THE NON-ACETYLATED N-TERMINAL SEQUENCE OF THE GREEN TOMATO ATYPICAL MOSAIC VIRUS COAT PROTEIN (G-TAMV)*

W. A. Rombauts

Virus Laboratory, University of California, Berkeley, California 94720

Received April 18, 1966

The N-terminal end of G-TAMV protein has no acetyl group. The sequence is NH-Pro-Tyr-Thr-Ile-AspNH_2-Ser-Pro-, which is identical, as far as investigated, to the recently reported N-terminal sequence of the U2 strain (Wittmann, 1965a).

In 1962, Knight and coworkers isolated from diseased tobacco leaves two different virus strains related to the common ("wild" or "vulgare") strain of tobacco mosaic virus (TMV). They called them yellow tomato atypical mosaic virus (Y-TAMV) and green tomato atypical mosaic virus (G-TAMV), according to the color of the mottling they cause on leaves of Nicotiana tabacum var. Turkish (Turkish tobacco). Both of these strains were obtained from plants which had originally been inoculated with material from a leaf of tobacco infected with tomato atypical mosaic virus (TAMV), originally described by Miller (1953). The total amino acid composition of the coat protein of these strains showed 18 net differences, while the Y-TAMV and G-TAMV differed from the "wild" TMV strain respectively by 8 and 17 net exchanges. Furthermore, they reported that Y-TAMV had a C-terminal serine and, at the N-terminus, the same sequence as "wild" TMV, i.e., N-acetyl-Ser-Tyr-. The C-terminal residue of G-TAMV was found to be threonine, which is the same as in "wild" TMV, but attempts to detect the acetylated or free N-terminal residue were unsuccessful

*This investigation was supported by grant GB 3107 from the National Science Foundation.

On the basis of the total amino acid composition of the coat proteins, Tsugita (1962) classified Y-TAMV with Dahlemense strain in his group "B", and G-TAMV in a separate group, called "C". Later, Funatsu (1964) published the amino acid composition of some peptides isolated from a tryptic digest of G-TAMV coat protein. Since then, no further characterization of G-TAMV coat protein was attempted. In this work, we now report the sequence of seven residues at the N-terminal end, which is particularly noteworthy, because it represents one of the few cases where a plant virus protein is not acetylated.

Materials and Methods

Virus coat protein was a gift from Professor C. A. Knight, and was prepared as described by Knight et al. (1962). Two different batches were used: the first one was prepared in March, 1963, the other one in January, 1966. Amino acid analyses of the protein were made after 24 and 72 hours of hydrolysis in 6 N HCl at 108°C, on the 150 x 1 cm and 15 x 1 cm columns of the Beckman/Spinco analyzer according to Spackman et al. (1958). Sequence determinations were made with the Edman method (1950). The paper strip adaptation of Fraenkel-Conrat (1954) was used as described by Schroeder et al. (1963). In solution, the new procedure of Edman, as described by Doolittle (1964) was used.

The C-terminal residue was determined with carboxypeptidase (Sigma, 2x crystallized). The digestion was made for four hours at room temperature on 0.2 μM of protein, the enzyme to protein ratio being 1:1000. Identification of the amino acids split off was done on the analyzer.

The phenylthiohydantoin (PTH) amino acids were identified after paper chromatography on the A, D and E solvents of Sjöquist (1953), and Edman and Sjöquist (1956). Quantitation, when necessary, was done spectrophotometrically according to Sjöquist (1959).

Paper chromatographic identification was occasionally verified by hydrolyzing the phenylthiohydantoins back to the constituent amino acids

in 0.1 N NaOH at 120°C for 12 hours, and detecting them on the analyzer (Van Orden and Carpenter, 1964). No attempt to quantitate recoveries in this system was made, as B. Veniček (personal communication) has found that this requires special care in removing even trace amounts of air from the hydrolysates.

Results

Table I gives the amino acid composition of the coat protein of G-TAMV and, for comparison, the amino acid composition determined earlier by Knight et al. (1962).

Table II gives the partial sequence found by the phenylisothiocyanate method at the N-terminal end of the chain. In all three experiments, the proline residue at the free N-terminus was identified unambiguously after chromatography of its phenylthiohydantoin in solvent D. The recovery after elution from the paper was 0.6 mole per peptide chain in experiment one and two, and 0.68 in experiment three. In all cases, its absorption spectrum was checked in a Cary recording spectrophotometer and found to be that of a typical phenylthiohydantoin amino acid, with maximum at 269 mμ and minimum at 240 mμ. Its identity was double-checked by determination on the amino acid analyzer after hydrolysis. Recovery in this case, however, was only about 30%.

