with 4.4 g. of N^{α},N^{ω}-dicarbobenzoxy-L-arginine and 2.2 g. of N,N-dicyclohexylcarbodiimide.[12] After the resulting suspension has been stirred at room temperature for about 8 hr., 1.5 ml. of water is added and the stirring continued for an additional 30 minutes. The white precipitate of dicyclohexylurea (2.2 g.) is removed by filtration, the filtrate evaporated to dryness and the residue treated with ethanol and evaporated to dryness. Upon trituration of the residual material with a small amount of cold ethanol, 3.6 g. of the product is obtained; m.p. 146–147°, unchanged by mixed m.p. with the material prepared by method A above; $[\alpha]^{25}D$ −12.2° (4% in chloroform).

Method C.—Four and four-tenths grams of dry, finely pulverized N^{α},N^{ω}-dicarbobenzoxy-L-arginine is dissolved in a mixture of 40 ml. of anhydrous chloroform and 1.4 ml. of anhydrous triethylamine, with shaking and occasional heating at 40–50°. The resulting solution is cooled to 0°, treated with 1.0 ml. of ethyl chloroformate and after storage at 0° for 5 min., treated with 1.4 ml. of triethylamine.[13] After additional storage for 30 min. at room temperature, the reaction mixture is washed twice with water and twice with dilute acetic acid, dried over anhydrous sodium sulfate and concentrated to dryness. Removal of the final traces of chloroform is achieved by the addition of a small amount of ethanol to the residue followed by a further concentration to dryness. The crystalline residue is triturated with a small amount of ethanol and filtered over suction; yield 3.6 g. After recrystallization from ethanol, the yield is 3.0 g.; m.p. 146–147°, unchanged by mixed m.p. with the products obtained by methods A and B above; $[\alpha]^{25}D$ −12.3° (4% in chloroform).

(12) The *p*-toluenesulfonic acid is here employed to form a salt with the N^{α},N^{ω}-dicarbobenzoxyarginine and thus not only permit the solution of this latter material in dioxane, wherein it is generally insoluble, but also facilitate its reaction with the dicyclohexylcarbodiimide reagent.

(13) As the addition of ethyl glycinate *in lieu* of triethylamine leads to the same yield of N^{α},N^{ω}-dicarbobenzoxyanhydroarginine, in the absence of detectable acyldipeptide ethyl ester formation, it may be concluded that the mixed anhydride formed by the interaction of ethyl chloroformate with N^{α},N^{ω}-dicarbobenzoxyarginine undergoes immediate intramolecular reaction with the nitrogen atom of the guanidino moiety.

[CONTRIBUTION FROM THE VIRUS LABORATORY, UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA]

Studies on the Amino Acid Sequence of Tobacco Mosaic Virus (TMV) Protein. IV. The Amino Acid Sequences of an Eicosapeptide and a Heptadecapeptide Isolated from a Tryptic Digest of TMV Protein[1]

BY DUANE T. GISH[2]

RECEIVED FEBRUARY 10, 1961

The amino acid sequence of an eicosapeptide, ileu·ileu·gluNH_2·val·glu·aspNH_2·gluNH_2·ala·aspNH_2·pro·thr·thr·ala·glu·thr·leu·asp·ala·thr·arg, representing positions 93 through 112 of TMV protein, and the amino acid sequence of a heptadecapeptide, ser·ser·phe·glu·ser·ser·ser·gly·leu·val·try·thr·ser·gly·pro·ala·thr, representing the C-terminal portion, or positions 142 through 158, of TMV protein, have been determined.

Previous papers from this Laboratory[3–9] have reported the results of amino acid sequence studies on TMV protein. Based upon this work and, as yet, unpublished work, plus a few sequences tentatively accepted from the literature,[10] a complete amino acid sequence for TMV protein was proposed.[11] Anderer, *et al.*,[10] have proposed a nearly complete structure for TMV protein. At the time of the publication of their paper, Anderer and coworkers had not established the question of the amidation of 18 of the aspartic and glutamic acid residues and had not established the sequences of the amino acids occupying positions 26 and 27 nor those occupying positions 98 through 102 of the protein chain. Work was proceeding concurrently in Berkeley, and we wish to report the sequence of an eicosapeptide, the amino acids of which occupy positions 93 through 112 of the protein, and the sequence of a heptadecapeptide, the amino acids of which occupy positions 142 through 158, or the C-terminal portion of the protein. These peptides were isolated from a tryptic digest of the protein. The amino acid sequence of the C-terminal peptide has in part been previously published[8] and *in toto* referred to[12] and has been confirmed by Anderer, *et al.*[10]

The eicosapeptide, designated DCA-I-3 or Peptide 8 in a previous paper,[11] will be referred to in this paper as Peptide 8, and the heptadecapeptide, referred to in previous papers as the C-terminal peptide,[8,12] will be referred to as Peptide 12.

Peptides 8 and 12 were found in the fraction insoluble in the 2-butanol–0.1 *M* dichloroacetic acid system used for the countercurrent distribution of the tryptic digest of TMV protein.[3] These two components were separated by countercurrent distribution in the system 1-butanol–pyridine–0.1% acetic acid (5:3.5:12).[13] The distribution pattern after 1000 transfers is shown in Fig. 1. Although these peptides could not be recovered quantitatively from a tryptic digest of the protein due to some coprecipitation during the isolation of Peptide 1, or "I"-peptide,[3] and to the fact that Peptide 12 contains a tryptophyl–threonine bond very sensitive to traces of chymotryptic activity in the trypsin, they were recovered in quite sub-

(1) This paper has been aided by a U. S. Public Health Service Grant.

(2) The Upjohn Co., Kalamazoo, Michigan.

(3) D. T. Gish, L. K. Ramachandran and W. M. Stanley, *Arch. Biochem. Biophys.*, **78**, 433 (1958).

(4) L. K. Ramachandran and D. T. Gish, *J. Am. Chem. Soc.*, **81**, 884 (1959).

(5) K. Narita, *Biochim. Biophys. Acta*, **28**, 184 (1958).

(6) A. Tsugita, *ibid.*, **38**, 145 (1960).

(7) D. T. Gish, *ibid.*, **35**, 557 (1959).

