Proteolytic Cleavage of Encephalomyocarditis Virus Capsid Region Substrates by Precursors to the 3C Enzyme

GRiffith D. Parks,† Jean C. Baker, and Ann C. Palmenberg*

Institute for Molecular Virology, University of Wisconsin, Madison, Wisconsin 53706

Received 24 August 1988/Accepted 30 November 1988

Picornaviruses produce multiple viral proteins by proteolytic processing of a giant precursor polyprotein (5, 15). The first cleavage event occurs cotranslationally as ribosomes reach the middle, or P2 region, of the genome. This primary cleavage releases a large capsid region precursor (P1 or L-P1-2A) from the amino end of the growing polypeptide chain (15, 18, 19). For poliovirus, polypeptide 2A has been shown to be responsible for this activity (19), but the protease catalyzing the analogous event during cardiovirus and aphthovirus translation has still not been identified.

In all picornaviruses, however, most subsequent viral processing is carried out by viral protease 3C, which is derived from the carboxy-terminal (P3) region of the polyprotein (2, 10). Mature 3C enzyme is released from its precursors through a unique pattern of monomolecular or bimolecular self-cleavage steps which are inititated soon after translation of the P3 region is completed (11, 21). Cell-free reactions with encephalomyocarditis virus (EMC) have demonstrated that the released enzyme can catalyze as many as nine successive reactions within remaining viral proteins (10, 13).

For EMC, the natural 3C cleavage sites invariably occur between amino acid pairs containing Gln or Glu, followed by Gly or Ser (7, 9). Many of these sites are also flanked by proline (Pro) residues at the P2 or P1' position, but it is not clear how the enzyme finds or recognizes appropriate cleavage junctions within its polyprotein, as other apparently identical sequences are not predictably viable substrates.

We have previously reported the construction of EMC cDNA clones, whose RNA transcripts express active viral proteins in vitro, including the 3C enzyme and its L-P1-2A capsid region substrate (12, 13). Since EMC RNA and its cloned derivatives translate with unusual efficiency in cell extracts (e.g., reticulocyte extracts), the system provides a powerful assay for the analysis of enzyme activity and substrate specificity. For example, we have shown that EMC 3C is indeed responsible for nine polyprotein cleavage events and that cleavage site recognition by this enzyme can (by substitution) extend to Gin-Ala but not to Gin-Val, Gln-Glu, Lys-Ala, Lys-Val, Lys-Glu, or Lys-Gly amino acid pairs. Our engineered mutations also suggest that a flanking Pro residue is not a definitive requirement for a cleavage site (13).

We now report extended experiments which demonstrate that several P3 region precursors of EMC 3C also have proteolytic activity in vitro. Site-specific substitutions were introduced into the cleavage sites flanking the 3C protease to effectively lock the enzyme into the desired precursor forms. The resulting mutant proteins were assayed for their ability to self-process and to cleave the capsid region precursors in vitro. Additionally, we have constructed six new cleavage site sequence substitutions within the P3 region and tested them for their ability to serve as substrates in monomolecular or bimolecular reactions with 3C protease.

MATERIALS AND METHODS

Nomenclature. EMC proteins are designated by standard L-4-3-4 nomenclature (16), with the exception of 1AB (VP0), 1C (VP3), and 1D (VP1) where traditional names are used interchangeably. Because of cloning site selection, viral coding sequences within plasmids pE5P3 and its derivatives (pE5P3r, pE5P3I, and pE5P3I) are missing 33 codons from the 5' ends of their P3 regions relative to those of wild-type sequences (amino end of the peptide). Resultant peptides containing the truncated 3A segments are designated with primes (P3', 3A'BC, and 3A'B). Protease fraction F2 is a clarified extract of EMC-infected HeLa cells (12). Cleavage sites are denoted as . . . P1P1'P2P2' . . . where scission occurs between P1 and P1'.

