Sequence of the A-Protein of Coliphage MS2

II. ISOLATION AND SEQUENCE DETERMINATION OF CHYMOTRYPTIC PEPTIDES*

(Received for publication, March 30, 1976, and in revised form, April 26, 1977)

JOËL S. VANDEKERCKHOVE,† JAN G. GIELEN, AND MARC C. VAN MONTAGU§
From the Institute of Histology and Genetics, State University Ghent, Ledeganckstraat 35, B-9000 Ghent, Belgium

A total of 94 different peptides have been isolated from chymotryptic digests of underivatized (35 mg) and of reduced, carboxymethylated, and citraconylated (48 mg) A-protein. Their amino acid sequences provide information for 375 residues of the total of 393 amino acids of the A-protein polypeptide chain. Abnormal specificity of chymotrypsin in the digestion of citraconylated A-protein was observed.

The preceding communication (1) reports the isolation of the A-protein, partial NH₂- and COOH-terminal sequence determinations, and also a characterization of the tryptic peptides. In the present paper we discuss the peptides obtained by chymotryptic hydrolysis of the A-protein. Peptides were first separated by gel filtration on Sephadex G-25, followed by a three-dimensional combination of paper electrophoresis and chromatography. Peptides, detected with fluorescamine, were characterized by amino acid sequence analysis.

The recovery of methionine- and histidine-containing peptides, isolated from a digest of carboxymethylated and citraconylated A-protein was generally found to be low. This could be circumvented by a chymotryptic digestion of underivatized A-protein. We also found distinct differences in the specificity of chymotrypsin towards derivatized and underivatized A-protein. This phenomenon proved to be very useful for the ordering of some chymotryptic peptides.

With the information obtained from 94 different peptides isolated from both digestions, we report the sequence of 375 residues of the total of 393 residues which have been assigned to the MS2 A-protein polypeptide chain (2).

**EXPERIMENTAL PROCEDURES**

The materials and procedures were those described in the preceding paper (1), with the following additions: subtilisin was obtained from the Carlsberg Laboratory; leucine aminopeptidase was a product of Sigma; and 2-methylmaleic acid anhydride was obtained from Fluka.

**Derivatization of A-Protein**

A-protein of MS2 was purified as described in the preceding paper (1). Reduction and carboxymethylation was carried out according to Crestfield et al. (3). After extensive dialysis against 0.01 n acetic acid, guanidinium HCl and pyrophosphate were added to a final concentration of 6 M and 0.1 M, respectively. After adjusting the pH to 8.7, citraconylation was carried out as described by Gibbons and Perham (4). Forty-eight milligrams of carboxymethylated and citraconylated A-protein was obtained as a suspension in 0.5% NH₄HCO₃.

**Chymotryptic Digestions**

**Digest I**—Carboxymethylated and citraconylated A-protein was treated with chymotrypsin in 0.5% NH₄HCO₃ at 37°. The enzyme-substrate ratio was 1:50. After 3 h, the pH was lowered by addition of formic acid to pH 3.5 and decitraconylation was carried out overnight at 37° in the presence of a few drops of toluene. The precipitate remaining at the end of the decitraconylation process was removed by centrifugation and washed with 7% acetic acid. The combined supernatants were lyophilized and taken up in 4 ml of 7% HAc. The remaining chymotryptic "core" was solubilized in 1 ml of phenol/formic acid (1:1, by volume) mixture and reprecipitated by adding 15 ml of ethanol and by adjusting the pH to 4.5 with pyridine. The washed "core" was kept at −20° after lyophilization from a 5% formic acid solution.

**Digest II**—Underivatized A-protein (35 mg) was suspended in 3 n NH₄OH and diluted with water to 0.2 n. After pH adjustment with acetic acid to pH 8.0, digestion was carried out for 12 h at an enzyme:substrate ratio of 1:50. The supernatant at pH 8.0, containing the soluble peptides, was lyophilized and taken up in 4 ml of 7% acetic acid. The insoluble "core" was not used for further structural work since, upon polyacrylamide gel electrophoresis in dodecyl sulfate, it was found to consist of a complex mixture of intact A-protein and large polypeptides.

