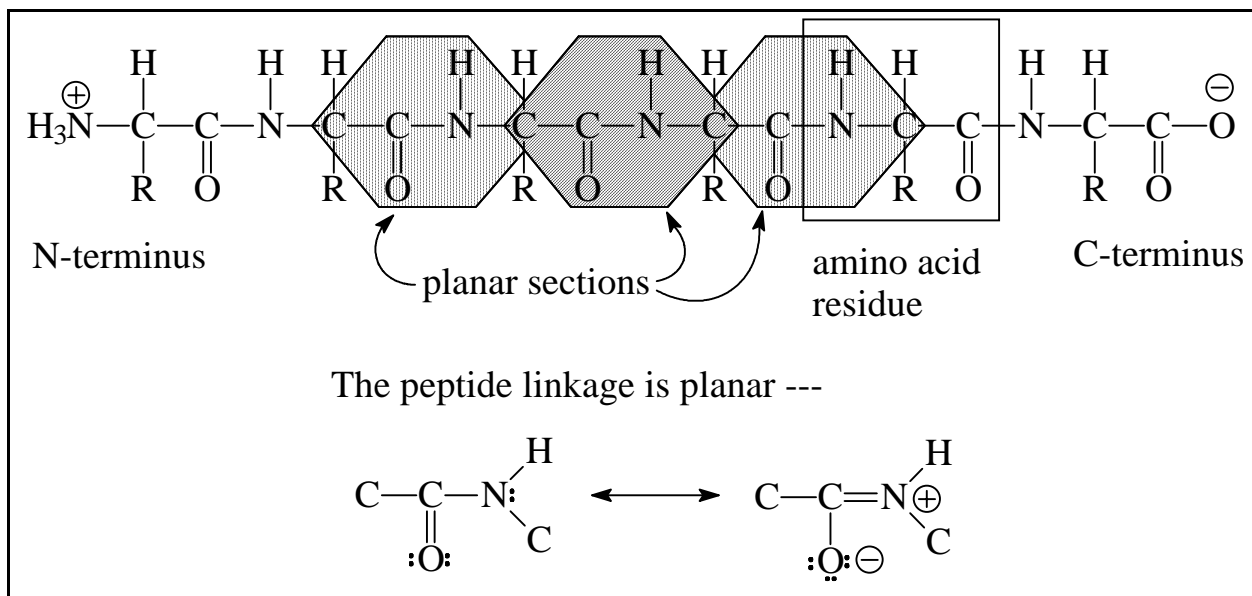


## Polypeptides and Proteins

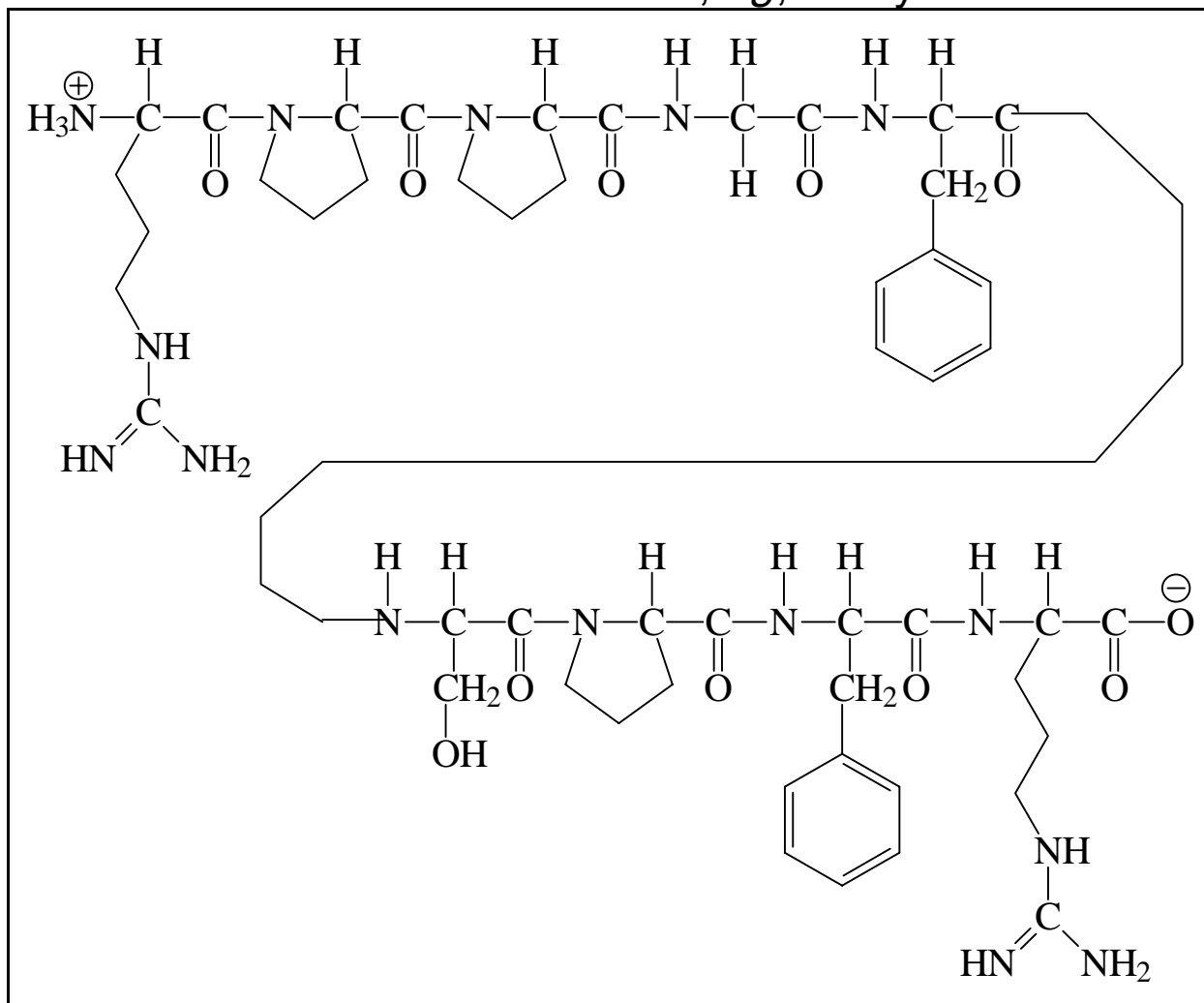
These molecules are composed, at least in part, of chains of amino acids. Each amino acid is joined to the next one through an amide or *peptide* bond from the carbonyl carbon of one amino acid “residue” to the  $\alpha$ -amino group of the next. At one end of the chain there will be a free or protonated amino group: the N-terminus. At the other end there will be a free carboxyl or carboxylate group: the C-terminus. By convention the N-terminus is drawn at the left end, as shown below in a generic hexapeptide (6 amino acid residues).



