The Complete Amino Acid Sequence of Papain

ADDITIONS AND CORRECTIONS*

RONALD E. J. Mitchell, IRWIN M. CHAIKEN,† and EMIL L. SMITH‡

From the Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, California 90024

SUMMARY

The active cysteine of papain was labeled with 14C-iodoacetate and the cysteine residues were reduced and coupled with unlabelled iodoacetate. The heptacarboxymethyl papain was then maleylated and hydrolyzed with trypsin. Key peptides were isolated from this hydrolysate which have permitted completion of the amino acid sequence of the protein. Thus, an earlier tentative and incomplete version of this sequence has been corrected and shown to be in accord with the studies of Drenth et al. (DRENTH, J., JANSONIUS, J. N., KOEKOEK, R., SWEN, H. M., and WOLThERS, B. G., Nature, 218, 929 (1968)) by x-ray crystallographic methods.

In earlier publications on papain from this laboratory the amino-terminal (1) and carboxyl-terminal sequences (2), as well as the sequences of peptides from various types of hydrolysates of papain derivatives, were elucidated (3–10). Further, the active site cysteine residue and the three disulfide bridges were identified and a tentative amino acid sequence for the protein was proposed (11). At that time, it was clearly noted that the sequence was tentative, inasmuch as some of the overlaps were not as conclusive as desired, and that two gaps existed in the sequence, one of these being in the region following the active site cysteine residue and the three disulfide bridges.

Despite these deficiencies in providing a complete sequence of papain, the availability of the above information permitted Drenth et al. (12) to interpret the results of their x-ray studies of the crystalline enzyme and to deduce a model of the enzyme. From their work it was indicated that two interior sections of the sequence were transposed, that 11 residues were missing in one indicated gap in the sequence, and that a few additional minor discrepancies existed between the x-ray data and the chemically determined sequence analysis.1

In order to complete the studies and to reconcile the sequence deduced by use of the x-ray and the chemical methods, we have reinvestigated certain portions of the sequence of this protein. For the sake of brevity, we report only those portions of the sequence that have been useful in completing or reconciling the investigations from the two laboratories. After our work was essentially complete, Hussain and Lowe (13), using different methods, independently reported on the sequence of the missing residues near the active cysteine; however, they did not ascertain whether 1 residue was present as glutaminyl or glutamyl nor did they report on other portions of the sequence.

EXPERIMENTAL PROCEDURE

Reaction of Papain with Iodoacetic Acid—Twice recrystallized native papain (100 μmoles, 2.3 g (14)) was activated under nitrogen for 5 min with a 100-fold excess of cysteine in the presence of a 20-fold excess of EDTA in phosphate buffer at pH 7.2 and 37°. 2-14C-Iodoaceatic acid (498 μmoles) containing 6.5 μCi was added to the activated protein and allowed to react for 20 min at 37° under nitrogen. The completely inhibited monocarboxymethylated papain was separated from excess reagents by chromatography in 0.01 M sodium phosphate buffer at pH 7.2 on a column, 4 × 150 cm, of Sephadex G-50F and yielded 0.68 μmole of mono-carboxymethyl cysteine per μmole of protein based on 14C content. The yield was 76.7 μmoles based on A278 (E1% cm = 25) (15).

The above material was incubated for 30 min at 37° in a solution 6 M in guanidine, 0.2 M in sodium phosphate buffer, and 0.002 M in EDTA at pH 7.2. Dithioerythritol (595 mg, 3920 μmoles) was added and, after a further 10 min of incubation under nitrogen, 3.58 g of unlabeled iodoacetic acid were added. The reaction mixture was maintained at pH 7.2 and 37° for 1 hour. Exhaustive dialysis against water resulted in precipitation of the product (hepta-S-carboxymethyl papain) in which all 7 half-cystine residues (16) were carboxymethylated.

Maleylation of Hepta-S-carboxymethyl Papain—An aqueous suspension of hepta-S-carboxymethyl papain was adjusted to pH 9.0 at 40°. Solid maleic anhydride (17) was added slowly until the protein was completely dissolved (approximately 2 hours). The solution was stirred continuously and the pH was maintained between 8.5 and 9.2 with 6 N sodium hydroxide. The maleylated, carboxymethylated product was dialyzed against water adjusted to pH 9 with ammonia.

