Laboratory Synthesis of Peptides

Synthesis of polypeptides in the laboratory is much more complicated than the synthesis of the polyamides described in Chapter 24 because more than one monomer must be used and the order of their attachment must be carefully controlled. For example, suppose we want to prepare the dipeptide Gly-Ala. We cannot just heat a mixture of glycine and alanine. Although this would produce some of the desired dipeptide, it would also form a host of other products, including other dipeptides and tripeptides, as shown in the following equation:

\[
\begin{align*}
\text{Gly} & \quad \text{CH}_3 \\
& \quad \text{NH}_2\text{CH}_2\text{CO}_2\text{H} \\
& \quad \text{Gly} \quad \text{Ala} \\
& \quad \text{NH}_2\text{CH}_2\text{CO}_2\text{H} + \text{NH}_2\text{CH}_2\text{CO}_2\text{H} \\
& \quad \text{Gly-Gly} \\
+ & \quad \text{NH}_2\text{CH}_2\text{CO}_2\text{H} + \text{NH}_2\text{CH}_2\text{CO}_2\text{H} \\
& \quad \text{Ala-Ala} \\
& \quad \text{Gly-Ala} \\
& \quad \text{Ala-Gly} \\
& \quad \text{Gly-Gly} + \text{Ala-Ala} + \text{Gly-Ala-Gly} + \text{Gly-Ala-Ala} + \ldots
\end{align*}
\]

The problem with this reaction is that there are too many reactive functional groups: two amino groups and two carboxylic acid groups. To prepare the desired dipeptide in a controlled manner, it is necessary to restrict the reactivity of the compounds so that only the amino group of alanine and the carboxylic acid group of glycine are available to react. As usual, this is accomplished by the use of protecting groups on the functional groups in which reaction is not desired. The synthesis of the dipeptide is performed by first protecting the
amino group of glycine and the carboxylic acid group of alanine. To form the amide bond, the carboxylic acid group of the protected glycine is activated by conversion to a more reactive derivative (see Section 19.6). Then the protected alanine is added. In the final step the protecting groups must be removed without breaking the newly formed amide bond.

Recall that an amino group is commonly converted to an amide for protection (see Section 23.4). However, there must be some special feature of the protecting group that allows it to be removed without also cleaving the amide bonds of the peptide. Although a large number of different protecting groups has been developed, the most common ones are the tert-butoxycarbonyl (BOC) group and the benzzyloxy carbonyl group. The BOC group is added by the use of an anhydride and is readily removed by treatment with dilute acid:

The benzzyloxy carbonyl group is added by using an acyl chloride and can be removed by hydrogenolysis:

PROBLEM 26.19
Show all of the steps in the mechanisms for these reactions:
The carboxylic acid group is usually protected as an ester. Because an ester is considerably more reactive than an amide, it is possible to remove this protecting group by hydrolysis without cleaving the amide bond of the peptide.

For a carboxylic acid and an amine to form an amide, the carboxylic acid usually must be activated; that is, it must be converted to a more reactive functional group. Conversion to an acyl chloride is a common way to accomplish this for normal organic reactions (see Chapter 19). However, acyl chlorides are quite reactive and do not give high enough yields in peptide synthesis because of side reactions. Therefore, milder procedures for forming the amide bond are usually employed. In one method the carboxylic acid is reacted with ethyl chloroformate (a half acyl chloride, half ester of carbonic acid) to produce an anhydride. Treatment of this anhydride with an amine results in the formation of an amide:

\[
\text{RCOH} + \text{ClCOEt} \rightarrow \text{RCOEt} + \text{HCl}
\]

**PROBLEM 26.20**

Show all of the steps in the mechanisms for these reactions:

\[
\begin{align*}
\text{a) } & \text{RCOH} + \text{ClCOEt} \rightarrow \text{RCOEt} + \text{HCl} \\
& \text{Formed} \\
\text{b) } & \text{RCOEt} + \text{R'NH}_2 \rightarrow \text{RCNHR'} + \text{CO}_2 + \text{EtOH}
\end{align*}
\]

Explain why the second product is not formed in the preceding reaction.

Explain why the amine attacks the left carbonyl group rather than the right one.

The most common way to form the amide bond in peptide synthesis uses dicyclohexylcarbodiimide (DCC) as the coupling agent:

\[
\text{RCOH} + \text{NH}_2\text{R'} \rightarrow \text{RCNHR'} + \text{EtOH}
\]

The mechanism for amide formation promoted by DCC is shown in Figure 26.4. The carboxylic acid first reacts with DCC to form an intermediate that resembles an anhydride, but with a carbon–nitrogen double bond in place of one of the carbonyl groups.
When this pseudo-anhydride is reacted with an amine, an amide is produced along with the by-product, dicyclohexylurea. Many other coupling reagents and protecting groups have been developed, but the ones shown here are the most common and illustrate the principles involved.

**Active Figure 26.4**

**MECHANISM OF AMIDE FORMATION USING DICYCLOHEXYLCARBODIIMIDE.** Test yourself on the concepts in this figure at OrganicChemistryNow.

