Protein Structure: Three-dimensional structure

Background on protein composition:

Two general classes of proteins

- Fibrous - long rod-shaped, insoluble proteins. These proteins are strong (high tensile strength). Examples: keratin, hair, collagen, skin nails etc…
- Globular - compact spherical shaped proteins usually water-soluble. Most hydrophobic amino acids found in the interior away from the water. Nearly all enzymes are globular… an example is hemoglobin

Proteins can be simple - no added groups or modifications, just amino acids

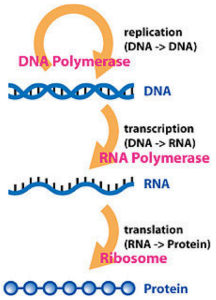
Or proteins can be conjugated. Additional groups covalently bound to the amino acids. The naked protein is called the apoprotein and the added group is the prosthetic group. Together the protein and prosthetic group is called the holoprotein. Ex. Hemoglobin

Levels of Structure

Type of Structure:	Defines:	Type of Bonds:
Primary (1°)	Order of amino acids	Covalent bonds
Secondary (2°)	Local structure (α-helix, β-sheet, loop)	Non-covalent interactions, disulfides
Tertiary (3°)	Overall fold, 2° elements organize and compact—low surface to volume ratio	Non-covalent interactions, disulfides
Quaternary (4°)	Subunit organization, dimers, macromolecular assembly	Non-covalent interactions, disulfides

How do we know a protein's primary sequence?

Many older chemical methods, but since the human genome project we now rely primarily on the central dogma of molecular biology:



The Nucleotide Sequence Tells Us the Amino Acid Sequence!

- Translation software (ExPASy Translate is free!) will translate your DNA sequence to an amino acid sequence.

If you don't know the gene…sequence the protein

1. Eliminate any quaternary structure--chromatography
2. Denature the protein (urea, GdnHCl--chaotropes) get rid of any secondary and tertiary structure
3. Cleave any Disulfides (reduction—beta-mercaptoethanol)
4. Sequence

Acid Hydrolysis—busts all the peptide bonds

Edman Degradation—N-terminal analysis

Enzyme Hydrolysis

Trypsin—cleaves C-terminally after Arg, Lys

Chymotrypsin—cleave C-terminally after Phe,Tyr,Trp

Sanger Sequencing—first methodology

We'll learn in more detail how to determine the amino acid sequence of a protein using mass spectrometry during our protein expression and purification lecture!

Primary Sequence: Mutations

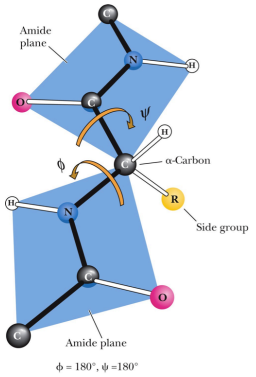
Mutations in our DNA occur for different reasons:

- Mistakes during DNA replication
- Environmental DNA damage

The mistakes (e.g. a Cytosine to Guanine point mutation), will be observed also at the mRNA and amino acid level. What do you think these mutations do to the overall structure of the protein? It's function?

BUT, it's important to note that some proteins that have **high sequence homology have divergent functions**, e.g. the serine proteases elastase and thrombin

Primary Structure → Secondary/Tertiary Structure → FUNCTION!

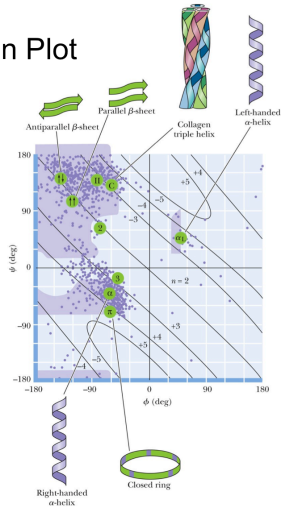


Protein structure is defined by the peptide plane and the 2 degrees of freedom around the peptide bond

Two basic predictable forms of secondary structure
Alpha helix and beta pleated sheet
-Other forms are specialized (collagen)
- Beta sheets, tight turns and beta bulge

Ramachandran Plot

- Ramachandran plot shows the values of the torsional angles and allows prediction of the conformation.
- The angle will depend on the two amino acids. The steric hindrance between the functional groups determine the angle of stability.
- There are only three real means of conformation the sheets and helices.
- Glycine and proline alter the possible angles.

