6

Peptide Synthesis

INTRODUCTION

The chemical synthesis of a peptide presents a variety of challenges to the organic chemist. Many side-chain groups in the constituent amino acids have to be protected and subsequently unblocked, peptide bonds synthesized without racemization of the optically active constituents, side reactions prevented, and at the end of the synthesis a homogeneous product must be obtained from a mixture that may well contain highly analogous side products. The classical criteria of success has to be applied to peptide synthesis just as in any other organic synthesis; that is, success is judged by the observation that the physical, chemical, and biological properties of the product match those of the natural parent compound.

Peptide synthesis has evolved to the stage where small peptides (up to 20 to 30 residues) have become a cornerstone in examining structure–function relationships in peptide hormones (almost all of which have been chemically synthesized) by the systematic variation of one or more residues. In another application, the putative epitope of a viral antigen has been explored by systematic synthesis of a series of analogs with different amino acid substitutions in each position of a heptapeptide epitope, first identified by synthesis of all 208 possible overlapping peptides covering the 213-residue viral coat protein antigen.

Synthesis of long peptides has not been successfully achieved by chemical means and with the advent of cloning and site-specific mutagenesis as an approach to manipulating protein structures, much of the impetus to long peptide synthesis has been removed. Among the larger peptides successfully synthesized is the 57-residue apolipoprotein C-I, the protein constituent of the very low density human plasma

lipoproteins. The synthesized protein activated lecithin:cholesterol acyltransferase to the same extent as the native protein and bound similar amounts of dimyristoyl phosphatidylcholine. In an ambitious, yet to be completed project, the synthesis of a hypothetical protein that would have a predicted β-barrel secondary–tertiary structure has been undertaken. The synthesis of this approximately 80-residue protein has been reported, but the determination of the crystal structure is still in progress.

In this chapter we discuss some of the problems encountered in peptide synthesis, the basic protocols that have been developed to accomplish it, and some of the alternative approaches that are being used to provide the answers to protein structure–function questions that the original workers in the field of peptide synthesis envisaged being able to answer about proteins in general.

CHEMICAL APPROACHES

Two different approaches have been developed which share many of the experimental challenges of synthesis, and they are schematically shown in Fig. 6-1: the solution method (part A) and the solid-phase support method (part B). In both methods the carboxyl group of the amino acid on the soon-to-be N-terminal side of the peptide must be activated to allow peptide-bond formation during the coupling phase of the reaction. In the solution method the carboxyl group of the C-terminal side and the amino group of the N-terminal side must both be protected to prevent unwanted side reactions. In the solid-phase support method the carboxyl group of the C-terminal side is involved in immobilization to the support and does not need protection. With both approaches any reactive groups in the amino acid side chains have to be protected. After coupling, the blocking groups (and in the case of the solid-phase method the immobilization linkage) must be removed to give the final dipeptide product. To achieve larger peptides, these processes can be repeated sequentially, although purification of intermediates to remove unwanted-side-product polypeptides is advisable.

It is quite a simple process to synthesize a peptide up to five or six residues by such procedures. If the final end product is a much larger peptide, the fragment condensation approach becomes attractive. In this approach, shown schematically in Fig. 6-2, two half-fragments are synthesized, purified, and subsequently ligated.

The advantage is that purification of the final product, peptide C in the scheme, is quite easy since the reaction mixture contains only peptides A and B (or derivatives thereof produced during the ligation reaction) and the desired product peptide C, which is much larger than the reactants. Although this process sounds simple, it has some potential problems associated with the ligation procedures.

Reactive Group Protection

We now review briefly some of the procedures used in carboxyl or amino-terminal protection, in coupling and ligation, and in side-chain protection, as well as immobilization processes employed in the solid-phase method.