When this pseudo-anhydride is reacted with an amine, an amide is produced along with the by-product, dicyclohexylurea. Many other coupling reagents and protecting groups have been developed, but the ones shown here are the most common and illustrate the principles involved.

**PROBLEM 26.21**

Show the products of these reactions:

a) \( t\text{-BuOCOOCR-Bu} + \text{NH}_2\text{CH}_2\text{COH} \xrightarrow{\text{Et}_3\text{N}} \)

b) \( \text{NH}_2\text{CH}_2\text{CNHCH}_2\text{COCH}_2\text{Ph} \xrightarrow{t\text{-BuOCOOCR-Bu} + \text{Et}_3\text{N}} \text{H}_2 \xrightarrow{\text{HCl}} \text{H}_2\text{O} \)
Let's consider the synthesis of Gly-Ala. First, the amino group of glycine must be protected. (Today, amino acids can be purchased in BOC protected form.)

The carboxyl group of alanine must also be protected:

\[
\text{H}_2\text{C} \quad \text{O} \quad \text{NH}_2\text{CHCOH} + \text{CH}_3\text{OH} \quad \xrightarrow{[\text{HA}]} \quad \text{H}_2\text{C} \quad \text{O} \quad \text{NH}_2\text{CHCOCH}_3
\]

Then the carboxyl group of the protected glycine must be activated:

\[
\text{t-BuOCOCO}_\text{t-Bu} + \text{NH}_2\text{CH}_2\text{COH} \quad \xrightarrow{\text{HCl}} \quad \text{t-BuOCNHCH}_2\text{COH}
\]

The next step is the formation of the amide bond by reaction of the activated, N-protected glycine with the carboxyl-protected alanine:

\[
\text{t-BuOCNHCH}_2\text{COEt} + \text{NH}_2\text{CHCOCH}_3 \quad \xrightarrow{\text{HCl}} \quad \text{t-BuOCNHCH}_2\text{C} \equiv \text{NHCHCOCH}_3
\]

To complete the synthesis, it is necessary only to remove the protecting groups. The BOC group is removed by treatment with aqueous acid. This reaction occurs under very mild conditions that do not also hydrolyze the ester group. Note that this dipeptide, deprotected at the N-terminal, can be reacted with another N-protected, carboxyl-activated amino acid to produce a tripeptide. These steps can be repeated to produce a tetrapeptide, and so on.

Finally, the carboxyl protecting group is removed by saponification:

\[
\text{t-BuOCNHCH}_2\text{C} \equiv \text{NHCHCOCH}_3 \quad \xrightarrow{\text{HCl}} \quad \text{NH}_2\text{CH}_2\text{C} \equiv \text{NHCHCOCH}_3
\]

\[
\text{NH}_2\text{CH}_2\text{C} \equiv \text{NHCHCOCH}_3 \quad \xrightarrow{\text{NaOH}} \quad \text{NH}_2\text{CH}_2\text{C} \equiv \text{NHCHCOH}
\]

Of course, other protecting groups and other coupling methods could be employed. In addition, things are a little more complicated because many of the amino acids have
functional groups in their side chains that must also be protected. However, the general ideas of the process are provided in the preceding procedure.

**PROBLEM 26.22**
Show a synthesis of Ala-Phe-Gly-Gly starting from the individual amino acids.

As is apparent from this example, the synthesis of a polypeptide requires numerous steps. At each step, the product must be isolated, and if you have worked in an organic chemistry laboratory, you are certainly aware of how much time and energy are required to isolate and purify a product. This makes a polypeptide synthesis quite tedious. In addition, the mechanical losses that occur in each isolation step contribute to lower yields for the overall process. Motivated by these problems, R. B. Merrifield developed a method, called *solid phase synthesis*, that makes the preparation of a polypeptide much easier. Merrifield was awarded the 1984 Nobel Prize in chemistry for this work. Let’s see how it works.

Merrifield’s idea was to attach the initial amino acid molecules to the surface of insoluble polymer beads. To isolate the product after the next amino acid has been attached, it is necessary only to collect the beads by filtration and wash them to remove any remaining reagent. Not only is the isolation procedure fast and simple, but mechanical losses are minimized.

The insoluble polymer that is used for solid phase syntheses is the chloromethylated, divinylbenzene cross-linked polystyrene that was described in Section 24.10. The beads of this material have CH₂Cl₂ groups bonded to the phenyl groups that are part of the polymer chains on the surface of the beads. As we saw in Section 24.10, functional groups on the surface of such beads exhibit normal chemical reactions. To synthesize a polypeptide, the carboxyl group of the C-terminal amino acid, with its amine group protected, is first bonded to the polymer by an ester linkage. The amino group of this amino acid is deprotected. Then the next amino acid, with its amine group protected, is attached using DCC to promote coupling. This cycle of deprotection of the amino group followed by coupling with another N-protected amino acid is repeated until all of the amino acids of the desired polypeptide have been added. At each step, isolation of the product is accomplished by simply collecting the beads by filtration. Finally, the peptide is cleaved from the polymer bead. An example of a solid phase synthesis is provided in Figure 26.5.