Tryptic Hydrolysis of Maleylated Hepta-S-carboxymethyl Papain—The completely inhibited papain was activated under nitrogen for 10 min at 37°. Trypsin (180 μg, 1.5 μmole) was added and, after a further 10 min of incubation under nitrogen, 3.82 g of unlabeled iodoacetic acid were added. The reaction mixture was maintained at pH 7.2 and 37° for 1 hour. Exhaustive dialysis against water resulted in precipitation of the product (maleylated hepta-S-carboxymethyl papain) in which all 7 half-cystine residues (16) were carboxymethylated.

In earlier publications on papain from this laboratory the amino-terminal (1) and carboxyl-terminal sequences (2), as well as the sequences of peptides from various types of hydrolysates of papain derivatives, were elucidated (3–10). Further, the active site cysteine residue and the three disulfide bridges were identified and a tentative amino acid sequence for the protein was proposed (11). At that time, it was clearly noted that the sequence was tentative, inasmuch as some of the overlaps were not as conclusive as desired, and that two gaps existed in the sequence, one of these being in the region following the active site cysteine residue.

Despite these deficiencies in providing a complete sequence of papain, the availability of the above information permitted Drenth et al. (12) to interpret the results of their x-ray studies of the crystalline enzyme and to deduce a model of the enzyme. From their work it was indicated that two interior sections of the sequence were transposed, that 11 residues were missing in one indicated gap in the sequence, and that a few additional minor discrepancies existed between the x-ray data and the chemically determined sequence analysis.1

In order to complete the studies and to reconcile the sequence deduced by use of the x-ray and the chemical methods, we have reinvestigated certain portions of the sequence of this protein. For the sake of brevity, we report only those portions of the sequence that have been useful in completing or reconciling the investigations from the two laboratories. After our work was essentially complete, Hussain and Lowe (13), using different methods, independently reported on the sequence of the missing residues near the active cysteine; however, they did not ascertain whether 1 residue was present as glutaminyl or glutamyl nor did they report on other portions of the sequence.

EXPERIMENTAL PROCEDURE

Reaction of Papain with Iodoacetic Acid—Twice recrystallized native papain (100 μmoles, 2.3 g (14)) was activated under nitrogen for 5 min with a 100-fold excess of cysteine in the presence of a 20-fold excess of EDTA in phosphate buffer at pH 7.2 and 37°. 2-14C-Iodoacetate acid (498 μmoles) containing 6.5 μCi was added to the activated protein and allowed to react for 20 min at 37° under nitrogen. The completely inhibited monocarboxymethylated papain was separated from excess reagents by chromatography in 0.01 M sodium phosphate buffer at pH 7.2 on a column, 4 × 150 cm, of Sephadex G-50F and yielded 0.68 μmole of mono-carboxymethyl cysteine per μmole of protein based on 14C content. The yield was 76.7 μmoles based on A278 (E1% cm = 25) (15).

The above material was incubated for 30 min at 37° in a solution 6 M in guanidine, 0.2 M in sodium phosphate buffer, and 0.002 M in EDTA at pH 7.2. Dithioerythritol (595 mg, 3920 μmoles) was added and, after a further 10 min of incubation under nitrogen, 3.58 g of unlabeled iodoacetic acid were added. The reaction mixture was maintained at pH 7.2 and 37° for 1 hour. Exhaustive dialysis against water resulted in precipitation of the product (hepta-S-carboxymethyl papain) in which all 7 half-cystine residues (16) were carboxymethylated.

Maleylation of Hepta-S-carboxymethyl Papain—An aqueous suspension of hepta-S-carboxymethyl papain was adjusted to pH 9.0 at 40°. Solid maleic anhydride (17) was added slowly until the protein was completely dissolved (approximately 2 hours). The solution was stirred continuously and the pH was maintained between 8.5 and 9.2 with 6 N sodium hydroxide. The maleylated, carboxymethylated product was dialyzed against water adjusted to pH 9 with ammonia.

Tryptic Hydrolysis of Maleylated Hepta-S-carboxymethyl Papain—The completely inhibited papain was activated under nitrogen for 10 min at 37°. Trypsin (180 μg, 1.5 μmole) was added and, after a further 10 min of incubation under nitrogen, 3.82 g of unlabeled iodoacetic acid were added. The reaction mixture was maintained at pH 7.2 and 37° for 1 hour. Exhaustive dialysis against water resulted in precipitation of the product (maleylated hepta-S-carboxymethyl papain) in which all 7 half-cystine residues (16) were carboxymethylated.

