SOLID PHASE SYNTHESIS

Nobel lecture, 8 December, 1984

by

BRUCE MERRIFIELD

The Rockefeller University, 1230 York Avenue,
New York, N.Y. 10021-6399

The proteins, as the Greek root of their name implies, are of first rank in living systems, and their smaller relatives, the peptides, have now also been discovered to have important roles in biology. Among their members are many of the hormones, releasing factors, growth factors, ion carriers, antibiotics, toxins, and neuropeptides. My purpose today is to describe the chemical synthesis of peptides and proteins and to discuss the use of the synthetic approach to answer various biological questions.

The story begins with Emil Fischer (1) at the turn of this century when he synthesized the first peptide and coined the name. The general chemical requirements were to block the carboxyl group of one amino acid and the amino group of the second amino acid. Then, by activation of the free carboxyl group the peptide bond could be formed, and selective removal of the two protecting groups would lead to the free dipeptide. Fischer himself was never able to find a suitable reversible blocking group for the amine function, but his former student Max Bergmann, with Zervas, was successful (2). Their design of the carbobenzoxy group ushered in a new era. When I began working on the synthesis of peptides many years later this same general scheme was universally in use and was very effective, having led, for example, to the first synthesis of a peptide hormone by Du Vigneaud in 1953 (3). It soon became clear to me, however, that such syntheses were difficult and time consuming and that a new approach was needed if large numbers of peptides were required or if larger and more complex peptides were to be made.

SYNTHESIS ON A SOLID MATRIX

One day I had an idea about how the goal of a more efficient synthesis might be achieved. The plan (4) was to assemble a peptide chain in a stepwise manner while it was attached at one end to a solid support. With the growing chain covalently anchored to an insoluble matrix at all stages of the synthesis the peptide would also be completely insoluble and, furthermore, would be in a suitable physical form to permit rapid filtration and washing. Therefore, after completion of each of the synthetic reactions the mixture could be filtered and thoroughly washed to remove excess reactants and by-products. The interme-

diate peptides in the synthesis would thus be purified by a very simple, rapid procedure rather than by the usual tedious crystallization methods. When a multistep process, such as the preparation of a long polypeptide or protein is contemplated, the saving in time and effort and materials could be very large. The fact that all of the steps just described are heterogeneous reactions between a soluble reagent in the liquid phase and the growing peptide chain in the insoluble solid phase led to the introduction of the name "solid phase peptide synthesis".

The general scheme for solid phase synthesis is outlined in Fig. 1. It begins

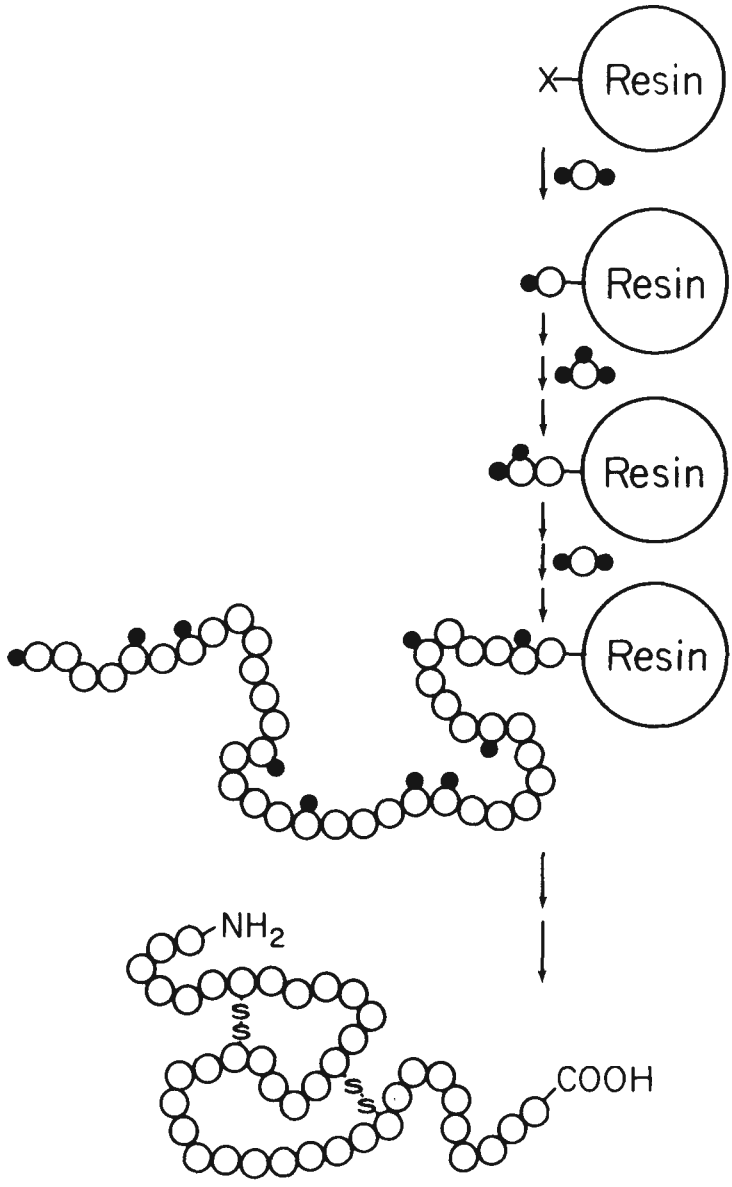


Fig. 1. The general scheme for solid phase synthesis.

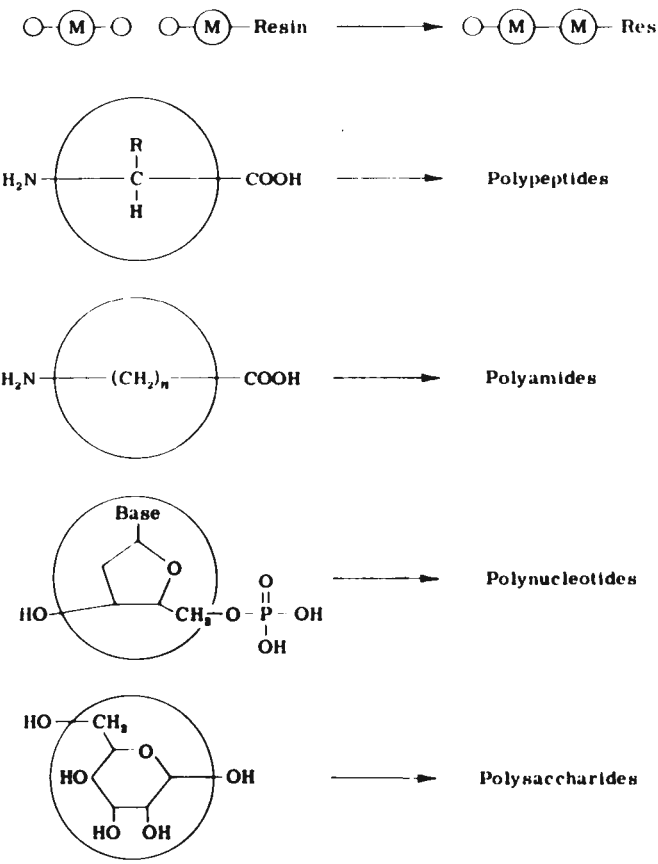


Fig. 2. Monomer units for solid phase synthesis

with an insoluble particle, indicated by the large circles, which is functionalized with a group, X. The first monomer unit, small circles, is blocked at one end and at the reactive side chain groups (black dots) and anchored to the support by a stable covalent bond. The α protecting group is then removed and the second monomer unit is added to the first by a suitable reaction. In a similar way the subsequent units are combined in a stepwise manner until the entire polymeric sequence has been assembled. Finally, the bond holding the chain to the solid support is selectively cleaved, together with the side chain protecting groups, and the product is liberated into solution. Such a system offers four main advantages: it simplifies and accelerates the multistep synthesis because it is possible to carry out all the reactions in a single reaction vessel and thereby avoid the manipulations and attendant losses involved in the repeated transfer of materials; it also avoids the large losses which normally are encountered during the isolation and purification of intermediates; it can result in high yields of final products through the use of excess reactants to force the individual reactions to completion; and it increases solvation and decreases aggregation of the intermediate products. It only remained to translate the general idea into a workable set of reactions.

Although the plan was originally conceived as a way to synthesize peptides, the general scheme does not specify the nature of the monomer units. It soon became apparent that the technique should be applicable to units other than amino acids, such as those shown in Fig. 2. We extended it to the synthesis of depsipeptides (5) and other laboratories succeeded in synthesizing polyamides (6), polynucleotides (7) and polysaccharides (8). In principle the monomer may be any bifunctional compound that can be selectively blocked at one end and activated at the other. In addition, the solid support idea can be applied to a variety of conventional reactions in organic chemistry to aid in directing the course of the reaction or in the separation of the products from reagents and by-products. It also led to the solid phase sequencing technique.