Note that the backbone chain of amino acid residues consists of a series of flat “plates” that “join” at the  $\alpha$ -carbons of the amino acid residues. It is the tetrahedral  $\alpha$ -carbons that allow the chain to bend and have some

flexibility. The plates are flat because the amide (or peptide) C-N bond is planar owing to its double bond character as shown by resonance in the figure above.

Because peptide molecules are large a shorthand method is used to indicate their structures, *eg*, bradykinin —



As you can see, it is simpler to use the three letter abbreviations for the amino acid residues than it is to draw the usual structure:

Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg.

Note that the N-terminus is at the left end and the C-terminus is at the right end; it is critical that we hew to this convention when writing structures using the three letter abbreviations —

Arg-Phe-Pro-Ser-Phe-Gly-Pro-Pro-Arg would be a different molecule.

### Determination of a polypeptide structure —

1. Which amino acid residues are present?
2. How many of each?
3. What is their sequence?

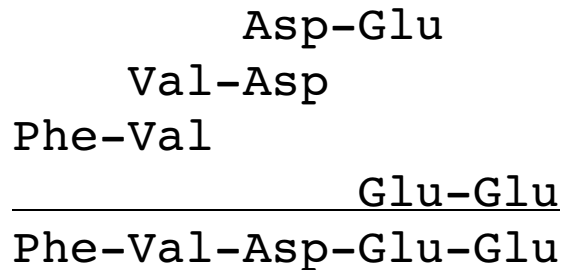
(1) may be determined by complete acid hydrolysis and qualitative analysis of the hydrolysate.

(2) may be determined by complete acid hydrolysis, quantitative analysis of the hydrolysate, and the determination of the MW of the polypeptide.

Consider a polypeptide, MW = 637. Complete hydrolysis of 637 mg (1.00 mmole) gives: Asp, 133 mg (1.00 mmole); Phe, 165 mg (1.00 mmole); Val, 117 mg (1.00 mmole); Glu, 294 mg (2.00 mmole). Therefore, polypeptide = Asp-Glu<sub>2</sub>-Phe-Val, not in sequence.

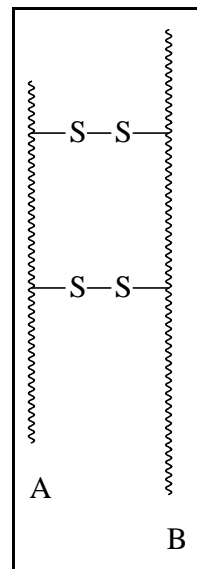
The sequence (3) may be determined by partial hydrolysis. Suppose the following units (among others) are found after partial hydrolysis of the above polypeptide: Asp-Glu, Val-Asp, Phe-Val, Glu-Glu.

We can extrapolate to the overall sequence as follows:

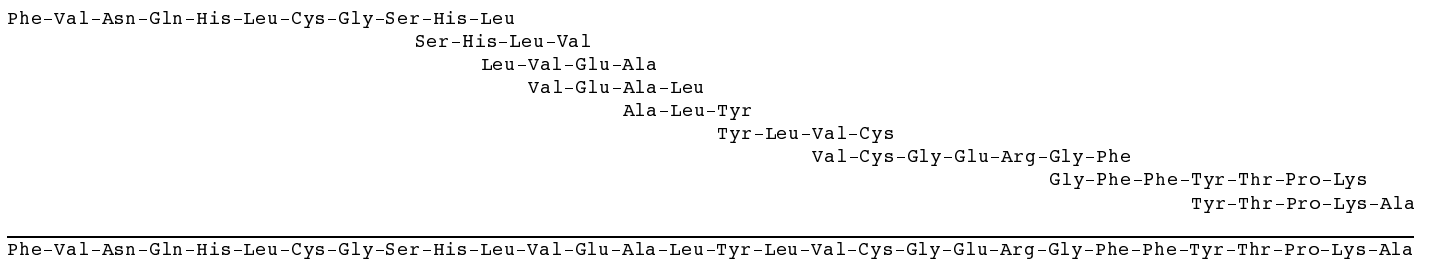


Therefore, the sequence is Phe-Val-Asp-Glu-Glu.

Another --- real life --- example would be the determination of the sequence of amino acid residues in the B-chain of insulin. [Insulin consists of two polypeptide chains, A and B, held together by two disulfide bridges, C-S-S-C. A disulfide bridge results from oxidative coupling of the S-H groups of two cysteine residues.]



First the disulfide bridges would be cleaved reductively, regenerating cysteine SH groups, and separating the two chains. Then the chain of interest, B in this case, would be partially hydrolyzed. The result obtained when this was done follows:

