Organic Chemistry III

Mohammad Jafarzadeh

Faculty of Chemistry, Razi University

Organic Chemistry, *Structure and Function* **(7th edition)**

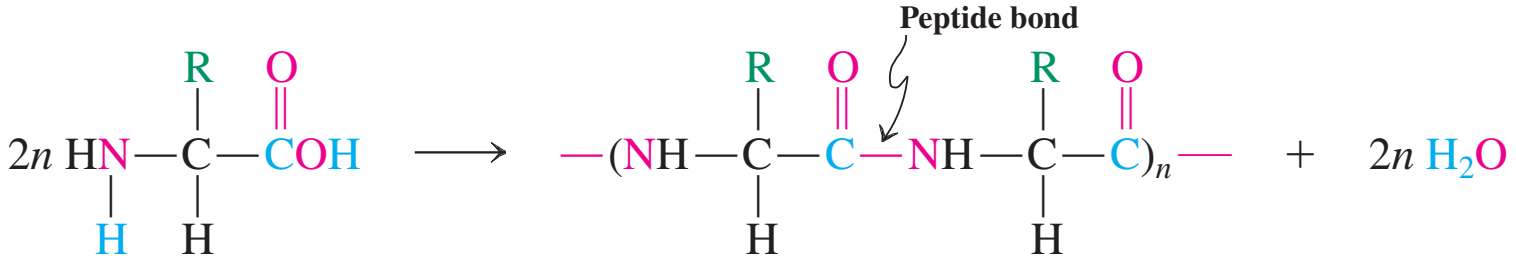
By *P. Vollhardt* and *N. Schore*, Elsevier, 2014

26-4 PEPTIDES AND PROTEINS: AMINO ACID OLIGOMERS AND POLYMERS

Amino acids are very versatile biologically because they can be polymerized. 2-Amino acids are the monomer units in **polypeptides**. Long polypeptide chains are called **proteins** (somewhat arbitrarily defined as >50 amino acids) and are one of the major constituents of biological structures. Proteins serve an enormous variety of biological functions; these functions are often facilitated by the twists and folds of the component chains.

The polymer forms by repeated reaction of the carboxylic acid function of one amino acid with the amine group in another to make a chain of amides. The amide linkage joining amino acids is also called a **peptide bond.**

The oligomers formed by linking amino acids in this way are called **peptides. T**wo amino acids give rise to a **dipeptide,** three to a **tripeptide,** and so forth. The individual amino acid units forming the peptide are referred to as **residues.**



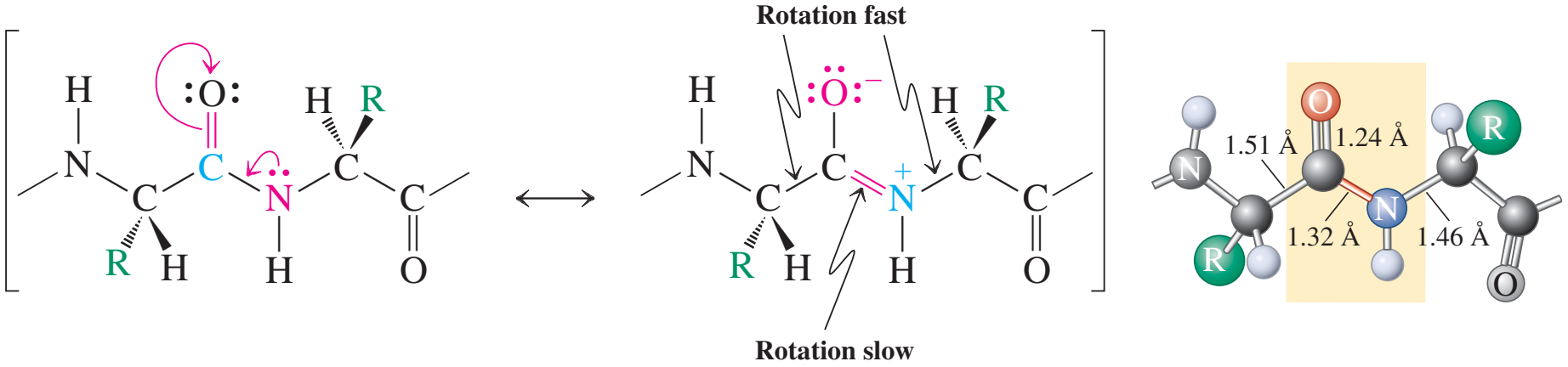
2-Amino acid
(α-Amino acid)

Polyamino acid
(Polypeptide)

The peptide bond is fairly rigid at room temperature and planar, a result of the conjugation of the amide nitrogen lone electron pair with the carbonyl group. The N–H hydrogen is almost always located trans to the carbonyl oxygen, and rotation about the C–N bond is slow because the C–N bond has partial double-bond character (1.32 Å), between the length of a pure C–N single bond (1.47 Å) and that of a C–N double bond (1.27 Å).

On the other hand, the bonds adjacent to the amide function enjoy free rotation. Thus, polypeptides are relatively rigid but nevertheless sufficiently mobile to adopt a variety of conformations. Hence, they may fold in many different ways. Most biological activity is due to such folded arrangements; straight chains are usually inactive.

Planarity Induced by Resonance in the Peptide Bond



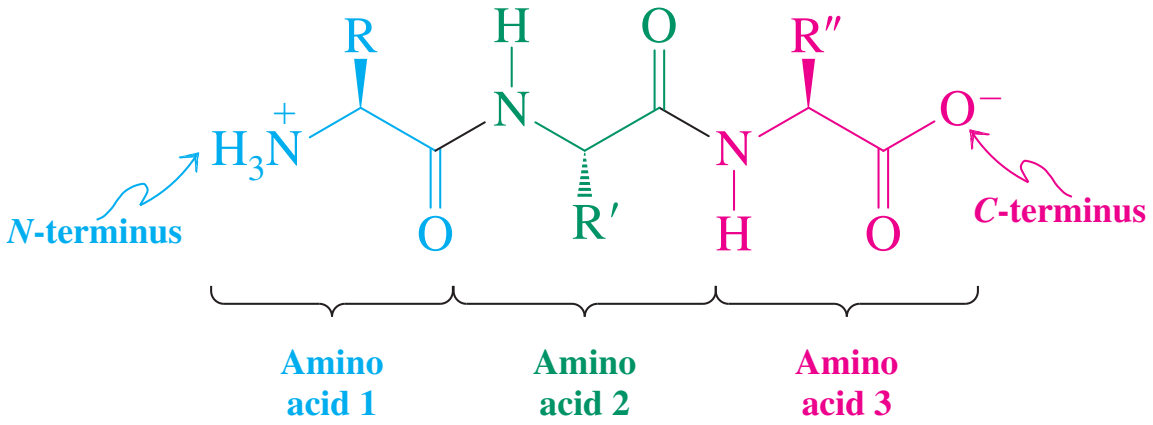
Polypeptides are characterized by their sequence of amino acid residues

In drawing a polypeptide chain, the **amino end,** or ***N*-terminal amino acid,** is placed at the left. The **carboxy end,** or ***C*-terminal amino acid,** appears at the right.

The configuration at the C2 stereocenters is usually presumed to be *S*.

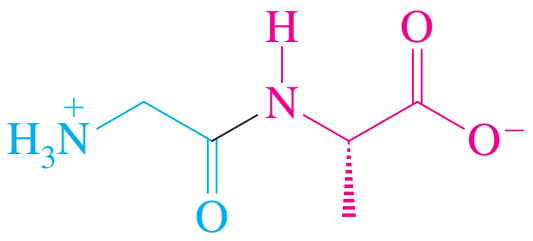
The chain incorporating the amide (peptide) bonds is called the **main chain,** the substituents R, R', and so forth, are the **side chains.**

How to Draw the Structure of a Tripeptide

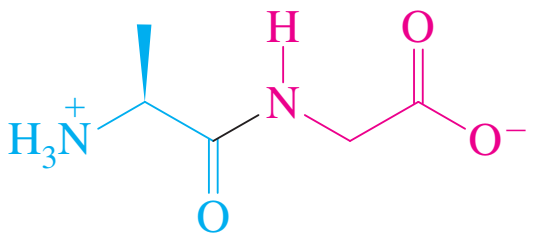


The naming of peptides is straightforward. Starting from the amino end, the names of the individual residues are simply connected in sequence, each regarded as a substituent to the next amino acid, ending with the *C*-terminal residue.

