THE AMINO ACID SEQUENCE OF THE MAJOR COAT PROTEIN SUBUNIT OF THE FILAMENTOUS VIRUS Xf

Blas FRANGIONE, Yasutsugu NAKASHIMA*+, William KONIGSBERG+ and Robert L. WISEMAN++

Department of Pathology, Irvington House Institute, New York University Medical Center, New York, NY 10016,
+Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06250 and
++Department of Biochemistry, The Public Health Research Institute of The City of New York, Inc., New York, NY 10016, USA

Received 21 August 1978

1. Introduction

Xf is a filamentous bacterial virus which has *Xanthomonas oryzae* as host [1,2]. This virus can be distinguished from Pfl and fd, two other filamentous viruses of current structural interest, by its length, host specificity and the amino acid composition of its coat protein [3]. Despite differences, all three viruses contain circular single-stranded DNA (reviewed [4]). Gel electrophoresis in sodium dodecyl sulfate has shown that in each virus most of the protein consists of small subunits of a single major coat protein with 2% or less from other protein components and according to X-ray diffraction patterns of intact virions, these major protein subunits form a cylindrical sheath around each circular DNA molecule [5].

The sequence of the major coat protein is known for fd [6–8] and for Pfl virus [9]. To facilitate an accurate structural analysis of the filamentous virus Xf and to allow comparisons with the coat proteins from other filamentous viruses, the primary structure of the Xf major coat protein has been determined and is reported here.

2. Materials and methods

2.1. *Virus protein*

Purified Xf virus [5] was disrupted with redistilled phenol and the protein extracted in the phenol phase [6] was extensively dialyzed against 0.015 M sodium citrate and finally against distilled water. The resulting protein suspension was freeze-dried.

2.2. *Sequence analysis*

Cleavage by cyanogen bromide (CNBr), partial acid hydrolysis with HCl and digestions with pronase, subtilisin, carboxypeptidase A (CPA) and thermolysin were carried out by standard methods [10]. The resulting peptides were separated by ion-exchange chromatography and paper electrophoresis [11]. Amino-terminal analyses were done with dansyl chloride [12] and the dansyl-Edman technique [12] was used to sequence small peptides. Automated amino acid sequencing was performed with a Beckman 890C Sequencer using a peptide program (Beckman 102974) with dimethylallylamine–trifluoroacetic acid buffer. Phenylthiohydantoin amino acids were identified as desired. Amino acid analysis was done with a Durrum D-500 automatic amino acid analyzer [13]. The mass spectral analysis of peptides was carried out by E. Henson (Children's Hospital Medical Center, Boston) and by W. McMurray (Department of Physical Sciences, Yale University).

3. Results

The major coat protein of Xf consists of 44 amino acids. The amino acid sequence of this protein is shown in fig.1 and was deduced from the following results. Treatment with CNBr split the protein into

* Present address: National Institute of Environmental Health Sciences, Research Triangle Park, Durham, NC, 27709, USA

