PROTEIN ENGINEERING

LECTURE 01: INTRODUCTION TO PROTEIN

Proteins are one of the most amazing group of molecules in the human body. They are complex combinations of smaller chemical compounds called ***amino acids***. These are like the bricks or building blocks of a building. There are about 22 amino acids. Some can be made in the body from other amino acids, or in other ways. However, at least 8 or 10 of them must be obtained from our diet.

All of the amino acids contain oxygen, hydrogen, carbon and nitrogen. Many also contain sulfur. Like building blocks, millions of these amino acids are chained together, stuck together, and folded over each other to form complicated shapes and structures to create millions of critical body components.

Metaphorical definition.

Proteins are associated with motion, the basic quality of animal life. While plant life is more concerned with sugars and carbohydrates, animal life is more concerned with proteins. (However, all life contains both proteins and carbohydrates.)

WHAT ARE PROTEINS USED FOR IN OUR BODIES?

They are used to make most everything in our bodies. Here are a few of their main uses.
Hormones. Many of the hormones such as insulin and progesterone are proteins.
Oxygen transport. Hemoglobin, a blood protein, carries oxygen to the cells.
Cellular repair. Heat shock proteins help rebuild our cells after stress.
Binding and transport of nutrients. Transferrin, metallothionine, ceruloplasmin and other transport proteins bind to minerals and other things, and carry them throughout the body.
Movement. Muscle protein (such as myoglobin) is responsible for our ability to move.
Holding genetic information. Proteins such as RNA (ribonucleic acid) and DNA (deoxyribonucleic acid) in the nuclei of our cells are responsible for the genetic code.
Structural proteins. Proteins are also essential for the body structure. Bone consists of a protein matrix that fills with calcium and other minerals. Other structural proteins include collagen, cartilage, elastin and keratin that form the skin and other structures.

Enzymes. All enzymes are proteins. Thousands of enzymes facilitate every chemical reaction in the body.
Conversion to fuel. Proteins may also be converted to sugar or fat to be used as fuel for the body. This is not ideal, but it does occur in some people.

PROTEIN ENGINEERING

Protein Functions

Structure

some proteins provide structural support: collagen, hair, crystallins (eyes)

Transport

some proteins are reponsible for the transportation of smaller molecules from one part of the body to another, transport across cell membranes, etc. An example is hemoglobin, which transports oxygen from the lungs to cells throughout the body.

Storage

Myoglobin is an example of a storage protein. Myoglobin stores oxygen in muscles so that during exercise a ready supply of oxygen is available in the muscle tissue.

Hormones

Some hormones are proteins; insulin is an example. Hormones serve as chemical messengers, carrying signals from one part of the body to another.

Signal Transduction

Some proteins mediate the interpretation of cellular signals, like hormones. In order to "receive" the molecular signal, the cell has a "receptor" protein that interacts with the hormone and relays the message to the appropriate cellular "machinery."

Catalysis - Enzymes

Enzymes catalyze the chemical reactions that allow cells to function.

Amino acids - the building blocks of protein

During the process of digestion the proteins in our food are broken down into their constituent amino acids which are in turn absorbed by the blood capillaries and transported to the liver. The amino acids are then synthesized into proteins or stored as fat or glycogen for energy. Each gram of protein produces approximately 4 Calories. Many proteins function as enzymes and others:

- Form the structural framework of various parts of the body - Keratin in skin and hair
- Function as hormones - Insulin
- Serve as antibodies
- Transport vital substances throughout the body - hemoglobin
- Serve as contractile elements in muscle tissues - actin & myosin

Proteins

All proteins are broken down into single amino acids by the digestive enzymes in either the stomach or in the small intestine and then reassembled into the specific new proteins by other enzymes within the body. The cells within the digestive tract can only absorb single amino acids, known as free form, and very small chains of two or three amino acids called peptides. In the same way that glycogen is made up of many molecules of glucose put together, protein is made up and linked together by many amino acids. These amino acids are linked together in long and varied chains by peptide bonds. The types of peptide bonds are:

- Dipeptide: Two amino acids joined by a peptide bond
- Tripeptide: Three amino acids joined by peptide bonds
- Oligopeptide: Four to 10 amino acids joined by peptide bonds
- Polypeptide: More than 10 amino acids going up to as many as 100 amino acids joined by peptide bonds

A combination of more than 100 amino acids joined by peptides finally forms a protein.

