

Theiler's Virus Genome Is Closely Related to That of Encephalomyocarditis Virus, the Prototype Cardiovirus

SIMONA OZDEN, FREDERIC TANGY, MARIO CHAMORRO,† AND MICHEL BRAHIC*

Département de Virologie, Institut Pasteur, 75724 Paris Cedex 15, France

Received 19 May 1986/Accepted 2 September 1986

Theiler's virus causes a persistent demyelinating infection of the mouse central nervous system. Our study of the molecular mechanism of persistence led us to sequence 1925 nucleotides located at the 3' end of the viral genome. We observed extensive homologies between this region and the corresponding region of encephalomyocarditis virus, the prototype cardiovirus, and only some homologies with the 3' ends of foot-and-mouth disease virus, rhinovirus, and poliovirus genomes.

Theiler's murine encephalomyelitis virus (TMEV), a picornavirus, causes asymptomatic enteric infections and neurological diseases in mice (15). Chiefly on this basis, it has been classified in the enterovirus group of picornaviruses (12). When inoculated intracerebrally, TMEV causes two types of central nervous system diseases. The highly virulent GD VII and FA strains are responsible for acute fatal encephalomyelitis, whereas other strains (e.g., TO, DA, and BeAn) cause persistent central nervous system infections accompanied by primary demyelination (4, 10). This last condition closely resembles human multiple sclerosis. Persistent infection of the central nervous system in the face of an immune response, a central question of pathogenesis, seems to result from host-imposed restriction of viral RNA replication in infected glial cells (3). To pursue our analysis of viral RNA metabolism in infected glial cells, we studied the RNA polymerase and protease genes (P3 region) of both virulent (GD VII) and persistent (DA) strains of TMEV. In a preliminary step, we cloned and sequenced ≈ 2 kilobases of RNA located at the 3' extremity of the TMEV strain GD VII genome. We observed that this region presents a high degree of homology with the corresponding region of encephalomyocarditis virus (EMC), the prototype cardiovirus. Some degree of homology with the genomes of representative aphtho- and enteroviruses was also found.

cDNA clones of the TMEV RNA genome were obtained by reverse transcription by using oligo(dT) as a primer (6). Clones corresponding to the 3' region of the genome, identified by hybridization with a [32 P]poly(A) probe, were inserted in the *Pst*I site of pBR322 after oligo(dG):oligo(dC) tailing. One clone (PBT 4) which contained a poly(A) tract at the 3' extremity and 2 kilobases of upstream sequence was subcloned in the *Sma*I site of M13mp10 vector after random shearing (shotgun procedure) (5). Single-stranded templates were sequenced by the dideoxynucleotide procedure of Sanger et al. (13). The sequence of clone PBT 4 was

compiled by using the computer program of Staden (14). On average, each nucleotide was sequenced seven times. The nucleotide and deduced amino acid sequences of clone PBT 4 are presented in Fig. 1. Clone PBT 4 extends 1925 nucleotides away from the 3' end of the viral genome. It contains a single reading frame terminated by a UGA codon 135 nucleotides before the 3' end. The cleavage site that separates the protease (3C) from the RNA polymerase (3D) during maturation of the polyprotein precursor was identified (position 418 in Fig. 1) by analogy with other picornaviruses. In most cases, this cleavage occurs between a glutamine and a glycine residue (8, 11). The identity of this cleavage site is further suggested by the peptide sequence surrounding position 418 of clone PBT 4 (Glu-Pro-Gln/Gly-Ala) which is identical to that surrounding the EMC 3C/3D cleavage site. Therefore, clone PBT 4 covers 417 nucleotides belonging to the protease gene, the entire RNA polymerase gene (1373 nucleotides), and 135 nucleotides of noncoding 3' sequences.

The nucleotide and deduced amino acid sequences of clone PBT 4 were compared by using a computer program (9) with picornavirus sequences (1, 2, 8, 11) available in data banks (Gene Bank, National Biomedical Research Foundation). The results at the amino acid level in the form of a dot plot matrix are shown in Fig. 2. A dot represents four identical consecutive amino acids in a window of five amino acids. Extensive homologies were found between TMEV and EMC sequences (57.7%). Some homologies were observed between TMEV and foot-and-mouth disease virus (A 10) (38%), probably indicating regions essential for the function of the proteins. Limited homologies were also found between TMEV and rhinovirus type 14 sequences (29.3%) and between TMEV and poliovirus type 1 sequences (30%). Hydrophobicity plots (7) further confirmed the relationship of TMEV and EMC RNA polymerase genes (data not shown).