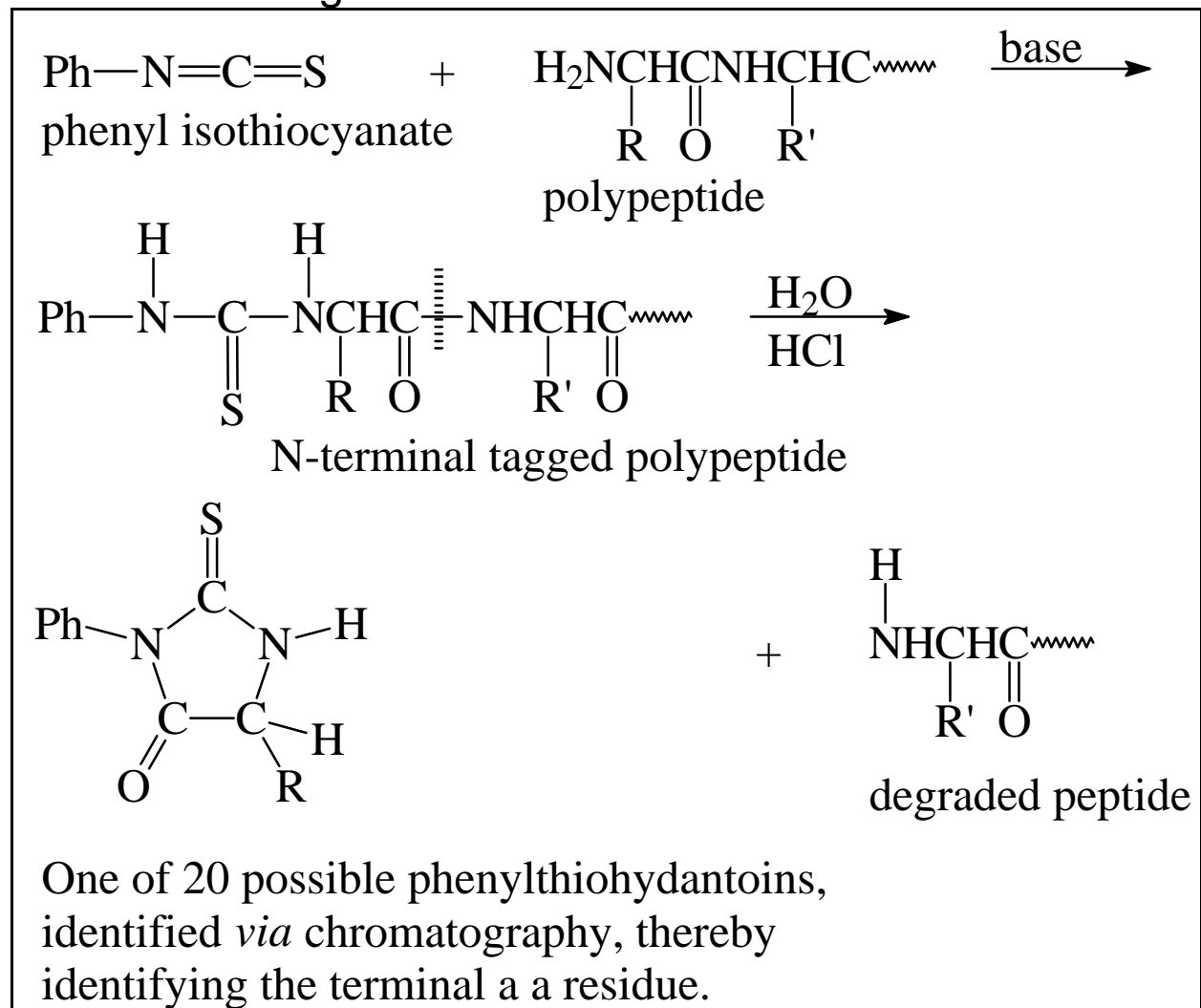


But how is the sequence in the fragments determined?

Answer: Terminal residue analysis.

Some reagents will react specifically with one terminus or the other, thus enabling one to determine the nature of the terminal residue. In some cases, the terminal residue can be cleaved off and the remainder of the polypeptide remains (more or less) intact. The remaining polypeptide can be subjected to terminal residue analysis, *etc.*

The Edman Degradation —



*Polypeptides*: Smaller than proteins.

MW < ~7-10,000 or < ~50 amino acid residues.

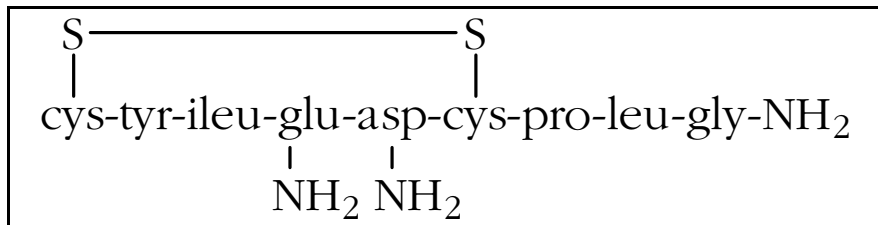
Polypeptides are usually flexible: they usually do not have the *secondary* structure characteristic of proteins. (Nor do they have *tertiary* or *quaternary* structure.)

Examples:

bradykinin: helps regulate blood pressure; pain.

methionine enkephalin: Tyr-Gly-Gly-Phe-Met,  
morphine like analgesic.

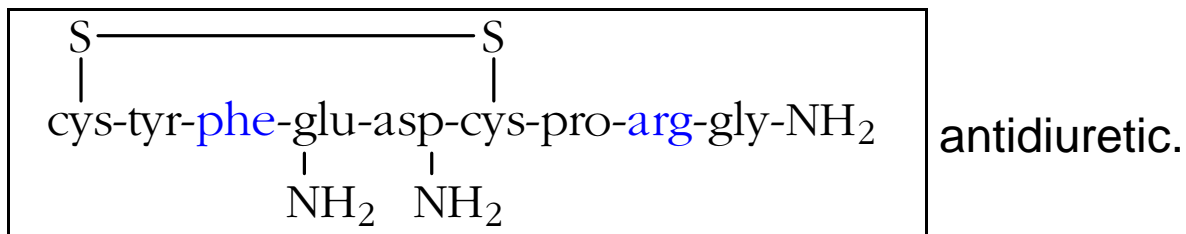
oxytocin:



causes uterine contraction and lactation;

not species specific—  
labor induced via chicken oxytocin.

vasopressin (note the structural similarity to oxytocin):



*Proteins* —

### Classification by Gross Chemical Structure

*Simple*: consist only of amino acid residues.

*Conjugated*: consist of polypeptide portion(s) and non-polypeptide portion(s) called *prosthetic group(s)*, eg lipoproteins, glycoproteins, nucleoproteins, metalloproteins.

### Classification by Gross Morphology

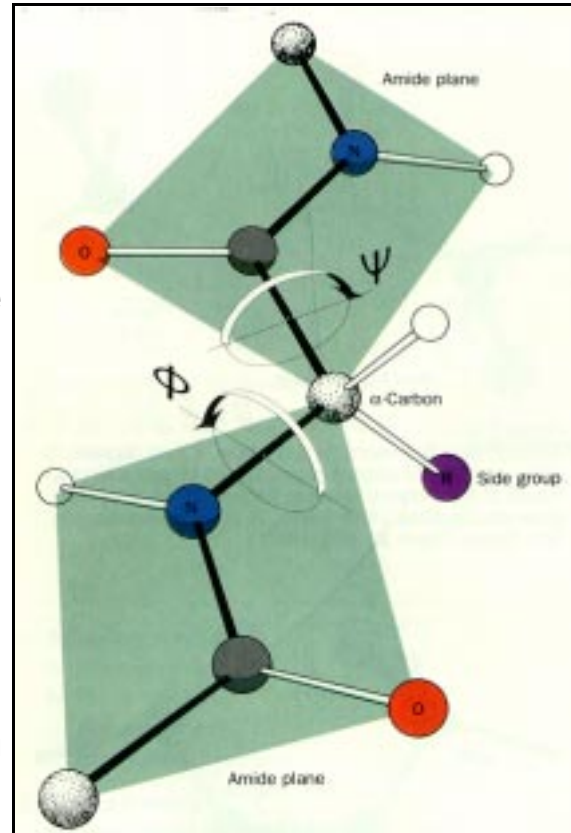
*Fibrous*: insoluble in water, resist hydrolysis, eg, sericine (silk fibroin), hard keratin (hair, nails, claws), soft keratin (skin), myosin (muscle), collagen (connective tissue, bones).

*Globular*: soluble or colloidal in water, eg, hemoglobin, enzymes, antibodies, some hormones.

