PROTEOLYTIC PROCESSING OF THE N-TERMINAL REGION OF THE EQUINE ARTERITIS VIRUS REPLICASE

Eric J. Snijder, Alfred L.M. Wassenaar, and Willy J.M. Spaan

Department of Virology, Institute of Medical Microbiology
Faculty of Medicine, Leiden University
Postbus 320, 2300 AH Leiden, The Netherlands

ABSTRACT

A papainlike cysteine protease (PCP) domain in the N-terminal region of the equine arteritis virus (EAV) replicase was identified by *in vitro* translation and mutagenesis studies. The EAV protease was found to direct an autoproteolytic cleavage at its C-terminus which leads to the production of an approximately 30K N-terminal replicase product (nsp1) containing the PCP domain. Amino acid residues Cys^{164} and His^{230} of the EAV replicase polyprotein were identified as the most likely candidates for the role of PCP catalytic residues. It was shown that cleavage occurs *in cis* between Gly^{260} and Gly^{261}.

INTRODUCTION

Equine arteritis virus (EAV) is an enveloped positive-stranded RNA virus. Its isometric nucleocapsid core contains a nonsegmented 12.7 kb genome (for a recent review: see reference 1). The morphological characteristics and genome size of EAV are most comparable to those of togaviruses. However, we have recently described[2,3] that the EAV replication strategy is similar to that of coronaviruses (for a review: see reference 4) and toroviruses[5,6], which possess 25-31 kb positive-stranded RNA genomes. Among their common features are a polycistronic genome organization, the same basic gene order (5'-replicase gene-envelope protein genes-nucleocapsid protein gene-3'), and the production of a 3'-coterminal nested set of 4 to 7 subgenomic mRNAs. The 5' part of the genomes of these viruses is occupied by two large open reading frames (ORF1a and ORF1b) which encode the viral replicase[2,6,7,8,9]. Both ORF1a and ORF1b are expressed from the genomic RNA, the latter by means of a ribosomal frameshifting mechanism[2,6,10]. The predicted

Coronaviruses, Edited by H. Laude and J.F. Vautherot
Plenum Press, New York, 1994

ORF1b products of corona-, toro-, and arteriviruses contain a number of homologous protein domains[2,6,8] which, in addition to the other similarities described above, indicate that these viruses are evolutionarily related. We have therefore proposed these viruses to be members of a coronaviruslike superfamily of positive-stranded RNA viruses[2,6]. The large EAV replicase gene product (345K) is presumed to be a polyprotein precursor which is posttranslationally cleaved into smaller functional units. In the 187K ORF1a amino acid sequence putative papainlike cysteine protease and trypsinlike serine protease domains were identified[2] (Fig. 1). We have now initiated a study of the coronaviruslike replicase using the relatively small EAV replicase gene as a model.

RESULTS

Reconstruction and *in vitro* translation of EAV ORF1a

The analysis of the posttranslational processing of the EAV replicase polyprotein was started by reconstructing ORF1a from seven overlapping cDNA clones. In transcription vector pEAV1a a full-length cDNA copy of ORF1a is located downstream of the T7 RNA polymerase promoter. In addition, a set of pEAV1a deletion mutants (p1aΔ1 through p1aΔ8) was produced which contained termination codons for translation at various positions in ORF1a. To create termination codons for translation, *Nhe*I linkers (5' CTAGCTAGCTAG 3') were inserted into the following pEAV1a restriction sites: *Sac*I (nt 858), *Hin*dIII (nt 1501), *Kpn*I (nt 1802), *Sal*I (nt 2608), *Nhe*I (nt 2878), *Apa*I (nt 3688), *Eco*RV (nt 4263), and *Bam*HI (nt 5115). Nucleotide (nt) and amino acid (aa) sequence numbers refer to the EAV genomic and protein sequences which we have published previously[2].

Vector pEAV1a and the p1aΔ*n* constructs were used to transcribe a set of 3'-truncated RNAs from which an increasing part of ORF1a could be translated. *In vitro* translation was carried out in a rabbit reticulocyte lysate (Promega) in the presence of [^{35}S]methionine (1 h at 30°C). A direct SDS-PAGE analysis of the translation products of the p1aΔ*n* series and pEAV1a is shown in Fig. 1. Only p1aΔ1, which contained a termination codon in the center of the putative PCP domain, produced a protein of the predicted size. All other constructs gave rise to a prominent product of about 30K and to accompanying bands which were smaller than predicted from the (partial) ORF1a amino acid sequence. These data indicated that an approximately 30K protein was cleaved from the N-terminus of the EAV ORF1a product. The putative PCP domain, which resided in this cleavage product, could be involved in this proteolytic event.

