

Selective Cleavage of Proteins

Myron J. Waxdal

The elucidation of the primary structure of proteins requires that they be selectively cleaved into smaller, more workable fragments or peptides. The methods used to accomplish these cleavages may be divided into two categories: chemical and enzymatic. In each case, the major emphasis is on the specificity and efficiency of these cleavages. For small proteins or peptides, one or two methods may provide the peptides suitable for manual or automated sequence analysis. However, larger proteins re-

quire the use of several methods to obtain all the peptides necessary for the establishment of a unique sequence. In certain instances both chemical and enzymatic techniques are combined. These mixed techniques may restrict enzymatic cleavage to fewer bonds by blocking normal sites of cleavage or they may provide new sites. The structure of a γ G1 immunoglobulin will be discussed as an example of a protein requiring several approaches for the elucidation of its amino acid sequence.

The purpose of this paper is to provide nonspecialists with some insights into how the amino acid sequences of large proteins are determined. The techniques and procedures which will be discussed here have been selected to serve as examples and are not the only or necessarily the best of their kind. As an example, certain aspects of the elucidation of the complete amino acid sequence of a γ G1 immunoglobulin, Eu (Edelman *et al.*, 1969), will be presented.

To provide a background for discussion let us briefly look at some of the problems posed by protein Eu, which is comprised of about 25,000 atoms. It contains 1320 amino acid residues, 127 potential tryptic peptides, 16 disulfide bonds, and has multiple polypeptide chains. Direct sequence analysis is pitifully inadequate. This protein has more seryl residues in one of its chains than the total number of amino acid residues usually determined by the use of an automated sequenator. The major task in determining the structure of so large a protein is to divide it into workable pieces so that finally one has the peptides which are amenable to modern techniques for determining amino acid sequence. One must then order these structural peptides to reconstruct the complete original sequence.

We will dwell on the means of dividing a protein specifically and efficiently into reasonable numbers of smaller pieces. No attempt will be made to enumerate all of the procedures that have been utilized, but one will be made rather to discuss some of the problems and the general approaches toward solving them.

SPECIFIC CLEAVAGE OF PEPTIDES AND PROTEINS

To simplify the isolation of each of the required pieces, which is probably the greatest task in protein structural work, and then to overlap or determine the order of the pieces, one tries to work with reasonably simple mixtures. Hence, it is desirable to cleave the protein chain at selected

“rare” amino acid residues, which will produce a mixture with a limited number of pieces to be isolated.

The choice of agent to split at these rare residues must be specific for these rare residues so that extraneous contaminating pieces do not arise. The agent must also be efficient in breaking the peptide bond at these residues; moreover, high yields of the pieces are needed for further processing. Inefficient cleavage gives rise to horrendous mixtures which are impossible to separate, as well as being wasteful of the starting protein or peptide.

Once we have obtained fragments of a protein by using one method, they must be arranged in their original order. This is accomplished by cleaving the protein by another method to yield a second set of fragments which overlap the first. Structural studies on this second set will then provide the data for ordering the first set. In practice, one method frequently does not provide all the necessary pieces, and several methods are needed to divide the protein into smaller pieces.

Two types of agents have been used for these procedures: chemical reagents which cleave peptide bonds and the proteolytic enzymes. Each of these methods and an approach which uses both chemical and enzymatic means to produce new sets of pieces will be discussed.

Chemical Methods. This topic has been reviewed comprehensively by Spande *et al.* (1970).

Acid hydrolysis, usually *in vacuo* with 6 N HCl at 105° C for 24 hr (Moore and Stein, 1963), is used to degrade a protein or peptide completely into its constituent amino acids. Unfortunately, acid hydrolysis also destroys tryptophan and converts glutamine and asparagine into glutamic and aspartic acid, respectively. This hydrolysis allows one to carry out an amino acid analysis of the protein or peptide. Complete acid hydrolysis does not provide pieces for amino acid sequence studies, however. Individual peptide bonds are hydrolyzed at different rates, depending upon which amino acid residues form each bond. Hence, under milder conditions hydrolysis predominantly occurs at the most labile of these peptide bonds. This approach, partial acid hy-

The Rockefeller University, New York, N.Y. 10021

drolisis, was used extensively in the earlier work on protein structure (Kay and Schroeder, 1954; Sanger *et al.*, 1955; Schroeder *et al.*, 1954). Because of its lack of specificity and low yields, partial acid hydrolysis is now generally limited to the structure determination of small peptides.

Another general mechanism, intramolecular nucleophilic attack of the peptide carbonyl function on an activated side chain, has proved fruitful. One approach uses *N*-bromosuccinimide (NBS) to effect the cleavage of the peptide bond on the carboxyl side of the aromatic amino acids tryptophan, tyrosine, and histidine (Ramachandran and Witkop, 1967). The probable mechanism of this cleavage in the case of tyrosine is shown in Figure 1. Unfortunately, this method suffers badly from partial cleavages resulting in low yields, and NBS also attacks the sulfur-containing amino acids. In general, it is not possible to restrict cleavages to a single type of residue. Although this method is unsuitable for general application, like partial acid hydrolysis, it may be valuable in special cases or in the determination of the structure of small peptides.