In summary, the sequences of the P3 regions of TMEV, EMC, foot-and-mouth disease virus, rhinovirus, and poliovirus were compared at both the RNA and polypeptide levels. Because of the extensive homologies existing between the sequences of TMEV and EMC, we propose that the taxonomic position of TMEV be reconsidered and that this virus be classified as a cardiovirus.

* Corresponding author.

† Present address: Department of Microbiology, University of California, San Francisco, CA 94143.

ArgSerValAsnArgSerGlyAlaGluThrAspLeuThrPheValLysValThrLysGlyProLeuPheLysAspAsnValAsnLysPheCysSerAsnLysAspAspPheProAlaAr,
GCGUUCGUAUAUCGUCAGGAGCUGAAACCGACCUCUACAUCGUGAAGGUACUAAAGGACCACUCUCAAAGGCAAUGUGAACAAAGUUUUGCUCAAACAGGACGAUUUCCUCUAG
. 100
AsnAspThrCysTyrArgAspIleAsnThrGlyLeuAlaPheValTyrSerGlyAsnPheLeuIleGlyAsnGlnProValAsnThrThrThrGlyAlaCysPheAsnHisCysLeuHis
CAAUGACACCGUGUUAACCGGAUAUGAACACUGGAUUGGCCUUCGUGUAUUCGGUAAUUUCUGAUUGGCAAUCCACUGUGAACACAACAACUGGAGCCUUCACACCACUGCCUCCA
. 200
TyrArgAlaGlnThrArgArgGlyTrpCysGlySerAlaIleIleCysAsnValAsnGlyLysLysAlaValTyrGlyIleHisSerAlaGlyGlyGlyGlyLeuAlaAlaAlaThrIle
CUAUCGAGCUCAAACUCGACGUGGUUGGUGUGUUCUGCCAUCAUCUGCAAUGUUAACGGCAAAAAAGCUGUUUACGGAAUGCACUCUCCUGGAGGCGGAGGCCUUCGCGCCUACCAU
. 300
IleThrArgGluLeuIleGluAlaAlaGluLysSerMetLeuAlaLeuGluProGlnGlyAlaIleValAspIleSerThrGlySerValIleHisValProArgLysThrLysLeuArg
CAUCACCAAGAGUUGAUGAGCAGCUGAGAGUCUAUGUUGGCCUGGAACCGCAAGGUGCCAUUGGACAUUCCACAGGACUUGUGUACAGUCCCAAGAAAGACCAAAACGUGAG
. 400
ArgThrValAlaHisAspValPheGlnProLysPheGluProAlaValLeuSerArgTyrAspProArgThrAspLysAspValAspValAlaPheSerLysHisThrThrAsnIle
GAGAACAGUCGCAUGACGCUUUAACCAACCAAAUUCGAACCCUGCAGUUCUGCCGUAUUGACCCUCCGACCAAGGAUGUAUGUUGGCCUUCUCCAAACACACUACUAAACAU
. 500
GluSerLeuProProIlePheAspIleValCysGlyGluTyrAlaAsnArgValPheThrIleLeuGlyLysAspAsnGlyLeuLeuThrValGluGlnAlaValLeuGlyLeuSerGly
CGAAAGCUUGCCUCAAUCUUUGACAUUGUCUGCGGUAUAACGCUAACCCUGUUUACCAUCCUUGGUAAGACAAACGGGUCUUAACCGUUGAACAGGCGUGCUUGCCUUGCGGG
. 600
MetAspProIleGluLysAspThrSerProGlyLeuProTyrThrGlnGlnGlyLeuArgArgThrAspLeuLeuAspPheAsnThrAlaLysIleThrProGlnLeuAspTyrAlaHis
CAUGGACCCCAUGGAGAAGGACCCUCCCGGAUUGCCUACACUCAACAAGGACUCAGACGAAACUGACUUCUGGAUUUCAACACUGCCAAAAGACACCCCAUUGGACUAGCCCA
