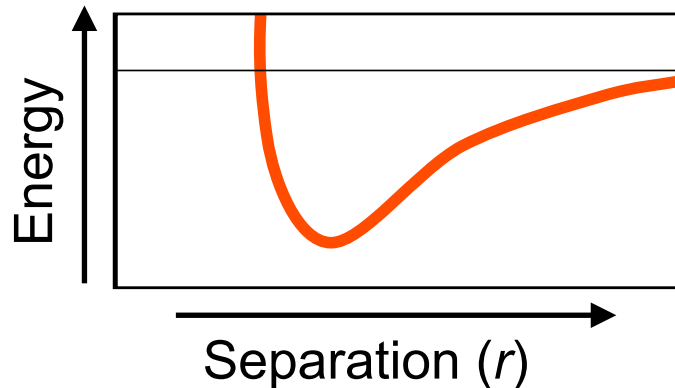


- Favorable interaction between an electronegative atom (e.g., N or O) and a hydrogen bound to another electronegative atom
- Result of multiple electrostatic and van der Waals interactions
- Very sensitive to geometry of the atoms (distance and alignment)
- Strong relative to typical van der Waals or electrostatic forces
- Critical to protein structure

Hydrogen bonds



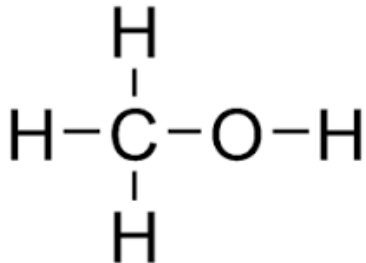
- Favorable interaction between an electronegative atom (e.g., N or O) and a hydrogen bound to another electronegative atom
- Result of multiple electrostatic and van der Waals interactions
- Very sensitive to geometry of the atoms (distance and alignment)
- Strong relative to typical van der Waals or electrostatic forces
- Critical to protein structure



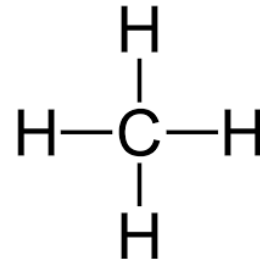
Hydrophilic vs. hydrophobic

- Hydrophilic molecules are polar and thus form hydrogen bonds with water
 - Polar = contains charged atoms. Molecules containing oxygen or nitrogen are usually polar.
- Hydrophobic molecules are apolar and don't form hydrogen bonds with water

Hydrophilic (polar)



Hydrophobic (apolar)



Hydrophobic effect

- Hydrophobic molecules cluster in water
 - “Oil and water don’t mix”

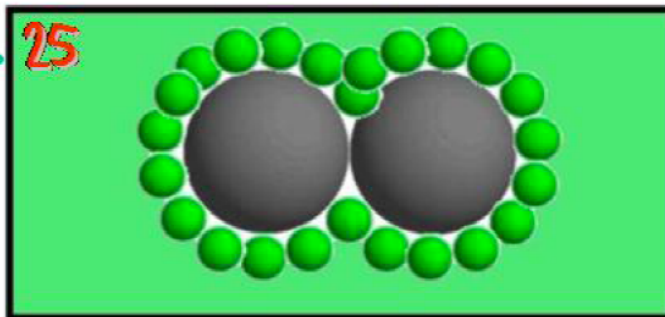
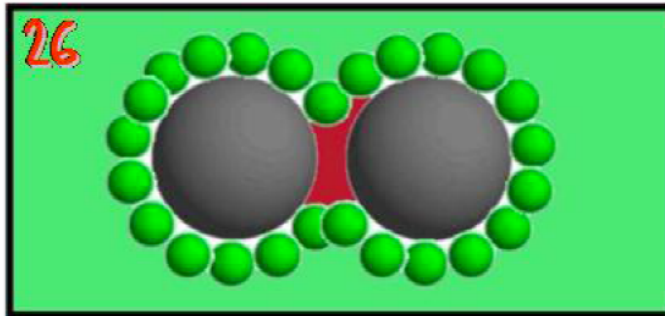
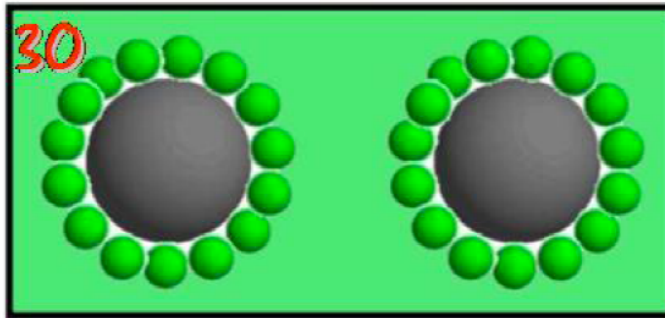


<http://science.taskermilward.org.uk/mod1/KS4Chemistry/AQA/Module2/Mod%20%20img/Oil-in-Water18.jpg>

- This is critical to protein structure

EXPLAINING HYDROPHOBICITY

Number of unhappy water molecules



- Water molecules next to solute cannot move freely.

- They are ordered and have less entropy. They are unhappy.

- The system changes so that fewer water molecules are in the surface layer.

- The hydrophobic solutes aggregate.

©Michael Levitt 04

- Slide from Michael Levitt
- We will discuss entropy next week. If this isn't clear now, don't worry.

Protein structure: a more detailed view

“Levels” of protein structure

- Primary structure: sequence of amino acids
- Secondary structure: local structural elements
- Tertiary structure: overall structure of the polypeptide chain
- Quaternary structure: how multiple polypeptide chains come together

Protein structure: a more detailed view

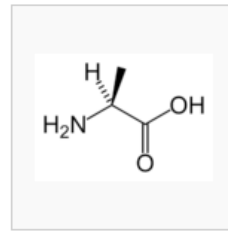
Properties of amino acids

Proteins are built from amino acids

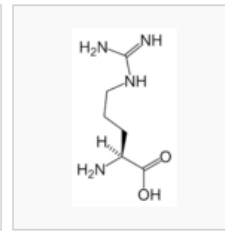
- 20 “standard” amino acids
- Each has three-letter and one-letter abbreviations (e.g., Threonine = Thr = T; Tryptophan = Trp = W)

The “side chain” is different in each amino acid

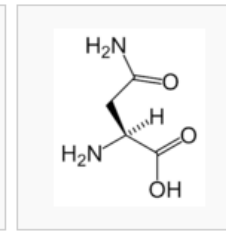
All amino acids have this part in common.



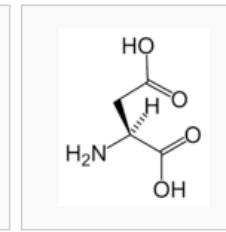
L-Alanine
(Ala / A)



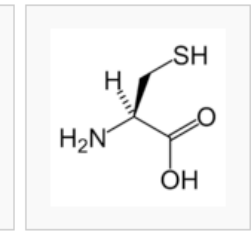
L-Arginine
(Arg / R)



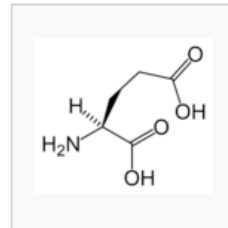
L-Asparagine
(Asn / N)



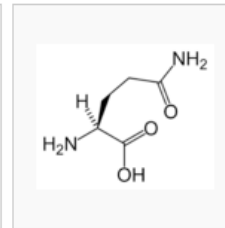
L-Aspartic acid
(Asp / D)



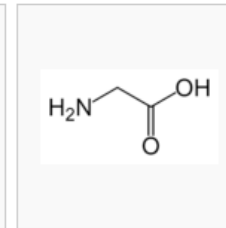
L-Cysteine
(Cys / C)



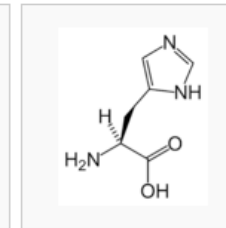
L-Glutamic acid
(Glu / E)