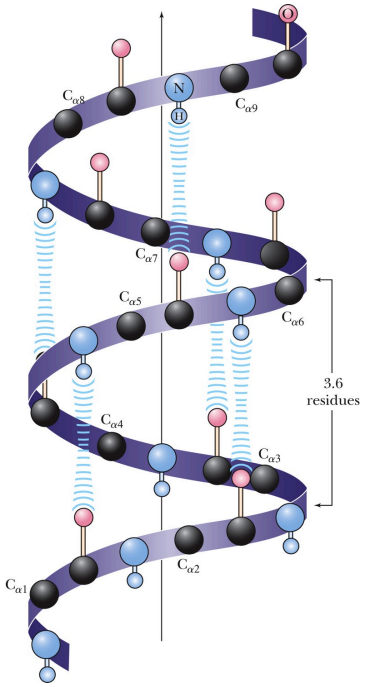


α-helix

Discovered by Linus Pauling
- 2 Nobel prizes.
- Discovered folding while sick in bed!
- Treated poorly as a result of his anti-nuclear stands - 2nd NB prize
- Pushed high doses of vitamin C

- The helix is a right handed twist of the backbone - notice when we are looking at this the side groups are NOT considered
- Notice where the amino acids are.
- Hydrogen bonding occurs between the carbonyl and the amino group four residues away. The bonding takes place within the same chain.

- A run of proline residues lead to breaking the helix structure. Wh
- Formation of α-helices are governed by **HYDROGEN BONDS**!
- One helix turn is 3.6 amino acid residues, and involves 13 atoms from the O to the H of the H bond
- For an a-helix of n residues, there are n-4 hydrogen bonds!
- **Helix capping**—the last 4 amide hydrogens/carbonyl oxygens cannot H-bond. Proteins compensate by folding other parts of the protein to facilitate hydrogen bonding.
-

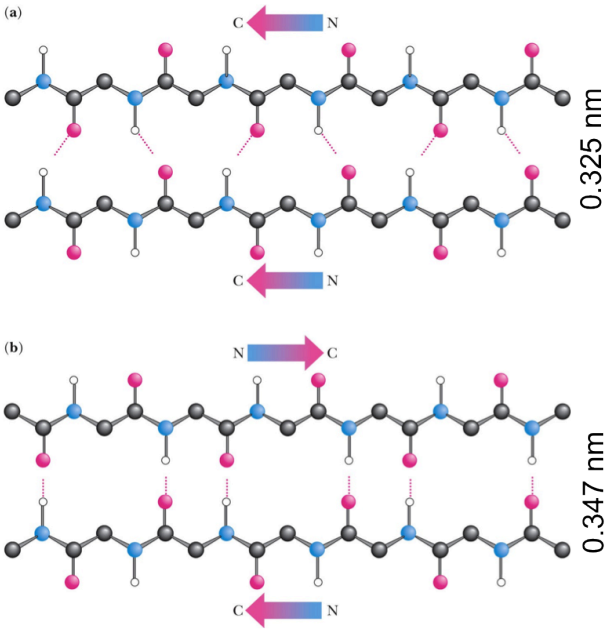


H-bonds all point in the same direction
Amide bond has a **dipole moment**—cumulatively, the helix has a large dipole moment

β-Pleated Sheet
Notice that there are no "turns" pictured here—the sheet here is not a contiguous amino acid sequence!
- Side chains perpendicular to plane of the sheet
- Defined by a different hydrogen-bonding network
- Note: hydrogen bonds occur **inter**strand

Parallel: Adjacent chains run in the same direction
- Bent H-bonds
- Normally large, >5 strands
- Hydrophobic side chains distributed on both sides of sheet

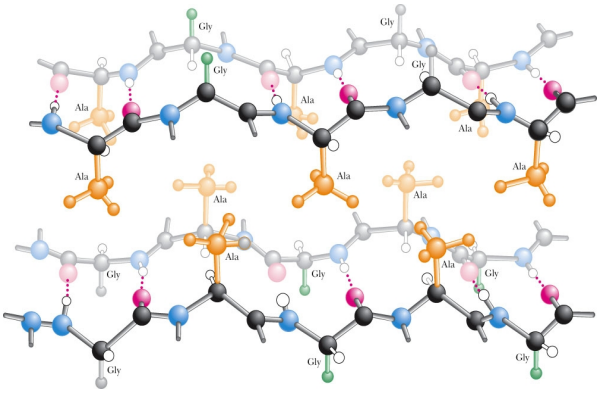
Anti-Parallel: Adjacent chains run in the opposite direction
- More extended H-bond conformation
- Can consist of 2 strands
- Hydrophobic on same side—alternates hydrophilic and hydrophobic in primary sequence



Silk and keratin

Anti parallel conformations are stronger - alignment of H bonding.