Analysis and mutagenesis of the EAV PCP domain

Typical papainlike cysteine proteases show a requirement for at least one cysteine and one histidine residue[11]. On the basis of amino acid sequence comparison of the ORF1a sequence and cellular and viral papainlike thiol proteases, Cys^{164} had been proposed as active site residues of an EAV papainlike cysteine protease[2]. His^{219} and His^{230} were considered the most likely candidates for the role of catalytic His. To prove that the EAV PCP domain was responsible for the observed proteolytic processing of the ORF1a protein, amino acid substitutions were introduced into the ORF1a sequence. Transcription vector pCP0 (Fig. 2), encoding a 46K product which is cleaved into 30K and 16K polypeptides, was used as a basis for these experiments. Derivatives of this construct were used to test

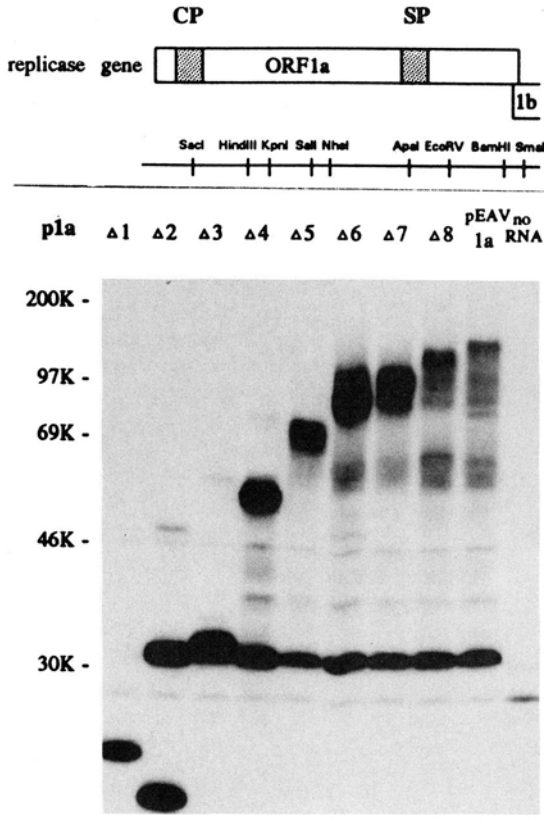

Fig. 1. *In vitro* translation results from EAV ORF1a expression constructs.
A schematic representation of the ORF1a region of the EAV replicase gene is shown. The positions of the predicted papainlike cysteine protease (CP) and trypsinlike serine protease (SP) domains are indicated. The restriction sites were used in the construction of the expression plasmids p1aΔ1 through p1aΔ8 and pEAV1a. Plasmids p1aΔ1 through p1aΔ8 and pEAV1a were used for *in vitro* transcription of a 3'-truncated set of RNAs from which an increasing part of ORF1a could be translated. The RNAs were translated in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine and direct analysis of translation products was performed by SDS-PAGE.

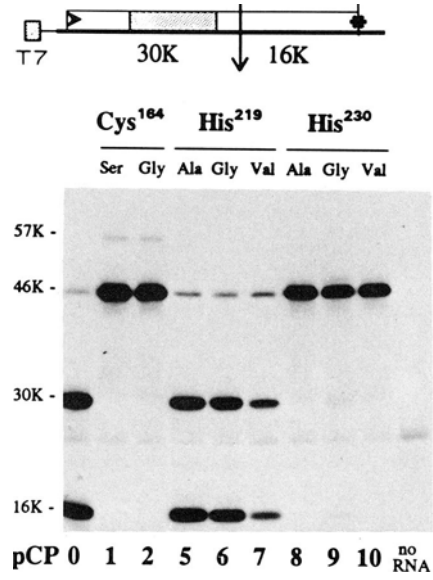

Fig. 2. Identification of possible active site residues of the EAV papainlike cysteine protease.
Amino acid substitutions were introduced into construct pCP0 (p1aΔ2 from Fig. 1) and tested by *in vitro* translation and SDS-PAGE. Mutations are indicated at the top of each lane; the name of the corresponding expression plasmid is shown at the bottom of the lane. The 46K full-length translation product and its 30K and 16K cleavage products are indicated.

the influence of substitutions at the positions of Cys^{164}, His^{219}, and His^{230}. Nucleotide changes were introduced by oligonucleotide-directed mutagenesis and mutations were tested by *in vitro* transcription and translation (Fig. 2).