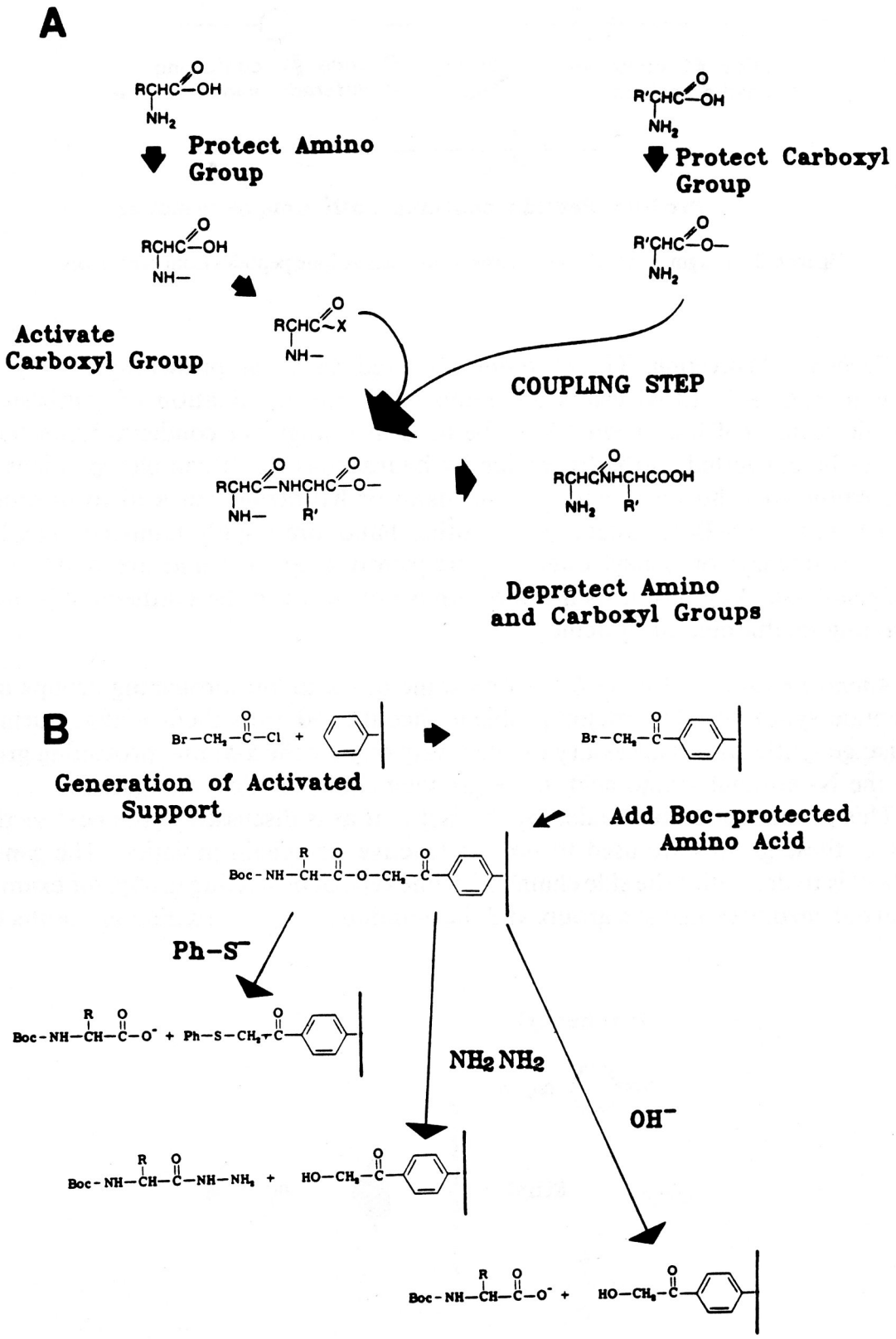


Figure 6-1 Outline of alternative methods for peptide synthesis: (A) solution method; (B) solid-phase support method.

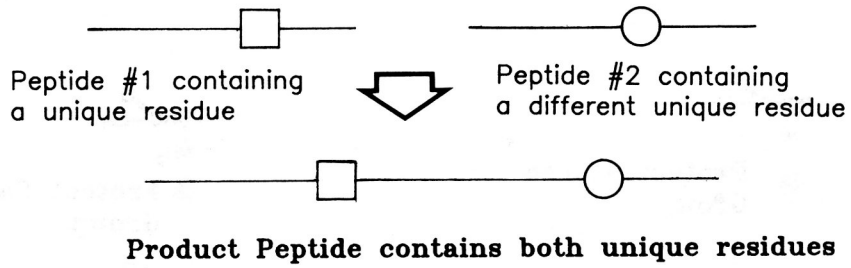


Figure 6-2 Fragment condensation used to synthesize long peptides from precursors.

Carboxyl Protection. Three commonly used carboxyl protecting groups are shown in Fig. 6-3. Ethyl esters are removed at the termination of synthesis by saponification, or if the product is to be used in a fragment condensation scheme, they can be converted to the hydrazide by hydrazinolysis. If complex peptides are being synthesized, however, exposure to alkali or hydrazine can lead to unwanted side reactions. *tert*-Butyl esters, on the other hand, are readily removed acidolytically. Nitrobenzyl or benzyl esters are frequently employed and are removed by hydrogenolysis. As a result, such protection is not suited to the synthesis of peptides containing methionine or cysteine.

Amino Protection. Figure 6-4 shows some of the amino-protecting groups used in peptide synthesis. The major problem encountered with them is that during a multistage synthesis it is necessary to selectively remove the α-amino protecting group from the N-terminal amino acid in the growing chain to allow elongation.

This problem is compounded by the fact that as is discussed in the next section, many of these groups are used to protect reactive side-chain moieties. The general protocol is to derivatize the side chains with one type of protecting group, for example, the carbobenzoxy protection group, and the α-amino group with a different protecting

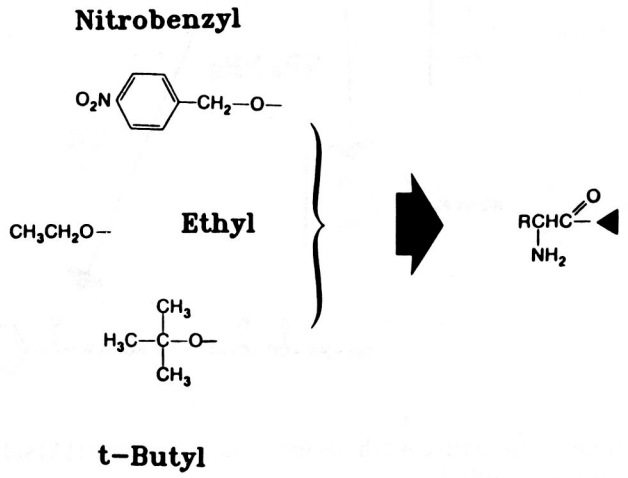


Figure 6-3 Carboxyl protecting groups used in peptide synthesis.