* This work was aided by Grant GM 11061 from the National Institute of General Medical Sciences, United States Public Health Service.
† Present address, Laboratory of Chemical Biology, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Maryland 19614.
‡ To whom inquiries and reprint requests should be sent.
§ It is a pleasure to acknowledge the courtesy and cooperation of Dr. J. Drenth in informing us of his results before publication.

1 It is a pleasure to acknowledge the courtesy and cooperation of Dr. J. Drenth in informing us of his results before publication.

2 Only one carboxymethyl group contains 14C.
**Papain**—The entire sample of maleylated hepta-S-carboxymethyl papain was warmed to 40° and adjusted to pH 7.8. A sample of 35 mg of trypsin, previously treated with L-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone, was added and the digestion mixture was maintained at pH 7.8 with 2 n sodium hydroxide. A further 17 mg of the treated trypsin were added after 32 min. The hydrolysis was complete in 56 min, after which 20 μmoles of diisopropylfluorophosphate were added to inactivate the trypsin.

**Other Methods**—Amino acid analysis of peptides was performed on a Spinco automatic analyzer after hydrolysis for 24 hours, unless otherwise noted, under vacuum in 6 N HCl. Radioactivity was counted in a Packard model 3003 Tri-Carb scintillation counter. Radioactive peptides were located on paper with a Packard model 7201 radiochromatogram scanner coupled to a Packard model 385 recording rate meter.

Paper chromatography was performed on sheets of Whatman No. 3MM paper and, unless otherwise noted, run for approximately 16 hours in 1-butanol-acetic acid-water (40:6:15, v/v/v). Electrophoresis was done on Whatman No. 3MM paper at 2500 volts for 45 min under Varsol; the solvents have been previously described (18).

The buffers for elution of peptides from Dowex 50 columns were as follows: pH 3.3: 16.1 ml of pyridine, 279 ml of acetic acid, 200 ml of 1-propanol, diluted with water to 1 liter; pH 4.9: 161 ml of pyridine, 115 ml of acetic acid, 200 ml of 1-propanol diluted to 1 liter. Column elution profiles were determined after alkaline hydrolysis of peptides by measuring the absorption of aliquots at 570 nm after reaction with ninhydrin. All samples from column eluates were dried or concentrated at low temperature under vacuum by rotary evaporation.

The Edman degradation procedure was performed as described by Kasper and Smith (18). Phenylthiohydantoin derivatives were hydrolyzed to free amino acids with 6 N HCl at 150° for 12 hours. The free amino acids were identified by paper chromatography or clectrophoresis or on the amino acid analyzer.

**RESULTS**

In order to facilitate following the results of the present studies, the complete primary amino acid sequence of papain is shown in Fig. 1.

It is now evident that part of the early difficulties in studying the sequence of this protein, first undertaken more than 15 years ago, is the fact that most of the inactive derivatives of this proteinase are insoluble in water (9, 19). This rendered difficult the problem of obtaining satisfactory hydrolysis by proteolytic enzymes, particularly trypsin, and hence also the isolation of peptides in reasonable yields. The necessarily prolonged times used for hydrolysis frequently led to secondary cleavage produced by traces of other proteolytic enzymes. In

![Amino Acid Sequence of Papain](image-url)
the present study these problems were circumvented by use of
the completely carboxymethylated papain which had been fully
maleylated by the method of Butler et al. (17). This soluble
derivative was readily hydrolyzed by trypsin in less than 1 hour
(see above).

Inasmuch as the sequence had already been definitively
established through residue 26 (8, 11) (see Fig. 1), it could be
predicted that hydrolysis would occur at arginyl residue 8, and
that the tryptic peptide (Tm-2) from the adjacent portion of
the sequence would include the active cysteinyl residue 25 labeled
with 14C as carboxymethyl cysteine.

Tryptic peptides from the maleylated protein are numbered
on the basis of the number of arginyl residues from the amino
terminus of papain at which hydrolysis had occurred. Number-
ing of other peptides is arbitrary.