The most specific and efficient chemical cleavage of peptide bonds uses cyanogen bromide (CNBr) to cleave the bond on the carboxyl side of methionyl residues (Gross and Witkop, 1961). The mechanism of CNBr cleavage is similar to that of *N*-bromosuccinimide and is shown in Figure 2. Again, the electrons from the carbonyl group attack an activated side chain to form an intermediate, which is then hydrolyzed to break the original peptide into two new peptides, one ending in homoserine lactone and the other beginning with a new amino terminal residue. CNBr cleavage is specific only for methionyl residues and the conversion to homoserine lactone is nearly 100% complete. However, this method also has certain problems. If the bond to be cleaved is formed by a Met-Ser or Met-Thr sequence, the yield of the two derived peptides may drop to near zero, even though all the methionine is converted and homoserine is found after acid hydrolysis and amino acid analysis (Cunningham *et al.*, 1968; Schroeder *et al.*, 1969; Waxdal *et al.*, 1968a,b). The actual yield of derived peptides appears to involve the choice of solvent as well as the choice of protein. The use of 70% formic acid as a solvent gives rapid cleavage and good yields except with Met-Thr and Met-Ser bonds. In 70% trifluoroacetic acid these bonds are cleaved to a larger extent, but the cleavage of all bonds proceeds at a lower rate (Schroeder *et al.*, 1969). Nonetheless, cyanogen bromide cleavage is the chemical method which most closely meets the criterion of specificity and efficiency in the breakdown of proteins and peptides.

Enzymatic Methods. The alternative to chemical cleavage is the use of proteolytic enzymes to digest the proteins or peptides into smaller pieces. A range of proteolytic enzymes has been isolated; some demonstrate specificity for a single type of peptide bond, and others show little differentiation among peptide bonds. In order to discuss some of these

Table I. Proteolytic Enzymes for Cleavage of Proteins

Very wide specificity	Wide specificity	Narrow specificity
Pronase	Chymotrypsin	Trypsin
Subtilisin	Pepsin	Clostripain
	Papain	Thrombin
	Thermolysin	
	Streptococcal proteinase	

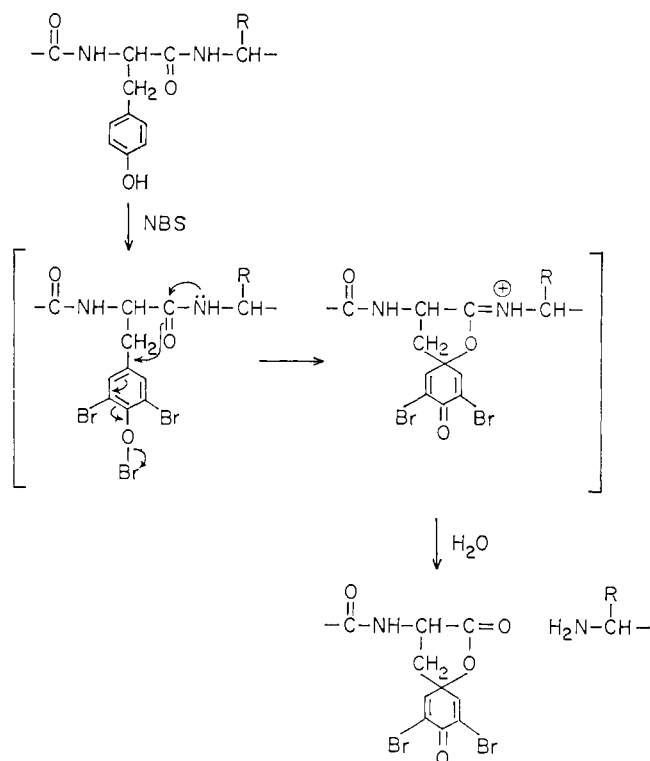


Figure 1. The cleavage of a tyrosyl bond by *N*-bromosuccinimide (NBS)

enzymes briefly, I have divided them into three groups (Table I) based upon the range of peptide bonds which they hydrolyze. Most of the proteolytic enzymes used are active in the neutral pH ranges and therefore the protein or peptide substrate must be soluble at neutral pH values. All too often the substrates are not soluble in this range or become insoluble after a few peptides have been split off. These problems can severely restrict the applicability of many of the usual enzymatic digestions.