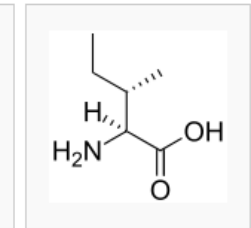
L-Glutamine
(Gln / Q)



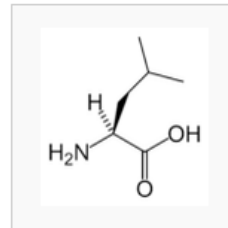
Glycine
(Gly / G)



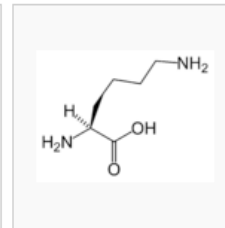
L-Histidine
(His / H)



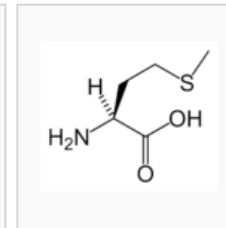
L-Isoleucine
(Ile / I)



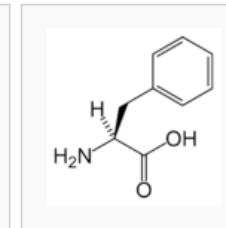
L-Leucine
(Leu / L)



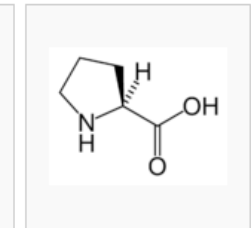
L-Lysine
(Lys / K)



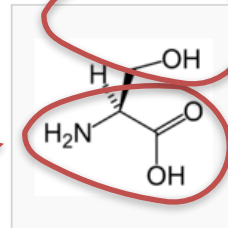
L-Methionine
(Met / M)



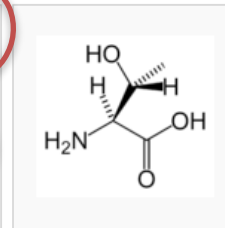
L-Phenylalanine
(Phe / F)



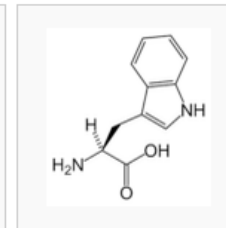
L-Proline
(Pro / P)



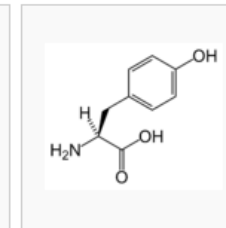
L-Serine
(Ser / S)



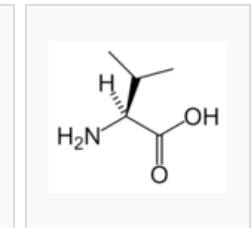
L-Threonine
(Thr / T)



L-Tryptophan
(Trp / W)



L-Tyrosine
(Tyr / Y)

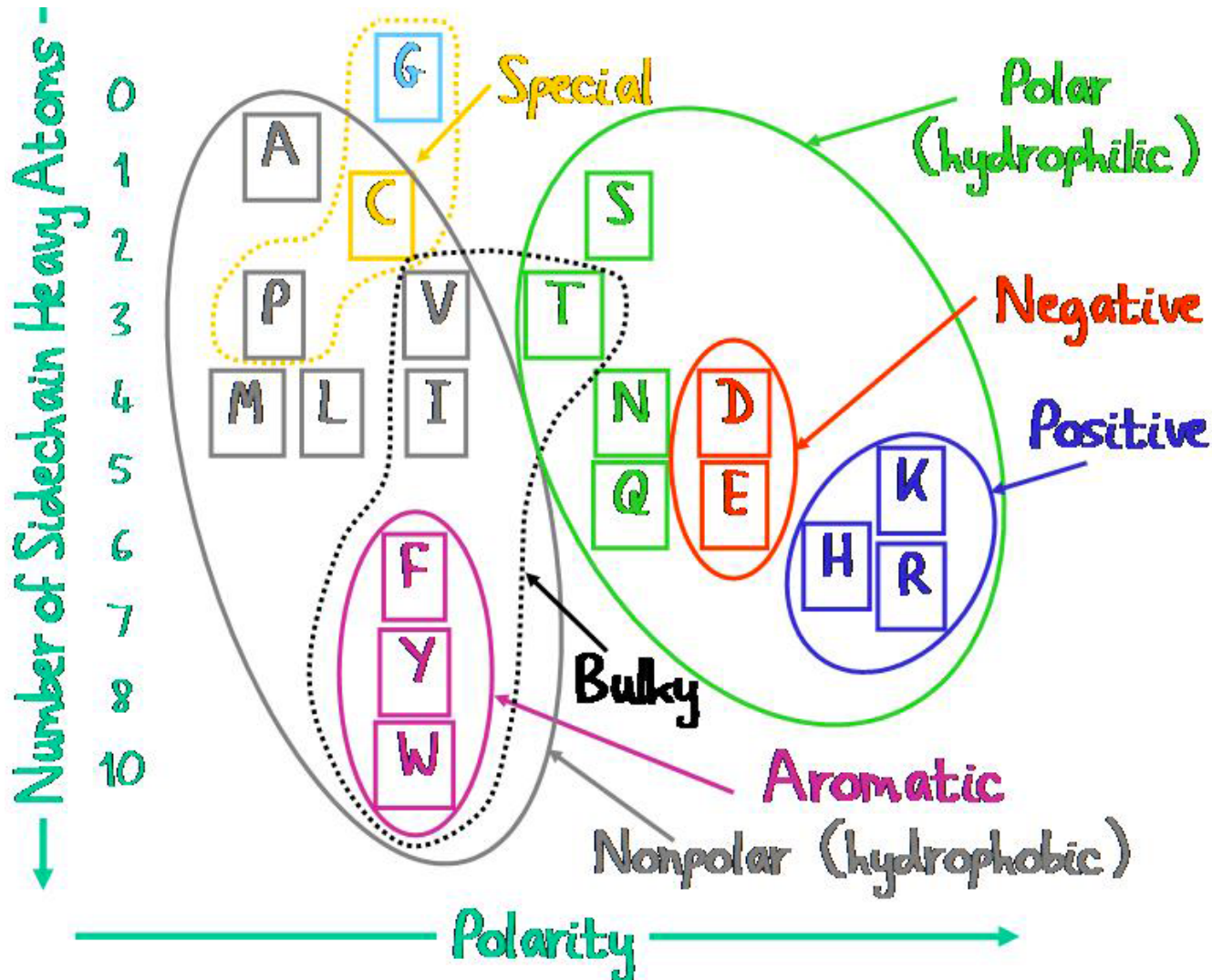


L-Valine
(Val / V)

Amino acid properties

- Amino acid side chains have a wide range of properties. These differences bring about the 3D structures of proteins.
- Examples:
 - Large side chains take up more space than small ones
 - Hydrophobic side chains want to be near one another
 - Hydrophilic side chains form hydrogen bonds to one another and to water molecules
 - Negatively charged (acidic) side chains want to be near positively charged (basic) side chains

Amino acid properties



There are many properties.

They cluster logically.

