

The Amino Acid Sequence of Egg White Lysozyme

ROBERT E. CANFIELD*

From the Laboratory of Cellular Physiology and Metabolism, National Heart Institute, National Institutes of Health, Bethesda 14, Maryland

(Received for publication, December 10, 1962)

The separation and characterization of peptic and chymotryptic peptides of egg white lysozyme by phosphocellulose chromatography has been reported (1). In addition, a group of 18 different tryptic peptides of lysozyme, the total composition of which approximates that of the whole protein, has been characterized (2). This communication reports further studies that define a unique sequence for these tryptic peptides, representing the entire polypeptide chain of egg white lysozyme. Additional data are reported from which the complete amino acid sequence may be deduced.

The general plan that has been followed in these experiments has been to study, in detail, the chymotryptic and peptic fragments that contained lysine and arginine in order to obtain information that would establish the sequence of the complete set of tryptic peptides. To complement these data, the isolated tryptic peptides were digested with chymotrypsin, subtilisin, and pepsin to provide a matching set of peptides. Once the tryptic peptides were aligned, the rest of the peptides could then be assigned to their respective regions, and selective end group studies could be used to establish the exact position of each of the amino acid residues along the entire polypeptide chain.

EXPERIMENTAL PROCEDURE

Endopeptidases—The general procedure employed in these enzymic digestions was to dissolve 0.5 to 2 μ moles of a purified peptide in 0.5 ml of either 0.1 M ammonium bicarbonate (for trypsin, chymotrypsin, or subtilisin digestion) or 5% aqueous formic acid (for pepsin digestion). The equivalent of 25 μ l of a 1% solution of the desired proteolytic enzyme was added, and the mixture kept at 37° for 6 hours. The digestion was terminated by applying the mixture over a 7-inch line to Whatman No. 3 chromatography paper which was 18½ inches wide and 48 inches long. Separation of the peptides was accomplished by electrophoresis at pH 3.6 or in 4% formic acid at 5000 volts for 3 to 5 hours, depending on the type of separation desired (3, 4). After the peptides were located by staining, they were eluted in a closed, humid chamber with 5% aqueous pyridine. No eluting buffer has been found that gives a uniformly high yield of peptides, but the pyridine-water system seemed to be most generally applicable. A fraction of the eluate was then taken for amino acid analysis, and the remainder was saved for end group studies.

The chymotrypsin, pepsin, and trypsin were prepared as noted earlier (1, 2). The subtilisin was a commercial crystalline product obtained from the Nagarse Company, Osaka, Japan.

* Present address, Department of Medicine, Columbia University Medical School, New York, New York.

Exopeptidases—Leucine aminopeptidase and carboxypeptidase were obtained from the Worthington Biochemical Corporation and were Batches 5913A and B and 626/8, respectively.

The leucine aminopeptidase was dissolved at a concentration of 0.4 mg/100 ml in 0.1 N trimethylamine acetate at pH 8 and made 0.002 M in MnCl_2 . After activation at 37° for 30 minutes, aliquots were frozen until used. When serial leucine aminopeptidase digestion was employed to obtain information on the amino-terminal sequence, 0.2 to 0.3 μ mole of peptide was dissolved in 100 μ l of trimethylamine acetate at pH 8.5, and 10 μ l of 0.02 M MnCl_2 were added. Then 25 μ l of the above leucine aminopeptidase solution were added, and 10- μ l aliquots were removed at timed intervals thereafter. These were immediately applied to an origin line at one end of a sheet of Whatman No. 3MM paper (18½ × 48 inches). If a 1-inch applicator is used, a total of 12 separate applications (including standard amino acid mixtures) can be made on one paper. The amino acids were separated by electrophoresis in 4% aqueous formic acid with the use of an applied voltage of 5000 volts and, after drying, the papers were stained with a CdCl_2 -ninhydrin mixture (4). Strips, containing the separated components of each single application, were cut out, and the color produced by each amino acid spot was estimated with a Spinco analytrol equipped with an integrating, recording densitometer (4). When analyses of aliquots taken at 1- to 2-minute intervals were compared, the order of the first two or three NH_2 -terminal residues could usually be determined by the increasing order of their color, but on occasion, a slowly released residue was followed by an almost simultaneous increase in those residues adjacent to it. In certain instances, the sequential release of amino acids was followed quantitatively on an automatic amino acid analyzer.

Carboxypeptidase crystals were suspended and centrifuged twice in distilled water, and the pellet was then dissolved in 2 M NH_4HCO_3 at 23°. Diisopropyl fluorophosphate, 1 μ l per ml, was added, and the solution was gently stirred at 23° for ½ hour (5). The solution was centrifuged again to remove traces of insoluble material and aliquots were frozen until used. Carboxypeptidase digestions were carried out at 23° at an enzyme concentration of 1 mg per ml. Amino acid release was determined by the same methods as noted for leucine aminopeptidase.

The preparation of both leucine aminopeptidase and carboxypeptidase solutions was similar to that used by Potts *et al.* (6).

Amide Determination—Digestion with leucine aminopeptidase was also used for amide group determinations except that the digestion was allowed to proceed for a much longer period of time. The enzymic hydrolysates were separated by electro-

phoresis in formic acid as noted earlier, and the papers were stained with the ninhydrin-collidine color dip (1). Asparagine migrates to a position between threonine and methionine and yields a tan color on reaction with the ninhydrin-collidine stain. Glutamine moves just ahead of glutamic acid and produces a blue color. In instances where the peptide composition was complicated, the leucine aminopeptidase digest was analyzed on an automatic amino acid analyzer, by which method the relative quantities of aspartic and glutamic acids could be compared with values obtained following complete acid hydrolysis of an equal aliquot of the same peptide. Glutamine and asparagine emerge together with serine on the 50° column (7). When serine was known to be absent from the peptide, the appearance of a peak in this region, during analyses of leucine aminopeptidase digests, could be associated with a corresponding decrease in the recovery of aspartic or glutamic acids.

Edman Degradation—The Edman reaction was effected by the procedure described earlier (2). With small peptides, amino acid analyses were performed by electrophoresis in the 4% formic acid system, and the papers were stained with CdCl_2 -ninhydrin dip (4). This type of analysis permits one to perform as many as three Edman cycles on 0.2 μmole of an uncomplicated peptide, since the amino acid analysis can be performed with only 0.02 μmole . When the composition of the peptides suggested that the electrophoresis system might not yield adequate results, the analyses were performed on an automatic amino acid analyzer. This was especially indicated for peptides containing two or three residues of the same amino acid. Fig. 1 illustrates the results of an Edman procedure used to determine the amino-terminal sequence of 0.5 μmole of the peptide for which the composition after acid hydrolysis was (Gly, Asp, Val, Thr, Glu, Ala). Each densitometer tracing is shown above the portion of electrophoresis strip from which the peaks were recorded. Frequently, a visual comparison of the stained electrophoresis strips is adequate to determine the sequence of residues removed. The sequential decline of glycine, threonine, and aspartic acid in E-1, E-2, and E-3 is quite definite: E-0 is a sample taken prior to the initial Edman reaction.