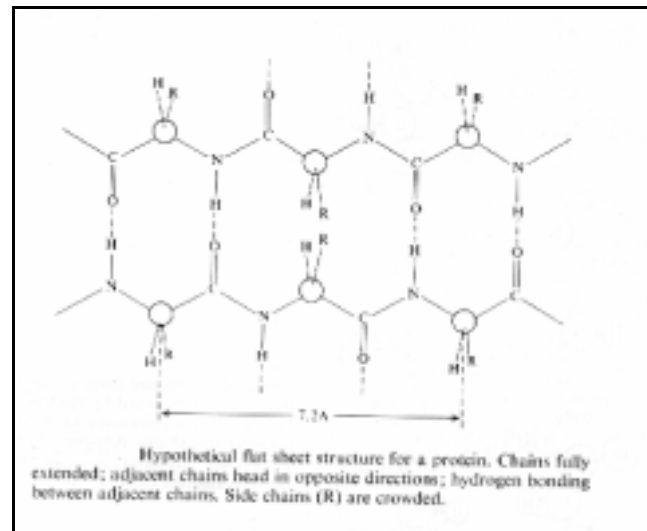
## Protein Structure

*Primary:* sequence of amino acid residues.

*Secondary:* way in which segments of polypeptide backbone are oriented in space; depends on planarity of peptide bond and, often, H-bonding. The only places in the protein chain where there is torsional flexibility is at the  $\alpha$ -carbons: the angles  $\Phi$  and  $\Psi$  can change causing one plane to rotate relative to the other.

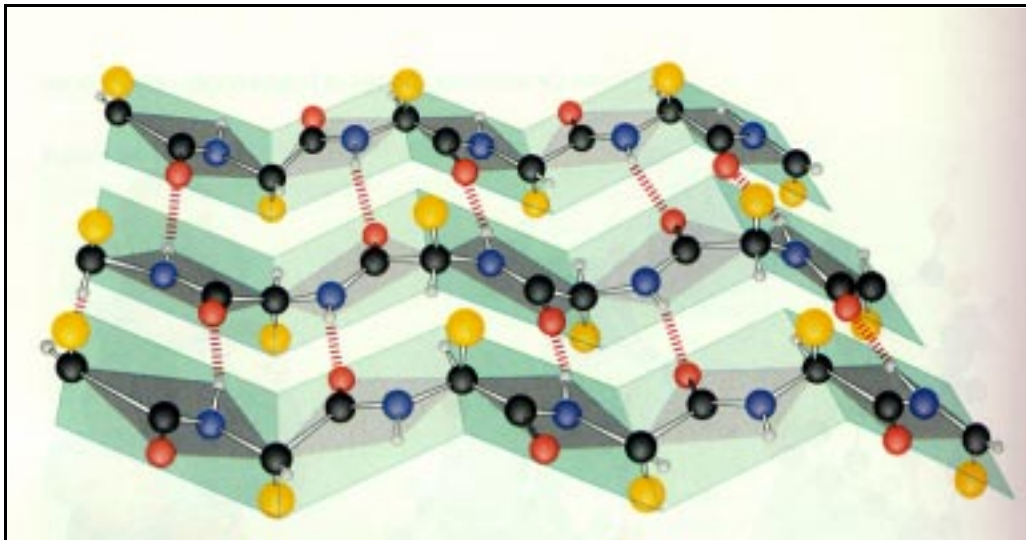


One could imagine that all of the planar "plates" in the protein backbone lie in the same plane. This would produce a flat strand. By snaking back and forth other parts of the strand could lie down alongside the first in parallel (or antiparallel) fashion creating a *flat sheet*. Or other protein molecules could lie down parallel (or antiparallel) to the first.





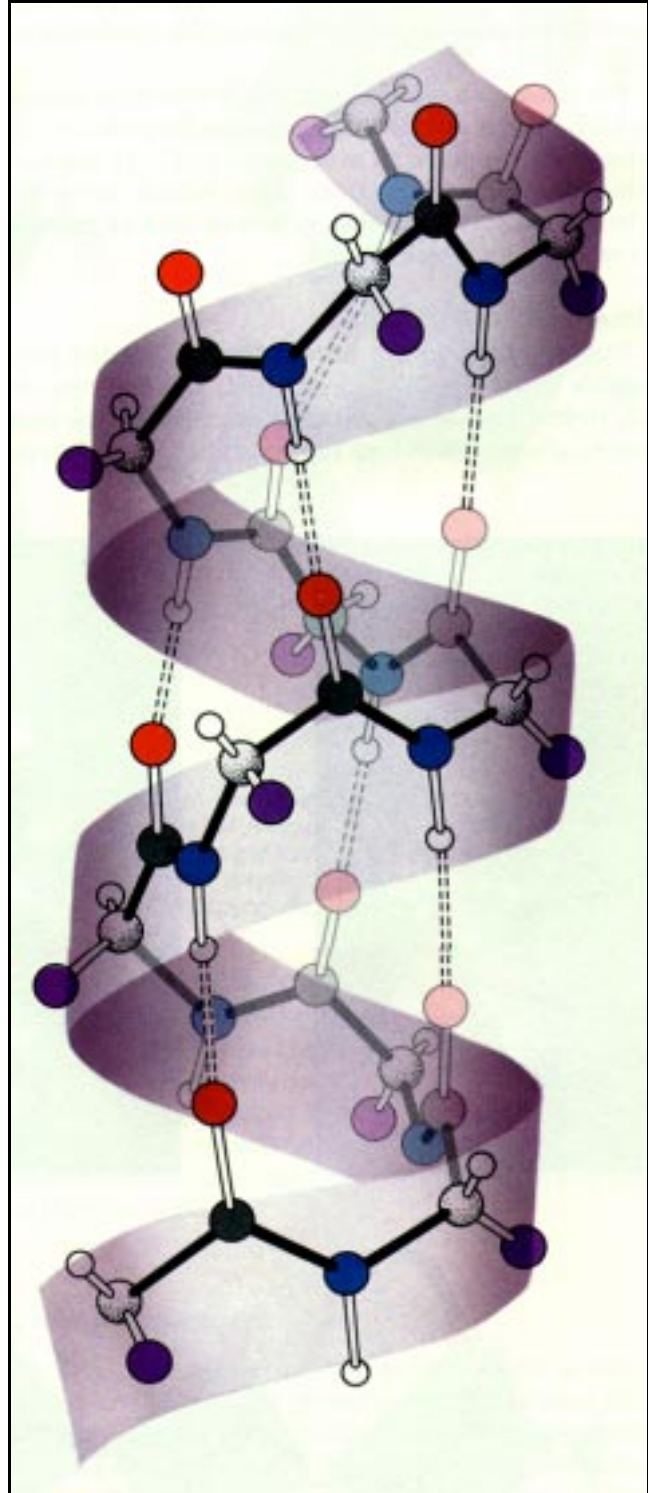
This flat sheet would allow hydrogen bonding between adjacent strands of the protein (favorable for formation of the sheet). However, the R groups sterically interfere with each other in this setup. This hypothetical structure has not been observed in nature.