Isolation of Tm-2 (Residues 9 through 41)

The tryptic hydrolysate of maleylated hepta-S-carboxymethyl
papain was fractionated on a column, 4 × 150 cm, of Sephadex
G-50F in 0.05 M pyridine (Fig. 2) and the fractions were pooled
as indicated. Fraction A, which presumably consisted of
undigested material or large fragments, was not studied further.
Fraction B was demaleylated by adjusting to pH 3.0 with 6
HCl and maintaining the solution at 40° for 60 hours. The
sample was dried, dissolved in 20 ml of formic acid, and imme-
diately diluted to 50 ml with 30% acetic acid. The resulting
solution was divided into three portions and each was chroma-
tographed on a column, 1.8 × 150 cm, of Sephadex G-50F in
30% acetic acid (Fig. 3). Fraction 1, eluting at the break-
through volume, proved to contain solely Peptide Tm-2. The
peptide did not move significantly on paper electrophoresis at
pH 1.9 or on paper chromatography with various solvents.

The composition of Peptide Tm-2 is given in Table I and
corresponds to the papain segment containing residues 9 through
41. It is noteworthy that prolonged hydrolysis in 6 N HCl was
necessary to achieve complete release of all of the isoleucine
and valine present in the peptide. Identification of this peptide as
representing residues 9 through 41 was facilitated by the fact
that it contains not only the radioactivity of the labeled cysteine
(residue 25) but also 1 of the 4 phenylalanine and 1 (residue 26)
of the 5 tryptophanyl residues of the protein. Total yield based
on radioactivity was 27 μmoles.

In earlier work two peptides containing phenylalanine had not
been accommodated in the sequence; these were Ala–Phe and
Ala–Phe–Ser (7). These peptides are obviously derived from
this region of the molecule as are additional residues of isoleu-

---

**Figure 2.** Fractionation of tryptic hydrolysis products of male-
ylated hepta-S-carboxymethyl papain on a column of Sephadex
G-50F (4 × 150 cm) run in 0.05 M pyridine at an elution rate of
1 ml per min. Fractions of 10 ml were collected. Ninhydrin ab-
sorbance at 570 nm and counts per min were determined on 20-μl samples.

**Figure 3.** Chromatography of Fraction B (Fig. 2) on a column of
Sephadex G-50F (4 × 150 cm) run in 30% acetic acid at the rate
of 1 ml per min. Fractions of 10 ml were collected. Ninhydrin
absorbance was determined at 570 nm on 50-μl samples; counts per
min on 20-μl samples.

---

**Table I**

<table>
<thead>
<tr>
<th>Amino acid composition of Peptide Tm-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Amino acid</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tryptophane</td>
</tr>
<tr>
<td>Lysine</td>
</tr>
<tr>
<td>Arginine</td>
</tr>
<tr>
<td>Carboxymethyl cysteine</td>
</tr>
<tr>
<td>Aspartic acid</td>
</tr>
<tr>
<td>Asparagine</td>
</tr>
<tr>
<td>Glutamic acid</td>
</tr>
<tr>
<td>Proline</td>
</tr>
<tr>
<td>Glycine</td>
</tr>
<tr>
<td>Alanine</td>
</tr>
<tr>
<td>Valine</td>
</tr>
<tr>
<td>Isoleucine</td>
</tr>
<tr>
<td>Leucine</td>
</tr>
<tr>
<td>Tyrosine</td>
</tr>
<tr>
<td>Phenylalanine</td>
</tr>
</tbody>
</table>

a Detected by the Ehrlich reaction.

b Values are uncorrected for losses on hydrolysis.
Peptide Tm-2 was hydrolyzed with acid and identified after analysis as undigested Peptide Tm-2. Analysis of a portion of Fraction B indicated a single component, Peptide Tm-2-S7, comprising residues 18 through 41 (Table II).

Fraction C, containing the smaller peptides, was chromatographed on a column (0.9 × 50 cm) of Dowex 50-X2 at 40°C. The equilibrating buffer consisted of 400 ml of pH 3.3 pyridine-acetic acid buffer in the mixing vessel and 400 ml of pH 4.9 pyridine-acetic acid buffer in the reservoir. Further washing with the pH 4.9 buffer yielded no more peptides (Fig. 5). The purification, composition, and sequence of each peptide obtained from the fractions designated in Fig. 5 are described below. For those peptides derived from the known portion of the sequence, residues 9 through 26, no detailed description is given.