Amino Acids

Amino acids are the building blocks of protein. The formation of protein can result in dehydration because water molecules are lost as amino acids combine to form more complex molecules. The body requires 20 different amino acids of which 8 are referred to as essential amino acids which cannot be synthesized by the human body. Animals and plants manufacture proteins which contain these essential amino acids. Non-essential amino acids can be synthesized by body but this does not mean they are unimportant, they are, it is just that the body is capable of producing sufficient to meet the demands for growth and tissue repair. It is therefore important that our diet contains appropriate levels of protein.

Essential Amino Acids

There are 8 essential amino acids. These are the amino acids that the body must obtain on a daily basis through the foods we eat. The essential Amino Acids are:

- Isoleucine, Leucine, Lysine, Methionine, Phenylaianine, Threonine, Tryptophan and Valine

Nonessential Amino Acids

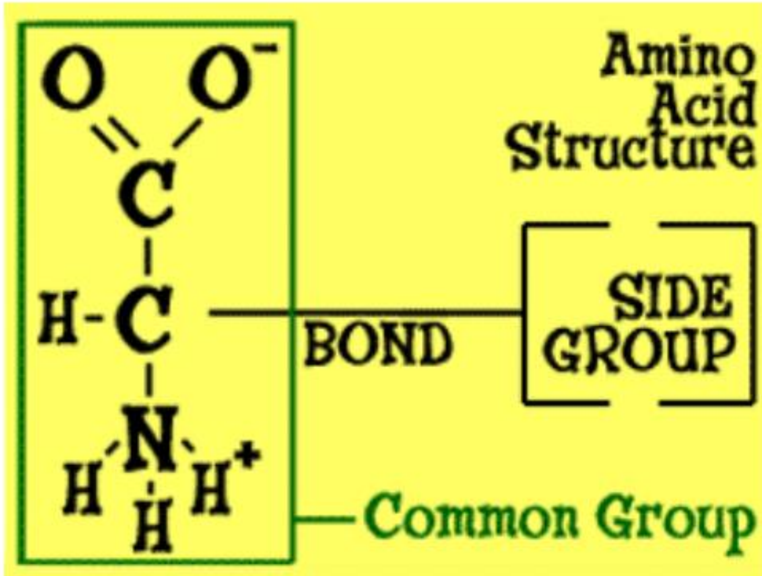
There are 14 nonessential amino acids. These are the amino acids that the body produces itself in the liver through a process known as transamination. Being called nonessential does not mean that these amino acids are unimportant. They form from compounds that are already in the body at a rate that meets the needs of normal growth and tissue repair. The non-essential Amino Acids are:

- Alanine, Arginine, Asparagine, Aspartic acid, Cysteine, Clutamic acid, Clutamine, Glycine, Histidine*, Proline, Serine and Tyrosine.

*Histidine is essential for babies but not for adults.

"What is an amino acid?"

There are more than fifty, and each one of them is a little different. **Amino acids** are used in every cell of your body to build the **proteins** you need to survive. All organisms need some proteins, whether they are used in muscles or as simple structures in the cell membrane. Even though all organisms have differences, they still have one thing in common: the need for basic chemical building blocks.



Amino acids have a two-carbon bond. One of the carbons is part of a group called the**carboxyl group** (COO^-). A carboxyl group is made up of one **carbon** (C) and two **oxygen** (O) atoms. That carboxyl group has a **negative charge**, since it is a carboxylic acid (-COOH) that has lost its**hydrogen** (H) atom. What is left — the carboxyl group — is called a **conjugate base**. The second carbon is connected to the amino group. Amino means there is an NH_2 group bonded to the carbon**atom**. In the image, you see a "+" and a "-". Those positive and negative signs are there because, in amino acids, one hydrogen atom moves to the other end of the molecule. An extra "H" gives you a positive charge.

Making Chains

Even though scientists have discovered over 50 amino acids, only 20 are used to make something called proteins in your body. Of those twenty, nine are defined as **essential**. The other eleven can be synthesized by an adult body. Thousands of combinations of those twenty are used to make all of the proteins in your body. Amino acids bond together to make long chains. Those long chains of amino acids are also called proteins.

