Sequence Determination of the Extreme 5' End of Equine Arteritis Virus Leader Region

ALI KHEYAR, GILLES ST-LAURENT, & DENIS ARCHAMBAULT

Université du Québec à Montréal, Département des Sciences Biologiques, Montréal, Québec, Canada, H3C 3P8

Abstract. The extreme 5' end of the leader sequence of four equine arteritis virus (EAV) strains was obtained by using rapid amplification of cDNA end method (5' RACE), and sequenced. Seventeen more nucleotides were added upstream of the 5' end of the EAV published genomic sequence. A common feature among the analyzed EAV isolates was the presence of an AUG start codon within the added sequence and the appearance of an intraleader open reading frame (ORF) of 111 nucleotides which was predicted to encode a peptide of 37 amino acids. The role of this putative intraleader ORF has yet to be determined.

Key words: equine arteritis virus, 5' terminus, intraleader ORF

Equine arteritis virus (EAV) is the etiologic agent of equine viral arteritis, a debilitating respiratory disease with the most severe form resulting in abortion from pregnant mares (1). Although the virus transmission is primarily via the respiratory route, the virus is also shed in the semen of persistently infected stallions which infect mares at the time of breeding (1). EAV is the prototype species of the Arterivirus group, which includes lactate dehydrogenase-elevating virus (LDV), porcine reproductive and respiratory syndrome virus (PRRSV) and simian hemorrhagic fever virus (SHFV) (2). The EAV genome is a positive, polyadenylated, single stranded RNA of approximately 12.7 kb (3). The replication strategy of EAV resembles that of corona- and toroviruses. During virus replication, a 3' end coterminal nested set of seven virus-specific RNAs is produced with a common leader sequence derived from the 5' end of the EAV genome (4). The leader sequence is joined to each open reading frame (ORF) by a junction sequence motif 5' UCAAC 3' (14). On the basis of primer-extension experiments, the EAV leader sequence has been predicted to be 207 (3) or 208 (4) nucleotides (nt) in length. However, sequence data obtained from cDNA clones derived from a genomic EAV cDNA library (3) failed to identify the most extreme 5'-terminal nt of the leader sequence.