DNP Determinations—Only a few NH_2 -terminal analyses were attempted by the DNP¹ method because the Edman degradation proved more useful. The DNP NH_2 -terminal residues that are reported were determined by the modified Levy procedure (8), and the chromatographic separation was achieved with 1.5 M potassium phosphate at pH 6.5 and *tert*-amyl alcohol-1 M NH_4OH (4:1) systems (9).²

Amino Acid Analysis—As noted above, the analysis of small quantities of amino acids by electrophoresis in 4% aqueous formic acid has been very useful in certain situations (4). The two principal advantages are the small quantity of amino acids required (approximately 0.02 μmole) and the ease with which many analyses can be performed in a short time. The quantitative analysis obtained after staining with ninhydrin- CdCl_2 dip and integration of the spots with a recording densitometer is not as accurate as the automatic amino acid analyzer and, therefore, has limited usefulness for peptides containing three or more residues of a single amino acid. In addition, the separation of valine from serine, and of methionine from proline, is difficult. Nevertheless, this has been a very valuable technique when

applied to the Edman degradation and to exopeptidase digestions in situations where the composition of the peptide was already known.

The remainder of the peptide compositions reported here were determined by the procedure of Spackman, Stein, and Moore (7). The neutral and acidic amino acids were separated on a 0.9- $\times$ -90-cm column, and the basic amino acids were separated on a 0.9- $\times$ -10-cm column. With increased buffer flow rate, three amino acid analyses can be performed in a day. Under these conditions, the adjacent peaks of threonine and serine, glycine and alanine, and tyrosine and phenylalanine overlap slightly, but it is still possible to integrate a representative area under each peak. Tyrosine and phenylalanine present the greatest difficulty, but none of the peptides analyzed on these columns contained both of these amino acids. Also, since there is only one histidine residue in lysozyme, the short column was generally only presented with the task of separating lysine, ammonia, and arginine.

The presence of tryptophan was determined by the Ehrlich reaction on the peptide spots, as noted earlier (1).

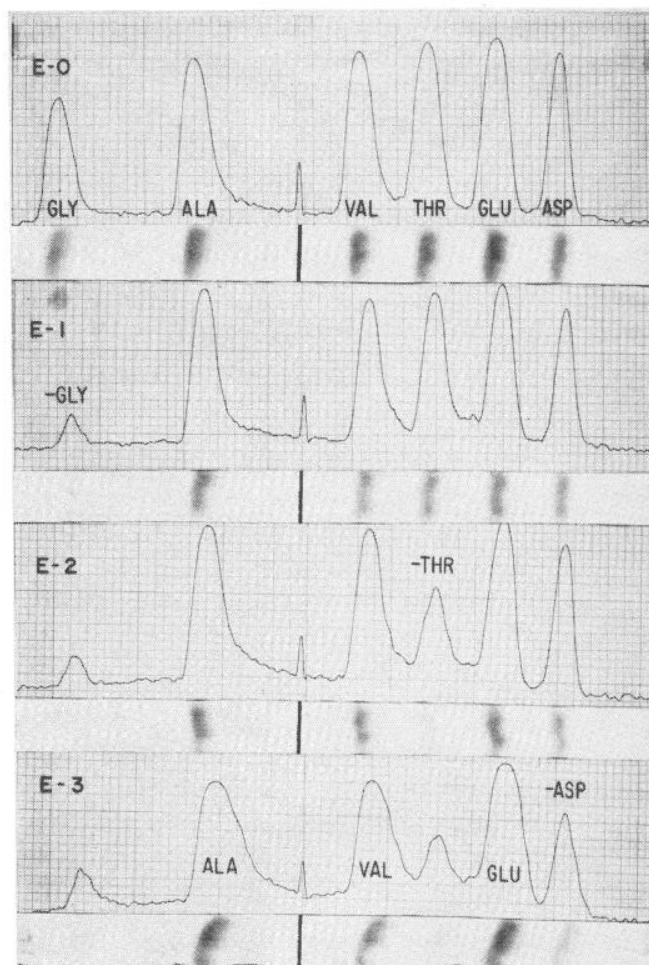


FIG. 1. Application of the paper electrophoretic method of amino acid analysis (Dreyer and Bynum (4)) to serial Edman degradation of the hexapeptide Gly-Thr-AspNH₂-Val-Glu-Ala obtained from P-18. Aliquots taken for analysis following completion of each Edman cycle are labeled E-1, E-2, and E-3. The filter paper strips are shown below the densitometer tracings.

¹ The abbreviations used are: DNP, dinitrophenyl derivative of amino acid; CMC, *S*-carboxymethylcysteine.

² B. Hartley, personal communication.

RESULTS

Since the results reported below include information obtained by a number of different experimental techniques (see "Experimental Procedure"), all of the data have been presented in a form that directly relates to the proposed sequence. Figs. 2, 3, 4, and 5 contain all the data that have been used to deduce the sequence of lysozyme. The sequence along the center of Figs. 2 to 5 is that proposed on the basis of the present studies. It is divided into regions which represent each of the tryptic peptides described in the accompanying paper (2). All of the other peptides that are represented in these figures are so positioned that their constituent amino acids are directly aligned with the portion of the lysozyme sequence from which the peptides are derived.

The endopeptidase digestions of the tryptic peptides are illustrated above the center line. The letters *C*, *S*, or *P* indicate whether chymotrypsin, subtilisin, or pepsin was used to produce the subpeptides. The vertical position of these small peptides in relation to their parent tryptic peptides is arbitrary. Each subpeptide has an identifying number which signifies the parent tryptic peptide and the enzymic digestion from which the partic-

ular smaller peptide was isolated. In general, all the peptides recovered from a given digestion mixture are illustrated. Slowly hydrolyzed tryptic peptides such as T-(3 + 4) are also indicated above the central sequence, as are the results from Edman degradations of the tryptic peptides.

The peptides resulting from peptic or chymotryptic digestion of reduced, carboxymethylated lysozyme are shown below the centers of the figures. Each of these is identified by the same number that was used in their original description (1). Since these two proteases are not completely specific in their action, two or more peptides were frequently produced from a given region. The pepsin and chymotrypsin peptides examined embrace every amino acid position in the sequence except for the tryptophan residue at position 63 which is represented by data from chymotrypsin digests only.