Essential Amino Acids: Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Threonine, Tryptophan, and Valine.
Nonessential Amino Acids: Alanine, Asparagine, Aspartic Acid, Glutamic Acid.
Conditional Amino Acids: Arginine (essential in children, not in adults), Cysteine, Glutamine, Glycine, Proline, Serine, and Tyrosine.

each amino acid different from the others. Of the 20 side groups used to make proteins, there are two main groups: **polar** and **non-polar**. These names refer to the way the side groups, sometimes called "R" groups, interact with the environment. Polar amino acids like to adjust themselves in a certain direction. Non-polar amino acids don't really care what's going on around them. The polar and nonpolar chemical traits allow amino acids to point towards water (**hydrophilic**) or away from water (**hydrophobic**). The growing chains can then begin to twist and turn when they are being synthesized.

Classification According to Charge and Polarity of Side Chains (R-Group)

Nonpolar (Hydrophobic) Side Chains

Alanine
Glycine
Leucine
Valine
Isoleucine
Phenylalanine
Tryptophan
Methionine
Proline

Uncharged Polar (Hydrophilic) Side Chains

Asparagine
Glutamine
Cysteine
Serine
Threonine
Tyrosine

Acid Side Chains

Aspartic Acid
Glutamic Acid

Basic Side Chains

Arginine
Histidine
Lysine

PROTEIN ENGINEERING

Physical properties of amino acids:

Colorless, crystals, soluble in water, insoluble in ether. All amino acids(except glycine) are optically active. Amphoteric (react as acidic and basic), (NH2 and COOH group).

Titration curve of amino acids

Introduction and principle: Amino acids are organic compounds containing both an amino group (- NH3) and a carboxyl group (- COOH).

Each of these amino acids has a side chain with distinctive chemical properties. Amino acids are grouping into five main classes based on the properties of their R group. These classes are non-polar aliphatic R groups, aromatic R groups, polar uncharged R groups, positively charged (basic) R group, and negatively charged (acidic) R groups. When amino acid is dissolved in water, it exists in solution as the dipolar ion, or "zwitterion".

+ H3N CH COO – R

A zwitterion can act as either an acid (proton donor) or a base (proton acceptor).

Substances having this dual nature are amphoteric. So, amino acids in aqueous solution are exist predominantly in isoelectric form. The characteristic pH at which the net electric charge is zero is called the isoelectric point (pI). So, an amino acid has a net negative charge at any pH above its pI, and has a net positive charge at any pH below its pI. Each amino acid has its own pI value. The Ionizable groups of amino acids act as weak acids or bases, giving off or taking on protons when the pH is altered. Simply, common amino acids are weak polyprotic acids. Titration curves are produced by monitoring the pH of given volume of a sample solution after successive addition of acid or alkali. Titration curves are usually plots of pH against the volume of titrant added or more correctly against the number of equivalents added per mole of the sample. Upon titration of amino acid with acid, it acts as a base, and upon titration with base, it acts as an acid. In this experiment, the amino acid represents either the conjugate base A- or the conjugate acid HA from in the Henderson-Hasselbalch equation, depending on the titration. For particular Ionizable group

For particular Ionizable group

$$pH = pKa + log \frac{unprotonated \quad form\,(conjugate \quad base)}{protonated \quad form\,(conjugate \quad acid)}$$

pKa for particular Ionizable group is equal the pH at which the Ionizable group is at its best buffering capacity. Considering applying the Henderson-Hasselbalch equation to the titration of glycine with acid and base, glycine has two Ionizable groups: a carboxyl group and an amino group, with pKa values of 2.34 and 9.6 respectively. In water at pH 6, glycine exists as a dipolar ion, or zwitterion, in which the carboxyl group is unprotonated (−COO -) and the amino group is protonated to give the substituted ammonium ion (−NH3 +).

Addition of the acid to the solution lowers the pH rapidly at first and then more slowly as the buffering action of the carboxyl is exerted. At pH 2.34 the pKa is reached, one-half the acid has been consumed, and the carboxyl group is half ionized and the most effective as a buffer. Titration of the amino group with base follows a similar curve into the alkali region. The intersection between the titration of the carboxyl group and the titration of the amino group describes in this case the point at which glycine has no net charge, and is called the isoelectric point (pI). For simple amino acids such as glycine, the pI is an average of the pKa 's of the carboxyl (2.34) and ammonium (9.6) groups. Thus the pI for glycine is calculated to be: (2.34 + 9.6) / 2 = 5.97 . If additional acidic or basic groups are present as side-chain functions, the pI is the average of the pKa's of the two most similar acids or bases. Most amino acids contain carboxyl and amino groups having pKa values similar to those of glycine. In addition to these groups, some amino acids contain other Ionizable groups, which introduce other "steps" or pKa values into their titration

curves.

