MAR 23 1964

BNL 7803

MASTER

THE CHEMISTRY OF PEPTIDES AND PROTEINS^{1, 2}

C. H. W. Hirs³

Biology Department, Brookhaven National Laboratory, Upton, N. Y.

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

INTRODUCTION

The principal concern of this review will be a consideration of the experience that has grown in the last few years from the application of chemical procedures to the elucidation of the covalent structure of peptides and proteins. A comprehensive review has not been attempted; rather publications selected for their value in developing the interest of the present survey have been gathered for analysis. The selection has been influenced in part by the extent to which certain topics have recently been discussed elsewhere. For instance, chemical studies with the hemoglobins will be mentioned only indirectly in view of Schroeder's review in 1963 (1). An attempt has also been made to limit the number of preliminary reports considered so as to generate more space in which to deal in greater depth with several subjects that appear to be of particular importance.

CHEMISTRY OF STRUCTURE DETERMINATION

Amino Acid Analysis

Automatic methods.--The automatic amino acid analyzer developed by Spackman, Moore & Stein (2) continues to play a pivotal role in the evolution of peptide and protein chemistry. As was to be expected, the method originally described has undergone various modifications to extend its usefulness [cf., for example, the improved resolution of basic amino acids described by Kominz (3)]. A major change in approach was introduced by Piez & Morris (4) who developed a single column gradient elution procedure. The principal shortcomings of the method (and modifications of it) are the relatively large, progressive change in the base-line value during the analysis, the uncertainty of the blank value in the lysine region, and the length of time required for the elution of arginine; virtues are its potential flexibility for specific applications, good resolution of the amino acids, and the avoidance of volumetric errors [cf. Walsh (5)] attendant on the application of samples to separate columns as required in the procedure of Spackman, Moore & Stein (2). At the ex-

pense of some resolution, especially in the tyrosine and phenylalanine region, the original technique of Spackman, Moore & Stein was modified by Spackman for use with columns approximately one-third as long as in the original method. This development particularly facilitated the analysis of peptides as less time and sample are required per analysis, and played an important part in work on the structure of hemoglobin (6-11), and in the reinvestigation of the amino acid sequence of bovine pancreatic ribonuclease (12, 13). In the most recent version described by Spackman (14), improved ion-exchange resins have permitted flow rates 50 percent greater than those previously employed with little loss in resolution compared to that obtained with longer columns. Three complete analyses may be effected in a single day. A modification of the procedure has been employed for over a year in the present writer's laboratory and involves the use of a bead-type sulfonated polystyrene resin (Chromobeads, Type III, Technicon Instruments Corp.). The resolution obtained is significantly better than with pulverized resins commonly in use. Hamilton (15, 16) has described greatly improved separations with carefully screened sulfonated polystyrene beads (Dowex 50-X8) in the range of 17 to 18 microns

in diameter. He has also pioneered in studies on the modification of the amino acid analyzer for microanalysis by electronic amplification of detector signals and by use of flow-cells that permit optical amplification of absorbance readings. In the present writer's experience very satisfactory amplification has been achieved with a commercially available, longitudinal flow-cell possessing a light path of 15 mm (Technicon Instruments Corp., "tubular" flow-cell). The amplification gained permits routine analysis of peptides and proteins in the range of 0.01 to 0.04 micromoles of amino acid with a precision comparable to that attained in the procedure originally developed by Spackman, Moore & Stein (2).

While the subject of amino acid analyzers is not one of the principal concerns of this review, a report on the status of the field would not be complete without mention of work aimed at the automation of data processing from amino acid analyzers. A cumulative integration technique with recorder display has been developed by R. J. Hill⁴ who has had it in use for over a year. Integrators that utilize a slope-detector in conjunction with an analog voltage to frequency conversion

counter have been developed commercially (e.g., Infotronics, Inc.), but results obtained with such equipment have not been published. Another approach under commercial development has been to digitize the recorder output of the amino acid analyzer for transmission to punched tape which is subsequently processed in a computer. This approach has been used successfully in the writer's laboratory for several months with a digitizing unit designed by A. Yonda. A high-speed computer (Merlin) has been programmed by H. Tape and D. Filmer to process the output. The results obtained have either equalled or surpassed in accuracy those obtained by the customary manual computation.

A completely different approach to microanalysis of amino acids has been described by Kirsten & Kirsten (17). The column dimensions are reduced to those of a coil of capillary tubing through which the eluents are forced by the controlled displacement of a syringe. The product developed with a ninhydrin reagent is determined in a photometer equipped with a longitudinal flow-cell. The sensitivity attained permits operation in the nanomole (10^{-9}) range of concentration.

Analysis for specific amino acids.--An improved procedure for

the determination of cysteine plus cystine as cysteic acid has been detailed by Moore (18). Fletcher & Robson (19) found that addition of S³⁵-cystine to the proteins insulin, lysozyme, and bovine serum albumin prior to acid hydrolysis permitted an accurate analysis for the combined contents of cysteine and cystine by isotope dilution. The cystine was recovered from the hydrolysates by ion-exchange chromatography.

Tower, Peters & Wherrett (20) have described a method for the separate determination of the dicarboxylic amino acids and their amides in proteins. A partial hydrolysate of the protein is obtained with the aid of a mixture of enzymes produced by activation of desiccated, defatted pancreas (Viokase). The partial hydrolysate is suitable for microbiological assays. Four separate operations are involved. (1) Asparagine is determined with a purified guinea-pig liver enzyme. (2) Glutamine, glutamic acid, and aspartic acid in combination are determined with a decarboxylase preparation from Cl. perfringens in which L-aspartate decarboxylase is activated with pyruvate and pyridoxal phosphate. (3) The

same microbial preparation treated with cetyl-trimethylammonium bromide to inactivate aspartate decarboxylase serves for the determination glutamic acid plus glutamine. (4) Finally, glutamine alone is determined with the aid of the glutaminase present in the Cl. perfringens preparation. The results obtained in these several determinations lead to the quantitation of each component. The technique was applied to a number of peptides and proteins with particularly satisfactory results. The method appears to have advantages over that described by Hill & Schmidt (21), who have employed a two-stage enzymic hydrolysis of proteins with papain and a mixture of leucine aminopeptidase and prolidase. The hydrolysates obtained were subjected to chromatography on an amino acid analyzer. Obviously, in this approach success depends critically on the completeness with which a protein is hydrolyzed to amino acids. It would appear from the results obtained with insulin and ribonuclease that it is difficult to attain complete liberation of amino acid residues adjacent to cystine in native proteins.

Determination of amino acid derivatives.--Dinitrophenyl amino acids continue to play an important part in the development of protein

chemistry. It is therefore appropriate to mention a significant advance in the chromatographic determination of these derivatives. Kesner and his colleagues (22) have described an automatic column method for both dinitrophenyl amino acids and peptides. The column is prepared under carefully standardized conditions from a slurry of acid-moistened silica gel in n-heptane. Elution is achieved with a gradient of t-amyl alcohol and methyl ethyl ketone in heptane; the eluants are saturated with water prior to transfer to the column by a constant volume pump. The resolution obtained is dependent on the extent of hydration of the silica gel: in an experiment illustrated by these authors, the resolution was satisfactory for all the derivatives except the two (leucine and isoleucine) that migrated most rapidly on the column. The effluent passes a flow-cell in a recording spectrophotometer, and under the conditions specified between 0.06 and 0.2 micromole quantities of the individual derivatives were determined with relative standard deviations between 5 and 9 percent. A complete analysis for all the common dinitrophenyl amino acids required five hours.

Gas-liquid chromatography of amino acid derivatives.--Attempts

continue to be made to develop procedures for the quantitative amino acid analysis of peptides and proteins by gas chromatography of suitable derivatives. There are three principal steps in the development of a satisfactory gas chromatographic approach to amino acid analysis: preparation of an appropriate stable derivative in reproducible, high yield; the column separation itself; and quantitation of the derivatives as they emerge from the column. At present it is apparent that the problems connected with the first two of these steps are close to solution in approaches that involve the n-butyl or n-amyl esters of the N-trifluoroacetyl amino acids. Zomzely, Marco & Emery (23) have described a procedure in which esterification was achieved at 50-60° in 1-butanol containing dimethylformamide (to solubilize the basic amino acids) and 5 percent anhydrous HCl. The water formed in the reaction was removed by addition of a three-fold excess of dibutoxypropane. Esterification was estimated more than 90 percent complete. Trifluoroacetylation was performed at 28° in methylene chloride (also containing dimethylformamide) and probably proceeded quan-

titatively. The method served to prepare the n-butyl esters of all the trifluoroacetyl amino acids, including cystine and tryptophan. The chromatographic separation was achieved on a column of Gas Chrom A coated with 1 percent of a neopentylglycol succinate polyester. With the exception of a partial overlap of the derivatives of aspartic acid and phenylalanine, excellent resolution of all the amino acid derivative was obtained. A hydrogen-flame detector was employed, and considerable variation in detector response among the derivatives required several range shifts during the recording of the chromatogram in order to keep peaks on scale. Johnson, Scott & Meister (24, 25) have described similar separations with the n-amyl esters of trifluoroacetyl amino acids. The further development of these procedures for amino acid analysis will require studies on the reproducibility with which the derivatives can be prepared, and a simple solution to the problem of determining the quantities of the derivatives from the detector response. Nonetheless, the high sensitivity of the gas chromatographic procedures and the rapidity with which analyses may be obtained hold promise that the methods in their present state will prove valuable

in certain specialized applications. In this regard it is of interest that Pisano, Vanden Heuvel & Horning (26) have described gas chromatographic separations of a number of dinitrophenyl amino acids and amino acid phenylthiohydantoins. The dinitrophenyl amino acids were converted to methyl esters with diazomethane prior to separation on columns of Gas Chrom P coated with between 0.75 to 1 percent of thermostable polymers. Three polymers were examined: a methyl silicone, a fluoroalkyl silicone, and a phenyl silicone. All gave satisfactory resolution with neutral and acidic amino acids, but the derivatives of serine, threonine, tryptophan, tyrosine, and histidine underwent decomposition on the columns, and the derivatives of lysine and arginine could not be chromatographed. Amino acid phenylthiohydantoins could be separated readily on the same columns. It was necessary to convert the aspartic acid and glutamic acid derivatives to their methyl esters prior to chromatography. Asparagine and glutamine phenylthiohydantoins underwent decomposition, though reproducible elution patterns were obtained when they were chromatographed. Serine and threonine phenylthiohydantoin could also not be recovered unchanged after chroma-

tography; and the phenylthiohydantoins of lysine and arginine could not be chromatographed at all. Despite these limitations the methods should prove useful in research involving these important amino acid derivatives.

End-Group Methods

The cyanate procedure.--An important advance in this field has been the introduction by Stark & Smyth (27) of the cyanate procedure, which was developed to take advantage of the convenience and precision of the amino acid analyzer. The peptide or protein (in 8 M urea if necessary) is allowed to react with potassium cyanate at pH 8 until all the amino groups have been fully carbamylated. The carbamylated product is exposed to dilute acid at 110°, whereupon the N-terminal amino acid is cleaved off as the hydantoin and the remainder of the molecule undergoes partial hydrolysis. When tryptophan is present as an end-group special conditions are used for the acid cleavage (inclusion of serotonin). Unlike peptides and amino acids the hydantoins of the acidic and neutral amino acids (including tryptophan) and the hydantoin of homocitrulline, formed in the carbamylation of N-terminal lysine, are not absorbed by

Dowex 50-X2 and are obtained when the solution is filtered through a small column of the resin in the acid form. An equilibrium exists between the hydantoins and the carbamyl amino acids. The equilibrium mixture from the parent compound, hydantoin itself, contains approximately 20 percent of carbamyl glycine; the proportion of carbamyl amino acid at equilibrium appears to be smaller with other amino acids. Being acidic substances the carbamyl derivatives of most amino acids are not retarded on Dowex 50 and therefore accompany the hydantoins in the filtration step. However, carbamyl homocitrulline is partially retarded and approximately 20 percent of this compound fails to appear in the hydantoin fraction. Pyrrolidone-carboxylic acid and peptides containing it are also not retarded by Dowex 50 in the acid form, but pretreatment in dilute acid at 100° suffices to almost completely hydrolyze pyrrolidone-carboxylic acid to glutamic acid. The hydantoins in the filtrate from the Dowex 50 column are subjected to alkaline hydrolysis and the amino acids are determined with an automatic analyzer. Elution of the Dowex 50 column with M ammonium hydroxide removes the hydantoins of histidine and arginine together with some free amino acids and peptides. The constituents

of the ammonium hydroxide eluate are fractionated on a second column of Dowex 50-X2 with 0.8 N HCl as eluent, when the hydantoins of histidine and arginine appear as separate zones admixed with peptides. However, peptides containing histidine and arginine have an additional positive charge in strong acid and do not elute with the hydantoins of these amino acids. The fractions containing histidine and arginine hydantoin are separately subjected to acid hydrolysis to regenerate the free amino acids.