. 700
SerLysLeuValLeuGlyValTyrGlyArgArgCysLeuProIleIlePheLysAspGlnIleArgProLeuHisLysIleHisGluAlaLysThrArgIleValAspValProProPhe
UUCCAAACUGGUACUCGCGGUUUAUGGACGAGCUGUUCUACCAAUCAUUUUUAAAGAUCAAUUGCGCCUUGCUAAGAUCCAGGACCAAAAACCCGGAUUGUUGUUGCCCGCUU
. 800
AlaHisCysIleTrpGlyArgGlnLeuLeuGlyArgPheAlaSerLysPheGlnThrLysProGlyLeuGluLeuGlySerAlaIleGlyThrAspProAspValAspTrpThrArgTyr
UGCCCAUCGUAUUUGGGAAGACAGCUUUGGAGCGUUCGCUUCCAAAUUCAAACUAAACUUGGACUUGAUCUGCAAUUGGACUGACCCCGGAUGUUGACUGGACCGGCUA
. 900
AlaAlaGluLeuSerGlyPheAsnTyrValTyrAspValAspTyrSerAsnPheAspAlaSerHisSerThrAlaIlePheGluCysLeuIleAsnAsnPhePheThrGluGlnAsnGly
UGCCGCGAGCUGAGCGGUUCAAACUACGUCUAUGAUGUAGAUUACUCCAAAUUGAUGCUUCCAUUUCUACUGCAAUGUUUGAUGUUUUAACAAUUCUUAACAGACAAAUGG
. 1000
PheAspArgArgIleAlaGluTyrLeuArgSerLeuAlaValSerArgHisAlaTyrGluAspArgArgValTyrSerTrpGlyProAlaSerGlyCysAlaAlaThrSerIleLeuAsn
AUUUCAGACAGCCAUUGCCGAGUACCUAGAUUCUGGUCUGGCGGACAUUGAUGGAGCAGCCGCGUGUAUAGCUGGGGGCCUGCUUGCGGUGGUGUCUACACGAGUUA
. 1100
ThrIleMetAsnAsnValIleIleArgAlaAlaLeuTyrLeuThrTyrSerAsnPheGluPheAspAspIleLysValLeuSerTyrGlyAspAspLeuLeuIleGlyThrAsnTyrGln
CACCAUCAUGAACAAUGUCAUAUUCGUGCUGCCUGUACCUUAUUAUCAAUUUUGAUAUUUGAUGAUAUUUAGGUUUUCCUACGGAGACGUUUUUUUGGAAACUUAUUACCA
. 1200
IleAspPheAsnLeuValLysGluArgLeuAlaProPheGlyTyrLysIleThrProAlaAsnLysThrThrThrPheProLeuThrSerHisLeuGlnAspValThrPheLeuLysArg
AAUUGAUUUUAAUCUUGUUAAGAAAGAUUAGCCCCUUCGCUUUAAGAUAUACUCCUGGCAACAGACCAUACUUUUCUUGACUCCUCCAUUUGCAAGAUUUUACCUUUCUAAAGAG
. 1300
ArgPheValArgPheAsnSerTyrLeuPheArgProGlnIleAspAlaValAsnLeuLysAlaIleValSerTyrCysLysProGlyThrLeuLysGluLysLeuIleSerIleAlaLeu
AAGAUUUGGAGAUUUUUAUUCUACUGUUCAGACCCUCAAUUGGAUGCUGUCAUUUUGAAAGCAAUGGUUAGCUACUGUAACCAAGGAAACAUUAGGCGAAACUUAUGUCCAUUGCUCU
. 1400
LeuAlaValHisSerGlyProAspIleTyrAspGluIlePheLeuProPheArgAsnValGlyIleValValProThrTyrAspSerMetLeuTyrArgTrpLeuSerLeuPheArg***
UCUGCCCGUUAUUCUGGACCAAGAUUUUUAUGAUGAGAUUUUCCUUCUUUUAAGGAUGUUGGAAUUGUUGUCCCAUUGAUAUUGCUUUUAGAGUUGGUUAGCUUUAUUUAGAUG
. 1500
ACAUCUUCUGCAUUGGACCAAGCUUUAACCUAGAACCCACUAGGUGUACGCGCCGCUUCACGUUGGAAUUCUUUUAAGCAAAGUUGUGUAGAUUCUUAUUUGGAAUUGAGAA
. 1600
CAACA poly A