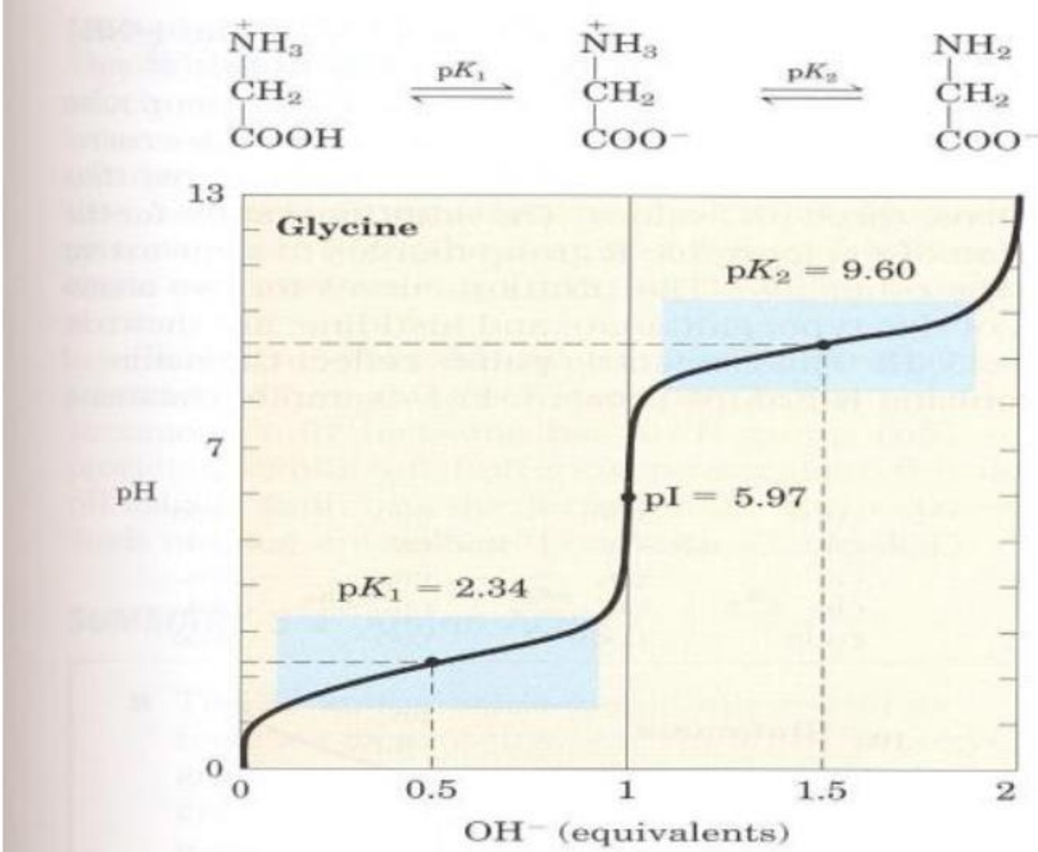
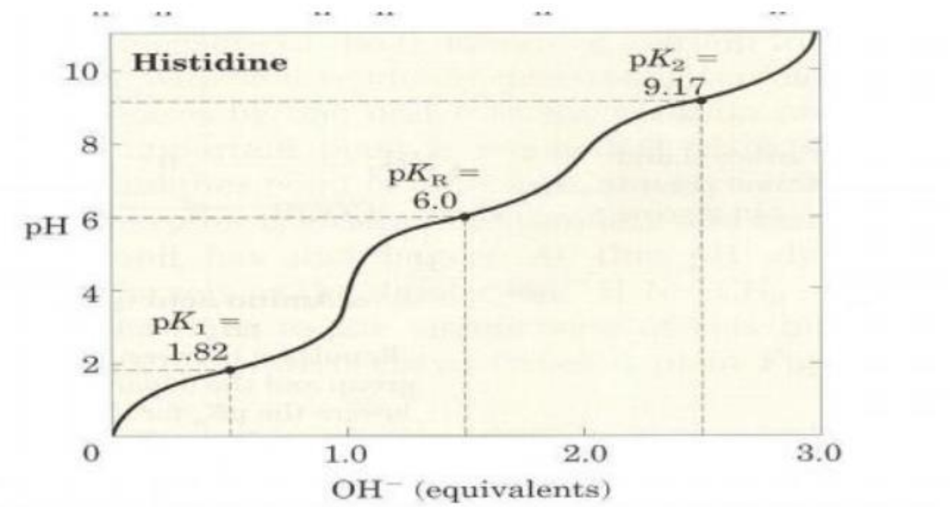
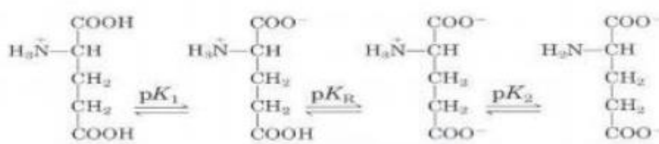