Slide from Michael Levitt

45

You don't need to memorize which amino acids have which properties

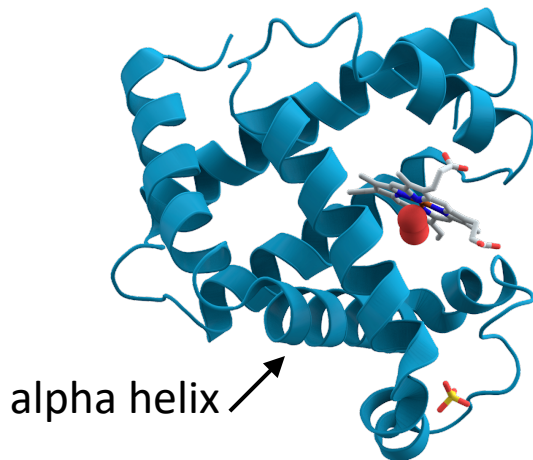
Protein structure: a more detailed view

Secondary structure

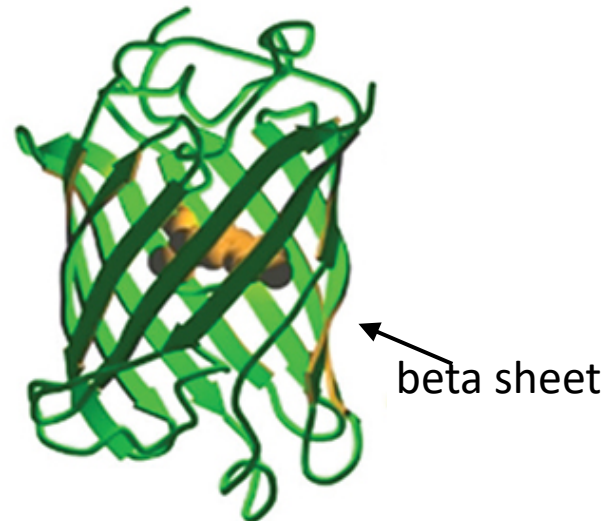
Secondary structure

- “Secondary structure” refers to certain local structural elements found in many proteins
 - These are energetically favorable primarily because of hydrogen bonds between backbone atoms
- Most important secondary structure elements:
 - alpha helix
 - beta sheet

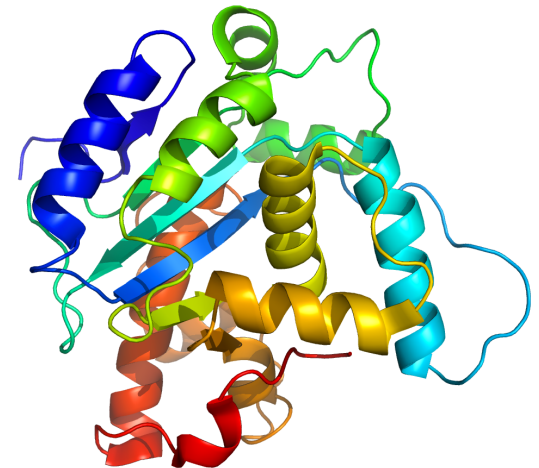
Myoglobin



Green Fluorescent Protein



Pop2p

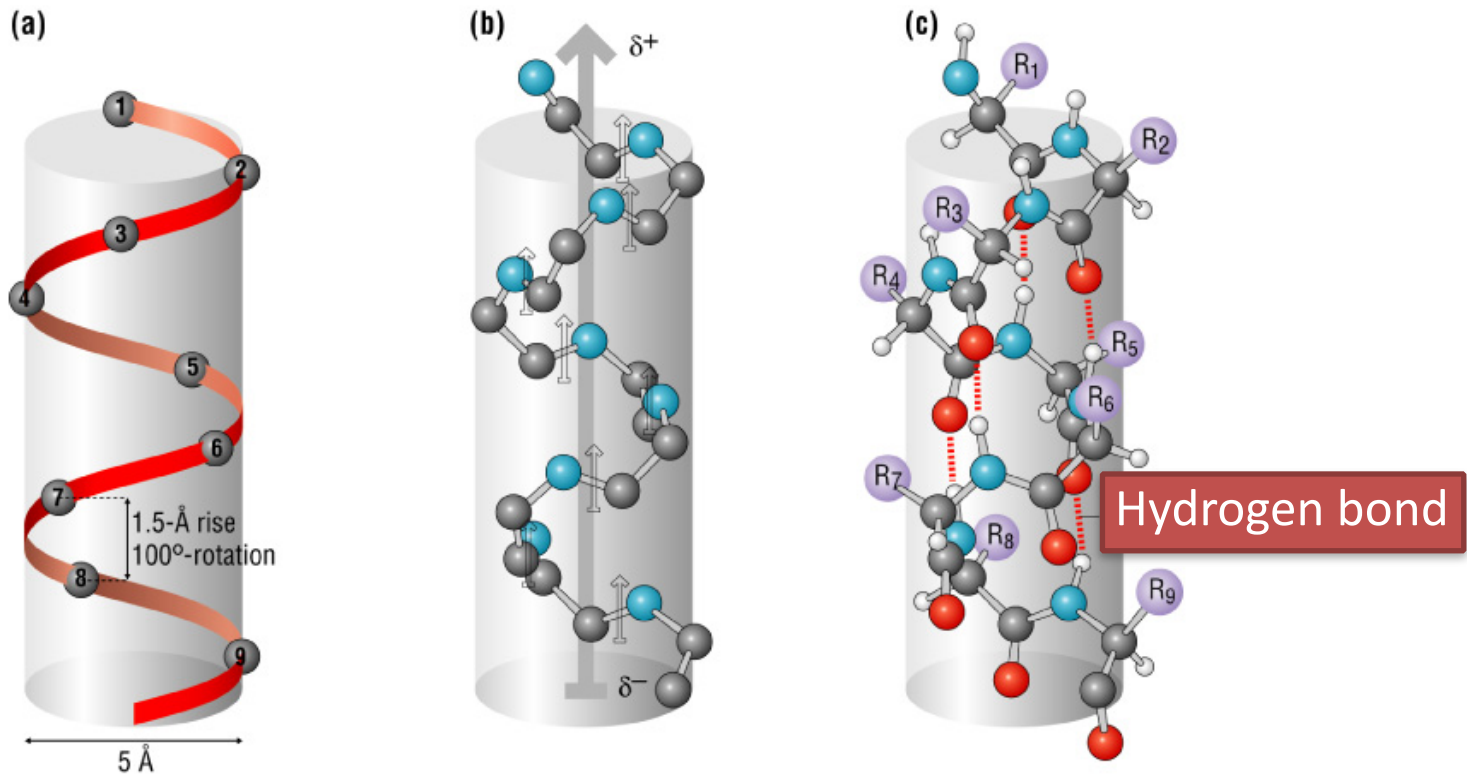


<https://upload.wikimedia.org/wikipedia/commons>,

http://www.biotech.com/assets/tech_resources/11596/figure2.jpg

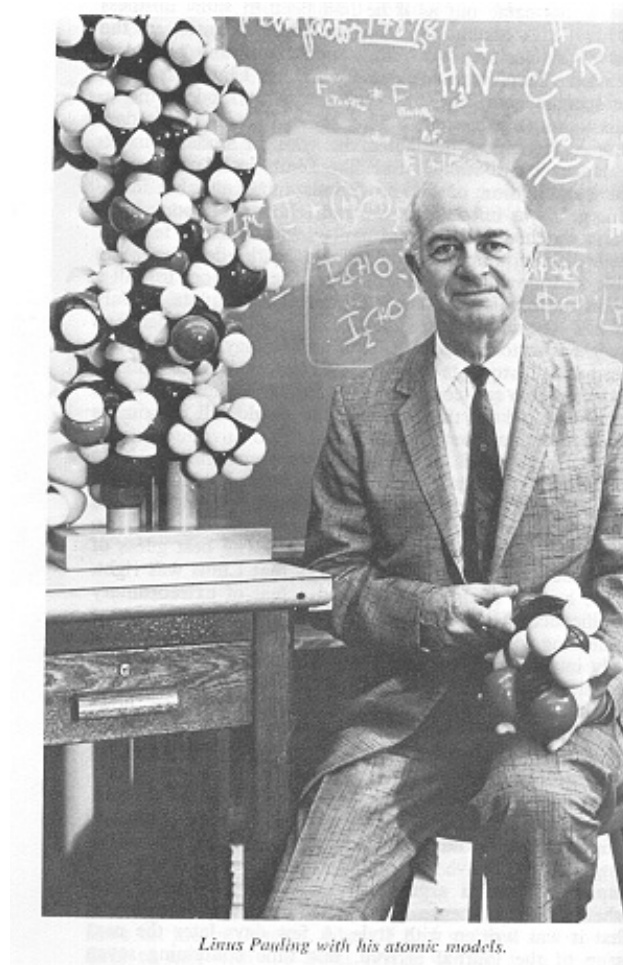
http://upload.wikimedia.org/wikipedia/commons/e/e6/Spombe_Pop2p_protein_structure_rainbow.png