**Separation of Soluble Chymotryptic Peptides**

The soluble chymotryptic peptides of Digests I and II were fractionated by the following procedure. Peptides were gel-filtered on a Sephadex G-25 column (200 × 0.9 cm) in 5% acetic acid. Fractions of 2.2 ml were collected. An aliquot was removed for examination by paper electrophoresis at pH 6.5. The results are given in Fig. 1. Fractions were pooled as indicated by solid bars. Each of these peptide groups was further fractionated by a "three-dimensional mapping" procedure. Peptides of each group were first subjected to electrophoresis in the solvent system 1-butanol/pyridine/acetic acid/water, 15:10:3:12 (by volume) at right angles to the original run. Peptides were located...
by a dilute fluorescamine stain (5) and eluted from paper with freshly prepared electrophoresis buffer at pH 6.5.

Separations of peptides derived from enzymatic cleavage of isolated chymotryptic peptides and used for amide assignment or for help in the sequence determinations were performed on thin layer cellulose plates.

**Digestions of Peptides with Proteolytic Enzymes**

Conditions for the digestion of peptides are given in Table I. All digestions were terminated by vacuum evaporation over P2O5 and NaOH. Carboxypeptidase A and B digestions were carried out in 0.2% NH4HCO3 as described by Ambler (6). Excess ammonia was removed by lyophilization in 20% pyridine before application on the amino acid analyzer. Leucine aminopeptidase digestion was as described by Light (7).

### Table I

**Conditions for digestion of peptides with proteolytic enzymes**

<table>
<thead>
<tr>
<th>Proteolytic enzyme</th>
<th>Amount of enzyme relative to substrate</th>
<th>Concentration of peptide Buffer</th>
<th>pH</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin</td>
<td>1.0 % w/w</td>
<td>NH4HCO3 (0.5%)</td>
<td>6.0</td>
<td>37°C</td>
</tr>
<tr>
<td>Subtilisin</td>
<td>2.0 % w/w</td>
<td>NH4HCO3 (0.5%)</td>
<td>7.8</td>
<td>25°C</td>
</tr>
<tr>
<td>Thermolysin</td>
<td>1.4 % w/w</td>
<td>NH4Ac (0.2 N HAc)</td>
<td>8.5</td>
<td>42°C</td>
</tr>
<tr>
<td>Staphylococcus aureus protease</td>
<td>2.5%</td>
<td>NH4HCO3 (0.5%)</td>
<td>7.8</td>
<td>37°C</td>
</tr>
</tbody>
</table>

**Fig. 1.** Fractionation of the soluble chymotryptic peptides of the carboxymethylated and citraconylated A-protein by gel filtration on Sephadex G-25 and electrophoresis at pH 6.5. Details of the column separation were given under "Separation of Soluble Chymotryptic Peptides." An aliquot of 150 μl of each fraction was applied to Whatman No. 3MM paper and subjected to pH 6.5 electrophoresis. Peptides were visualized with ninhydrin. Valine, glutamic acid, and aspartic acid were used as amino acid markers. Fractions were combined as indicated by solid bars (----). The pattern obtained from an exhaustive chymotryptic hydrolysis on derivatized A-protein (Digest II) was slightly different from the one shown; however, fractions could be combined in an identical way based on the position of some peptides easily recognized in both digests. Peptide group A, containing the largest peptides, gave a rather different pattern in both digests. Therefore, separation of these peptides was carried out differently, as is shown in Fig. 2 (for Digest I see A' group, for Digest II see A group). Peptide Cb22, migrating as indicated, was obtained in a pure form after one-dimensional electrophoresis.