Both the rather conservative substitution of Cys^{164} by Ser as well as the Cys^{164} to Gly mutation completely abolished proteolytic activity, indicating that this Cys residue is indeed essential for the protease function. Replacing His^{219} by either Ala, Gly, or Val did not affect cleavage to a significant extent. In contrast, the same set of substitutions at the position of His^{230} exhibited an effect similar to that observed after replacing Cys^{164}: the His^{230}-Val substitution completely inhibited proteolytic activity, and only traces of cleavage products could be detected after translation of the Ala^{230} and Gly^{230} mutants. These results confirmed that the 5' region of the EAV genome encodes a proteolytic domain which is responsible for the observed cleavage event.

Activity of the EAV PCP domain in *E. coli* and identification of the PCP cleavage site

The bacterial expression vector pGEX-2T (Pharmacia) was used to express the N-terminus of the EAV ORF1a polypeptide as part of a bacterial fusion protein. The 86K fusion protein contained a pGEX-derived 26K glutathione *S*-transferase (GST) moiety, followed by 9 aa encoded by the EAV 5' untranslated region, the 528 N-terminal aa of the ORF1a product (including the PCP domain), and 33 aa encoded by a short in-frame vector sequence. Remarkably, the construct did not only produce an 86K fusion protein: it also yielded 55K and 32K bands. In view of the data obtained with the p1aΔn series (see above) and the 26K size of the GST part of the fusion protein, it was concluded that the EAV PCP domain was functional in *E. coli*. This was confirmed by the introduction of mutations into the PCP domain of the fusion protein.

The 55K bacterial expression product represented the N-terminal cleavage product consisting of the GST part and the previously observed 30K EAV ORF1a protein. The amount of the 32K C-terminal cleavage product was sufficient to allow purification for N-terminal microsequencing. The following amino acid residues were found to constitute the N-terminus of the 32K cleavage product: Gly-Tyr-Asn-Pro-Pro-Gly-Asp-Gly-Ala. This sequence is present at aa position 261-269 in the EAV ORF1a protein, indicating that cleavage takes place between Gly^{260} and Gly^{261} in the EAV sequence.

The EAV PCP is a *cis*-acting protease

To test whether the EAV PCP can function *in trans*, the protease domain and a number of the mutants described above were used in *trans*-cleavage assays. To test posttranslational *trans*-cleavage, PCP-containing proteins were prepared by *in vitro* translation using unlabeled methionine, while labeled (uncleaved) substrates were prepared by translation of mutant constructs in the presence of [^{35}S]methionine. Translation reactions which contained substrates were mixed with equal volumes of protease-containing reactions. The 46K translation products derived from pCP1 and pCP10 (carrying the Cys^{164}-Ser and His^{230}-Val substitutions, respectively; see also Fig. 2) were used as substrates. They contained normal EAV cleavage sites (Gly^{260}-Gly^{261}), which remained unprocessed due to mutations in the protease domains. Though mixtures of substrates and proteases were incubated at 30°C for up to 7 h, both nsp1 and its precursor were unable to produce detectable amounts of 30K and 16K cleavage products. The possibility of cotranslational *trans*-cleavage was excluded, by mixing transcripts encoding substrates and proteases and translating them in the same reaction.

DISCUSSION

Although the available coronaviruslike replicase sequences have been searched for possible protease domains extensively, - even their cleavage sites have already been predicted[9,12,13] - there is little experimental evidence which supports these theoretical analyses. The presence of a number of ORF1a- and ORF1b-encoded proteins in MHV-infected cells has been reported[12,14]. However, coronavirus proteases have not yet been studied in detail, and replicase cleavage sites and processing pathways remain to be elucidated.

In this paper the first detailed analysis of a protease domain which is located in a coronaviruslike replicase is reported. The identification of a Cys and a downstream His as putative active site residues indicates that this EAV protease is related to the cellular papainlike enzymes, and can therefore be added to the short but growing list of viral papainlike cysteine proteases. In a recent review, Gorbalenya *et al.*[15] discriminated between a group of viral PCPs which mediate the production of a single N-terminal cleavage product (the so-called 'leader proteases') and PCPs which are thought to be 'main' proteases involved in multiple processing steps. Of the latter group only the PCP domain residing in the Sindbis virus (SIN) nsp2 has actually been shown to possess proteolytic activity[16]; it is responsible for the production of the non-structural SIN proteins from a polyprotein precursor.