The alpha helix



*Image from "Protein Structure and Function"
by Gregory A Petsko and Dagmar Ringe*

The alpha helix



Linus Pauling



Berkeley
UNIVERSITY OF CALIFORNIA

Reserved for
Nobel Laureate

Nobel Laureate Reserved Space
Parking Permit
Required At All Times

Violators will be cited and / or towed
Per UCB violation codes 101-1208 VC 2261(a)
For Towed Vehicles Call 510.642.4600

②
COLL

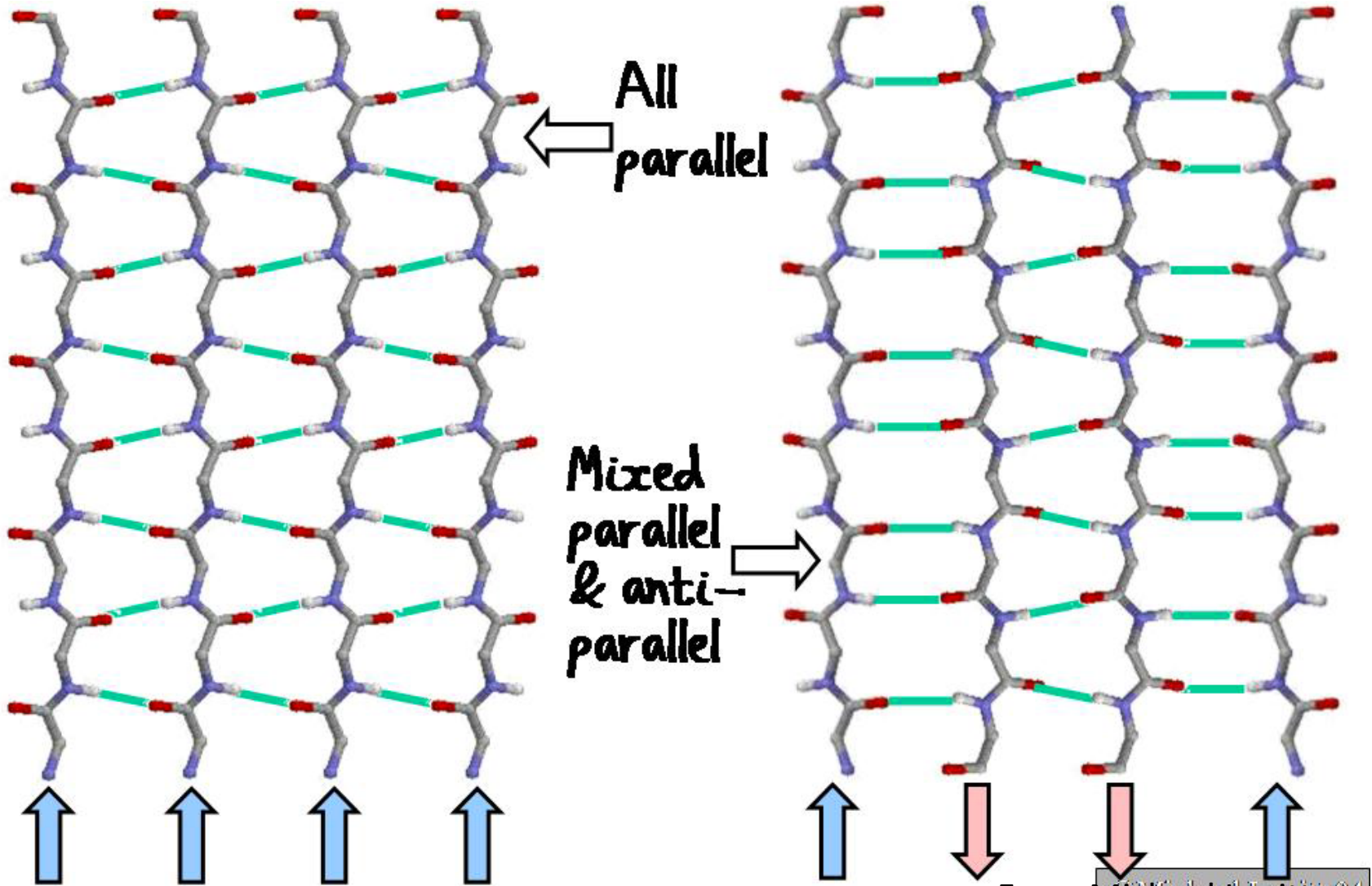
RESERVED

RESERVED PARKING
COLLEGE OF CHEMISTRY
9/30/2016
6:30 AM-12:00 PM
Stephen Isaacs

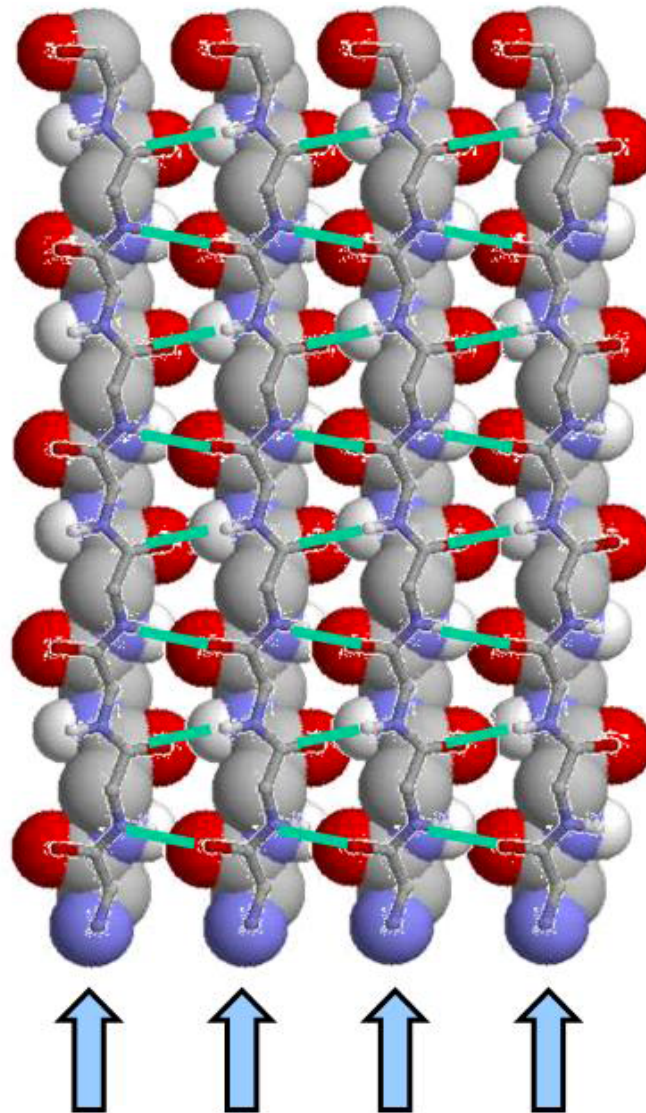
RESERVED PARKING
COLLEGE OF CHEMISTRY
9/30/2016
1:00 PM-6:00 PM
Prof. Ron Dror