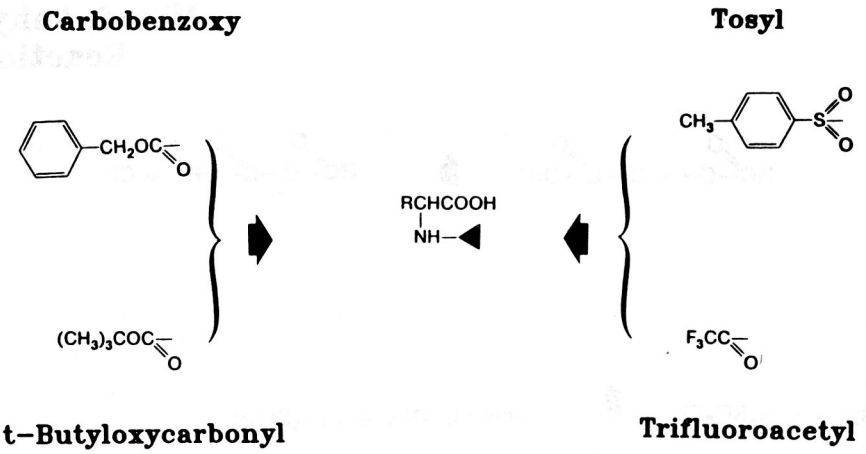


Figure 6-4 Amino protecting groups used in peptide synthesis.

group, for example, the *t*-butyloxycarbonyl protection group (or vice versa), and selectively remove the α-amino protecting group.

In peptides lacking cysteine or methionine the carbobenzoxy group can be removed in the presence of *t*-butyloxycarbonyl groups by hydrogenolysis. The converse situation, selective removal of *t*-butyloxycarbonyl in the presence of carbobenzoxy, is more difficult, but has been achieved by treatment with 98% formic acid or β-mercaptoethanesulfonic acid.

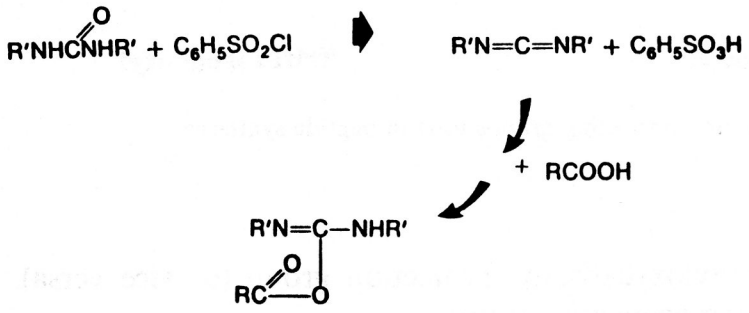
Side-Chain Protection. Table 6-1 shows a variety of approaches for the protection of the reactive side chains in trifunctional amino acids. As can be seen, different protection groups have been used in different cases for the same amino acid side chain. There are no general rules governing protection; in some cases protection is not even used, but this can lead to increased risk of unwanted side reactions.

TABLE 6-1 Protection and deprotection of trifunctional amino acids[a]

Amino acid	Protection[b]	Deprotection
Lys	Boc(1) Z(4,5)	TFA(1) HF(4,5)
Arg	NO_2(2,3,4)	H_2/Zn/HCl(2) H_2/Pd(3) HF(4)
His	Trt(2) Z(2,3)	H_2/HOAc(2) H_2/Pd(3)
Asp	Bu^t(1,3) Bzl(3,4)	TFA(1,3) H_2/Pd(3) HF(4)
Glu	Bu^t(2) Bzl(3,4)	HCl(2) H_2/Pd(3) HF(4)
Ser	Bu^t(1,2) Bzl(3,4)	TFA(1) HCl(2) HF(3,4)
Thr	Bu^t(1,2) Bzl(4)	TFA(1) HCl(2) HF(4)
Tyr	Bu^t(1,2) Bzl(4)	TFA(1) HCl(2) HF(4)
Cys	Trt(2) Bzl(2,4) Acm(5)	$Hg(OAc)_2$(2,5) Na/NH_3(2) HF(4)

[a] 1, glucagon; 2, calcitonin; 3, secretin; 4, RNAse-A; 5, RNAse-S.
[b] Bu^t, *t*-butyl; Z, carbobenzoxy; Trt, trityl; Bzl, benzyl; Acm, acetamidomethyl.

Mixed Anhydride Reaction



Carbodiimide Reaction

Figure 6-5 Activation of carboxyl groups for coupling in peptide synthesis.

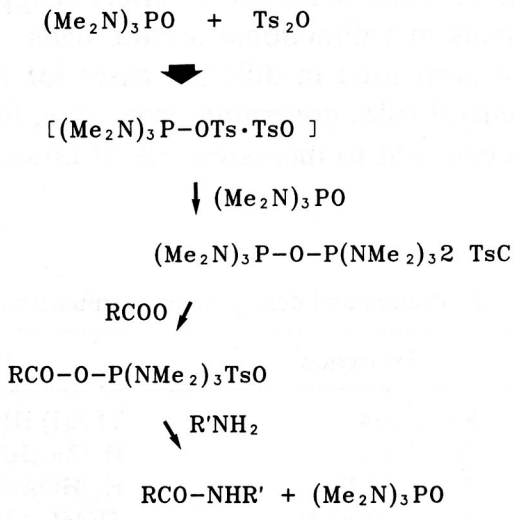


Figure 6-6 Hexamethylphosphoramide approach to coupling.