Positions determined by exopeptidase digestion are designated L for leucine aminopeptidase or C for carboxypeptidase, followed by a numeral which indicates the order in which the residues appeared. End group determinations by the DNP method are labeled as such, and E is used to designate those residues removed by the Edman degradation procedure. A

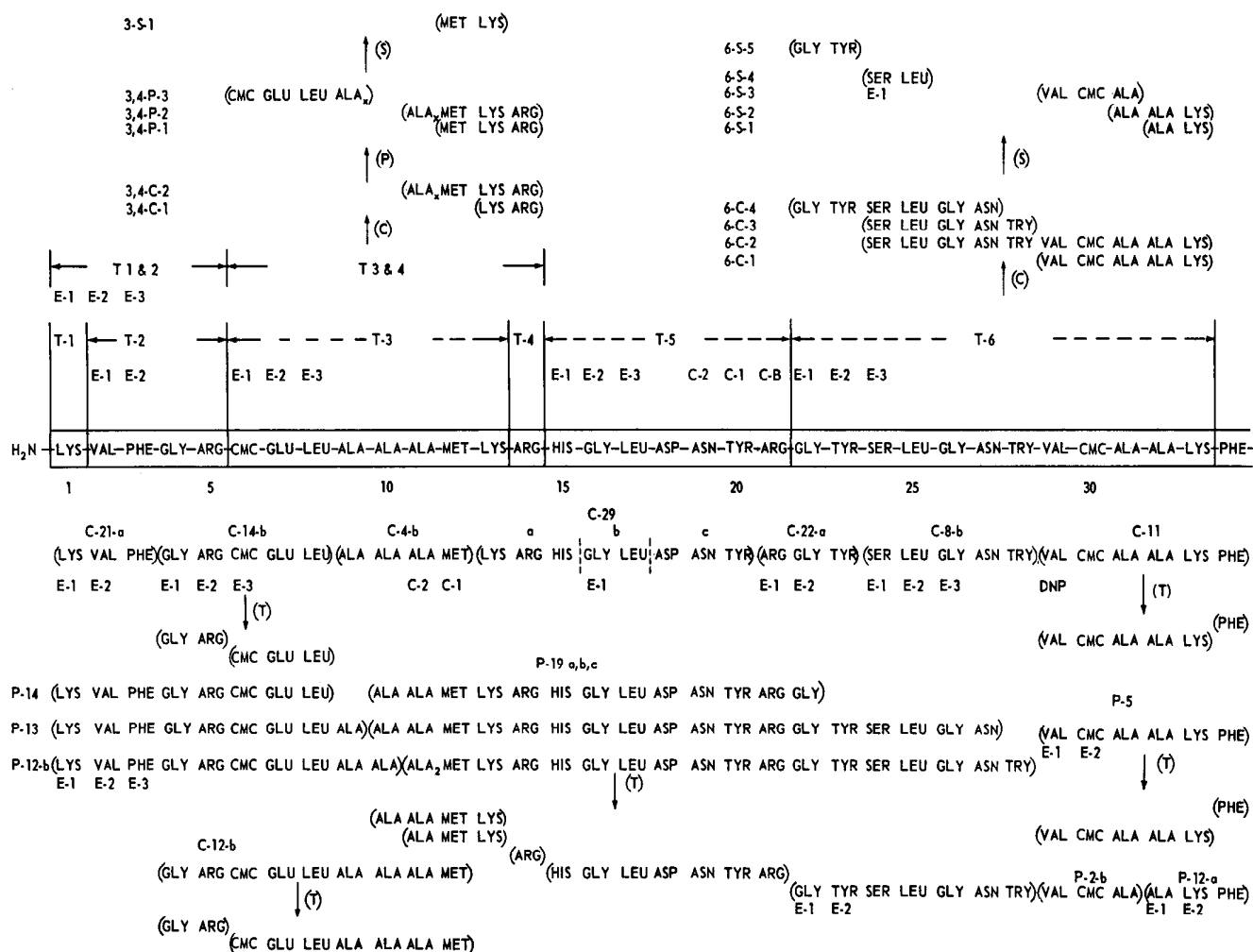


FIG. 2

Figs. 2, 3, 4, and 5. The peptide data that have been used to deduce the amino acid sequence for lysozyme. Each figure represents approximately one-quarter of the molecule, beginning with the NH₂-terminal region in Fig. 2 and ending with the COOH-

terminal region in Fig. 5. In instances where a particular peptide spans a region represented in two figures, a discontinuity in the peptide is represented by a +. The arrangement of each figure and the symbols employed are discussed in the text.