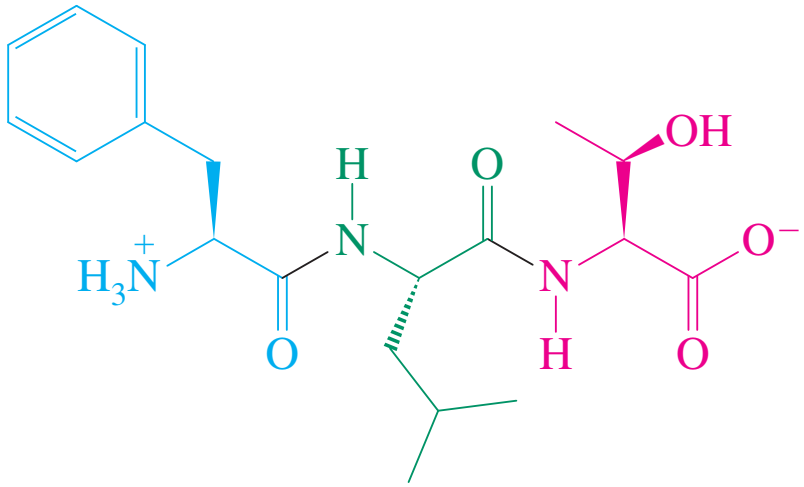
Because this procedure rapidly becomes cumbersome, the three-letter abbreviations listed in Table 26-1 are used for larger peptides.



Glycylalanine
Gly-Ala



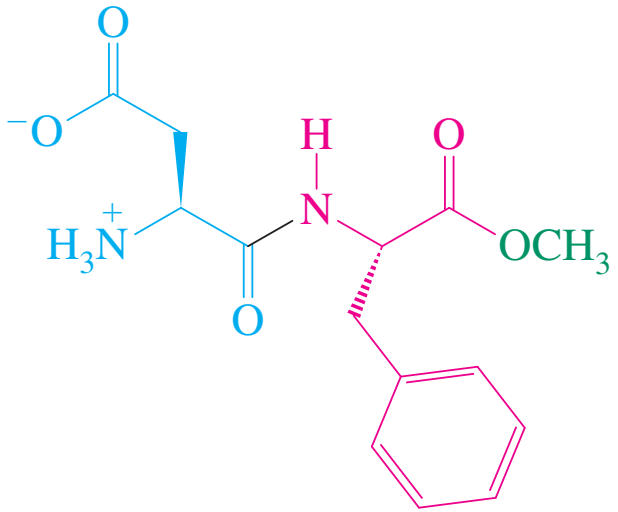
Alanylglycine
Ala-Gly



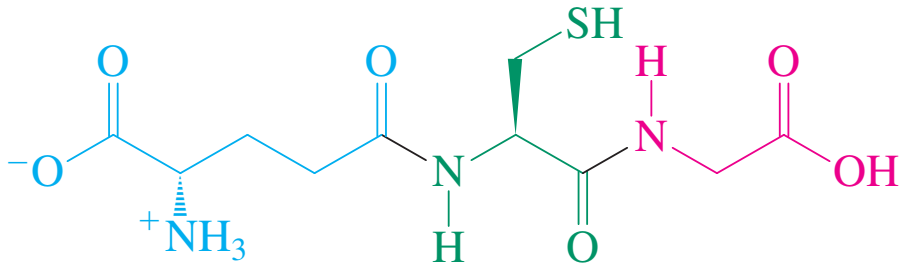
Phenylalanylleucylthreonine
Phe-Leu-Thr

A dipeptide ester, aspartame, is a low-calorie artificial sweetener (NutraSweet). In the three-letter notation, the ester end is denoted by OCH_3.

Glutathione, a tripeptide, is found in all living cells, and in particularly high concentrations in the lens of the eye. It is unusual in that its glutamic acid residue is linked at the γ-carboxy group (denoted γ-Glu) to the rest of the peptide. It functions as a biological reducing agent by being readily oxidized enzymatically at the cysteine mercapto unit to the disulfide-bridged dimer.



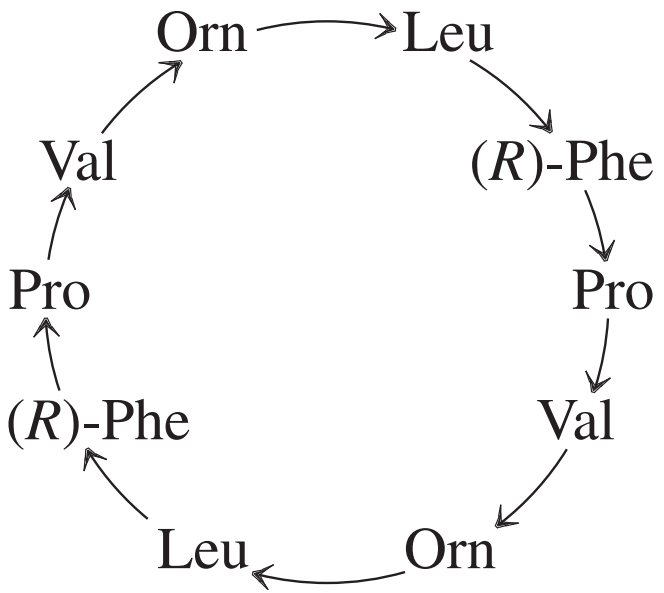
Aspartylphenylalanine methyl ester
Asp-Phe-OCH_3
(Aspartame)



γ-Glutamylcysteinylglycine
γ-Glu-Cys-Gly
(Glutathione)

Gramicidin S is a cyclic peptide, topical antibiotic, constructed out of two identical pentapeptides that have been joined head to tail.

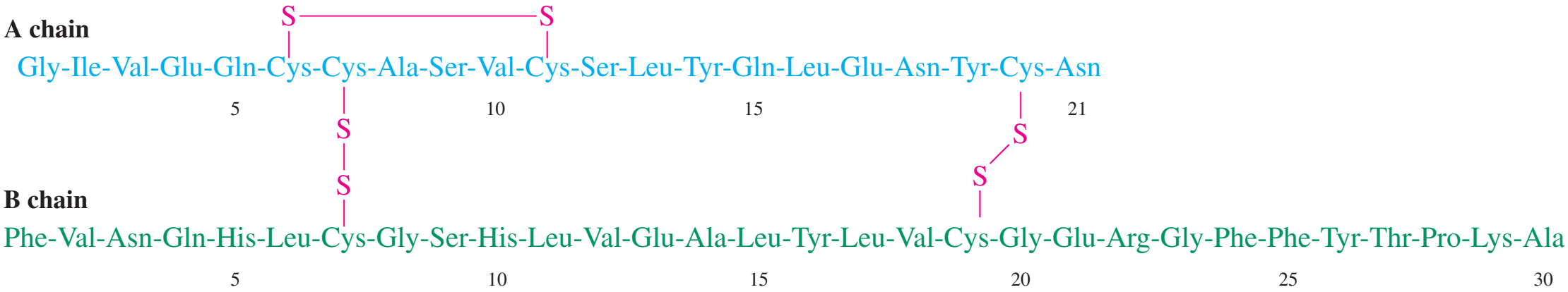
It contains phenylalanine in the *R* configuration and a rare amino acid, ornithine [Orn, a lower homolog (one less CH_2 group) of lysine]. The amino acids are linked (amino-to-carboxy direction), indicating by arrows.



Gramicidin S

Insulin illustrates the three-dimensional structure adopted by a complex sequence of amino acids. This protein hormone is an important drug in the treatment of diabetes because of its ability to regulate glucose metabolism.

Insulin contains 51 amino acid residues incorporated into two chains, denoted A and B. The chains are connected by two disulfide bridges, and there is an additional disulfide linkage connecting the cysteine residues at positions 6 and 11 of the A chain. Both chains fold up in a way that minimizes steric interference and maximizes electrostatic, London, and hydrogen-bonding attractions. These forces give rise to a fairly condensed three-dimensional structure.



Because synthetic methods gave only low yields, insulin used to be isolated from the pancreas of slaughtered cows and pigs, purified, and sold as such. In the 1980s, the development of genetic engineering methods allowed the cloning of the human gene that codes for insulin.

