

## Letter to the Editor

# Nucleotide Sequence Between the Peplomer and Matrix Protein Genes of the Porcine Transmissible Gastroenteritis Coronavirus Identifies Three Large Open Reading Frames

PAUL A. KAPKE<sup>1</sup>, FRANK Y.T. TUNG<sup>2</sup>, AND DAVID A. BRIAN<sup>2</sup>

<sup>1</sup>*United States Department of Agriculture, Agricultural Research Service, National Animal Disease Center, Ames, IA 50010, USA*

<sup>2</sup>*Department of Microbiology, The University of Tennessee, Knoxville, TN 37996-0845, USA*

Received November 9, 1988

Accepted November 10, 1988

Requests for reprints should be addressed to Paul A. Kapke, U.S. Department of Agriculture, Agricultural Research Service, National Animal Disease Center, Ames, IA, 50010, USA

The nucleotide sequence between the peplomer and matrix protein genes in the genome of the Purdue strain of porcine transmissible gastroenteritis coronavirus (TGEV) was determined by sequencing parts of six cDNA clones. Open reading frames potentially encoding proteins of 7,711, 27,711, and 9,241 Da were identified (Fig. 1). The sequence for this region of the genome for the same strain of virus was published by Rasschaert et al. (3), but our sequence differs by two bases, one of which results in a major change in the properties of the second open reading frame (ORF). G in our sequence at position 433, rather than T, enlarges the second open reading frame from 165 to 244 amino acids and establishes a sequence context more favorable for initiation of translation (4). C in our sequence at position 606, rather than T, changes Leu to Pro. Each open reading frame is preceded by a sequence that is similar to the CTAAAC intergenic sequence thought to be required for leader-primed transcription (1,3). The enlargement of the second open reading frame from 165 amino acids (18,833 Da) to 244 amino acids (27,711 Da) resolves two concerns raised by Rasschaert et al.(3) namely a) the second ORF in their sequence would require initiation of translation 570 bases downstream from the CTAAAC intergenic sequence (or 249 bases downstream of the CTAAAT sequence that we propose is used), which is an unusually long distance. b) The second ORF in their sequence is not large enough to encode the 24 kD polypeptide translated *in vitro* from TGEV mRNA 3 by Jacobs et al. (5), an mRNA

```

1 TAAATTTAAATGTTAATCTCATCTCGTATAATAGCAGTTGTTTCTGCTAGAGAATTTTGTAAAGTAGTGAATAAAGCTTTAAGAACTAAACTT
101 ACGAGTCATTACAGGTCCTGTATGGACATTTGTCAAATCCAATTACACATCCGTAGATGCTGACTTGACGCACTTGATTTGGCATGCTTTGGTGAAGTC
M D I V K S I Y Y S V D A V L D E L D C A Y F A V Y F
201 TTAAGTAGAATTAAGCACTGGTAAATTAATCTGTGTAGGTTTGGTGACACACTTCTGCTGCTAAGGATAAAGCATATGCTAAGCTTGGTCTCTC
L K V E F K T V G K L L V G C I G F G D Y L L L A A K D K A Y A R L G L L S
301 CATTAITGAGAAGTCAATAGTCATAGTGTGTTAATATCATTAAACACAAAAACCCAAAGCATTAAAGTGTACAAAACAATTAAGAGAGATTATAG
I I E E V N S H I V V