In this study, the extreme 5' end of the EAV leader sequence of four EAV strains is described, with a prediction of an intraleader open reading frame (ORF) encoding a short peptide of 37 amino acids within the full length leader sequence. To this end, the EAV Bucyrus reference strain (5), and the EAV laboratory 87AR-A1, 86NY-A1 and Vienna strains (6), with low cell passage levels ranging from two to four, were used. The Bucyrus strain was isolated from tissues of a fetus aborted during an abortion endemic episode in standardbreed horses (5), while the others were isolated from nasal swab (Vienna) or semen of infected horses (87AR-A1, 86NY-A1) (6). Each virus strain was plaque-purified and propagated for one to two additional passages in rabbit kidney (RK-13) cells (7). After three freeze-thaw cycles, cell culture supernatant was centrifuged at 5000 × g for clarification. Virion RNA was extracted from 300 μl of infected cell culture supernatant by the guanidium isothiocyanate method (8). The RNA sample was resuspended in 20 μl of diethylpyrocarbonate (DEPC)-treated water containing 15 units of hu-
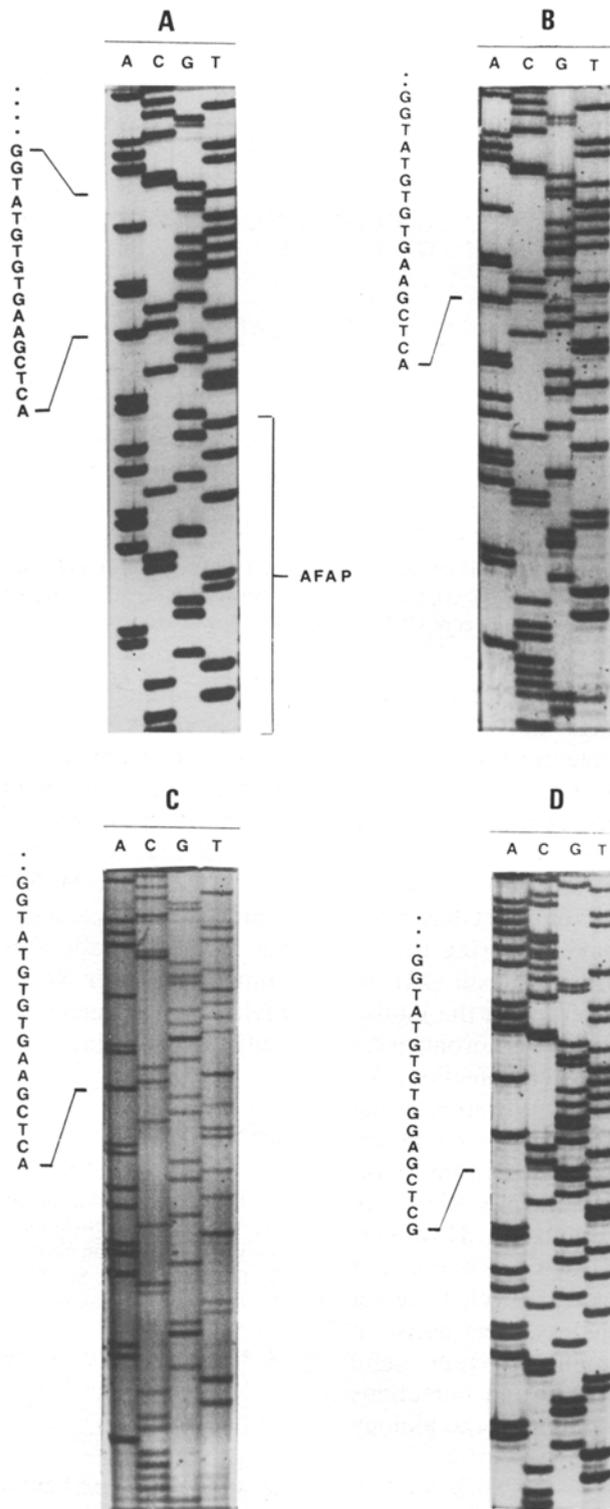
man placental RNase inhibitor (Pharmacia), and kept at $-70°C$ until used.

To determine the 5′ end of the EAV leader sequence, the single strand ligation to single-stranded cDNA (SLIC) method (9) was used. For this purpose, the 5′-Amplifinder RACE Kit (Clontech) was used to amplify the 5′ distal end of the EAV leader sequence. Briefly, purified EAV RNA was preincubated at 65°C for 5 min and reverse transcribed with avian myeloblastosis virus (AMV) reverse transcriptase (RT) using the cDNA synthesis procedure described in the manufacturer's protocol. The RT reaction was primed with 20 pmol of the antisense oligonucleotide primer PEV-L1 (5′GTGGAGCCGTC CACTTC 3′) which is complementary to nt 306 to 323 of the EAV genome (3) (Genebank accession number X 53459). This primer sequence is located downstream from the 3′ end of the leader sequence in the ORF 1a genomic region. After RNA hydrolysis, the first single-stranded cDNA (ss-cDNA) was purified using the GENO-BIND™ kit (Clontech). The 3′ end blocked AmpliFINDER anchor sequence was ligated to the 3′ end of ss-cDNA at room temperature for 18 h with T4 RNA ligase. The single-stranded ligation product was then used in the polymerase chain reaction (PCR) procedure. The oligonucleotide primers used in the PCR consisted of antisense primer PEV-L02 (5′ACCCGTCAAGCCA CAAGATG 3′), which is complementary to an internal sequence (nt 165 to 145) of the EAV leader sequence (3), and the sense Ampli-FINDER anchor primer (AFAP). The PCR reaction was then performed using a thermal cycler programmed for 30 successive cycles at conditions recommended by the supplier. The PCR products were then cloned into the *Sma* 1 cleaved pBluescript II KS+ plasmid vector (Stratagene) and sequenced by the dideoxynucleotide chain-termination method (10).