Coupling and Ligation Procedures

Several methods have been used for the sequential coupling of activated amino acids to the growing chain of a synthetic peptide. The two most common are the mixed-anhydride method and the carbodiimide method, both illustrated in Fig. 6-5. After the activation step the coupling proceeds as in Fig. 6-1. These methods, especially the mixed-anhydride method, are quite adequate for the stepwise addition of amino acids without the danger of racemization. However, anhydrides of peptides (as compared to anhydrides of single amino acids) frequently undergo racemization, making this method of coupling not very useful in fragment condensation reactions (i.e., ligation). When fragments are ligated using the mixed-anhydride approach, the peptide to be added to the existing amino terminal usually has a C-terminal glycine or proline which cannot undergo racemization on anhydride formation.

An alternative approach involves activated derivatives of hexamethylphosphoramide and is outlined in Fig. 6-6. This method produces little or no racemization and the products are readily separable. The reaction has been used successfully to couple asparagine, glutamine, methionine, tyrosine, or tryptophan containing peptides without the need for side-chain protection. When serine, threonine, or histidine is present, protection is necessary.

Attachment to Solid-Phase Support

In the solid-phase support method of peptide synthesis, several types of linkage of the growing chain to the immobile support have been used.

Ester Linkage. The phenacyl ester linkage procedure, illustrated in Fig. 6-7, is typical of this approach. The matrix copoly(styrene-divinylbenzene) is activated by bromo acetylation and then esterified with amino-protected amino acid. The resultant ester linkage is stable to acidolytic removal of the amino protecting group, but after completion of the synthesis is readily cleaved by sodium hydroxide, ammonia, or hydrazine, as indicated.

Amide Linkage. A variety of linkages have been used, including the carboxamide and sulfonamide linkages illustrated in Figs. 6-8 and 6-9, respectively. In these procedures it should be noted that the carboxamide linkage is cleaved by hydrofluoric acid (HF) to give an amide derivative, while the sulfonamide linkage gives the sodium salt of the C-terminal amino acid.

Other Linkages. Although the examples discussed use the carboxyl group of the first (i.e., C terminal) amino acid in the linkage, it is possible to immobilize via the amino group as indicated in Fig. 6-10, although synthesis must then proceed in an N-to-C direction as opposed to the more usual C-to-N manner.

Some syntheses have been reported which even use a functional side chain of an amino acid as the point of attachment to the solid support. Such procedures are somewhat esoteric and are not recommended as general approaches to peptide synthesis since they lead to increased problems in protecting the backbone amino and

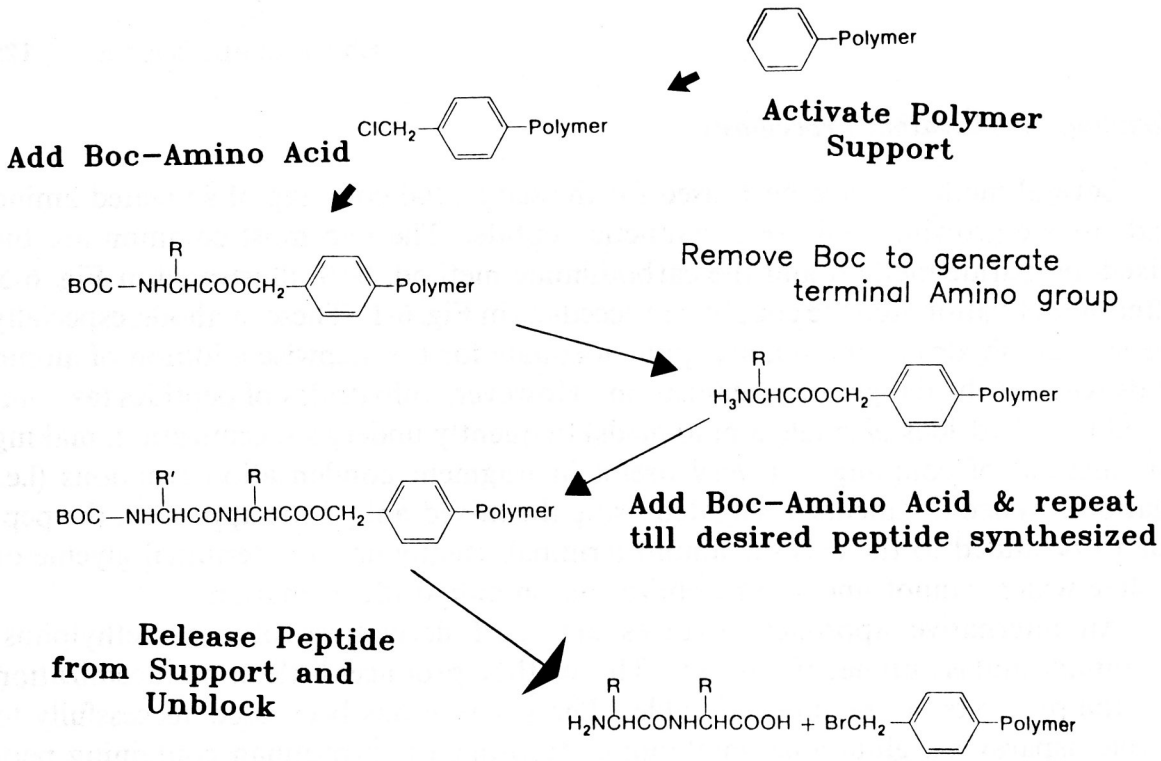


Figure 6-7 Preparation and use of bromoacetyl resin.