Plasmid DNAs. Standard cloning techniques were used (8). All enzymes and vector pUC19 DNA were purchased from New England Biolabs, Inc. Plasmids pE5P3, pE5A1, and pE5LVP0 have been described previously (12, 13). To construct pUCP3, plasmid pE5P3 DNA (2 μg) was digested to completion with PstI and SmaI. The resulting 963-base-pair fragment which encodes 3B, 3C, and part of 3D (bases

* Corresponding author.
† Present address: Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, IL 60208.
Site-specific mutagenesis. Site-specific mutagenesis was carried out as described previously (6, 13, 22). Plasmid M13-3CD was constructed for mutagenesis of the 3C-3D junction. Briefly, DNA from pUCP3 was digested with BglII and EcoRI and the resulting 470-base-pair fragment (encoding the 3C-3D cleavage site) was gel purified. It was then inserted into the BamHI-EcoRI sites of the replicative form of M13mp18. Likewise M13-3BC was constructed for mutagenesis of the 3B-3C site by digestion of pUCP3 with HindIII and XhoI. The resulting 643-base-pair fragment was then purified and inserted into the HindIII-SalI sites of M13mp18. Both recombinant M13 viruses contained EMC minus-sense sequences.

Single-stranded M13 DNA templates were enriched in uracil prior to mutagenesis reactions (6). Mutagenesis was performed (100-μl reaction; 25 mM Tris hydrochloride [pH 7.5], 30 mM NaCl, 13 mM MgCl2, 1.2 mM β-mercaptoethanol, 0.7 mM deoxynucleoside triphosphate, 0.35 mM rATP) with 0.5 μg of single-stranded, uracil-enriched DNA template, Klenow polymerase (5 U), T4 DNA ligase (6 U), and a synthetic oligonucleotide primer (3 to 35 pMol). The primer sequences are illustrated in Fig. 1. After 30 min at 4°C, followed by 2 h at 25°C, the mixture was used to transfect Escherichia coli JM101, and plaques were screened by dideoxynucleotide sequencing (9). The frequency of primer-generated mutations ranged from 65 to 80%.

Bacteriophage containing desired mutations were plaque purified. The EMC-specific segment was then excised from the replicative form of M13 by digestion with either PstI plus BglII (3B-3C junction) or with EcoRI plus XhoI (3C-3D junction) and substituted for the wild-type sequences in pESP3. Plasmid pESP3r encodes a Gin-to-Arg substitution at the 3B-3C junction. Plasmid pESP3I encodes a Gly-to-Ile substitution at the 3C-3D junction. For pESP3rI, corresponding wild-type sequences in pESP3r were replaced with the BglII-XhoI fragment from pESP3I, producing a plasmid containing both mutations. All mutations were reconfirmed by dideoxynucleotide sequencing before use in protease assays.

In vitro transcription, cell-free translation, and gel electrophoresis. RNA transcripts were generated from BamHI-linearized plasmid DNAs in reactions with bacteriophage T7 RNA polymerase as previously described (12). Resulting transcription products were precipitated with ethanol, dried under vacuum, and dissolved in water (at about 1 μg/μl) before use. Cell-free protein synthesis in reticulocyte extracts was programmed with the transcription products as described previously (12, 18). [35S]methionine (1,100 Ci/mmol; Dupont, NEN Research Products) or [3H]leucine (131 Ci/mmol; Amersham Corp.) was included at approximately 1 or 5 μCi/μl, respectively. Translation reactions were stopped by the addition of pancreatic RNase and cycloheximide (to 0.33 mg/ml each). Protein gel electrophoresis, autoradiography, and fluorography were carried out by standard methods (10).

RESULTS

Mutagenesis of the 3B-3C cleavage junction. Synthetic oligonucleotide primers (Fig. 1) were used to introduce site-specific substitutions into the nucleic acids encoding the Gin-Gly pair at the EMC 3B-3C cleavage junction of plasmid pESP3. One resulting clone, pESP3r (pESP3 plus Arg), contained a Arg-Gly combination at this site and was chosen

for analysis of its proteolytic properties. Wild-type (pESP3) and pESP3r plasmid DNAs were transcribed in vitro, and the RNA products were used to program [3H]leucine-labeled rabbit reticulocyte translation reactions. RNA produced from pESP3 directed synthesis of a P3' precursor, which was rapidly cleaved to yield 3CD, 3D, 3C, and 3AB (Fig. 2A, lanes 1 and 2). These bands, as shown previously, represent normal P3 region processing by the endogenous 3C sequences encoded within this plasmid (12).

In contrast, translation of pESP3r-directed RNA gave a somewhat different pattern of bands (Fig. 2A, lanes 3 through 5). The P3′ precursor, and its 3D and 3A′/BC cleavage products were the only proteins evident (lane 3), even after extended incubation (lane 4). Addition of highly active exogenous protease did not change the processing pattern (lane 5), indicating that the altered 3B-3C site could not be cleaved in bimolecular or monomolecular reactions. Instead, large amounts of 3A′/BC produced show that the Arg residue in the P1 position of the cleavage site was inhibitory and thus that the naturally occurring Gin must be very important for recognition and processing of EMC cleavage sites (13).