**Fig. 2.** Separation of soluble chymotryptic peptides of MS2 A-protein. Each peptide group, obtained by combination of the Sephadex G-25 column fractions as indicated by solid bars (Fig. 1), was further fractionated on Whatman No. 3MM paper. Peptides were first subjected to electrophoresis at pH 6.5. Less complex mixtures were resolved by an additional chromatographic step in the solvent system 1-butanol/pyridine/acetic acid/water (B: A: P: W) = 15:10:3:12 (by volume) (groups A' and E). For some groups it was necessary to perform an additional electrophoresis step at pH 3.5 (groups C and D). The A and B peptide groups, with a more complex composition, were first divided into bands containing peptides with almost the same electrophoretic mobilities at pH 6.5. Each of these bands was further fractionated by a suitable combination of electrophoresis at pH 3.5 or pH 2.0 and descending paper chromatography. The relative electrophoretic mobilities of the peptide bands were calculated according to the mobility of aspartic acid in the first electrophoresis step and were as follows: Aa1, 0.25–0.30; Aa2, 0.21–0.25; Ab1, 0.45–0.55; Ab2, 0.30–0.30; Ba1, 0.39–0.43; Ba2, 0.33–0.39; Ba3, 0.27–0.32; Ba4, 0.22–0.27; Bb1, 0.11–0.14; Bb2, 0.8–0.9; Bb3, 0.7–0.8; Bb4, 0.45–0.55; Bb5, 0.30–0.40. The neutral peptides of groups A and B were separated by electrophoresis at pH 3.5 and chromatography. Peptides not completely resolved by this procedure could ultimately be purified by electrophoresis at pH 3.5 or 2.0, depending on the pH used during the previous electrophoresis step. The map of group A' is composed entirely of large peptides from Digest I. Peptides from the exhaustive Digest II were not present in this group. Methyl green (M.G.) and Pyronin (Pyr) were color markers used during electrophoresis. Valine was used as an amino acid reference during paper chromatography. f indicates the chromatographic solvent front.

**Methods for amino acid composition analysis and for sequential degradation are described in the preceding communication (1).** Amides were assigned from the molecular weight and the relative electrophoretic mobility at pH 5.5 according to Offord (8). In cases where more than one possible amide occurred in a given peptide, the mobilities before and after successive removal of residues by Edman degradation were measured. Alternatively, amides were positioned after complete hydrolysis of the peptides with leucine aminopeptidase or carboxypeptidase A. The position of tryptophan and histidine was obtained directly by degradation with carboxypeptidase A. Aliquots were removed after appropriate digestion times and mixed with a known amount of norleucine serving as internal reference. One part was subjected to acid hydrolysis and the other part was analyzed as such. From
Sequence of MS2 A-Protein

Ch1 (+1) Met-Arg-Ala-Phe
Ch5    Ser-Thr-Leu-Asp-Arg-Glu-Asn-Glu-Thr-Phe
Ch16 (+1) Val-Pro-Ser-Val-Arg-Val-Tyr
Ch22   Ala-Asp-Gly-Glu-Thr-Glu-Asp-Asn-Ser-Phe
Ch32 (+1) Ser-Leu-Lys-Tyr
Ch36   Arg-Ser-Asn-Trp
Ch40 (+1) Thr-Pro-Gly-Arg-Phe
Ch45   Asn-Ser-Thr-Gly-Ala-Lys-Thr-Lys-Gln-Trp

Ch55    His-Tyr-Pro-Ser-Pro-Tyr
Ch61 (+1) Ser-Arg-Gly-Ala-Leu
Ch66    Ser-Val-Thr-Ser-Ile-Asp-Gln-Gly-Ala-Tyr
Ch76    Lys-Arg-Ser-Gly-Ser-Ser-Trp
Ch83 (+1) Gly-Arg-Pro-Tyr
Ch87 (-1) Glu-Glu-Lys-Ala-Gly-Phe
Ch93 ( o) Gly-Phe

ICh95   Ser-Leu-Asp-Ala-Arg
ICh100  Ser-Cys(Cm)-Tyr
Ch103 ( o) Ser-Leu-Phe-Pro-Val-Ser-Gln-Asn-Leu

Ch112 ( o) Thr-Tyr

Residues 114 through 137 : see Fig. 4
Ch138 ( o) Asn-Leu-Gly-Val-Ala-Leu

Residues 144 through 153 : see Fig. 4
Ch150 ( o) Gly-Leu-Asp-Phe

Residues 154 through 164 : see Fig. 4
IICCh165 Thr-Ala-Ala-Arg-Arg-Gly-Asn-Trp
ICh165 Thr-Ala-Ala-Arg
ICh169 Arg-Gly-Asn-Trp and Arg-Ser-Asn-Trp
Ch173 (+1) Arg-Gln-Ala-Leu
Ch177 (+1) Arg-Tyr

Ch179    Leu-Ala-Leu-Asn-Glu-Asp-Arg-Lys-Phe
Ch180 ( o) Ala-Leu-Asp-Glu-Asx-Arg-Lys-Phe
Ch188 (+3) Arg-Ser-Lys-Mts
Ch192    Val-Ala-Gly-Arg-Trp
Ch197    Leu-Glu-Leu-Gln-Phe
Ch202    Gly-Trp-Leu-Pro-Leu
Ch207    Met-Ser-Asp-Ile-Gln-Gly-Ala-Tyr