When appropriately modified the procedure is capable of dealing with N-terminal cysteine, cystine, and half-cystine residues. Terminal cysteine undergoes carbamylation on both the amino group and the thiol side-chain. In the subsequent cyclization in acid, a significant fraction of the S-carbamyl group is lost by hydrolysis and a mixture of cysteine and S-carbamylcysteine hydantoins is present in the neutral hydantoin fraction. Acid hydrolysis regenerates the free amino acid, which is measured as cystine. N-terminal cystine likewise is finally determined as the free amino acid, but can be distinguished from cysteine end-groups by

oxidation with performic acid subsequent to the carbamylation step but prior to cyclization, since S-carbamylcysteine residues cannot be oxidized with performic acid but are sensitive to acid hydrolysis. Cysteic acid hydantoin is determined in a separate chromatogram on Dowex 1-X8 with 0.2 N HCl as developer. N-terminal half-cystine appears as cysteic acid hydantoin when oxidation with performic acid and carbamylation precede cyclization and chromatography; when the order of these manipulations is reversed so that carbamylation, cyclization, and separation of the hydantoin fraction precede oxidation, no cysteic acid hydantoin is found. The cyanate procedure may also be applied to performic acid-oxidized proteins. However, cysteic acid peptides formed during the cyclization step can accompany the hydantoin fraction and must be eliminated by an additional chromatogram on Dowex 1-X8 with 0.5 M acetic acid as developer, during which the hydantoins are separated from the strongly adsorbed cysteic acid peptides.

The cyanate procedure gives very satisfactory results with most end-groups. After alkaline hydrolysis arginine end-groups are finally mea-

sured as ornithine, and isoleucine and threonine as a mixture of natural and allo diastereoisomers. The poorest recoveries are obtained with serine and threonine, which appear in yields of 20 and 30 percent respectively. Alkaline hydrolysis of hydantoin fractions containing these amino acids gives rise to glycine. With threonine hydantoin, glycine is formed in a yield approximating that of threonine. The method was tested with cytochrome c and the protein of tobacco mosaic virus, in both of which the N-terminal amino acid is acetylated, and no end-group was found. The N-terminal half-cystine residue in α -chymotrypsinogen was obtained quantitatively. The relative simplicity of the procedures involved and their high degree of reproducibility will undoubtedly lead to the widespread acceptance of the method for quantitative work. It appears to have certain advantages over the dinitrophenylation method of Sanger [cf. Porter (28)], notably in that the recoveries of most of the end-groups are significantly higher. Improved procedures for fractionating and determining dinitrophenyl amino acids have been mentioned in a previous section of this review, and should, in conjunction with careful

technique to minimize photo-decomposition [cf. Russell (29)], serve to place the dinitrophenylation method on a reproducible quantitative footing.

Reduction and Alkylation of Proteins

Accumulating experience indicates that the most satisfactory reagent thus far employed for the reduction of disulfide bonds in proteins is β -mercaptoethanol. Thioglycollic acid has been shown by White (30) to contain poly-thioglycollides which, as thioesters, may acylate amino groups in the protein. It is therefore necessary to remove thioglycollides prior to the use of the reagent by treatment with an amine such as methylamine. Sodium borohydride in 8 M urea, at the pH of the reagent, has been shown by Crestfield, Moore & Stein (31) to achieve a quantitative reduction of disulfide bonds in ribonuclease, but to lead also to the formation of fragmentation products of lower molecular weight (as detected by the appearance, after carboxymethylation with iodoacetate, of several components moving with greater retardation than reduced, carboxymethylated ribonuclease itself on a column of Sephadex G-75 operated

with a dissociating solvent). The reactions responsible for the fragmentation have not been identified but probably involve the reduction by the hydride of certain reactive peptide bonds; the esters formed would undoubtedly undergo some hydrolysis under the alkaline conditions prevailing during the reduction.

Oxidative sulfitolysis according to the procedure of Swan (32) has been applied to the cleavage of disulfide bonds in several proteins, including α -chymotrypsinogen (33), α -lactalbumin (34), and insulin (35). The principal problems encountered in the application of this procedure are the difficulty of forcing the reaction to completion [at least with α -chymotrypsinogen; cf. Hartley (36)], and the instability of S-sulfo-cysteine at extremes of pH, which compels manipulative restrictions and special requirements for quantitative analysis.

In general, complete reduction of all the disulfide bonds in a protein requires that the tertiary structure be abolished first. It is common practice to achieve this end by performing the reduction with a large excess of reducing agent in the absence of oxygen and in solutions

that contain high concentrations of urea or guanidinium chloride. Mercaptoethanol is itself an effective denaturing agent, and Thompson has described a technique for reducing proteins in solutions containing 40 percent mercaptoethanol (37). The undesired formation of cyanate ion in concentrated urea solutions requires that these be prepared fresh and preferably be passed through a mixed-bed deionizing column immediately prior to use; alternatively an excess of methylamine may be added. Stark, Stein & Moore (38) have shown there is otherwise a danger that modification of ϵ -amino groups may occur. The reduced protein may be isolated simply by acid precipitation in organic solvents, or by gel-filtration under acid conditions as described by Anfinsen & Haber (39) with reduced ribonuclease. It is frequently convenient to perform the subsequent alkylation step without first isolating the reduced protein, but this approach has the disadvantage that excess reducing agent present from the first step must be taken into account.

The alkylation of thiol groups has most frequently been accomplished at pH 8 to 8.5 with iodoacetate. The reaction is essentially

stoichiometric and is usually performed by adding iodoacetate in the calculated quantity or in slight excess over the total equivalents of thiol present. Excess iodoacetate is subsequently removed by an addition of mercaptoethanol. Though they take place more slowly than the carboxymethylation of thiol groups, sulfonium salt formation from methionine side-chains and carboxymethylation of imidazole and ϵ -amino groups are potential side-reactions at pH 8.5, and the conditions for alkylation must be determined accurately to minimize such side-reactions. A careful study of the reduction and carboxymethylation of ribonuclease and insulin has been presented by Crestfield, Moore & Stein (31). These authors have emphasized a further difficulty attendant on the use of iodoacetate for alkylation: the iodide formed may be the source of detrimental concentrations of iodine if the product is not isolated rapidly and/or the solutions are exposed to light. Free halogen may cause modification of thioether, imidazole, phenolic, and indole side-chains. Under the conditions used by Crestfield, Moore & Stein there was no oxidation of thioether groups to sulfoxides in ribonuclease, and tryptophan

determinations with β -lactoglobulin and lysozyme that had been reduced and carboxymethylated under the same conditions revealed that this amino acid had not been attacked.

Most of the problems connected with the use of iodoacetate are avoided when acrylonitrile is substituted for the alkylation. Weil & Seibles (40) introduced the use of this reagent for blocking protein thiol groups. These authors studied the reduction and cyanoethylation of α -lactalbumin and β -lactoglobulin at pH 8, and demonstrated a quantitative conversion of cystine to S-cyanoethylcysteine without side-reactions of any sort. Reduction of these proteins with the usual excess of mercaptoethanol was achieved without the need for denaturing agents; cyanoethylation was conducted with twice the calculated quantity of reagent required to combine with all the thiol present. Plummer & Hirs (41) have employed the method for studies on the structure of the glycoprotein ribonuclease B. Reduction was attained with mercaptoethanol in 8 M urea, and quantitative cyanoethylation of the cysteine residues was observed at pH 8. The oligosaccharide moiety in this enzyme was not affected.

Studies with ribonuclease A showed that at pH 9.5 acrylonitrile reacted slowly with amino groups, but no modification of imidazole or thioether groups was detected. The fact that S-cyanoethylcysteine residues confer a lower solubility to peptides than residues of S-carboxymethylcysteine appears to be the principal limitation of the method.

Specific Cleavage of Peptide Chains

Trypsin continues to be the most specific reagent available for the cleavage of polypeptide chains. The chymotryptic activity of crystalline samples obtained by ammonium sulfate fractionation of acid extracts of pancreas arises in part from contaminating chymotrypsin and in part from an inherent "chymotryptic" activity of the enzyme itself (42). It has become customary to selectively inhibit adventitious chymotrypsin by acid treatment at elevated temperatures, or by reaction with approximately 0.1 molar equivalent of di-isopropylphosphorofluoridate. A chromatographic separation of the two enzymes has been described by Maroux, Ravery & Desnuelle (43), and provides an unambiguous method for precisely determining the chymotrypsin contamination in a trypsin preparation.

The high selectivity of trypsin action has led Hofmann (44) to explore the value of the technique originally proposed by Lindley (45) for the specific cleavage of peptides at cysteine residues. In this procedure the cysteine residues are alkylated with β -halo-ethylamines to give S- β -aminoethylcysteine residues. Lindley demonstrated that peptides containing such residues were substrates for trypsin. Application of the procedure to reduced trypsinogen has been valuable (44) in studies of the amino acid sequence of this protein. The reaction of β -halo-ethylamines with nucleophiles involves the intermediate formation of ethylene imine. Cole [cf. Raftery, Lovenberg & Cole (46)] has demonstrated that ethylene imine may be used advantageously in place of halo-ethylamines in the Lindley procedure. Earlier approaches to the creation of trypsin-sensitive sites in polypeptides by modification reactions have been reviewed by Witkop (47).

The action of trypsin at lysine residues is blocked when the ϵ -amino group is appropriately substituted. The use of blocking groups that may subsequently be removed by mild procedures affords an opportunity

for the selective degradation of polypeptide chains in stages, first at arginine and subsequently at lysine residues. To date, the most promising results appear to have been obtained with the trifluoroacetyl group, which, as Goldberger & Anfinsen (48) have shown, may be introduced by reaction of the protein at pH 10 with ethyl thioltrifluoroacetate. The reaction was carried out with performic acid-oxidized ribonuclease, and tryptic hydrolysis of the product afforded the five fragments to be expected from the primary structure. The blocking groups may be removed by treatment with M piperidine at 0°. Walsh has reported (44) the use of the trifluoroacetylation procedure in his studies on the structure of trypsinogen. An alternative, but probably less useful procedure, also introduced in Anfinsen's laboratory, involves the use of carbon disulfide. Reaction of the protein at pH 10 with this reagent effects modification of the α - and ϵ -amino groups to give dithiocarbamates. The sensitivity to oxidation of dithiocarbamates necessitates special precautions. The protecting groups are readily removed by lowering the pH. Merigan, Dreyer & Berger (49) demonstrated the value of the

method with the S-peptide from ribonuclease S, and applied it to obtain a number of fragments from reduced, carboxymethylated egg-white lysozyme, one of which was to prove useful to Canfield (50, 51, 52) in the elucidation of the amino acid sequence of this enzyme.

A further approach to selective blocking of amino groups has been suggested by the work of Hunter & Ludwig (53) and of Ludwig & Byrne (54). Hunter & Ludwig found that methyl acetimidate and methyl benzimidate react readily with the amino groups in amino acids and peptides in the pH range between 7 and 11 to form amidines. The substitution does not alter the charge of the amino group, a factor of importance in maintaining the solubility of the product when the reaction is applied to proteins. Simple alcohols, thiols, p-hydroxyphenyl compounds, carboxylic acids, imidazole and guanidine derivatives did not react with the imido-esters at measurable rates. The products formed by reaction of insulin at pH values between 7 and 10 with large excesses of the imido-esters methyl acetimidate and methyl β -dimethylaminopropionimidate were studied in detail. The extent of substitution was measured by dinitrophenylation,

which showed that the modified insulins were virtually devoid of free ϵ -amino groups, but contained approximately 20 percent of unreacted N-terminal phenylalanine, and possibly small quantities of unreacted terminal glycine. Electrophoretic examination of the products showed the presence of a minor component the formation of which was found to be attributable to an undefined side-reaction that involved the N-terminal phenylalanyl sequence of the B-chain. The modified insulins were cleaved by trypsin only at the single arginine residue in the B-chain. A preparation formed by treatment of performic acid-oxidized insulin with methyl acetimidate was held in a solution of ammonium acetate buffer at pH 11.3, room temperature, for eight hours. Complete removal of the blocking groups without detectable side-reactions was achieved by this treatment. Hydrazine acetate at pH 9 was also found to be capable of displacing acetimidyl groups.

The search for purely chemical methods of high selectivity has received encouragement and stimulation through the successes attained [and recently reviewed (47)] by Witkop and his students. In particular,

their cyanogen bromide cleavage reaction (55, 56) for the fission of peptide chains at methionine residues promises to be invaluable. Koshland, Strumeyer & Ray (57) used the reaction to identify the methionine residue in α -chymotrypsin that becomes modified when the enzyme is exposed to the action of hydrogen peroxide at pH 3. Edmundson (58) has cleaved sperm whale myoglobin at the two methionine residues to obtain fragments essential for the determination of the amino acid sequence of this protein by chemical means. Cyanogen bromide cleavage has also played a part in the structural analysis of trypsinogen (44).