(8) D. T. Gish, *Biochem. Biophys. Res. Communs.*, **1**, 67 (1959).

(9) D. T. Gish, *J. Am. Chem. Soc.*, **82**, 6329 (1960).

(10) F. A. Anderer, H. Uhlig, E. Weber and G. Schramm, *Nature*, **186**, 922 (1960).

(11) A. Tsugita, D. T. Gish, J. Young, H. Fraenkel-Conrat, C. A. Knight and W. M. Stanley, *Proc. Natl. Acad. Sci.*, **46**, 1463 (1960).

(12) A. Tsugita and H. Fraenkel-Conrat, *ibid.*, **46**, 636 (1960).

(13) H. Rasmussen and L. C. Craig, *J. Am. Chem. Soc.*, **81**, 5003 (1959).

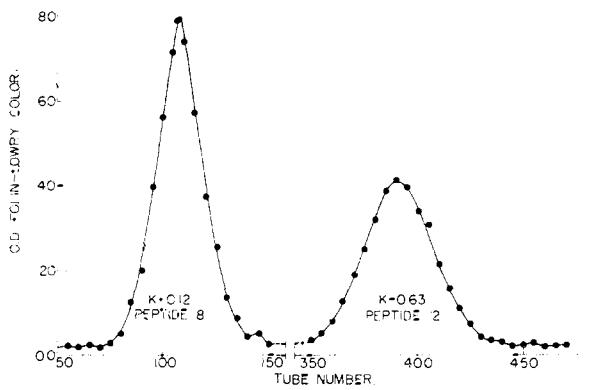


Fig. 1.—Countercurrent distribution of Peptides 8 and 12 in 1-butanol–pyridine–0.1% acetic acid (5:3.5:12); 1000 transfers.

stantial yields which left no doubt that they represented portions of the protein chain.

Peptide 8 was found to contain 20 amino acid residues rather than 19 as reported by Anderer, *et al.*[10] They had reported the sequence of residues 93 through 96 to be ileu·glu·val·glu (the asterisk indicates that the presence or absence of an amide group is unknown). The present work has proven the sequence to be ileu·ileu·gluNH$_2$·val·glu. This was shown conclusively by the amino acid composition of a 72-hr. hydrolyzate and by Edman degradation[14] of this pentapeptide isolated from a subtilisin digest of Peptide 8. The finding of an additional isoleucine residue increased the total number of residues found in TMV protein to 158, as reported by the Berkeley group,[11] rather than 157 as proposed by Wittmann and Braunitzer[15] and by Anderer, *et al.*[10] This also confirmed the analytical data of Ramachandran[16] and Tsugita and Fraenkel-Conrat,[12] who reported that TMV protein contained 9 isoleucine residues rather than 8 as proposed by the German workers. The resistance of the isoleucyl–isoleucine bond in this peptide to acid hydrolysis may account for the fact that it was missed by Wittmann and Braunitzer[15] and by Anderer, *et al.*[10] Indeed, as shown in Table I, the amino acid composition of a 16-hr. hydrolyzate of Peptide 8 according to the dinitrophenyl (DNP) method[17] revealed approximately two moles of the leucines. The N-terminal isoleucine was split off in the dipeptide isoleucyl–isoleucine, the DNP-derivative of which chromatographed in the Levy system practically identically to isoleucine, and thus was determined as isoleucine. In addition to the N-terminal isoleucine, one mole of leucine was found. Only after Edman degradation of the N-terminal pentapeptide derived from Peptide 8, as previously mentioned, was the second isoleucine found.

Treatment of Peptide 8 with a mixture of carboxypeptidases A and B indicated a C-terminal sequence of . . . . .(glu, thr, leu) asp·ala·thr·arg.

(14) P. Edman, *Acta Chem. Scand.*, **4**, 277 (1950).
(15) H. G. Wittmann and G. Braunitzer, *Virology*, **9**, 726 (1959).
(16) L. K. Ramachandran, *ibid.*, **5**, 244 (1958).
(17) A. L. Levy, *Nature*, **174**, 126 (1954).

TABLE I

AMINO ACID COMPOSITION OF PEPTIDE 8

Amino acid	Molar ratios	Assumed no. of residues
thr	4.11	4
pro	1.00	1
ala	2.75	3
val	0.90	1
leu-ileu[a]	1.65	2
asp	3.34	3
glu	4.37	4
arg	0.75	1

[a] Ileu was shown to be N-terminal. Correct total for the leucines was 3 (see text).

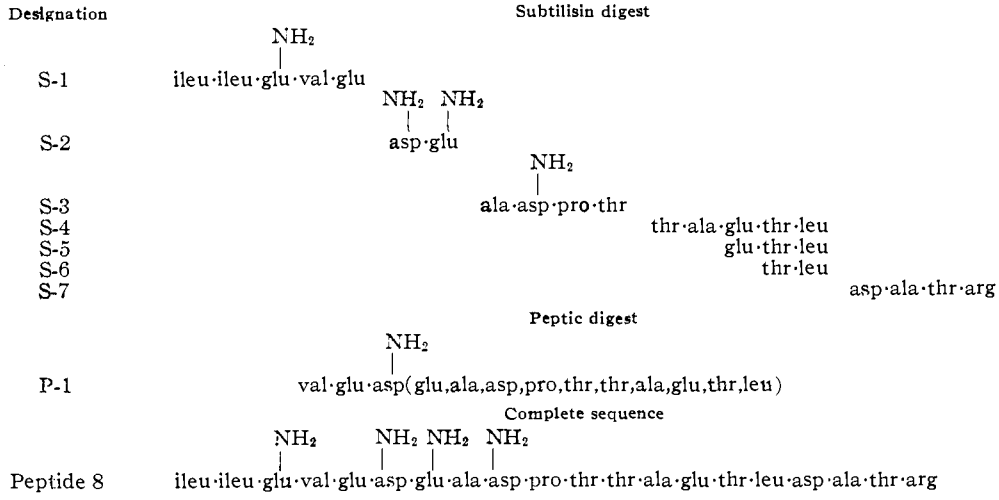
The peptide was resistant to leucine aminopeptidase. The peptide was hydrolyzed with subtilisin and the peptides so obtained were isolated and sequences determined. The sequences found are shown in Table II. From the fact that isoleucine was shown to be N-terminal, the results of the treatment of the peptide with the carboxypeptidases, and the sequences of the peptides from the subtilisin digest, the sequence of Peptide 8 could be fixed except for the relative positions of the aspNH$_2$·gluNH$_2$ (S-2) and the ala·aspNH$_2$·pro·thr (S-3) peptides. The peptide P-1 (Table II) was isolated from a peptic digest of Peptide 8 and degraded by the Edman technique for a sufficient number of steps to establish the relative positions of S-2 and S-3. The results of the sequence studies permitted the postulation of the sequence of Peptide 8 (positions 93–112 of TMV protein) to be ileu·ileu·gluNH$_2$·val·glu·aspNH$_2$·gluNH$_2$·ala·aspNH$_2$·pro·thr·thr·ala·glu·thr·leu·asp·ala·thr·arg.