IICCh215 (-1) Glu-Met-Leu
IICCh218 (+2) Thr-Lys-Val-Mts
Ch222    Leu-Gln-Glu-Phe
Ch226 ( o) Leu-Pro-Met
ICh241   Asp-Gly-Arg-Leu-Ser-Tyr-Pro-Ala-Ala-Asn-Phe

Ch242 ( o) Arg-Gly-Leu
Ch245 ( o) Ser-Tyr-Pro-Ala-Ala-Asn-Phe
Ch267    Ile-Asn-Asp-Ala-Arg
Ch275 ( o) Leu-Ser-Ser-Leu
Ch279 ( o) Gly-Ile-Leu-Asn-Pro-Leu
Ch285    Gly-Ile-Val-Trp
Ch289 ( o) Gly-Lys-Val-Pro-Phe
Ch294a   Ser-Phe-Val-Val-Asp-Trp
Ch294b ( o) Ser-Phe
Ch296    Val-Yal-Asp-(Trp)
Ch340 ( o) Gly-(Trp)
IICCh358 (+1) Arg-Gly-Val-Gln
IICCh362 Ser-Va (Trp)-Pro-Thr-Thr-Gly-Ala-Tyr

Ch371 (+1) Val-Lys-Ser-Pro-Phe
Ch387 (+2) Ile-Arg-Gln-Arg-Leu
Ch392 (+1) Ser-Arg
Ch319 Sp319 Met-Ser-Arg-Thr-Val-Thr-Asp-Val-Ile-Thr-Gly-Glu
IICCh319 Sp331 Ser-Ile-Ile-Val-Asp-Ala-Pro-Cyr

Yet-Ile-Ile-Ser-Val-Asp-Ala-Pro-Yr
Fig. 4. Amino acid sequence of regions in the MS2 A-protein form overlapping chymotryptic peptides. The meaning of the signs is explained in the section "Peptides Nomenclature and Abbreviations." The charge on peptides, necessary for the assignment of the amides, is given in parentheses: (-) for acidic; (+) for basic; (0) for neutral.

Fig. 3. Amino acid sequence of chymotryptic peptides of the A-protein of bacteriophage MS2. Peptides, the sequence of which is given without further explanation, are discussed in detail in the miniprint supplement. The sequences of the remaining peptides could be obtained directly using dansyl-monitored Edman degradation and net charge determination. For the meaning of the signs reference is made to the section "Peptide Nomenclature and Abbreviations." The charge of the peptides at pH 6.5, necessary for the assignment of the amides, is given in parentheses: (+) for basic; (-) for acidic; (0) for neutral.
Sequence of MS2 A-Protein

Residues 300 through 318

Leu-Leu-Pro-Val-Gly-Asn-Met-Leu-Glu-Gly-Leu-Thr-Ala-Pro-Val-Gly-Cys-Ser-Tyr
ICH300

ICH300a

ICH300b

ICH300T1300 ICH300 T1303(o) ICH300 T1307ICH300 T1310(-1)

ICH300 T1310b ICH300 T1314(-1)

Residues 342 through 357

Thr-Val-Glu-Arg-Gln-Gly-Thr-Ala-Lys-Ala-Gln-Ile-Ser-Ala-Met-His
ICH342

ICH342 T1342(o) ICH342 T1348(+1) ICH342 T1353(o)

ICH342 T1349ICH342 T1351(o)

ICH342 Tr342(o) ICH342 Tr346(+1) ICH342 Tr351(o)

ICH342a

ICH342b

ICH342c

ICH342c Sp342 ICH342c Sp345(+1)

**Fig. 4—Continued**

these results we were able to calculate the amount of amino acid liberated per mol of peptide. Tryptophan was determined quantitatively on a Biotronik amino acid analyzer using Durrum resin DC4A.

Peptides showing a positive Ehrlich reaction on paper are considered to contain tryptophan.

All data presented of the dansyl monitored Edman degradation of peptides were unique. In cases where ambiguous or unclear results were obtained, no data are given and the sequence was deduced by other methods (secondary proteolytic hydrolysis, carboxypeptidase A and B degradation, and extensive leucine aminopeptidase hydrolysis).