RY

The beta sheet



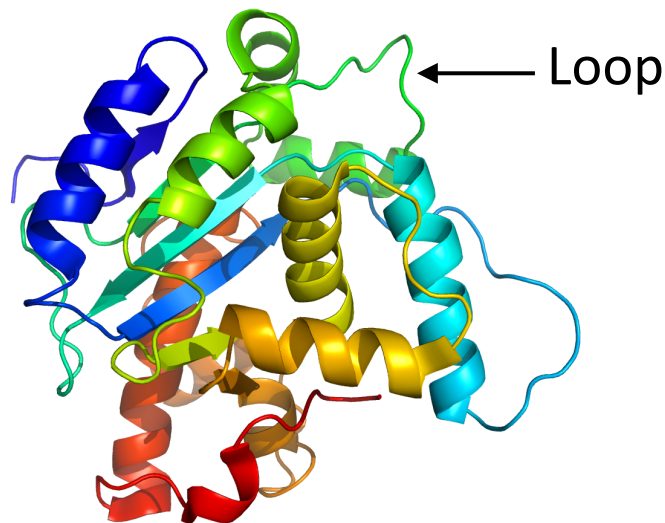
The beta sheet



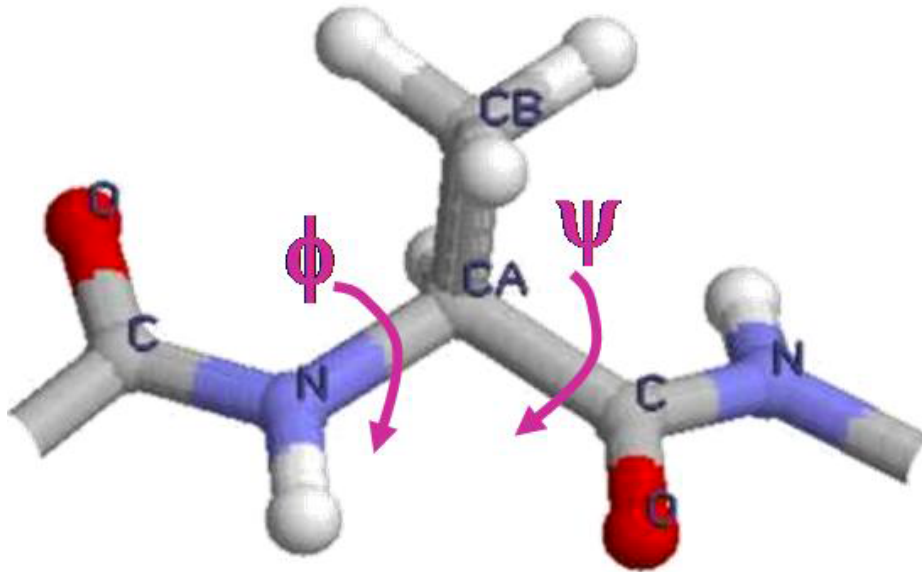
From Michael Levitt

Other secondary structure

- There are several less common secondary structures
- Regions connecting well-defined secondary structure elements are often referred to as “loops”



BACKBONE DEGREES OF FREEDOM



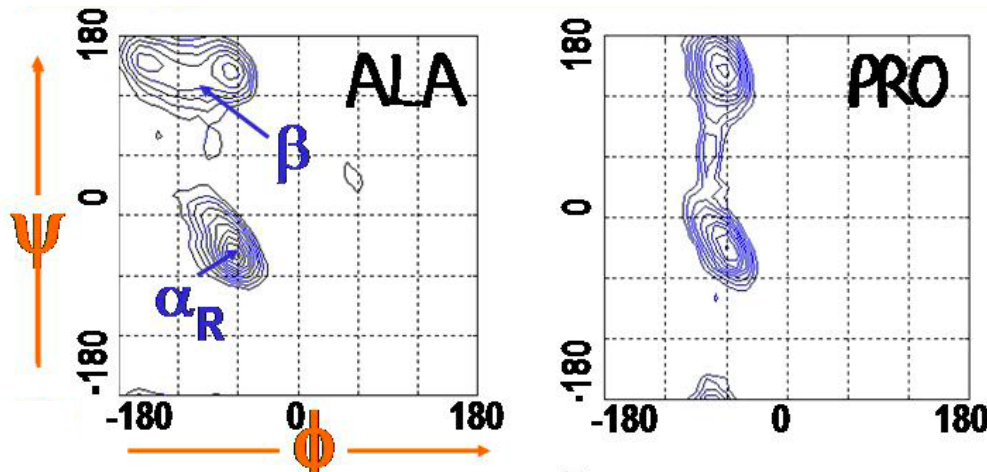
- The torsion angle rotating about the N-CA bond is called ϕ
- The torsion angle rotating about the CA-C bond is called ψ
- Together they are the (ϕ, ψ) angles

From Michael Levitt

- The remaining backbone bond (N-C, the “peptide bond”) is rigid

Ramachandran diagrams

- A plot showing a distribution in the (Φ , Ψ) plane is called a Ramachandran diagram
 - Such a diagram can be a scatterplot, or a two-dimensional histogram visualized as a contour map or heat map
 - For example, one might make a Ramachandran diagram for many residues of the same amino acid type
- Some amino acid types have distinctive Ramachandran diagrams



Ala is typical
Pro is unusual

Image from
Michael Levitt

- Alpha helices and beta sheets have characteristic Ramachandran diagrams

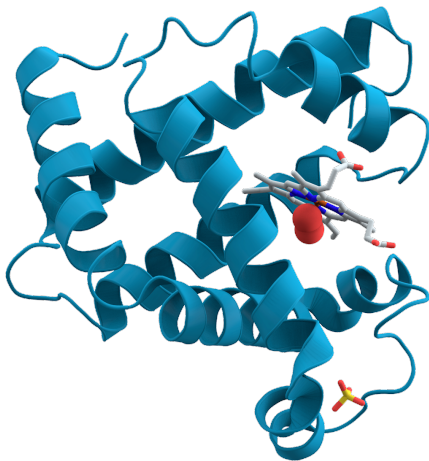
Protein structure: a more detailed view

**Tertiary structure, quaternary structure,
and domains**

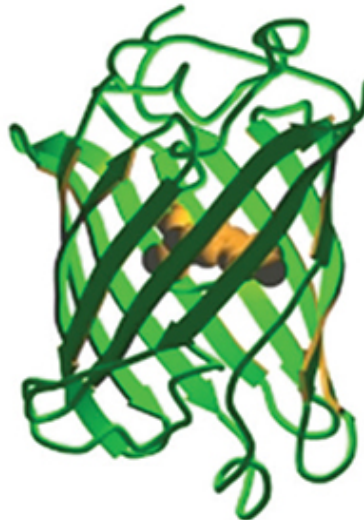
Tertiary structure

- Tertiary structure: the overall three-dimensional structure of a polypeptide chain