SOLID PHASE PEPTIDE SYNTHESIS

A detailed scheme for the synthesis of peptides is shown in Fig. 3. Each of the steps has been modified in many ways, but the chemistry shown here has served well and has been applied to the synthesis of large numbers of peptides (9). The carboxyterminal amino acid is blocked at the amino end by a tert-butyloxycarbonyl (Boc) group and is covalently attached to the resin support as a benzyl ester via the chloromethyl group. Side chain functional groups must

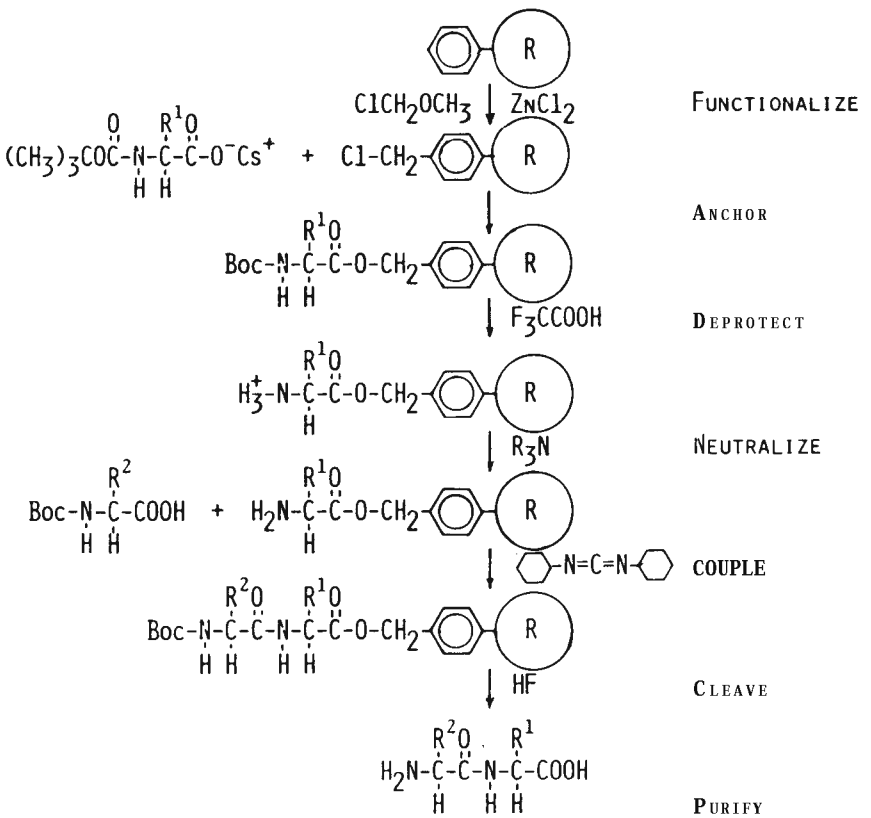


Fig. 3. A scheme for solid phase peptide synthesis.

also be blocked, usually with benzyl-based derivatives. The synthesis depends on the differential sensitivity of these two classes of protecting groups to acid, which is greater than 1000: 1. The Boc group is completely removed with 50% trifluoroacetic acid in dichloromethane, with minimal loss of the anchoring bond or of the other protecting groups. The resulting α amine salt is neutralized with a tertiary amine such as diisopropylethylamine and the free amine of the resin-bound amino acid is then ready to couple with a second Boc-amino acid. The latter must be activated for the reaction to occur. The simplest and most often used procedure is activation with dicyclohexylcarbodiimide (10) as shown, but active esters (11), anhydrides (12), and many other activated derivatives have been successfully applied. All of these reactions are carried out under non-aqueous conditions in organic solvents that swell the resin and accelerate the rates. Dichloromethane and dimethylformamide are the solvents of choice.

To extend the peptide chain the deprotection, neutralization and coupling steps are repeated for each of the succeeding amino acids until the desired sequence has been assembled. Finally, the completed peptide is deprotected and cleaved from the solid support. With the chemistry described here, this is accomplished by treatment with a strong anhydrous acid such as HF (13). The free peptide is then purified by suitable procedures.

It is very important that the repetitive steps proceed rapidly, in high yields, and with minimal side reactions in order to prevent the accumulation of excessive amounts of by-products. Much of our effort has been directed toward developing and evaluating these requirements.

SOLID PHASE NUCLEOTIDE SYNTHESIS

Similar schemes for the solid phase synthesis of oligonucleotides have now been developed which are rapid and give relatively high yields (14, 15). They employ protected nucleotides as monomer units and make use of either phosphotriester or phosphite triester chemistry. One such procedure is outlined in Figure 4. The resin is first functionalized with an aminomethyl group and the nucleotide derivative is coupled to it, through a spacer, by a stable amide bond. In this example the 5′hydroxyl is esterified to the spacer and the 3′hydroxyl is temporarily blocked with a dimethoxytrityl (DMT) group. The latter is removed with acid or $ZnBr_2$ and the chain is extended at the 3′ end by coupling with the next protected nucleotide by activation with 1-(mesitylene-2-sulfonyl)-3-nitro-1, 2, 4-triazolide (MSNT). The completed oligonucleotide is finally cleaved from the solid support and deprotected by treatment with a base such as NH_4OH or tetramethylguanidine and then with hot acetic acid. The products are readily purified by ion exchange chromatography or by electrophoresis where the desired product always has the greatest negative charge. I will not deal further with polynucleotides and their use in site directed mutagenesis or with synthetic genes in this presentation but will concentrate instead on peptides and proteins.

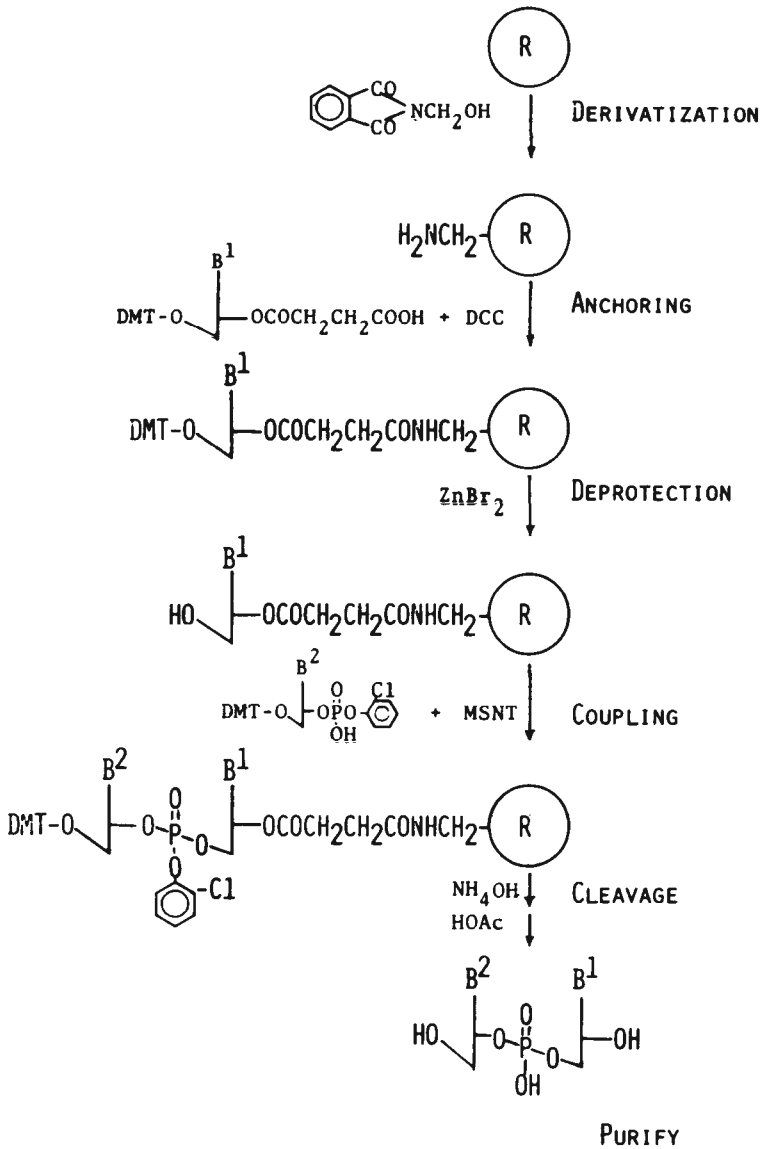
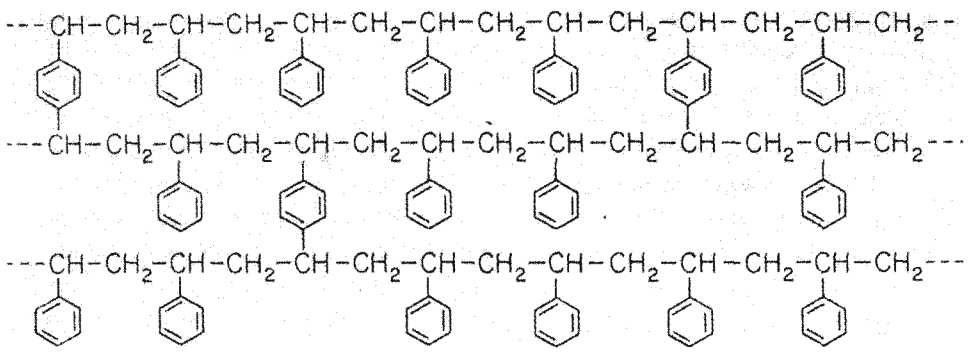


Fig. 4. A scheme for solid phase nucleotide synthesis.

THE SUPPORT

The first requirement for the development of solid phase synthesis was a suitable support. After examination of many potential supports it was found that the most satisfactory one was a gel prepared by suspension copolymerization of styrene and 1% of divinylbenzene as crosslinking agent (4). The resulting spherical beads (Fig. 5) are about 50 μm in diameter when dry, but in organic solvents such as dichloromethane they swell to 5 or 6 times their original volume. Furthermore, as peptide chains grow the dry volume increases to accomodate the added mass and, most importantly, the swollen volume



Styrene–divinylbenzene Copolymer

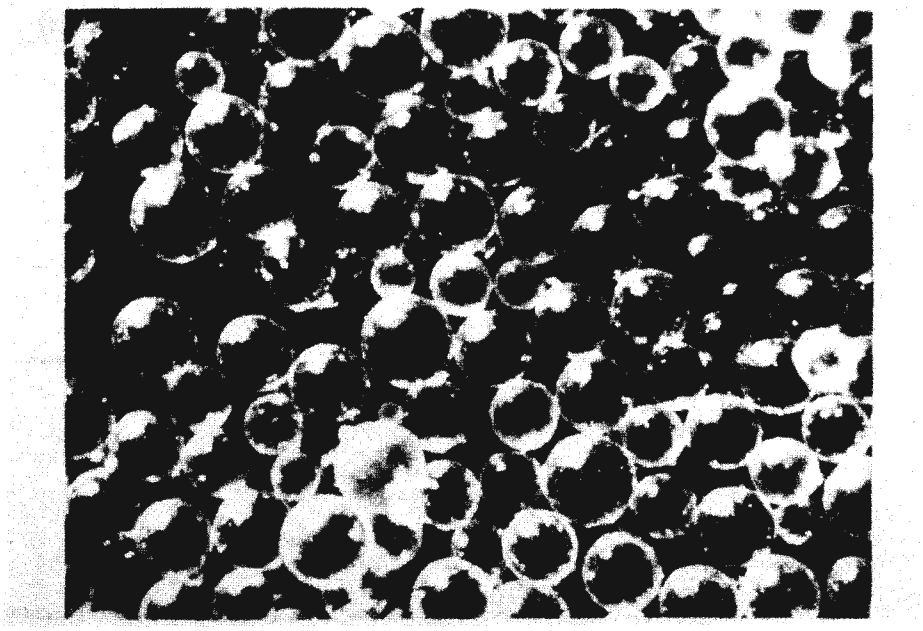


Fig. 5. Copoly(styrene-1%-divinylbenzene) resin.