Fig. 5. Chromatography of Fraction C (Fig. 4) on a column (0.9 × 50 cm) of Dowex 50-X2 at 40°C. The linear gradient consisted of 400 ml of pH 3.3 pyridine-acetic acid buffer in the mixing vessel and 400 ml of pH 4.9 pyridine-acetic acid in the reservoir. The column was run at 18 ml per hr and 3-ml fractions were collected. Ninhydrin absorbance was measured at 570 nm on 30-ml aliquots. An aliquot of Fraction A (Fig. 4) was hydrolyzed with acid and identified after analysis as undigested Peptide Tm-2. Analysis of a portion of Fraction B indicated a single component, Peptide Tm-2-S7, comprising residues 18 through 41 (Table II).

Fraction C, containing the smaller peptides, was chromatographed on a column (0.9 × 50 cm) of Dowex 50-X2 at 40°C. The equilibrating buffer consisted of 400 ml of pH 3.3 pyridine-acetic acid buffer in the mixing vessel and 400 ml of pH 4.9 pyridine-acetic acid buffer in the reservoir. Further washing with the pH 4.9 buffer yielded no more peptides (Fig. 5). The purification, composition, and sequence of each peptide obtained from the fractions designated in Fig. 5 are described below. For those peptides derived from the known portion of the sequence, residues 9 through 26, no detailed description is given.

Peptide Tm-2-S10—Fraction b was further purified by electrophoresis at pH 3.6. Analysis showed the following composition: Val, 1.07(l); Thr, 0.83(l); Ile, 0.91(l); Glu, 1.12(l). Treatment with carboxypeptidase A at pH 8.1 failed to release any amino acids. One step of the Edman degradation followed by acid hydrolysis of the phenylthiohydantoin derivative yielded valine. A second step of the Edman degradation showed no amino acid after acid regeneration. A third step gave isoleucine after acid hydrolysis of the phenylthiohydantoin derivative. The residual amino acid was identified as glutamic acid by paper chromatography. The sequence is, therefore, Val-Thr-Ile-Glu. The intact peptide proved to be strongly anionic when subjected to electrophoresis at pH 6.5, indicating the presence of a glutamyl residue.

Peptide Tm-2-S2—Fraction c was further fractionated by electrophoresis at pH 1.9 and consisted mainly of Val-Thr-Pro-Val (residues 13 through 16) and free glycine. Peptide Tm-2-S5—Fraction d was further purified by paper chromatography and yielded mainly Lys-Asn-Gln-Gly (residues 17 through 20).

Peptide Tm-2-S4—Fraction e was also separated by paper chromatography and consisted of free alanine and the peptide Glu-Lys-Gly-Val (residues 9 through 12).

Peptides Tm-2-S8 and Tm-2-S9—Fraction g, fractionated by...
electrophoresis at pH 1.9, consisted of Peptide Tm-2-SS, Ala-Phe (residues 27 through 28), plus another peptide which gave the composition: Ser, 0.90; Ala, 1.10. A single step of the Edman degradation released free alanine. The sequence of Peptide Tm-2-S9 is, therefore, Ser-Ala (residues 29 through 30).

**Peptide Tm-2-SS**—Fraction j was further purified by paper chromatography and yielded the peptide Thr-Pro-Val-Lys (residues 14 through 17).

**Peptide Tm-2-S11**—Fraction k was subjected to electrophoresis at pH 3.6. A fast moving positively charged component was further purified by paper chromatography. Acid hydrolysis for 120 hours followed by amino acid analysis gave the following composition: Gly, 1.00(1); Ile, 1.88(2); Lys, 1.12(1). Hydrolysis for 48 hours yielded only 1.50 eq of isoleucine, indicating the presence of an Ile-Ile bond. One cycle of Edman degradation followed by acid hydrolysis of the phenylisothiocyanate derivative yielded glycine. Treatment with carboxypeptidase B released only lysine. The sequence is therefore Gly-Ile-Ile-Lys (residues 36 through 39).

**Peptide Tm-2-SS**—Fraction l contained only one peptide which was identified as Lys-Asn-Gln (residues 17 through 19).

**Peptide Tm-2-S4**—Fraction m was further fractionated by electrophoresis at pH 1.9 and consisted mainly of the peptide Val-Lys-Asn (residues 16 through 18).