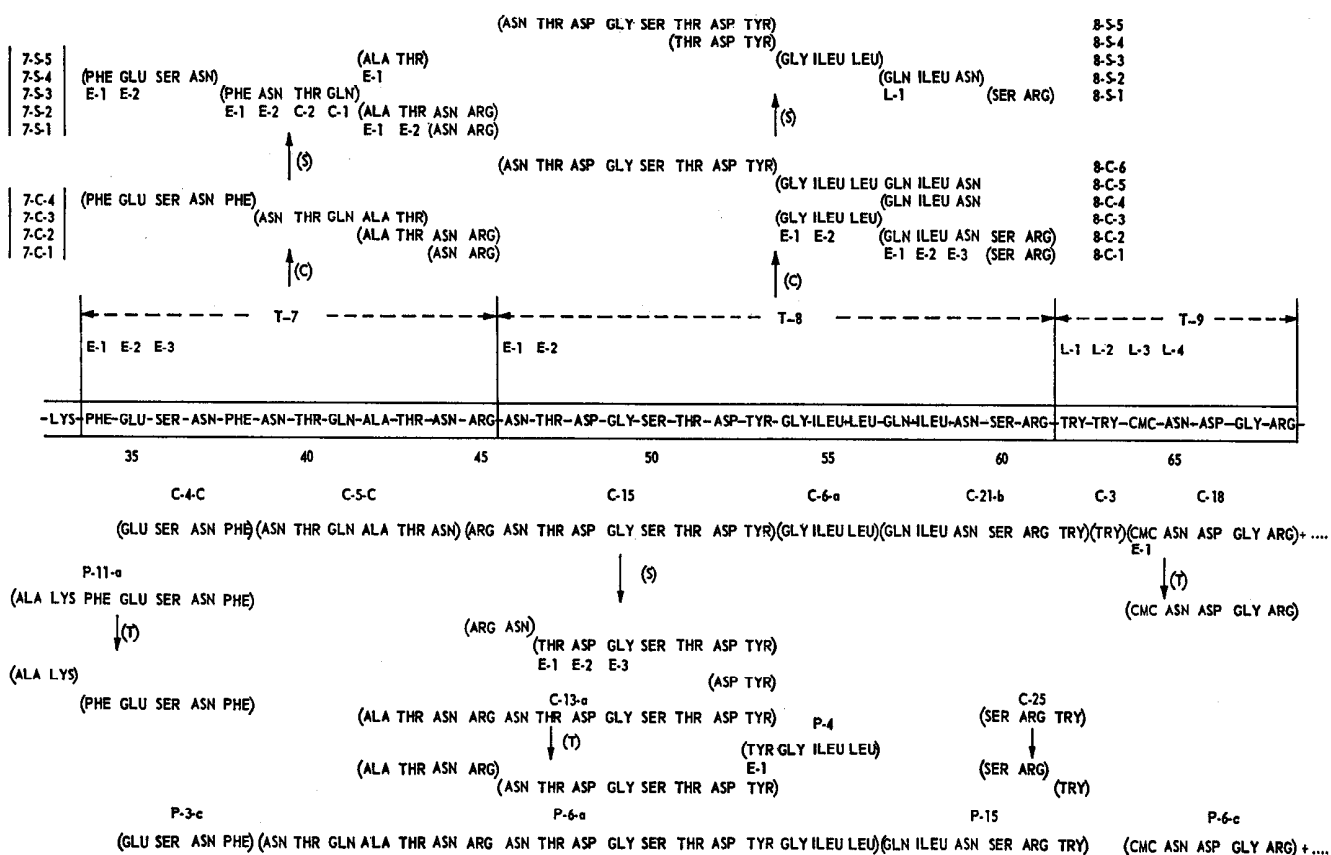


FIG. 3

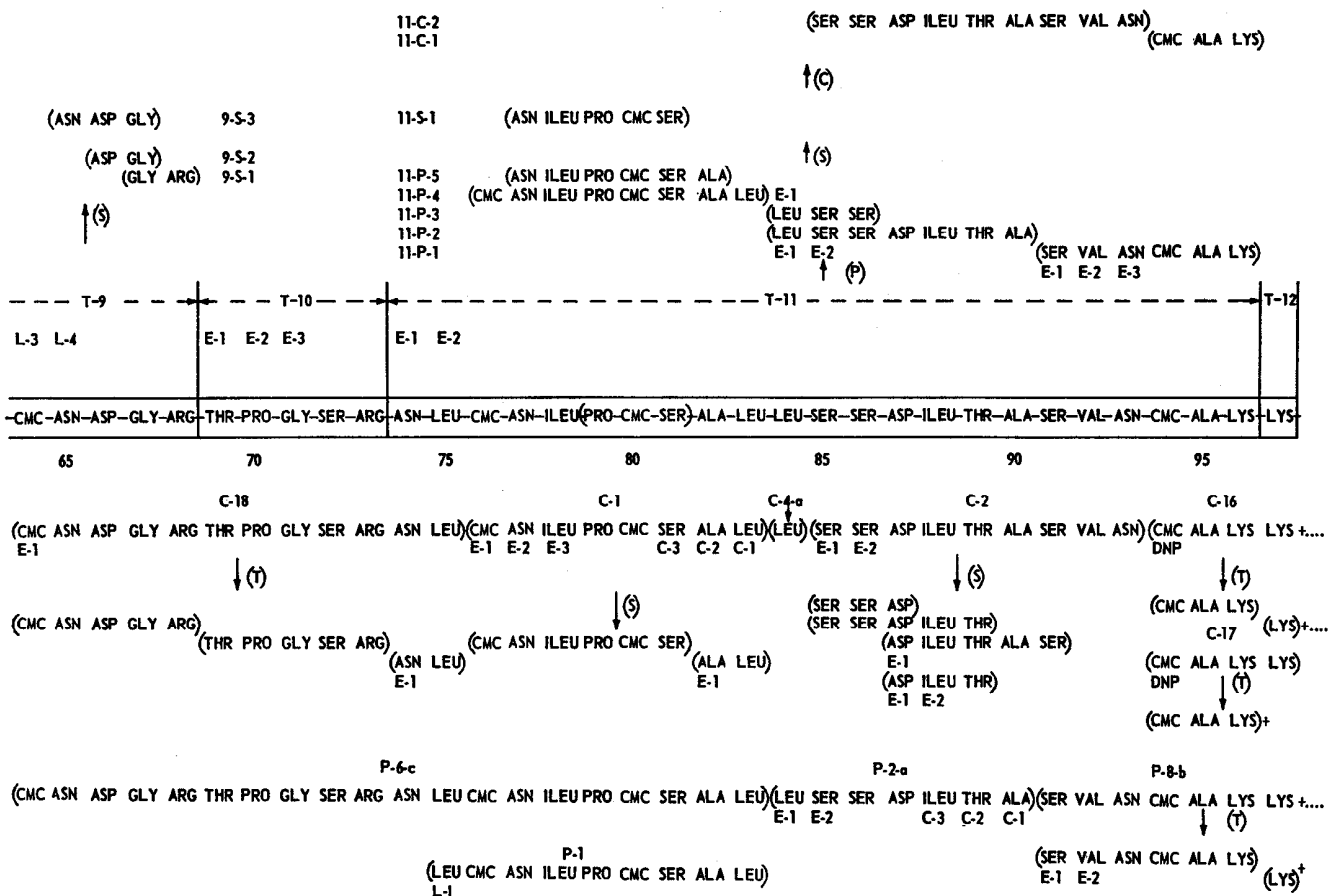


FIG. 4

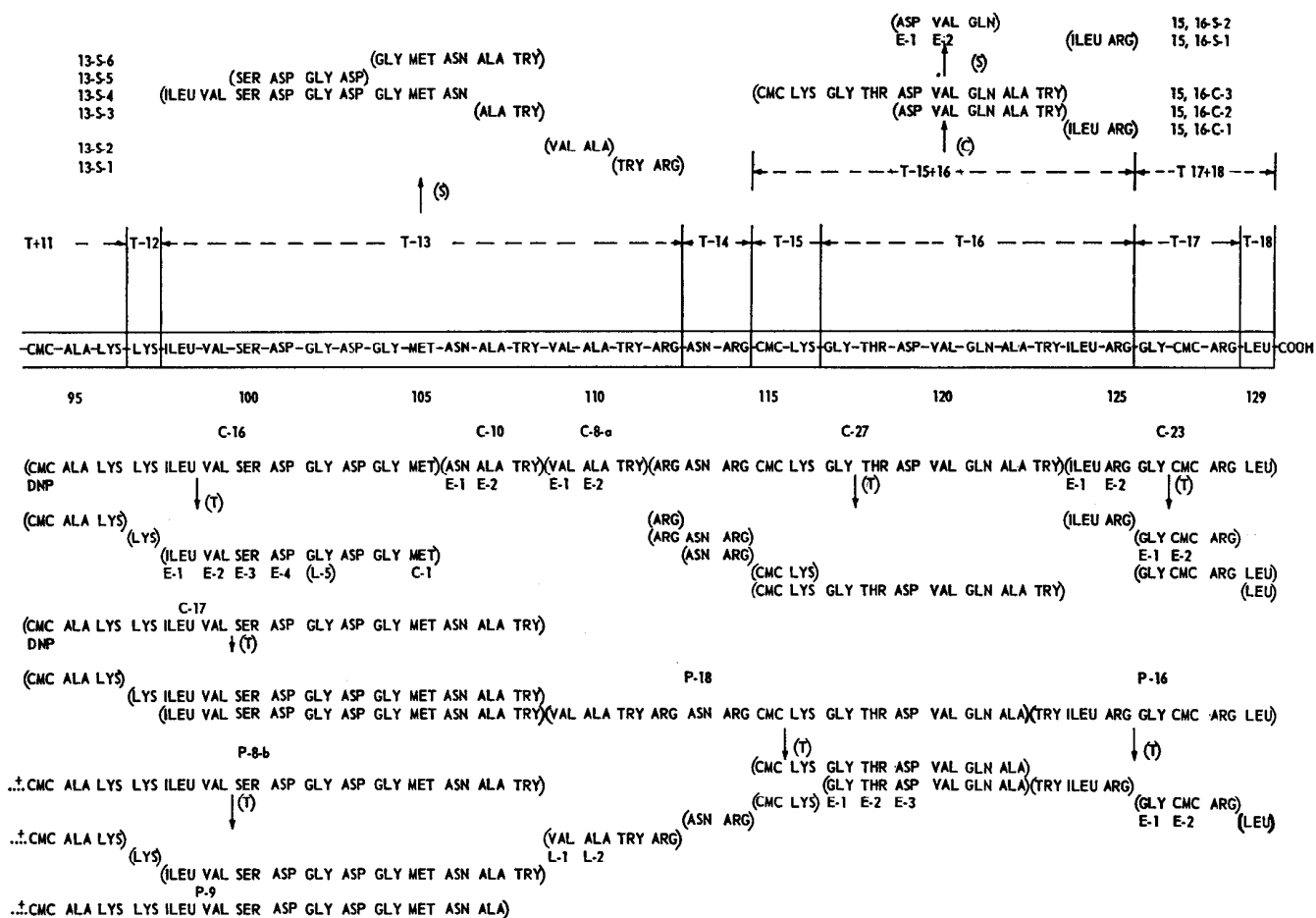


Fig. 5

numeral following the letter E indicates the step of the degradation during which an amino acid was removed. Although each peptide is illustrated by its proposed sequence, specific knowledge of the internal sequence of a given peptide is only implied when the particular residue is marked in the above manner.

The complete amino acid analyses of the tryptic, chymotryptic, and peptic peptides of reduced, carboxymethylated lysozyme are reported in the accompanying papers (1, 2). The composition of most of the other peptides illustrated in these figures was determined on 90- and 10-cm Amberlite IR-120 columns (see "Experimental Procedure"). Amino acid analysis by paper electrophoresis was only used in situations where the parent peptide did not contain more than one residue of any single amino acid.

The determination of the existence of the amide forms of aspartic or glutamic acid was generally attempted on small peptides that had been shown to contain only one aspartic or glutamic acid residue following acid hydrolysis. In certain instances, where several of these residues are present, the results of the analyses are noted in the text. The amide groups are indicated in each peptide in Figs. 2, 3, 4, and 5, but evidence for their existence was generally obtained from one peptide, and in each instance that peptide is identified in the text.

Some of the abbreviations used to indicate amino acid residues in Figs. 2 to 5 differ from the generally accepted forms. These differences (*Asn* for asparagine, *Gln* for glutamine, and *CMC*

for *S*-carboxymethylcysteine) have been introduced to simplify the construction of the figures.

Arrangement of Tryptic Peptides of Lysozyme and Their Amino Acid Sequence—Fraenkel-Conrat *et al.* (10) have shown that lysozyme is a single polypeptide chain by demonstrating that the reduced, alkylated protein has the same molecular weight as the native form. The repeated observations (11–20) of single NH_2 -terminal and COOH -terminal residues confirm this finding.

Schroeder (21) isolated a DNP-tetrapeptide with the sequence Lys-Val-Phe-Gly, and Acher *et al.* (22) located arginine as the 5th residue. Edman (23) has confirmed this sequence and extended it through the 14th position. His sequence is in agreement with that reported here.