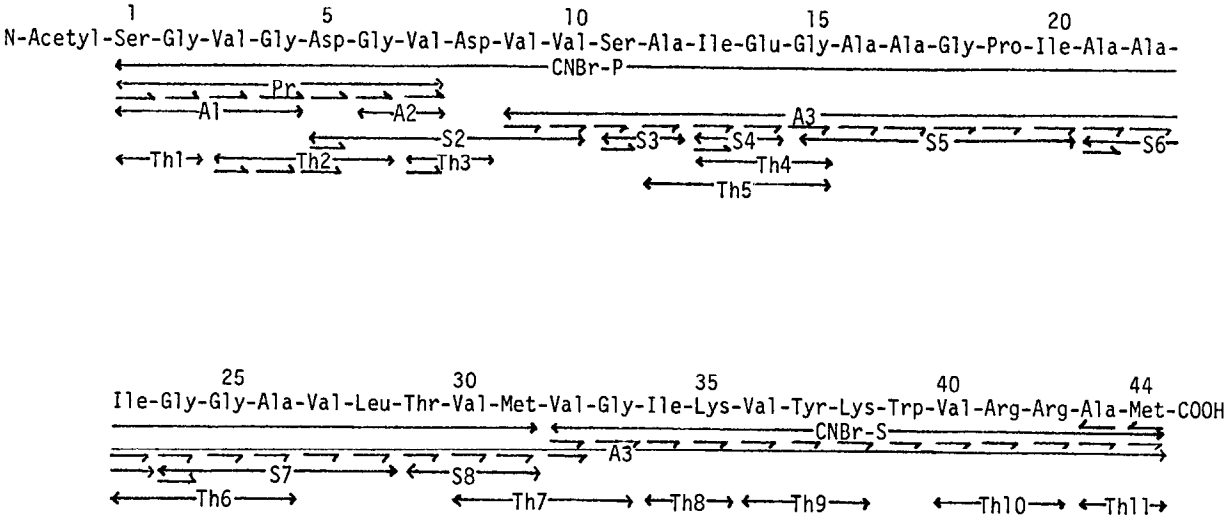

Fig.1. Proposed amino acid sequence of the Xf major coat protein from analysis of CNBr fragments (P and S), fragments from partial acid hydrolysis (A), pronase (Pr), subtilisin (S) and thermolysin (Th) peptides. Sequences: (—) sequences established by automatic sequencing; (——) sequences established by dansyl-Edman degradation and mass spectrometry; (←—) sequences determined by CPA digestion.

two fragments of differing solubility in 10% formic acid, CNBr-S and CNBr-P. The more soluble CNBr-S fragment composed of 13 amino acids was readily sequenced by automated methods. The virus protein contained two methionine residues, one of which was at the carboxy-terminus as shown by the release of alanine (0.3 residues/mol) and methionine (0.5 residues/mol) by CPA digestion of the intact protein. This observation is consistent with the identification of methionine as the carboxy-terminus [3]. The Xf protein was found to have an amino-terminus unreactive to dansyl chloride. Pronase digestion of the intact protein gave rise to a 7 residue peptide (Pr) also with an amino-terminus unreactive to this reagent. A sequence for this peptide was determined by mass spectral analysis and the blocked amino-terminal residue shown to be *N*-acetyl-serine. Partial acid hydrolysis of the protein released two residues of aspartic acid and a fragment (A3) which started at valine (position 9). The results of sequencing this fragment provided the overlap between the CNBr fragments.

The CNBr-P fragment was digested with subtilisin and 7 peptides (S2–S8) identified and ordered in the sequence according to their amino acid composition and amino-terminal residue. Peptide S2 was shown to be an overlapping peptide between the pronase peptide and the major fragment (A3) released by acid hydrolysis. Finally, thermolysin digestion of the intact protein gave 11 peptides of which two (Th2 and Th3) provided an amino acid sequence from valine (position 3) to aspartic acid (position 8). This sequence confirmed the sequence of the Pr peptide and together with the subtilisin peptides, the CNBr fragments and the automated sequencing of the CNBr-S fragment and the largest fragment (A3) from acid hydrolysis enabled a unique sequence to be derived for the Xf coat protein.

4. Discussion

The major coat protein subunit of Xf virus consists of a single chain of 44 amino acids with a mol. wt 4343 calculated from the sum of the constituent residue weights. There are no cysteine or histidine residues in the protein, but there is a high content of alanine, glycine and valine which amount to 24 residues, over half the total number of amino acids. An amino acid content of between 44 and 50 residues

and this type of amino acid composition are general features of the major coat protein subunits of all filamentous viruses.

The sequences of the coat proteins of 5 filamentous viruses are now known: fd, fl [14] and ZJ/2 [15] are either identical or show extremely close homology with each other while Pfl and Xf exhibit major differences. Because of the sequence differences, it is not possible to make an unequivocal case for the evolution of these protein genes from a common ancestral gene. Nevertheless, all the coat proteins examined have the negatively-charged amino acids localized to the amino-terminal region, the hydrophobic residues predominating in the central region and the positively charged residues localized to the carboxy-terminus. This uneven distribution of acidic, basic and hydrophobic amino acids may enable the major coat protein to be sequestered in the inner bacterial membrane prior to virus assembly as found with other filamentous viruses [16,17]. The amino-termini are presumed to be at the outer surface of the virion and the carboxy-termini at the inner surface of the protein shell, with the basic amino acids associating with the central DNA core. However, both the single tryptophan in the Xf subunit (position 39) and the single tryptophan in the fd subunit (position 26) are accessible to *N*-bromosuccinimide ([4], R.L.W. and L.A. Day unpublished results).