The EAV PCP clearly belongs to the group of leader proteases, which (among others) includes the potyvirus helper component protease[17,18] (HC-Pro) and the p29 protease of the hypovirulence-associated virus (HAV) of the chestnut blight fungus[19,20]. Although the overall sequence similarity between the EAV domain and these two well-studied leader PCPs is very limited, some striking similarities can be observed. The spacing between the putative catalytic Cys and His residues is similar and clearly different from the spacing in cellular papainlike proteases. In addition, all three proteases efficiently cleave a Gly-Gly dipeptide which is located 30-41 aa downstream of the putative active site His residue.

We have failed to detect other cleavage events after *in vitro* translation of ORF1a-specific RNAs. This is surprising since, in addition to the PCP domain, the larger constructs from the p1aΔn series (p1aΔ7 and p1aΔ8) and the full-length ORF1a construct contain the coding information for the putative EAV serine protease[2] (aa position 1090-1210). Although several additional bands were observed (Fig. 1), none of them can be explained as the obvious result of proteolytic processing.

We are currently preparing antibodies specific for ORF1a-encoded proteins, which will enable us to study the *in vivo* processing of the ORF1a product. Using an anti-peptide rabbit antiserum directed against the N-terminal part of the ORF1a product, we have recently identified the 30K EAV nsp1 in infected cells. This may enable us to study the function of nsp1 which contains a cysteine-rich N-terminal region, in addition to the C-terminal PCP domain. Also EAV nsp2 (starting with Gly^{261}) will be the subject of future study.

ACKNOWLEDGMENTS

We thank Johan den Boon, Twan de Vries, Alexander Gorbalenya, and R. Amons for technical assistance and helpful discussions.

REFERENCES

1. P.G.W. Plagemann and V. Moennig, Adv. Virus Res. 41:99.(1991).

2. J.A. den Boon, E.J. Snijder, E.D. Chirnside, A.A.F. de Vries, M.C. Horzinek and W.J.M. Spaan, J. Virol. 65:2910 (1991).
3. A.A.F. de Vries, E.D. Chirnside, P.J. Bredenbeek, L.A. Gravestein, M.C. Horzinek and W.J.M. Spaan, Nucleic Acids Res. 18:3241 (1990).
4. W.J.M. Spaan, D. Cavanagh and M.C. Horzinek, J. Gen. Virol. 69:2939 (1988).
5. E.J. Snijder, M.C. Horzinek and W.J.M. Spaan, J. Virol. 64:331 (1990).
6. E.J. Snijder, J.A. den Boon, P.J. Bredenbeek, M.C. Horzinek, R. Rijnbrand and W.J.M. Spaan, Nucleic Acids Res. 18:4535 (1990).
7. M.E.G. Boursnell, T.D.K. Brown, I.J. Foulds, P.F. Green, F.M. Tomley and M.M. Binns, J. Gen. Virol. 68:57 (1987).
8. P.J. Bredenbeek, C.J. Pachuk, J.F.H. Noten, J. Charité, W. Luytjes, S.R. Weiss and W.J.M. Spaan, Nucleic Acids Res. 18:1825 (1990).
9. H.J. Lee, C.K. Shieh, A.E. Gorbalenya, E.V. Koonin, N. la Monica, J. Tuler, A. Bagdzhadzhyan and M.M.C. Lai, Virology 180:567 (1991).
10. I. Brierley, P. Diggard and S.C. Inglis, Cell 57:537 (1989).
11. L. Polgar and P. Halasz, Biochem. J. 207:1 (1982).
12. M.R. Denison, P.W. Zoltick, J.L. Leibowitz, C.J. Pachuk and S.R. Weiss, J. Virol. 65:3067 (1991).
13. A.E. Gorbalenya, E.V. Koonin, A.P. Donchenko and V.M. Blinov, Nucleic Acids Res. 17:4847 (1989).
14. M.R. Denison, P.W. Zoltick, S.A. Hughes, B. Giangreco, A L. Olson, S. Perlman, J.L. Leibowitz and S.R. Weiss, Virology 189:274 (1992).
15. A.E. Gorbalenya, E.V. Koonin and M.M.C. Lai, FEBS Lett. 288:201 (1991).
16. J.H. Strauss and E.G. Strauss, Semin. Virol. 1:347 (1991).
17. J.C. Carrington and K.L. Herndon, Virology 187:308 (1992).
18. C.-S. Oh and J.C. Carrington, Virology 173: 692 (1989).
19. G.H. Choi, D.M. Pawlyk and D.L. Nuss, Virology 183:747 (1991).
20. E.V. Koonin, G.H. Choi, D.L. Nuss, R. Shapira and J.C. Carrington, Proc. Natl. Acad. Sci. USA 88:10647 (1991).

The data presented in this paper are derived from a publication by the same authors in the Journal of Virology, volume 66, pages 7040-7048 (published by the American Society for Microbiology in december 1992).