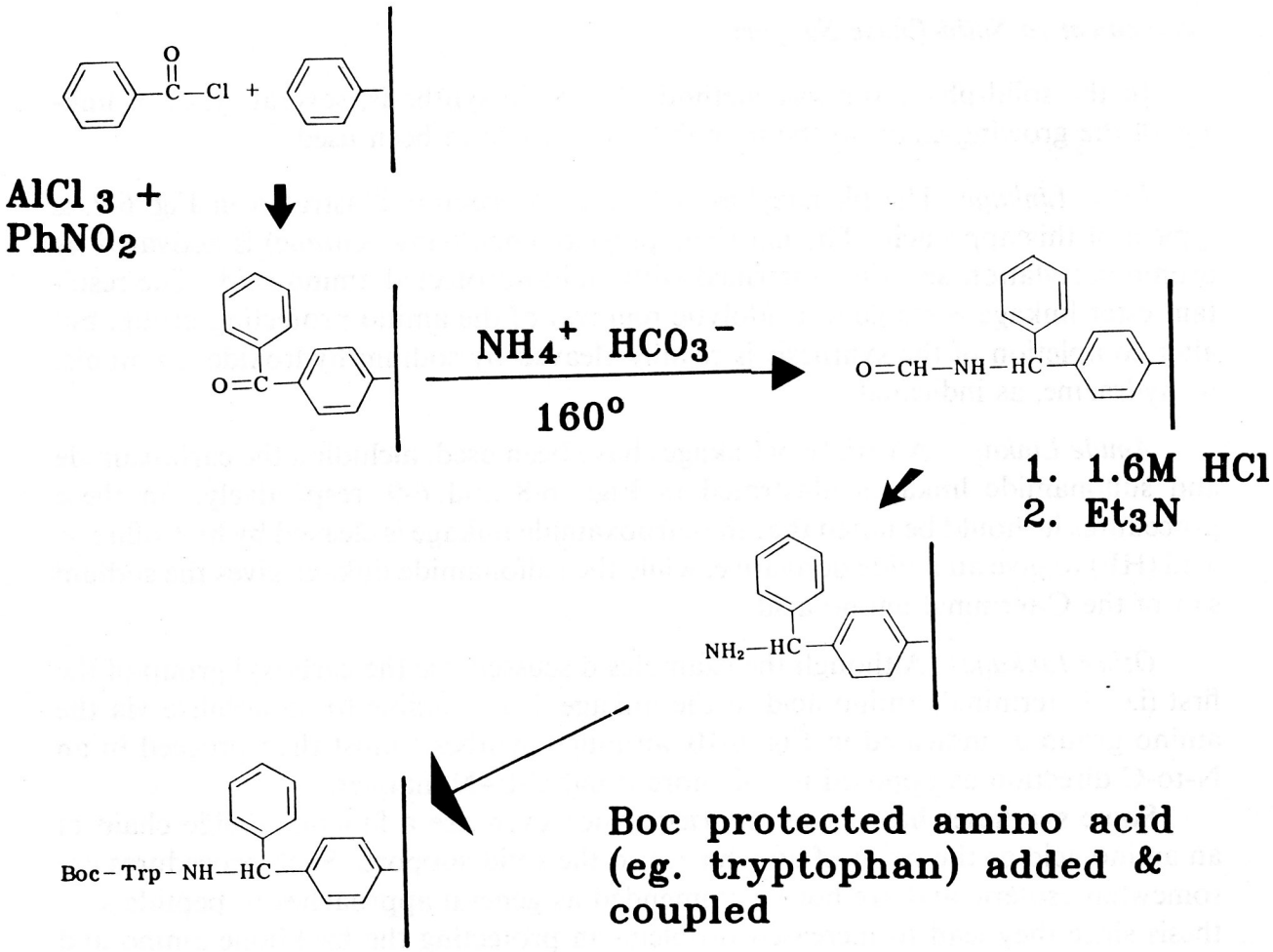


Figure 6-8 Carboxamide linkage used in benzhydrylamine support.