continues to increase. Values up to 25 fold have been measured and calculations indicate that the maximum expansion should be about 200 fold (16). This means that the polystyrene matrix and the pendant peptide are highly solvated during the chemical reactions and are freely accessible to diffusing reagents. The reactions occur not only at the surface of the bead but, in major part, within the interior of the crosslinked polymeric matrix. This could be demonstrated by autoradiography of a cross section of a bead containing a synthetic tritium-labeled peptide (17). At this resolution the silver grains were located uniformly throughout the bead, although the distribution is not known at the molecular level. Because of the solvation and swelling of the beads, the reac-

tions are very fast, with half times in the order of seconds for both the coupling and the deprotection steps. Current efforts to evaluate the effects of mass transfer and diffusion indicate that they are not rate limiting. We believe the solid matrix not only does not have detrimental effects on the synthesis but actually has beneficial effects in certain instances. One of the well recognized difficulties with the classical synthesis in homogeneous solution is insolubility of some intermediates. This problem can be overcome in many cases by the use of solid supports, where the peptide chain and the lightly crosslinked polymer chain become intimately mixed and exert a mutual solvating effect on one another. It becomes thermodynamically less favorable for the peptide to self aggregate and it remains available for reaction. For this to occur the solvated state of the bound peptide needs only to be favorable relative to the amorphous unsolvated state within the peptide-resin matrix (16). Similar solubilizing properties of linear polymers for covalently attached components are known, but the effect will be greater for a lightly cross-linked polymer network. The phenomenon can be illustrated by the synthesis of oligoisoleucines (18). The standard solution synthesis failed after the tetrapeptide stage because of aggregation and insolubility, whereas the chain could be extended up to 8 residues on linear polyethyleneglycol. A solid phase synthesis proceeded smoothly at least as far as the dodecamer, where the experiment was stopped. There is very significant polymer chain motion in these crosslinked polystyrene resins. Both ^{1}H and ^{13}C NMR measurements (19) have shown that the motional rates for the aromatic groups and the aliphatic backbone atoms in CH_2Cl_2 are high and equivalent to linear soluble polystyrene ($\tau_c 10^{-8}$ sec). The α carbon ^{13}C resonances of model resin-supported peptides were as sharp as the solvent peak in CH_2Cl_2 or dimethylformamide and similar to small molecules in solution (τ_c 10^{-10} sec). A variety of chemical experiments also have shown polymer flexibility. For example short resin-bound peptides that were too far apart on average to reach one another if the resin were rigid could be shown to react to the extent of 99.5% indicating considerable motion of the polystyrene segments within the matrix (20).

Many other solid supports have also been examined and several have been satisfactory for peptide synthesis. These have included polymethylmethacrylate, polysaccharides, phenolic resins, silica, porous glass, and polyacrylamides, but only the latter have seen widespread use (21). Comparative studies with polystyrene and polyacrylamide have shown that they can be equally effective, even with difficult peptides.

AUTOMATION

The ability to purify after each reaction by simple filtration and washing and the fact that all reactions could be conducted within a single reaction vessel appeared to lend themselves ideally to a mechanized and automated process. Initially, a simple manually operated apparatus was constructed (Fig. 6). This system was first used to work out the methodology and to synthesize bradykinin (22) angiotensin (23) oxytocin (24), and many other small peptides. In order to accelerate the process we undertook the design and construction (25)

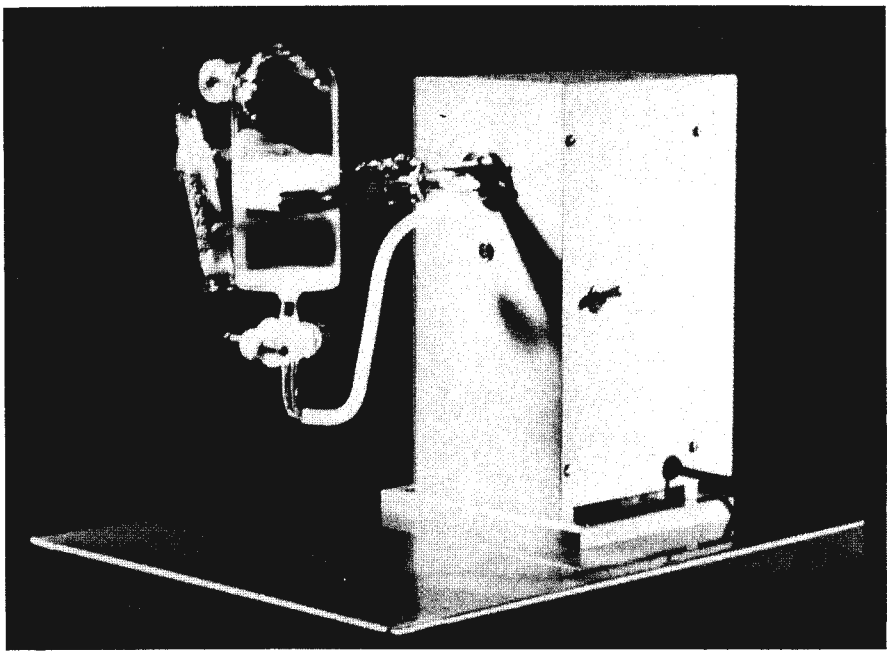


Fig. 6. A manually operated synthesis apparatus

of the automated instrument shown in Fig. 7. The essential features were the reaction vessel, containing the resin with its growing peptide chain, and the necessary plumbing to enable the appropriate solvents and reagents to be pumped in, mixed, and removed in the proper sequence. These mechanical events were under the control of a simple stepping drum programmer and a set of timers. In the past few years many commercial instruments have been constructed in several countries. They differ considerably in detail, particularly in the sophistication of the electronic program mechanisms but are designed to carry out the same chemistry.

THE SYNTHESIS OF RIBONUCLEASE A

The idea of chemically synthesizing an enzyme must have occurred to many people over the years. There was a time when such a thought would have been unacceptable even on philosophical grounds, but from the period when enzymes were shown to be proteins and proteins were shown to be discrete organic molecules it was a goal that chemists could begin to think about. If an enzyme could be made in the laboratory, then it should become possible to learn new things about how these large and very complex molecules function. Specific changes could be made in their structures that could not be made readily by altering the native protein and data should be forthcoming that would supplement the information already obtained from the natural enzymes themselves. In this regard, a quotation from Fischer in 1906 (26) is pertinent: "Whereas cautious professional colleagues fear that a rational study of this class of compounds [proteins], because of their complicated structure and their

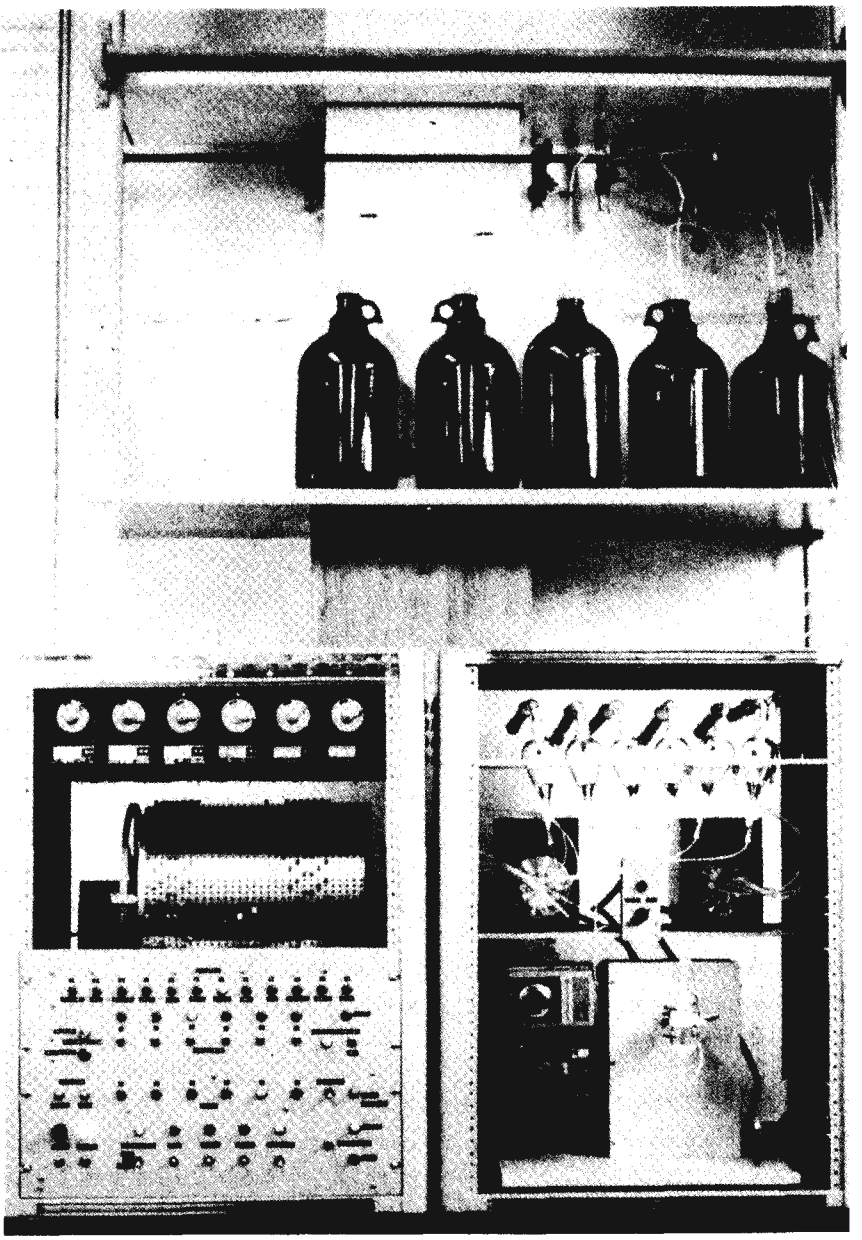


Fig. 7. An automated peptide synthesizer.

highly inconvenient physical characteristics, would today still uncover insurmountable difficulties, other optimistically endowed observers, among which I will count myself, are inclined to the view that an attempt should at least be made to besiege this virgin fortress with all the expedients of the present; because only through this hazardous affair can the limitations of the ability of our methods be ascertained."