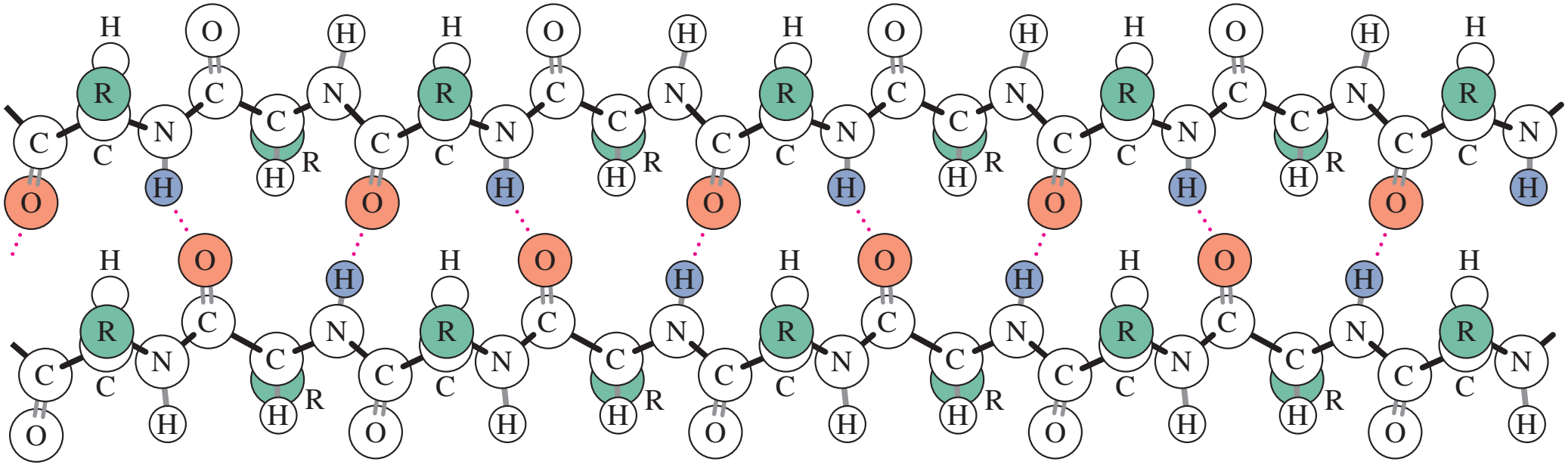
Proteins fold into pleated sheets and helices: secondary and tertiary structure

Insulin and other polypeptide chains adopt well-defined three-dimensional structures. Whereas the sequence of amino acids in the chain defines the **primary structure,** the folding pattern of the chain induced by the spatial arrangement of close-lying amino acid residues gives rise to the **secondary structure** of the polypeptide.

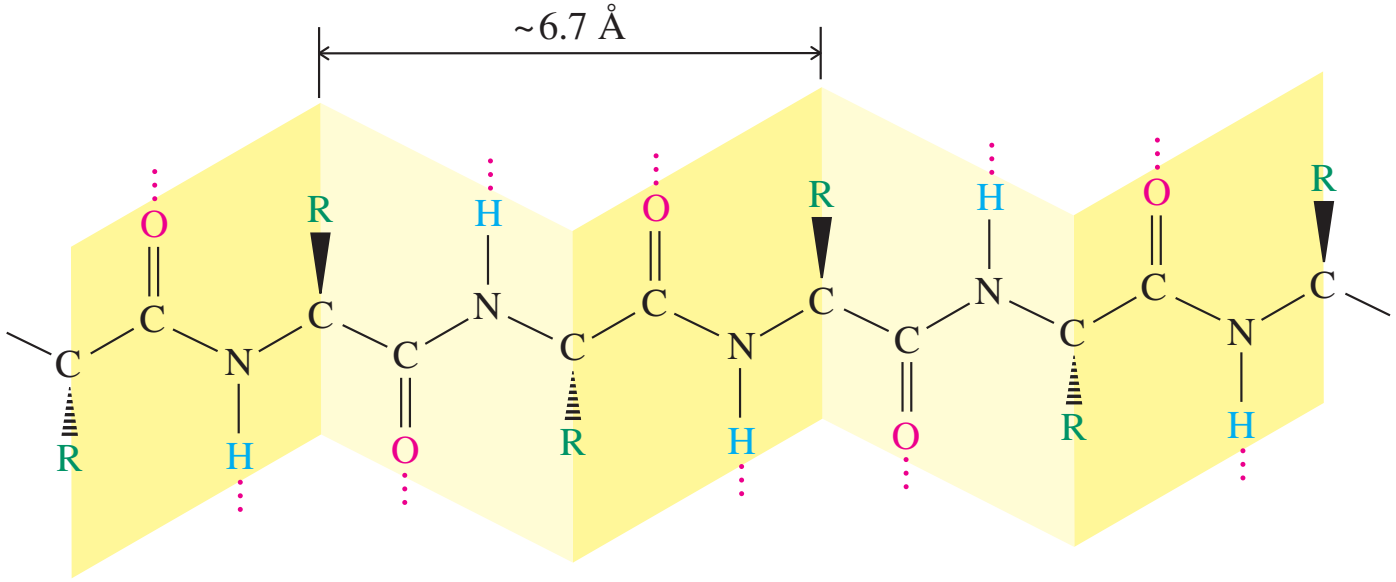
The secondary structure results mainly from the rigidity of the amide bond and from hydrogen (and other noncovalent) bonding along the chain(s). Two important arrangements are the pleated sheet, or β **configuration**, and the α **helix**.

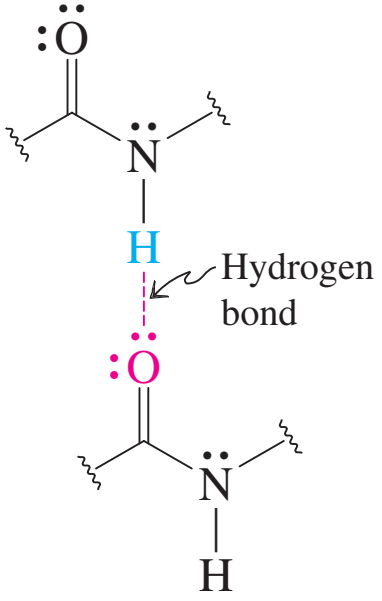
In the **pleated sheet** (also called β **sheet**), two chains line up with the amino groups of one peptide opposite the carbonyl groups of a second, thereby allowing hydrogen bonds to form.

Such bonds can also develop within a single chain if it loops back on itself. Multiple hydrogen bonding of this type can impart considerable rigidity to a system. The planes of adjacent amide linkages form a specific angle, a geometry that produces the observed pleated-sheet structure, in which the R groups protrude above and below at each kink.



A

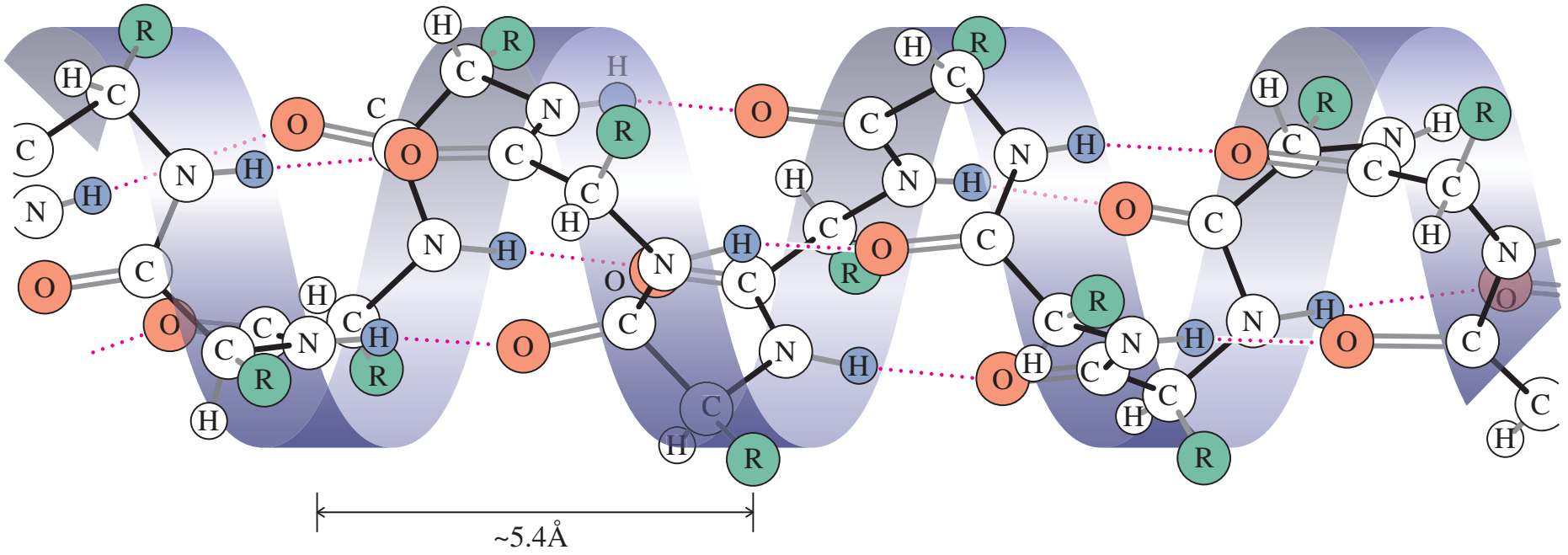




The *α* **helix** allows for intramolecular hydrogen bonding between nearby amino acids in the chain: The carbonyl oxygen of each amino acid is interacting with the amide hydrogen four residues ahead.