The amino acid composition of Peptide 12 by the DNP method is shown in Table III. Serine was shown to be N-terminal. It may be noted that 6 of the 16 serine residues of TMV protein are found in the C-terminal heptadecapeptide. Five of the remaining serine residues are found within the first 15 amino acid residues from the N-terminal end of TMV protein.[10,11] Thus 11 of the 16 serine residues of TMV protein are concentrated at the N-terminal and C-terminal portions of the molecule. Whether this feature of the structure of the TMV protein subunit contributes importantly to the ability of the subunits to polymerize to the protein rod or to other physical and chemical properties of the virus remains to be shown.

When Peptide 12 was treated with carboxypeptidase A, threonine, and only threonine, was released. Harris and Knight[20] had shown that threonine, and only threonine, is split from TMV by carboxypeptidase A. Niu and Fraenkel-Conrat[21] had isolated the peptide thr·ser·gly·pro·ala·thr from a chymotryptic digest of TMV protein and had shown that this peptide was derived from the C-terminal end of TMV protein. The fact that the heptadecapeptide reported in this paper had no C-terminal basic amino acid, although isolated from a tryptic digest, and that its behavior toward carboxypeptidase and its sequence at the C-terminal end (as described below) is the same as that found for TMV protein established that it represents the

TABLE II

SEQUENCES OF PEPTIDES FROM PEPTIDE 8



sequence of the 17 amino acids at the C-terminal end of TMV protein.

Treatment of Peptide 12 with N-bromoacetamide, which splits tryptophyl peptide bonds,[22] liberated a peptide with N-terminal threonine. thus demonstrating the presence of a tryptophyl–threonine bond in the peptide. Hydrolysis of Peptide 12 with leucine aminopeptidase released the N-terminal serine plus a second mole of serine, one mole of phenylalanine and 0.6 mole of glutamic acid. Peptide 12 was hydrolyzed with chymotrypsin and the products were separated and their amino acid sequences were determined. The results are shown in Table IV. The establishment of the sequences of these peptides, the presence of an N-terminal serine, a C-terminal threonine and a tryptophyl-threonine bond in Peptide 12 permitted the postulation of the structure shown in Table IV.

TABLE III

AMINO ACID COMPOSITION OF PEPTIDE 12

Amino acid	Molar ratios	Assumed no. of residues
ser	0.43[a] + 4.96	6
leu	1.07	1
phe	1.00	1
gly	1.98	2
val	0.92	1
pro	1.15	1
thr	1.85	2
ala	0.96	1
glu	1.01	1
try[b]	1.0	1

[a] Value for N-terminal residue. [b] Determined spectrophotometrically[18] and colorimetrically[19] before acid hydrolysis.

Methods and Materials

Enzyme Preparations.—The leucine aminopeptidase, carboxypeptidase A, chymotrypsin and pepsin used were those preparations previously described.[4] The carboxypeptidase B was prepared by Dr. L. K. Ramachandran according to the procedure of Folk and Gladner.[23] The subtilisin was a commercial preparation of Al-proteinase, 20×10^4 PUN, from Nagase, Osaka, Japan. All digestions were carried out at 37°. The buffer used with leucine aminopeptidase was a *p*H 8.55 0.075 *M* veronal buffer containing $MgCl_2$ at 0.05 *M* and that used with the carboxypeptidases was a *p*H 8 0.02 *M* $NaHCO_3$ buffer containing NaCl at 0.1 *M*.

N-Terminal and Amino Acid Analyses.—The fluorodinitrobenzene reagent[24] was used for the determination of N-terminal amino acids and for the amino acid analyses. The general procedure described by Fraenkel–Conrat, Harris and Levy[25] was used with modifications previously described.[4] Since correction factors for the recovery of amino acids from peptides as their DNP derivatives are not available, in general the molar ratios of the amino acids were based directly on the ratio of the optical densities of the DNP amino acids. The recovery of the N-terminal amino acid as its DNP derivative is quite variable and the molar ratio of the N-terminal amino acid based on the recovery of its DNP derivative is generally 0.4–0.8. Well known exceptions are glycine, which usually is recovered in an amount about 10–20% of that of the other amino acids, and proline, which is completely lost. Also, it was noted that when a dinitrophenylated peptide along with the by-products of the reaction, was hydrolyzed (generally hydrolyses were carried out in constant boiling HCl at about 105° for 16 hr.), small but variable amounts of artifacts chromatographing as the DNP derivatives of glutamic acid, glycine, serine and lysine (di-substituted) were found. These artifacts apparently arise as the result of the presence, in the hydrolysis reaction mixture, of a by-product of the dinitrophenylation, probably dinitrophenol, since no such artifacts were noted when the DNP peptide was purified before hydrolysis. Tsugita also has reported[26] the presence in the hydrolyzate of a DNP peptide of artifacts chromatographing as glutamic acid, glycine and serine when the amino acids, after removal of the N-terminal DNP amino acid, were chromatographed on an ion-exchange column. For the reasons cited above, two analyses generally were performed on each peptide. One analysis, in which dinitrophenylation was carried out both before and after acid hydrolysis, was performed to establish the N-terminal amino acid, and a second analysis, in which only the amino acid hydrolyzate of the peptide was dinitrophenylated, was performed to establish the amino acid composition. In order to distinguish between leucine and