With the development of solid phase peptide synthesis and its automation the time seemed right to attempt the total synthesis of an enzyme. Dr. Bernd Gutte and I selected bovine pancreatic ribonuclease A because it was a small stable protein of known amino acid sequence (27), and the three dimensional structure was known from X-ray diffraction studies (28). Much of the detailed mechanism by which this enzyme hydrolyzes and depolymerizes ribonucleic acid was also known. The purpose of a chemical synthesis of this 124-residue molecule was, first, simply to demonstrate that a protein with the high catalytic activity and specificity of a naturally occurring enzyme could be synthesized in the laboratory. For the long range the more important purpose was to provide a new approach to the study of enzymes. We believed it should be possible to modify the structure and to alter the activity and the substrate specificity of the enzyme.

The synthesis (29) was carried out on a copoly(styrene-1%-divinylbenzene)-resin support using the general methods described above. The C-terminal Boc-Val was anchored to the solid matrix by a benzyl ester bond, the usual benzyl-based side chain protecting groups were used and the Boc group provided the reversible N^{α} protection. The deprotection steps were with trifluoroacetic acid and the coupling reactions were with dicyclohexylcarbodiimide activation. Fig. 8 shows the final protected derivative of ribonuclease. It contained a total of 67 side chain protecting groups and had a molecular weight of 19,791. The synthesis is summarized in Table I. The overall yield after several purification procedures was about 3% based on the original amount of valine attached to the resin. There was a large (83%) loss of chains during the assembly of the peptide chain due to partial instability of the anchoring bond, and the accumulated losses during HF cleavage from the resin and the purification steps were another 80%. The crude cleaved product was air oxidized to form the four disullide bonds and the monomer fraction was isolated by gel filtration. The monomers with incorrect disullide pairing or incorrect folding were digested by trypsin and the small fragments were separated from the stable protein with the correct structure. An ammonium sulfate fractionation gave the final purified enzyme possessing approximately 80% specific activity compared with native ribonuclease A. We could not claim that our product was completely pure or that the synthesis constituted a structure proof for RNase, only that the molecule showed a close chemical and physical resemblance to the native protein and that it was a true enzyme. The chemical and physical comparisons were based on amino acid analysis, enzyme digestions, peptide maps, paper electrophoresis, gel filtration, ion-exchange chromatography, and antibody neutralization. At that time we did not have HPLC or an affinity chromatography system.

Table II summarizes the activity data at various stages of purification of the synthetic enzyme. Both the specific activity and the total number of units of RNase increased as the purification proceeded, indicating either that inhibitory impurities were being removed or that the molecule was gradually refolding into a conformation that more closely resembled the native structure. The substrate specificity of the synthetic enzyme was consistent with that to be

Fig. 8. Protected-ribonuclease-resin.

Summary of Ribonuclease A Synthesis

Stage of Synthesis	Overall Yield	
	mg	%
Boc-Val-Resin	2000	100
Deprotect, Neutralize, Couple } Repeat 123 times		
17% ↓		
Protected RNase-Resin	3430	17
Cleave and Deprotect HF		
71% ↓		
Crude RNase $(SH)_8$	697	12
Sephadex G-75		
53% ↓		
RNase A (monomer fraction)	373	6.4
69% Trypsin digestion, Sephadex G-50 ↓		
RNase A (Trypsin resistant fraction)	256	4.4
66% $(NH_4)_2SO_4$ fractionation ↓		
RNase A	169	2.9

Table I

Ribonuclease A Activity

Purification stage	Specific activity (%)	Total activity $\frac{\text{mg RNase}}{\text{2 g resin}}$
HF cleavage	2	14
Sephadex G-75	9	33
IRC-50	13	53
Trypsin	61	156
$(NH_4)_2SO_4$	78	132

Table II

expected for RNase A: it was able to cleave both large (RNA) and small (C>p) substrates and therefore to catalyze both the transphosphorylation and the hydrolysis steps; it was specific for D-ribose instead of D-deoxyribose and for a pyrimidine instead of a purine at the 3′ position of the phosphodiester substrate (Table III). The Km values toward RNA were also the same for the natural and synthetic enzymes.

The purified RNase A was compared on a CM-cellulose column with natural RNase A and with reduced and reoxidized natural RNase A. They were identical by this criterion, which was the one first used by White (30) to show that RNase A after reduction and reoxidation of the disulfide bonds was indistinguishable from the native enzyme. His was the demonstration that led to the hypothesis that the primary structure of the protein determined its tertiary structure (31). Our synthesis provided a new kind of evidence for this hypothesis. The fact that the only information put into the synthesis was the linear sequence means that the primary structure must be sufficient to direct the final folding of the molecule into its active tertiary structure. The synthesis of an active enzyme containing no substituents except amino acids also provided a new proof for the now well established belief that enzymatic activity can be attributed to a simple protein containing no other components.

Table III. Substrate Specificity of Synthetic Ribonuclease A

Substrate		Activity
Ribonucleic acid	(RNA)	78
Deoxyribonucleic acid	(DNA)	0
2',3'-Cyclic-cytidine phosphate	(C>p)	65
2',3'-Cyclic-guanosine phosphate	(G>p)	0
5'-(3'-guanylyl)-cytidylic acid	(GpCp)	0
5'-(3'-adenylyl)-adenylic acid	(ApAp)	0

STRUCTURE-FUNCTION STUDIES ON RIBONUCLEASE

The synthesis of ribonuclease provided answers to several fundamental questions and laid the foundation for new studies on the relation of structure to function in the enzyme. The classic S-peptide/S-protein system discovered by Richards (32) provided an ideal way to study such relations because a small peptide (residues 1-20) and a large protein component (residues 21-124) could be combined noncovalently with regeneration of nearly full enzymatic activity. The extensive work from the Hofmann (33) and Scoffone (34) laboratories on the synthesis of the S-peptide and its combination with the natural S-protein had already provided a great amount of information about the role of individual residues in the N-terminal region of the enzyme. We undertook to study this region of RNase by total synthesis (Fig. 9). During the initial synthesis we removed samples after coupling Cys^{26} and again after Ser^{21} and in that way prepared synthetic S-protein (21-124) and S-protein (26-124). The

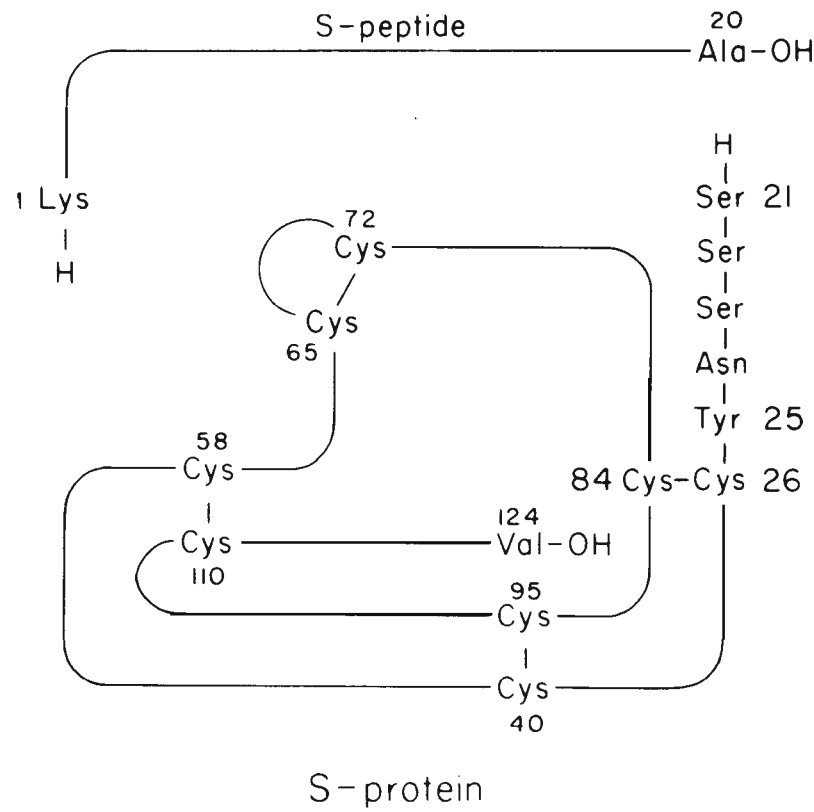


Fig. 9. The S-peptide/S-protein system

partially purified proteins were reduced at their 4 disulfide bonds, each mixed with synthetic S-peptide, reoxidized and assayed for enzymatic activity. Each of the crude mixtures was found to have as much activity as the product derived from native S-protein by the same treatment. From these data it was concluded first, that S-protein had been synthesized and second, that the live residues 2 1-25 were definitely not necessary for the binding and reactivation to occur. Earlier X-ray data (35) had predicted that the serines at positions 21, 22 and 23 would probably not be necessary, but Asn[24] and Tyr[25] appeared to be involved in a total of 5 hydrogen bonds in RNase S and it was expected that they might be necessary for the formation of an active complex. The synthetic studies showed that they were not.