**PROBLEM 26.23**
Show a synthesis of Ala-Phe-Gly-Gly using the Merrifield solid phase method.

As mentioned earlier, the ease of isolation of the product is the major advantage of the solid phase method. However, this is also a disadvantage because the product, the growing polypeptide attached to the polymer, is never purified until it is finally cleaved from the polymer. Each time a coupling reaction does not proceed in 100% yield, a small amount of the final product will be missing one of the amino acids. These impurities build up with the number of steps in the synthesis and become more and more difficult to remove as the polypeptide becomes larger. For this reason the yield of each step must be as high as possible. Current methods provide yields of 99.5% or better in each step.

Another advantage of the solid phase method is that the steps are repetitive and can be automated. An automatic peptide synthesizer is commercially available. It consists of a flask with a fritted-glass filter at the bottom along with bottles containing solutions of BOC-protected amino acids and the other necessary reagents and wash solutions. The polymer with the first amino acid attached is placed in the flask, and the order of
The C-terminal amino acid, BOC-protected phenylalanine, is reacted with the chloromethylated polymer. The carboxylate anion acts as a nucleophile in an $S_N$2 displacement of Cl.

The BOC group is removed by treatment with acid.

The next amino acid, BOC-protected alanine, is coupled using DCC.

The cycle is repeated to add the next amino acid. The BOC group is removed by treatment with acid.

BOC-protected glycine is coupled using DCC.

Finally, the tripeptide is cleaved from the polymer by treatment with HF. The benzylic bond is broken in an $S_N$2 reaction. The acid treatment also removes the BOC group.

Figure 26.5

THE MERRIFIELD SOLID PHASE SYNTHESIS OF GLY-ALA-PHE.
the amino acids in the desired polypeptide is programmed into the machine. The synthesizer then adds the next amino acid and DCC, reacts the mixture for the appropriate period of time, and removes the reagents by filtration. After the polymer beads have been washed, the next reagent is added and the process continues. The nonapeptide hormone, bradykinin, has been synthesized in 85% yield in 27 hours using such a device. As another example, bovine pancreatic enzyme ribonuclease A, with 124 amino acids, was synthesized in 17% yield in 6 weeks. After purification this material showed 78% of the bioactivity of the natural enzyme. However, this is an example in which the synthesis went particularly well. More commonly, large polypeptides like this contain many impurities that cannot readily be removed and therefore have a bioactivity that is much less than the natural protein. Solid phase synthesis is best for smaller polypeptides, up to about 50 amino acid residues. For larger peptides it is usually best to synthesize fragments of 50 amino acids or less and then couple these by solution phase reactions after they have been purified.

26.8 Protein Structure

The properties of a protein depend primarily on its three-dimensional structure. The sequence of amino acids in the polypeptide chain is termed its primary structure. Its secondary structure is the shape of the backbone polypeptide chain. Remember that each amide group is planar, but the chain can have various conformations about the bond between the \( \alpha \)-carbon and the nitrogen. The tertiary structure is the overall three-dimensional shape of the protein, including the conformations of the side chains.

The primary structure of the protein—that is, the sequence of its amino acids—determines its secondary and tertiary structure. Each protein usually has one shape that has the lowest energy. Individual molecules of the protein naturally and reproducibly fold into this lowest-energy conformation. The interactions that determine the preferred shape include hydrogen bonding involving the hydrogens on the amide nitrogens and the carbonyl groups, disulfide bonds, and interactions of the side chains with other side chains and the aqueous environment of the protein. The protein usually folds so that charged side chains—those of aspartic acid, glutamic acid, lysine, arginine, and histidine—are on the surface of the protein, so they can be solvated by water. Polar side chains also prefer the surface, where they are exposed to the aqueous environment, whereas nonpolar side chains, with their hydrophobic groups, prefer to pack together in the interior of the protein to avoid contact with water.

Two common elements of secondary structure that comprise about 50% of the structure of an average protein are the \( \alpha \)-helix and the \( \beta \)-pleated sheet. In an \( \alpha \)-helix the polypeptide chain assumes a helical shape that is held in place by hydrogen bonding between the NH group of one amino acid residue and the carbonyl oxygen four residues removed. There are 3.6 amino acid residues per turn of the helix. A simplified picture of an \( \alpha \)-helix is shown in Figure 26.6.

In \( \beta \)-pleated sheets the hydrogen bonding occurs between parts of the polypeptide chain separated by larger distances. In one version, the antiparallel \( \beta \)-pleated sheet, the parts of the chain in the sheet run in opposite directions; that is, the N-terminal to C-terminal directions are opposite, as shown in Figure 26.7. In a parallel \( \beta \)-pleated sheet, the N-terminal to C-terminal directions are the same. In either case, when viewed from the edge, the sheet takes on a pleated shape. In an actual protein, \( \alpha \)-helices and \( \beta \)-pleated sheets, connected by loops and turns, pack together to give the overall three-dimensional shape of the protein.