Antiparallel  $\beta$ -pleated sheet. Black = C, white = H, blue = N, Red = O, Yellow = R (side chain).

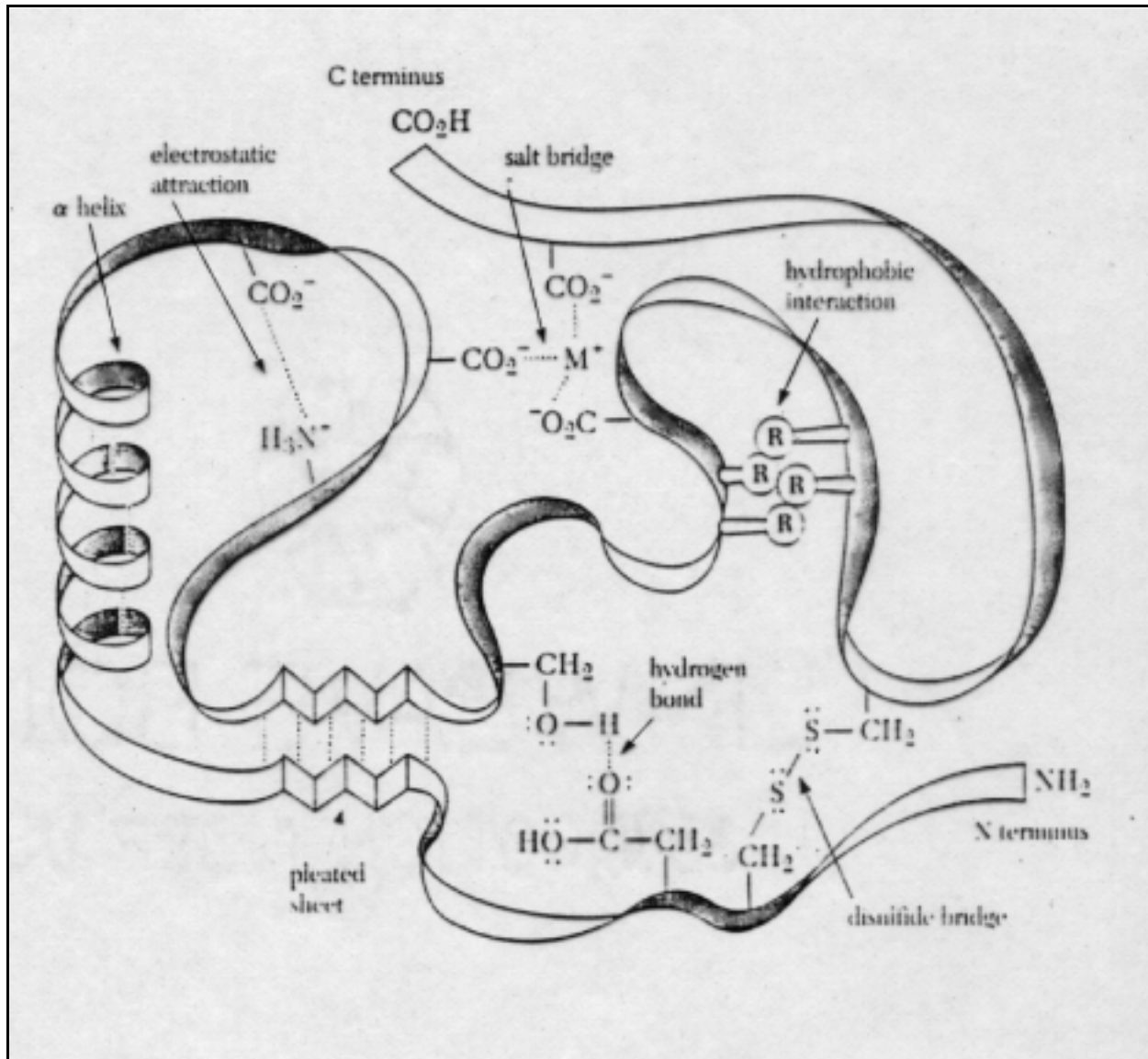
The steric interference between the R groups in the flat sheet can be reduced if the sheet becomes “pleated”. In this arrangement the angle between adjacent plates is not zero. Hydrogen bonding is still possible and now the R groups can splay away from each other to some extent. This setup works if the R groups are small. This is the case for silk fibroin which does form the pleated sheet structure.

Another geometry which is possible is the helix. This arrangement allows hydrogen bonding in directions roughly parallel to the helical axis (dotted bonds) and the R groups are oriented outward from the helix, away from each other. This happens in nature. Several varieties of helix are known; the most common form is known as the  $\alpha$ -helix. It is a right-handed helix (if it were a wood screw, turning it clockwise would cause it to be driven into the wood). [The mirror image of a right-handed helix is a left handed one. The mirror images of the L-amino acids in the right-handed helix would be D-amino acids. The chirality of the amino acid residues determines the chirality of the helix!]



$\alpha$ -Helix. Black = C, white = H, blue = N, red = O, purple = R.

*Tertiary*: way in which protein as a whole is twisted in three dimensions; may depend on disulfide bonds, H-bonds, electrostatic attractions, hydrophobic interactions.