Several years earlier I had been interested in the question ofwhether or not a peptide from the carboxyl end of RNase might function in a manner similar to that of the S-peptide at the amino end. Consequently, the RNase 111- 124 tetradecapeptide was synthesized and purified. RNase was inactivated by carboxymethylation of the imidazole ring of His[119] Attempts to reactivate the enzyme by addition of the synthetic peptide were uniformly unsuccessful. Somewhat later in a separate study Lin et al. (36) succeeded in preparing a series of shortened RNases. They made RNase 1-120, RNase 1-119 and

RNase 1-118 by enzymatic digestion. When the synthetic peptide 111-124 was assayed in the presence of these inactive proteins, high enzymatic activity was generated (37) and it became clear that a system existed at the C-terminus of RNase that was similar to the one at the N-terminus.

We then made the interesting discovery that the C-terminal peptide 111-124 containing His^{119}, the N-terminal peptide 1-20 containing His^{12}, and the central protein component 21-118 containing Lys^{41} could be mixed together non-covalently and ribonuclease activity would be generated. Therefore, three components each containing one of the known residues required for enzymatic activity could bind together and form the specific well ordered structure necessary for substrate binding and catalytic activity.

A series of synthetic studies was then undertaken to define the roles of some of the individual residues in the C-terminal region. These can be summarized and discussed by referring to Fig. 10. When peptides shorter than 111-124 were prepared and combined with RNase 1-118, both the binding constant and the activity were progressively decreased and peptide 117-124 was inactive, indicating that each residue contributed to the binding energy (38). However, even in the complex 1-118 + 116-124 there were 3 overlapping residues. It was then found that the complex 1-115 + 116-124, in which there were no overlapping residues, had a binding constant 100 times larger (39). In these experiments it could also be shown that Tyr^{115} was not necessary for enzymatic activity.

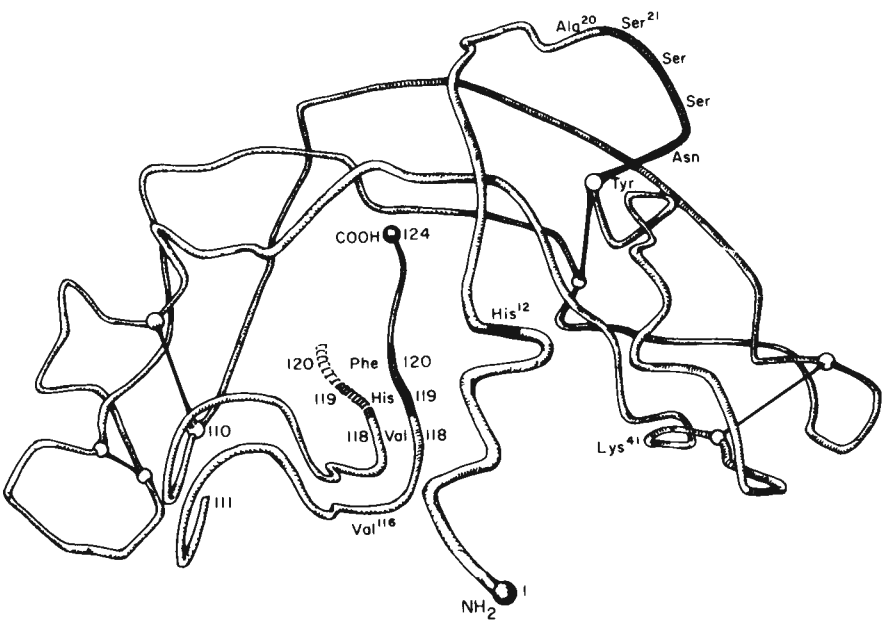


Fig. 10. A 3-dimensional representation of ribonuclease fragments 1-20, 21-118 and 111-124 summarizing the synthetic structure-function studies.

Phenylalanine-120 was shown to be important in stabilizing the ribonuclease structure by transition temperature studies and to interact with the pyrimidine substrate by X-ray and NMR studies. Our synthetic analog work on the 1-118 + 111-124 system showed that replacement of Phe^{120} by Leu^{120} or Ile^{120} reduced the binding by 5 and 17 fold and reduced the maximum enzymatic activity to 10 % and 5 % respectively, indicating that the aromatic side chain of phenylalanine was of considerable importance in binding the peptide to the protein (40). It could only partially be replaced by a hydrophobic aliphatic chain, indicating an inexact alignment of the catalytic site. The small residue Ala^{120} and the bulky aromatic residue Trp^{120} were inactive. Replacement of Phe^{120} by an aromatic residue of similar size, Tyr^{120}, in the 111-124 peptide gave a complex with 1-118 that was fully active toward C>p as substrate and 190% as active toward U>p (41). A semisynthetic enzyme with enhanced activity was a novel finding. Km and Ki data led to the conclusion that Phe^{120} does not have a unique role in the binding of substrate but is important for stabilizing the peptide-protein complex and the native enzyme itself. Nevertheless, the presence of substrate increased the binding constant between 1-118 and 111-124 by a factor of 50.

Similar experiments with the aspartic acid residue at position 121 have shown that it can be replaced partially (~20%) by glutamic acid, but the Asn^{121} and Ala^{121} analogs did not show measurable binding. Removal of Val^{124} from RNase A does not affect the enzymatic activity and removal of Val^{124} from S-protein does not reduce the activity of the complex with S-peptide. In contrast, omission of Val^{124} from the C-terminal tetradecapeptide produced an essentially inactive complex with RNase 1-118, indicating an important hydrophobic interaction necessary for peptide-protein binding. The smaller aliphatic residue Ala^{124} could only restore half of the binding energy (41).

X-ray data (42) indicate that the uracil and cytosine residues of RNA and the cyclic nucleotides probably bind to ribonuclease through the series of hydrogen bonds shown in Fig. 11. For uracil the hydroxyl of Thr^{45} is a hydrogen acceptor and for cytosine it is a hydrogen donor. Conversely the hydroxyl of Ser^{123} is a donor for uracil and an acceptor for cytosine. We reasoned that if these two hydroxyls were blocked as methyl ethers they could only be hydrogen acceptors and if replaced by Ala they could be neither donor nor acceptor. A suitable combination of these residues in replacement analogs might, therefore, lead to a synthetic ribonuclease with altered substrate specificity. Such analogs have been made for Ser^{123} (43).

The tetradecapeptide containing Ala^{123} gave a complex with RNase 1-118 that showed appreciable selectivity for substrates containing cytosine relative to those containing uracil (either the 2',3' cyclic nucleotides or polynucleotides) (Table IV). Replacement with 0-methylserine did not result in differential substrate specificity. It was concluded that a hydrogen bond between the hydroxyl of Ser^{123} and the C^4 amino group of cytosine is not important for substrate binding and catalytic activity, but that the hydrogen bond between the hydroxyl of Ser^{123} and the C^4 carbonyl of uracil contributes significantly to the binding and activity; when Ser is replaced by Ala the H-bond is absent and

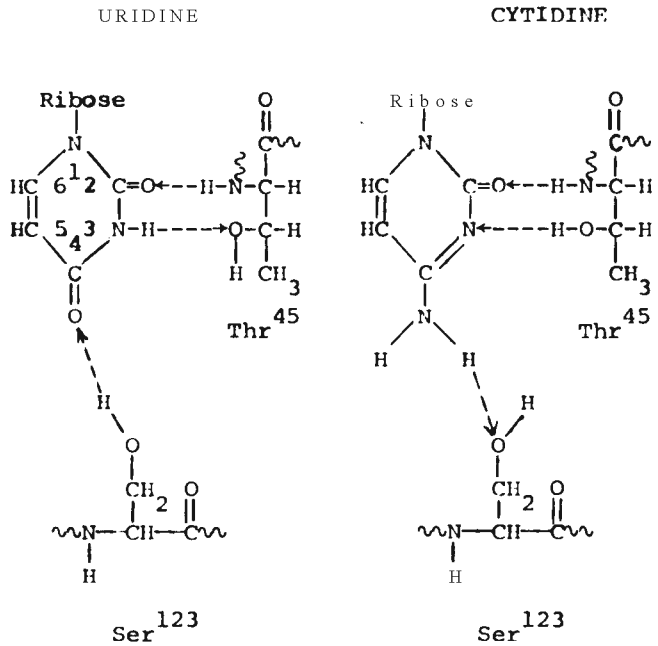


Fig. II. Proposed hydrogen bonding of uracil and cytosine substrates to ribonuclease.

the activity is reduced. The corresponding studies with replacement of Thr^{45} by Ala^{45} and $Ser(Me)^{45}$ involve total synthesis of the enzyme and these much more difficult experiments have not yet been completed. We believe that the substrate binding at Thr^{45} is much tighter than at Ser^{123} and that changes at this residue will lead to much greater substrate selectivity.

Table IV. Substrate Selectivity of (Ala[123])-RNase Complex

Enzyme	Selectivity (ks/Km) C>p / U>p
RNase A (natural)	4.6
[Ser^{123}]-RNase 111−124 + RNase 1−118	5.0
[Ala^{123}]-RNase 111−124 + RNase 1−118	19

RECENT IMPROVEMENTS IN SOLID PHASE PEPTIDE SYNTHESIS

Although the earlier solid phase chemistry was very useful for these studies on ribonuclease, it was clear that there was a need for improvement in several areas. One was the mode of attachment of the peptide to the resin. If the strategy of differential stability toward acid for the N^{α} and C^{α} groups was to be

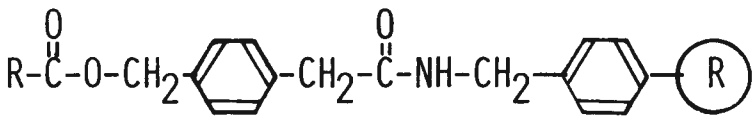


Fig. 12. Acyloxymethyl-Pam-resin

continued, a more acid stable anchoring bond was needed. We predicted that the insertion of an acetamidomethyl group between the benzyl ester and the polystyrene matrix would increase the stability of the benzyl ester to trifluoroacetic acid by a factor of approximately 25 to 400 times. When such a linkage was finally constructed it was found to be 100 times more stable (44). A new synthesis of aminomethyl-resin was first developed in which N-hydroxymethylphthalimide and polystyrene resin were reacted under acid catalysis with F_3CSO_3H, HF, or S_nCl_4(45). This product was then coupled with a derivative of the C-terminal amino acid. Thus, **N^{α}-Boc-aminoacyloxymethylphenylacetic** acid was prepared and activated with dicyclohexylcarbodiimide for the reaction. The product was the acyloxymethylphenylacetamidomethylcopoly(styrene-1%-divinylbenzene) resin (acyloxymethyl-Pam-resin) (Fig. 12). This new preparation has the advantages that it is more acid stable, and it is made from purified, well characterized intermediates, which give a cleaner product with fewer side reactions. It is free of chloromethyl groups that can give rise to quaternization and ion exchange reactions and is free of hydroxyl groups that can lead to peptide chain terminations via trifluoroacetylation (46).