Sequence analysis of the EAV leader sequence region of 3 clones of the Bucyrus strain revealed the presence of 17 additional nt (5′ ACT CGAAGTGTGTATGG 3′) absent from the published sequence (3) and located at the 5′ end (Fig. 1A). The same additional 17 nt were also obtained from several clones of the 87AR-A1, and Vienna EAV strains (Fig. 1B and C). The results obtained with the 86NY-A1 strain showed a se-

quence similar to the other EAV strains we studied with the exception of two substitutions $(G \rightarrow A)$ at positions $+1$ and $+7$ (Fig. 1D). Sequencing of PCR products from the four virus strains analyzed confirmed that the missing most extreme 5′-terminal nt were of EAV origin, since it was part of the nt sequence of the EAV genome according to the published data (3). With the addition of these 17 nucleotides, the length of the EAV leader sequence would be 206 nt without the junction site motif (Fig. 2). Thus, the EAV leader sequence is identical in length to that reported for PRRSV (11) and is longer than that of LDV (156 nt) (12) or SHFV (202 nt) (13). However, if a cap structure exists, thus an additional G would represent the 207th base, as already suggested (3). In any case, EAV and LDV (14) are, at present, the only two arteriviruses for which the entire leader sequence is determined, in comparison to the PRRSV genome (11) and SHFV genome (13), for which the first two 5′ nucleotides of the leader sequence remain to be determined.

An important observation found in this study was the presence of a unique AUG initiator codon located at nucleotide positions $+14$ to $+16$. The optimal sequence context for efficient recognition as an internal initiation signal for RNA translation was determined to be

$$\overset{-3}{C} \quad C \quad A/G \quad C \quad C \quad \underline{A} \quad \underline{U} \quad \underline{G} \quad \overset{+4}{G}$$

with the -3 and $+4$ nucleotide positions being of primary importance (14). On the basis of the later definition, the initiator codon is designated as strong or weak depending on the presence or absence of these residues at these positions (15). The -3 and $+4$ nucleotide positions of the EAV leader sequence are occupied by U and G, respectively. This suggests that the EAV intraleader AUG might be a weak initiator codon and therefore may provide a suboptimal context for translation initiation (14). Moreover, a UAG terminator codon occurs at positions $+125$ to $+127$ of the EAV published genomic sequence (3), thereby showing the presence of an intraleader ORF. This intraleader ORF sequence of 111 nt in length is predicted to encode a short peptide of 37 amino acids (Fig. 2). The estimated molecular mass of this peptide is 4.6 kDa.

Fig. 1. Sequence determination of the extreme 5' end of the leader region of various EAV isolates. A: Bucyrus reference strain; B: 87AR-A1 strain; C: Vienna strain; D: 86NY-A1 strain. Seventeen additional residues (5' ACTCGAAGTGTGTA TGG 3') are present at the 5' end of the EAV Bucyrus strain genome. The assigned GenBank Database accession numbers are U46944 (Bucyrus), U46945 (86NY-A1), U46946 (87AR-A1), and U46947 (Vienna).

```
          *              *              *              *
  1 ACTCGAAGTGTGTATGGTGCCATATACGGCTCACCACCATATACACTGCA
          M  V  P  Y  T  A  H  H  H  I  H  C  K


          *              *              *              *
 51 AGAATTACTATTCTTGTGGGCCCCTCTCGGTAAATCCTAGAGGGCTTTCC
     N  Y  Y  S  C  G  P  L  S  V  N  P  R  G  L  S


                  *              *              *         *
101 TCTCGTTATTGCGAGATTCGTCGTTAGATAACGGCAAGTTCCCTTTCTTA
     S  R  Y  C  E  I  R  R  .


          *              *              *              *
151 CTATCCTATTTTCATCTTGTGGCTTGACGGGTCACTGCCATCGTCGTCGA


201 TCTCTA
```

Fig. 2. Complete sequence of the EAV Bucyrus strain leader region. The 5' end 17 nt added to the previously published sequence (3) are underlined. The intraleader ORF sequence is located between ATG start and TAG stop codons (bolded and underlined). The corresponding putative intraleader ORF-encoded amino acid sequence (bottom line) is also shown.