The remarkable selectivity of very weak acid at 100° in preferentially liberating free aspartic acid from proteins has been known since 1950 (59). The reaction has been studied extensively by Schultz has been (60) and applied with considerable success recently in achieving useful cleavages for the determination of the amino acid sequences of relatively small peptides. Thus Schroeder et al. (61) have described experiments in which this selectivity was utilized in the study of several peptides obtained by tryptic hydrolysis of the γ -chain from human fetal hemoglobin.

Tsung, Funatsu & Young (62) have applied the reaction to a study of certain peptides from the protein of tobacco mosaic virus.

Reactions of Specific Amino Acid Residues

Serine.--Dixon & Weitkamp (63) have achieved a noteworthy replacement of the N-terminal serine residue in corticotropin by glycine. Oxidation of corticotropin with periodate gave a product, separated from minor contaminants by chromatography, in which the terminal serine residue had been oxidized to glyoxylic acid. When this product was held in solution with glutamic acid in the presence of cupric ion transamination occurred in good yield with the formation of an N-terminal glycine residue. The modified hormone contained a small quantity of material probably derived by oxidation of methionine to the sulfoxide, but this could be removed by reduction.

Phenylalanine.--Wilchek & Patchornik (64) studied the possibility of reducing phenylalanine side-chains in peptides with lithium in methylamine at -60° to obtain tetrahydro derivatives in which the single double bonds remaining in the side-chains would be in an allylic relation to the

carbonyl group of the peptide bond. Such reduced phenylalanyl residues should be susceptible to N-bromosuccinimide with cleavage of the peptide bond at the carbonyl group in a reaction similar to that undergone by tryptophan, histidine, and tyrosine residues, and by residues of the amino acids allyl-glycine and methallyl-glycine (47). Acetyl-DL-phenylalanyl-glycine and acetyl-DL-phenylalanyl-DL-valine were reduced at -60° . The desired products were isolated in yields close to 50 percent. The tetrahydro-peptides were treated with bromine in 50 percent acetic acid or with N-bromosuccinimide, and cleavage to the extent of 65 percent occurred as estimated from the quantity of glycine or valine formed.

Cysteine.---The thiol group of cysteine residues in peptides and proteins react rapidly with fluorodinitrobenzene at neutral or slightly acid pH. Sokolovsky, Wilchek & Patchornik (65) have synthesized p-nitrobenzyloxycarbonyl-glycyl-(S-dinitrophenyl)-L-cysteine methyl esters, S-dinitrophenyl-glutathione, S,N-bis-dinitrophenyl-glutathione (and its dimethyl ester), and p-nitrobenzyloxycarbonyl-glycyl-(S-dinitrophenyl)-L-cysteinyl-glycine. All of these compounds underwent β -elimination when

dissolved in 0.2 N NaOH. The reaction was complete within 30 minutes and could be followed by measurement of the dinitrothiophenol formed, either spectrophotometrically or by iodimetry. The dehydroalanine peptides produced in the reaction are known to be particularly susceptible to hydrolysis by either acids or bases. Patchornik & Sokolovsky (66) have explored the possibility of cleaving dehydroalanine peptides by oxidation. p-Nitrobenzyloxycarbonyl-glycyldehydroalanine underwent quantitative oxidation to p-nitrobenzyloxycarbonyl-glycinamide with one mole of iodine at pH values between 1 and 6. However, the methyl ester of this peptide was not attacked, nor was the dehydroalanine peptide γ -L-glutamyl-dehydroalanyl-glycine. By contrast all three of these compounds were oxidized by one mole of bromine in acid solution. p-Nitrobenzyloxycarbonyl-glycyl-dehydroalanine and its methyl ester formed the amide in 80 to 90 percent yield, while γ -L-glutamyl-dehydroalanyl-glycine gave glutamine in 87 percent yield.

Wood & Catsimpoolas (67) identified 2-iminothiazolidine-4-carboxylic acid in an acid hydrolysate formed from bovine serum albumin that

had been exposed to the action of cyanide ion at pH 7. The thiazolidine formation was explained when it was found that N,N'-diacetylcystine gave N-acetylcysteine and 2-iminothiazolidine-4-carboxylic acid when exposed to cyanide under similar conditions. The yield of thiol in the reaction attained 72 percent in 28 hours at 37°. Cyclization of the intermediate thiocynoalanine to the iminothiazolidine-carboxylic acid was accompanied by elimination of the acetyl group. Elimination of thiocyanate ion competed with the cyclization and reached 20 percent in 21 hours when the ethyl ester of N,N'-diacetylcystine was similarly treated. When oxidized glutathione was subjected to the action of cyanide at pH 7 for 21 hours 93 percent of the theoretical quantity of reduced glutathione and 96 percent of the expected glutamic acid were formed. The remaining fragment, 2-iminothiazolidine-4-formylglycine was identified by chromatography of its dinitrophenyl derivative. Oxytocin was cleaved by cyanide in apparently excellent yield to the expected fragments. The yield of thiol formed in the reaction was only 20 percent of the theoretical in spite of precautions taken to remove oxygen. Similarly, when ribonuclease was treated

with cyanide only 5 percent of the expected thiol was liberated. It is possible that reoxidation by oxidants in the reagents was responsible for the low thiol titers; the disulfides formed would of course immediately have undergone further reaction with cyanide. The mixture formed by cyanide-cleavage^{of} ribonuclease was separated by electrophoresis and chromatography on paper into nine components, eight of which were shown to contain 2-iminothiazolidine-4-carboxylic acid and therefore probably corresponded to the fragments formed by cleavage at each of the eight half-cystine residues. The further exploration of the potentialities of this reaction will be awaited with great interest.

Amino Acid Sequence Determination

Exopeptidases.--Leucine aminopeptidase and the carboxypeptidases continue to play an important part in amino acid sequence determinations, as well as in the allocation of amide groups in sequences that contain residues of the dicarboxylic amino acids. Impurities in crude leucine aminopeptidase preparations include enzymes with prolidase, prolinase, and carboxypeptidase activities, the presence of which may lead to diffi-

culties in sequence determinations [cf. Smyth, Stein & Moore (13)]. A simplified preparation of highly purified ($C_1 = 40 - 70$) leucine aminopeptidase from hog kidney has been described by Fasold, Linhart & Turba (68). The key steps in the procedure appear to be a continuous electrophoretic separation followed by chromatography on DEAE-cellulose. The same authors (69) have described experiments in which the B-chain of insulin was subjected to the action of the enzyme in a thin-film dialysis apparatus under conditions that allowed continuous removal of the perfusate. Amino acid analysis of fractions collected on a time basis were performed by paper chromatography, and the results permitted the determination of the sequence of the first five residues. Anderer & Handschuh (70) mention the use of an aminopeptidase preparation from yeast which proved to be of value in work with the peptide Ser.Ileu.Thr. Thr.Pro.Ser.Gln.Phe from tobacco mosaic virus protein.

Edman degradation.---The acid catalyzed cyclization-cleavage of phenylthiocarbamyl peptides is perhaps the most important single reaction employed by the peptide sequence analyst. Several relatively recent devel-

opments of both theoretical and practical importance in the application of the reaction require consideration in this review. Some of these have been considered briefly by Richards (71).

The greatest economy of material in the Edman degradation is attained when the consequences of its successive application to a peptide or protein are monitored by identification of the phenylthiohydantoin cleaved off at each stage. Following Fraenkel-Conrat (72) and Harris (73) a number of procedures have been described in which the reactions required for the sequential removal of amino acids are performed on microgram quantities of the peptide supported on adsorbents such as paper or silica-gel. The phenylthiohydantoins may then be identified directly by paper or thin-layer chromatography. Among these procedures that of Schroeder and his colleagues (61) represents a noteworthy advance in the direct micro-application of the Edman degradation. It involves drying approximately 0.2 micromoles of peptide onto a small piece of filter paper which is subsequently soaked with a solution of phenyl-iso-thiocyanate in dioxane. The coupling reaction is catalyzed by expos-

ing the paper to the vapors of a pyridine, dioxane, water mixture at 40°.

After coupling, the paper is dried and excess reagent and diphenylthio-

urea extracted with benzene. Cyclization is accomplished by drying the

paper again, after which it is exposed at room temperature for seven

hours to the vapors from glacial acetic acid and concentrated HCl at a

total pressure of 100 mm. The paper is dried and extracted with acetone

to remove the phenylthiohydantoin. A variation is used when the hydant-

toins of histidine, cysteic acid, and arginine are formed, since these

compounds are not extracted by acetone. The phenylthiohydantoin is

recovered from the acetone extract and are identified by chromatography

on paper with the aid of the solvents developed by Sjöquist (74, 75).

An important feature of the procedure is a special chamber in which the

numerous drying steps are performed. The phenylthiohydantoin is found

to be particularly sensitive to oxidative destruction on the paper strips

at high humidities and in the presence of the oxidizing pollutants present

in city air. A number of precautions necessary for the reproducible chroma-

tography of the phenylthiohydantoin on paper were also described. The

procedure clearly represents the outcome of extensive experience. It was applied in the determination of the amino acid sequence of the γ -chain of human fetal hemoglobin. While the technique, as applied, was essentially qualitative, the generally excellent yields obtained at each step permitted unambiguous interpretation of the results with some peptides over as many as eight stages. The phenylthiohydantoins of serine, threonine, asparagine, and glutamine were identified as such; no difficulty was experienced with deamidation of glutamine and asparagine. Removal of glutamic acid and aspartic acid residues proceeded in the smallest yields.

When larger quantities of peptides are available the progress of the Edman degradation may be followed subtractively by analysis of a portion of the peptide remaining after the application of each stage.

In much recent work it has been common practice to combine both the direct and the subtractive approach to obtain unequivocal identification of the residue removed. Important progress in the application and understanding of the subtractive approach is due to the researches of Konigsberg &

Hill (7), and of Smyth, Stein & Moore (12). The contributions of the latter authors were made when the need to reinvestigate the amino acid sequence of ribonuclease in positions 11-18 became apparent from the work of Gross & Witkop (55), and of Allende & Richards (76). Thanks to Konigsberg & Hill a highly satisfactory procedure for the successive application of the cyclization to peptides has been developed. Reaction with phenyl-iso-thiocyanate in 50 percent ethanolic N-ethylmorpholine acetate buffer (pH 9) assures complete acylation of the α -amino group. For cyclization the crude phenylthiocarbamyl peptide is dissolved in anhydrous trifluoroacetic acid, in which such derivatives are readily soluble, and the solution is kept at room temperature. In control experiments it was found that several large tryptic peptides from hemoglobin were recovered unchanged from solution in trifluoroacetic acid; solvolysis therefore appears to be negligible (cf. Tascher & Kupryszewski (77) concerning studies on the N-acidolysis of peptide bonds). Experiments with synthetic dipeptides revealed that at 0° the phenylthiocarbamyl derivatives were cleaved quantitatively in two minutes or less. The reaction

required an hour as measured by the rate at which free amino acids appeared with the peptides Val.Gly and Ser.Gly. When the procedure finally adopted, with a cyclization time of one hour at room temperature, was applied routinely for sequence determination to peptides from hemoglobin the yield of degraded peptide in the cyclization step varied between 60 and 80 per cent. Further exposure to trifluoroacetic acid did not increase the recovery of peptides when relatively low yields were obtained. A ninhydrin-negative product that on amino acid analysis was found to possess the same composition as the starting peptide made up the bulk of the unusable residue. The nature of this product was not elucidated.

When removal of an amino acid brings a glutamine residue into the N-terminal position, cyclization to a derivative of pyrrolidone-carboxylic acid is possible. Smyth, Stein & Moore (12) have shown that in anhydrous trifluoroacetic acid at 25° this cyclization proceeds very slowly, if at all. By contrast, in glacial acetic acid saturated with anhydrous HCl at 100°, the conditions used by the present writer in earlier research on the amino acid sequence of ribonuclease (78), the cycliza-

tion to pyrrolidone-carboxylic acid peptides is almost complete; glutamine itself was quantitatively cyclized. Application of the step-wise procedure is of course blocked to the extent that pyrrolidone formation takes place. Moreover, as Smyth, Stein & Moore (12) showed in model experiments with glycylserine, the conditions for the cyclization in acetic acid saturated with HCl at 100° are such as to facilitate esterification of serine and threonine hydroxyl functions. Glycylserine was 90 percent esterified in one hour. When an O-acetyl serine (or threonine) residue becomes N-terminal, O- to N-acyl migration takes place at the alkaline pH maintained during phenylthiocarbamylation in the next stage of the degradation, with protection of the amino group from further reaction. An additional consequence of the exposure of peptides to acetic acid saturated with HCl at 100° was found by Smyth, Stein & Moore: the acidic peptide $\text{Pyrglu.His.MeSO}_2\text{.Asp.Ser.Ser.Thr.Ser}$ (Pyrglu = pyroglutamyl residue) from ribonuclease was completely transformed to a neutral peptide, probably by imide formation at the aspartic acid residue. It is likely that peptides that contain internal asparagine residues would

also undergo this reaction. Moreover when the pyroglutamyl peptide from ribonuclease was subjected to the same acid treatment after attempted phenylthiocarbamylation a product was recovered in which the serine content was approximately 0.6 residues and the threonine content about 0.3 residues smaller than in the starting material. The loss of serine and threonine remains without explanation, but was due to the conditions of cyclization, since the peptide was recovered unchanged from trifluoroacetic acid. An additional unexplained feature of the cyclization in acetic acid saturated with HCl at 100° is that in peptides containing tyrosine residues a progressive loss of this amino acid occurs as the degradation continues over several stages. This destruction of tyrosine residues is not observed in trifluoroacetic acid (12).