There are 3.6 amino acids per turn of the helix, two equivalent points in neighboring turns being about 5.4 Å apart.

The C=O and N–H bonds point in opposite directions and are roughly aligned with the helical axis. The (hydrophobic) R groups point away from the helix.

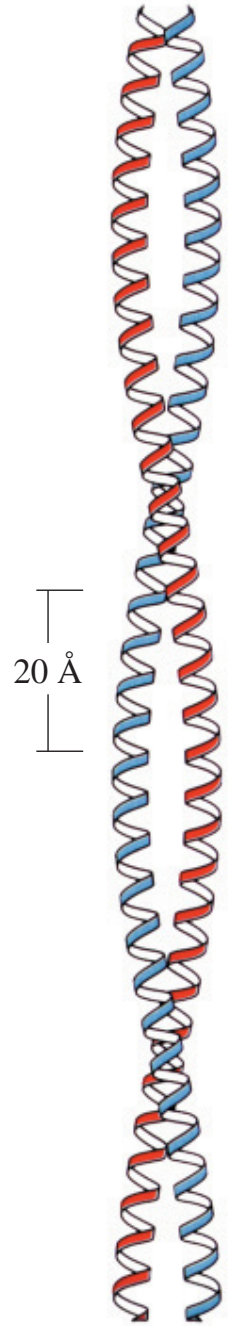


Tertiary structure: Further folding, coiling, and aggregation of polypeptides is induced by distant residues in the chain end. A variety of forces, all arising from the R group, come into play to stabilize such molecules, including disulfide bridges, hydrogen bonds, London forces, and electrostatic attraction and repulsion.

There are also **micellar effects**: The polymer adopts a structure that maximizes exposure of polar groups to the aqueous environment while minimizing exposure of hydrophobic groups (e.g., alkyl and phenyl), the “hydrophobic effect”.

Pronounced folding is observed in the **globular proteins,** many of which perform chemical transport and catalysis (e.g., myoglobin and hemoglobin). In the **fibrous proteins,** such as myosin (in muscle), fibrin (in blood clots), and α-keratin (in hair, nails, and wool), several α helices are coiled to produce a **superhelix**.

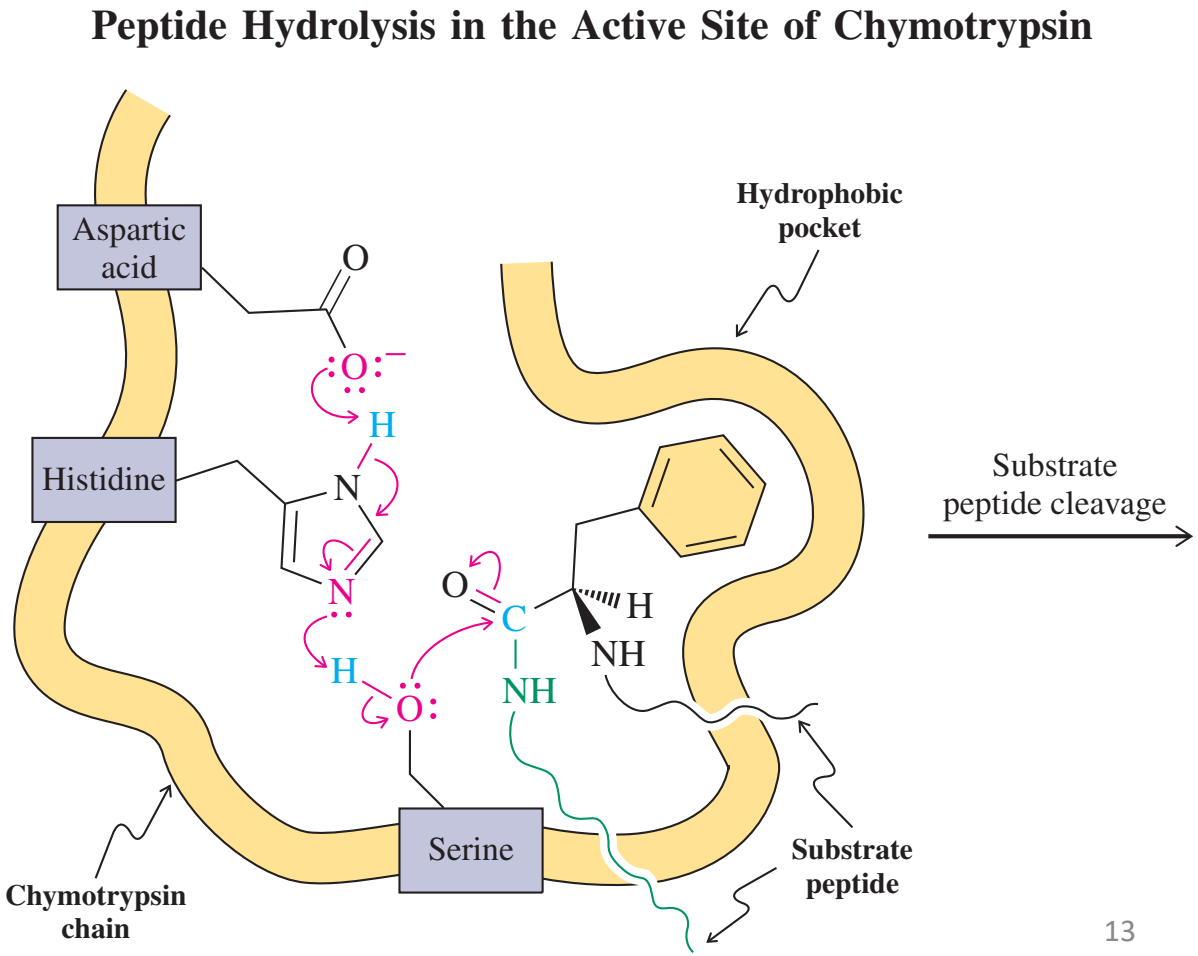
The tertiary structures of enzymes and transport proteins (carry molecules from place to place) have three-dimensional pockets, called **active sites** or **binding sites.** The size and shape of the active site provide a highly specific “fit” for the **substrate** or **ligand,** the molecule on which the protein carries out.



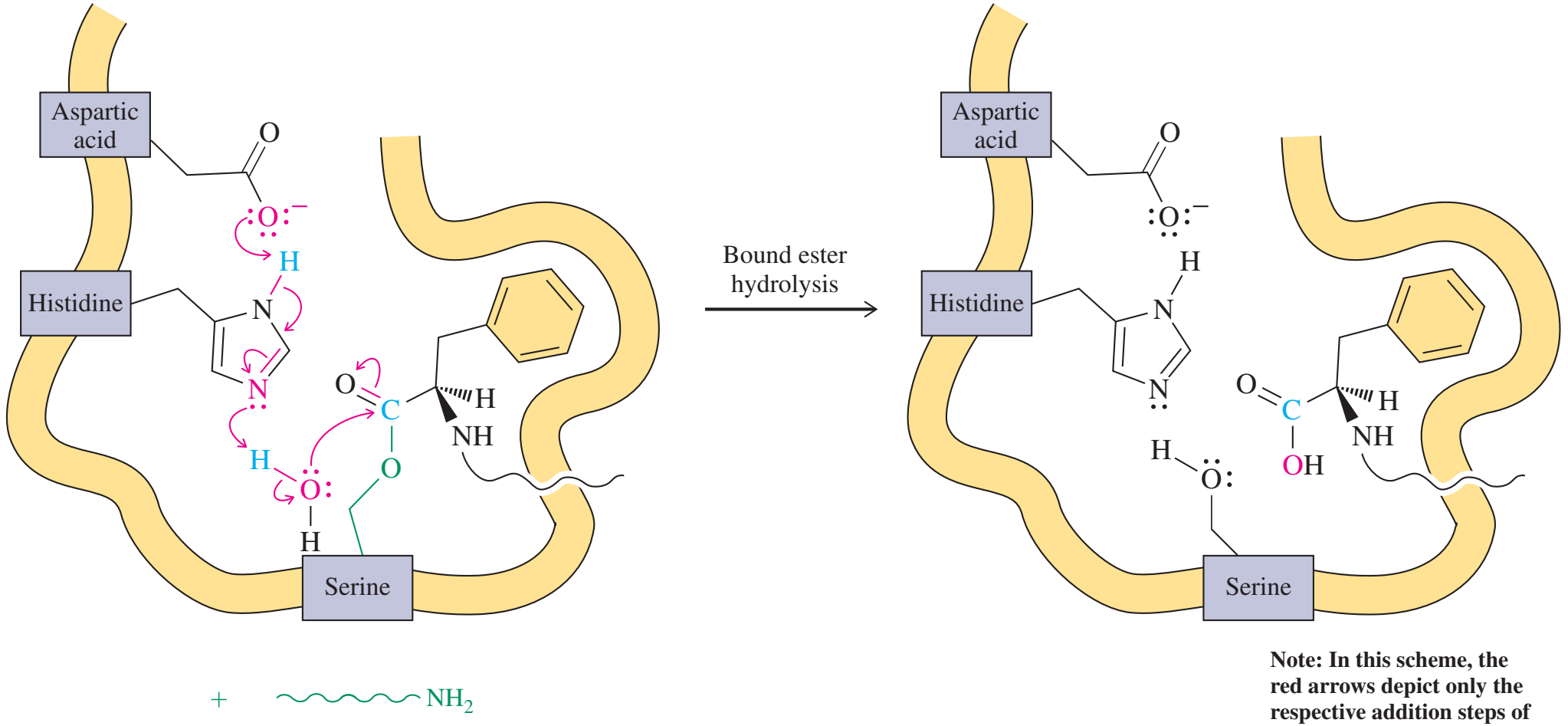
The inner surface of the pocket typically contains a specific arrangement of the side chains of polar amino acids that attracts functional groups in the substrate by hydrogen bonding or ionic interactions. In enzymes, the active site aligns functional groups and additional molecules in a way that promotes their reactions with the substrate.