**Peptide Tm-2-S12**—Fraction o contained a single component. Acid hydrolysis for 96 hours followed by amino acid analysis gave the composition: Ile, 0.91(1); Arg, 1.09(1). One step of the Edman degradation followed by acid hydrolysis of the phenylisothiocyanate derivative yielded isoleucine. The sequence is Ile-Arg (residues 40 through 41).

**Peptide Tm-2-S7-A1**—Peptide Tm-2-S7 (from Fraction B, Fig. 4) was hydrolyzed under vacuum at 110º with 1.0 N HCl for 2 hours in order to obtain an additional overlapping peptide. The hydrolysate was chromatographed on a column, 1.8 x 150 cm of Sephadex G-25F in 30% acetic acid (Fig. 6).

Fraction B (Fig. 6) was further purified by electrophoresis at pH 3.6 and by paper chromatography. Peptide Tm-2-S7-A1, after acid hydrolysis for 120 hours, gave the following composition: Ala, 1.10(1); Val, 1.95(2). For brevity, other components which were isolated from this acid hydrolysate are not described since they did not contribute materially to the sequence deduced.

**Sequence of Residues 9 through 41**—The peptides isolated from Peptide Tm-2 are shown in Fig. 7 with the complete sequence indicated. No comment is required concerning residues 9 through 26 since this portion of the sequence had been established earlier (8, 11). The sequences Ala-Ala-Phe and Ala-Phe-Ser Gly(S)-Phe allowed extension to include Ser-Ala (Peptide Tm-2-SS) since there is only 1 seryl residue in this portion of the structure. Peptide Tm-2-S7-A1 is Ala-Val-Val since the sequence Ala-Phe-Ser-Ala has already accounted for both alanine residues of Peptide Tm-2-SS and this allows extension through Peptide Tm-2-S10, Val-Thr-Ile-Glu, insomuch as there are only 2 valyl residues in this region of the peptide. Peptide Tm-2-S11, Gly-Ile-Ile-Lys, is placed by difference since Ile-Arg (Peptide Tm-2-S2-S12) must be COOH-terminal to account for the site of tryptic hydrolysis of the maleylated protein.

**Peptide Tm-10 (Residues 112 through 140)**

Fraction C (Fig. 2) was demaleylated as previously described for Fraction B. A precipitate formed and was removed by...
Fig. 7. The complete sequence of Peptide Tm-2 (residues 9 through 41, Fig. 1). The sequence containing residues 9 through 26 had been reported earlier (8, 11).

Fig. 8. Chromatography of Fraction D (Fig. 2) on a column of DEAE-Sephadex A-50 (0.9 × 50 cm) with a linear gradient of sodium bicarbonate, 0.1 M to 0.5 M in a total volume of 400 ml. Fractions of 4 ml were collected at a flow rate of 0.5 ml per min. The location of Peptide Tm-10 is shown. Ninhydrin absorbance was measured on 0.25-ml aliquots, radioactivity on 0.05-ml samples.

Centrifugation. The supernatant solution was dried, dissolved in 2 N pyridine, and chromatographed at 35° on a column, 0.9 × 160 cm, of Bio Rad AG1 X2 in acetate form, equilibrated with 2 N pyridine. Washing with 2 N pyridine eluted Peptide Tm-12 at the front.

Acid hydrolysis for 72 hours gave the analysis: Ile, 0.93(1); Lys, 1.00(1); Arg, 1.07(1). A single step of the Edman degradation followed by paper chromatography of the regenerated free amino acid showed isoleucine. The remaining dipeptide was hydrolyzed for 72 hours; paper chromatography showed only lysine and arginine in equal amounts. Since lysine residues were blocked during the tryptic hydrolysis the sequence is Ile-Lys-Arg and this peptide must be preceded by an arginyl residue.