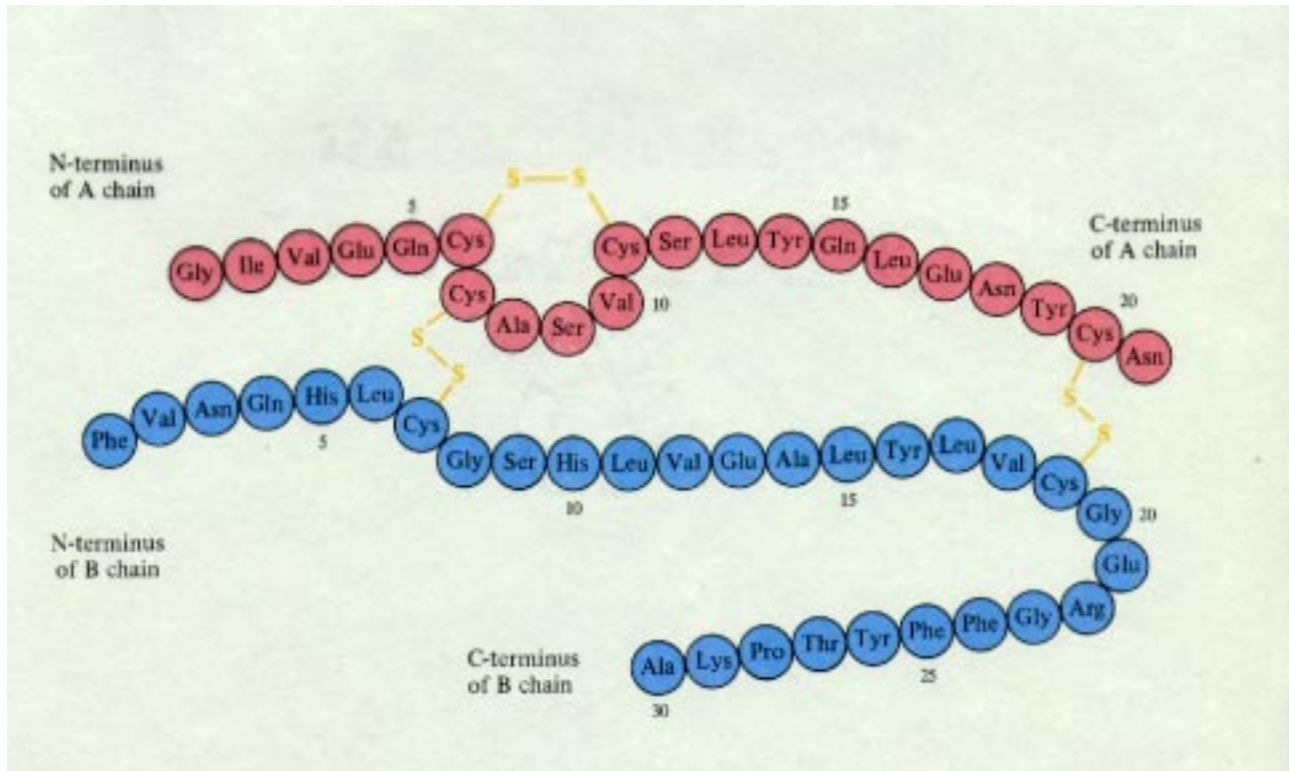
The principal factors affecting protein tertiary structure.

*Quaternary*: aggregate structures held together by nonbonded interactions or occasionally disulfide links.

Let's look at some proteins!

## The Primary Structure of Bovine Insulin —

Insulin is a protein hormone produced in the pancreas. It is essential for the regulation of carbohydrate metabolism.



### Species                      Amino Acid Residues

	A-8	A-9	A-10	B-30
Cow	Ala	Ser	Val	Ala
Sheep	Ala	Gly	Val	Ala
Horse	Thr	Gly	Ileu	Ala
Human	Thr	Ser	Ileu	Thr
Pig	Thr	Ser	Ileu	Ala
Whale	Thr	Ser	Ileu	Ala

I have never seen a discussion of the secondary or tertiary structure of insulin. It is a small protein so its secondary structure may be somewhat flexible, although one would expect it to have conformations of lower and higher energy with one or several of lower energy predominating. With regard to tertiary structure the disulfide links would certainly be important.

### Collagen —

Collagen is found in all multicellular animals and is the most abundant protein of vertebrates. It is a fibrous protein, the fibers of which have great tensile strength. It is found in the connective tissue of teeth, bones, ligaments, tendons and cartilage.

### Primary structure —

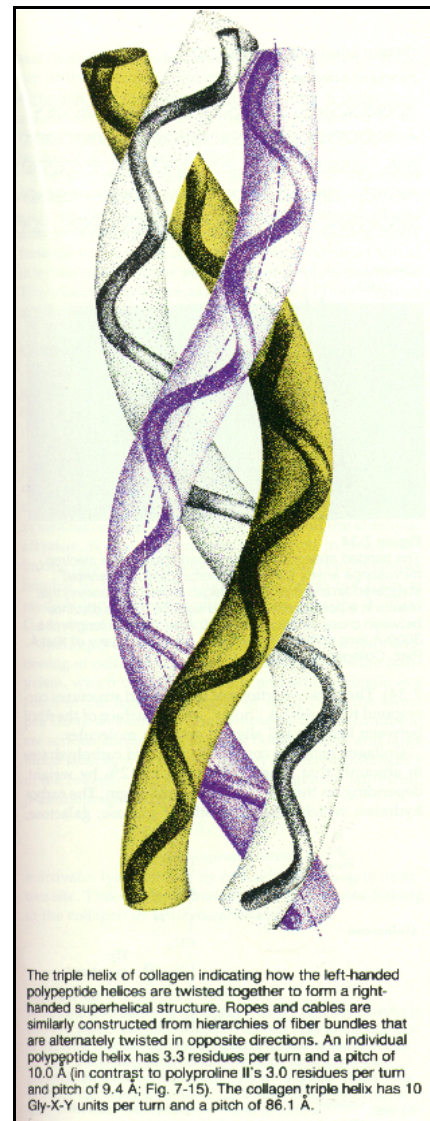
Collagen is rich in glycine (R=H). This enables the polypeptide strands to fit tightly to other strands without steric difficulties. The other major components are proline and hydroxyproline.

### Secondary structure —

A left-handed helix.

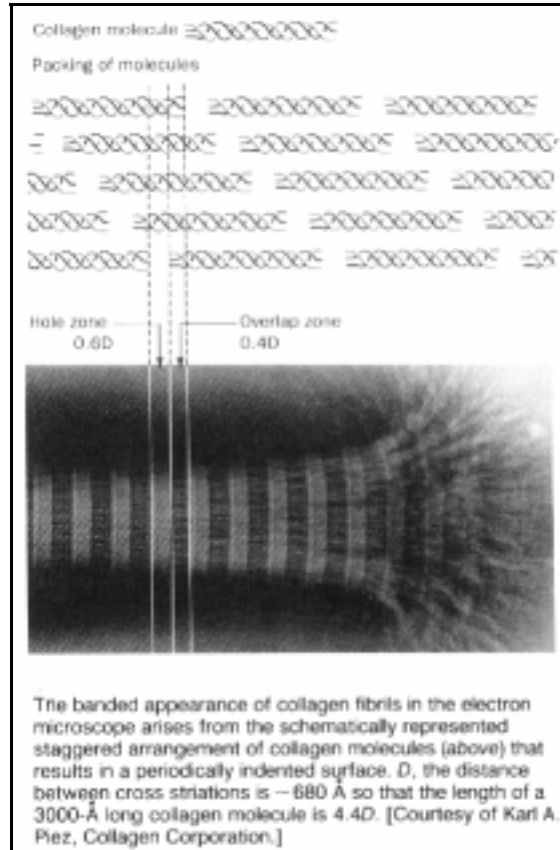
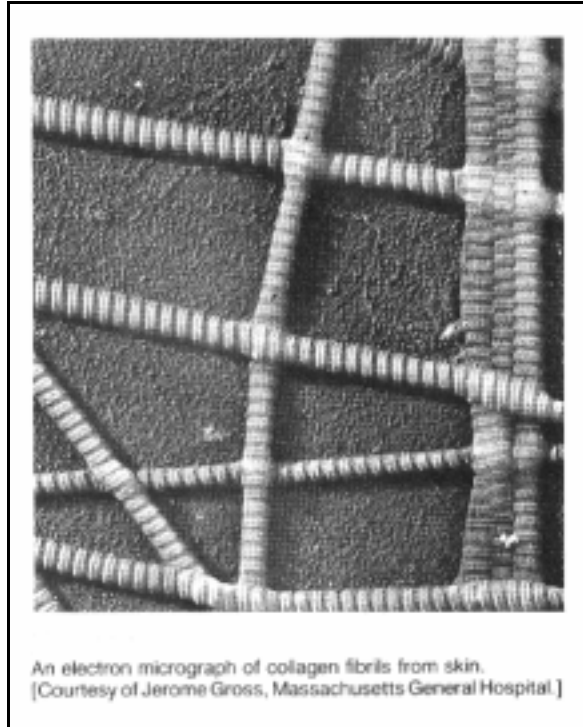
### Tertiary structure —