A further indispensable contribution made by Konigsberg & Hill (7) was the introduction of an ion-exchange step subsequent to the cleavage reaction to separate the degraded peptide in pure form from the other reaction products, particularly derivatives of the parent peptide, rather than to continue the degradation with a mixture at the next stage. The

introduction of the ion-exchange step greatly facilitates the interpretation of the results obtained in the subtractive approach. The mixture obtained in the cyclization is brought into dilute acetic acid. Small quantities of the phenylthiocarbamyl derivative of the terminal amino acid, formed by hydrolysis of residual quantities of the first intermediate in the cyclization, the corresponding anilinothiazolinone [cf. Edman (79)], are likely to accompany the phenylthiohydantoin. Nonpolar phenylthiohydantoins and phenylthiocarbamyl amino acids are extracted with benzene. The solution is applied to a small column of Dowex 50-X2, which retains peptides and the phenylthiohydantoins of the basic amino acids. The neutral phenylthiohydantoins and phenylthiocarbamyl amino acids, and the unidentified ninhydrin-negative product(s) formed in the cyclization are not retained. The adsorbed substances are fractionated by successive elution of the column with dilute acetic acid, and with pyridine acetate buffers of increasing molarity at pH 5. The residual peptide is usually eluted in 0.17 M buffer. Smyth, Stein & Moore (13) found that peptides containing cysteic acid were not retarded suffi-

ciently for adequate fractionation on the small column of resin routinely employed. When this difficulty was encountered, it was generally found to be possible to perform three stages of the degradation without fractionation after each stage, and then to subject the peptide mixture to chromatography on much longer columns of Dowex 50-X2 to obtain a pure peptide.

The recognition of the shortcomings of cyclization in acetic acid saturated with HCl at 100° placed in doubt many of the sequences previously arrived at in ribonuclease by the present writer (78) on the basis of subtractive Edman degradations in which these conditions had been employed routinely. In consequence, Smyth, Stein & Moore (13) repeated under the improved conditions all the Edman degradations that could have led to erroneous conclusions in the earlier work, and included in their reinvestigation most of the remainder as well. The results obtained brought to light several errors in the sequence and ultimately led to the following corrections: residues 11-18, Gln.His.Met.Asp.Ser. Ser.Thr.Ser; 69-71, Gln.Thr.Asn; 87-89, Thr.Gly.Ser; 94-96, Asn.Cys.Ala;

and 101-103, Gln.Ala.Asn. The revised sequence in the region of residues 11-18 was fully corroborated by the independent studies of Potts et al. (80). The use of glacial acetic acid saturated with HCl at 100° for the cyclization step has been avoided in recently reported sequence analyses. For example, in his studies with peptides from cytochrome c, Margoliash (81) employed a mixture of glacial acetic acid and concentrated HCl at 40°, and a similar procedure was employed by Canfield in work on the amino acid sequence of lysozyme. Canfield (52) noted the recovery of the phenylthiohydantoins of serine and threonine under such conditions of cyclization.

COVALENT STRUCTURES OF PROTEINS

Selected Studies on Individual Proteins

Clupeine.--Ando et. al. (82) in a preliminary communication have announced the determination of the amino acid sequence of one of the components of the clupeine from the Pacific herring. The total clupeine was fractionated on alumina into two chromatographic components, one of which, component Z, appeared to be homogeneous as judged by its

behavior in counter-current distribution. It possessed a single amino terminal residue of alanine per molecule of empirical formula (Ala₃, Ser₃, Pro₂, Val₂, Arg₂₁), (molecular weight, approximately 5000). The clupeine was subjected to the action of trypsin, and the peptides formed were fractionated on Amberlite IRC-50 with the aid of an elution program that involved sodium borate buffers at pH 8.1 with increasing concentrations of sodium chloride. Quantitative recovery of the peptides was achieved. The peptides, mostly di- and tripeptides of arginine, were detected with the Sakaguchi reaction and with ninhydrin. Their amino acid sequences were obtained largely by end-group procedures. The problem of obtaining overlaps of the tryptic peptides was elegantly solved by application of the sulfuric acid-catalyzed N- to O-acyl migration at the three serine residues in the molecule. Clupeine Z was kept in concentrated sulfuric acid at 25° for four days, the product was acetylated, and the acetylated, rearranged poly-peptide was subjected to mild alkaline hydrolysis (0.3 N sodium carbonate at 30° for four hours). The four fragments, separated on Amberlite IRC-50 in acid solvents, were charac-

terized by analysis and by determination of their N- and C-terminal residues. The sequential release of amino acids from these larger fragments during carboxypeptidase hydrolysis was also determined. The combined results led to the sequence shown in Figure 1. Tryptic cleavage at the two Arg.Pro bonds in the molecule, as was to be expected from the well-known resistance of such bonds to the enzyme, proceeded relatively slowly and was responsible for the presence in the hydrolysate of peptides that were useful in the derivation of the sequence.

Insulin.--Kotaki (83,84) has investigated the amino acid sequence of one of the components, component II (85), of the insulins of the bonito, and has derived partial formulae for both of the peptide chains in the molecule. These are shown in Figure 2 and were arrived at by procedures very similar to those employed by Sanger and his colleagues in their classical investigations. Dixon and his colleagues (86) have cleaved the disulfide bonds in insulins from various animals by sulfitolysis. The S-sulfo derivatives of the A- and B-chains were separated by electrophoresis on paper. When the separated S-sulfo derivatives were

mixed with an excess of mercaptoethanol, and the solution was exposed to air, regeneration of beef insulin in a yield between 1 and 2 percent occurred. More recently Wilson, Wardlaw & Dixon (87) have shown that it is possible in this way to form "hybrid" insulins from the separated chains of the insulins from cattle and codfish. The yield of the "hybrid" molecules was comparable to that obtained when chains from a single species were recombined.

Cytochrome c.--The amino acid sequence of horse heart cytochrome c has been elucidated through the combined researches of Margoliash & Smith (88, 89, 81), and of Tuppy & Kreil (90). The protein possesses an acetylated N-terminal amino acid residue, the presence of which was demonstrated by both groups, as well as by Titani, Narita & Okunuki [cf. Tuppy & Kreil (90)]. The amino acid composition of cytochrome c is noteworthy for the absence of serine and the presence of only two cysteine residues, both of which are used to attach, by thioether bonds, the vinyl side-chains of the protoporphyrin moiety. The composition was determined by Margoliash et al. (88) by analysis of both acid and total enzymic

hydrolysates. Negligible quantities of serine were present in acid hydrolysates, but hydrolysates produced by the consecutive action of papain at pH 6, and a mixture of leucine aminopeptidase and prolidase at pH 8 contained the equivalent of two residues of serine. The origin of the serine is uncertain, for the subsequent work unequivocally demonstrated the absence of this amino acid from the protein. The authors have suggested the possibility that the presence of manganous ion (used as the metal activator for the aminopeptidase) and the heme from the protein may have led to the formation of serine from the thioether-bonded cysteine residues.

The amino acid sequence was determined from studies of the peptides in tryptic and chymotryptic hydrolysates of the denatured protein. Margoliash & Smith (89) obtained the chymotryptic peptides at pH 7.8. With an initial enzyme concentration of 2 percent the rate of hydrolysis slowed down considerably after the cleavage of eight bonds. On doubling the enzyme concentration, the rate increased sharply, to diminish again after approximately 12 bonds had been cleaved. A final

addition of enzyme at this stage again increased the rate, and the reaction was stopped after the cleavage of 16 bonds, at which point the hydrolysis was still progressing at a significant rate. This behavior is typical of a proteolysis by chymotrypsin in which bonds with very widely differing specific rates of hydrolysis are being cleaved under conditions where autolysis of the enzyme may take place. An initial fractionation of the peptides was achieved by chromatography on the pyridine form of Dowex 50-X2. The starting buffer, 0.2 M pyridine acetate at pH 3.1, served to elute the first ten peptide fractions. A linear gradient of pH and buffer concentration between pH 3.1 and 5.0 and 0.2 M and 2 M was applied after 13 column volumes; it effected the elution of a further 18 peptide fractions. Continued elution with 2.0 M pyridine acetate at pH 5.0 furnished an additional seven fractions. Finally, the column was washed with 3.0 M trimethylamine to elute two very strongly adsorbed peptide zones. Of the total of 37 fractions eluted 24 were needed in later studies. However, all the fractions present were used to construct a peptide "map" obtained by electrophoresis.

(at pH 6.4) and chromatography in a butanol-pyridine-dilute acetic acid solvent. The peptide "map" was invaluable in revealing the homogeneity of the fractions obtained by chromatography on Dowex 50 and was used to devise conditions for the preparative separation of individual peptides, either by paper chromatography or electrophoresis alone, or sometimes in combination. Most of the peptides were purified to the extent that they contained less than 10 percent impurity as judged by analysis with an amino acid analyzer. A total of 28 fragments were ultimately defined by this approach. Of these 13 were obtained in final over-all yields greater than 25 percent. Among the remainder, several were larger peptides subsequently shown to be incomplete hydrolysis products from which two or more of the smaller peptides could be derived. In other cases the low yield could be attributed to incomplete multiple cleavages and to the presence in the peptides of a relatively high content of basic or hydrophobic residues, attributes which contribute a high affinity for sulfonated polystyrene resins.

The portion of the cytochrome c molecule associated with the

heme group appeared in the chymotryptic hydrolysate as a heme peptide fraction which could not be eluted from Dowex 50-X2 except under drastic conditions. A separate experiment was performed to obtain this fraction by chromatography over IRC-50 at pH 8.4 in ammonium acetate buffer. The majority of the peptides were eluted with the initial buffer; the colored fraction was eluted at pH 9.5. The fragments containing heme in this fraction were found to be sufficiently large (or aggregated) to permit isolation and some purification by dialysis. Subsequent estimation of the average molecular weight of the heme peptide fraction from measurements of the heme content gave a value of about 4000. Analysis revealed the heme peptide fraction to be impure and further attempts to isolate a single peptide from it by a variety of methods were unsuccessful. Accordingly, it was subjected to tryptic hydrolysis. The derived peptides were separated by two-dimensional electrophoresis and chromatography on paper, and the separations were repeated a number of times to accumulate sufficient quantities of the peptides for further study. Eight principal fragments were obtained, of which one was free histidine, and two were heme

peptides that contained nine and 12 residues. The heme peptide of nine residues had the same amino acid composition as the heme peptide identified by Tuppy & Bodo (91, 92, 93) in tryptic hydrolysates of cytochrome c, while the peptide of 12 residues corresponded in composition, with the addition of a lysine residue, to a peptide obtained by Margoliash, Frohwirt & Weiner (94) from peptic hydrolysates of the parent protein. The larger heme peptide contained two of the three valine residues in cytochrome c and could be shown to derive from the amino-terminal portion of the chain. The remaining tryptic fragments of the heme peptide fraction derived from sequences carboxyl-terminal to the sequence of the heme peptides in the manner shown in Figure 3. The demonstration of free histidine and of a peptide that was probably Gly.Gly.Lys.His in the tryptic hydrolysate indicated that the original heme peptide fraction contained a subfragment formed by a chymotryptic cleavage at the histidine residue shown in position 26. Indeed a peptide identical in composition to the 28-36 residue fragment (cf. Figure 4), but containing an additional lysine residue, was isolated from the

Dowex 50-X2 eluates.

The residues demonstrated as constituents of the heme-peptide fraction and in nonoverlapping peptides from the Dowex 50-X2 separation completely accounted for the composition of the protein. Margoliash (81) proceeded with systematic sequence analyses of the chymotryptic peptides. The general approach was to obtain as much information as possible by enzymic methods before resort was made to sequential degradation. Large peptides that contained basic residues were frequently subjected to tryptic hydrolysis to derive subfragments. In other instances the larger peptides were also subjected to the action of papain. The consequences of these enzymic hydrolyses were routinely examined by two-dimensional electrophoresis and chromatography on paper. As in the separation of the principal chymotryptic peptides themselves, the results of these analytical separations were used in the design of suitable preparative separations on paper. The compositions of the subpeptides were determined quantitatively. On the other hand, the hydrolysis of the principal peptides and of their subfragments by either leucine aminopeptidase or

carboxypeptidase A was studied semiquantitatively by paper chromatography. The chromatograms frequently provided unequivocal identification of asparagine and glutamine residues. In many other instances the electrophoretic behavior of peptides provided information upon which the allocation of amide groups to residues of the dicarboxylic amino acids could proceed with considerable assurance. In peptides that contained proline cessation of leucine aminopeptidase action was regarded as evidence that successive removal of amino acid residues had progressed to within one residue of the proline position (residues 44 and 71 in Figure 4).