In the following discussion, the existence of a single NH_2 -terminal lysine residue, which is NH_2 -terminal to a tetrapeptide sequence, Val-Phe-Gly-Arg, is assumed on the basis of the reports quoted above.

T-(1–3)—The amino-terminal lysine residue is only slowly released by trypsin and is usually isolated as the NH_2 -terminal residue of T-(1 + 2). Edman degradation has confirmed the presence of the sequence, Val-Phe, in the second and third positions. A similar result was obtained from Edman reaction with Peptide C-21-a. Finally, a long stretch of the amino-terminal sequence is fixed by the peptides which span Peaks P-12, P-13, and P-14. The NH_2 -terminal sequence of P-12-b is also Lys-Val-Phe.

C-14-b is the only chymotryptic peptide cleaved by trypsin to yield the fragment Gly-Arg, the COOH-terminal sequence of T-2. This fragment is coupled to CMC-Glu-Leu, which is the NH₂-terminal sequence of T-3 as derived by Edman degradation. The Edman degradation of C-14-b and trypsin cleavage of C-12-b confirm this overlap. C-4-b has a unique composition which can only fit T-3 and serial carboxypeptidase digestion places the methionine residue in the COOH-terminal position. The COOH-terminal sequence of T-3 is deduced from a series of peptides produced by digestion of T-3 and T-(3 + 4). The compositions of Peptides 3-S-1, 3,4-P-1, and 3,4-C-1 indicate that this sequence is Met-Lys-Arg, and that the terminal arginine is the only residue derived from T-4.

Leucine aminopeptidase digestion of T-3 yields glutamic acid in equimolar ratio with the other residues and no glutamine; *i.e.* nothing appears in the serine position on the amino acid analyzer.

T-(4-6)—The series of Peptides P-19-a,b,c, after digestion with trypsin, yield (a) COOH-terminal fragments of T-3 which must then represent the NH₂-terminal portion of P-19; (b) free arginine which must be T-4 on the basis of the evidence previously cited; (c) the complete tryptic peptide T-5; and (d) a COOH-terminal fragment (*i.e.* without lysine or arginine) the composition and amino-terminal sequence of which fit the NH₂-terminal sequence of T-6. These results serve to define the position of the third through sixth tryptic peptides.