A right-handed helix.



## Quaternary structure —

A cabling of three right-handed helicies. However, this continues to higher levels of organization as we can see in the following photomicrographs.

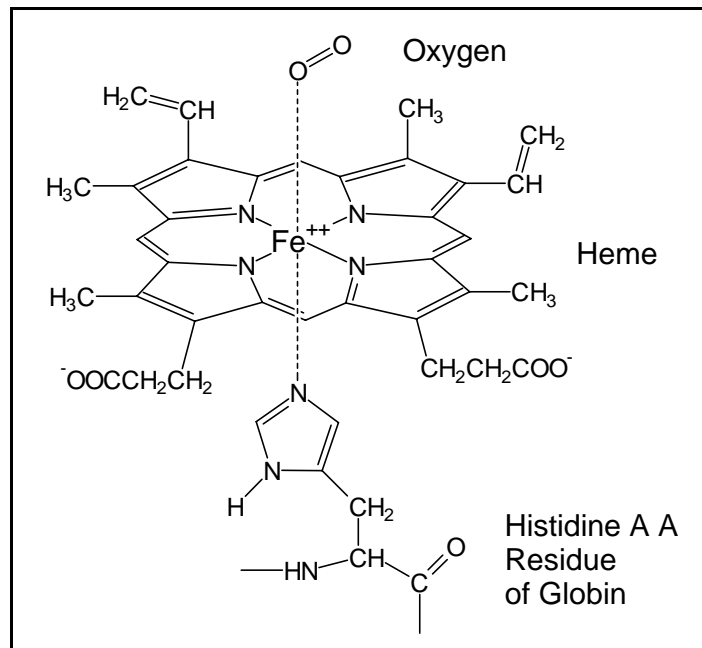


## Hemoglobin —

Hemoglobin is a complex, conjugated, metalloprotein. It is what makes red blood cells red. It carries oxygen throughout the body from the lungs.

Each hemoglobin protein consists of four units, two  $\alpha$ -units and two  $\beta$ -units. This is its quaternary structure.

Each unit consists of a globin and a heme. The globin is a chain of amino acid residues. The heme is a porphyrin containing one iron atom. The iron atom carries the oxygen molecule. The presence of the non-polypeptide heme makes hemoglobin a conjugated protein.



Each type of unit,  $\alpha$  and  $\beta$ , has its own particular shape composed of various twists and turns. This is its tertiary structure.

Within each unit the polypeptide backbone adopts certain conformations going down the chain – in many places this is a helix. This is the secondary structure of the protein.

Finally, the sequence of the amino acids in the  $\alpha$ - (141 residues) and  $\beta$ -globins (146 residues) is the primary structure.



Oxygenated hemoglobin. The amino acid residues are shown by number in each globin, starting with #1 at the N-terminus. This image was drawn looking down the C-2 axis. Ignore the large grey arrows.



Over 100 abnormal varieties of hemoglobin have been reported in humans. The abnormalities are in the globins. One of these is the replacement of the glutamic acid residue at position #6 in the  $\beta$ -globins by a valine residue:

Normal: Val-His-Leu-Thr-Pro-Glu-Lys ---> 146

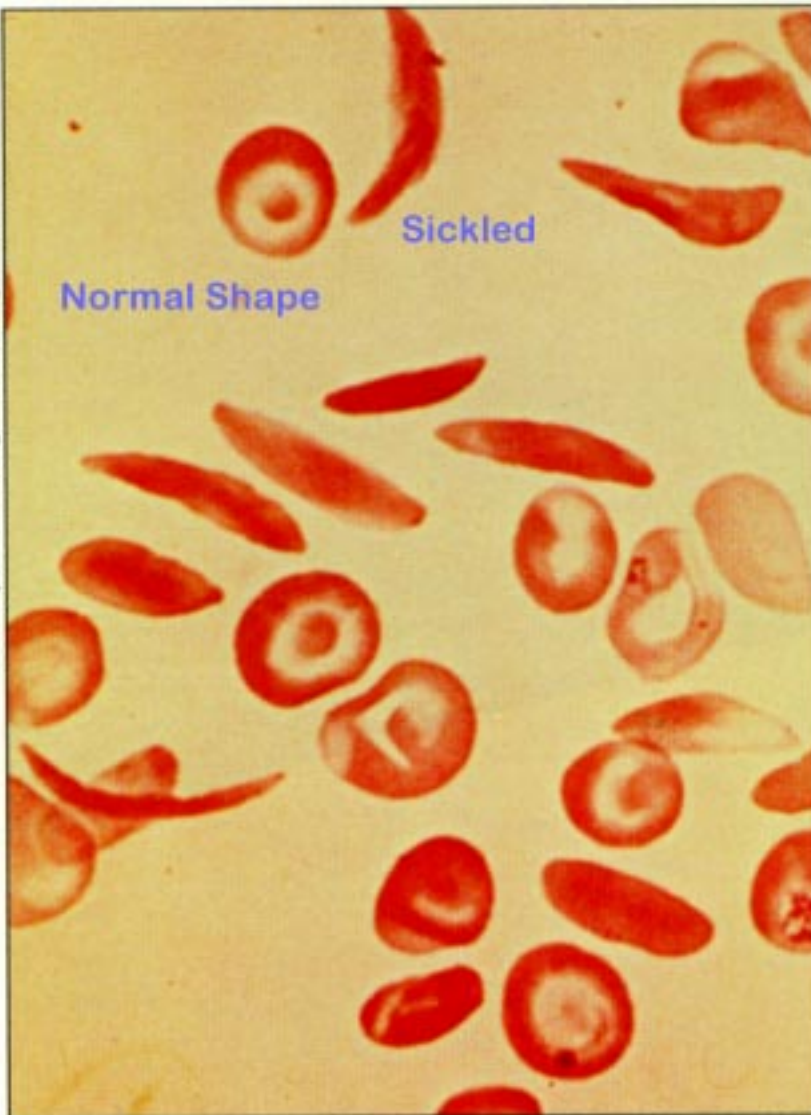
Abnormal: Val-His-Leu-Thr-Pro-Val-Lys ---> 146

This change causes the hemes associated with the  $\beta$ -globins to be unable to carry oxygen. As a result cells become oxygen starved.