Edman degradations were performed routinely by reaction of the peptides at 40° with phenyl-iso-thiocyanate in 50 percent pyridine to which a small quantity of trimethylamine was added. Cyclization was effected in a mixture of five parts glacial acetic acid and one part concentrated HCl at 40°. Most cyclizations were relatively rapid and required one to two hours; with glycine end-groups, however, the yields reached 80 percent only after 12 hours. The phenylthiohydantoins were identified

by paper chromatography (74, 95), while a portion of the residual peptide fraction in most of the stages was subjected to amino acid analysis. The largest number of stages through which the degradation was carried was five; only one of these involved a glutamine residue, which was identified in a yield of approximately 50 percent by difference analysis and by characterization of the phenylthiohydantoin.

A peptide of considerable interest was the one designated C-XIX a, shown subsequently to be Acetyl.Gly.Asp.Val.Glu.Lys.Gly.Lys.Lys.Ileu.Phe. This peptide was not attacked by leucine aminopeptidase, behaved as a neutral peptide on electrophoresis, and contained no α -amino group detectable by dinitrophenylation or by application of the Edman degradation. These properties suggested the presence of an acyl group at the N-terminus. A carefully purified sample of the peptide was subjected to acid hydrolysis. Gas chromatography of the products that could be volatilized from the hydrolysate revealed the presence of acetic acid. The demonstration of the presence of an acetyl group identified peptide C-XIX a as the peptide originating from the amino-terminal seg-

ment of the peptide chain.

Tuppy & Kreil (90) obtained the tryptic peptides of denatured horse heart cytochrome c at pH 8.0 and were able to show that in spite of acid pretreatment of their trypsin preparation a number of typically chymotryptic cleavages were unavoidable. The amino acid composition of the protein showed that complete tryptic cleavage should have afforded 20 to 22 peptides, including the tryptic heme-peptide already characterized by Tuppy & Bodo (91, 92, 93), but the existence of several Lys.Lys and one Arg.Lys sequences ^{that were} incompletely cleaved increased the number of tryptic fragments to 29. An initial separation of these fragments was attained by two-dimensional electrophoresis and chromatography: the electrophoretic step in pyridine acetate buffer at pH 4.7, and the chromatographic step with butanol-acetic acid-water (4:1:5). The peptides were then subjected to electrophoresis at either pH 6.4 (pyridine acetate) or in dilute formic acid at pH 2 for final purification. Methionine-containing peptides were pretreated with thioglycollic acid to reduce sulfoxide formed during previous steps, as described by Harris &

Roos (96). The purity of the peptides was judged on the basis of the results of end-group analysis by dinitrophenylation. When pure the peptides were subjected to amino acid analysis by qualitative procedures. The regular spacing and relatively large number of basic amino acid residues in the peptide chain was particularly advantageous. Three of the peptides were dodecapeptides; the remainder were smaller, for the most part di-, tri-, and tetrapeptides. The amino acid sequence of one of the dodecapeptides was established; the sequences of the remaining two were partially elucidated. The sequences of all but two of the smaller peptides were determined. These investigations involved partial hydrolysis experiments with enzymes or with concentrated HCl at 37°. The fragments produced were separated electrophoretically, and their N-terminal amino acids and compositions determined. The enzymes used included chymotrypsin, pepsin, subtilisin, and oxytocinase [cf. Tuppy & Wintersberger (97)]. Of particular interest in the investigation was the work connected with the identification of the N-terminal sequence in the protein. This was present in the tryptic peptide designated T 20, which contained the amino

acids (Gly, Asp, Glu, Val, Lys). On cleavage with pepsin it afforded two fragments: T 20 P 1 and T 20 P 2. The former proved to be Val.Glu.Lys, while T 20 P 2 was ninhydrin-negative and contained glycine and aspartic acid; it could be detected by chlorination or with the aniline-glucose reagent of Michl & Hogenauer (98). On hydrazinolysis T 20 P 2 yielded glycine hydrazide, aspartic acid, and acetyl hydrazide. When the peptide was heated in 0.01 N HCl at 105° for 1.5 hours [cf. (60)] cleavage occurred in good yield to acetyl glycine and aspartic acid. Acetyl glycine was identified by chromatographic comparison with a series of simple acyl glycines. Peptide T 20 was thus revealed to be Acetyl.Gly.Asp.Val.Glu.Lys.

The detailed information obtained by study of the tryptic and chymotryptic peptides permitted the unambiguous determination (99) of the complete amino acid sequence shown in Figure 4. A similar approach has been employed by Matsubara & Smith (100) to elucidate the amino acid sequence of human heart cytochrome c. A chromatographic separation essentially identical to that utilized by Margoliash & Smith (89) for the sep-

aration of the chymotryptic peptides from the horse protein was employed in the separation of both the chymotryptic and tryptic peptides from the human material. The amino acid sequences of the chymotryptic peptides were determined, and the relative alignment of these peptides was established from the composition of the tryptic peptides, and from consideration of the homology that must exist in the structures of the protein from two mammalian species. In this regard it is of interest that Matsubara & Smith detected the presence of a peptide Gly.Glu.Asp.Thr.Leu.Leu in the chymotryptic hydrolysate from human heart cytochrome c. This peptide could not have been derived from the amino acid sequence developed for the protein, and since the peptide was isolated in relatively low yield from a sample that originated in a large number of individuals, the possibility exists that a genetically-determined variant of the protein was present in the preparation analyzed. On this basis, the more prevalent chymotryptic peptide in the sample was Gly.Glu.Asp.Thr.Leu.Met (residues 60 to 65). The investigation of differences among cytochrome c molecules from a wide range of different species is currently under

investigation, and the results will be of great interest in relation to the function of the molecule as well as in connection with studies of biochemical evolution (101).

Lysozyme.--Jollès and his colleagues have for several years been engaged in a comprehensive study of the covalent structure of the lysozyme of hen egg-white (102-108). An independent examination of the amino acid sequence was undertaken more recently by Canfield (50, 51, 52) in Anfinsen's laboratory, and the availability of his results makes a review of the present position in this field appropriate. Canfield cleaved the disulfide bonds of lysozyme by reduction with mercaptoethanol in 10 M urea solution. The reduced protein was carboxymethylated in 8 M urea in essentially quantitative yield; no side-reactions with lysine, histidine, or methionine side-chains were observed. Reduced, carboxymethylated lysozyme was subjected to tryptic hydrolysis to afford a total of 23 fragments in unspecified, but evidently relatively high yield. These fragments were fractionated by chromatography on columns of Dowex 50-X₄ operated with pyridine acetate buffers. A gradient of pH and

buffer concentration between pH 3.8, 0.1 M acetate and pH 5.2, 2 M acetate served to elute all but one of the products in the temperature range from 35 to 50°: the remaining peptide was eluted with ammonium acetate solution. The majority of the tryptic peptides were obtained in a homogeneous condition, but when mixtures were present, these were separated by electrophoresis on paper at pH 3.6. The results of quantitative amino acid analyses demonstrated that the tryptic fragments provided for an accounting of all the amino acid residues in the molecule. Five of the peptides isolated represented incomplete hydrolysis products: prolongation of the exposure of the protein to trypsin decreased the levels of these peptides, but attendant slow cleavages at bonds other than the carbonyl bonds of lysine and arginine residues led to diminished yields of the principal peptides. The tryptic fragments were identified in this manner as the amino acids leucine (1 equivalent), arginine (1 equivalent), lysine (2 equivalents), and 14 peptides that varied in chain-length from two to 22 residues. Four of the principal peptides contained tryptophan, the content of which was in each instance determined by analysis of a total

hydrolysate produced by the action of leucine aminopeptidase. The complete analysis of the tryptic peptides accurately defined the amino acid composition of the protein and defined the number of residues in the single polypeptide chain of this enzyme as 129. The composition derived from the peptides agreed precisely with amino acid analyses performed with the intact protein.

The relative alignment of the tryptic peptides along the chain was established by study of the peptides produced from the reduced, carboxymethylated enzyme with chymotrypsin and pepsin. These peptides were chromatographed on columns of phosphorylated cellulose by Canfield & Anfinsen (50), who have described some of the advantages of this exchanger for the fractionation of peptides. They note, in particular, the high exchange capacity and the relative lack of buffering power in the pH range between 2.5 and 5.5, which permits the use of eluents of low ionic strength. It would appear, however, that the recovery of peptides from phosphocellulose is not quantitative. The separation of the tryptic peptides from reduced, carboxymethylated lysozyme on phosphocellulose was

not as satisfactory as on sulfonated polystyrene, and the recovery of some peptides was notably low. The chymotryptic peptides were separated by gradient elution with ammonium acetate buffers of increasing molarity in the pH range from 3.8 to 8.35. The use of ammonium acetate buffers permitted the determination of the absorbancy of the effluent fractions at 280 m μ , and thus aided in the identification of peptides that contained tyrosine and tryptophan residues. The ammonium acetate was removed by volatilization at 50° with no apparent effect on the distribution of side-chain carboxyl and carboxamide residues. The peptic hydrolysis products from reduced, carboxymethylated lysozyme were separated by an elution program with pyridine acetate buffers similar to that used in the separation of the tryptic peptides on Dowex 50-X4. Mixtures of the chymotryptic and peptic fragments still present after the fractionation on phosphocellulose were separated by electrophoresis on paper at pH 3.6.

In the subsequent analysis Canfield (52) relied heavily on the technique of "reciprocal" cleavages: the tryptic peptides were subjected to chymotryptic and peptic hydrolyses, while the chymotryptic

and peptic peptides were fragmented with trypsin. These reciprocal enzymic hydrolyses were frequently augmented by further studies with the principal peptides, or their subfragments, in which subtilisin was used to obtain relatively large numbers of small peptides. The mixtures produced in these various enzymic hydrolyses were resolved by electrophoresis on paper. The composition of many of the relatively simple subfragments could be determined by the paper electrophoretic procedure of Dreyer & Bynum [referred to in Dreyer (109)], instead of with an amino acid analyzer. In this procedure the amino acids are subjected to high-voltage electrophoresis in 4 percent aqueous formic acid for 20 minutes. Subsequent to staining with a modified ninhydrin reagent, the papers are scanned with an integrating densitometer. The procedure has the advantage that several analyses may be performed simultaneously in a very short time with approximately 0.02 micromole samples of peptide. However, the separation of valine from serine, and of proline from methionine is difficult, and, as may be expected, the over-all accuracy is not as satisfactory as that obtainable with an automatic analyzer. It would

appear that the principal value of the method will be in the monitoring of degradation experiments with relatively small and thoroughly purified peptides, the compositions of which have already been determined by analyzer methods.

Canfield employed the subtractive technique of Edman degradation and time studies of the release of free amino acids from peptides by leucine aminopeptidase and the carboxypeptidases A and B for his sequence determinations. The attendant amino acid analyses were frequently performed by the method of Dreyer & Bynum. Unfortunately, an objective appraisal of these sequence determinations is difficult. The values obtained in the quantitative or semiquantitative analyses were not presented. The nature and variety of the subfragments obtained enzymically from the principal peptides made unnecessary the performance of more than four successive stages of the Edman procedure on any one peptide. None of the steps involved glutamine. The degradations were performed under conditions similar to those described by Margoliash (81), and are illustrated with densitometer tracings obtained in three successive steps with

the peptide Gly.Thr.Asn.Val.Glu.Ala. Removal of glycine in the first stage proceeded in high yield; approximately 40 percent of the threonine was missing from the residue after the second stage; and about 20 percent of the asparagine residue in the third position was removed in the final stage. The allocation of amide groups proceeded almost entirely from a consideration of the results obtained by quantitative amino acid analysis of leucine aminopeptidase hydrolysates of appropriate peptides. The final conclusions are summarized in the sequence shown in Figure 5. Further work will be required to define the sequence of the residues shown in positions 79, 80, and 81. The evidence for the sequence between residues 98 and 105 derives from the study of the principal tryptic fragment of the chymotryptic peptide that originated from residues 94 to 105 (cf. Figure 5). Residues 98 to 101 were removed by Edman degradation, and the methionine residue in position 105 was revealed by the action of carboxypeptidase A. Leucine aminopeptidase hydrolysis of this peptide was accompanied by the liberation of equimolar quantities of glycine and aspartic acid before any free methionine was detected. This

demonstrates that glycine in the sequence must be preceded by at least one aspartic acid residue. The reliability of the identification of the aspartic acid residue in position 101 thus assumes importance, and Canfield has emphasized the desirability of independent checks of the 101-102 sequence.