Although no single enzyme has been described which is as effective as complete acid hydrolysis for splitting all peptide bonds, there are enzyme mixtures which can very nearly reduce a protein or peptide completely into its constituent amino acids. The most common of these is a mixture of enzymes from *Streptomyces griseus* known as pronase. Another nonspecific proteolytic enzyme is subtilisin. These

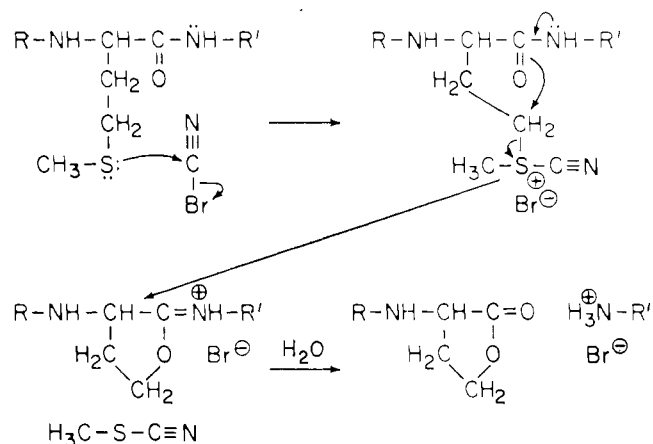


Figure 2. The cleavage of a methionyl bond by cyanogen bromide (CNBr)

enzymes, along with other enzymes which degrade peptides from the amino or carboxyl terminal end, may be used to digest a peptide nearly completely to free amino acids (Hill and Schmidt, 1962). As we noted above, complete acid hydrolysis destroys tryptophan and converts asparagine and glutamine to the corresponding free acids. In contrast, complete enzymatic digests allow one to determine the number of these amino acids in a peptide or a protein.

The second group of enzymes, showing a more restricted specificity toward the peptide bonds which they hydrolyze, is most useful in degrading a protein or large peptide into smaller peptides. Probably the best known of this group is chymotrypsin, which cleaves peptide bonds on the carboxy terminal side of tryptophan and tyrosine, and at a lower rate cleaves phenylalanine, leucine, isoleucine, and methionine bonds. Chymotrypsin may hydrolyze the peptide bonds of asparagine, glutamine, and sometimes other residues at even lower rates (Goldberg *et al.*, 1951; Green and Neurath, 1954; Hirs *et al.*, 1960; Kaufman and Neurath, 1949).

Among the other enzymes in this group are pepsin, papain, thermolysin, and streptococcal proteinase. Pepsin is an enzyme which is active from pH 2 to pH 4 and thus may be used when the substrate is insoluble at neutral pH values (Green and Neurath, 1954).

The remaining group is composed of those proteolytic enzymes which show the most specificity in the bonds which they hydrolyze. The best known of these is trypsin. Because trypsin only splits peptide bonds on the carboxyl side of arginyl and lysyl residues (Bergman, 1942), it has frequently been the first enzyme used in the structure determination of a new protein. Another enzyme, clostripain, apparently is specific only for arginyl bonds, but has not been commonly used in amino acid sequence studies (Mitchell, 1968; Mitchell and Harrington, 1968). Another enzyme, thrombin, shows specificity toward both arginine and lysine in model compounds, but apparently splits only at arginyl bonds when used to degrade proteins (Lundblad, 1970). The latter two enzymes may prove to be a great value in future structural studies on proteins.

In its native state, each protein is folded into a specific three-dimensional structure. In general this structure is more resistant to enzymatic attack than the same protein in an unfolded state. However, this specific structure may have certain peptide bonds which are highly susceptible to specific enzymes, or regions which are highly susceptible to general enzymatic attack. There are many examples of these highly susceptible bonds in the literature. The conversion of a zymogen such as trypsinogen to the active enzyme trypsin by splitting of a single specific peptide bond (Davie and Neurath, 1955) is a well known example.

Aside from the activation of zymogens, there are several other well documented cases of highly susceptible bonds. Subtilisin converts ribonuclease to the S-protein and S-peptide (Richards and Vithayathil, 1959). Gamma globulin is split by papain (Fleischman *et al.*, 1963) or trypsin (Edelman *et al.*, 1968) into the Fab and Fc pieces.

We have seen how either chemical or enzymatic means may be utilized to divide a protein or large peptide into smaller fragments which are amenable to further structure determination. In both approaches we also have the means available to efficiently degrade the protein into very small nonspecific peptides or free amino acids. On the other hand, we have both chemical and enzymatic means to cleave the protein at a few highly selective peptide bonds. Between these two extremes there are procedures which are not as

specific or as efficient for all proteins, but which are of great use in specific cases.

Combined Methods. If we combine these chemical and enzymatic degradation procedures, most proteins can be broken into the pieces necessary to determine a complete amino acid sequence. However, if we chemically modify the points of enzymatic cleavage, we can create a new pattern of splitting and produce new, more useful, fragments of the protein. This type of mixed technique is increasingly referred to as "peptide surgery." Some of the advantages of peptide surgery are: to provide overlapping pieces from other digests; to solubilize material which is otherwise insoluble at neutral pH values where most of the enzymes are active; and to cleave a protein selectively into a limited number of fragments which may be less troublesome to fractionate and order than those from a more extensive enzymatic digestion.

Currently, peptide surgery utilizes the specificity of the enzyme trypsin. The first type of chemical modifications for peptide surgery blocks the ϵ -amino group of lysyl residues and hence restricts tryptic digestion to the arginyl residues. There are a number of reagents used for this purpose, each of which may also block the α -amino group of the protein (Figure 3).