401 AAAAAGTCATTCTAAATTCATCGGAAAAATGATGGTGGACTTTTCTTAGTACTCTGAGTTTGTAAITGTTAGTAAACCATCTATGTTAATAACA
M A G I G G L F L S F F V I V S N H S I V N H
501 CAGCAATGTGCATCATATACAACAAGACGTTGTATAGTACAACAGCATCAGGTTGTAGTGTAGAACAACAACTATTCCGAGAGTTCAGCATCCG
Y A N V H H I G Q E R V I V Q D H Q V V S A R F D N Y A P E F C A G C
601 TGTACCCTTGTATCTTTCTAGCTTTGTACCGTAGTACAACCTTAAAGAGCTGTGCGGCATCTAAATGTTAAGATTTTATCAATGACACTTTTAGGA
P M L Y A Y I A Y R S Y N F K T E V G C I T A A T G I T A A G A T T T A T C A A T G A C A C T I T T A G G A
701 CCTATGCTTATAGCATATGGTACTACATTTGATGGATTTGTAACAACACTGCTTATCTTTAAGATTTGCTACTTAGCATACTTTTGGTATGTAATA
P M L Y A Y I A Y R S Y N F K T E V G C I T A A T G I T A A G A T T T A T C A A T G A C A C T I T T A G G A
801 GTAGGTTGAATTTATTTATACAATAACAACGACACTCATGTTTGTACATGGCAGAGCTGCACCGTTATGAGAAGTTCACAGCTCTATTTATGTCAC
S R F E F I A Y I L Y N I T Y L M F V H G R A A P F M R S S H S S I Y V T
901 ATTGTATGGTGGCATAAATATATGTTTGTGAATGACCTCAGGTTGCATTTGTAGACCTATGCTTGTAAAGCATAGCAATACGTTGGCTTAGCTCATGCT
L Y T G E F I A Y I L Y N I T Y L M F V H G R A A P F M R S S H S S I Y V T
1001 GATCTAAGTGTAGTACAGCAGTGAACCTCTCAATGGTGATTTATTTATGATTTTACAGGAGCCGCTAGTCCGTTTACAAATGCAGCTTTTCTC
D L Y V V R A V E L L N G D F F I T T A T T A T T A C A G G A G C C G T A G T C C G T T T A C A A T G C A G C T I T T C T C
1101 AGGCGGTTCTAAACGAAATGACTTAAAGAAGAGAGAACCACTAGCTATGACGTTTCTAGGCGATTGACTGTCTAGATGACAATGGAATGGTCA
Q A V L W E I D L K E E E D H Y M T F P R A L T V I D D N G M V
1201 TTAACATCATTTCTGGTTCCTGTGTAATTAATGATATTACTTTCAATAGCATTGCTAAATAATTAAGCTATGCATGGTGTGTTGCAATTAGG
I N I L F M F L L I I L L L S I A L L N I I R L C M V C C N L G
1301 AAGGACAGTATTATTTGTCAGCCCAACATGCTTACGATGCTATAAGAATTTATGCGAATTAAGCATACAACCCGATGGAGCACTCTGCTTGA
R V I I V P A Q H A Y D A Y K N F M R I K A Y N P D G A L L A
1401 ACTAAACAAAATG
M

```

Fig. 1. Nucleotide sequence between the peplomer and matrix protein genes of TGEV and the deduced amino acid sequences for the three large open reading frames. The nucleotide sequence begins with the TAA stop codon of the peplomer gene (1, and unpublished data) and ends with the ATG start codon of the matrix protein (2). The consensus intergenic sequences are underlined.

that maps in this region of the genome. The base differences we report were obtained from two separate cDNA clones.

### Acknowledgments

This work was supported by grant AI-14367 from the National Institute of Allergy and Infectious Disease and by grant 82-CRSR-2-1090 from the U.S. Department of Agriculture.

### References

1. Raschaert D. and Laude H., *J Gen Virol* 68, 1883-1890, 1987.
2. Kapke P.A., Tung F.T.Y., Hogue B.G., Brian D.A., Woods R.D. and Wesley R.D., *Virology* 165, 367-376, 1988.
3. Rasschaert D., Gelfi J. and Laude H., *Biochemie* 69, 591-600, 1987.
4. Kozak M., *Microbiol Rev* 47, 1-45, 1983.
5. Jacobs J., Van der Zeijst B.A.M. and Horzinek M.C., *J Virol* 57, 1010-1015.