The results obtained by Jollès and his colleagues have for the most part been described in outline only, and definitive reports are therefore awaited with considerable interest. Lysozyme was reduced with thioglycollic acid, and the reduced protein was carboxymethylated (103). The reduced, carboxymethylated lysozyme proved to be much more readily attacked by trypsin and chymotrypsin than were the heat-denatured or performic acid-oxidized lysozyme employed in the earlier investigations of this group (102). Tryptic hydrolysates of reduced, carboxymethylated lysozyme were fractionated by chromatography over the sodium form of Dowex 50-X2 under conditions similar to those used for the separation of the tryptic peptides from performic acid-oxidized ribonuclease (110). The peptide fractions were examined by electrophoresis on paper at pH

3.6 and 6.5, and by paper chromatography in a butanol, aqueous formic acid solvent. When the need for further purification of a peptide fraction became apparent the results of these preliminary tests were employed as guides for the design of preparative separations for each major component. The amino acid analyses were in part obtained manually by the procedure of Moore, Spackman & Stein (111), and more recently with an amino acid analyzer. In addition to free leucine, lysine, and arginine, all but one of the series of principal tryptic peptides also identified by Canfield (51) were characterized. This peptide was Canfield's T-9, Try.Try.CMCys.Asn.Asp.Gly.Arg, which was probably not eluted from Dowex 50-X2 under the conditions employed by Jollès et. al (106). However, when Jollès and his colleagues subjected their chymotryptic peptide CR 19 to tryptic hydrolysis they obtained a peptide differing from Canfield's T-9 only in that it possessed one less residue of tryptophan, and from its unique composition deduced that it represented a portion of a tryptic peptide suspected to be outstanding from a consideration of the amino acid composition of the parent protein.

The tryptic peptides were subjected to further hydrolyses with chymotrypsin and pepsin in a manner analogous to that subsequently described in more detail by Canfield. The subfragments obtained in these experiments were subjected to Edman degradation, and to hydrolytic studies with carboxypeptidase A. The majority of the fragments studied in this manner were apparently tri- or tetrapeptidases, the structures of which were readily deduced (106). When enzymic hydrolyses failed to furnish small peptides, resort was made to partial acid hydrolysis (12 N HCl at 37°). The results presented, however, indicate that several of the larger tryptic peptides were not studied in detail. A series of chymotryptic peptides were obtained initially by procedures similar to those used for the tryptic peptides. In later work the chymotryptic peptides were fractionated on columns of Dowex 1-X2: several peptides containing aromatic and/or basic residues that failed to elute from Dowex 50-X2 were obtained in addition to those peptides already characterized. Twelve of the chymotryptic peptides characterized contained basic amino acid residues and were therefore important for the estab-

lishment of the arrangement of the tryptic peptides along the chain. The amino acid sequences of a dipeptide, three tripeptides, a pentapeptide, and a hexapeptide in this group were determined, while partial sequences were arrived at for the rest. The arrangement of tryptic peptides as far as residue 114 was derived by a consideration of the partial sequences of the chymotryptic peptides that contained basic amino acids. The two tryptic peptides in the region from residues 115 to 125 were connected with the aid of the composition of a peptide isolated from a tryptic hydrolysate of the reduced, acetylated protein, in which the action of trypsin was limited to the carbonyl bonds of the arginine residues. This connection was confirmed by the chymotryptic peptide Arg.Asn. The combined results of these studies in 1961 led to the proposal of a partial sequence that showed the placement of 100 residues in the chain (107). Further studies of the tryptic peptides increased this total to 119 by 1963 (108).

Jollès and his colleagues have also undertaken (108) a study of the disulfide bonds in lysozyme by an approach essentially similar to

that employed by Spackman, Stein & Moore (112) with ribonuclease. The products from an initial peptic hydrolysis of the enzyme at pH 1.9 were subjected to chymotrypsin action at pH 6.75. The peptides formed were chromatographed on a column of Dowex 50-X2. One of the cystine-containing zones was found to contain a component that on performic acid oxidation furnished only two cysteic acid-containing peptides. These were identified from the amino acid sequence as peptides that originated from the sections of the molecule around the first and second of the eight half-cystine residues.

The results obtained in the sequence analyses by the two groups are in essential agreement. Differences between the conclusions reached are shown in Table 1. An earlier formula shown by Jollès (107) contained 130 residues: the additional residue, a serine residue, was located around position 117.

Papain.--Detailed reports of the progress made towards the elucidation of the amino acid sequence of papain have been presented by the group in Smith's laboratory (129-134). The research has been hampered by

the resistance of the protein to attack by proteolytic enzymes. Papain denatured by heat or acid, performic acid-oxidized papain, and carboxymethylated papain cannot be cleanly digested with trypsin or chymotrypsin. The peptides characterized from the protein, with one or two exceptions, have all been obtained in relatively low yield. Moreover, many of the peptides from papain could not be eluted satisfactorily from ion-exchange adsorbents. This difficulty was particularly pronounced with tryptophan peptides, only one of which was isolated in a satisfactory yield. These and similar difficulties have had as their consequence the fact that a complete accounting of all the residues in the molecule in terms of well-defined peptides has still to be attained. Nonetheless, it is a tribute to the skill and perseverance of this group that of the 185 amino acid residues in papain, 148 may now be placed in known sequences, many of which have been connected by appropriate overlap peptides. The research with papain should serve as a warning to those who may be complacent enough to delude themselves into the comforting belief that the elucidation of the amino acid sequences of proteins has by now become a matter of routine.

Protein of tobacco mosaic virus.--Essentially complete amino

acid sequences for the protein of tobacco mosaic virus were put forward in 1960 by Anderer et al. (113) and by Tsugita et al. (114). The "Tubingen group has published final memoirs of their researches (115, 116, 70).

Unfortunately, in their desire to provide a concise summary of the very large number of experiments performed, the German authors have presented their results schematically and have therefore deprived others of the benefits of much valuable experience. Since it appears likely that amino acid sequence elucidation will remain a preoccupation of an increasing number of biochemists, this reviewer wishes to enter a plea that sequence work be documented with full experimental details, and, wherever possible, with quantitative data. Otherwise, as in the research reported by Anderer and his colleagues, it will be impossible to reach an objective appraisal of the conclusiveness of the sequence derived.

Anderer & Handschuh (70) adjusted a tryptic hydrolysate of the virus protein to pH 4.7 to obtain a precipitate which was subsequently found to consist of two tryptic peptides, designated fractions 1 and 2.

Fraction 1 was the peptide of 41 residues recognized in earlier work (117, 118) to contain the terminal N-acetyl group. It was separated from fraction 2 (residues 123-134), another sparingly soluble peptide, by gel-filtration over Sephadex G-50. The remaining tryptic peptides were obtained in high yield by a slight modification of the procedure of Wittman & Braunitzer (119) with columns of Dowex 1-X2, as originally applied with notable success by Braunitzer and his students for the fractionation of peptides from hemoglobin [cf. Braunitzer et al. (120, 121) and Schroeder et al. (122)]. Tryptic peptide 4 (residues 47-61) possesses an amino-terminal glutamine residue and underwent cyclization in part to a pyro-glutamyl peptide during isolation. This reaction complicated the separation of peptides 2, 3, and 4, and an additional gel-filtration step was necessary for the complete resolution of the mixture.

The amino acid sequences of the tryptic peptides were determined by methods monitored with an amino acid analyzer. Subfragments of the larger peptides were obtained by the action of chymotrypsin or pepsin, and were separated by chromatography either on Dowex 50-X2 or

Dowex 1-X2 columns, or, in a few instances, by high-voltage preparative electrophoresis on paper. The structures of the subfragments and of the smaller principal peptides were determined largely by studies of the release with time of free amino acids in hydrolyses with leucine aminopeptidase or the carboxypeptidases, and by sequential degradations. In some instances, important information was also obtained by partial acid hydrolysis. The Edman degradations were performed by a slight modification of the procedure of Eriksson & Sjöquist (123), and were carried to a maximum of four stages. The cyclization of the phenylthiocarbamyl peptides took place in a mixture of one volume of water and two volumes of glacial acetic acid saturated with HCl at 40°. The phenylthiohydantoins were apparently frequently identified by Sjöquist's procedure (75), and after each stage of the degradation a portion of the residual peptide was routinely subjected to amino acid analysis. The results indicate that no difficulties were encountered in the routine application of the degradation. Particular attention was focussed on fraction 1, and the amino acid sequences of several subfragments of this large peptide

were reinvestigated [cf. (118)]. A major source of difficulty in the earlier experiments was the sequence around the cysteine residue in position 27 (cf. Figure 6). Cysteine-containing peptides were obtained in very poor yield after fractionation on Dowex 50. This problem was obviated when fraction 1 was subjected to performic acid-oxidation. Though the cause of the modification of the tryptophan residue in position 17, the oxidation of fraction 1 permitted the isolation in good yield of peptic peptides that contained residues 27-29, 27-31, and 24-29.

Anderer (115) fractionated a chymotryptic hydrolysate of the protein on columns of the sodium form of Dowex 50-X2 under conditions similar to those used with peptides from ribonuclease (110). The separation afforded 18 peptide fractions. Most of these were dinitrophenylated, and the mixture of dinitrophenyl peptides was in each instance resolved by paper chromatography in one of two solvents. These studies led to the characterization of 37 peptides, for each of which the N-terminal group and the amino acid composition were determined by the procedure

of Levy (124). Partial sequences of the peptides were obtained by study of the products formed from them by partial acid hydrolysis, and by the action of carboxypeptidase A. A peptide of particular interest was obtained in fraction 11 and represented the portion of the parent molecule between residues 88 and 108. The composition of this peptide was determined with an amino acid analyzer. The slow liberation of isoleucine from the Ileu.Ileu sequence in positions 93-94 was revealed by time studies of the content of isoleucine in acid hydrolysates of the peptide. The detection of an additional isoleucine residue that had been missed in the earlier work enabled Anderer to correct the total composition of the protein by one residue and thereby brought the empirical formula in line with that derived by the Berkeley group (114). An erroneous composition was also derived in the earlier work for the chymotryptic peptide from positions 82-84. This peptide was handled as its dinitrophenyl derivative. The extended hydrolysis requisite for the liberation of leucine from the Leu.Leu sequence was attended by decomposition of dinitrophenylalanine, and led to the conclusion that the peptide was DNP-Ala.Leu.Leu.

Leu. The additional leucine residue ascribed to the peptide was subsequently assigned to position 76 by the isolation of peptides from residues 73-77 and 73-81.

The chymotryptic peptides characterized provided information about the sequences around the two lysine and around six of the 11 arginine residues. The remaining arginine residues were present in peptides that could not be eluted from columns of Dowex 50-X2 operated in the sodium form [cf. Jollès et al. (106)]. A similar problem was encountered when a peptic hydrolysate of the virus protein (116) was fractionated on columns of the sodium form of Dowex 50-X2. Some 45 peptides, many of which were formed in very low yield, were characterized, and partial sequences were deduced for them by the procedures employed in the study of the chymotryptic peptides. Additional peptides that increased the representation of amino acids to include all the residues in the protein except the tripeptide sequence around the cysteine residue in position 27 were obtained by chromatographic fractionation of peptic hydrolysates on columns of Dowex 1-X2 by the methods of Braunitzer (121). The chroma-

tograms were surveyed by ninhydrin analysis and by tests with the Sakaguchi reaction. Only fractions that contained arginine or lysine were studied further.

The sequence finally arrived at by the ["]Tubingen workers (70) differs from that proposed by Tsugita et al. (114) by an inversion of the residues in positions 25-26, and by differences in the locations of certain amide groups. The formula shown by Tsugita et al. shows an amide group at position 22, but not at position 26; an amide group at position 98, but not at position 97; and no amide group at position 106. The most recent researches of the Berkeley workers have not been documented. The earlier work of Gish & Ramachandran (125-128) was concerned with a study of the tryptic peptides, and Narita (117) firmly established the presence of the terminal acetyl group. The initial fractionation of the tryptic peptides was achieved by extensive counter-current distributions (125), after which individual peptides were separated by electrophoresis in cellulose powder blocks. The nine peptides investigated contained 85 residues, present in position 42-46, 47-61, 62-68, 69-71, 92-93, 93-112, 113-122, 135-141, and 142-158. The compositions of

most of these peptides were determined by the dinitrophenylation procedure of Levy (124). Analysis of the peptide from positions 62-68 by this procedure led to the conclusion that it contained only one aspartic acid residue. This error was recognized subsequently (114). The sequences in the tryptic peptides studied by Gish & Ramachandran (126) were determined by a study of subfragments formed in further hydrolyses with either chymotrypsin, pepsin, or subtilisin. Many of these studies were also monitored by amino acid analyses performed by the procedure of Levy (124).

The composition of the peptide from residues 93-112, however, was determined with an automatic analyzer, and the presence of two isoleucine residues was clearly recognized (128). The sequence in positions 135-138 was inferred to be Gly.Thr.Ser.Gly from time studies of the release of amino acids by the action of leucine aminopeptidase on the tryptic peptide from residues 135-141. Subsequent re-examination of the peptide furnished results that required revision to the sequence shown in Figure 6.