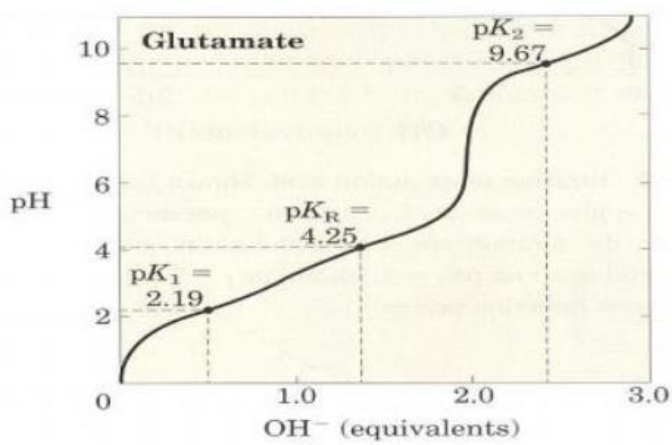
FIGURE 3–10 Titration of an amino acid. Shown here is the titration curve of 0.1 M glycine at 25 °C. The ionic species predominating at key points in the titration are shown above the graph. The shaded boxes, centered at about $pK_1 = 2.34$ and $pK_2 = 9.60$, indicate the regions of greatest buffering power.



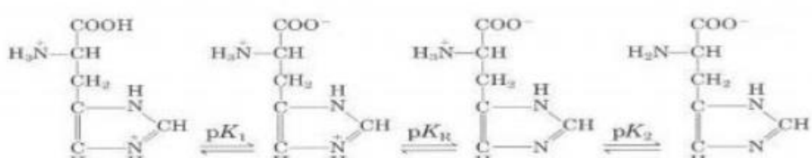
(b)

FIGURE 3–12 Titration curves for **(a)** glutamate and **(b)** histidine. The pK_a of the R group is designated here as pK_R.





(a)



PROTEIN ENGINEERING

Based on Nutritional requirement, amino acids can be divided into 3 types.

1. Essential Amino acids (EAA)
2. Non-Essential Amino acids (NEAA)
3. Semi Essential Amino acids (SEAA)

1. Essential Amino acids (EAA):

Some of the amino acids doesn't synthesize in the human body. It should be supplied through diet. They are required for proper growth and maintenance of the individual.

Eg:

MATT VIL PHLy

(or)