Peptide Nomenclature and Abbreviations

Peptides are designated by two letters indicating the enzyme used for hydrolysis of the A-protein and for further hydrolysis of the peptides (Ch for chymotrypsin, Tr for trypsin, Tl for thermolysin, Su for subtilisin, and Sp for *Staphylococcus aureus* protease). These letters are followed by arabic numbers indicating the position of the NH₂-terminal residue of the peptide in the A-protein chain. Different peptides having their NH₂-terminal residue at the same position in the A-protein (peptides produced by partial splitting at the COOH terminus) are further characterized by suffixes a, b, c, etc. Since the specificity of chymotrypsin towards underivatized and citraconylated A-protein was different, some peptides were only recovered in one of both digests. This type of peptide is indicated by a roman number preceding the name of the peptide: I, for peptides recovered from the chymotryptic digest on reduced, carboxymethylated, and citraconylated A-protein; II, for peptides isolated from underivatized A-protein.

The sign → in figures underneath a residue denotes that the nature of the residue has been established by the dansyl-monitored Edman procedure. The sign + above the amino acid residue indicates a sequence determination from a leucine aminopeptidase digest. The sign – above the amino acid residue indicates that carboxypeptidase A and/or B hydrolysis was used for the sequence determination. The position of residues given between brackets was

\(^1\) The abbreviation used is: dansyl, 6 dimethylaminonaphthalene 1-sulfonyl.
Sequence of MS2 A-Protein

The results of Edman degradation and the methods described under "Experimental Procedures" indicated the separation patterns of peptides derived from Digest I and Digest II. Peptides which were only isolated from Digest I are designated by roman number I, while peptides only recovered in Digest II are indicated as II. The amino acid composition, net charge at pH 6.5, and relative yields of the peptides obtained by chymotryptic digestion of the soluble A-protein (Digest I) on the one hand, and the different peptides shown in Fig. 2 are presented in Table II.

### Table II

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Amino Acid Composition</th>
<th>Net Charge (pH 6.5)</th>
<th>Relative Yields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide 1</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Peptide 2</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Peptide 3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Peptide 4</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Values for threonine and serine have not been corrected for destructive loss during hydrolysis. The presence of tryptophan was estimated by a positive Ehrlich reaction. The yields are expressed as percentages (nmol/nmol) of the protein digested, with no correction for losses, known or unknown in the multistage purification.
**Sequence of MS2 A-Protein**

### Table II—Continued

<table>
<thead>
<tr>
<th>12348</th>
<th>12347</th>
<th>12342</th>
<th>12351</th>
<th>12353</th>
<th>12350</th>
<th>12345</th>
<th>12344</th>
<th>12340</th>
<th>12336</th>
<th>12337</th>
<th>12338</th>
<th>12334</th>
<th>12335</th>
<th>12352</th>
<th>12339</th>
<th>12333</th>
<th>12331</th>
<th>12332</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartic acid</td>
<td>1.00 (1)</td>
<td>1.10 (1)</td>
<td>0.84 (1)</td>
<td>1.34 (1)</td>
<td>1.11 (1)</td>
<td>0.84 (1)</td>
<td>1.34 (1)</td>
<td>0.84 (1)</td>
<td>1.34 (1)</td>
<td>0.84 (1)</td>
<td>1.34 (1)</td>
<td>0.84 (1)</td>
<td>1.34 (1)</td>
<td>0.84 (1)</td>
<td>1.34 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threonine</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alanine</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td>1.00 (1)</td>
<td>2.04 (8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valine</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td>0.91 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methionine</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0.91 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyrosine</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td>1.00 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proline</td>
<td>0.91 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparagine</td>
<td>0.91 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM cysteine</td>
<td>0.91 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trpapolan</td>
<td>0.91 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td>0.80 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total residues</td>
<td>45</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield (%)</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net charge at pH 6.5</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
due to the small amount of protein available, the sequences were mainly based on data obtained from the dansyl-monitored Edman procedure. Additional verification came from further digestion of the primary peptides or from results of carboxypeptidase and leucine aminopeptidase degradations.

We wish to point out that, except for peptide ICh169, all soluble chymotryptic peptides used for the integration of the final sequence of the polypeptide chain of MS2 A-protein were obtained in pure form. This conclusion could be made from the amino acid composition, the complete agreement of electrophoretic mobility, molecular weight, and net charge in the O'ford plot (8) and also from the unambiguous results obtained during Edman degradation.