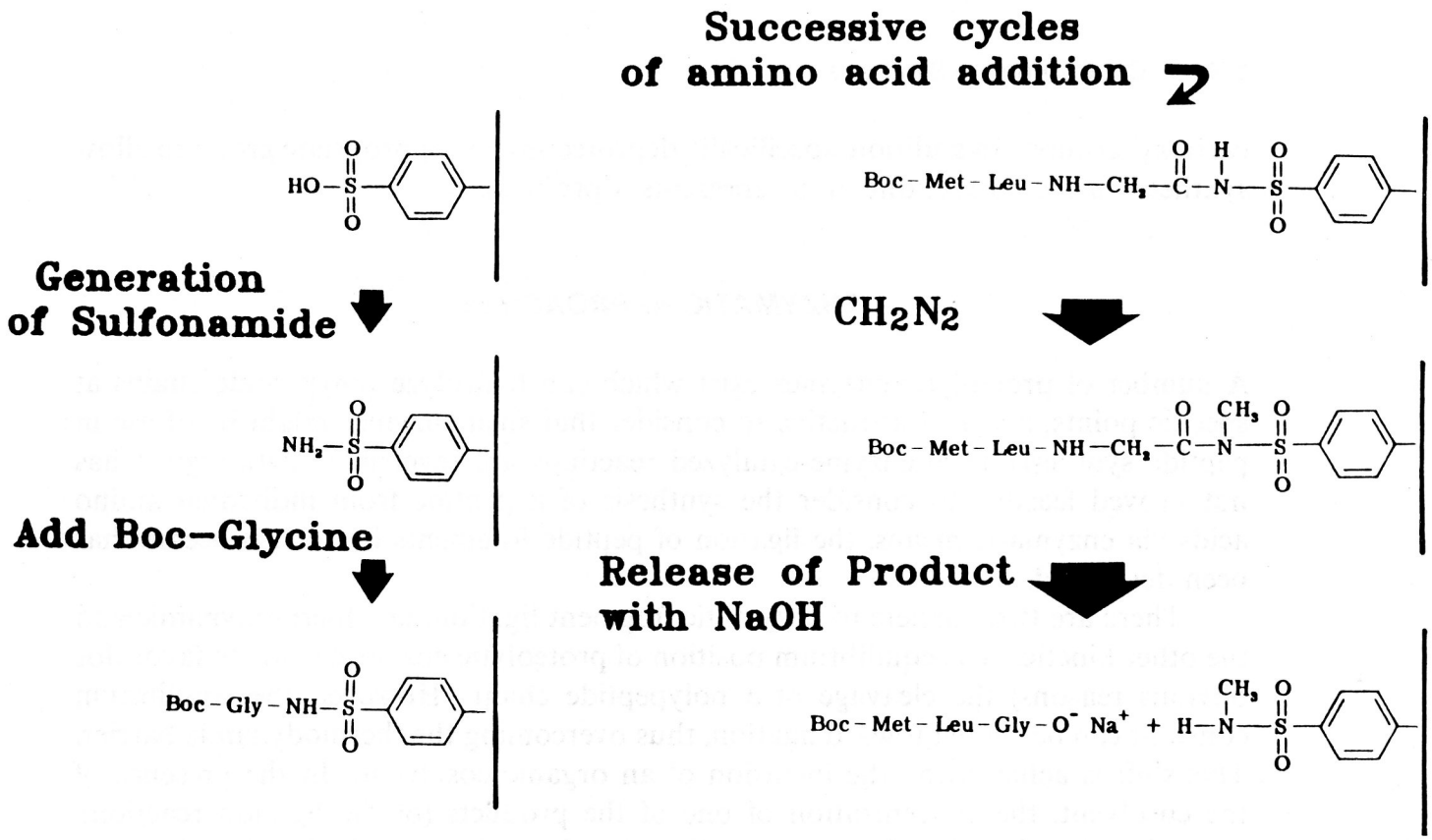


Figure 6-9 Sulfonamide linkage used in tripeptide synthesis.