PVT TIM HALL

M= Methionine **A**=Arginine **T**=Threonine **T**=Tryptophan **V**=Valine

I=Isoleucine **L**=Leucine **P**=Phenylalanine **H**=Histidine **L**=Lysine

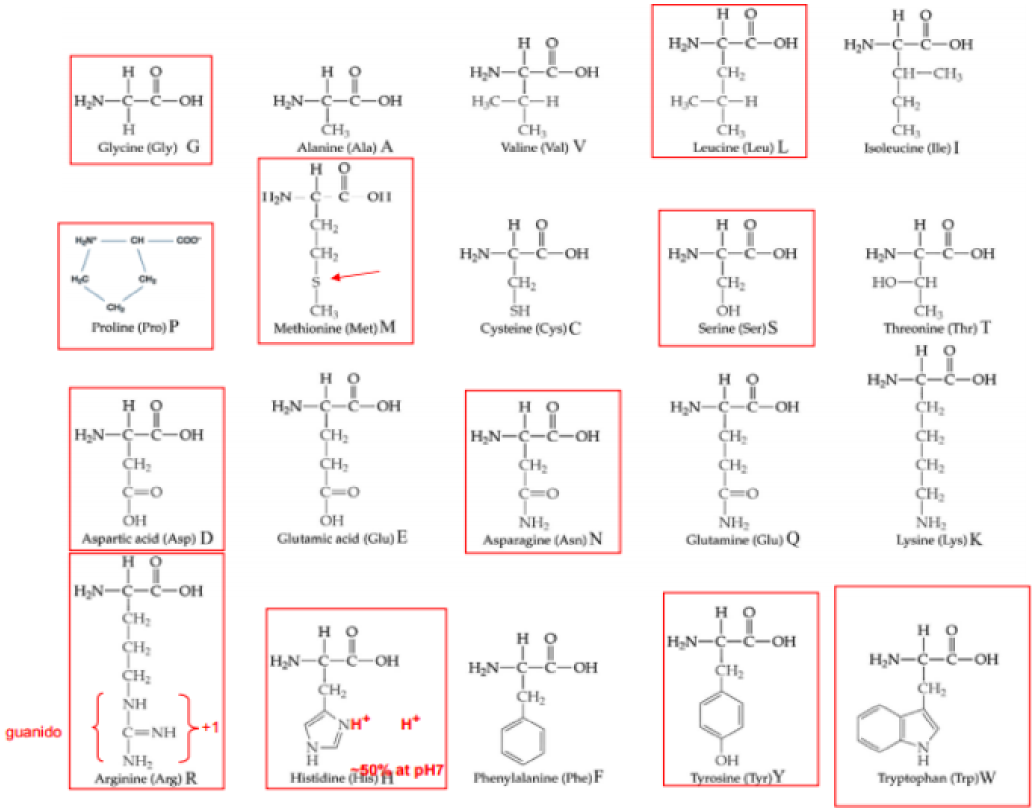
2. Non-Essential Amino acids (NEAA):

The body can synthesize about 10 amino acids to meet the biological needs, hence they need not be consumed in the diet.

Eg: Gly, Ala, Ser, Cys, Asp, Asn, Glu, Gln, Tyr and Pro.

3. Semi-Essential Amino acid:

Histidine and *Arginine* are semi essential amino acids. Growing children require them in food. But they are not essential for the adult individual.



Classification of Amino Acids

(based on polarity)

- Hydrophobic / non-polar R group: Glycine, alanine, valine, leucine, isoleucine, methionine, proline, phenylalanine, tryptophan
- Polar R group (net charge 0 at pH 7.4): Serine, threonine, cysteine, tyrosine, asparagine, glutamine, histidine
- Polar R group (Charged ion at pH 7.4): aspartate, glutamate, lysine, arginine

Insulin

Background

Insulin is a small protein (more specifically a hormone) that lowers blood sugar. Insulin was discovered by Fredrick Banting, Charles Best, J.J.R. Macleod and James Collip (Strakoshc, 2005). At that time, it was know that diabetes was a disease that caused people who were affected to have difficulty controlling their blood sugar levels. Upon their discovery of insulin, a way to battle diabetes was found.

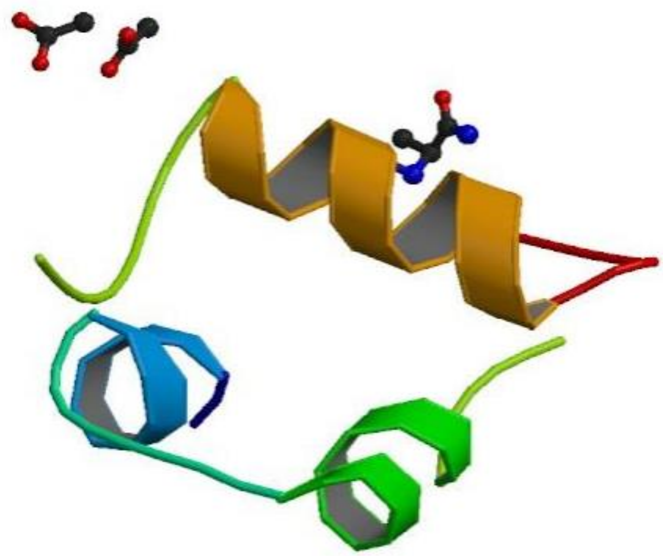


Figure 1. Stucture of hyperactive human insulin analogue courtesy of Protein Data Bank. The bottom strand is the A subunit and the top strand is the B subunit.

Structure of Insulin

The primary structure of insulin is made from two polypeptide chains named subunit A and B. Subunit A consists of 21 amino acids, whereas subunit B consists of 30 amino acids. These chains are connected by two disulfide bridges as seen in **figure 2**. Insulin also forms quaternary structure by creating diamers using hydrogen bonds and hexamers by bonding with two zinc ions (Bowen, 1999). Insulin The final product of insulin only consists of 51 amino acids so it is quite small compared to other proteins. Insulin's small size allows it to be a ligand for other proteins appropriately named insulin receptors.