The ϵ -amino groups of the lysyl residues may be specifically reacted with *O*-methylisourea and converted to homoarginine (Hunter and Ludwig, 1962). Although the blocked lysyl residue continues to maintain its positive charge, it is no longer susceptible to tryptic digestion. Amino groups may be acetylated (Weil and Telka, 1957) or blocked with a trifluoroacetyl group (Goldberger and Anfinsen, 1962). In both cases the positive charge is lost, and the trifluoroacetyl group may be removed later and the fragment redigested with trypsin to cleave at these lysyl residues (Goldberger and Anfinsen, 1962). Fluorodinitrobenzene also may be used to block the amino groups with an aromatic group and remove the positive charge (Redfield and Anfinsen, 1956). This derivative is frequently rather insoluble.

The positive charge on the lysyl residues may be replaced with a negative charge, usually accompanied by a marked change in solubility properties. If a stable derivative is wanted, succinic anhydride (Li and Bertsch, 1960) is a good reagent. If one wishes to remove the blocking groups at a

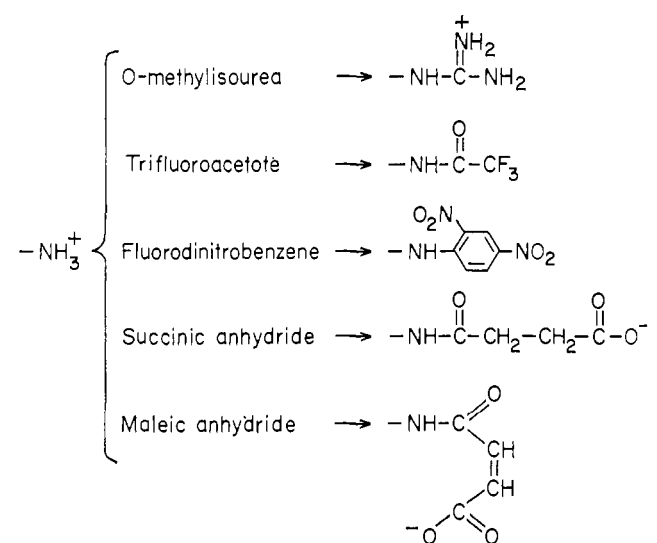


Figure 3. Reagents for blocking the ϵ -amino group of a lysyl residue. See text for discussion and references

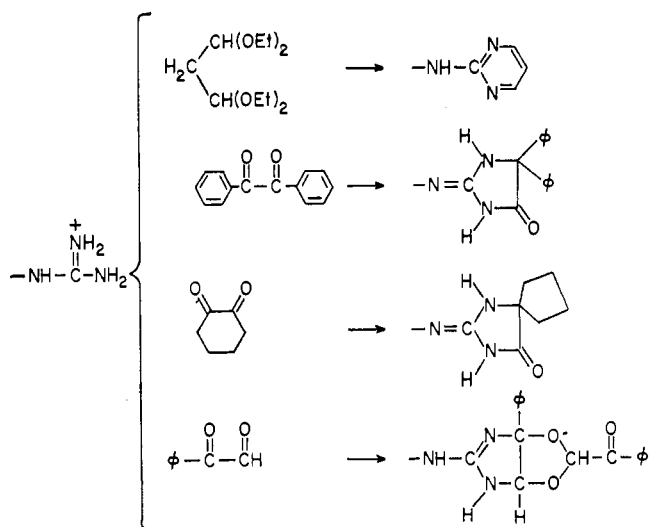


Figure 4. Reagents for blocking the guanido group of arginyl residues, 1,1,3,3-tetraethoxypropane (King, 1966); benzil (Itano and Gottlieb, 1963); 1,2-cyclohexandione (Toi *et al.*, 1965, 1967); phenylglyoxal (Takahashi, 1968)

later stage and digest at the lysyl residues, maleic anhydride (Butler *et al.*, 1967) or one of its derivatives may be used (Dixon and Perham, 1968). If 2-methylmaleic anhydride is used, deblocking occurs overnight and the extremely labile 2,3-dimethyl maleoyl groups are removed in hours (Dixon and Perham, 1968).

In each case we have blocked the lysyl residues to restrict tryptic cleavage of the protein or peptide to the arginyl residues, thus limiting the number of pieces produced.

Conversely, the arginyl residues may be blocked to restrict tryptic cleavage to the lysyl residues (Figure 4). The lower three of these reagents each utilizes a dione function to react with the guanido group of the arginyl residues. Phenylglyoxal, which reacts under the most moderate conditions, yields the only derivative which may be treated to yield back the free guanido group. The blocking group is removed by raising the pH to 8 or above (Takahashi, 1968).

Thus far we have blocked either lysyl or arginyl residues to restrict the positions of tryptic cleavage of the polypeptide chain. The opposite approach, to cause new positions of tryptic cleavage, has also been utilized. Ethyleneimine is used to convert cysteine to *S*-(β -aminoethyl)cysteine (Figure 5; Cole, 1967). This derivative is similar in structure to lysine, with the sulfur replacing a methylene group. Trypsin will now split the amide bond on the carboxyl side of this new residue (Cole, 1967).