An example is the active site of chymotrypsin, a mammalian digestive enzyme responsible for the degradation of proteins in food.

Chymotrypsin accomplishes the hydrolysis of peptide bonds at body temperature and at physiological pH. Recall that ordinary amide hydrolysis requires much more drastic conditions.



Moreover, the enzyme recognizes specific peptide linkages that are targeted for selective cleavage, such as the carboxy end of phenylalanine residues.



Note: In this scheme, the red arrows depict only the respective addition steps of the two addition–elimination sequences.

The enzyme has four important, close-lying parts, all of which work together to facilitate the hydrolysis reaction: a hydrophobic pocket, and the residues of aspartic acid, histidine, and serine.

The hydrophobic pocket helps to bind the polypeptide to be “digested” by attraction of the hydrophobic phenyl substituent of one of its component phenylalanine residues.

With the phenyl group held in this pocket, the three amino acid residues cooperate in a proton-transfer relay sequence to effect nucleophilic addition–elimination of the serine hydroxy group to the carbonyl function of Phe, releasing the amine part of the cleaved polypeptide.

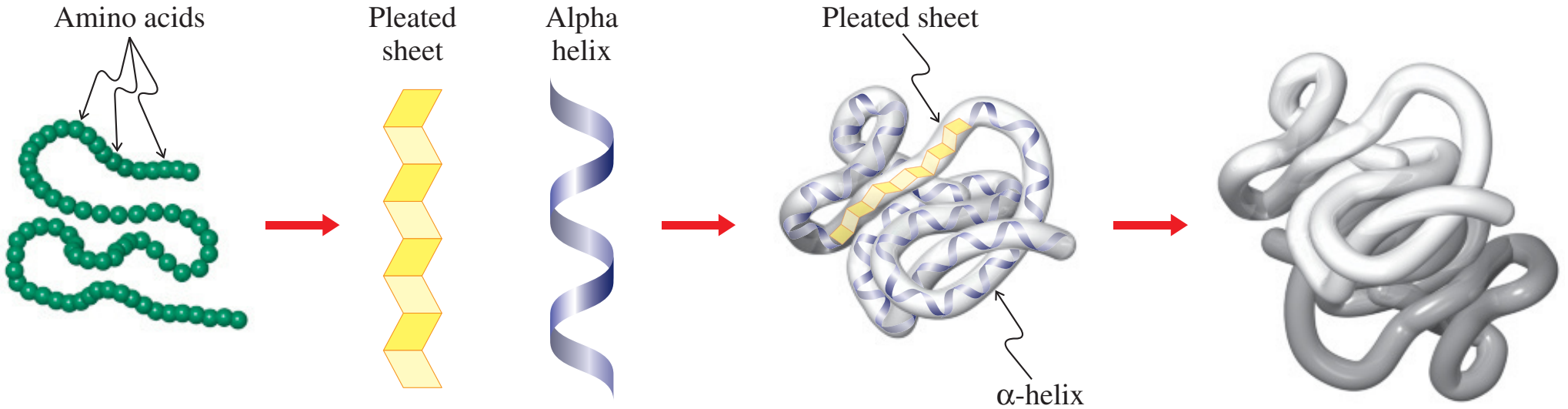
The remainder of the substrate is held by an ester linkage to the enzyme, positioned to undergo ester hydrolysis by a water molecule.

This reaction is aided by a tandem proton-transfer sequence similar to that used for hydrolyzing the peptide bond. With the link to the enzyme broken, the carboxy segment of the original substrate is now free to leave the intact active site of chymotrypsin, making room for another polypeptide.

Denaturation, or breakdown of tertiary structure, usually causes precipitation of the protein and destroys its catalytic activity. Denaturation is caused by exposure to excessive heat or extreme pH values. Think, for example, of what happens to clear egg white when it is poured into a hot frying pan or to milk when it is added to lemon tea.

Some molecules, such as hemoglobin, also adopt a **quaternary structure**, in which two or more polypeptide chains, each with its own tertiary structure, combine to form a larger assembly.

The Progression from Primary to Secondary, Tertiary, and Quaternary Polypeptide Structures



26-5 DETERMINATION OF PRIMARY STRUCTURE: AMINO ACID SEQUENCING

Biological function in polypeptides and proteins requires a specific three-dimensional shape and arrangement of functional groups, which necessitate a definite amino acid sequence.

The determination of the primary structure of a protein, called amino acid or polypeptide **sequencing,** can help us to understand the protein's mechanism of action.

First, purify the polypeptide

Several techniques can separate polypeptides on the basis of size, solubility in a particular solvent, charge, or ability to bind to a support.

Dialysis, the polypeptide is separated from smaller fragments by filtration through a semipermeable membrane.

Gel-filtration chromatography, uses a carbohydrate polymer in the form of a column of beads as a support. Smaller molecules diffuse more easily into the beads, spending a longer time on the column than large ones do.

Ion-exchange chromatography, a charged support separates molecules according to the amount of charge that they carry.

Electrophoresis is based on electric charge. A spot of the mixture to be separated is placed on a plate covered with a thin layer of chromatographic material (such as polyacrylamide) that is attached to two electrodes.

When the voltage is turned on, positively charged species (e.g., polypeptides rich in protonated amine groups) migrate toward the cathode, negatively charged species (carboxy-rich peptides) toward the anode. The separating power of this technique is extraordinary.

More than a thousand different proteins from one species of bacterium have been resolved in a single experiment.

Affinity chromatography exploits the tendency of polypeptides to bind very specifically to certain supports by hydrogen bonds and other attractive forces. Peptides of differing sizes and shapes have differing retention times in a column containing such a support.

Second, determine which amino acids are present

When the polypeptide strand has been purified, the next step in structural analysis is to establish its composition. To determine which amino acids and how much of each is present in the polypeptide, the entire chain is degraded by amide hydrolysis (6 N HCl, 110 ^{o}C, 24 h) to give a mixture of the free amino acids.