Edman degradation of T-5 established His-Gly-Leu as the NH₂-terminal sequence. Serial digestion with leucine aminopeptidase rapidly removes these three residues followed by the slow but simultaneous rise of equimolar amounts of aspartic acid, asparagine, and tyrosine. An aliquot of T-5 was digested for 6 hours with carboxypeptidase B to remove its COOH-terminal arginine and carboxypeptidase A was then added. Tyrosine promptly appeared, but nothing else. Addition of large quantities of carboxypeptidase A and warming to 37° led to the appearance of a small but definite amount of asparagine. Therefore, aspartic acid has been assigned to position 18 and asparagine to position 19. This assignment is in keeping with the usual experience with leucine aminopeptidase digestion. Thus, asparagine was readily removed from the inverted sequence AspNH₂-Asp at positions 65 and 66.

Peak C-29 emerged with the alkaline wash from the column employed for chromatography of the chymotrypsin digest, and this fraction probably also contained active chymotrypsin. After lyophilization, the product yielded the three peptides, C-29-a,b,c, which almost certainly emanate from this region as a single peptide, subsequently digested by chymotrypsin in the test tubes containing the column effluent.

Edman degradation of T-6 shows Gly-Tyr-Ser to be the NH₂-terminal sequence, and this information permits the use of the sequence of C-22-a to confirm overlap T-(5-6). The composition of C-8-b requires placement of this peptide within T-6. The compositions of 6-C-3 and 6-C-4 are consistent with this placement.

Edman degradation of C-8-b extends this sequence to Ser-Leu-Gly-AspNH₂-Tyr. In this reconstruction, a comparison of 6-C-3 and 6-C-4 is used to position the tryptophan residue. End group studies on the lysine-containing fragments of C-11 and P-5 (which, from their compositions, can only be derived from the same part of the molecule as T-6), when considered in conjunction with the compositions of 6-C-1,2,3, permit com-

pletion of this region as Val-CMC-Ala-Ala-Lys. The fragments, P-2-b and P-12-a, confirm this, and Edman degradations of P-12-a identify the unique NH₂-terminal phenylalanine residue of T-7 and, therefore, the next tryptic overlap. Peptides P-5, P-11, and C-11, following trypsin cleavage, produce further evidence for this overlap.

Digestion of C-8-b with leucine aminopeptidase yields asparagine and no aspartic acid.

T-(7-8) (*Fig. 3*)—The Edman degradation of T-7 establishes Phe-Glu-Ser in the NH₂-terminal positions, and 7-S-4 places AspNH₂ as the next residue. Peptide 7-C-4 indicates the next residue to be phenylalanine, and the studies on 7-S-3 extend this sequence to AspNH₂-Phe-AspNH₂-Thr-GluNH₂ (see amide identifications below). The third cycle of the Edman reaction on 7-S-3 resulted in a decrease of only 10% in threonine content as compared to glutamic acid. If glutamine were occupying position 40, it is likely that pyrrolidone formation could have led to this relatively unsuccessful third Edman step (24). However, a serial carboxypeptidase digestion of 7-S-3 indicated, after 8 minutes of digestion, a 250% greater area under the serine (or glutamine) peak than the threonine peak. It can be stated that the serine peak did not contain asparagine because no phenylalanine was noted (liberation of asparagine would necessarily have liberated phenylalanine). The residual dipeptide, Phe-AspNH₂, was not identified. The COOH-terminal fragments of T-7, namely 7-C-1, 7-S-1, 7-S-2, and 7-S-5, establish the remainder of the sequence to be GluNH₂-Ala-Thr-AspNH₂-Arg.

Digestion of 7-S-3 with leucine aminopeptidase liberated glutamine and asparagine. A similar digestion of 7-S-4 liberated glutamic acid and asparagine. Peptide T-C-1 has the same electrophoretic mobility at pH 3.6 and color characteristics as T-14 and, therefore, contains asparagine.

Peptide C-13-a conveniently provides the T-(7-8) overlap by yielding a COOH-terminal fragment of T-7 and an NH₂-terminal fragment of T-8 after digestion with trypsin. When an aliquot of T-8 from the Dowex column (2) is subjected to electrophoresis at pH 3.6, or carried through a peptide-mapping procedure, it fails to give a colored spot with ninhydrin. This has been noted by others who were able to demonstrate the existence of the peptide, following paper electrophoresis, by the use of peptide bond stains.³ However, when an aliquot of the same fraction is digested with chymotrypsin or subtilisin, a peptide (8-S-5) containing the NH₂-terminal sequence is isolated following electrophoresis at pH 3.6, and this peptide produces a yellow color when stained with the ninhydrin-collidine dip (1). Furthermore, the Edman reaction is quite successful when carried out on the same batch of peptide that failed to stain on the peptide maps. The Edman results establish (AspNH₂-Thr) as the NH₂-terminal sequence of T-8. The composition of peptide C-15 indicates that it must be derived from the same portion of the molecule as T-8, and the composition of 8-S-5 places it in the NH₂-terminal region. Subtilisin digestion of C-15 yields a fragment, Arg-AspNH₂, which has the electrophoretic mobility of T-14, stains blue (inverted sequence for T-14), and yields asparagine following digestion with leucine aminopeptidase. This asparagine residue occupies position 46. Three Edman cycles on another subtilisin fragment of C-15 provide the sequence Arg-AspNH₂-Thr-Asp-Gly. The NH₂-terminal

³ T. C. Merigan, personal communication.

tyrosine on P-4 establishes its position as COOH-terminal in C-15 since there is only one tyrosine residue in T-8. The Asp-Tyr subtilisin fragment of C-15, and 8-S-4, leave only the single serine residue unaccounted for, and this residue must occupy position 50. The peptide C-15 thus has the sequence Arg-AspNH₂-Thr-Asp-Gly-Ser-Thr-Asp-Tyr.

Digestion of an aliquot of C-15 with leucine aminopeptidase produced the following (in micromoles): Asp, 1.71; Thr, 1.60; Ser (or color equivalent), 1.37; Gly, 0.86; Tyr, 0.78. These data indicate the presence of one asparagine and two aspartic acid residues. Since position 46 has been shown to be asparagine, 48 and 52 must then be aspartic acid residues. Quantitative amino acid analysis of a leucine aminopeptidase digestion of 8-C-6 yielded comparable results.

P-4, the composition of which is compatible only with T-8, links 8-S-3 and 8-C-3 with the sequence Tyr-Gly-Ileu-Leu derived by Edman degradation. The COOH-terminal fragment 8-C-2 occupies all the remaining positions of T-8, and results from Edman degradation of this peptide complete the sequence of T-8 as (GluNH₂-Ileu-AspNH₂-Ser-Arg). Digestion of C-21-b with leucine aminopeptidase yields glutamine and asparagine. The peptide, P-6-a, was found to have a composition which fits the region of residues 39 to 56. Unfortunately, it was recovered in very small yield following preparative electrophoresis, thus preventing further study.