An alternative protecting group strategy is to make use of an orthogonal system (47) in which the N^{α}, C^{α}, and side chain groups represent three different classes of compounds that are cleavable by three different kinds of reactions. In that way any one of the functional groups can be selectively removed in the presence of the other two. Figure 13 illustrates such a system in which the anchoring o-nitrobenzyl ester is photolabile but stable to acid or nucleophiles, the side chain groups are based on tert-butyl derivatives that are very acid labile but stable to light or nucleophiles, and the N^{α} protecting group is the dithiasuccinoyl group which is removed by nucleophilic thiols but is

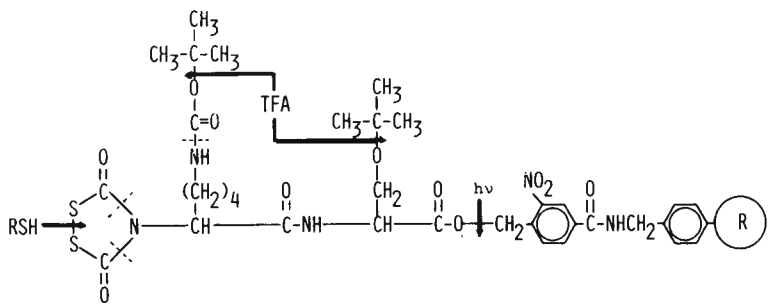


Fig. 13. An orthogonal protecting group scheme

stable to acid and photolysis. This scheme has recently been put to the test and found to give excellent results (48).

Anhydrous hydrogen fluoride, the usual cleavage reagent for solid phase peptide synthesis, is a very strong acid (H_0-10.8) and is known to promote a number of side reactions. In particular it leads to the formation of carbonium ions which then can alkylate tyrosine, tryptophan, methionine and cysteine residues of the peptide. In addition, HF can protonate and dehydrate the side chain carboxyl of glutamic acid residues with formation of the very reactive acylium ion, which has been shown to acylate the aromatic rings of anisole and other scavengers present in the mixture. Activated glutamic residues can also form pyrrolidone (pyroglutamic)containing products. Aspartyl residues can close in HF to the aspartimide derivative and subsequently open to produce β–aspartyl residues. All of these undesired reactions result from the S_N1 mechanism of the cleavage reaction under the usual conditions (90% HF + 10% anisole, 0° C, 1 hr). We reasoned that if conditions could be found that would change the reactions to an S_N2 mechanism in which the acidolysis is aided by a nucleophile and carbocation is never formed (Fig. 14) it should be possible to minimize or avoid these problems. Dr. James Tam and W.F. Heath, a graduate student, have succeeded in developing such conditions and in demonstrating marked improvements in solid phase peptide synthesis (49).

The problem was to find a suitable weak base which would reduce the acidity function of the HF but which would remain largely unprotonated and nucleophilic under the resulting acidic conditions. It should be a weaker base than the groups to be cleaved so that they would be largely protonated under the same conditions. Dimethylsulfide (DMS) was found to be an ideal base for this purpose. It has a pKa of -6.8 compared with values of -2 to -5 for the benzyl ethers, esters and carbamates to be cleaved. It is a good solvent for HF and it is volatile and easily removed from the reaction mixture. A 1: 1 molar

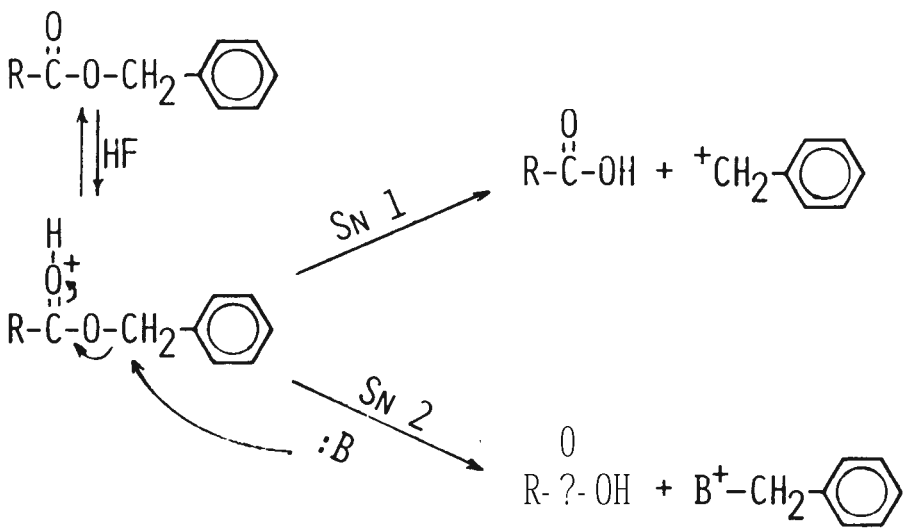


Fig. 14. The S_N1 and S_N2 acidolysis mechanisms.

mixture of HF and DMS (1:3 by volume) was determined by Hammett indicators to have an He between -4.6 and -5.2. The mechanisms of removal of various benzyl-based protecting groups by HF/DMS mixtures were tested by kinetic and product analysis experiments. Based on earlier work with H_2SO_4 hydrolysis of alkyl acetates (50), a sharp upward break in the rate constant was expected when the acid concentration was increased. At the break point the mechanism changed from S_N2 to S_N1. A similar change was found in the cleavage of O-benzyl serine by HF/DMS mixtures; above 50% HF by volume the rate increased rapidly, indicating the change in mechanism. Product analysis for the deprotection of tyrosine benzyl ether as a function of HF concentration is shown in Fig. 15. Above 15% HF in DMS the yield (after 1 hr, 0° C) of tyrosine was quantitative and the other product was the benzyldimethylsulfonium salt. In the range of 40-50% HF the amount of sulfonium salt began to decrease and the level of the undesirable byproduct, 3-benzyltyrosine, increased. Again, there was a change from the S_N2 to the S_N1 mechanism around 40-50% HF. The reactions were accelerated in the presence of 5-10% of cresol. We selected 25% HF/65% DMS/10% cresol as the best reagent and refer to it as "low HF".

This reagent was also effective in preventing acylium ion formation in glutamyl and aspartyl peptides and avoided the acylation and imide side reactions. It was also found to be very effective in converting methionine

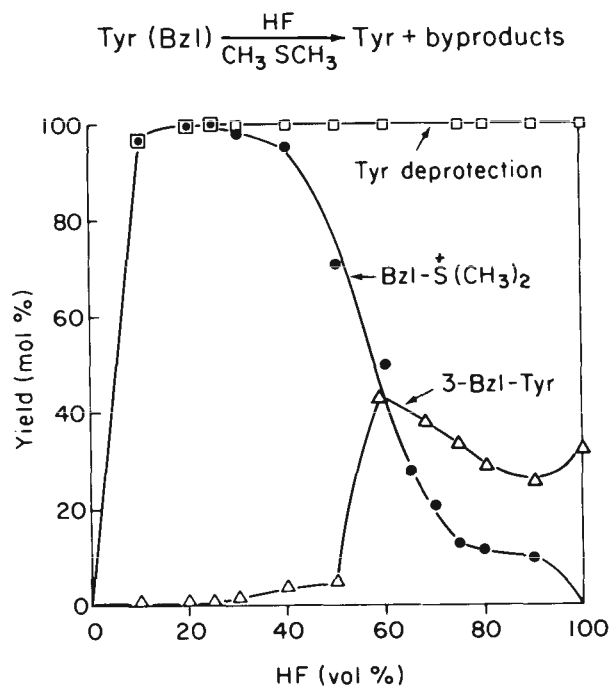


Fig. 15. Product analysis for the deprotection of tyrosine benzyl ether in mixtures of HF and dimethylsulfide.

sulfoxide to methionine, with formation of dimethylsulfoxide as coproduct. Furthermore, in the presence of 5% of thiol such as thiocresol a nearly quantitative removal of the formyl protecting group from the indole nitrogen of tryptophan was possible. The byproduct was $HC(SR)_3$. The last two reactions do not occur at high (90%) HF due to protonation of the reagents.

The derivatives Arg(Tos), Arg(NO_2), Cys(4-MeBzl), or Asp(OcHex) will not be deprotected under the low HF conditions, and peptides containing these and certain other residues must be retreated in high HF (90%) after removal of the DMS. However, since most of the potential carbonium ions in the peptide will already have been trapped as the less reactive dimethylsulfonium salt, the byproducts of the reaction are still greatly reduced.

THE NEED TO PAY ATTENTION TO DETAILS

I cannot emphasize enough how important it is to be attentive to even the smallest of details if one expects to synthesize a peptide of high quality. The principal byproducts of solid phase peptide synthesis can be classified as termination, deletion, or modification peptides. Much effort has gone into identifying these problems, developing ways to quantitate them, and finding ways to eliminate them. First of all, it is important to begin with clean, well characterized resins, clean amino acid derivatives, and clean solvents. Most of the known side reactions can now be eliminated or greatly minimized if the proper coupling methods and conditions are selected (51). It is important to monitor coupling reactions to determine that they have proceeded to completion so that deletion peptides missing one or more residues will be avoided. The quantitative ninhydrin reaction (52) is useful for that purpose and can detect the presence of 0.1% unreacted chains (i.e. 99.9% coupling). After a peptide chain has been assembled it can be analyzed by solid phase sequencing methods (53) to quantitate the levels of preview and therefore of deletion sequences (54). Except for special cases, racemization is not usually a problem in stepwise solid phase synthesis, but sensitive methods for its detection are available (55). If the various precautions alluded to here are taken, satisfactory results can be expected in most instances.