The mixture is then separated and its composition recorded by an automated **amino acid analyzer.**

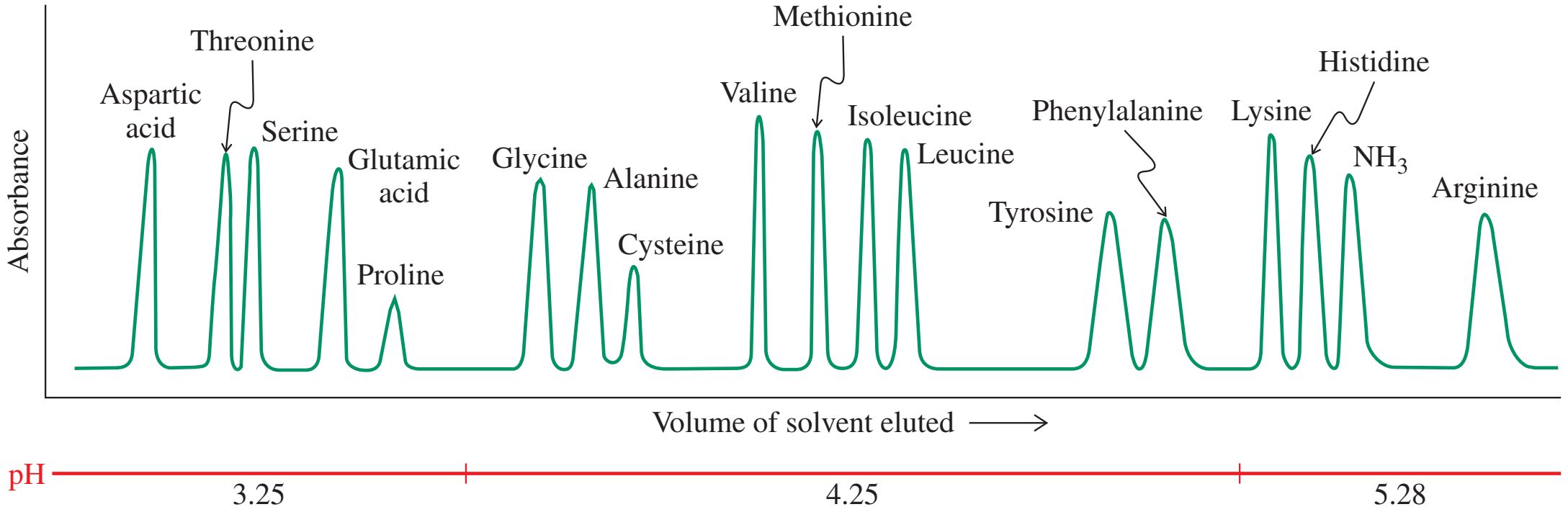
This instrument consists of a column bearing a negatively charged support, usually containing carboxylate or sulfonate ions. The amino acids pass through the column in slightly acidic solution. They are protonated to a greater or lesser degree, depending on their structure, and therefore are more or less retained on the column.

This differential retention separates the amino acids, and they come off the column in a specific order, beginning with the most acidic and ending with the most basic.

At the end of the column is a reservoir containing a special indicator.

Each amino acid produces a violet color whose intensity is proportional to the amount of that acid present and is recorded in a chromatogram. The area under each peak is a measure of the relative amount of a specific amino acid in the mixture.

The amino acid analyzer can readily establish the composition of a polypeptide.



Sequence the peptide from the amino (*N*-terminal) end

Determine the order in which the individual amino acids are bound to one another—the amino acid sequence.

Several different methods can reveal the identity of the residue at the amino end. One such procedure is the **Edman* degradation,** and the reagent used is phenyl isothiocyanate, $C_6H_5N{=}C{=}S$ (a sulfur analog of an isocyanate).

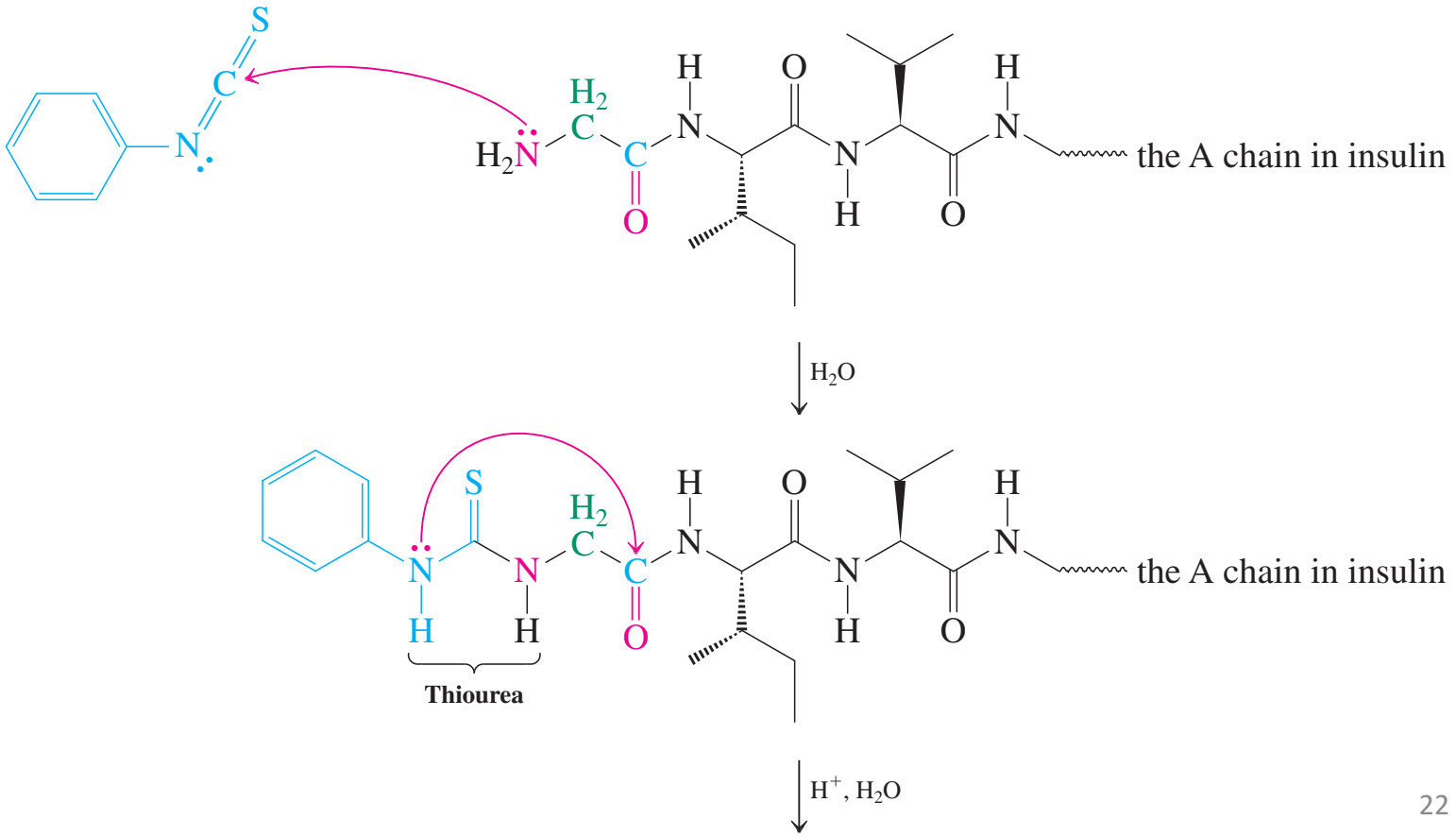
Isocyanates are very reactive with respect to nucleophilic attack, and the same is true of their sulfur analogs. In the Edman degradation, the terminal amino group adds to the isothiocyanate reagent to give a thiourea derivative.

Mild acid causes extrusion of the tagged amino acid as a phenylthiohydantoin, leaving the remainder of the polypeptide unchanged. The phenylthiohydantoins of all amino acids are well known, so the *N*-terminal end of the original polypeptide can be readily identified.