Wild-type and pESP3r-derived P3 region products were tested for their ability to process EMC capsid precursors by incubation with [35S]methionine-labeled L-P1-2A substrate (Fig. 2B, lane 1). After incubation, the capsid region cleav-

FIG. 1. Map of EMC P3 region and engineered processing sites. (A) Structure of clone pESP3 and processing of the P3 region of the EMC polyprotein. Vpg, Pro, and Pol denote the genome-linked viral protein, protease, and polymerase sequences, respectively. 5 NC denotes the EMC 5′ noncoding sequences contained within pESP3 (12). (B) Positions and sequences of the synthetic oligonucleotide primers used to introduce substitutions at the 3B-3C (left) and 3C-3D (right) cleavage junctions. Nucleotide substitutions are shown below the primers with the corresponding amino acid changes shown above the wild-type sequence. Also shown are the putative polypeptides encoded by clones pESP3r, pESP3l, and pESP3rl. X denotes mutated cleavage sites within these proteins.
age patterns generated by the wild-type (Fig. 2B, lane 2) or mutant protease (lane 3) preparations were virtually identical, including bands representing P1-2A, P1, 1ABC, E1 (ε1), 1AB, 1D, 1C, 2A, and L. Generation of this pattern requires four separate 3C-directed cleavage events within the L-P1-2A precursor (9, 13). Protein E1 (also known as ε1) is an unmodified form of 1AB and minor variations in its intensity do not reflect differential proteolytic activity (10, 11). We conclude that although the pE5P3r-derived 3C protease cannot be liberated from its 3A’BC precursor, the trapped enzyme still actively and correctly catalyzed all four capsid region processing steps.

Mutagenesis of the 3C-3D cleavage junction. To determine whether other P3 precursor forms retained similar activity, we used a second synthetic primer (Fig. 1) to introduce different substitutions in the 3C-3D cleavage site of pE5P3. Clone pE5P3 (pE5P3 plus Ile), encoded a Gln-Ile at this position instead of the wild-type Gln-Gly. RNA transcripts from pE5P3 were translated in cell extracts to yield 3A’B (results not shown, off bottom of figure) and 3CD (Fig. 3A, lane 3) as the only products. As with clone pE5P3r, the introduced mutation effectively prevented cleavage at the altered site and allowed accumulation of a larger form of the protease, in this case, protein 3CD.

To produce a stable version of the P3’ protein, the mutations within pE5P3r and pE5P3I were combined within a single construction to yield pE5P3rI (pE5P3 plus Arg plus Ile). Unexpectedly, cell-free expression of RNA from pE5P3rI produced not only the predicted P3’ protein but also a second band, migrating slightly faster than P3’ (Fig. 3A, lane 4). We believe this additional protein represents 3BCD, a species not normally observed in any of our cell-free reactions or in EMC-infected cells (11, 18). Tryptic peptide mapping experiments are currently under way to confirm the identity of this band (as 3BCD).

Clone-derived proteins from pE5P3I and pE5P3rI were tested for their ability to process the viral L-P1-2A capsid precursor in protease assays similar to those described in the legend to Fig. 2. Both mutant P3 region protein samples cleaved the capsid precursor into all of the expected products (Fig. 3B, lanes 2 and 3) with efficiencies equivalent to that of the wild-type enzyme (lane 1).

EMC 3C cleavage site specificity. We have previously shown that the EMC 3C enzyme is capable of processing cleavage sites composed of Gln followed by Gly, Ala, or Ser (13). As part of the experiments described above, several other site-specific substitutions were engineered into the 3C-3D junction, replacing the Gly residue with either Cys, Thr, or Tyr. Surprisingly, when expressed in reticulocyte extracts, precursors harboring 3C-3D junctions composed of Gln-Cys were spontaneously cleaved to produce 3C. The cleavage was inefficient relative to the wild-type sequence, but a weak band of 3C was nevertheless present after prolonged incubation. Substitution of a Thr or Tyr residue in the P1’ position eliminated self-cleavage at this junction, as did the Ile substitution already described. Likewise, substitution of Leu for Gln at the 3B-3C site resulted in an inactive cleavage site (gels not shown). The results with all altered EMC P3 region cleavage site combinations are summarized in Fig. 4. Cleavage site substitutions previously reported at the 1C-1D junction (shaded boxes) are included for com-