α -Chymotrypsinogen and chymotrypsin.--Reports of progress in

studies on the covalent structure of bovine α -chymotrypsinogen have been presented by Hartley (36) and by the group associated with Sorm (135-137).

The principal difficulty experienced in these studies, as it has been with papain, is that all of the approximately 240 residues in the molecule have not been identified in the peptide fragments thus far characterized. In this context research in Desnuelle's laboratory aimed at the separation of the three peptide chains in α -chymotrypsin (138, 139) has therefore assumed considerable importance, and, as pointed out by Hartley (36), may provide a more favorable access to the sequence problem. Performic acid-oxidized α -chymotrypsin (138) was subjected to gel-filtration in dilute acid on a column of Sephadex G-50. The relatively short A-chain was cleanly separated from the C-chain, which, in turn, was satisfactorily resolved from a mixture that appeared to consist of both B- and C-chains. Subsequently the same group described a very satisfactory separation of the B- and C-chains from S-sulfo-chymotrypsin by chromatography over DEAE-cellulose (139).

Preliminary reports have appeared from Sorm's laboratory (137) in which the identification of the disulfide bonds in chymotrypsinogen is described.

Trypsinogen.--Preliminary reports of structural studies with bovine trypsinogen have appeared. In Neurath's laboratory (140, 141, 44) the principal effort has been a study of peptides derived from S-sulfo-trypsinogen. The group with Sorm (142, 143) are studying DFP-trypsin.

Tryptophan synthetase.--The A protein of the tryptophan synthetase of E. coli has been isolated in pure form by Yanofsky and his students (144). Considerable significance has attached itself to investigations of the primary structure of this protein in view of the availability of altered A proteins formed by mutants of E. coli. The A protein has several noteworthy features that make it attractive for the study of gene-protein relationships. The enzyme occurs in relatively high concentrations in the organism, is readily isolated in pure form, and its single polypeptide chain (molecular weight 29000) is devoid of

tryptophan. Carlton & Yanofsky (145) have obtained the amino acid sequence of the N-terminus, Met.Glu.Arg.Tyr, by Edman degradation (modification of the Fraenkel-Conrat & Schroeder procedures). This sequence was confirmed by the isolation of Met.Glu.Arg. from a tryptic hydrolysate, and Met.Glu.Arg.Tyr from a chymotryptic hydrolysate of the A protein in unpublished experiments performed with Dowex 1-X2 columns. Carlton & Yanofsky (146) have defined the carboxyl-terminal sequence as Ala.Ala.Thr.Arg.Ser. The protein is attacked only relatively slowly by carboxypeptidase A, and quantitative liberation of the terminal serine residue could not be achieved unless carboxypeptidase B was simultaneously present. Two-dimensional separations by chromatography and electrophoresis on paper of tryptic hydrolysates of the parent and carboxypeptidase-treated protein revealed that two peptides, TP 20 and TP 25, were absent from the hydrolysate of the treated protein. The amino acid sequences of these peptides were shown by Edman degradation to be Ala.Ala.Thr.Arg, and Ala.Ala.Thr.Arg.Ser, respectively. The carboxyl-terminal serine residue in the A protein was also detected by hydrazinolysis. Further progress in

the sequence analysis will be greatly facilitated by the availability of well standardized peptide "maps" for tryptic, chymotryptic, and combined tryptic-chymotryptic hydrolysates of the protein (147).

Fibrin.--An advance of considerable interest in the study of fibrinogen and fibrin has been attained through the research of Clegg & Bailey (148). The fibrinogen molecule of molecular weight 330,000 consists of a dimer, each molecule of which contains three peptide chains. Clegg & Bailey subjected fibrin prepared from fibrinogen by the action of a snake-venom preparation ("reptilase") to oxidative sulfitolysis. The three peptide chains were cleanly separated by chromatography in solvents that contained 8 M urea on columns of CM-cellulose at pH 5.3. The three chains were identified by end-group analysis (dinitrophenylation), the results of which corresponded to those to be expected from the terminal amino acid composition of the starting material.

FOOTNOTES

1. The literature survey in connection with this review was concluded
in November, 1963.
2. The author wishes to thank Mrs. A. Campo for library work.
3. Research carried out at Brookhaven National Laboratory under the
auspices of the U. S. Atomic Energy Commission.
4. Personal communication.

LITERATURE CITED

1. Schroeder, W. L., Ann. Rev. Biochem., 32, 301 (1963).
2. Spackman, D. H., Stein, W. H., and Moore, S., Anal. Chem., 30, 1190 (1958).
3. Kominz, D. R., J. Chromatog., 9, 253 (1962).
4. Piez, K. A., and Morris, L., Anal. Biochem., 1, 187 (1960).
5. Walsh, K. A., and Brown, J. R., Biochim. Biophys. Acta, 58, 596 (1962).
6. Guidotti, G., Hill, R. J., and Konigsberg, W., J. Biol. Chem., 237, 2184 (1962).
7. Konigsberg, W., and Hill, R. J., J. Biol. Chem., 237, 2547 (1962).
8. Hill, R. J., and Konigsberg, W., J. Biol. Chem., 237, 3151 (1962).
9. Konigsberg, W., and Hill, R. J., J. Biol. Chem., 237, 3157 (1962).
10. Goldstein, J., Konigsberg, W., and Hill, R. J., J. Biol. Chem., 238, 2016 (1963).
11. Konigsberg, W., Goldstein, J., and Hill, R. J., J. Biol. Chem., 238, 2028 (1963).

12. Smyth, D. G., Stein, W. H., and Moore, J., J. Biol. Chem., 237, 1845 (1962).
13. Smyth, D. G., Stein, W. H., and Moore, J., J. Biol. Chem., 238, 227 (1963).
14. Spackman, D. H., Federation Proc., 22, 244 (1963).
15. Hamilton, P. B., Ann. N. Y. Acad. Sci., 102, 55 (1962).
16. Hamilton, P. B., Anal. Chem., 35, 2055 (1963).
17. Kirsten, E., and Kirsten, R., Biochem. Biophys. Res. Commun., 7, 76 (1962).
18. Moore, J., J. Biol. Chem., 238, 235 (1963).
19. Fletcher, J. C., and Robson, A., Biochem. J., 84, 439 (1962).
20. Tower, D. B., Peters, E. L., and Wherrett, J. R., J. Biol. Chem., 237, 1861 (1962).
21. Hill, R. L., and Schmidt, W. R., J. Biol. Chem., 237, 289 (1962).
22. Kesner, L., Muntwyler, E., Griffin, G. E., and Abrams, J., Anal. Chem., 35, 83 (1963).
23. Zomzely, C., Marco, G., and Emery, E., Anal. Chem., 34, 1414 (1962).

24. Johnson, D. E., Scott, S. J., and Meister, A., Anal. Chem., 53, 669 (1961).
25. Johnson, D. E., Scott, S. J., and Meister, A., Abstr. Am. Chem. Soc., p. 48 C (Chicago, Ill., 1961).
26. Pisano, J. J., Vanden Heuvel, W. J. A., and Horning, E. C., Biochem. Biophys. Res. Commun., 7, 82 (1962).
27. Stark, G. R., and Smyth, D. G., J. Biol. Chem., 238, 214 (1963).
28. Porter, R. R., in Methods in Enzymology, 4, 221 (Academic Press, New York, 1962).
29. Russell, D. W., Biochem. J., 87, 1 (1963).
30. White, F. H., J. Biol. Chem., 235, 383 (1960).
31. Crestfield, A. M., Moore, S., and Stein, W. H., J. Biol. Chem., 238, 622 (1963).
32. Swan, J. M., Nature, 180, 643 (1957).
33. Pechère, J. F., Dixon, G. H., Maybury, R. H., and Neurath, H., J. Biol. Chem., 233, 1364 (1958).
34. Weil, L., and Telka, M., Arch. Biochem. Biophys., 71, 204 (1957).

35. Leach, S. J., and Swan, J. M., Australian J. Chem., 15, 365 (1962).
36. Hartley, B. S., Brookhaven Symp. Biol., 15, 85 (1962).
37. Thompson, E. O. P., and O'Donnell, I. J., Biochim. Biophys. Acta,
53, 447 (1961).
38. Stark, G. R., Stein, W. H., and Moore, S., J. Biol. Chem., 235,
3177 (1960).
39. Anfinsen, C. B., and Haber, E., J. Biol. Chem., 236, 1361 (1961).
40. Weil, L., and Seibles, T. S., Arch. Biochem. Biophys., 95, 470
(1961).
41. Plummer, T. H., and Hirs, C. H. W., Abstr. Am. Chem. Soc., p. 35 A
(Denver, Colorado, 1964).
42. Inagami, T., and Sturtevant, J. M., J. Biol. Chem., 235, 1019 (1960).
43. Maroux, S., Rivery, M., and Desnuelle, P., Biochim. Biophys. Acta,
56, 202 (1962).
44. Hofmann, T., Walsh, D. A., Kauffman, D. L., and Neurath, H., Federa-
tion Proc., 22, 528 (1963).
45. Lindley, H., Nature, 178, 647 (1956).

46. Raftery, M. A., Lovenberg, W., and Cole, R. D., Abstr. Am. Chem. Soc., p. 34 A (Denver, Colorado, 1964).
47. Witkop, B., Advan. Protein Chem., 16, 221 (1961).
48. Goldberger, R. F., and Anfinsen, C. B., Biochemistry, 1, 401 (1962).
49. Merigan, T. C., Dreyer, W. J., and Berger, A., Biochim. Biophys. Acta, 62, 122 (1962).
50. Canfield, R. E., and Anfinsen, C. B., J. Biol. Chem., 238, 2684 (1963).
51. Canfield, R. E., J. Biol. Chem., 238, 2691 (1963).
52. Canfield, R. E., J. Biol. Chem., 238, 2698 (1963).
53. Hunter, M. J., and Ludwig, M. L., J. Am. Chem. Soc., 84, 3491 (1962).
54. Ludwig, M. L., and Byrne, R., J. Am. Chem. Soc., 84, 4160 (1962).
55. Gross, E., and Witkop, B., J. Am. Chem. Soc., 83, 1510 (1961).
56. Lawson, W. B., Gross, E., Foltz, C. M., and Witkop, B., J. Am. Chem. Soc., 84, 1715 (1962).
57. Koshland, D. E., Strumeyer, D. H., and Ray, W. J., Brookhaven Symp. Biol., 15, 101 (1962).
58. Edmundson, A. B., Nature, 198, 354 (1963).

59. Partridge, S. M., and Davis, H. F., Nature, 165, 62 (1950).
60. Schultz, J., Allison, H., and Grice, M., Biochemistry, 1, 694 (1962).
61. Schroeder, W. A., Shelton, J. R., Shelton, J. B., Cormick, J., and Jones, R. T., Biochemistry, 2, 992 (1963).
62. Tsung, C. M., Funatsur, G., and Young, J. D., Arch. Biochim. Biophys.,
in press.
63. Dixon, H. B. F., and Weitkamp, L. R., Biochem. J., 84, 462 (1962).
64. Wilchek, M., and Patchornik, A., J. Am. Chem. Soc., 84, 4613 (1962).
65. Sokolovsky, M., Wilchek, M., and Patchornik, A., Bull. Res. Council Israel, Proc. Israel Chem. Soc., 11A, 79 (1962).
66. Patchornik, A., and Sokolovsky, M., Bull. Res. Council Israel, Proc. Israel Chem. Soc., 11A, 80 (1962).
67. Wood, J. L., and Catsimpoilas, N., J. Biol. Chem., 238, PC 2887
(1963).
68. Fasold, H., Linhart, P., and Turba, F., Biochem. Z., 336, 182 (1962).
69. Fasold, H., Linhart, P., and Turba, F., Biochem. Z., 336, 191 (1962).
70. Anderer, F., and Handschuh, D., Z. Naturforsch., 17b, 536 (1962).

71. Richards, F. M., Ann. Rev. Biochem., 32, 269 (1963).
72. Fraenkel-Conrat, H., J. Am. Chem. Soc., 76, 3606 (1954).
73. Fraenkel-Conrat, H., Harris, J. I., and Levy, A. L., Methods Biochem. Anal., 2, 393 (1955).
74. Sjöquist, J., Acta Chem. Scand., 7, 447 (1953).
75. Sjöquist, J., Biochim. Biophys. Acta, 41, 20 (1960).
76. Allende, J. E., and Richards, F. M., Biochemistry, 1, 295 (1962).
77. Taschner, E., and Kupryszewski, G., Roczniki Chem., 36, 63 (1962):
Chem. Abstr., 57, 15225 (1962).
78. Hirs, C. H. W., Moore, S., and Stein, W. H., J. Biol. Chem., 235,
633 (1960).
79. Edman, P., Acta Chem. Scand., 10, 761 (1956).
80. Potts, J. T., Berger, A., Cooke, J., and Anfinsen, C. B., J. Biol. Chem., 237, 1851 (1962).
81. Margoliash, E., J. Biol. Chem., 237, 2161 (1962).
82. Ando, T., Iwai, K., Ishii, S.-I., Azegami, M., and Nakahara, C.,
Biochim. Biophys. Acta, 56, 628 (1962).