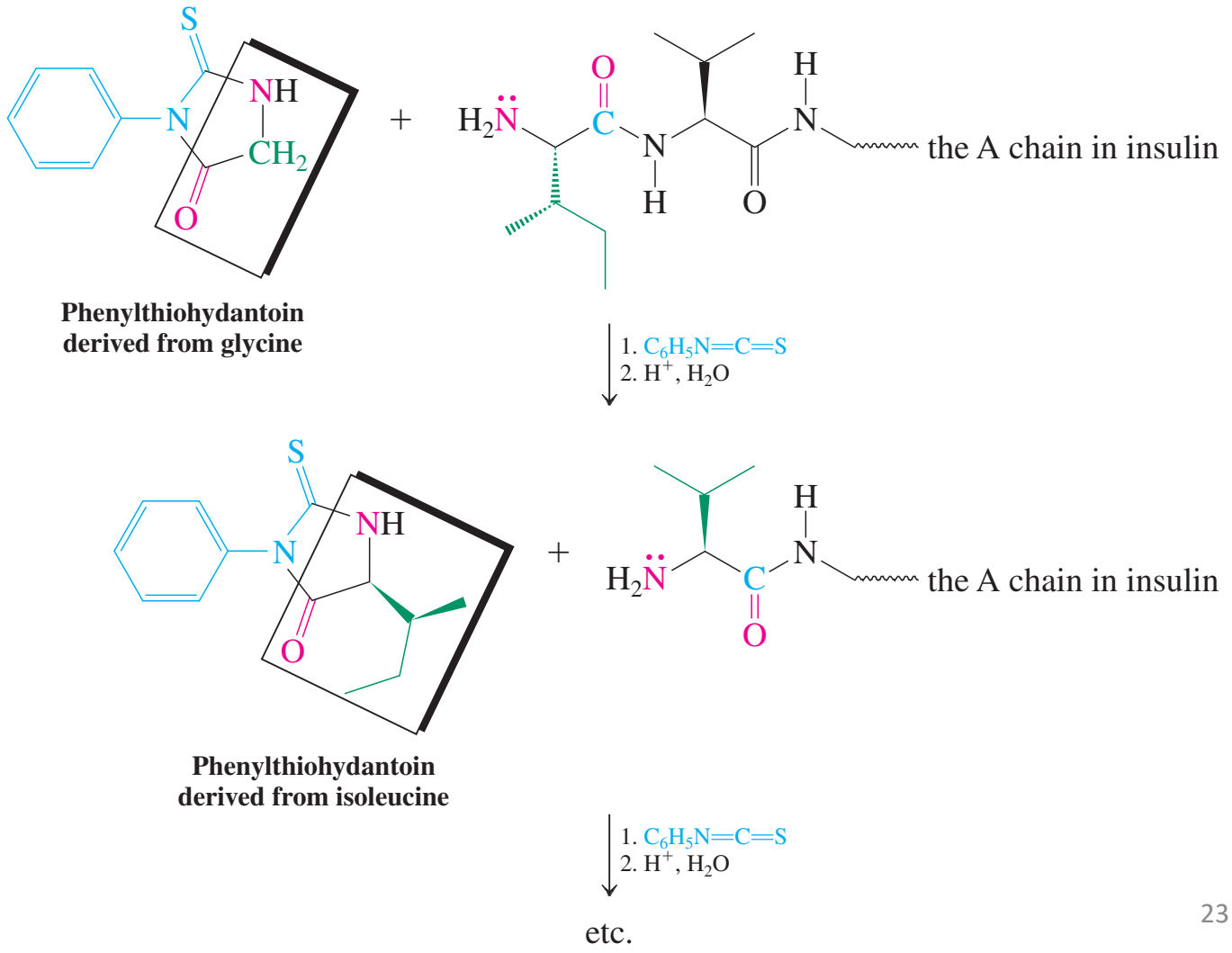
The new chain, carrying a new terminal amino acid, is now ready for another Edman degradation to tag the next residue, and so forth.

The entire procedure has been automated to allow the routine identification of polypeptides containing 50 or more amino acids.

Edman Degradation of the A Chain of Insulin



The buildup of impurities becomes a serious impediment. The reason for this drawback is that each degradation round is not completely quantitative, thus leaving small quantities of incompletely reacted peptide admixed with the new one.



The chopping up of longer chains is achieved with enzymes

The Edman procedure allows for the ready sequencing of only relatively short polypeptides. For longer ones (e.g., those with more than 50 residues), it is necessary to cleave the larger chains into shorter fragments in a selective and predictable manner. These cleavage methods rely mostly on hydrolytic enzymes.

Trypsin, a digestive enzyme of intestinal liquids, cleaves polypeptides only at the carboxy end of arginine and lysine. A more selective enzyme is clostripain, which cleaves only at the carboxy end of arginine. In contrast, chymotrypsin, which is found in mammalian intestines, is less selective and cleaves at the carboxy end of phenylalanine, tryptophan, and tyrosine.

Table 26-2 Specificity of Hydrolytic Enzymes in Polypeptide Cleavage

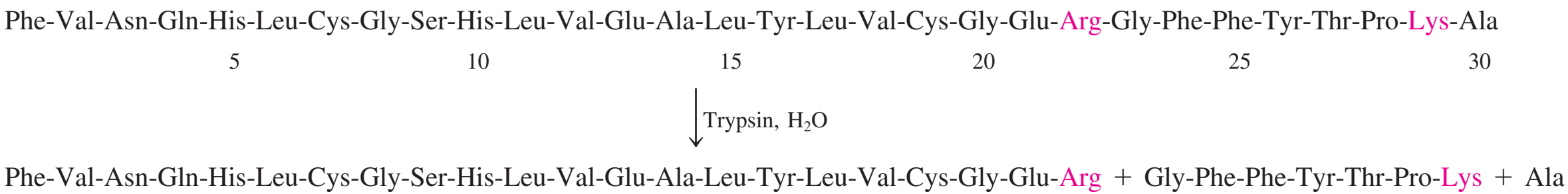
Enzyme	Site of cleavage
Trypsin	Lys, Arg, carboxy end
Clostripain	Arg, carboxy end
Chymotrypsin	Phe, Trp, Tyr, carboxy end
Pepsin	Asp, Glu, Leu, Phe, Trp, Tyr, carboxy end
Thermolysin	Leu, Ile, Val, amino end

Other enzymes have similar selectivity. In this way, a longer polypeptide is broken down into several shorter ones, which may then be sequenced by the Edman procedure.

After a first enzymatic cleavage, sequences of segments of the polypeptide under investigation are determined, but the order in which they are linked is not. For this purpose, selective hydrolysis is carried out a second time, by using a different enzyme that provides pieces in which the connectivities broken under the first conditions are left intact, so-called *overlap peptides.*

The solution is then found by literally “piecing” together the available information like a puzzle.

Selective Hydrolysis of the B Chain of Insulin by Trypsin



Protein sequencing is made possible by recombinant DNA technology

Despite the success of the techniques for polypeptide sequencing described so far, allowing for the structure elucidation of hundreds of proteins, their application to large systems (i.e., those containing more than 1000 residues) is an expensive, laborious, and time-consuming business.

Progress in this field was significantly impeded until the advent of recombinant DNA technology. The sequences of the four bases in DNA—adenine, thymine, guanine, and cytosine—are directly correlated to the amino acid sequences of the proteins encoded by genes or the corresponding messenger RNA.

Modern developments have led to the rapid automated analysis of DNA, and the knowledge gained can be immediately translated into primary protein structure. In this way, tens of thousands of proteins have been sequenced in the past few years.

26-6 SYNTHESIS OF POLYPEPTIDES: A CHALLENGE IN THE APPLICATION OF PROTECTING GROUPS

In a sense, the topic of peptide synthesis is a trivial one: Only one type of bond, the amide linkage, has to be made, but achieving selectivity poses great problems.

Consider even as simple a target as the dipeptide glycylalanine.

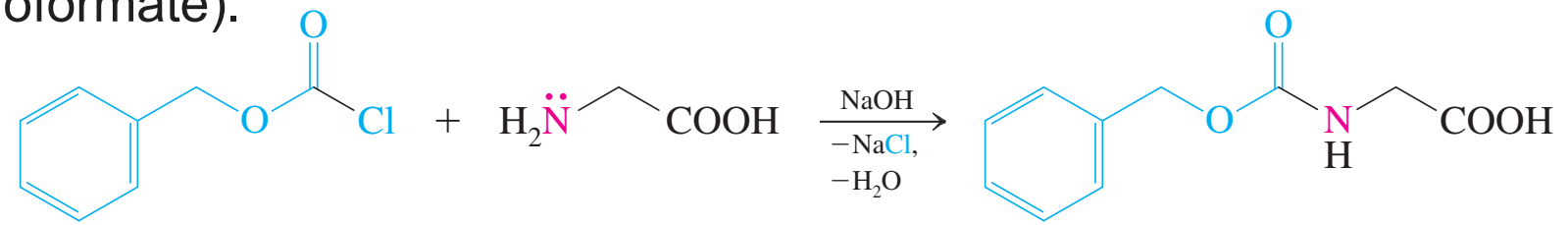
Just heating glycine and alanine to make the peptide bond by dehydration would result in a complex mixture of di-, tri-, and higher peptides with random sequences. Because the two starting materials can form bonds either to their own kind or to each other, there is no way to prevent random oligomerization.

An Attempt at the Synthesis of Glycylalanine by Thermal Dehydration

Selective peptide synthesis requires protecting groups

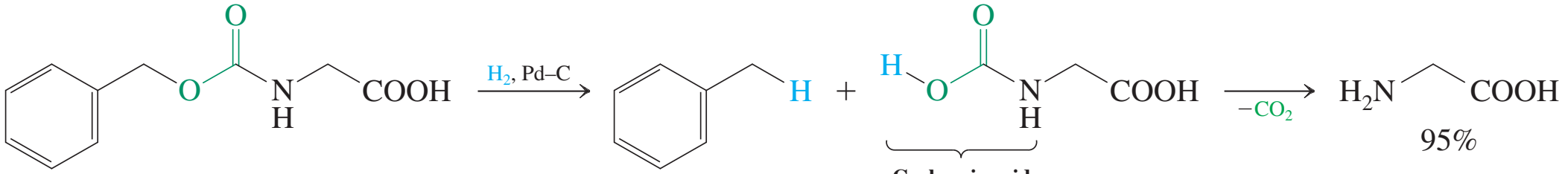
To form peptide bonds selectively, the functional groups of the amino acids have to be protected. There are several amino- and carboxy-protecting strategies available.