After three degradation steps with the paper strip method, so many PTH-amino acids appeared on the chromatogram that no identification could be made with certainty.

However, application of the new procedure of Edman, as described by Doolittle (1964), gave much cleaner degradations. The isoleucine residue in position four in experiment three was determined unambigously after hyrolysis on the analyzer. As indicated by Van Orden and Carpenter (1964), PTH-isoleucine after alkaline hydrolysis is converted partly into alloisoleucine and partly into isoleucine (ratio about 3:2). PTH-leucine, on the contrary, gives only leucine. In our analysis, only alloisoleucine and isoleucine were present in ratio 3:2.

Table I

Amino Acid Analyses of G-TAMV Coat Protein

	Our Analyses[a]				Knight et al. (1962)[b]			
	24 h	72 h	Extrapol. Zero Time	Nearest Integer Values	24 h	72 h	Extrapol. Zero Time	Integer[c] Values
Lys	1.0	1.0	-	1	1.0	1.0	-	1
His	-	-	-	-	-	-	-	-
Arg	8.2	7.6	-	8	7.9	7.5	-	8
Asp	22.3	22.4	-	22	21.9	22.0	-	22
Thr	18.1	17.5	18.6	19	18.2	17.4	18.6	19
Ser	9.1	8.1	9.7	10	9.0	8.1	9.5	10
Glu	15.9	16.1	-	16	16.0	16.0	-	16
Pro	9.5	9.7	-	10	10.1	10.2	-	10
Gly	4.1	4.1	-	4	4.1	4.1	-	4
Ala	18.2	18.0	-	18	18.0	18.1	-	18
Val	11.5	11.8	-	12	11.0	11.8	-	12
1/2 Cys	.7	trace	-	1	-	-	-	1
Met	1.8	2.0	-	2	1.9	1.7	-	2
Ile	6.3	7.7	-	8	6.3	7.2	-	8
Leu	10.7	11.1	-	11	10.7	11.0	-	11
Tyr	5.5	5.7	-	6	5.7	5.8	-	6
Phe	7.6	8.0	-	8	7.8	7.8	-	8
Try	Not det.			(2)	-	-	-	2
Total				158				158

[a] Serine and threonine were extrapolated to zero time, and valine, isoleucine and leucine values were calculated on the basis of the 72 h results only.

[b] Cysteine was determined by Anson's (1942) uric acid colorimetric method, and tryptophan was estimated by the spectrophotometric method of Goodwin and Morton (1946).

[c] Values in this column are nearest integer values except in three instances when independent judgments, allowing more adequately for specific hydrolysis difficulties, were made.

Table II

N-Terminal Sequences of G-TAMV Coat Protein

Exp. 1	Batch 1963	Pro-Tyr-Thr- (Pro) (Tyr)	paper strip method
Exp. 2	Batch 1966	Pro-Tyr-Thr (Pro) (Tyr)	paper strip method
Exp. 3	Batch 1966	Pro-Tyr-Thr-Ile-AspNH$_2$-Ser-Pro- (Asp) (Gly) (Gly) (Ser) (Ala) (Ala) (Leu) (Val)	in solution

Determinations were made on 8-12 mg of protein, corresponding to 0.5-0.7 μM. Minor spots in the chromatograms are indicated between brackets for each step.

Asparagine in position five was identified on paper in solvent E in overall yield of about 35%, with a trace of aspartic acid and serine also present. Serine in position six was identified in solvent E. The overall yield here was rather low (5 to 10%). This is to be expected because of partial formation of PTH-dehydroserine (Ilse and Edman, 1963). Small amounts of glycine and alanine were also present.

Proline in position seven was identified in solvent D and the recovery was 15%. Traces of leucine, valine, glycine and alanine (10% of the proline amount) were present.

Our recoveries are of the same order of magnitude as those obtained by Doolittle (1964), who cites an average recovery per step of about 80%. After seven degradation cycles, the experiment was stopped because of difficulties in identification of the subsequent residues.

Most other coat proteins from plant viruses so far investigated, have an acetylated N-terminal end. It therefore appeared important to verify that no accidental breaks had occurred during preparation of the protein, producing a spurious N-terminal proline residue, since X-proline peptide bonds are known to be less stable than others. If this were true, additional C-terminal residues should also have been formed. After carboxypeptidase treatment of our material, threonine was identified as the only significant free amino acid. The yield of 0.83 mole per mole protein corresponds to the

average recovery of C-terminal threonine from common TMV coat protein by this procedure.

Conclusions

From the reported evidence we conclude that the G-TAMV coat protein has a free N-terminus, and that the sequence at the N-terminal end of the chain is as follows: NH-Pro-Tyr-Thr-Ile-AspNH_2-Ser-Pro-.