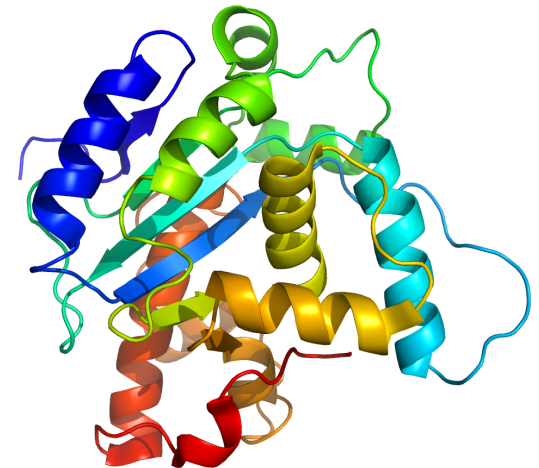
Myoglobin



Green Fluorescent Protein



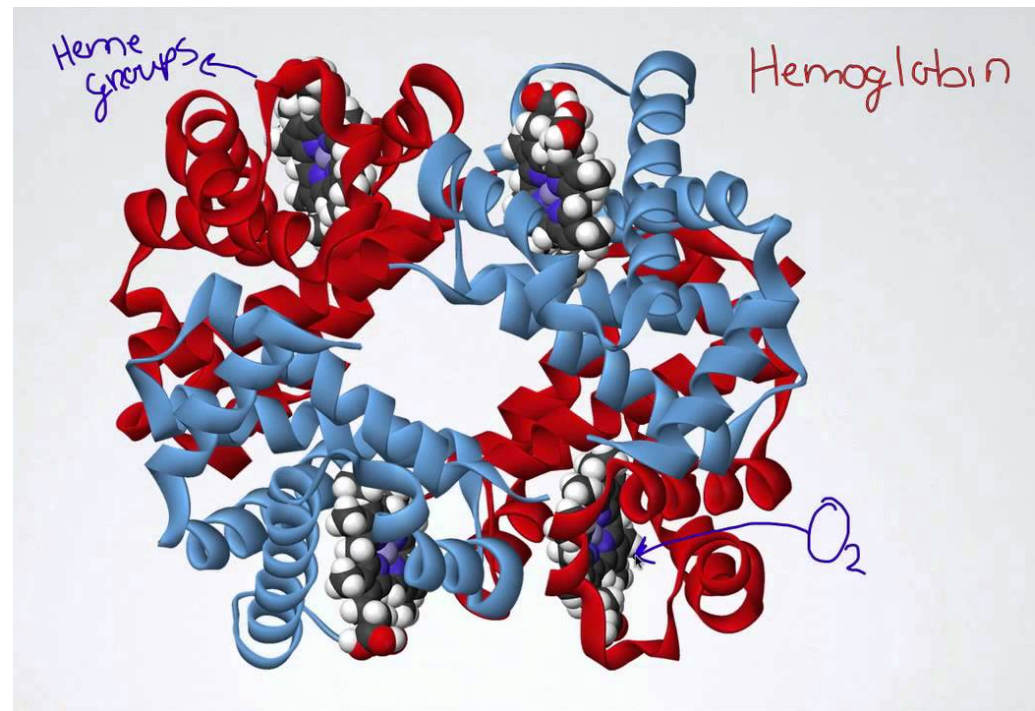
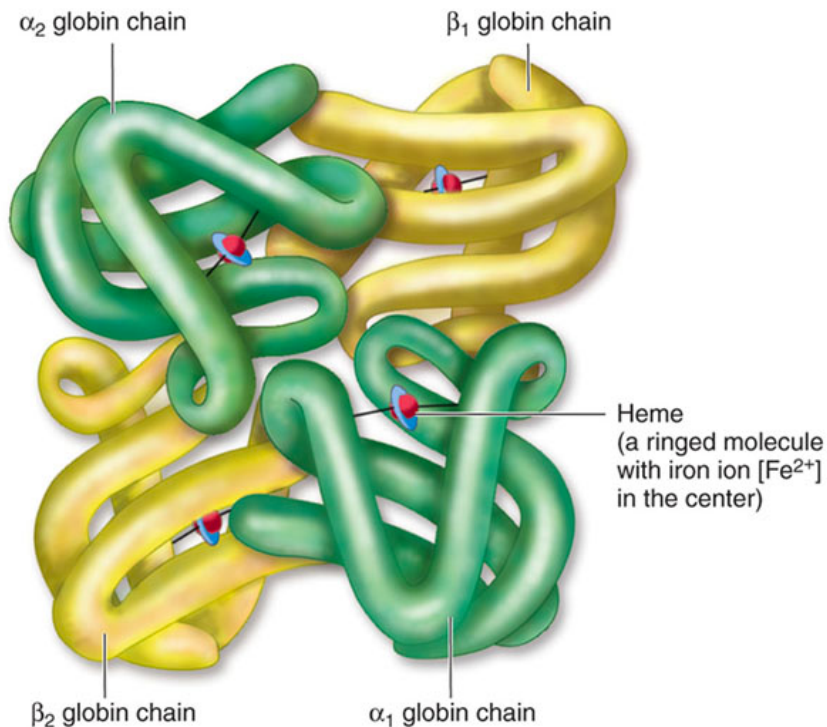
Pop2p



Quaternary structure

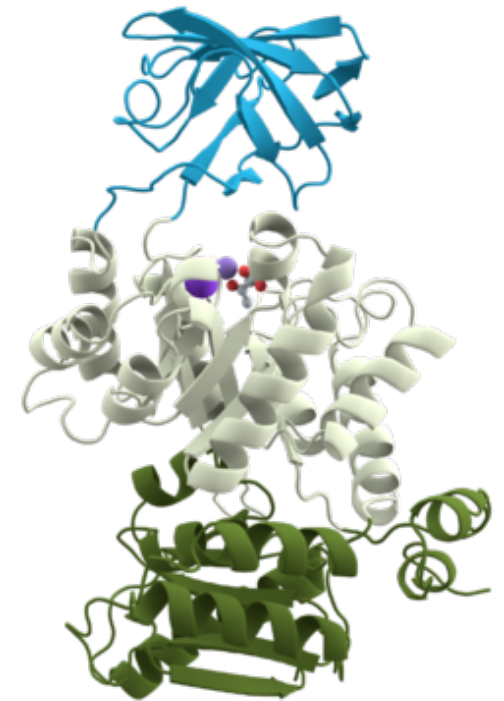
- Quaternary structure: the arrangement of multiple polypeptide chains in a larger protein

Molecular Structure of Hemoglobin



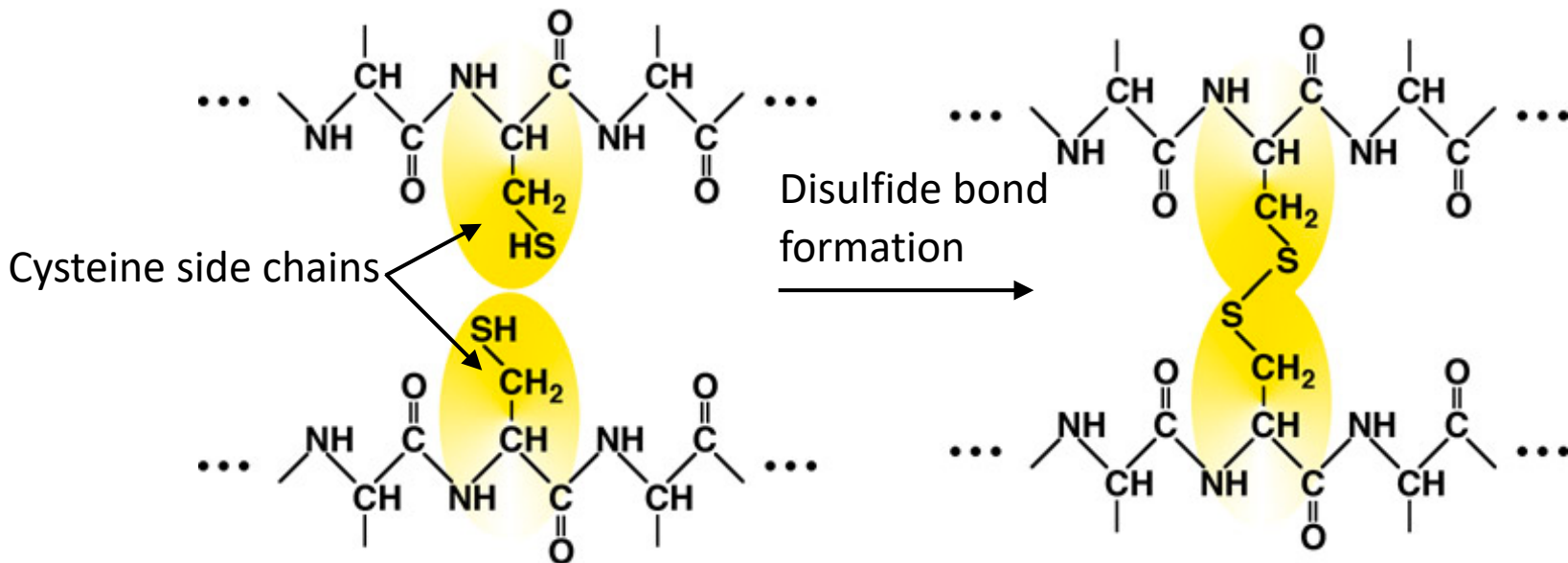
Domains

- Large proteins often consist of multiple compact 3D structures called *domains*
 - Many contacts with a domain. Few contacts between domains.
 - “Domain \approx blob”
- One polypeptide chain can form multiple domains, and a single domain may include portions of several polypeptide chains