The amino end is frequently blocked by a **phenylmethoxycarbonyl group** (abbreviated **carbobenzoxy** or **Cbz**), introduced by reaction of an amino acid with phenylmethyl chloroformate (benzyl chloroformate).



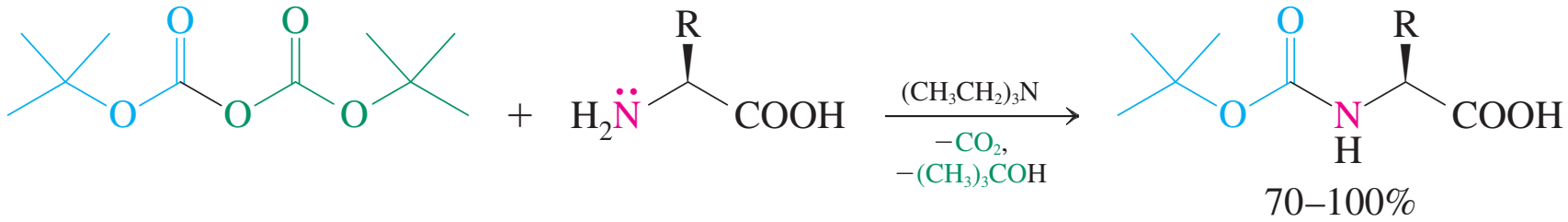
Phenylmethyl chloroformate (Benzyl chloroformate) | **Glycine** | **Phenylmethoxycarbonylglycine (Carbobenzoxyglycine, Cbz-Gly)**

The amino function is deprotected by hydrogenolysis, which initially furnishes the carbamic acid as a reactive intermediate. Decarboxylation occurs instantly to restore the free amine.



Carbamic acid function

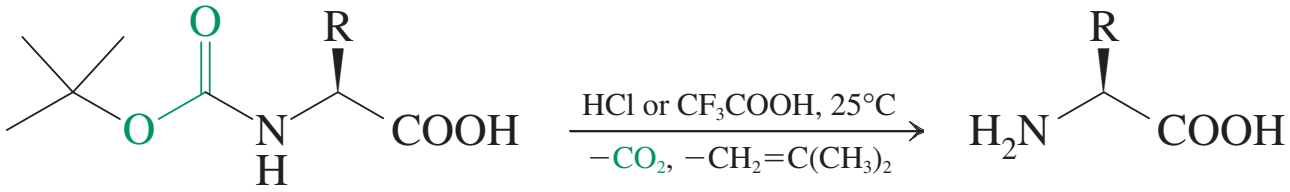
Another amino-protecting group is **1,1-dimethylethoxycarbonyl (*tert*-butoxycarbonyl, Boc),** introduced by reaction with bis(1,1-dimethylethyl) dicarbonate (di-*tert*-butyl dicarbonate). Similar to carbamic acids, the leaving group, $(CH_3)_3COCOOH$, is an unstable monoester of carbonic acid. It decomposes spontaneously to CO_2 and $(CH_3)_3COH$.



Bis(1,1-dimethylethyl) dicarbonate (Di-*tert*-butyl dicarbonate)

1,1-Dimethylethoxy-carbonylamino acid (*tert*-Butoxycarbonylamino acid, Boc-amino acid)

Deprotection in this case is achieved by treatment with acid under conditions mild enough to leave peptide bonds untouched.



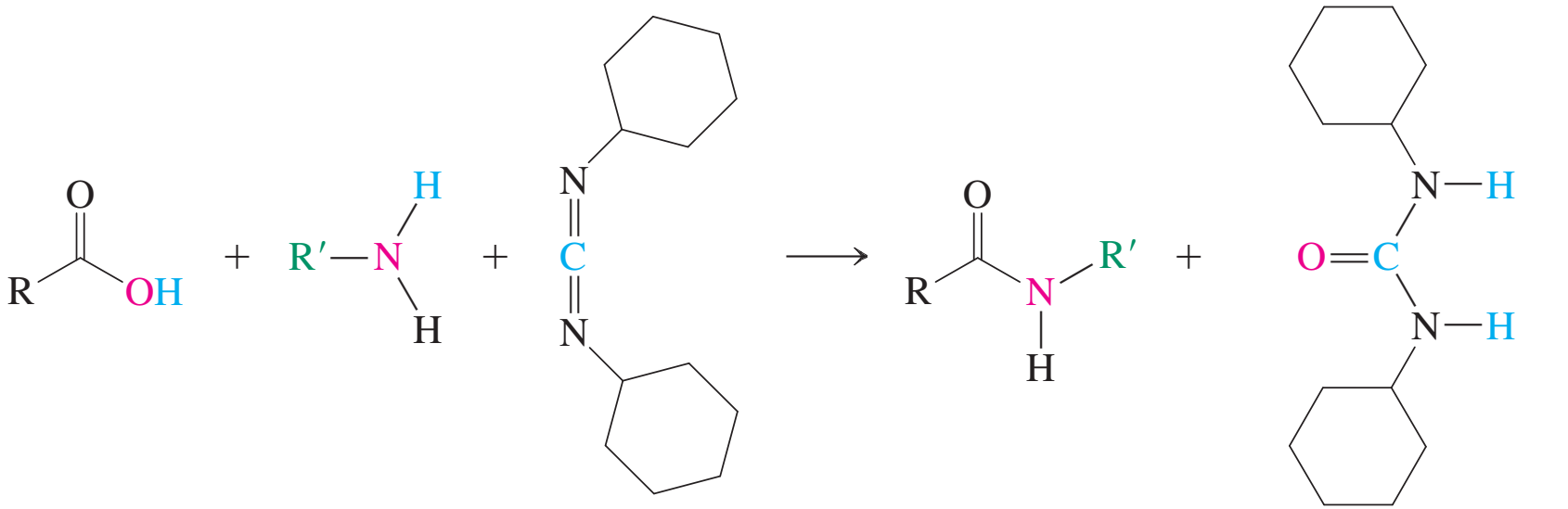
The carboxy terminus of an amino acid is protected by the formation of a simple ester, such as methyl or ethyl. Deprotection results from treatment with base. Phenylmethyl (benzyl) esters can be cleaved by hydrogenolysis under neutral conditions.

Peptide bonds are formed by using carboxy activation

With the ability to protect either end of the amino acid, we can synthesize peptides selectively by coupling an amino-protected unit with a carboxy-protected one. Because the protecting groups are sensitive to acid and base, the peptide bond must be formed under the mildest possible conditions.

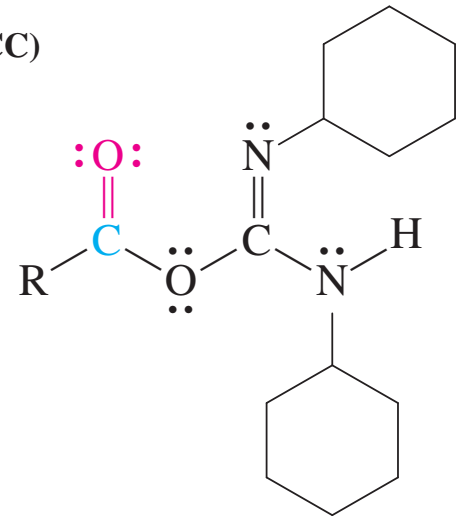
Special carboxy-activating reagents are used. The most general of these reagents is **dicyclohexylcarbodiimide (DCC).** The electrophilic reactivity of this molecule is similar to that of an isocyanate; it is ultimately hydrated to *N*,*N*'-dicyclohexylurea.

The role of DCC is to activate the carbonyl group of the acid to nucleophilic attack by the amine. This activation arises from the formation of an ***O*-acyl isourea,** in which the carbonyl group possesses reactivity similar to that in an anhydride.



Dicyclohexyl-carbodiimide (DCC)

***N,N′*-Dicyclohexylurea**



An *O*-acyl isourea

Synthesis of glycylalanine: amino-protected glycine is added to an alanyl ester in the presence of DCC. The resulting product is then deprotected to give the desired dipeptide.

Preparation of Gly-Ala