- Often found in silk
- R groups can interact - glycine and alanine



Unique structure of collagen

A special helical protein

- biological significance - fibrous, structural component
- Type I collagen is found in bone, tendon and skin, II in cartilage and III in blood vessels

- Glycine R group face inside others outside
- up to 30% are proline or hydroxyproline - important for maintaining secondary structure
- hydroxyprolines involved in H bonding of three strands together
- helical structure formed by three left handed helices twisted to form a right handed superhelix (gives strength)
- hydrogen bonding between 3 helices
- (thus the glycine)
- covalent bonding of lysine between strands necessary for strength

Hydroxylations on pro are performed by an enzyme called prolyl hydroxylase, which is an enzyme that requires vitamin C as a cofactor in the reaction.
Absence of vitamin C in the diet reduces hydroxylation of pro, and collagen fibres begin to break down and new collagen not formed properly.
Lack of vitamin C causes scurvy because collagen fibres are not formed properly, and this causes skin lesions, weakened gums so teeth fall out etc.

A special helical protein

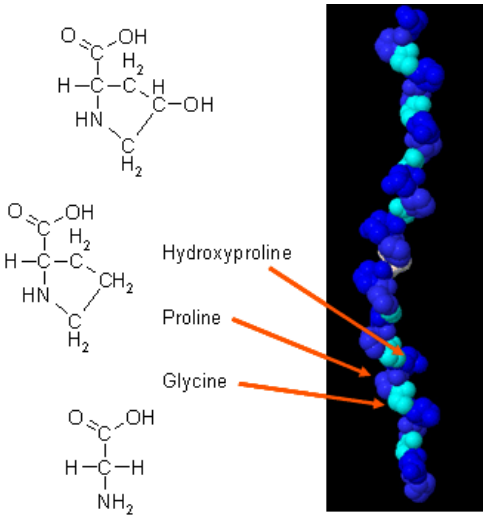
Equally important is hydroxy-lys catalysed by lysine hydroxylase. Attached to the lys residues are three sugars gal-gal-glu, and these enable H-bonding to occur between triple helices, which is essential for stability of the greater complex that binds fibers together to form a matrix bed to binds cells to the matrix and form a tissue.

Collagen Related Disease

Loss of flexibility with age is likely due to increased amount cross-linked collagen compared to younger tissue

■Scurvy – problems with sea voyages, lack of food other than salted meats.

- Symptoms include, swollen gums, loose teeth, small black-and-blue spots on the skin, and bleeding from small blood vessels are among the characteristic signs of scurvy.
- Caused when vitamin C (ascorbic acid) is lost from diet
- Vit C is needed to keep Iron reduced in the active site of prolyl hydroxylase. This is the enzyme responsible for conversion of proline to hydroxyproline. The H bonding of hydroxyproline is vital for the connective protein's function
- In 1795, the British Royal Navy provided a daily ration of lime or lemon juice to all its men. English sailors to this day are called "limeys", for lime was the term used at the time for both lemons and limes.



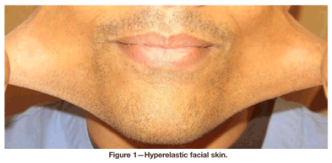


Figure 1—Hyperelastic facial skin.

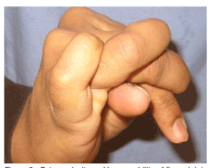


Figure 2—Extreme laxity and hypermobility of finger joints.

■Several heritable diseases result from mutations in the collagen

Marfan's Syndrom and Ehler's-Danlos syndromes - inherited disorder of connective tissue which affects many organ systems, including the skeleton, lungs, eyes, heart and blood vessels. All resulting from various mutation in collagen and other fibril associated proteins, ultimately affecting the structure and molecular interaction.

"Random coils" are not random

The segments of a protein that are *not helices or sheets* are traditionally referred to as "random coil", although this term is misleading:
Most of these segments are neither coiled or random
They are usually organized and stable, but don't conform to any frequently recurring pattern
Random coil segments are strongly influenced by side-chain interactions with the rest of the protein

Tertiary Structures in Globular Proteins: Packing

- Secondary structural elements form, then pack together to create tertiary structure
- Packing excludes water from the center of the folded protein—hydrophobic amino acids are buried in the center, no longer have to be solvated by water
- Cavities form that are complementary to any small molecules (or other large proteins)—binding sites/active sites (in the case of enzymes)----FUNCTION!