**Determination of the Amino Acid Sequence of the Chymotryptic "Core" Peptide**—The chymotryptic "core" was isolated from Digest I (see "Experimental Procedures"). Upon electrophoresis in 10% and 20% polyacrylamide dodecyl sulfate-containing gels (results not shown), the "core" migrated as a single polypeptide with a molecular weight of 2,300. Tryptic "core" peptides of MS2 coat protein with respective molecular weights of 2,670 and 3,926 were used as reference peptides. Although the relationship between molecular weight and relative mobility is not always linear for polypeptides in this molecular weight range, these data allowed the conclusion that the "core" is either a single polypeptide chain free of intact A-protein and other large polypeptides, or a mixture of peptides apparently with similar molecular weights. The amino acid composition as presented in Table II (peptide ICh319) was calculated for a polypeptide with a molecular weight of 2,300.

Upon digestion of the chymotryptic "core" with *Staphylococcus aureus* protease under the conditions given in Table I, three peptides were recovered in high yield. Their separation is presented in Fig. 5. Peptide material was not found at the origin, either by spraying the paper with fluorescamine or by elution with formic acid followed by acid hydrolysis. Peptides ICh319 SP319 and ICh319 SP319 showed identical amino acid compositions. Both forms of the same peptide, showing different chromatographic and electrophoretic behavior, could be explained by partial oxidation of the NH₂-terminal methionyl residue (8).

The amino acid sequences of the secondary peptides of ICh319 were derived by dansyl-monitored Edman degradation and complete leucine aminopeptidase hydrolysis (for assignment of amides) and is presented in Fig. 3. Since peptide ICh319 SP331 contained COOH-terminal tyrosine, we deduced the order ICh319 SP319-ICh319 SP331.

The sum of the compositions of peptides ICh319 SP319 and ICh319 SP331 differs only slightly from the composition of the primary "core" from which they are derived. No peptides could be recovered from the *Staphylococcus aureus* protease map of ICh319 in yields higher than 7% of the yield of ICh319 SP331. These findings practically exclude the existence of other fragments as parts of peptide ICh319. The complete proof for the sequence of ICh319 by assaying another type of proteolytic cleavage could not be given since no material was left. However, data from thermolytic peptides T1323 and T1319, presented in the following paper (9), provide additional support for the proposed sequence of peptide ICh319.

**DISCUSSION**

A uniform distribution of aromatic and leucine residues, providing the major cleavage sites for chymotrypsin, resulted in the formation of peptides of suitable size for purification on paper. We noticed strong trypptic-like specificity towards carboxymethylated and citraconylated A-protein. This phenomenon was observed in the arginine-containing stretches presented in Table III. Cleavage at the COOH-terminal side of arginine were mostly quantitative (peptides derived from tryptic-like cleavage were recovered with approximately the same yield as the other chymotryptic peptides) and could not directly be explained as due to chemical modification-induced negative charges. Cleavage of some arginine-containing groups in the primary "core" can be observed from the data presented in Table III, the action of chymotrypsin appears to be random or to be related to a particular secondary or tertiary structure in the immediate environment of the susceptible bond in the citraconylated protein. We would like to point out that the same batch of chymotrypsin was also used for the digestion of underivatized A-protein (Digest II), where this type cleavage was not found. It is obvious that the splitting at arginine was not a result of trypsin activity in chymotrypsin. Advantage was taken of the differences in chymotrypsin specificity towards underivatized A-protein and reduced, carboxymethylated, and citraconylated A-protein to make the alignment of the peptides presented in Fig. 4 (alignment of residues 114-137, 144-153, 154-164, and 229-240).

Citraconylation allows chymotrypsin to reach cleavage sites more uniformly, producing a relative homogeneous "core." This "core" was shown to consist of a single polypeptide chain of 21 amino acids (ICh319). However, it is possible that other insoluble peptides were produced by chymotryptic digestion and we cannot exclude the possibility that other hydrophobic peptides forming the chymotryptic "core" are washed out by ethanol precipitation from a phenol/formic acid solution. Chymotryptic hydrolysis of underivatized A-protein results in a very heterogeneous insoluble "core" which could not be used for further sequence studies. Although more soluble peptides were produced by exhaustive hydrolysis than by short term cleavage, less sequence information was gathered from Digest II. This is because sulfhydryl groups had not been protected and cysteine-containing peptides were lost; and because there was failure to cleave at arginine residues.