T-(9-11) (Figs. 3 and 4)—The peptide, C-25, provides the overlap T-(8-9) since it is cleaved by trypsin to yield Ser-Arg and free tryptophan. The peptide Ser-Arg is the unique COOH-terminal sequence of T-8 and, similarly, tryptophan is NH₂-terminal only in the tryptic peptide, T-9.

The peptides C-24 and C-25 appeared in two distinct peaks from the phosphocellulose column (1), but they have the same composition following acid hydrolysis. It is possible that the difference between these two is attributable to the presence of two tryptophan residues (in positions 62 and 63) in peptide C-24, and only a single tryptophan residue (position 62) in C-25. Sufficient material for quantitative tryptophan analyses was not available. Analysis of peptide P-15, following leucine aminopeptidase digestion, indicated the presence of only one residue of tryptophan per molecule.

Quantitative amino acid analysis of a serial leucine aminopeptidase digestion of T-9 yielded two tryptophan residues followed by *S*-carboxymethylcysteine, and then asparagine, which emerges from the Amberlite IR-120 column in the serine position. Low yields of aspartic acid were also found. The COOH-terminal sequence is established by the compositions of 9-S-1, 2, and 3 (Fig. 4). The NH₂-terminal *S*-carboxymethylcysteine residue of C-18 (which is an Ehrlich-negative peptide) also confirms the sequence Try-Try-CMC-AspNH₂-Asp-Gly-Arg for T-9.

Tryptic cleavage of C-18 yields the COOH-terminal portion of T-9, the complete tryptic peptide T-10, and a COOH-terminal fragment AspNH₂-Leu, with asparagine being NH₂-terminal (the evidence for the amide function on this residue is noted below). Three Edman degradation cycles on T-10 establish the sequence of this peptide as Thr-Pro-Gly-Ser-Arg.

The AspNH₂-Leu fragment of C-18 (noted above) is unique as the NH₂-terminal sequence of T-11, and identifies this overlap. The composition of peptide P-6-c indicates that it contains elements of T-9, T-10, and T-11. The presence of the only two

proline residues of lysozyme in P-6-c also indicates that peptides C-1 and P-1 must have originated near the NH₂-terminal region of T-11. The alignment of C-1 and P-1 shown in Fig. 4 is the only one that satisfactorily accounts for the composition of P-6-c and the results of Edman degradation of T-11. Three Edman cycles on C-1 extend the NH₂-terminal sequence of T-11 to AspNH₂-Leu-CMC-AspNH₂-Ileu. Digestion of the AspNH₂-Leu fragment of C-18 and of C-1 with leucine aminopeptidase yielded asparagine in both instances.

The determination of the order of amino acid residues for positions 79 to 81 has offered particular difficulties. Subtilisin digestion of C-1 liberates a COOH-terminal fragment, Ala-Leu, the sequence of which was defined by removal of an NH₂-terminal alanine through Edman degradation. Serial carboxypeptidase digestions liberate first leucine, then alanine, and finally, serine and a fraction that appeared to be *S*-carboxymethylcysteine on ion exchange chromatography and paper electrophoresis. However, analysis of this latter fraction, following acid hydrolysis, indicated it to be residual peptide material. Thus, it appears that carboxypeptidase liberates leucine, alanine, small amounts of serine, and a residual peptide that mimics CMC.⁴ Serial leucine aminopeptidase digestion of C-1 rapidly liberates equimolar amounts of *S*-carboxymethylcysteine and asparagine followed by much smaller but equal amounts of the rest of the components of the peptide. The assignment of proline to position 79 would account for the slowing of digestion by leucine aminopeptidase at the penultimate isoleucine residue and leaves the second residue of *S*-carboxymethylcysteine to occupy position 80. Further study of 11-S-1 would have been helpful, but unfortunately, there was only a sufficient quantity for amino acid analysis. Thus, the sequence for positions 79 to 81 is inferred to be Pro-CMC-Ser, but the evidence is quite limited. To emphasize this fact, that region has been placed in parentheses in Fig. 4.

It is necessary at this point to turn to the COOH-terminal region of T-11 where tryptic fragments of C-16 and C-17, both of which were shown to contain NH₂-terminal *S*-carboxymethylcysteine by dinitrophenylation, establish CMC-Ala-Lys as the sequence. This is extended to Ser-Val-AspNH₂-CMC-Ala-Lys by serial Edman degradation on 11-P-1 and a tryptic fragment of P-8-b. Serial carboxypeptidase digestions of C-2 (analyzed by both column and paper methods) show equal amounts of asparagine and valine followed by serine. The peptide, C-2, then overlaps with P-8-b, and study of the fragments produced from subtilisin digestion of C-2, when considered together with its NH₂-terminal Ser-Ser sequence, allows one to deduce the entire sequence of residues 85 to 93. Serial carboxypeptidase digestion of P-1 confirms this, and sequential Edman degradations of the same peptide place leucine in position 84. This is probably the leucine residue, C-4-a.

Digestion of 11-P-1 with leucine aminopeptidase liberates asparagine from position 93. Similar digestion of P-2-a gives the following analysis (in micromoles): Asp, 0.45; Thr, 0.45; Ser (or chromatographic equivalents), 0.95; Ala, 0.43; Ileu, 0.49; Leu, 0.51, indicating that position 87 is occupied by aspartic acid.

T-(11-13) (Figs. 4 and 5)—The peptides C-16, C-17, P-8-b,

⁴ The author gratefully acknowledges assistance in interpretation of these data from Dr. John T. Potts.

and P-9 are all cleaved by trypsin into a COOH-terminal fragment of T-11, free lysine (which is the only residue of T-12) and a fragment the composition of which uniquely places it as an NH₂-terminal fragment of T-13. Four Edman cycles on this latter fragment of C-16 establish Ileu-Val-Ser-Asp for that region. A serial leucine aminopeptidase digestion of this fragment released equimolar amounts of glycine and aspartic acid before any methionine appeared (*i.e.* before the bond between the two COOH-terminal residues was cleaved). Since glycine and aspartic acid appear simultaneously and position 101 has been shown to contain aspartic acid, glycine has been assigned to position 102. Carboxypeptidase digestion of this same fragment yields only methionine, and this fact, coupled with the composition of 13-S-5 and 13-S-6, completes the sequence for that region as Ser-Asp-Gly-Asp-Gly-Met. A leucine aminopeptidase digest of the methionine-containing fragment from C-16 did not contain detectable asparagine on amino acid separation by paper electrophoresis. Both aspartic acid residues have, therefore, been listed as aspartic acid.