Large Proteins Fold Modularly

- Large globular proteins (normally >250 AAs), made up multiple distinct structures (DOMAINS) that are usually stable by themselves in aqueous solution
- Typical domain structures are hydrophobic cores with hydrophilic surfaces

Quaternary Structure: Driving Forces

The Bad News:

- Considerable entropy loss when subunits come together
- Loss of translational degrees of freedom
- Residues that were able to move at the subunit interface are now restricted

The Great News:

- Increased Van der Waals contacts—but nearly as many are lost with water as are made with the new oligomer
- Increased hydrophobic interactions—the money maker (roughly 100-200kJ/mol)
- Polar interactions at the interface
- Salt bridges/disulfides

Quaternary Structure

Homomultimeric
Identical Subunits

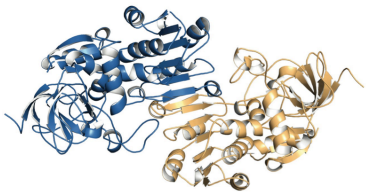
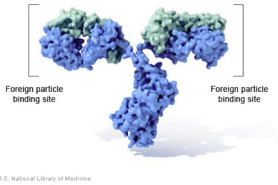


Figure 6.41

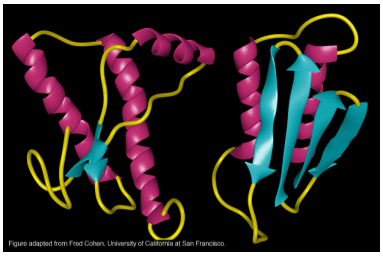
Liver Alcohol Dehydrogenase
(This guy is probably important to you…)

Heteromultimeric
Non-identical Subunits



Immunoglobulin G
(Antibodies—also important to you)

What kind of ***inter***molecular forces are we dealing with?



Quaternary Interactions Gone Awry: Amyloidoses

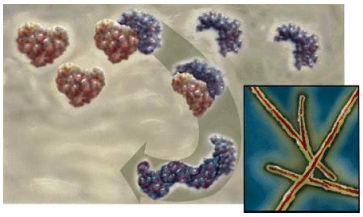
- We've covered interactions between already folded subunits.
- ***Unfolded or misfolded*** monomers can glob together to form aggregate structures, which organized into cross-β-sheets (amyloid)
- Amyloidoses are a major health problem in the ageing population (Alzheimer's disease, Systemic amyloidoses, etc.)

MAD COW DISEASE

- The prion protein exists in two forms. The normal, protein (PrPc) can change its shape to a harmful, disease-causing form (PrPSc). The conversion from PrPc to PrPSc then proceeds via a chain-reaction. When enough PrPSc proteins have been made they form long filamentous aggregates that gradually damage neuronal tissue. The harmful PrPSc form is very resistant to high temperatures, UV-irradiation and strong degradative enzymes.

Prion diseases arise in three different ways

1. Through horizontal transmission from e.g. a sheep to a cow (BSE).
2. In inherited forms, mutations in the prion gene are transmitted from parent to child.
3. They can arise spontaneously.



Route of infection

When cows are fed with offals prepared from infected sheep, prions are taken up from the gut and transported along nerve fibers to the brain stem. Here prions accumulate and convert normal prion proteins to the disease-causing form, PrPSc. Years later, BSE results when a sufficient number of nerve cells have become damaged, affecting the behavior of the cows.

Remember: Function is Dependent on Structure!

- **Heating** up proteins, adding **chaotropes** (molecules that disrupt noncovalent bonds—urea, guanadine HCl (GdmHCl)) OR reducing disulfide bond w/ β-mercaptoethanol causes proteins to lose their *quaternary*, *tertiary* and *secondary* structure and ***UNFOLD***.

Protein Folding - Secondary structure forms, then the protein begins to compact itself until it reaches the lowest energy state possible.

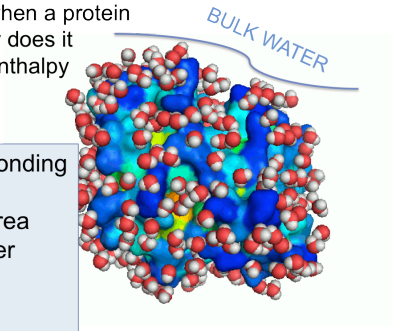
- Proteins fold spontaneously and on short (nanosecond!) timescales—but theoretically, a protein has a large number of degrees of freedom (all those ψ and ϕ angles in the peptide bonds), and if all possibilities were sampled it would take FOREVER (well, longer than the age of the earth for a 100-mer protein)—**Levinthal's Paradox**

How Does Protein Stability Arise?