Table III

Amino acid compositions of Peptide Tm-10 and related peptides

<table>
<thead>
<tr>
<th>Residue</th>
<th>Peptide Tm-10</th>
<th>Peptide Tm-10d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21 hrs</td>
<td>72 hrs</td>
</tr>
<tr>
<td>Lysine</td>
<td>0.09</td>
<td>1.02(1)</td>
</tr>
<tr>
<td>Arginine</td>
<td>1.01</td>
<td>0.91(1)</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>2.35</td>
<td>3.00(3)</td>
</tr>
<tr>
<td>Serine</td>
<td>1.80*</td>
<td>1.49*2</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>6.02</td>
<td>6.13(6)</td>
</tr>
<tr>
<td>Proline</td>
<td>1.98</td>
<td>2.15(2)</td>
</tr>
<tr>
<td>Glycine</td>
<td>2.03</td>
<td>2.05(2)</td>
</tr>
<tr>
<td>Alanine</td>
<td>3.05</td>
<td>4.07(4)</td>
</tr>
<tr>
<td>Valine</td>
<td>3.22</td>
<td>3.89(4)</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>1.00</td>
<td>0.98(1)</td>
</tr>
<tr>
<td>Leucine</td>
<td>4.02</td>
<td>4.02(4)</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>3.00</td>
<td>2.97(3)</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>1.01</td>
<td>0.95(1)</td>
</tr>
</tbody>
</table>

* Uncorrected.
Residues 64 through 83

In the proposed tentative sequence for papain (11) it was known that no overlaps were available in the sequence now known to represent residues 60 through 83 (12) and to include the tripeptide, Ser-Ala-Leu, which was not connected to the remainder of the sequence, nor was it known whether any residues were missing in this portion of the sequence. This was due to the isolation in low yield and relatively poor state of purity of the tryptic Peptide A-T-XXXII (9). In order to attempt to complete this portion of the sequence, Drs. W. H. Evans and D. P. Botes studied a tryptic digest of aminoethylated papain. Only a brief summary of their experiments concerning this portion of the sequence is given here in order to indicate the evidence for placing the tripeptide, Ser-Ala-Leu, and the missing dipeptide Gln-Leu.

Papain was fully reduced in 10 M urea with a 100-fold excess of redistilled mercaptoethanol and then aminoethylated with excess ethyleneimine (20). Excess reagents were removed and the $\lambda$-aminoethylated papain (90 $\mu$moles) was digested with two additions of 20 mg of trypsin in 0.1 M NH$_4$HCO$_3$ at pH 7.9 for 4 hours. The digest was fractionated on Dowex 1-X2 by the methods described by Schroeder and Robberson (21). The peptide in question emerged in a fraction eluted near pH 8; it was further purified on Sephadex G-50 in a solvent containing 20% acetic acid and 15% propanol. Its composition indicated 

<table>
<thead>
<tr>
<th>No.</th>
<th>Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>Asp-Gly-Gly-Tyr-Pro-Trp-Ser-Ala-Leu-Gln-Leu-Val-Ala-Gln-Tyr-Gly-Ile-Ile-Tyr-Arg</td>
</tr>
<tr>
<td>70</td>
<td>(Pa-1)</td>
</tr>
<tr>
<td>75</td>
<td>(Pa-2)</td>
</tr>
<tr>
<td>80</td>
<td>(Pa-3)</td>
</tr>
<tr>
<td>83</td>
<td>(Pa-4)</td>
</tr>
<tr>
<td>64</td>
<td>(Pa-5)</td>
</tr>
<tr>
<td>66</td>
<td>(Pa-6)</td>
</tr>
<tr>
<td>73</td>
<td>(Pa-7)</td>
</tr>
</tbody>
</table>

The complete sequence of residues 64 through 83. A chymotryptic peptide including residues 62 through 69 had been isolated earlier (11); the sequence has been corrected on the basis of the work of Drenth et al. (12). The sequence of residues 75 through 83 has been reported earlier (11). The peptides obtained from a papain (Pa) digest have permitted completion of this portion of the sequence.

Since there is only 1 serine residue in the entire peptide. Peptides Pa-4, Pa-5, and Pa-6 permit placing Gln-Leu immediately following Ser-Ala-Leu, thus closing the other gap. Residue 73 is glutamine as much as Peptide Pa-4 was neutral. Furthermore, Peptide Pa-2 was ninhydrin-negative and thus was obtained as pyrroldnoneacetylglutamine, explaining why this peptide was previously overlooked in the chymotryptic digest of the entire protein as well as of Peptide A-T-XXXII (9).