83. Kotaki, A., J. Biochem. (Japan), 51, 301 (1962).
84. Kotaki, A., J. Biochem. (Japan), 53, 61 (1963).
85. Yamamoto, M., Kotaki, A., Okuyama, T., and Satake, K., J. Biochem. (Japan), 48, 84 (1960).
86. Dixon, G. H., and Wardlaw, A. C., Nature, 188, 721 (1960).
87. Wilson, S., Dixon, G. H., and Wardlaw, A. C., Biochim. Biophys. Acta, 62, 483 (1962).
88. Margoliash, E., Kimmel, J. R., Hill, R. L., and Schmidt, W. R., J. Biol. Chem., 237, 2148 (1962).
89. Margoliash, E., and Smith, E. L., J. Biol. Chem., 237, 2151 (1962).
90. Tuppy, H., and Kreil, G., Monatsh. Chem., 93, 780 (1962).
91. Tuppy, H., and Bodo, G., Monatsh. Chem., 85, 807 (1954).
92. Tuppy, H., and Bodo, G., Monatsh. Chem., 85, 1024 (1954).
93. Tuppy, H., and Bodo, G., Monatsh. Chem., 85, 1182 (1954).
94. Margoliash, E., Frohwirt, N., and Weiner, E., Biochem. J., 71, 559 (1959).
95. Edman, P., and Sjöquist, J., Acta Chem. Scand., 10, 1507 (1956).

96. Harris, J. I., and Roos, P., Biochem. J., 71, 434 (1959).
97. Tuppy, H., and Wintersberger, E., Monatsh. Chem., 91, 1001 (1960).
98. Michl, H., and Hogenauer, M., J. Chromatog., 2, 380 (1959).
99. Margoliash, E., Smith, E. L., Kreil, G., and Tuppy, H., Nature, 192,
1125 (1961).
100. Matsubara, H., and Smith, E. L., J. Biol. Chem., 238, 2732 (1963).
101. Chan, S.-K., Needleman, S. B., Stewart, J. W., Walasek, O. F., and
Maroliash, E., Federation Proc., 22, 658 (1963).
102. Jollès-Thaureaux, J., Jollès, P., and Fromageot, C., Biochim. Biophys.
Acta, 27, 298 (1958).
103. Jollès, P., and Jollès, J., Bull. Soc. Chim. Biol., 40, 133 (1958).
104. Jollès, P., Jollès, J., and Jauregui, J., Biochim. Biophys. Acta,
31, 96 (1959).
105. Jollès, J., and Jollès, P., Biochim. Biophys. Acta, 45, 407 (1960).
106. Jollès, J., Jollès, P., Jauregui-Adell, J., Bull. Soc. Chim. Biol.,
42, 1319 (1960).
107. Jollès, J., and Jollès, P., Compt. Rend. Acad. Sci., 253, 2773 (1961).

108. Jollès, J., Jauregui-Adell, J., and Jollès, P., Biochim. Biophys. Acta, 71, 488 (1963).
109. Dreyer, W. J., Brookhaven Symp. Biol., 13, 243 (1960).
110. Hirs, C. H. W., Moore, S., and Stein, W. H., J. Biol. Chem., 219, 623 (1956).
111. Moore, S., Spackman, D. H., and Stein, W. H., Anal. Chem., 30, 1185 (1958).
112. Spackman, D. H., Stein, W. H., and Moore, S., J. Biol. Chem., 235, 648 (1960).
113. Anderer, F. A., Uhlig, H., Weber, E., and Schramm, G., Nature, 186, 922 (1960).
114. Tsugita, A., Gish, D. T., Young, J., Fraenkel-Conrat, H., Knight, C. A., and Stanley, W. M., Proc. Natl. Acad. Sci. (U. S.), 46, 1463 (1960).
115. Anderer, F. A., Z. Naturforsch., 17b, 526 (1962).
116. Anderer, F. A., Z. Naturforsch., 17b, 530 (1962).
117. Narita, K., Biochim. Biophys. Acta, 28, 184 (1958).

118. Anderer, F. A., Weber, E., and Uhlig, H., Z. Naturforsch., 15b, 79 (1960).
119. Wittmann, H. G., and Braunitzer, G., Virology, 9, 726 (1959).
120. Braunitzer, G., Rudloff, V., and Hilschmann, N., Z. Physiol. Chem., 331, 1 (1963).
121. Rudloff, V., and Braunitzer, G., Z. Physiol. Chem., 323, 129 (1961).
122. Schroeder, W. A., Jones, R. T., Cormick, J., and McCalla, K., Anal. Chem., 34, 1570 (1962).
123. Eriksson, S., and Sjoquist, J., Biochim. Biophys. Acta, 45, 290 (1960).
124. Levy, A. L., Nature, 174, 126 (1954).
125. Gish, D. T., Ramachandran, L. K., and Stanley, W. M., Arch. Biochem. Biophys., 78, 433 (1958).
126. Ramachandran, L. K., and Gish, D. T., J. Am. Chem. Soc., 81, 344 (1959).
127. Gish, D. T., J. Am. Chem. Soc., 82, 6329 (1960).
128. Gish, D. T., J. Am. Chem. Soc., 83, 3303 (1961).

129. Light, A., and Smith, E. L., J. Biol. Chem., 235, 3144 (1960).
130. Light, A., and Smith, E. L., J. Biol. Chem., 235, 3151 (1960).
131. Light, A., Glazer, A. N., and Smith, E. L., J. Biol. Chem., 235,
3159 (1960).
132. Kimmel, J. R., Kato, G. K., Paiva, A. C. M., and Smith, E. L.,
J. Biol. Chem., 237, 2525 (1962).
133. Light, A., J. Biol. Chem., 237, 2535 (1962).
134. Light, A., and Smith, E. L., J. Biol. Chem., 237, 2537 (1962).
135. Keil, B., Meloun, B., Vanecek, J., Kostka, V., Prusik, Z., and
Sorm, F., Biochim. Biophys. Acta, 56, 595 (1962).
136. Keil, B., Prusik, Z., Moravek, L., and Sorm, F., Collection Czech.
Chem. Commun., 27, 2946 (1962).
137. Keil, B., Prusik, Z., and Sorm, F., Biochim. Biophys. Acta, 78,
559 (1963).
138. Van Hoang, D., Rovey, M., and Desnuelle, P., Biochim. Biophys.
Acta, 58, 613 (1962).

139. Van Hoang, D., Rovey, M., Guidoni, A., and Desnuelle, P., Biochim. Biophys. Acta, 69, 188 (1963).
140. Walsh, K. A., Kauffman, D. L., and Neurath, H., Biochemistry, 1, 893 (1962).
141. Walsh, K. A., Kauffman, D. L., and Neurath, H., Biochim. Biophys. Acta, 65, 540 (1962).
142. Mikes, O., Holeysovsky, V., Tomasek, V., and Sorm, F., Collection Czech. Chem. Commun., 26, 1048 (1961).
143. Mikes, O., Holeysovsky, V., Tomasek, V., Keil, B., and Sorm, F., Collection Czech. Chem. Commun., 27, 1964 (1962).
144. Henning, V., Helinski, D. R., Chao, F. C., and Yanofsky, C., J. Biol. Chem., 237, 1523 (1962).
145. Carlton, B. C., and Yanofsky, C., J. Biol. Chem., 237, 1531 (1962).
146. Carlton, B. C., and Yanofsky, C., J. Biol. Chem., 238, 636 (1963).
147. Helinski, D. R., and Yanofsky, C., Biochim. Biophys. Acta, 63, 10 (1962).
148. Clegg, J. B., and Bailey, K., Biochim. Biophys. Acta, 63, 525 (1962).

TABLE I

DIFFERENCES IN REPORTED AMINO ACID SEQUENCES IN LYSOZYME

Residues	40 41 42	58 59	92 93
Jollès (108)	Gln.Ala.Thr	Asn.Ileu	Asn.Val
Canfield (52)	Thr.Gln.Ala	Ileu.Asn	Val.Asn

NH₂-Ala-Arg-Arg-Arg-Arg-Ser-Arg-Arg-Ser-Arg-Pro-
-Val-Arg-Arg-Arg-Arg-Pro-Arg-Arg-Val-Ser-Arg-Arg-
-Arg-Arg-Ala-Arg-Arg-Arg-Arg-COOH

FIG. 1. Amino acid sequence of clupeine Z
from the Pacific herring (82).

NH_2 -Gly-Ileu-(His, Glu-Glu, CySO_3H , (CySO_3H , His, Lys, Pro)-

CySO_3H , Leu)-Phe-Glu-Leu-Glu-Asp-Tyr- CySO_3H -Asn-COOH

NH_2 -Ala-Ala-Asn-(Pro, His, Leu)- CySO_3H -[Gly, Ser, His, -

(Leu-Val-Glu-Ala-Leu)]-Tyr-Leu-(Val, CySO_3H , Gly, Glu)-

Arg-Gly-Phe-Phe-Tyr-Glu-Pro-Lys-COOH

FIG. 2. Partial formulae for the oxidized A- and B-chains, respectively, of component II of the insulins from the bonito (83,84). The residues that differ from those in bovine insulin are underlined.

FIG. 3. Heme-peptide fraction from chymotryptic hydrolysate of horse heart cytochrome c (89). The points of tryptic cleavage are indicated by the arrows. The residues are numbered according to their position in the amino acid sequence of the parent protein.

Acetyl-Gly-Asp-Val-Glu-Lys-Gly-Lys-Lys-Ileu-Phe-Val-Gln-Lys-Cys-Ala-Gln-Cys-His-Thr-
 20 25 30 35 40
 -Val-Glu-Lys-Gly-Gly-Lys-His-Lys-Thr-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe-Gly-Arg-Lys-Thr-
 45 50 55 60
 -Gly-Gln-Ala-Pro-Gly-Phe-Thr-Tyr-Thr-Asp-Ala-Asn-Lys-Asn-Lys-Gly-Ileu-Thr-Try-Lys-Glu-
 65 70 75 80
 -Glu-Thr-Leu-Met-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ileu-Pro-Gly-Thr-Lys-Met-Ileu-Phe-
 85 90 95 100
 -Ala-Gly-Ileu-Lys-Lys-Lys-Thr-Glu-Arg-Glu-Asp-Leu-Ileu-Ala-Tyr-Leu-Lys-Lys-Ala-Thr-Asn-
 -Glu-COOH

FIG. 4. The amino acid sequence of horse heart cytochrome c (88, 89, 91, 90, 99).

Residues that differ in human heart protein are indicated by underlining (100).

FIG. 5. The amino acid sequence of hen egg-white lysozyme (50, 51, 52, 106, 108) as deduced from experiments with the reduced, carboxymethylated protein. The locations of the eight carboxymethylcysteine (CMC) residues are marked by the Roman numerals.

5 10 15 20
 Acetyl-Ser-Tyr-Ser-Ileu-Thr-Thr-Pro-Ser-Gln-Phe-Val-Phe-Leu-Ser-Ser-Ala-Try-Ala-Asp-Pro-Ileu-Glu-Leu-
 25 30 35 40 45
 -Ileu-Asn-Leu-CySH-Thr-Asn-Ala-Leu-Gly-Asn-Gln-Phe-Gln-Thr-Gln-Gln-Ala-Arg-Thr-Val-Gln-Val-Arg-Gln-
 50 55 60 65 70
 -Phe-Ser-Gln-Val-Try-Lys-Pro-Ser-Pro-Gln-Val-Thr-Val-Arg-Phe-Pro-Asp-Ser-Asp-Phe-Lys-Val-Tyr-Arg-Tyr-
 75 80 85 90 95
 -Asn-Ala-Val-Leu-Asp-Pro-Leu-Val-Thr-Ala-Leu-Leu-Gly-Ala-Phe-Asp-Thr-Arg-Asn-Arg-Ileu-Ileu-Gln-Val-
 100 105 110 115 120
 -Gln-Asp-Gln-Ala-Asn-Pro-Thr-Thr-Ala-Gln-Thr-Leu-Asp-Ala-Thr-Arg-Arg-Val-Asp-Asp-Ala-Thr-Val-Ala-
 125 130 135 140
 -Ileu-Arg-Ser-Ala-Asp-Ileu-Asn-Leu-Ileu-Val-Glu-Leu-Ileu-Arg-Gly-Thr-Gly-Ser-Tyr-Asn-Arg-Ser-Ser-Phe-
 145 150 155
 -Glu-Ser-Ser-Ser-Gly-Leu-Val-Try-Thr-Ser-Gly-Pro-Ala-Thr-COOH

FIG. 6. The amino acid sequence of the protein of tobacco mosaic virus (70, 114, 115, 116).