Disulfide bonds

- One particular amino acid type, cysteine, can form a covalent bond with another cysteine (called a disulfide bond or bridge)
- Apart from the bonds within an amino acid residue and the peptide bonds that connect residues, disulfide bonds are the most common covalent bonds within a protein
- In a typical cellular environment, disulfide bonds can be formed and broken quite easily



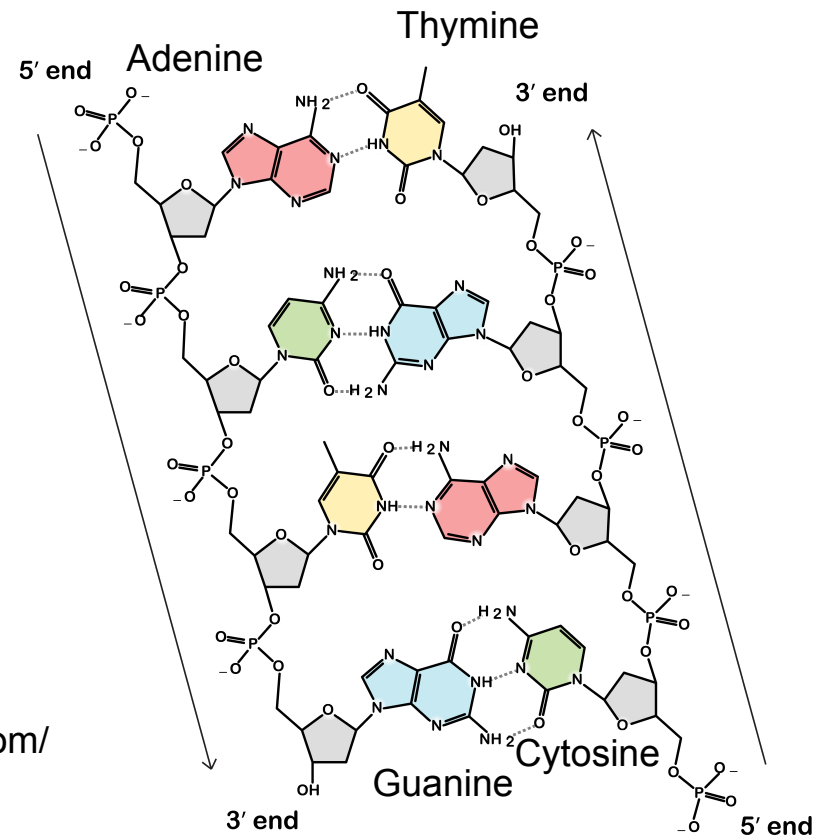
Structures of other biomolecules

What determines the structure of other biomolecules?

- The physical interactions that determine protein structure also determine the structures of other biomolecules
 - More generally, the great majority of the material covered in this course for proteins applies to other biomolecules as well

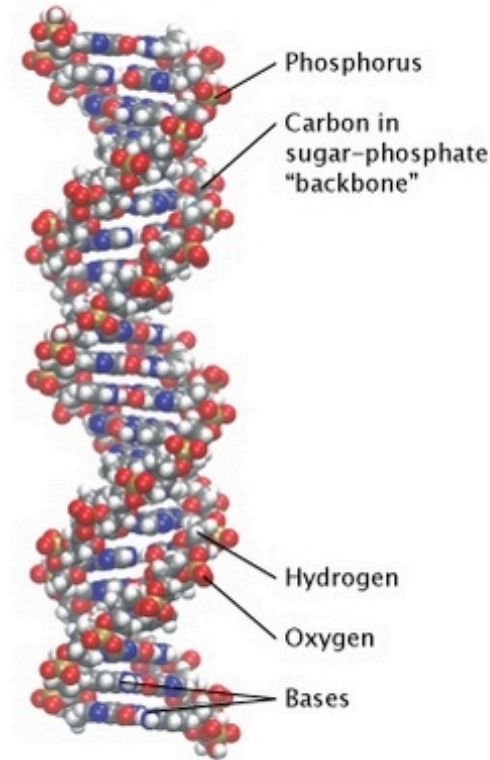
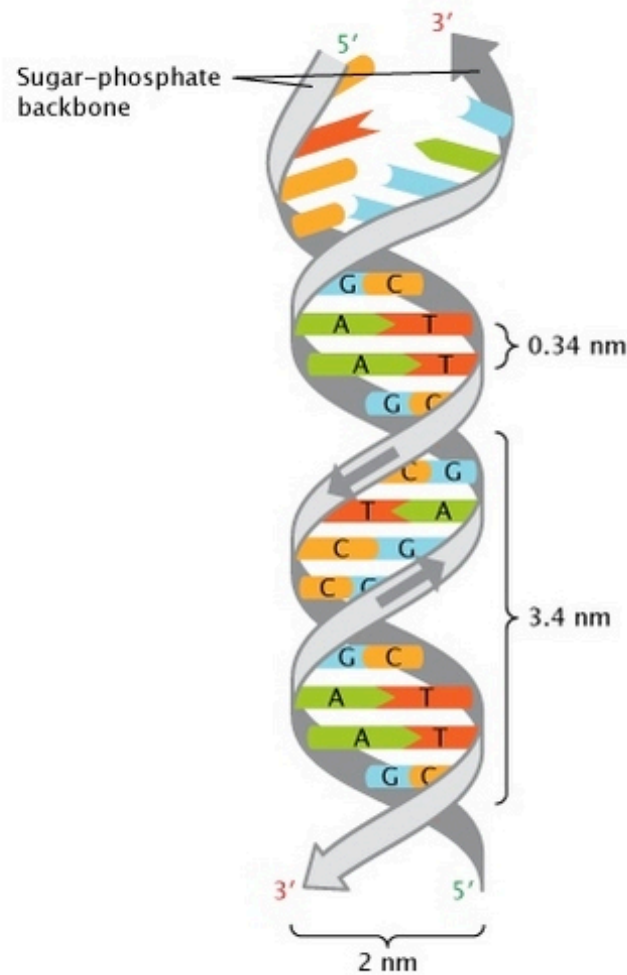
DNA

- DNA (deoxyribonucleic acid) stores the genetic code
- DNA, like protein, is a string of units with a uniform backbone
 - The units are nucleotides, instead of amino acid residues
 - Different nucleotides contain different nucleobases (“bases”) instead of side chains
- Only four common DNA bases
 - Adenine pairs with Thymine
 - Guanine pairs with Cytosine



DNA

- DNA forms one dominant 3D structure: a double helix
 - DNA usually acts more as information storage than as “machinery”
 - Long stretches of double helix can form coarser-scale structures, as we’ll see later on





Cambridge, 1953. Shortly before discovering the structure of DNA, Watson and Crick, depressed by their lack of progress, visit the local pub.



CARTOONSTOCK.com

Search ID: shrn2169

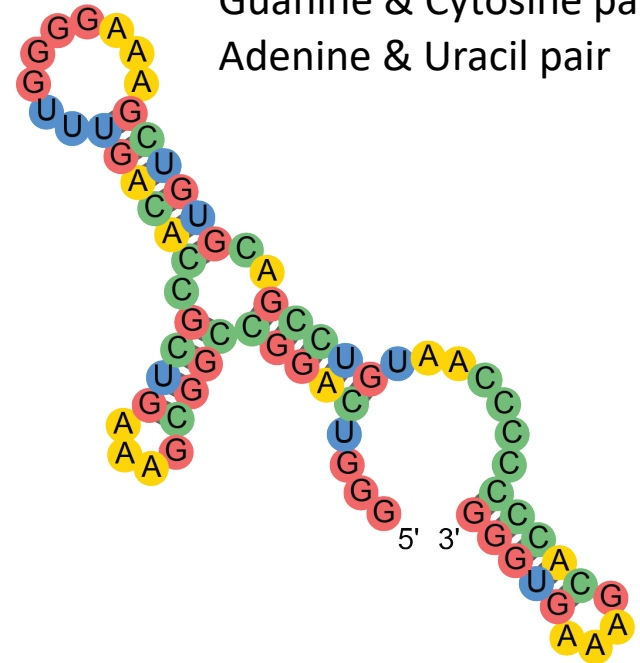
"IT'S NOT SUPPOSED TO BE A
TRIPLE HELIX, IS IT?"

