Biomolecular Structure, Including Protein Structure

• Please interrupt if you have questions, and especially if you're confused!

Outline

- Visualizing 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 (dominated by proteins because of their versatility in structure and function)



60000

APPROV

37

CONTRACTOR OF

Sec.

CONCERCISION OF

Contraction of the local data

STATE OF STATE

AL.

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 (closepacked).
- There are a lot of atoms. Simplified visual representations help you figure out what's going on.

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.



- http://www.rcsb.org/pdb/home/home.do
- A collection of (almost) all published experimental structures of biomacromolecules (e.g., proteins)
- Each identified by 4-character code (e.g., 1rcx)
- Currently ~134,000 structures. 90% 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
 - VMD: mol pdbload 1rcx



Chemical (two-dimensional) structure of proteins

Two-dimensional (chemical) structure vs. three-dimensional structure

- Two-dimensional (chemical) 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 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



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



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 each atom specify its geometry



 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 energy 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 bond graph. Energy needed to change equilibrium — largely geometry and sterics.
 - Non-bonded forces: act between all pairs of atoms

Bonded: Bond length stretching

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



Bonded: Bond angle bending

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



Bonded: Torsional angle twisting

Certain values of each torsional angle are preferred over others.



Non-bonded: Electrostatic interaction



- Like charges repel.
 Opposite charges attract.
- Acts 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
 atoms it's connected to.

Non-bonded: 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

Water molecules form hydrogen bonds

- Water molecules form extensive hydrogen bonds with one another and with protein atoms
- The structure of a protein depends on the fact that it is surrounded by water



http://like-img.com/show/hydrogen-bond-water-molecule.html

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


Hydrophobic effect

- Hydrophobic molecules cluster in water
 - "Oil and water don't mix"



http://science.taskermilward.org.uk/mod1/KS4Chemistry/AQA/Module2/Mod%202%20img/Oil-in-Water18.jpg

• This is critical to protein structure

EXPLAINING HYDROPHOBICITY





- •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.

#Midhael Leviff (04

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

If we think about folding rates (TIME!)



Cyrus Levinthal's famous "paradox"

Proteins can fold on the order of milliseconds, clearly it does not sample the large number of torsional angles.

Even in the '60s, the idea of calculating energies of protein structures using computers, taking the sum of torsions and interaction energies, exist. Hydrophobic effect is the dominant driving force for folding

(using solvent transfer model for energy)



Kauzmann 1959; Tanford 1962

Protein structure: a more detailed view

"Hierarchies" 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



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



Slide from Michael Levitt

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



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

http://www.biotek.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



Lines Parling with his clove, small.

Linus Pauling

https://www.msu.edu/course/lbs/333/fall/images/PAULING.JPG

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

• The torsion angle rotating about the CA--C bond is called W

Together they are the
(\$\overline{\phi}\$,\$\overline{\phi}\$) 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



Quaternary structure

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



Domains

- Large proteins often consist of multiple compact 3D structures called *domains*
 - Many contacts with a domain. Few contacts between domains.
 - "Domain ≈ blob"
- One polypeptide chain can form multiple domains



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

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 only common covalent bonds within a protein
- In a typical cellular environment, disulfide bonds can be formed and broken quite easily



http://www.crc.dk/yeast/yeasthome/yeasthome/images/ls_jpgs/fig2.jpg ⁶⁰

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

other common biomolecules: DNA/RNA Small molecules Sugars A cell is like a stranded spaceship... it might be able to harvest external resources, but it has to have the ability to build its own parts. In biology, this is completely done through biopolymers.



63

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



Khan Academy (<u>https://ka-perseus-images.s3.amazonaws.com/</u>9d1d07df110f35ba532c792c73bceb164679a165.png)

DNA

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



http://www.nature.com/scitable/content/ne0000/ ne0000/ne0000/104944953/73_1_2.jpg



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

Taking advantage of DNA base pairing specificity for structural engineering: DNA origami



nature.com



https://laboratoryinfo.com/dna-vs-rna/

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 welldefined, varied 3D structure
 - Example: RNA in the ribosome



http://www.tbi.univie.ac.at/~pkerp/forgi/_images/1y26_ss.png



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)

Sugars: glycans

- The base units are called "mono-saccharides".
- When they are linked through glycosidic bond, they are called glycans. Yet another biopolymer that is essential for life!



Sugars: glycans

- Here are 9 used by mammalian cells to build Nlinked glycans (this means they are attached to the amino acid "N" (asparagine).
- Important for a large number of cellular functions including immunorecognition and protein folding.
- cellulose and chitin are also glycans (polysaccharides)