(18) T. W. Goodwin and R. A. Morton, *Biochem. J.*, **40**, 628 (1946).
(19) J. R. Spies and D. C. Chambers, *Anal. Chem.*, **21**, 1249 (1949).
(20) J. I. Harris and C. A. Knight, *Nature*, **170**, 613 (1952).
(21) C. I. Niu and H. Fraenkel-Conrat, *Arch. Biochem. Biophys.*, **59**, 538 (1955).
(22) L. K. Ramachandran and B. Witkop, *J. Am. Chem. Soc.*, **81**, 4028 (1959).
(23) J. E. Folk and J. A. Gladner, *J. Biol. Chem.*, **231**, 379 (1958).
(24) F. Sanger, *Biochem. J.*, **39**, 507 (1945).
(25) H. Fraenkel-Conrat, J. I. Harris and A. L. Levy, in "Methods of Biochemical Analysis," Vol. II, Ed. D. Glick, Interscience Publishers, Inc., New York, N. Y., 1954, p. 359.
(26) A. Tsugita, personal communication.

TABLE IV

SEQUENCES OF PEPTIDES FROM PEPTIDE 12

Designation	Sequence
C-1	ser·ser·phe
C-2	ser·ser·phe·glu·ser ser·ser·gly·leu
C-3	glu·ser·ser·ser·gly·leu
C-4	val·try
C-5	thr·ser·gly·pro·ala·thr
Peptide 12	ser·ser·phe·glu·ser·ser·ser·gly·leu·val·try·thr·ser·gly·pro·ala·thr

isoleucine, whose DNP derivatives are not separated by the Levy solvent system,[17] the DNP derivatives of those amino acids were hydrolyzed in concentrated NH_4OH at 105° for 2 hr. and the liberated amino acids were chromatographed on paper in the 2-butanol–3% NH_4OH (3:1) system.[27]

Chromatography and Electrophoresis of Peptides.—The solvent system used for paper chromatography of peptides was the 1-butanol–acetic acid–water (4:1:1) system. Paper electrophoresis was performed on the Durrum-type[28] electrophoresis apparatus of the Spinco Division of Beckman Instruments, Inc. Whatman 3 MM paper was used and for preparative scale separations paper up to 8″ width was employed. In preparative separations for both chromatography and electrophoresis, guide strips were cut from each edge and stained to locate the zones. In addition to the ninhydrin reagent, the Sakaguchi test for arginine according to Acher and Crocker[29] and the *p*-dimethylaminobenzaldehyde test for tryptophan[30] were employed.

The apparatus described by Paigen[31] was used for the electrophoretic separations on cellulose blocks. The cellulose, Whatman standard grade, was washed twice with water and twice with buffer before use. The apparatus was kept in a refrigerated box at about 5° during a run. The buffers employed, both for paper and cellulose block electrophoresis, were a pyridine–acetic acid buffer of *p*H 5.5 (24.3 ml. of pyridine and 5.7 ml. of glacial acetic acid per liter) and 33⅓% acetic acid, *p*H 1.9.

Purification and Elucidation of the Structures of Peptides.—In the discussions below, as partial sequences are disclosed, these are indicated in brackets. In accordance with common usage, in known sequences the amino acid designations are separated by dots, and in unknown sequences the amino acid designations are enclosed in parentheses and are separated by commas.

Countercurrent Distribution of Peptides 8 and 12.—About 5.25 g. (approximately 375 μmole, taking into account the prior removal of Peptide 1 or I- peptide) of the fraction of a tryptic digest of TMV protein soluble at *p*H 4.7[3] was extracted with about 300 ml. each of upper and lower phases of the 2-butanol–0.1 *M* dichloroacetic acid solvent system used for the countercurrent distribution (CCD) of the tryptic digest soluble in this system.[3] The insoluble residue was removed by centrifugation and washed twice with about 25 ml. of each phase. The insoluble material then was dissolved in water by adjusting the *p*H to 9, the solution was filtered and the material was precipitated by adjusting the *p*H to 2.5. This procedure was repeated and the material was washed with water and then dried by lyophilization. The material weighed 570 mg. The amount of Peptide 12 in this mixture, indicated by its tryptophan content as measured spectrophotometrically,[18] amounted to about 125 μmole. Much of the remainder of Peptide 12 probably precipitated with Peptide 1 at *p*H 4.7 and some (about 30 μmole) was found as a chymotryptic split product in the fraction soluble in the CCD solvent. The amount of Peptide 8 present in this mixture according to its arginine content as measured by the Sakaguchi test[32] was about 200 μmole.

For the CCD of the above material in the 1-butanol–pyridine–0.1% acetic acid (5:3.5:12) system[13] the material was extracted with a total of about 125 ml. each of upper and lower phases. Material which failed to dissolve was removed by centrifugation, suspended in water and dried by lyophilization. This material weighed 213 mg. (about 100 μmole) and according to its amino acid composition was practically pure Peptide 8. The soluble fraction was placed in a 500-tube automatic CCD apparatus[33] and distributed for 1000 transfers. The distribution curve as measured by the Folin–Lowry test[34] is shown in Fig. 1. The solution in tubes 355–430 (Peptide 12, K = 0.63) was removed, concentrated on a rotary evaporator and lyophilized. It weighed 201 mg. (118 μmoles). The distribution was continued with the material remaining in the machine (Peptide 8) for a total of 3600 transfers. The material (K = 0.12), according to its distribution curve, appeared to be homogeneous. It was recovered in the same manner as Peptide 12; weight, 193 mg. (about 80 μmoles).

Peptide 8.—The amino acid composition of Peptide 8 isolated by CCD is shown in Table I. The analysis was performed according to the DNP method on a 16 hr. hydrolyzate. As previously mentioned, only a fraction of the isoleucyl–isoleucine bond was split after 16 hr. and thus only about two moles of leucine–isoleucine was recovered, although later the peptide was shown to contain two isoleucine residues and one leucine residue. Isoleucine was shown to be N-terminal.