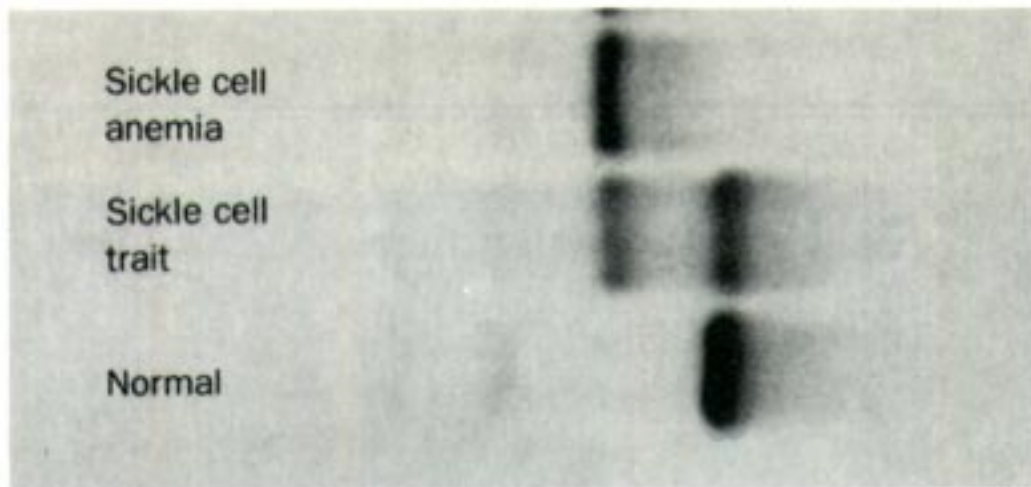
This alteration also causes the hemoglobin to crystallize and this, in turn, causes some red blood cells to stiffen and deform into a crescent shape. These cells can block blood flow through capillaries. Furthermore, the body recognizes these cells as abnormal and destroys them, resulting in anemia. (Life of abnormal cells ~ 16 days; normal ~ 120 days.) This condition, along with the shape of the cells gives rise to the name of this genetic disease: sickle cell anemia.

*After releasing their load of oxygen, some of the red blood cells of sickle cell patients become stiff and sickled, as shown in this photo (amplified 1,500 times). The misshapen cells often get stuck in small blood vessels, causing extreme pain and damage.*

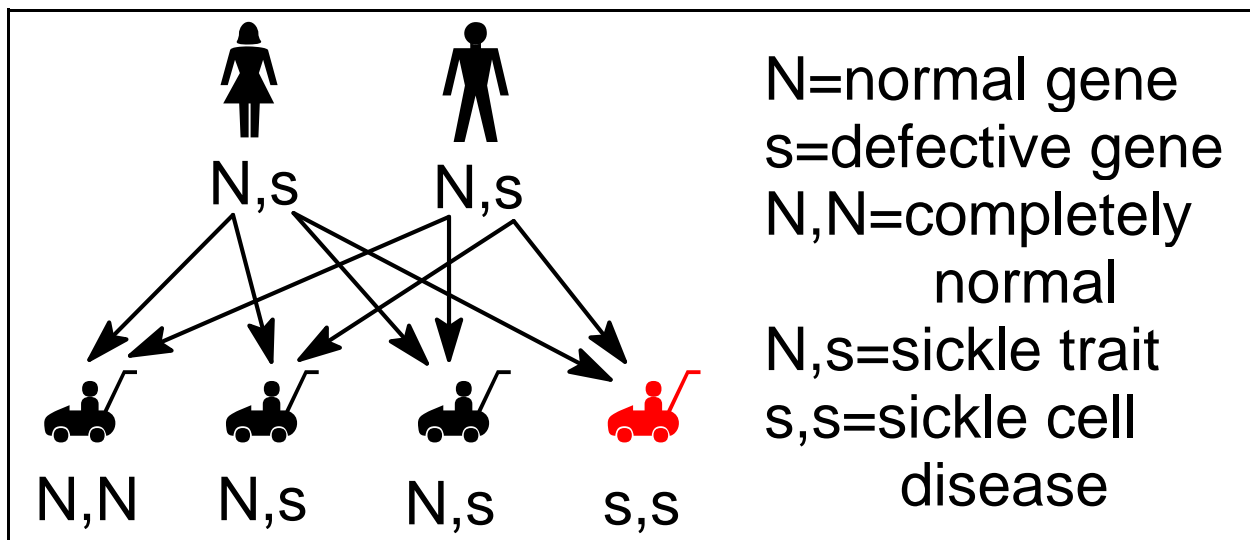
Courtesy of G.P. Rodgers, C.T. Noguchi, A.N. Schechter, NIH



People who have inherited two sickle cell genes (one from mother, one from father) will have the disease. Those who have inherited one normal and one defective gene will have some abnormal hemoglobin, but usually live normal lives; they are said to have the trait, and can pass the gene on to their children.



The electrophoretic pattern of hemoglobins from normal individuals and those with the sickle-cell trait and sickle-cell anemia. [From Montgomery, R., Dryer, R. L., Conway, T. W., and Spector, A. A., *Biochemistry, A Case Oriented Approach* (4th ed.), p. 87. Copyright © 1983 C. V. Mosby Company, Inc.]



Why hasn't this gene disappeared from the gene pool?  
Wouldn't it be removed by (Darwinian) natural selection?

Who gets the s gene?

People who live, or whose ancestors lived, in areas where malaria is, or was, prevalent.

Malaria?

The malarial parasite spends part of its life cycle in red blood cells. People who are N,s are somewhat resistant to malaria. Thus, the presence of malaria confers some advantage to the s gene. This gene did not survive long-term in populations where malaria is not endemic. The map below shows the relationship between malaria and the presence of the sickle gene.

# Map of Misery