The recovery of some hydrophilic peptides was low compared to other peptides. This deficiency has to be related to the irreversible adsorption of the peptides to the Sephadex or...
Arginine-containing amino acid stretches quantitatively cleaved at the COOH terminus of arginine upon chymotrypsin hydrolysis of carboxymethylated and citraconylated MS2 A-protein

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Arg</td>
<td>Ser</td>
<td>Cys</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Asn</td>
<td>Arg</td>
<td>Ala</td>
<td>Ser</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Ala</td>
<td>Arg</td>
<td>147-148</td>
<td>Ser</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Arg</td>
<td>168-169</td>
<td>Arg</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Val</td>
<td>Arg</td>
<td>232-233</td>
<td>Gln</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Arg</td>
<td>271-272</td>
<td>Leu</td>
<td>Ala</td>
<td></td>
</tr>
</tbody>
</table>

Digest I (on citraconylated A-protein) and Digest II yielded exactly the same methionyl and histidyl peptides; although the recovery was clearly higher for Digest II. A possible partial chemical modification during carboxymethylation or citraconylation, or both, at the methionine and histidine residues could be one reason for the low recovery of this type of peptide in Digest I. However, we were not able to characterize known or unknown methionyl and histidyl derivatives during amino acid analysis of these peptides.

By calculation of the electrophoretic mobilities at pH 6.5, all histidyl residues showed the contribution of one positive charge except in peptide Ch55, with NH₂-terminal histidine. The same phenomenon was observed during the sequence elucidation of Escherichia coli ribosomal protein S21, where a His-Ala-Lys peptide with a single positive charge was isolated (10). In both cases, characterization of the first residue by dansylation failed, while the Edman degradation appeared to be normal. In the same way, we observed that during the dansyl-monitored Edman degradation of histidine-containing peptides, the histidine was not detected as dansyl derivative except when it was at the COOH terminus. The low pK of the α-NH₂ group of the histidyl residue is probably because of the severe decrease of dansylation efficiency.

Acknowledgments — We wish to thank Dr. M. Osborn and Professor K. Weber of the Max Planck Institut, Göttingen, Germany for advice and critical reading of this manuscript. We thank Mr. A. Lenaerts for his expert assistance. The work has been carried out in the laboratory of Professor M. Sebrayns.

REFERENCES
Sequence of MS2 A-Protein

7771

Sequence of MS2 A-Protein
Sequence of MS2 A-Protein

Position 1-128

The sequence of MS2 A-Protein was determined by Edman degradation after cleavage of the protein with cyanogen bromide. The sequence is:

Met-Ser-Glu-Gly-Ala-Lys-Tyr

Met-Ser-Glu-Gly-Ala-Lys-Tyr

Position 129-255

Hydroxylation reaction

Met-Ser-Glu-Gly-Ala-Lys-Tyr

The Edman degradation was performed on the peptide shown above. The amino acid sequence was determined to be Met-Ser-Glu-Gly-Ala-Lys-Tyr. The sequence was confirmed by the Edman degradation of the peptide obtained from the cleavage with cyanogen bromide.

Position 256-373

Alignment of residues 229 through 234

Met-Ser-Glu-Gly-Ala-Lys-Tyr

The sequence of the peptide is shown below. The sequence is: Met-Ser-Glu-Gly-Ala-Lys-Tyr. The sequence was confirmed by the Edman degradation of the peptide obtained from the cleavage with cyanogen bromide.

Position 374-500

Sequence of MS2 A-Protein

The sequence of MS2 A-Protein was determined by Edman degradation after cleavage of the protein with cyanogen bromide. The sequence is:

Met-Ser-Glu-Gly-Ala-Lys-Tyr

The Edman degradation was performed on the peptide shown above. The amino acid sequence was determined to be Met-Ser-Glu-Gly-Ala-Lys-Tyr. The sequence was confirmed by the Edman degradation of the peptide obtained from the cleavage with cyanogen bromide.
Sequence of the A-protein of coliphage MS2. II. Isolation and sequence determination of chymotryptic peptides.
J S Vandekerckhove, J G Gielen and M C Van Montagu


Access the most updated version of this article at http://www.jbc.org/content/252/21/7761

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/252/21/7761.full.html#ref-list-1