RNA

- RNA (ribonucleic acid) is a string of nucleotides, like DNA
- RNA, however, generally occurs as a single string (strand) rather than paired strands
- RNA bases often pair with other bases in the same RNA strand
 - Much work on RNA structure focuses on the “secondary structure”: which bases pair with one another
 - Note that “secondary structure” has different meanings for RNA and protein
- RNA can form machines with well-defined, varied 3D structure
 - Example: RNA in the ribosome

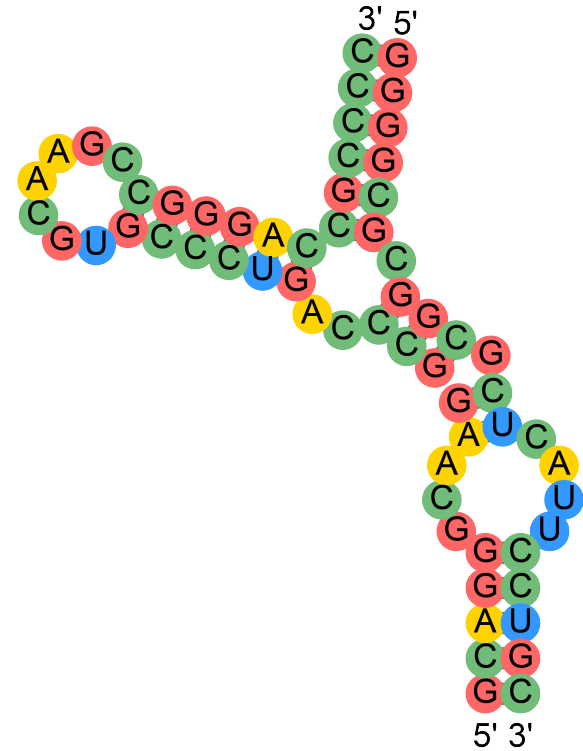
Typically:

Guanine & Cytosine pair
Adenine & Uracil pair



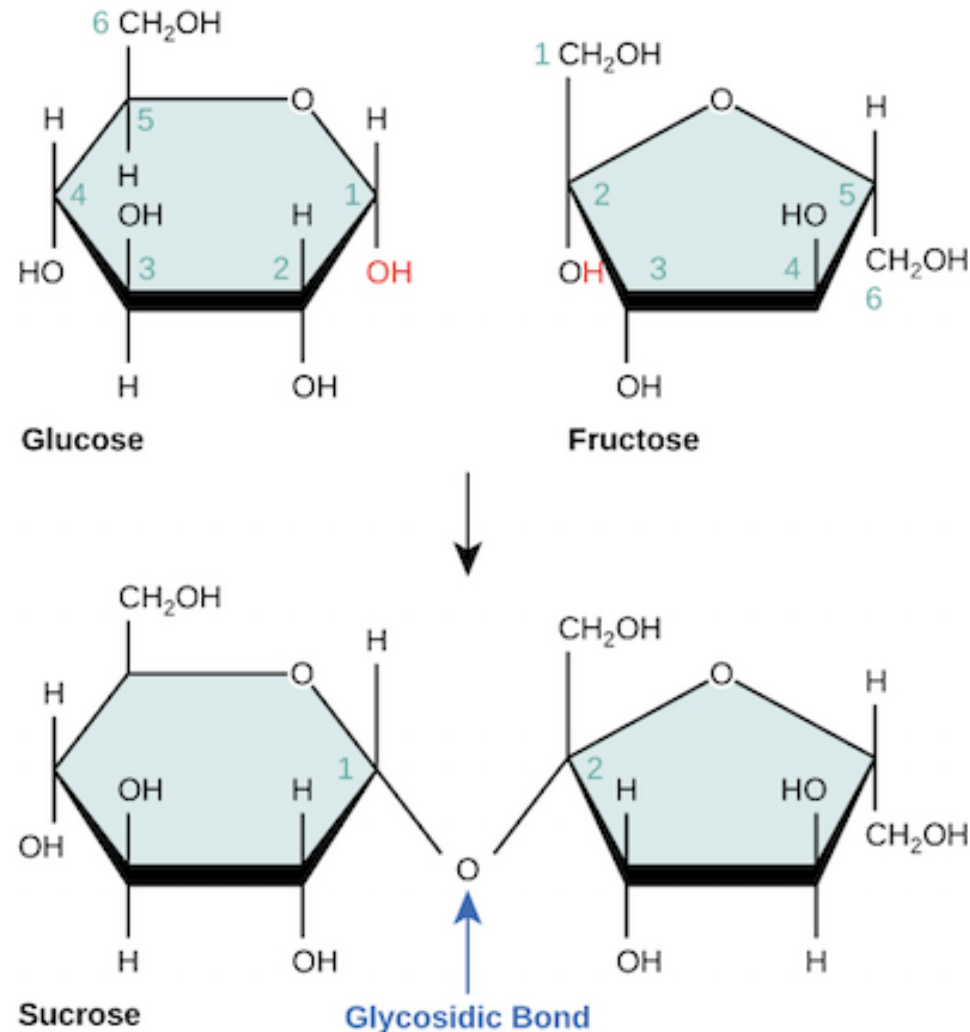
RNA

- Frequently, a single RNA is made up of multiple strands
 - Bases pair across strands
 - Secondary structure often includes multiple strands



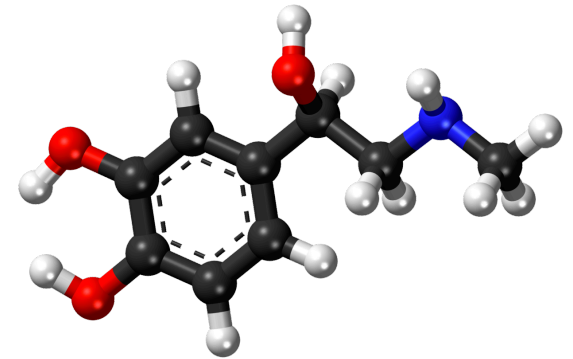
Glycans (e.g. carbohydrates)

- The base units are called “mono-saccharides”
- When they are linked through glycosidic bond, they are called glycans
- Examples: starch, cellulose, chitin
- In cells, glycans are often attached to proteins (“glycosylation”)



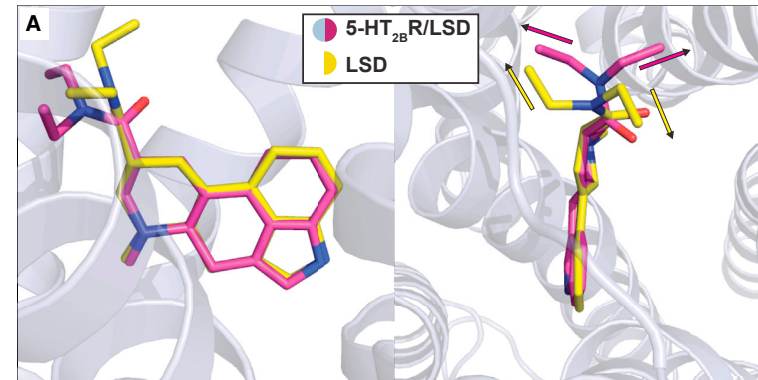
Small molecules

- Most drugs and many hormones, neurotransmitters, and other natural signaling molecules are “small molecules” (~100 atoms or fewer)
- Cambridge Structural Database is a repository of small molecule 3D structures, generally from x-ray crystallography
- However, these molecules are usually highly flexible and thus likely to take on a different 3D structure when bound to a protein



Adrenaline (epinephrine)

https://upload.wikimedia.org/wikipedia/commons/thumb/7/76/Epinephrine_ball-and-stick_model.png



LSD on its own (yellow) and receptor-bound (magenta)

Wacker et al., *Cell* (2017)