It is well known that members of both coronaviruses and arteriviruses cause persistent infections in their respective host (16–18). It has been reported that short ORFs within the 5' leader region of some eukaryotic mRNAs attenuate the rate of translation initiation at the downstream ORF (19). In fact, a translation-attenuating intraleader ORF of 33 nt has been described in bovine coronavirus during persistent infection (18). In contrast, a small ORF of 18 amino acids that was observed in the mouse hepatitis virus (which is also a coronavirus) leader sequence during a persistent infection, has been demonstrated to enhance translation of the downstream ORF (20). It is thus possible, by analogy with these virus systems, that the intraleader ORF could be involved in EAV RNA replication and/or translation regulation. However, the presence and role of such an intraleader ORF-encoded peptide in EAV life cycle have yet to be determined. Nevertheless, determination of the complete 5' end sequence provide useful information in construction of an infectious cDNA for a better understanding of the biology of EAV.

Acknowledgments

References

1. Timoney P.J. and McCollum W.H., Vet Clinics North Amer:Equine Practice *9*, 295–309, 1993.
2. Virology Division News., Arch Virol *135*, 227–237, 1994.
3. Den Boon J.A., Snijder E.J., Chirnside E.D., De Vries A.A.F., Horzinek M.C., and Spaan W.J.M., J Virol *65*, 2910–2920, 1991.
4. De Vries A.A.F., Chirnside E.D., Bredenbeek P.J., Gravestein L.A., Horzinek M.C., Spaan W.J.M., Nucleic Acids Res *18*, 3241–3247, 1990.
5. Doll E.R., Knappenberger R.E., and Bryans J.T., Cornell Vet *47*, 69–75, 1957.
6. Murphy T.W., McCollum W.H., Timoney P.J., Klingeborn B.W., Hyllseth B., Golnik W., and Erasmus B., Vet Microbiol *32*, 101–115, 1992.
7. Saint-Laurent G., Morin G., and Archambault D., J Clin Microbiol *32*, 658–665, 1994.
8. Chomzynski P. and Sacchi N., Anal Biochem *162*, 156–159, 1987.

9. Apte A.N. and Siebert P.D., BioTechniques *15*, 890–893, 1993.

10. Sanger F., Nicklen S., and Coulson A.R., Proc Natl Acad Sci (USA), *74*, 5463–5467, 1977.

11. Conzelmann K.-K., Visser N., Van Woensel P., and Thiel H.J., Virology *193*, 329–339, 1993.

12. Chen Z., Faaberg K.S., and Plagemann P.G.W., J Gen Virol *75*, 925–930, 1994.

13. Zeng L., Godeny E.K., Methven S.L., and Brinton M.A., Virology *207*, 543–548, 1995.

14. Kozak M., J Mol Biol *196*, 947–950, 1987.

15. Kozak M., J Cell Biol *108*, 229–241, 1989.

16. Plagemann P.G.W. and Moennig V., Adv Virus Res *41*, 99–192, 1992.

17. Perlman S., Jakobsen G., Olsen A.L., and Afifi A., Virology, *175*, 418–426, 1990.

18. Hofmann M.A., Savithra D.S., and Brian D., Proc Natl Acad Sci (USA) *90*, 11733–11737, 1993.

19. Kozak M., J Cell Biol *115*, 887–903, 1991.

20. Chen W. and Baric R.S., J Virol *69*, 7529–7540, 1995.