About 2.0 mg. of Peptide 8 was dissolved in 0.4 ml. of buffer and treated with 50 μl. of carboxypeptidase B and 200 μg. of carboxypeptidase A for 40 hr. The molar ratios of the amino acids released are shown in Table V. A partial sequence at the C-terminal end of peptide 8 indicated by these data (only arginine was released when the peptide was treated with carboxypeptidase B alone) was.....(glu, thr,leu) asp·ala·thr·arg. [*ileu* (*glu*, asp_2, thr_2, ala_2, *pro*, *val*) (*glu*, *thr*, *leu*) *asp·ala·thr·arg*].

TABLE V

MOLAR RATIOS OF AMINO ACIDS RELEASED FROM PEPTIDE 8 BY 40-HR. TREATMENT WITH CARBOXYPEPTIDASES A AND B

Amino acid	Molar ratios
arg	1.00
thr	1.32
ala	0.60
asp	.47
leu	.42
glu	.37

About 100 mg. of Peptide 8 was dissolved in 100 ml. of H_2O at *p*H 7 and the peptide was hydrolyzed with 1.7 mg. of subtilisin. The *p*H was maintained at 7 by intermittent addition of alkali, and the digestion was allowed to proceed for 24 hr. The digestion was followed by determination of N-terminal amino acids by the DNP method on aliquots withdrawn at various time intervals. The molar amounts of new N-terminal amino acid residues released by the subtilisin were: aspartic acid (or asparagine), 2; threonine, 1.5; alanine, 1; glutamic acid, fraction of a mole. The enzymic products were recovered by lyophilization.

The subtilisin digest was subjected to CCD in the system 1-butanol–0.5% acetic acid. The distribution was followed using the ninhydrin reaction.[35] After 1305 transfers, the material had separated into peaks with distribution coefficients of 0.02, 0.09, 0.15 and 0.36. The material in each

(27) J. F. Roland and A. M. Gross, *Anal. Chem.*, **26**, 502 (1954).

(28) F. G. Williams, Jr., E. G. Pickles and E. L. Durrum, *Science*, **121**, 829 (1955).

(29) R. Acher and C. Crocker, *Biochim. Biophys. Acta*, **9**, 704 (1952).

(30) I. Smith, *Nature*, **171**, 43 (1953).

(31) K. Paigen, *Anal. Chem.*, **28**, 284 (1956).

(32) H. Rosenberg, A. H. Ennor and J. F. Morrison, *Biochem. J.*, **63**, 153 (1956).

(33) L. C. Craig, W. Hausman, E. H. Ahrens, Jr., and E. J. Harfenist, *Anal. Chem.*, **23**, 1236 (1951).

(34) O. H. Lowry, N. J. Rosebrough, A. L. Farr and R. J. Randall, *J. Biol. Chem.*, **193**, 265 (1951).

(35) S. Moore and W. H. Stein, *ibid.*, **211**, 907 (1954).

fraction was recovered by separating the phases, adding a volume of ether to the upper phase and extracting with a volume of water in three portions. The aqueous phase and aqueous extracts from each fraction were combined, concentrated on a rotary evaporator and lyophilized.

The peptide fractions were purified by electrophoresis on cellulose or by paper chromatography. The methods and results for each fraction are summarized in Table VI. It

TABLE VI

SEPARATION AND PURIFICATION OF PEPTIDES FROM THE SUBTILISIN DIGEST OF PEPTIDE 8

CCD fraction	Peptide	Method of purification
K—0.02	S-2, S-7	P.C.
	S-3	
K—0.09	S-4	Elect., pH 5.5
K—0.15	S-5	Elect., pH 5.5
	S-6	
K—0.36	S-1	Elect., pH 5.5

will be noted that the CCD fraction with a K of 0.02 was separated into two fractions, peptide S-3 and a mixture of S-2 and S-7. The peptides S-2 and S-7 failed to separate in several systems tried, including paper chromatography and electrophoresis. The structures of these two peptides were resolved as described below without separation. The fraction with a K of 0.15 was also separated into two fractions, peptides S-5 and S-6. The amino acid compositions of the peptides are shown in Table VII.

of K = 0.02 indicated strongly that it was a mixture of two peptides, asp·ala·thr·arg and asp·glu, since treatment of Peptide 8 with a mixture of the carboxypeptidases, as previously related, had indicated a C-terminal sequence of (glu,thr,leu) asp·ala·thr·arg. About 0.5 μmole of this mixture in 0.3 ml. of buffer was treated with about 0.2 C_1 unit of leucine aminopeptidase for 16 hr. and the liberated amino acids were determined. The results were: $aspNH_2$, 0.93, $gluNH_2$, 1.00, and traces of asp, ala, thr and arg. The enzyme apparently split the peptide $aspNH_2$·$gluNH_2$ completely while splitting a very small amount of the peptide asp·ala·thr·arg. The peptide mixture was subjected to a one-step Edman degradation. After extraction of the PTH's for chromatography, the residual material was dinitrophenylated and DNP amino acids were extracted from DNP peptide material (DNP-ala·thr·arg) with ether and chromatographed. Chromatography of the PTH's in Sjöquist's solvent C revealed the presence, in approximately equal amounts, of the PTH's of aspartic acid and asparagine. Chromatography of the DNP amino acids revealed the presence of glutamine and glutamic acid in the ratio $gluNH_2$, 1.00, glu, 0.30, indicating that a fraction of the glutamine had been converted to glutamic acid during the treatment with 3 N HCl. The above results, along with results previously mentioned, permitted the unequivocal conclusion that the material was a mixture of the two peptides, $aspNH_2$·$gluNH_2$(S-2) and asp·ala·thr·arg (S-7) [*ileu·ileu·gluNH₂·val·glu(aspNH₂·gluNH₂,ala₂,asp,thr₂,pro)(glu,thr,leu)asp·ala·thr·arg*].

About 0.08 μmole of peptide S-3 in 0.2 ml. of buffer was treated with 0.1 C_1 unit of leucine aminopeptidase for 20

TABLE VII

AMINO ACID COMPOSITION OF SUBTILISIN-SPLIT PEPTIDES FROM PEPTIDE 8

	S-1		S-2, S-7	S-3	S-4	S-5	S-6
Amino acid	16 hr. Hydrol.	72 hr. Hydrol.[a]					
ileu	0.90[b] + 0.31	1.76					
val	1.00	1.00					
glu	1.84	1.94	0.83		0.96	0.90[b]	
asp			2.04[b]	0.85			
ala			1.08	.55[b]	0.99		
pro				.80			
thr			1.00	1.00	0.92[b] + 1.04	1.00	1.00[b]
leu					1.00	0.99	1.01
arg			0.74				

[a] Analyzed by ion-exchange chromatography.[37] [b] N-Terminal amino acid.