C-17 has three more residues than C-16 and, since both peptides have NH₂-terminal CMC residues, the peptide (Asp, Ala, Try) must be derived from the COOH-terminal end. This tripeptide has the same composition as C-10 which, on Edman degradation, was found to be Asp-Ala-Try. The composition of 13-S-3 is compatible with this sequence, and leucine aminopeptidase digestion of C-10 liberates asparagine from position 106. The association of 13-S-1 and 13-S-2 with C-8-a, which has the NH₂-terminal sequence Val-Ala, completes the sequence Val-Ala-Try-Arg.

Quantitative amino acid analysis of a leucine aminopeptidase digest of T-13 yields essentially equal areas under the aspartic acid and serine peaks, and this is consistent with the single serine and asparagine residues and two residues of aspartic acid that have been assigned to this peptide.

T-(14-18)—Trypsin digestion of P-18 liberates (a) a COOH-terminal fragment of T-13; (b) the complete tryptic peptides, T-14 and T-15; and (c) a fragment which by composition can only be the NH₂-terminal portion of T-16. Trypsin digestion of C-27 yields a similar group of peptides. In both instances, and in 15,16-C-3, peptides associating T-15 and T-16 have been isolated. Leucine aminopeptidase digestion of T-14 yields asparagine. The sequence for this region is thus Arg-AspNH₂-Arg-CMC-Lys. Three Edman cycles on the COOH-terminal fragment of P-18 (see Fig. 1), when compared with the data on 15,16-S-2, extends the sequence to Gly-Thr-Asp-Val-GluNH₂.

Trypsin digestion of peptides C-23 and P-16 liberates (a) peptides that by composition must be COOH-terminal fragments of T-16, (b) the tryptic peptide, T-17, and (c) free leucine. Since all of the tryptic peptides have thus been located within the lysozyme molecule, the COOH-terminal leucine residue of C-23 and P-16 is identical with the COOH-terminal residue of lysozyme (2). Edman degradation of C-23 locates the Ileu-Arg sequence, positions 124 and 125, and comparison with the composition of P-16 places tryptophan in position 123. Only alanine is left to occupy position 122. Edman degradation of the tryptic fragments of C-23 and P-16 permits completion of the sequence from positions 122 to 129 as Ala-Try-Ileu-Arg-Gly-CMC-Arg-Leu-COOH.

Digestion of T-16 with leucine aminopeptidase liberates aspartic acid and glutamine.

DISCUSSION

Data which are consistent with the NH₂-terminal and COOH-terminal portions of the amino acid sequence for lysozyme reported here have been mentioned earlier (11-23).

In an early study of Thompson (25, 26), a large number of lysozyme peptides, isolated after partial acid hydrolysis of performic acid-oxidized lysozyme, were analyzed by two-dimensional paper chromatography and studied with relation to end groups. On the basis of these data, 25 peptide sequences were proposed, containing from 3 to 6 amino acids each. Some of the results do not agree with those reported here and this may perhaps be explained by the fact that partial acid hydrolysis, in general, produced small peptides composed of the more abundant amino acids. Assignment of these to a given region of the chain of such a large molecule involved considerable uncertainty. However, attention must be directed to peptides containing unusual amino acids such as proline and methionine. The suggested sequence, Asp-Ala-Met-Lys-Cys-Arg, is considerably at variance with any region reported here, but the evidence presented for it contains elements of T-(3 + 4). The sequence Ser-Asp-Gly-Met-Asp could conceivably be mistaken for the Ser-Asp-Gly-Asp-Gly-Met-AspNH₂ sequence of T-13 since the differences following acid hydrolysis are quantitative and difficult to discern with paper chromatographic methods. The sequence Asp-Ileu-Pro-Cys is in agreement with that suggested here in tryptic peptide T-11.

The most significant published studies on the amino acid sequence of lysozyme are those which were initiated in the laboratory of Fromageot (22, 27-29) and carried on by Jollès *et al.* (15, 16, 30-35). In these publications, fragments of lysozyme produced by enzymic digestion have been reported and, in the most recent information that is available (35),⁵ they have been arranged in a continuous sequence from the amino to the carboxyl end of the polypeptide chain. Since the internal sequence of five regions containing 30 amino acids is not as yet available from the Paris group, the results of these studies and the present work cannot be completely compared. Also, the identification of five asparagine and glutamine residues has not been made. A discrepancy in the reported total number of amino acids is discussed in the accompanying paper (2). There are, however, certain differences which are already apparent.

1. In positions 40 to 42, the Jollès group report GluNH₂-Ala-Thr while Thr-GluNH₂-Ala is reported here.

2. The internal sequence of T-11 that has been assigned by Jollès differs from the present results in several ways. Position 76, reported here as a half-cystine residue, is asparagine in the Jollès sequence. There is agreement on a sequence Ala-Leu-Leu-Ser-Ser, but the residues in the sequence Val-AspNH₂ of positions 92 to 93 are inverted. The peptide Ser-Val-Asp-CySH-Ala found by Thompson (25) supports the model from this laboratory. There is agreement on a half-cystine residue in position 94, but the present sequence contains the other two half-cystine residues in positions 76 and 80. Although there is some uncertainty as to whether the latter half-cystine residue occupies position 79 or 80, these data do not support the presence of the second half-cystine residue noted by Jollès in position 93. Jollès reports four asparagine residues in T-11, but the present data indicate an aspartic acid residue in position 87.

⁵ P. Jollès, personal communication, September 1962.

- THAUREAUX, J., AND FROMAGEOT, C., *Bull. soc. chim. biol.*, **36**, 167 (1954).
30. JOLLÈS, P., AND THAUREAUX, J., *Compt. rend.*, **243**, 1685 (1956).
31. JOLLÈS-THAUREAUX, J., JOLLÈS, P., AND FROMAGEOT, C., *Biochim. et Biophys. Acta*, **27**, 298 (1958).
32. JOLLÈS, P., AND JOLLÈS, J., *Bull. soc. chim. biol.*, **40**, 1933 (1958).
33. JOLLÈS, P., JOLLÈS-THAUREAUX, J., AND FROMAGEOT, C., *Biochim. et Biophys. Acta*, **27**, 439 (1958).
34. JOLLÈS, P., JOLLÈS, J., AND JAUREGUI, J., *Biochim. et Biophys. Acta*, **31**, 96 (1959).
35. JOLLÈS, J., AND JOLLÈS, P., *Compt. rend.*, **253**, 2773 (1961).
36. TOWER, D. B., PETERS, E. L., AND WHERRETT, J. R., *J. Biol. Chem.*, **237**, 1861 (1962).