FIG. 1. Nucleotide and deduced amino acid sequences of 1925 bases from the 3' end of the TMEV genome. The proteolytic cleavage site between the 3C and 3D polypeptides is indicated by an arrow at position 418.

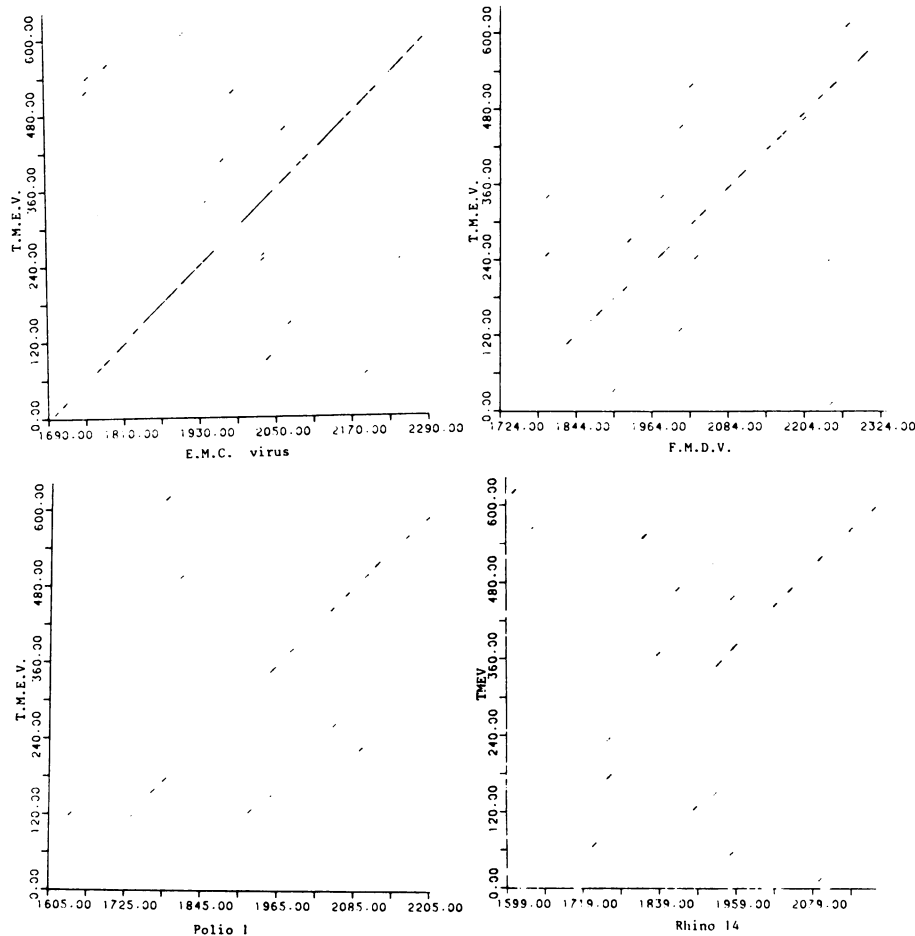


FIG. 2. Comparison of the amino acid sequence of the TMEV 3' region with those of representative picornaviruses by using dot plot matrices. Dots were scored when four consecutive amino acids (in a window of five) were identical. F.M.D.V., Foot-and-mouth disease virus.

We thank P. Sonigo and N. Tordo for their help with sequencing procedures.

This work was supported by grants from the National Multiple Sclerosis Society, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Fondation pour la Recherche Médicale, the Association pour la Recherche sur la Sclérose en Plaques, and the Institut Pasteur Fondation.

LITERATURE CITED

- Callahan, P. L., S. Mizutani, and R. J. Colonno. 1985. Molecular cloning and complete sequence determination of RNA genome of human rhinovirus 14. *Proc. Natl. Acad. Sci. USA* **82**:732-736.
- Carroll, A. R., D. J. Rowlands, and B. E. Clarke. 1984. The complete nucleotide sequence of the RNA coding for the primary translation product of foot and mouth disease virus. *Nucleic Acids Res.* **12**:2461-2472.
- Cash, E., M. Chamorro, and M. Brahic. 1985. Theiler's virus RNA and protein synthesis in the central nervous system of demyelinating mice. *Virology* **144**:290-294.
- Daniels, J. B., A. M. Pappenheimer, and S. Richardson. 1952. Observation on encephalomyelitis of mice (DA strain). *J. Exp. Med.* **96**:517-535.
- Deininger, P. L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* **129**:216-223.
- Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* **25**:263-269.
- Hopp, T. P., and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**:3824-3828.
- Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. Van der Werf, C. W. Anderson, and E. Wimmer. 1981. Primary structure, gene organisation and polypeptide expression of poliovirus RNA. *Nature (London)* **291**:547-553.
- Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* **227**:1435-1440.
- Lipton, H. L. 1975. Theiler's virus infection in mice: an unusual biphasic disease process leading to demyelination. *Infect. Immun.* **11**:1147-1155.
- Palmenberg, A. C., E. M. Kirby, M. R. Janda, N. L. Drake, G. M. Duke, K. F. Potratz, and M. S. Collett. 1984. The nucleotide and deduced amino acid sequences of the encephalomyocarditis viral protein coding region. *Nucleic Acids Res.* **12**:2969-2985.
- Rueckert, R. R. 1985. Picornaviruses and their replication, p. 705-738. *In* B. N. Fields (ed.), *Virology*. Raven Press, New York.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Staden, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res.* **10**:4731-4751.
- Theiler, M. 1937. Spontaneous encephalomyelitis of mice, a new virus disease. *J. Exp. Med.* **65**:709-719.