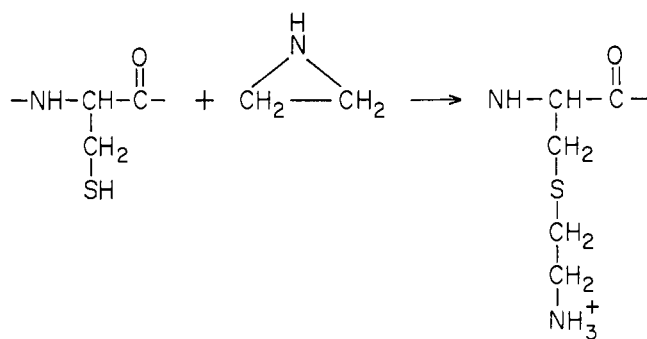


Figure 5. Conversion of cysteine to *S*-(β -aminoethyl)cysteine. See text for discussion and references

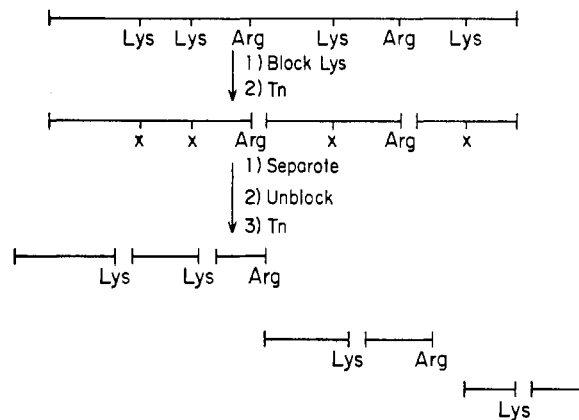


Figure 6. Example of peptide surgery on a hypothetical peptide

Peptide surgery is increasing in importance as the structures of larger and more difficult proteins are attacked. It can provide an efficient means of solubilizing large insoluble or "core" regions of proteins and specifically cleaving them into smaller workable fragments. This procedure allows one to break polypeptide chains only at arginyl residues (Figure 6), separate the pieces, unblock the lysyl residues, and redigest with trypsin to produce smaller specific peptides. A great advantage here is that each cleavage step produces a small number of pieces. The isolation and ordering of these pieces is much less difficult than if one had to separate and order a mixture of all the final peptides.

As more reagents are developed, peptide surgery will prove even more useful in the determination of protein structure.

THE AMINO ACID SEQUENCE OF EU

Now that we have enumerated some of the tools for breaking a protein into peptides suitable for direct sequence analysis and for ordering these peptides, let us briefly examine an example of how these methods were used in the determination of the complete covalent structure of immunoglobulin Eu. This work was performed by Gerald M. Edelman and his associates at The Rockefeller University (Edelman *et al.*, 1968, 1969; Waxdal *et al.*, 1968a,b; Gall *et al.*, 1968; Cunningham *et al.*, 1968, 1970; Gottlieb *et al.*, 1970; Rutishauser *et al.*, 1970; Bennett *et al.*, 1970; Gall and Edelman, 1970; Edelman, 1970).

First of all, highly susceptible bonds, both disulfide and peptide bonds, were specifically broken. The interchain disulfide bonds were reduced and alkylated to yield the four chains, two identical heavy chains and two identical light chains,

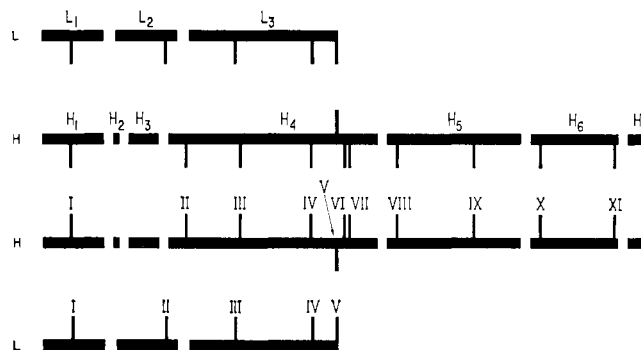


Figure 7. The ordered cyanogen bromide fragments of Eu. In the upper half of the model the fragments are numbered. In the lower half the positions of the half-cystinyl residues are indicated