The only C-terminal amino acid residue is threonine, which confirms the results of Knight et al. (1962). Our total amino acid composition results are also in agreement with those of these authors. The G-TAMV coat protein seems to be an exception among the plant virus coat proteins, all of which until recently had been reported to have an acetylated amino-terminal end. One possible exception is the coat protein of the satellite virus of tobacco necrosis virus (Reichman, 1964). Very recently, Wittmann (1965a) reported some amino acid sequences (1-61, 135-158) of the U2 strain protein which also was implied to have a free proline at the N-terminal end. In a following paper, Wittmann (1965b) published convincing evidence to substantiate his statement.

The N-terminal sequence of the G-TAMV strain is, as far as we have investigated it, identical with the corresponding sequence of the U2 strain, and the C-terminal residues are also the same. Thus, G-TAMV may well be identical with U2 as suggested by Wittmann, although the current incomplete evidence is insufficient to warrant this conclusion. Wittmann's statement(1965a), that U2 was cultured in Berkeley at the same time and place as Y-TAMV and hence picked up as a contaminant in the latter and called G-TAMV is incorrect. Virus Laboratory records show that U2 was not cultured or worked with at all during the first few years in which Y-TAMV and G-TAMV were under study in Berkeley. In fact, there were three years between the last culture of U2 and the discovery of G-TAMV.

Clearly, the origin of G-TAMV cannot be explained as contamination of Y-TAMV with U2, and other possibilities must be considered.

The absence of a terminal acetyl group sets G-TAMV apart from other typical plant viruses. If it is assumed, with Marchis-Mouren and Lipmann (1965), that acetylation takes place at the level of the finished protein, then this free N-terminus of G-TAMV could be attributed to the acetylase not acting on proline. Since prolyl bonds are resistant to many exo- and endo-proteases, the presence of an N-terminal proline might as such present an asset similar to that of acetylation in terms of the survival value for the virus.

Alternatively, the role of the acetyl group in plant virus proteins may conceivably correspond to that of the formyl group in bacterial protein synthesis (e.g. Adams and Cappechi, 1966), although the finding of several different acetylated amino acids in plant viruses does not favor this concept. If the origin of the acetyl group were related to chain initiation, then one would expect that it could have been secondarily lost from G-TAMV, the strain lacking an acetyl. This might then be attributable to a lesser stability of the acetyl-proline bond, as compared to other acetyl amino acids.

Acknowledgements

The author is grateful to Dr. C. A. Knight for the generous gift of the virus coat protein, and to Dr. W. M. Stanley, Dr. H. L. Fraenkel-Conrat and Dr. C. A. Knight for critical reading of the manuscript and discussion. He also thanks Dr. R. F. Doolittle for valuable advice about the Edman method.

References

Adams, J. M., and Cappechi, M. R. Proc. Natl. Acad. Sci. U. S. 55, 147 (1966).
Anson, M. L. J. Gen. Physiol. 25, 355 (1942).
Doolittle, R. F. Biochem. J. 94, 742 (1964).
Edman, P. Acta Chem. Scand. 4, 283 (1950).
Edman, P., and Sjöquist, J. Acta Chem. Scand. 10, 10 (1956).
Fraenkel-Conrat, H. J. Am. Chem. Soc. 76, 3606 (1954).
Funatsu, G. Biochemistry 3, 1351 (1964).
Goodwin, T. W., and Morton, R. A. Biochem. J. 40, 628 (1946).
Ilse, D., and Edman, P. Australian J. Chem. 16, 411 (1963).

Knight, C. A., Silva, D. M., Dahl, D., and Tsugita, A. Virology 16, 236 (1962).
Marchis-Mouren, G., and Lipmann, F. Proc. Natl. Acad. Sci. U. S. 53, 1147 (1965).
Miller, P. M. Phytopathology 43, 480 (1953).
Reichmann, M. E. Proc. Natl. Acad. Sci. U.S. 52, 1009 (1964).
Schroeder, W. A., Shelton, J. R., Shelton, J. B., Cormick, J., and Jones, R. T. Biochemistry 2, 992 (1963).
Sjöquist, J. Acta Chem. Scand. 7, 447 (1953).
Sjöquist, J. Arkiv Kemi 14, 291 (1959).
Spackman, D. H., Stein, W. H., and Moore, S. Anal. Chem. 30, 1190 (1958).
Tsugita, A. J. Mol. Biol. 5, 293 (1962).
Van Orden, H. O., and Carpenter, F. H. Biochem. Biophys. Res. Commun. 14, 399 (1964)
Wittmann, H. G. Z. Naturforsch. 20b, 1213 (1965a).
Wittmann, H. G. Z. Vererbungslehre 97, 204 (1965b).