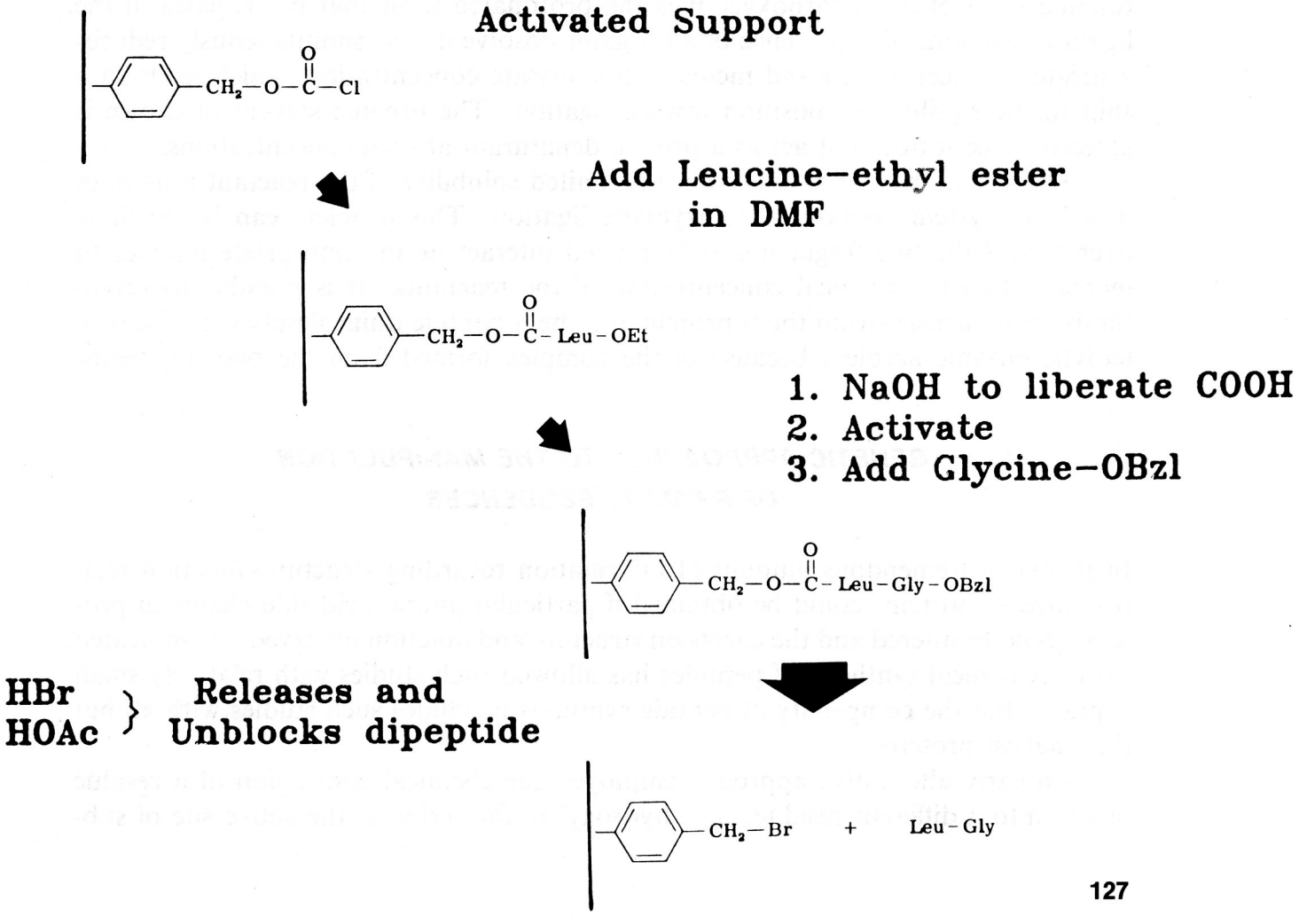


Figure 6-10 N linkage used in dipeptide synthesis.

carboxyl groups. In addition, specifically deprotecting the appropriate group to allow synthesis in the correct direction represents a problem.

ENZYMATIC APPROACHES

A number of proteolytic enzymes exist which can hydrolyze polypeptide chains at specific points, and it is attractive to consider that such enzymes might be of use in peptide synthesis since enzyme-catalyzed reactions are reversible. Although it has not proved feasible to consider the synthesis of a peptide from individual amino acids via enzymatic means, the ligation of peptide fragments by such processes has been developed.

There are two barriers to enzymatic fragment ligation, one thermodynamic and the other kinetic. The equilibrium position of proteolytic enzymes tends to favor (for obvious reasons) the cleavage of a polypeptide chain. However, the equilibrium constant can be shifted toward ligation, thus overcoming the thermodynamic barrier. This shift is achieved by the inclusion of an organic cosolvent. In the presence of the cosolvent, the concentration of one of the products (of the ligation reaction), water, is reduced, and at the same time the pK values of the terminal carboxyl groups are raised, which at a particular pH acts to raise the concentration of the protonated form of these carboxyls; it is the protonated form that participates in the ligation reaction. The presence of an organic cosolvent thus simultaneously reduces a product concentration and increases a substrate concentration, which leads to a shift in the equilibrium position toward ligation. The organic solvent of choice is glycerol since it does not act as a protein denaturant at high concentrations.

The kinetic barrier results from the limited solubility of the reactant molecules in solvent systems suitable for enzymatic ligation. This problem can be partially overcome if the two fragments to be ligated interact in an appropriate manner to increase the effective local concentration of the reactants. It is possible to resynthesize ribonuclease from the S-protein and the S-peptide quite simply with the proteolytic enzyme acrolein because of the complex formed from the two fragments.

GENETIC APPROACHES TO THE MANIPULATION OF PRIMARY SEQUENCES

In theory, a tremendous amount of information regarding structure–function relationships in proteins could be obtained if particular amino acid side chains in proteins could be altered and the effects on structure and function observed. As indicated earlier, chemical synthesis of peptides has allowed such studies with relatively small peptides, but the complexity of peptide synthesis precludes such studies with all but the smallest proteins.

An early alternative approach employed the chemical conversion of a residue in situ into a different residue. The hydroxyl of the serine at the active site of sub-

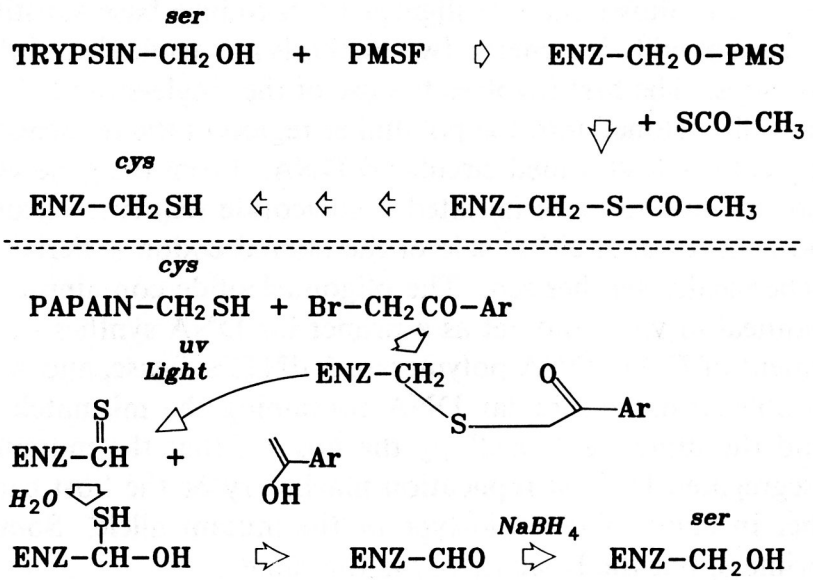


Figure 6-11 Chemical alteration of the serine or cysteine residue at the active site of trypsin and papain, respectively.