1 10 20
 ASP - ILE - GLN - MET - THR - GLN - SER - PRO - SER - THR - LEU - SER - ALA - SER - VAL - GLY - ASP - ARG - VAL - THR -
 30 40
 ILE - THR - CYS - ARG - ALA - SER - GLN - SER - ILE - ASN - THR - TRP - LEU - ALA - TRP - TYR - GLN - GLN - LYS - PRO -
 50 60
 GLY - LYS - ALA - PRO - LYS - LEU - LEU - MET - TYR - LYS - ALA - SER - SER - LEU - GLU - SER - GLY - VAL - PRO - SER -
 70 80
 ARG - PHE - ILE - GLY - SER - GLY - SER - GLY - THR - GLU - PHE - THR - LEU - THR - ILE - SER - SER - LEU - GLN - PRO -
 90 100
 ASP - ASP - PHE - ALA - THR - TYR - TYR - CYS - GLN - GLN - TYR - ASN - SER - ASP - SER - LYS - MET - PHE - GLY - GLN -
 110 120
 GLY - THR - LYS - VAL - GLU - VAL - LYS - GLY - THR - VAL - ALA - ALA - PRO - SER - VAL - PHE - ILE - PHE - PRO - PRO -
 130 140
 SER - ASP - GLU - GLN - LEU - LYS - SER - GLY - THR - ALA - SER - VAL - VAL - CYS - LEU - LEU - ASN - ASN - PHE - TYR -
 150 160
 PRO - ARG - GLU - ALA - LYS - VAL - GLN - TRP - LYS - VAL - ASP - ASN - ALA - LEU - GLN - SER - GLY - ASN - SER - GLN -
 170 180
 GLU - SER - VAL - THR - GLU - GLN - ASP - SER - LYS - ASP - SER - THR - TYR - SER - LEU - SER - SER - THR - LEU - THR -
 190 200
 LEU - SER - LYS - ALA - ASP - TYR - GLU - LYS - HIS - LYS - VAL - TYR - ALA - CYS - GLU - VAL - THR - HIS - GLN - GLY -
 210 214
 LEU - SER - SER - PRO - VAL - THR - LYS - SER - PHE - ASN - ARG - GLY - GLU - CYS

Figure 8. Complete amino acid sequence of the light chain. Methionyl residues are underlined and half-cystinyl residues are in boxes

1 10 20
 PCA - VAL - GLN - LEU - VAL - GLN - SER - GLY - ALA - GLU - VAL - LYS - LYS - PRO - GLY - SER - SER - VAL - LYS - VAL -
 30 40
 SER - CYS - LYS - ALA - SER - GLY - GLY - THR - PHE - SER - ARG - SER - ALA - ILE - ILE - TRP - VAL - ARG - GLN - ALA -
 50 60
 PRO - GLY - GLN - GLY - LEU - GLU - TRP - MET - GLY - GLY - ILE - VAL - PRO - MET - PHE - GLY - PRO - PRO - ASN - TYR -
 70 80
 ALA - GLN - LYS - PHE - GLN - GLY - ARG - VAL - THR - ILE - THR - ALA - ASP - GLU - SER - THR - ASN - THR - ALA - TYR -
 90 100
MET - GLU - LEU - SER - SER - LEU - ARG - SER - GLU - ASP - THR - ALA - PHE - TYR - PHE - CYS - ALA - GLY - GLY - TYR -
 110 120
 GLY - ILE - TYR - SER - PRO - GLU - GLU - TYR - ASN - GLY - GLY - LEU - VAL - THR - VAL - SER - SER - ALA - SER - THR -
 130 140
 LYS - GLY - PRO - SER - VAL - PHE - PRO - LEU - ALA - PRO - SER - SER - LYS - SER - THR - SER - GLY - GLY - THR - ALA -
 150 160
 ALA - LEU - GLY - CYS - LEU - VAL - LYS - ASP - TYR - PHE - PRO - GLU - PRO - VAL - THR - VAL - SER - TRP - ASN - SER -
 170 180
 GLY - ALA - LEU - THR - SER - GLY - VAL - HIS - THR - PHE - PRO - ALA - VAL - LEU - GLN - SER - SER - GLY - LEU - TYR -
 190 200
 SER - LEU - SER - SER - VAL - VAL - THR - VAL - PRO - SER - SER - SER - LEU - GLY - THR - GLN - THR - TYR - ILE - CYS -
 210 220
 ASN - VAL - ASN - HIS - LYS - PRO - SER - ASN - THR - LYS - VAL - ASP - LYS - ARG - VAL - GLU - PRO - LYS - SER - CYS -
 230 240
 ASP - LYS - THR - HIS - THR - CYS - PRO - PRO - CYS - PRO - ALA - PRO - GLU - LEU - LEU - GLY - GLY - PRO - SER - VAL -
 250 260
 PHE - LEU - PHE - PRO - PRO - LYS - PRO - LYS - ASP - THR - LEU - MET - ILE - SER - ARG - THR - PRO - GLU - VAL - THR -
 270 280
CYS - VAL - VAL - VAL - ASP - VAL - SER - HIS - GLU - ASP - PRO - GLN - VAL - LYS - PHE - ASN - TRP - TYR - VAL - ASP -
 290 300
 GLY - VAL - GLN - VAL - HIS - ASN - ALA - LYS - THR - LYS - PRO - ARG - GLU - GLN - GLN - TYR - ASX - SER - THR - TYR -
 310 320
 ARG - VAL - VAL - SER - VAL - LEU - THR - VAL - LEU - HIS - GLN - ASN - TRP - LEU - ASP - GLY - LYS - GLU - TYR - LYS -
 330 340
CYS - LYS - VAL - SER - ASN - LYS - ALA - LEU - PRO - ALA - PRO - ILE - GLU - LYS - THR - ILE - SER - LYS - ALA - LYS -
 350 360
 GLY - GLN - PRO - ARG - GLU - PRO - GLN - VAL - TYR - THR - LEU - PRO - PRO - SER - ARG - GLU - GLU - MET - THR - LYS -
 370 380
 ASN - GLN - VAL - SER - LEU - THR - CYS - LEU - VAL - LYS - GLY - PHE - TYR - PRO - SER - ASP - ILE - ALA - VAL - GLU -
 390 400
 TRP - GLU - SER - ASN - ASP - GLY - GLU - PRO - GLU - ASN - TYR - LYS - THR - THR - PRO - PRO - VAL - LEU - ASP - SER -
 410 420
 ASP - GLY - SER - PHE - PHE - LEU - TYR - SER - LYS - LEU - THR - VAL - ASP - LYS - SER - ARG - TRP - GLN - GLU - GLY -
 430 440
 ASN - VAL - PHE - SER - CYS - SER - VAL - MET - HIS - GLU - ALA - LEU - HIS - ASN - HIS - TYR - THR - GLN - LYS - SER -
 446
 LEU - SER - LEU - SER - PRO - GLY