We're losing a lot of entropy when a protein folds (floppy to ordered)—how does it overcome that loss (entropy-enthalpy compensation)?

$$\Delta G = \Delta H - T\Delta S$$

- Intramolecular hydrogen bonding
- Hydrophobic Collapse
- The reduction of surface area accessible to solvent (smaller solvent shell)
- Salt-bridges
- Van der Waals contacts



http://mspc.bii.a-star.edu.sg/tankp/help.html

Thinking About Proteins as Dynamic Structures

In solution, proteins are constantly in motion

Proteins can undergo ***conformational changes*** that involve the movement of whole sections of protein/side-chains/atoms—can be in response to stimuli (e.g. small molecule substrate)

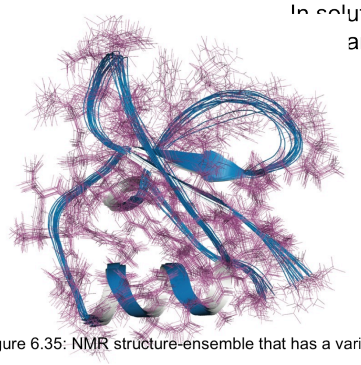
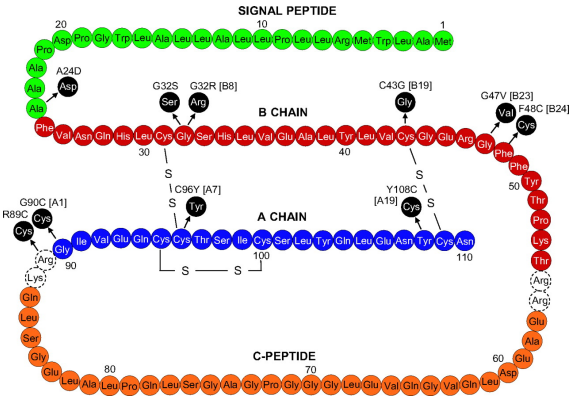


Figure 6.35: NMR structure-ensemble that has a variety of structures that fit the data

A Closer Look: Insulin What's Insulin?

Peptide hormone—also commonly referred to as a protein
2 chains: A- and B-chain, connected by ***inter***molecular disulfides
Pictured here is the **mature form** of the peptide (starts out as preproinsulin)



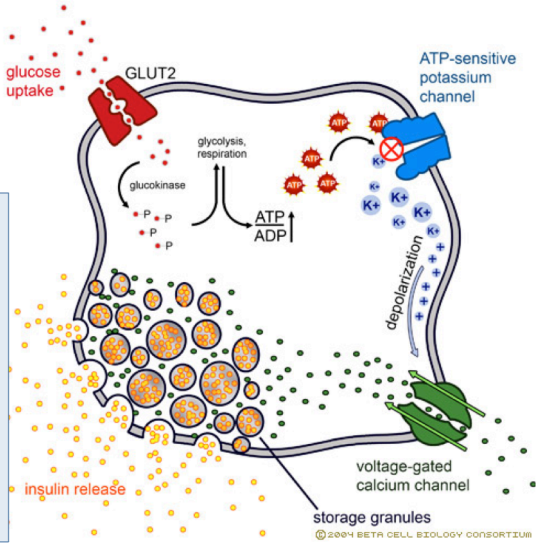
- Cleaved at the dashed circles (**what kind of amino acids are these?**) and in between the green and red chains by peptidases
- Some common mutations are depicted as black circles
- The preproinsulin is synthesized in pancreatic β-cells (inside the islets of Langerhans)
- The signal sequence is cleaved in the endoplasmic reticulum—prosinsulin
- ***Endopeptidases*** then cleave the C-chain away in the secretory vesicles, and we now have mature insulin.
- **Packaged in granules**
- **Association is driven by Hydrogen-bonding** between the C-termini of B chains
- Additionally, in the presence of zinc ions, insulin dimers associate into hexamers.
- **Monomers and dimers readily diffuse into blood, whereas hexamers diffuse poorly.**
- Absorption of insulin preparations containing a high proportion of hexamers is delayed