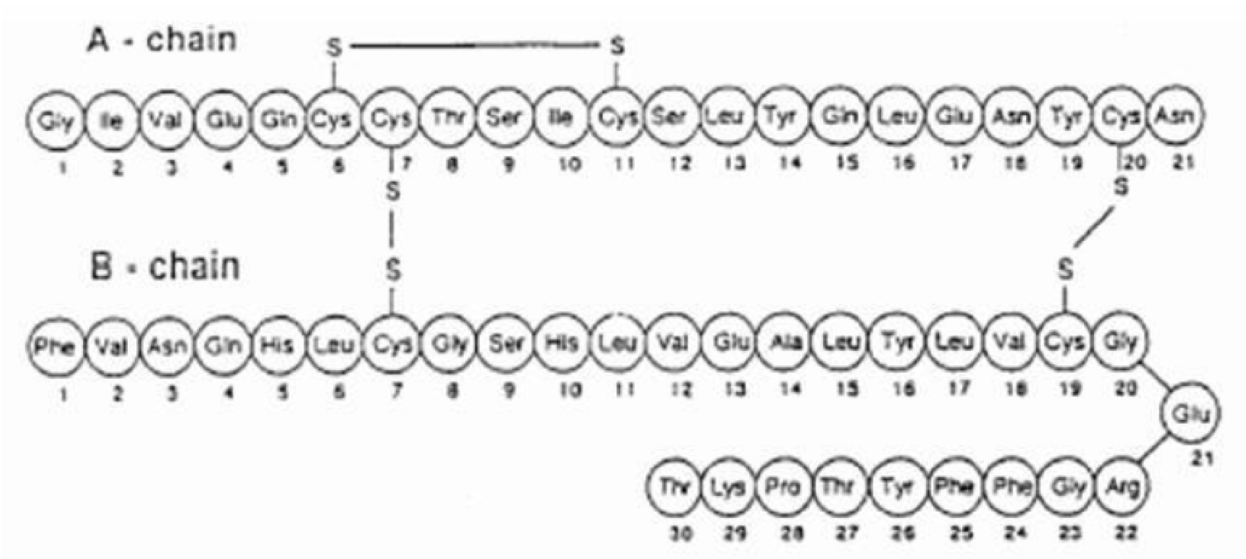
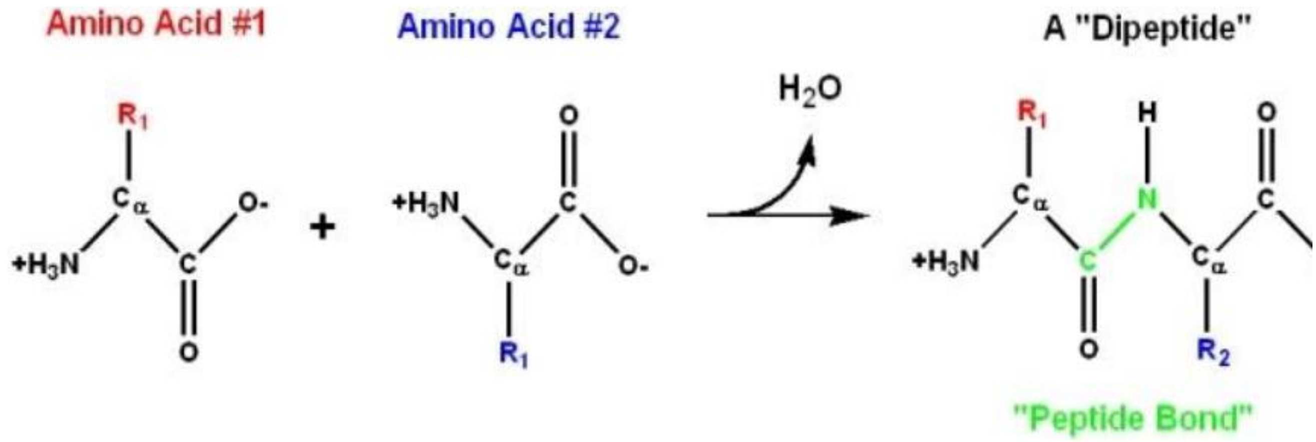


Figure 2. The primary structure of human insulin. As you can see, there are 21 amino acids in subunit A and 30 amino acids in subunit B.