Peptide S-1 was degraded by the Edman technique[14] according to the method described by Fraenkel-Conrat, Harris and Levy.[25] The subtractive method was used, in which the amino acid composition of the residual peptide after each step was determined. The analyses included the determination of the N-terminal amino acid. The results are recorded in Table VIII. The fact that the N-terminal isoleucine residue had been split off at step 1 was confirmed by chromatography of the first phenylthiohydantoin (PTH) in Sjöquist's solvent A.[36] A single spot, corresponding to the PTH of isoleucine, was found. The chromatography of the 3rd PTH in the same system revealed the presence of the PTH of glutamine, along with some PTH of glutamic acid, indicating that the 3rd residue was glutamine rather than glutamic acid. This was confirmed, as well as the fact that the C-terminal residue of S-1 was glutamic acid rather than glutamine, by treating the DNP derivative of S-1 with carboxypeptidase A. After 24 hr. incubation of 0.2 mg. of the DNP derivative with 100 μg. of the enzyme, the liberated amino acids were determined. The molar ratios found were: glu, 1.00, val, 0.18, and $gluNH_2$, 0.15. The above results established the sequence of S-1 to be ielu.ileu.glu NH_2·val. glu, allowing postulation of the partial sequence [*ileu·ileu·gluNH₂.val·glu (glu, asp₂, thr₂, ala₂, pro)(glu, thr, leu) asp·ala·thr·arg*].

The amino acid composition of one of the fractions (S-2, S-7) separated by paper chromatography from the material hr. and the amino acids released were determined. Found: ala, 1.00, $aspNH_2$, 0.08. Peptide S-3 then was subjected to hydrazinolysis.[38] About 0.15 μmole of the peptide was heated with 0.5 ml. of 95% hydrazine at 100° in a sealed tube for 10 hr. After removal of the excess hydrazine, the mixture was applied directly[11] to the ion-exchange column[37] and the C-terminal amino acid was determined. Found: thr, 0.11 μmole. The above results established the sequence of S-3 to be ala·$aspNH_2$·pro·thr. [*ileu·ileu·gluNH₂·val·glu (aspNH₂·gluNH₂, ala·aspNH₂·pro·thr) (thr, ala) (glu, leu) asp·ala·thr·arg*].

TABLE VIII

AMINO ACID COMPOSITIONS AND N-TERMINAL RESIDUES OF PEPTIDES FROM THE EDMAN DEGRADATION OF PEPTIDE S-1

	DNP analysis after		
Amino acid	Step 1	Step 2	Step 3
ileu	0.78[a]		
glu	2.00	0.68[a] + 1.00	1.00
val	1.00	0.70	0.46[a]

[a] N-Terminal amino acid.

Peptide S-4 was treated with carboxypeptidase A. A mole of leucine, and only leucine, was released. A sequence for S-4 of thr·ala·glu·thr·leu could thus be postulated, since the peptide S-6 was found to be thr·leu and the peptide S-5

(36) J. Sjöquist, *Acta Chem. Scand.*, **7**, 447 (1953).

(37) D. H. Spackman, W. H. Stein and S. Moore, *Anal. Chem.*, **30**, 1190 (1958).

(38) S. Akabori, K. Ohno and K. Narita, *Bull. Chem. Soc. Japan*, **25**, 214 (1952).

TABLE IX

AMINO ACID COMPOSITIONS OF PEPTIDES FROM THE CHYMOTRYPTIC DIGEST OF PEPTIDE 12

Amino acid	C-1 (K—0.25)	C-2 (K—0.29)	C-3 (K—0.18)	C-4[b] (K—0.68)	C-5 (K—0.06)
ser	0.54[a] + 1.00	0.39[a] + 4.30	2.61		1.00
phe	0.86	0.91			
glu		1.00	1.05[a]		
gly		1.01	0.97		1.15
leu		0.87	1.00		
val				1.00[a]	
try				0.70	
thr					0.97[a] + 1.00
pro					1.05
ala					0.88

[a] N-Terminal amino acid. [b] Only DNP-valine was recovered from an acid hydrolyzate. Hydrolysis with leucine aminopeptidase was used to obtain the composition shown.

was found to be glu (thr, leu) (see Table VII). The fact that glutamic acid was present rather than glutamine was confirmed by treating peptide S-5 with leucine aminopeptidase. About 0.3 mg. of the peptide in 0.2 ml. of buffer was digested with 0.1 C_1 unit of the enzyme for 28 hr. The amino acids released were glu, 1.00, thr, 0.07, leu, 0.62 [*ileu·ileu·gluNH₂·val·glu* (*aspNH₂·gluNH₂, ala·aspNH₂·pro·thr*) *thr·ala·glu·thr·leu·asp·ala·thr·arg*].

About 10 mg. of Peptide 8 was dissolved in 10 ml. of 0.01 M HCl, the pH was adjusted to 1.8 and the mixture was incubated with 0.5 mg. of pepsin for 16 hr. The solution was evaporated to dryness, and the peptides were chromatographed on Whatman 3MM paper (spread along a line 7" wide) in the system 1-butanol-acetic acid–water–pyridine (30:6:24:20).[39] The mixture was separated into five peptide zones. The peptide with an R_f of 0.34 (P-1) was found to have valine as its N-terminal amino acid and to have the composition val, 1.00, ala, 2.14, pro, 1.21, thr, 3.14, leu, 1.04, asp, 2.23, glu, 2.88. A partial sequence of val·glu (aspNH₂·gluNH₂, ala·aspNH₂·pro·thr)thr·ala·glu·thr·leu could be written based on results already obtained, and only the relative positions of aspNH₂·gluNH₂ (S-2) and ala·aspNH₂·pro.thr (S-3) needed to be fixed in order to establish the sequence of P-1 and thus establish the entire sequence of Peptide 8. Peptide P-1 was degraded by the Edman technique for 3 steps, and the amino acid composition of the residual peptide was determined. It was found to be asp, 1.19, glu, 2.08, thr, 2.44, ala, 1.86, pro, 1.00, leu, 0.85. This result indicated that one mole each of valine, glutamic acid and aspartic acid (or asparagine) had been removed as the PTH derivatives and thus fixed the sequence of P-1 to be val·glu·aspNH₂·gluNH₂·ala·aspNH₂·pro·thr·thr·ala·glu·thr·leu and the sequence of Peptide 8 to be ileu·ileu·gluNH₂·val·glu·aspNH₂·gluNH₂·ala·aspNH₂·pro·thr·thr·ala·glu·thr·leu·asp·ala·thr·arg.