The coat protein of Xf is the only example among these proteins in which the amino-terminus is blocked. The proline at the center of the Xf protein (position 19) is also unique. Since the major coat protein subunits in the other viruses are largely α-helical [6,18,19], the position of this proline should restrict α-helix formation and result in a different conformation for the Xf subunit. This is supported by spectroscopic measurements which indicate that in Xf virus, about half of the coat protein residues in the subunit are in the α-helix conformation compared to the almost complete α-helix content exhibited by the protein subunits in Pfl and fd [4].

Physicochemical measurements have demonstrated that the number of nucleotides per major coat protein subunit in Xf is different from Pfl and fd. In the case of Xf, the sequence molecular weight of the major coat protein enables a more accurate number of nucleotides/subunit to be obtained from measurements of g phosphorus/g virus [4,20]. A value of 2.02 ± 0.04 is now calculated instead of an earlier value of 1.92 which was based on a different Xf protein subunit molecular weight [3,4,20]. In this calculation, it is assumed that the minor protein components (~2% by wt) bind to 2% of the DNA. Other arrangements of the minor components can only increase this parameter up to 2.06 ± 0.04. The new value gives support to the contention that the nucleotides per subunit in the Xf virus is an integer value of 2.

The sequence of the Xf coat protein adds to the growing library of filamentous viral coat protein sequences. It is hoped that, when more of these sequences are known, meaningful comparisons can be made which will illuminate possible evolutionary relationships as well as provide additional insights concerning the construction of these virus particles.

Acknowledgements

We thank Mrs Ellen S. Rosenwasser and Mrs Joan Zaretzky for excellent technical assistance and Dr L. A. Day for helpful discussions. This work was supported in part by US Public Health Service Research Grants no. AM 01431, AM 02594, GM 12607 and AI 09049.

References

[1] Kuo, T.-T., Huang, T.-C., Wu, R.-Y. and Yang, C.-M. (1967) Bot. Bull. Acad. Sin. 8, 246–254.
[2] Kuo, T.-T., Huang, T.-C. and Chow, T.-Y. (1969) Virology 39, 548–555.
[3] Lin, J.-Y., Wu, C.-C. and Kue, T.-T. (1971) Virology 45, 38–41.
[4] Day, L. A. and Wiseman, R. L. (1978) in: Single-Stranded DNA Phages (Dressler, D. et al. eds) Cold Spring Harbor Lab., New York, in press.
[5] Marvin, D. A., Wiseman, R. L. and Wachtel, E. J. (1974) J. Mol. Biol. 82, 121–138.
[6] Asbeck, F., Beyreuther, K., Köhler, H., Von Wettstein, G. and Braunitzer, G. (1969) Hoppe-Seyler's Z. Physiol. Chem. 350, 1047–1066.
[7] Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H. and Fischer, D. (1973) Proc. Natl. Acad. Sci. USA 70, 1209–1213.
[8] Nakashima, Y. and Konigsberg, W. (1974) J. Mol. Biol. 88, 598–600.

[9] Nakashima, Y., Wiseman, R. L., Konigsberg, W. and Marvin, D. A. (1975) Nature 253, 68–71.
[10] Methods In Enzymology (1972) (Hirs, C. H. W. and Timasheff, S. N. eds) vol. 25, Academic Press, New York.
[11] Frangione, B. and Milstein, C. (1968) J. Mol. Biol. 33, 893–906.
[12] Hartley, B. S. (1970) Biochem. J. 119, 805–822.
[13] Michaelsen, T. E., Frangione, B. and Franklin, E. C. (1977) J. Biol. Chem. 252, 883–889.
[14] Bailey, G. S., Gillet, D., Hill, D. F. and Petersen, G. B. (1977) J. Biol. Chem. 252, 2218–2225.
[15] Snell, D. T. and Offord, R. E. (1972) Biochem. J. 127, 167–178.
[16] Trenkner, E., Bonhoeffer, F. and Gierer, A. (1967) Biochem. Biophys. Res. Commun. 28, 932–939.
[17] Smilowitz, H., Carson, J. and Robbins, P. W. (1972) J. Supramol. Struct. 1, 8–18.
[18] Day, L. A. (1966) J. Mol. Biol. 15, 395–398.
[19] Day, L. A. (1969) J. Mol. Biol. 39, 265–277.
[20] Wiseman, R. L. and Day, L. A. (1977) J. Mol. Biol. 116, 607–611.