SOME RECENT SYNTHESES OF PEPTIDES

Very large numbers of peptides have been synthesized in recent years by the techniques that have been discussed and I cannot begin to cover them here. From our own laboratory we have reported recent synthetic studies on apamin (56), thymosin α_1 (57) glucagon (58), and cecropin A (59, 60). For this discussion I have selected examples of syntheses that serve to illustrate certain areas of interest.

An excellent example of a synthetic peptide study leading to useful drugs is that of Manning and Sawyer on development of vasopressin analogs with high antidiuretic activity and essentially no remaining pressor activity for treatment of diabetes insipidus (61). The best was 1-deamino-[4-valine, 8-D-argininelvasopressin. They have also discovered, through synthesis, arginine vasopressin analogs that are strong inhibitors of both antidiuretic and pressor activity for

use in patients with hyponatremia due to excessive retention of water (62). The best was [1-(β-mercapto-β,β, cyclopentamethylenepropionic acid), 2-D-phenylalanine, 4-valine]-arginine-vasopressin.

In a few instances solid phase syntheses have been scaled up for commercial purposes. A good example is salmon calcitonin (63). It has been prepared in 50-100 g batches of highly purified peptide. This 32-residue hormone is highly effective for the treatment of Paget's disease and other conditions of hypercalcaemia.

The area of greatest current interest and activity is undoubtedly the synthesis of peptides for the elucidation of the immunogenic determinants of proteins and for the development of synthetic vaccines against viral and other infectious diseases. The work from Lerner's laboratory (64) has given an important impetus to this field. Synthetic antigens are also useful for the development of diagnostics and for the production of antibodies as aids in detecting and isolating unidentified gene products.

As an illustration of my emphasis on the importance of new chemistry and the need to pay attention to the details when utilizing solid phase peptide synthesis I would mention some new work on the epidermal growth factor (EGF) by Bill Heath (65). EGF stimulates cellular proliferation, inhibits gastric acid secretion and plays a role in embryonic development. The structure of this 53-residue peptide (66) is shown in Fig. 16. It is a hydrophobic, highly crosslinked, compact molecule that others have found very difficult to synthesize in the past. By using the newly developed Pam-resin support, several new protecting groups, pure reagents, the quantitative monitoring procedures, the new HF cleavage methods, and by taking all the other known precautions against side reactions, he succeeded in obtaining an essentially quantitative assembly of the peptide chain and a 97% cleavage yield, leading to a crude unpurified product that contained 65 % of the desired EGF. It could be readily isolated in a highly purified form which eluted from a C_{18} HPLC

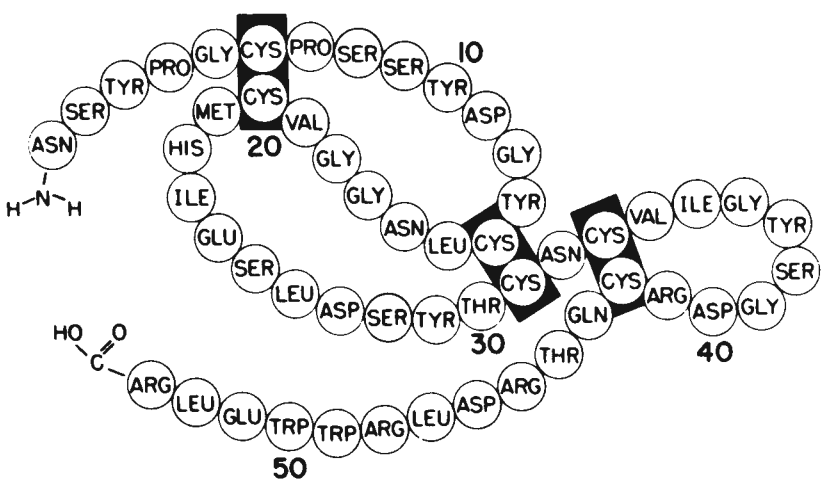


Fig. 16. Structure of mouse epidermal growth factor.

column at exactly the same time as natural EGF (Fig. 17). In the sensitive and discriminating Leydig cell growth assay the synthetic and natural EGF had identical activity.

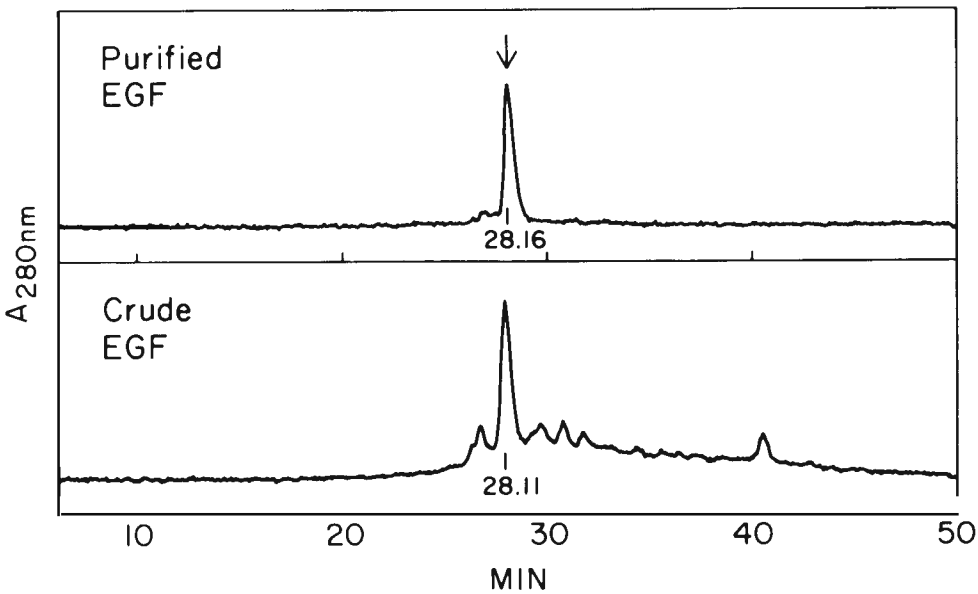


Fig. 17. HPLC analysis of synthetic EGF. The arrow indicates the position of natural EGF.

From the accumulated data presented, we conclude that the solid phase synthesis of peptides up to 50 or somewhat more residues can be readily achieved in good yield and purity; and this is a far better situation than I could have expected when this technique was first proposed.

As an example of a synthesis of a protein I have selected our recent studies on interferon. The sequence of human leucocyte interferon α_1 was first deduced from the DNA sequence of the cloned gene (67). It contains 166 amino acids with 5 cysteine residues (Fig. 18). The amino acid sequence of the isolated protein of human leucocyte interferon α_2 was also determined (68) and found to have only 155 residues. There is a high degree of homology between the two, but the latter has one deletion at Asp^{44} and is missing the last 10 residues predicted from the DNA sequence (Fig. 18). We have synthesized these two proteins and also their Ser^1 analogs and purified them by reduction, gel filtration, reoxidation, gel filtration, and affinity purification on a column of supported polyclonal antibodies to human leucocyte interferon (69). The synthetic proteins and the natural and recombinant interferon all had 10^8 to 10^9 units/mg in antiviral assays against a broad spectrum of cell lines. The development and duration of the antiviral state were also similar. Synthetic [Ser^1]IFN-α_2 and natural Hu-Le-IFN-α showed similar growth inhibition of K 562 cells. [Cys^1]IFN-α_2 and natural Hu-Le-IFN-α caused a similar increase of natural killer cell activity whereas synthetic [Ser^1]IFN-α_2 caused a decrease. All four synthetic interferons bind to and are eluted from polyclonal anti-Hu-Le-IFN-α antibodies under similar conditions.

```
                                 Glu                     Asp-Asn
α1  1   Cys-Asp-Leu-Pro          Thr-His-Ser-Leu         Arg-Arg-Thr-Leu-
α2                               Gln                     Gly-Ser

                                 Ser-Arg           Pro-Ser
   16   Met-Leu-Leu-Ala-Gln-Met          Ile-Ser           Ser-Cys-Leu-
                                 Arg-Lys           Leu-Phe

        Met                                                   Asp
   31       Asp-Arg-His-Asp-Phe-Gly-Phe-Pro-Gln-Glu-Glu-Phe       Gly-
        Lys                                                    -

                                 Pro-Ala      Ser                 Leu
   46   Asn-Gln-Phe-Gln-Lys-Ala          Ile      Val-Leu-His-Glu
                                 Glu-Thr      Pro                 Met

                                      Thr
   61   Ile-Gln-Gln-Ile-Phe-Asn-Leu-Phe     Thr-Lys-Asp-Ser-Ser-Ala-
                                      Ser

                        Asp                     Cys
   76   Ala-Trp-Asp-Glu     Leu-Leu-Asp-Lys-Phe     Thr-Glu-Leu-Tyr-
                        Thr                     Tyr

                                                Met     Glu-Glu-Arg
   91   Gln-Gln-Leu-Asn-Asp-Leu-Glu-Ala-Cys-Val     Gln
                                                Ile     Gly-Val-Gly

            Gly                     Asn-Ala
  106   Val     Glu-Thr-Pro-Leu-Met         Asp-Ser-Ile-Leu-Ala-Val-
            Thr                     Lys-Glu

        Lys             Arg                         Thr
  121       Lys-Tyr-Phe     Arg-Ile-Thr-Leu-Tyr-Leu     Glu-Lys-Lys-
        Arg             Gln                         Lys

  136   Tyr-Ser-Pro-Cys-Ala-Trp-Glu-Val-Val-Arg-Ala-Glu-Ile-Met-Arg-
                                                                     166
            Leu
  151   Ser     Ser-Leu-Ser-Thr  Asn-Leu-Gln-Glu-Arg-Leu-Arg-Arg-Lys-Glu
            Phe                   -   -   -   -   -   -   -   -   -   -
```

Fig. 18. Sequences of leucocyte interferons α_1 and α_2

These results are encouraging, but much more needs to be done to assure that even small proteins can be synthesized readily in high yield and purity. I think we can be optimistic about the future.