Peptide 12.—The amino acid composition of Peptide 12 is shown in Table III. About 0.1 μmole of the peptide in 0.1 ml. of buffer was treated with 20 μg. of carboxypeptidase A for 18 hr. The liberated amino acids then were determined. One mole of threonine, and only threonine, was released. About 0.2 μmole of Peptide 12 in 1 ml. of 70% acetic acid was treated with N-bromoacetamide essentially as described by Ramachandran and Witkop.[22] The N-terminal amino acids of the peptide fragments were then determined. Found: ser, 1.00, thr, 0.10. Thus, in addition to serine, which is N-terminal in Peptide 12, about 0.1 mole of newly liberated N-terminal threonine was found. The low yield of N-terminal threonine may have been due either to a slow rate of splitting of the tryptophyl–threonine bond or to the sensitivity of the N-terminal threonine to the reagent. In any case, the presence of a tryptophyl–threonine bond in Peptide 12 had been established [*ser* (*ser₅, glu, gly₂, leu, phe, val, try·thr, pro, ala*) *thr*].

About 0.2 mg. of Peptide 12 in 0.2 ml. of buffer was treated with about 0.2 C_1 unit of leucine aminopeptidase for 18 hr. and the liberated amino acids were determined. Found: ser, 1.82, phe, 1.00 and glu, 0.63. This indicated a sequence of ser (ser, phe) glu at the N-terminal end of Peptide 12.

Peptide 12 was digested with chymotrypsin. For this purpose, 40 mg. (about 20 μmoles) of the peptide was dissolved in 20 ml. of water at pH 7.8, and the mixture was treated with 0.8 mg. of the enzyme for 24 hr. The pH was maintained at or near 7.8 by the intermittent addition of N NaOH. Paper chromatography of an aliquot of the digest revealed at least five major products and two minor products. The mixture was fractionated by CCD in the system 2-butanol–0.5% acetic acid. The distribution was followed using the Folin-Lowry reaction[34] and the spectrophotometric method for tryptophan.[18] After 570 transfers a peptide with a K of 0.06 and a tryptophan-containing peptide with a K of 0.68 had separated from the remainder of the material. The solvent containing each of those peptides was removed, concentrated on a rotary evaporator and lyophilized to recover the material. The distribution then was carried on to 975 transfers with the remainder of the material, at which point a peptide with a K of 0.18 had completely separated and peptides with K's of 0.25 and 0.29 had only partially separated. The peptide materials from these two fractions were recovered and the peptides with K's of 0.25 and 0.29 were separated by electrophoresis on cellulose in 33⅓% acetic acid. After 18 hr. at 500 v. the separation of these two peptides into pure components was obtained. Two minor components were detected. The amino acid compositions and N-terminal amino acids of the peptides from the chymotryptic digest are given in Table IX.

Peptide C-1 was treated with carboxypeptidase A. About 0.25 mg. of the peptide was dissolved in 0.2 ml. of buffer and treated with 50 μg. of the enzyme for 16 hr. and the liberated amino acids were determined. One mole of phenylalanine, and only phenylalanine was released. This established the sequence of C-1 to be ser·ser·phe [*ser·ser·phe·glu* (*ser₂, gly, leu*)(*val, try·thr, ser, gly, pro, ala*) *thr*].

From the amino acid composition of peptide C-2 and the results of the treatment of Peptide 12 with leucine aminopeptidase, a partial sequence for peptide C-2 could be written as ser·ser·phe·glu (ser₂, gly, leu). The complete sequence could be deduced after the sequence of C-3 was established. About 0.2 μmole of C-3 in 0.2 ml. of buffer was treated with 100 μg. of carboxypeptidase A for 24 hr. and the liberated amino acids were determined. One mole of leucine, and only leucine, was released, establishing leucine as the C-terminal amino acid of peptide C-3 and thus of peptide C-2. About 0.15 μmole of peptide C-3 in 0.2 ml. of buffer was treated with 0.35 C_1 unit of leucine aminopeptidase for 24 hr. and the liberated amino acids were determined. Found: glu, 1.00, ser, 3.07, gly, 0.48, and leu, 0.34. These results indicated that the enzyme released the N-terminal glutamic acid, followed by three moles of serine, and then about half of the remaining gly·leu was split. The above results established the sequence of C-3 to be glu·ser·ser·ser·gly·leu and that of C-2 to be ser·ser·phe·glu·ser·ser·ser·gly·leu. [*ser·ser·phe·glu·ser·ser·ser·gly·leu* (*val, try·thr, ser, gly, pro, ala*) *thr*].

The sequence of peptide C-4 was established by the determination of its N-terminal amino acid and its composition. The N-terminal amino acid was established by the DNP method, but since acid hydrolysis destroys tryptophan the peptide was hydrolyzed with leucine aminopeptidase to establish its composition. About 0.05 μmole of the peptide in 0.3 ml. of buffer was treated with 0.1 C_1 unit of the

(39) S. G. Waley and J. Watson, *Biochem. J.*, **55**, 328 (1953).

enzyme for 18 hr. and the liberated amino acids were determined. The results are shown in Table IX. The above results established C-4 to be val·try [*ser·ser·phe·glu·ser·ser·ser·gly·leu·val·try·thr* (*ser, gly, pro, ala*) *thr*].