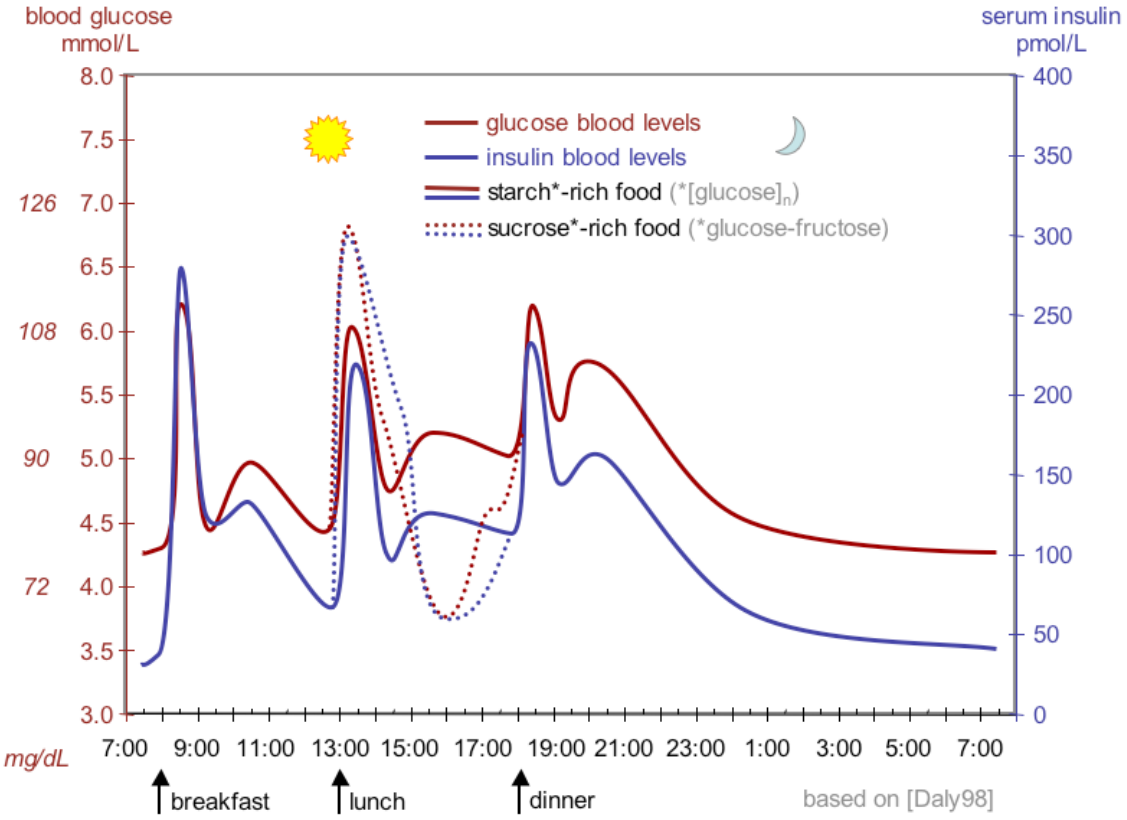
What Triggers Insulin Release?

Different stimuli cause the release of insulin

- Glucose***
- Sulfonylureas
- Arginine

1. Glucose uptake through GLUT2 receptor
2. Glucose is processed by glucokinase/glycolysis—generating ATP
3. Increased ATP closes K^+ channels, depolarizing the membrane and opening Ca^{2+} channels
4. Increased Ca^{2+} initiates exocytosis of insulin vesicles





What does Insulin do once it's released?
Insulin is a ***hormone***—it is released from beta cells and travels to other locations in the body through the bloodstream and binds to cells' **INSULIN RECEPTORS**.

What Does Insulin Do Once It's Released?

Increased lipid synthesis – insulin forces fat cells to take in blood lipids, which are converted to triglycerides

Increased esterification of fatty acids – forces adipose tissue to make fats (i.e., triglycerides) from fatty acid esters

Increased glycogen synthesis – insulin forces storage of glucose in liver (and muscle) cells in the form of glycogen; lowered levels of insulin cause liver cells to convert glycogen to glucose and excrete it into the blood.

Increased amino acid uptake – forces cells to absorb circulating amino acids—DIRECTLY IMPACTS PROTEIN SYNTHESIS and DNA REPLICATION

Not enough insulin—high levels of glucose—***hyperglycemia***
Too much insulin—low levels of gluclose—***hypoglycemia***
The liver and kidneys produce **insulinase**, which degrades circulating insulin