ACKNOWLEDGEMENTS

I owe a very special debt of gratitude to my teachers, Dr. M. S. Dunn of U.C.L.A. and Dr. D. W. Woolley of The Rockefeller University. Several of the past and present members of my laboratory have been referred to here, but to the many others who have not been specifically mentioned I am equally grateful because they all have contributed to the progress of our work. Finally, I wish to acknowledge the continuing support of The Rockefeller University and of the National Institutes of Health of the United States.

REFERENCES

1. Fischer, E. and Fourneau, E., *Ber. 34,* 2868 (1901).
2. Bergmann, M. and Zervas, L., *Ber, 65* , 1192 (1932).
3. Du Vigneaud, V., Ressler, C., Swan J. M., Roberts C. W., Katsoyannis, P. G. and Gordon, S., *J. Am. Chem. Soc., 75, 4879* (1953).
4. Merrilield, R.B., J. *Am. Chem. Soc., 8.5,* 2149 (1963).
5. Gisin, B. F., Merrilield, R. B. and Tosteson, D.C., J. *Am. Chem. Soc. 91,* 2691 (1969).
6. Kusch, P., *Angew. Chem., 78,* 611 (1966).
7. Letsinger, R. L. and Mahadevan, V., *J. Am. Chem. Soc., 87, 3526 (1965).*
8. Fréchet, J. M. and Schuerch, C., *J. Am. Chem. Soc., 93, 492* (1971).
9. Merrilield, R. B., *Advan. Enzymol., 32, 221 (1969).*
10. Sheehan, J. C. and Hess, G. P., *J. Am. Chem. Soc., 77,* 1067 (1955).
11. Bodanszky, M. and Sheehan, J. T., *Chem. Ind.* (London) 1423 (1964).
12. Wieland, T., Birr, C. and Flor, F., *Angew. Chem. Int. Ed. Engl. 10, 336* (1971).
13. Sakakibara, S. and Shimonishi, Y., *Bull. Chem. Soc. Jap. 38,* 1412 (1965).
14. Narang, *S.* A., *Tetrahedron, 39, 3* (1983).
15. Itakura, K., Rossi, J. J. and Wallace, R. B., *Ann. Reu. Biochem. 53, 323* (1984).
16. Sarin, V. K., Kent, S. B. H. and Merrilield, R. B., *J. Am. Chem. Soc., 102, 5463 (1980).*
17. Merrifield R. B. and Littau V., in "Peptides 1968", E. Bricas, ed., North-Holland Publ., Amsterdam, 1968, pp. 179-182.
18. Kent, S. B. H. and Merrilield, R. B., in "Peptides 1980", K. Brunfeldt, ed., Scriptor, Copenhagen, 1981, pp. 328-333.
19. Live, D. H. and Kent, S. B. H. in "Peptides: Structure and Function", V. Hruby and D. Rich, eds., Pierce Chem. Co., Rockford, Ill., 1983, pp. 65-68.
20. Bhargava, K. K., Sarin, V. K., Trang, N. L., Cerami, A. and Merrilield, R. B., *J. Am. Chem. Soc, 105, 3247* (1983).
21. Atherton, E., Clive, D. L. J. and Sheppard, R. C., *J. Am. Chem. Soc., 97,6584* (1975).
22. Merrilield, R. B., *Biochemistry, 3,* 1385 (1964).
23. Marshall, G. R. and Merrifield, R. B., *Biochemistry, 4, 2394* (1965).
24. Manning, M., *J, Am. Chem. Soc., 90,* 1348 (1968).
25. Merrifield, R. B., Stewart, J. M. and Jernberg, N., *Anal. Chem., 38,* 1905 (1966).
26. Fischer, E., *Ber., 39, 530* (1906). Translation taken from Greenstein, J. P. and Winitz, M., "Chemistry of the Amino Acids", Vol. 2, John Wiley, 1961, p. 1816b.
27. Hirs, C. H. W., Moore, S. and Stein, W. H., *J. Biol. Chem, 235, 633* (1960).
28. Kartha, G., Bello, J. and Harker, D., *Nature, 213, 862* (1967).
29. Gutte, B. and Merrilield, R. B., *J. Biol. Chem., 246,* 1922 (1971).
30. White, F. H. Jr., *J. Biol. Chem., 236,* 1353 (1961).
31. Anfinsen, C. B. and Haber, E., *J. Biol. Chem., 236,* 1361 (1961).
32. Richards, F. *M., Compt. Rend. Trav. Lab. Carlsberg Ser. Chim., 29, 329* (1955).
33. Finn, F. M. and Hofmann, K., *J. Am. Chem. Soc., 87, 645* (1965).
34. Rocchi, R., Marchiori, F., Moroder, L., Borin, G. and Scoffone, E., *J. Am. Chem. Soc., 91, 3927 (1969).*
35. Wyckoff, H. W., Hardman, K. D., Allewell, N. M., Inagami, T., Johnson, L. N. and Richards, F. M., J. *Biol. Chem., 242, 3984 (1967).*
36. Lin, M. C., J. *Biol. Chem., 245,6726* (1970).
37. Lin, M. C., Gutte, B., Moore, S. and Merrilield, R. B., *J. Biol. Chem., 245,* 5169 (1970).
38. Gutte, B., Lin, M. C., Caldi, D. G. and Merrilield, R. B., *J. Biol. Chem., 247, 4763* (1972).
39. Hayashi, R., Moore, S. and Merrilield, R.B., *J. Biol. Chem. 248, 3889 (1973).*
40. Lin, M. C., Gutte, B., Caldi, D. G., Moore, S. and Merrifield, R. B., *J. Biol. Chem., 247, 4768* (1972).
41. Merrilield, R. B. and Hodges, R. S., in "Proceedings Internat. Symp. Macromolecules", E. B. Mano, ed., Elsevier, Amsterdam, 1975, pp. 417-431.
42. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. and Richards, F. M., *J. Biol. Chem., 245, 305* (1970).
43. Hodges, R. S. and Merrifield, R. B., *J. Biol. Chem., 250,* 1231 (1975).

44. Mitchell, A. R., Erickson, B. W., Ryabtsev, M. N., Hodges, R. S. and Merrifield, R.B., J. *Am. Chem. Soc, 98, 7357* (1976).
45. Mitchell, A. R., Kent, S. B. H., Erickson, B. W. and Merrifield, R. B., *Tetrahedron Lett. , 3795 (1976).*
46. Kent, S. B. H., Mitchell, A R., Engelhard, M. and Merrifield, R. B., *Proc. Nat/. Acad. Sci. (USA,* 76, 2180 (1979).
47. Barany, G. and Merrifield, R. B., *J. Am. Chem. Soc ., 99, 7363* (1977).
48. G. Barany (Personal communication).
49. Tam, J. P., Heath, W. F. and Merrifield, R. B., *J. Am. Chem. Soc. 10.5, 6442 (1983).*
50. Yates, K. and McClelland, R. A., *Am. Chem. Soc., 89, 2686 (1967).*
51. Barany, G. and Merrilield, R. B., in "The Peptides" Vol. 2, E. Gross and J. Meienhofer, eds., Academic-Press, N.Y., 1979, pp. 1-284.
52. Sarin, V. K., Kent, S. B. H., Tam, J. P. and Merrilield, R. B., *Anal. Biochem., 117,* 147 (1981).
53. Laursen, R. A., *J. Am. Chem. Soc., 88, 5344 (1966).*
54. Niall, H. D., Tregear, G. W. and Jacobs, J., in "Chemistry and Biology of Peptides", J. Meienhofer, ed., Ann Arbor Press, MI., 1972, pp. 695-699.
55. Manning, J. M. and Moore, S., *J. Biol. Chem. , 243, 5591 (1968).*
56. Cosand, W. L. and Merrifield, R. B., *Proc. Natl. Acad. Sci. USA, 74,* 2771 (1977).
57. Wong, T. W. and Merrilield, R. B., *Biochemistry, 19, 3233* (1980).
58. *Mojsov, S.* and Merrifield, R. B., *Biochemistry, 20, 2950* (1981).
59. Merrifield, R. B., Vizioli, L. D. and Boman, H. *G., Biochemistry, 21,* 5020 (1982)
60. Andreu, D., Merrifield, R. B., Steiner, H. and Boman, H. G., Proc. *Natl. Acad. Sci. USA, 80,* 6475 (1983).
61. Manning, M., Balaspiri, L., Acosta, M. and Sawyer, W.H., *J. Med. Chem., 16, 975* (1973).
62. Manning, M., Klis, W. A., Olma, A., Seto, J. and Sawyer, W. H., *J. Med. Chem., 25, 414* (1982).
63. Hughes, J., U.S. Patent 3926-938. December 16, 1975.
64. Lerner, R. A., *Nature, 299, 592* (1982).
65. Heath, W. F., unpublished results.
66. Cohen, S., *J. Biol. Chem. 237,* 1555 (1962).
67. Mantei, N., Schwarzstein, M., Streuli, M., Panem, S., Nagata, S. and Weissmann, C., Gene, 10, 1 (1980).
68. Levy, W. P., Rubinstein, M., Shively, J., Del Valle, V., Lai, C. Y., Moschere, J,, Brink, L., Gerber, L., Stein, S. and Pestka, S., *Proc. Natl. Acad. Sci. USA, 78, 6186* (1981).
69. Krim, M., Mecs, I., Merrifield, E. L., Fox, F., Sarin, V. and Merrifield, R. B., TNO-ISIR Meeting, Heidelberg, Federal Republic of Germany, October 21-25, 1984.