From the work of Niu and Fraenkel-Conrat[21] already cited, the sequence of peptide C-5 could be assumed to be thr·ser·gly·pro·ala·thr. This sequence was confirmed. About 0.2 μmole of peptide C-5 in 0.2 ml. of buffer was treated with 0.35 C_l unit of leucine aminopeptidase for 20 hr. and the liberated amino acids were determined. The N-terminal threonine and one mole of serine were released. A partial sequence for peptide C-5 of thr·ser (gly, pro, ala) thr could thus be written. About 10 μmoles of C-5 in 10 ml. of buffer was treated with about 0.7 C_l unit of leucine aminopeptidase for 24 hr. About 0.35 C_l unit of the enzyme was added and the incubation was continued for an additional 24 hr. This treatment completely removed the N-terminal threonine and the adjacent serine. The residual peptide was purified first by paper chromatography on Whatman 3 MM paper and then by electrophoresis on the same type paper with 33⅓% acetic acid as buffer. The amino acid composition of the purified peptide was gly, 0.93, pro, 1.00, ala, 1.08, thr, 1.00. This peptide then was degraded two steps by the Edman technique and the amino acid composition of the residual peptide after each step was determined. Found: Step 1, gly, 0.34, pro, 0.89, ala, 1.00, thr, 1.05; step 2, gly, 0.20, pro, 0.18, ala, 1.00, and thr, 0.97. Although the cleavage at step 1 was only about 70% complete, these results conclusively showed that the sequence of the peptide was gly·pro·ala·thr (threonine was known to be C-terminal from the treatment of Peptide 12 with carboxypeptidase A). The sequence of peptide C-5 thus was established to be thr·ser·gly·pro·ala·thr, confirming the sequence postulated by Niu and Fraenkel-Conrat. The sequence of Peptide 12 therefore was established to be ser·ser·phe·glu·ser·ser·ser·gly·leu·val·try·thr·ser·gly·pro·ala·thr.

Acknowledgments.—The author is indebted to Miss Martha Langley and Miss Shirley Williams for valuable technical assistance and wishes to acknowledge gratefully the constant encouragement and interest of Dr. Wendell M. Stanley in this work and the helpful suggestions of Drs. Akira Tsugita, C. Arthur Knight and Heinz Fraenkel-Conrat, Berkeley, California.

[CONTRIBUTION FROM THE GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA]

Rearrangement of α-N-Acetyl-L-tyrosinhydrazide to 1-Acetyl-2-(L-tyrosyl)-hydrazine[1]

BY ABRAHAM N. KURTZ AND CARL NIEMANN[2]

RECEIVED MARCH 10, 1961

When an aqueous hydrochloric acid solution of α-N-acetyl-L-tyrosinhydrazide is heated, the preceding compound is rearranged, in good yield, to 1-acetyl-2-(L-tyrosyl)-hydrazine. Both the reaction and its product are of interest. The reaction offers promise of providing a new means for the identification of the carboxyl terminal α-amino acid residues of peptides and possibly for the stepwise degradation of peptides from their carboxyl terminal end. The product is the most effective readily reversible inhibitor of α-chymotrypsin discovered to date.

Aqueous solutions of α-N-acetyl-L-tyrosinhydrazide, a relatively water insoluble substrate of α-chymotrypsin,[3,4] are stable at 95–100° for periods up to and possibly exceeding 3 hr. However, when aqueous acidic solutions of the substrate were similarly treated a new and very water soluble species was formed. This latter substance was not a substrate but an inhibitor with an apparent enzyme–inhibitor dissociation constant at *p*H 7.9 smaller than any recorded previously.[5] It will be recalled that at *p*H 7.9 L-tyrosinhydrazide is a substrate of α-chymotrypsin,[4] that α-N-acetyl-L-tyrosine is a relatively poor inhibitor and that L-tyrosine is relatively insoluble in dilute aqueous acid.

An acidic solution of α-N-acetyl-L-tyrosinhydrazide maintained at 95–100° exhibited two parallel phenomena. One, the degree of inhibition of the α-chymotrypsin catalyzed hydrolysis of methyl acetyl-L-valinate in aqueous solutions at 25° and *p*H 7.90 and 0.1 *M* in sodium chloride by aliquots of the solution increased regularly with increased time of heating. Two, potentiometric titration, at 25°, of aliquots of the solution, showed that the concentration of the original conjugate acid, of $pK_A' = 3.2$, decreased while that of a new conjugate acid, of $pK_A' = 7.0$, increased with increased time of heating.

The kinetics of the above transformation were followed both enzymatically and potentiometrically. An aqueous solution 0.05 *M* in α-N-acetyl-L-tyrosinhydrazide and 0.05 *M* in hydrochloric acid was maintained at 96 ± 1°. Aliquots of the reaction mixture were removed at various time intervals and titrated potentiometrically at 25° with aqueous 0.125 *N* sodium hydroxide with the aid of a Di-Functional Recording Titrator.[6] The titer due to the appearance of the less acidic conjugate acid was calculated from the titration curves. Similarly, aliquots were assayed for inhibitor content using the system α-chymotrypsin–methyl acetyl-L-valinate, in aqueous solutions at 25.0° and *p*H 7.90 and 0.10 *M* in sodium chloride, with $[S]_0 = 40 \times 10^{-3}$ *M* and [E] = 0.1464 mg. protein-nitrogen per ml. Totally competitive inhibition was assumed and the size of the aliquot was adjusted to correspond to a maximum concentration of $[I] = 1.0 \times 10^{-3}$ *M* in the enzyme–substrate–inhibitor system, *i.e.*, for the initial reaction system at $t = 0$ the concentration of α-N-acetyl-L-tyrosinhydrazide was 1×10^{-3} *M* and at $t = \infty$, assuming 1:1 stoichiometry for the formation of the inhibitor, the concentration of the inhibitor was

(1) Supported in part by a grant from the National Institutes of Health, U. S. Public Health Service.

(2) To whom inquiries regarding this article should be sent.

(3) R. V. MacAllister and C. Niemann, *J. Am. Chem. Soc.*, **71**, 3854 (1949).

(4) R. Lutwack, H. F. Mower and C. Niemann, *ibid.*, **79**, 2179, 5690 (1957).

(5) R. J. Foster, H. J. Shine and C. Niemann, *ibid.*, **77**, 2378 (1955).

(6) The instrument used was manufactured by the International Instrument Co., Canyon, Calif. It can be used for constant or variable *p*H titrations.