Figure 9. Complete amino acid sequence of the heavy chain. Methionyl residues are underlined and half-cystinyl residues are in boxes

chains (Edelman *et al.*, 1968). A limited digest of the whole molecule with trypsin broke the heavy chains at a specific lysyl residue and the molecule fell into three pieces (Edelman *et al.*, 1968), similar to the Fab and Fc pieces obtained by digestion of γ -globulin with papain (Fleischman *et al.*, 1963). The two Fab pieces each contained an intact light chain and the amino terminal half of the heavy chain. The Fc piece was a dimer of the carboxyl terminal half of the heavy chain (Edelman *et al.*, 1968).

Each of these chains and enzymatic pieces and the intact molecule were cleaved into cyanogen bromide fragments. These fragments were isolated, characterized, and mapped into their original positions with the help of methionine containing overlap peptides obtained from tryptic digests of the light and heavy chains (Waxdal *et al.*, 1968a,b). This map of cyanogen bromide fragments is presented in Figure 7. There are 20 CNBr fragments in a molecule of Eu. But because there are two identical heavy chains and two identical light chains, the entire molecule is accounted for in ten unique fragments, L₁, L₂, and L₃ from the light chain and H₁ to H₇ from the heavy chain (Waxdal *et al.*, 1968a,b).

The isolation of these 10 cyanogen bromide fragments was a key step in the determination of the complete amino acid sequence of Eu. Instead of being faced with the Herculean task of conventionally determining the sequence of this molecule, the problem was converted to that of determining the sequence of ten smaller molecules. Hence, each fragment, which ranged in size from 6 to 165 amino acid residues, was treated as a sequence study in itself.

Most of the sequence data on the individual fragments came from tryptic, chymotryptic, and peptic peptides. Certain fragments also required treatment by pronase (H₁) or partial acid hydrolysis (H₂) to provide further peptides (Cunningham *et al.*, 1970). Peptide surgery including digestion with trypsin after blocking lysine residues or after the conversion of cysteine to aminoethyl cysteine was required on the three largest CNBr fragments [H₄ (Cunningham *et al.*, 1970), H₅, and H₆ (Rutishauser *et al.*, 1970)].

At each stage in this structure determination the most efficient and specific methods were brought to bear on the particular problems. The summation of the data from this eclectic approach established the complete sequence of the 660 unique amino acid residues in each half molecule of Eu.

The amino acid sequence of the 214 residues in each light chain is presented in Figure 8 and the 446 residues in each heavy chain is presented in Figure 9.

In the light chain Met-4 is followed in the sequence by a threonyl residue. This Met-Thr bond was cleaved by CNBr to less than 5% (Cunningham *et al.*, 1968). In the heavy chain Met-358 is also followed by a threonyl residue. The amount of CNBr cleavage of this bond, giving rise to H₅ and H₆, varied considerably from one experiment to another. In certain cases the yield of H₅ and H₆ was higher than 60% (Waxdal *et al.*, 1968a).

In the sequence studies of large proteins, the specific and efficient production of workable pieces and their purification is the major problem. In the majority of cases, no single cleavage method, chemical or enzymatic, is adequate to produce all of these required pieces. As in the case of Eu,

an eclectic approach utilizing several complementary methods can lead to the successful sequence analysis of large proteins.