**DISCUSSION**

It is now evident that the earlier transposition of two sections of the papain sequence (11) was due in part to the existence of certain repetitive dipeptide sequences, three with Ile-Lys and two with Ile-Arg. Our isolation of the unique peptide Ile-Lys-Arg, residues 189 through 191, thus serves to confirm the interpretation derived from the x-ray analysis by Drenth et al. (12). This tripeptide was not previously obtained and its isolation was rendered possible by the availability of the maleylation method (17) which blocks tryptic action at lysine residues. The sequence Ile-Lys-Arg must be preceded by an arginyl residue to explain its origin from the maleylated protein. The only sequences which fit this requirement are from Peptide D-C-XXXII (7), Ile-Arg (Ile,Lys), residues 187 through 190, and Peptide A-H-69b (10), Gly-$\gamma$-Tyr (Ile,Arg,Ile,Lys), residues 185 through 190. These peptides are derived from the sequence immediately preceding tryptic Peptide A-T-XXXIV (9), and the overlap is provided by Ile-Lys-Arg, described in this paper.

The maleylation method, which allowed preparation of a soluble, inactive derivative of papain, has also permitted facile isolation of Peptide Tn-2 containing 33 residues which overlaps with a portion of the NH$_2$-terminus and contained the residues necessary to complete this portion of the sequence. Our results on this sequence are in accord with those obtained by Husain and Lowe (13) who used different methods. Since these investigators used only acidic hydrolysis of the peptide, they could not determine the nature of residue 35, which we could clearly show is a glutamyl residue since enzymatic hydrolysis was used.

With the clarification of the nature of the 2 missing residues, Nos. 73 and 74 (Fig. 1), and the location of missing valine residue, No. 133, we have confirmed the number of residues assigned to the papain molecule by Drenth et al. (12).

It is perhaps useful to assess some of the reasons why, in earlier work, the amino acid analysis failed to indicate the correct number of total residues in the papain molecule. Although it

---

3 It was previously indicated (9) that Peptide XXXII was studied in a limited way and that the composition of the aminoterminal part should be considered as tentative. It is now recognized that the analysis was low by 1 residue each in serine and tyrosine but correctly indicated the presence of 2 residues each of leucine and glutamic acid.

4 We are grateful to Drs. Evans and Botes for allowing us to include results of their studies in this paper.
Amino acid composition of papain

These values are based on the sequence shown in Fig. 1.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No.</th>
<th>Amino acid</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysine</td>
<td>10</td>
<td>Glycine</td>
<td>28</td>
</tr>
<tr>
<td>Histidine</td>
<td>2</td>
<td>Alanine</td>
<td>14</td>
</tr>
<tr>
<td>Arginine</td>
<td>12</td>
<td>Valine</td>
<td>18</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>7</td>
<td>Isoleucine</td>
<td>12</td>
</tr>
<tr>
<td>Asparagine</td>
<td>12</td>
<td>Leucine</td>
<td>11</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>8</td>
<td>Tyrosine</td>
<td>19</td>
</tr>
<tr>
<td>Glutamine</td>
<td>12</td>
<td>Phenylalanine</td>
<td>4</td>
</tr>
<tr>
<td>Threonine</td>
<td>8</td>
<td>Tryptophan</td>
<td>5</td>
</tr>
<tr>
<td>Serine</td>
<td>13</td>
<td>Cysteine</td>
<td>1</td>
</tr>
<tr>
<td>Proline</td>
<td>10</td>
<td>Half-cystine</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>212</td>
</tr>
</tbody>
</table>

Acknowledgments—We are happy to acknowledge the assistance of Dorothy McNall for the amino acid analyses. We thank Dr. T. Cayle of the Wallerstein Laboratories for the dried papaya latex used for the preparation of the crystalline enzyme.

REFERENCES


In Table IV, we present the complete amino acid composition of papain, including the assignment of the amides, since this information has not been presented elsewhere. The calculated partial specific volume from this composition is 0.723. Inasmuch as earlier reported constants for many studies on papain were based on a lower molecular weight, such values should be corrected for a calculated value of 23,406, based on the sequence of the 212 residues shown in Fig. 1.
The Complete Amino Acid Sequence of Papain: ADDITIONS AND CORRECTIONS
Ronald E. J. Mitchel, Irwin M. Chaiken and Emil L. Smith


Access the most updated version of this article at http://www.jbc.org/content/245/14/3485

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/245/14/3485.full.html#ref-list-1