Amino acids are the monomeric building blocks of protein molecules

- The chemical bond that joins two amino acids involves an amide bond (also known as a "**peptide bond**") between the carbonyl group of the first amino acid and the amino group of the second amino acid.
- Bond formation involves the subsequent release of a water molecule

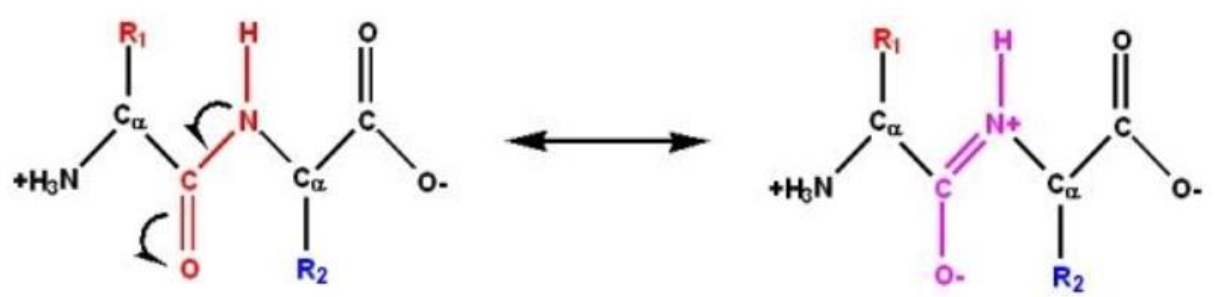
PROTEIN ENGINEERING



The joining of two amino acids results in the formation of a "dipeptide"

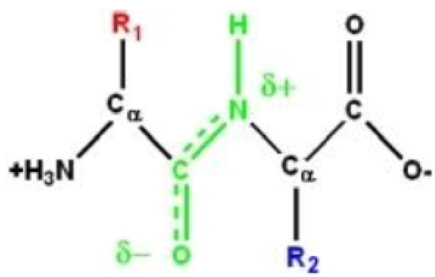
- The carboxyl group of the dipeptide (now known as the carboxyl terminus) is available for formation of a peptide bond with another amino acid
- Three amino acids connected by peptide bonds is referred to as a "tripeptide", next would come "tetrapeptide" and so on. "Polypeptide" or "Peptide" is a general term for any such polymeric molecule of non-defined length, and a "Protein" is generally "a long polypeptide". There is no clear distinction of when a polypeptide might become a protein. General, when people talk about "Peptides" they mean short polymers of 2-50 amino acids, and "Proteins" would be larger than 50 amino acids in length

The peptide bond (as drawn above) looks like it is a single bond, and therefore, free to rotate. However, the peptide bond has **partial double bond character** due to a resonance structure:

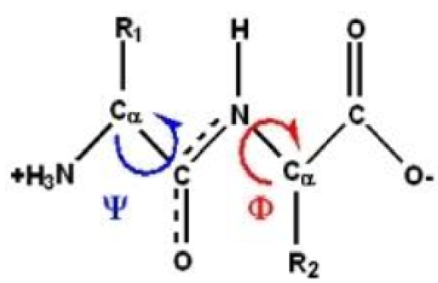


The resonance structures have the following consequences:

- Rotation around the peptide bond is restricted (88 kJ/mol energy required to rotate), therefore, it can be considered **rigid**
- The carbonyl oxygen is positioned ***trans*** to the amide hydrogen
- The amide nitrogen valence electron geometry has some sp^3 character, and is therefore intermediate between tetrahedral and trigonal planar. The **O-C-N-H atoms in the peptide bond are usually considered to be co-planar**.
- There is a partial positive charge on the amide nitrogen (+0.28), and a partial negative charge on the carbonyl oxygen (-0.28)



- While the peptide bond is rigid, there is a single bond between the C_α(1) and C atoms and the N and C(2) which are free to rotate. Thus, while the C_α (1) - C (O) - N (H) - C_α (2) atoms are all planar, this planar peptide bond has two degrees of rotational freedom.
- These two rotation angles are known by the greek letters **Φ** (phi) and **Ψ** (psi). Ψ is the rotation angle about the C_α (1) - C bond and **Φ** is the rotation angle about the N - C_α (2) bond. These are also referred to as "main chain" angles.



- **It is these two degrees of rotational freedom that allows polypeptides to fold up into unique conformations**