LITERATURE CITED

- Bennett, C., Konigsberg, W. H., Edelman, G. M., *Biochemistry* **9**, 3181 (1970).
 Bergman, M., *Advan. Enzymol.* **2**, 49 (1942).
 Butler, P. J. G., Harris, J. I., Hartley, B. S., Leberman, R., *Biochem. J.* **103**, 78 (1967).
 Cole, R. D., *Methods Enzymol.* **11**, 315 (1967).
 Cunningham, B. A., Gottlieb, P. D., Konigsberg, W. H., Edelman, G. M., *Biochemistry* **7**, 1983 (1968).
 Cunningham, B. A., Rutishauser, U., Gall, W. E., Gottlieb, P. D., Waxdal, M. J., Edelman, G. M., *Biochemistry* **9**, 3161 (1970).
 Davie, E. W., Neurath, H. N., *J. Biol. Chem.* **212**, 515 (1955).
 Dixon, H. B. F., Perham, R. N., *Biochem. J.* **109**, 312 (1968).
 Edelman, G. M., *Biochemistry* **9**, 3197 (1970).
 Edelman, G. M., Gall, W. E., Waxdal, M. J., Konigsberg, W. H., *Biochemistry* **7**, 1950 (1968).
 Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U., Waxdal, M. J., *Proc. Nat. Acad. Sci. U. S. A.* **63**, 78 (1969).
 Fleischman, J. B., Porter, R. R., Press, E. M., *Biochem. J.* **88**, 220 (1963).
 Gall, W. E., Edelman, G. M., *Biochemistry* **9**, 3197 (1970).
 Gall, W. E., Cunningham, B. A., Waxdal, M. J., Konigsberg, W. H., Edelman, G. M., *Biochemistry* **7**, 1973 (1968).
 Goldberg, H., Goldenberg, V., McLaren, A. D., *Biochim. Biophys. Acta* **7**, 110 (1951).
 Goldberger, R. F., Anfinsen, C. B., *Biochemistry* **1**, 401 (1962).
 Gottlieb, P. D., Cunningham, B. A., Rutishauser, U., Edelman, G. M., *Biochemistry* **9**, 3155 (1970).
 Green, N. M., Neurath, H. N., in *The Proteins*, H. Neurath and K. Bailey, Eds., Vol. II, Part B, Academic Press, New York, N. Y., 1954, p 1057.
 Gross, E., Witkop, B., *J. Amer. Chem. Soc.* **83**, 1510 (1961).
 Hill, R. L., Schmidt, W. R., *J. Biol. Chem.* **237**, 3151 (1962).
 Hirs, C. H. W., Moore, S., Stein, W. H., *J. Biol. Chem.* **235**, 633 (1960).
 Hunter, M. J., Ludwig, M. L., *J. Amer. Chem. Soc.* **84**, 3491 (1962).
 Itano, H. A., Gottlieb, A. J., *Biochem. Biophys. Res. Commun.* **12**, 405 (1963).
 Kaufman, S., Neurath, H. N., *Arch. Biochem.* **21**, 437 (1949).
 Kay, L. M., Schroeder, W. A., *J. Amer. Chem. Soc.* **76**, 3564 (1954).
 King, T. P., *Biochemistry* **5**, 3454 (1966).
 Li, C. H., Bertsch, L., *J. Biol. Chem.* **235**, 2638 (1960).
 Lundblad, R., Univ. North Carolina, personal communication (1970).
 Mitchell, W. M., *Science* **162**, 374 (1968).
 Mitchell, W. M., Harrington, W. F., *J. Biol. Chem.* **243**, 4683 (1968).
 Moore, S., Stein, W. H., *Methods Enzymol.* **6**, 1963 (1963).
 Ramachandran, L. K., Witkop, B., *Methods Enzymol.* **11**, 283 (1967).
 Redfield, R. R., Anfinsen, C. B., *J. Biol. Chem.* **221**, 385 (1956).
 Richards, F. M., Vithayathil, P. J., *J. Biol. Chem.* **234**, 1459 (1959).
 Rutishauser, U., Cunningham, B. A., Bennett, C., Konigsberg, W. H., Edelman, G. M., *Biochemistry* **9**, 3171 (1970).
 Sanger, F., Thompson, E. O. P., Kitai, R., *Biochem. J.* **59**, 509 (1955).
 Schroeder, W. A., Shelton, J. B., Shelton, J. R., *Arch. Biochem. Biophys.* **130**, 551 (1969).
 Schroeder, W. A., Kay, L. M., LeGette, J., Honnen, L., Green, F. C., *J. Amer. Chem. Soc.* **76**, 3556 (1954).
 Spande, T. F., Witkop, B., Degani, Y., Patchornik, A., *Advan. Protein Chem.* **24**, 98 (1970).
 Takahashi, K., *J. Biol. Chem.* **243**, 6171 (1968).
 Toi, K., Bynum, E., Norris, E., Itano, H., *J. Biol. Chem.* **240**, PC3455 (1965).
 Toi, K., Bynum, E., Norris, E., Itano, H., *J. Biol. Chem.* **242**, 1036 (1967).
 Waxdal, M. J., Konigsberg, W. H., Edelman, G. M., *Biochemistry* **7**, 1967 (1968a).
 Waxdal, M. J., Konigsberg, W. H., Henley, W. L., Edelman, G. M., *Biochemistry* **7**, 1959 (1968b).
 Weil, L., Telka, M., *Arch. Biochem. Biophys.* **71**, 473 (1957).

Received for review February 16, 1971. Accepted April 8, 1971. Presented at Symposium on Characterization of Proteins, Division of Agricultural and Food Chemistry, 160th ACS Meeting, Chicago, Ill., September 1970.