tilisin or trypsin could be converted to a sulfhydryl, and the sulfhydryl at the active site of papain converted to a serine residue. The major problem with such interconversions is specificity (actually lacking) and the limited range of alterations that can be attempted.

Although, as shown in Fig. 6-11, it has become possible to chemically alter serine or cysteine residues in active sites, other serine or cysteine residues may also be modified, and the harsh conditions employed for such alterations may lead to secondary reactions of other residues.

In an approach designed to radically alter the activity of the parent protein, some recent chemical modification work has been directed toward derivatizing, for example, the active-site cysteine in papain with flavin coenzymes. Such methods are designed to create new enzyme activities rather than provide insight into the functioning of the parent protein. In the particular case of papain it has become possible to derivatize the cysteine using 8-bromoacetyl-10-methylisoalloxazine with the resultant loss of proteolytic activity, but also with the generation of an effective oxidoreductase activity.

Genetic engineering techniques have allowed a wide range of such interconversions to be accomplished via a process known as site-directed mutagenesis. Many methods exist by which mutations can be introduced into a specific region of DNA sequence. Also, techniques are available to introduce specific mutations into a site where a restriction endonuclease can act. These are limited by a lack of specificity for changing a particular nucleotide and by the necessity for a restriction site (preferably unique) in the region one desires to change, respectively.

Oligonucleotide site-directed mutagenesis is certainly the most powerful tool available when considering sequence–function and sequence–structure relationships

in proteins because it allows the investigator to introduce base substitutions, insertions, and deletions at will. In general, two methods are available to introduce mutations at specific sites. The first involves the use of the single-stranded bacteriophage m_{13}. The gene is first cloned into the polylinker region of the replicative form and it propagates to yield single-stranded circular ϕ DNA. From the gene sequence of the protein around the residue to be mutated a nucleotide sequence encompassing two or three amino acid residues either side of the altered codon is derived and the oligonucleotide chemically synthesized. The oligonucleotide containing the mismatch is allowed to anneal in vitro and act as a primer for DNA synthesis using the large (Klenow) fragment of *E. coli* DNA polymerase I, dNTPs, ligase, and ATP to seal the circle. The double-stranded circular DNA containing the mismatch is introduced into *E. coli* and the mutation "fixed" by the host. Either the mutant can be propagated and segregated by host replication machinery or the host can "repair" the mismatch either in favor of the wild-type *or* the mutant allele. Some approaches employ two primers, but the basic theory is the same.

Oligonucleotide site-directed mutagenesis can also be accomplished using double-stranded plasmid DNA in a variety of ways. In all cases, the gene to be mutated is cloned into a plasmid vector. One method involves introducing a single-stranded nick (most often chemically) into the covalently closed circular molecule. All or part of the nicked strand is then degraded using (usually) exonuclease III, leaving a single circular strand of DNA. The same approach is then taken as with the single-stranded phage mutagenesis. An alternative to this is to denature the double-stranded plasmid DNA in the presence of two synthetic oligonucleotides that are complementary to the same strand, lie about 250 base pairs apart, and one of which contains the desired mutation. Klenow, dNTPs, ATP, and ligase are added and the resultant plasmid is introduced into an appropriate *E. coli* host. The mismatch in the plasmid is handled in a similar manner as that in the single-stranded phage.

Whether single-stranded ϕ or plasmid DNA is used in these methods, mutant clones can be detected by taking advantage of the fact that short duplex DNA containing a single mismatch is more easily denatured than a perfect match. Therefore, using the mutant synthetic oligonucleotide end-labeled with ^{32}P and conditions that only allow perfect matches, one can screen a large number of colonies in a filter hybridization for clones containing the mutated DNA.

UNANSWERED QUESTIONS

The synthesis of small peptides using either the solution or the solid-phase support approach has become almost commonplace and has allowed tremendous advances to be made in understanding structure–function relationships of the peptide hormones. The advent of genetic manipulation of primary sequences will permit similar advances to be made at the level of large proteins. The major challenge of peptide synthesis has shifted away from its original goals. (Although the techniques are now available to answer these questions, they have been applied in very few instances,

and much work remains to be done in this area). Peptide synthesis is now being used in a variety of ways that are increasingly important in examining various aspects of protein structure–function relationships. These range from the synthesis of defined short peptides which can then be crystallized to examine the influence of local primary structure on conformation, to the synthesis of peptides with the ability to serve as substrates for various post-translational modifications such as phosphorylation. The most esoteric challenge of peptide synthesis is perhaps the design and synthesis of a protein with predicted structure and function. From the academic standpoint this represents a new stage in the understanding of protein structure–function relationships and depends for its success on the principles of protein architecture described in Chaps. 9 to 11.