FIG. 2. Cleavage of L-P1-2A by pE5P3r-derived mutant proteins. (A) RNA transcripts from pE5P3 and pE5P3r were used to program [3H]leucine-labeled cell-free protein synthesis. After 45 min, RNase and cycloheximide were added and 3-μl samples were saved (lanes 1 and 3). Similar samples were mixed with 3 μl of buffer (phosphate-buffered saline [17]) (lanes 2 and 4) or with 2 μl of buffer plus 1 μl of nonradiolabeled fraction F2 EMC protease (lane 5) before further incubation at 30°C for 12 to 15 h. (B) [35S]methionine-labeled capsid precursor L-P1-2A was generated in vitro from clone pE5A1. Samples (1 μl) were saved (lane 1) or mixed with 2 μl of buffer plus 2 μl of the reactions shown in lanes 1 and 3 of panel A (lanes 2 and 3, respectively). After 12 to 15 h at 30°C, samples were analyzed by gel electrophoresis and visualized by fluorography (A) or autoradiography (B). Lane M is an EMC protein marker lane.
pleteness (13). Residues marked with asterisks occur at natural EMC sites.

**DISCUSSION**

It is well established that the EMC 3C enzyme in its free form can process the capsid region of its polyprotein (L-

\[ P_1 \]

\[ P_1' \]

\[ P_1'' \]

\[ P_1''' \]

\[ P_1'''' \]

...P2 - P1 - P1' - P2'...

FIG. 4. EMC cleavage site sequences tested in processing assays. EMC P3 region cleavage sites containing altered P1 and P1' residues were tested as substrates in cell-free assays as described in the legends to Fig. 2A and 3A. The sites were scored (+ or −) for their ability to be cleaved by 3C protease after incubation at 30°C for 12 to 15 h. Starred residues denote cleavage sequences which occur naturally in the polyprotein (9). Shaded boxes denote combinations tested at the 1C-1D site as described previously (13). The ES combination is the proposed natural site of the VP1-2A junction. NT, Not tested.

P1-2A) in a four-step cleavage cascade to yield proteins L, VP0, VP3, VP1, and 2A (10, 12, 13). Our experiments concerning possible activity of precursors to 3C were initiated at the suggestion of Richard Jackson (3). His careful time courses of cardiovirus processing in cell extracts clearly indicated that initial cleavages in the capsid region occurred before free 3C was detectable in the samples and correlated more closely with the appearance of 3ABC, normally a minor intermediate in the processing of the P3 region (3, 11). We have directly tested this hypothesis and shown that P3 region proteins harboring inactive cleavage sites which lock the 3C protease into various precursor forms are still fully active in processing the capsid region in vitro. Not only 3ABC but also 3CD and P3 (3BCD?) proved capable enzymes.

Thus, our data imply that the EMC 3C protease must be viewed as a group of enzymatic species, rather than a unique entity. It is very likely that in vivo, the different proteases may have various activities at different sites in their viral substrates. This has been shown to be the case for the related poliovirus, in which similar in vitro experiments have demonstrated that the 3CD protein, rather than 3C itself, is an essential requirement for complete processing of poliovirus P1 precursors (4, 20). For EMC, each of the tested P3 region enzyme forms (including the free enzyme) was capable of completing the entire L-P1-2A cleavage cascade. However, while our cell-free assays did not detect any overt differences in processing rates among the enzymes, our reactions were not designed as detailed kinetic studies. We are currently carrying out more-careful analyses with the four clone-derived forms of 3C and with various new substrate constructions derived from the P2 and the P1 regions in order to test enzyme preferences for particular cleavage sites.
Within the P3 region, synthesis of the putative 3BCD protein occurs at Glu-Gly, Glu-Ser, or Glu-Ser amino acid pairs, though of the 21 potential sites containing these sequences, only 9 actually serve as substrates, in vitro or in vivo (9). Although the structural elements dictating specific cleavage site selection have not been clearly established, it is reasonable to assume the reactions require a somewhat flexible peptide segment, located on or near the surface of the substrate protein (1, 7). Several of the uncleaved potential sites may simply be inaccessible to enzymes. However, the cleavage sequence itself must also play a significant role in site recognition, as evidenced by the limited repertory at the cleaved locations. We have tested some of the sequence-specific constraints on processing by introducing a series of amino acid substitutions into natural cleavage sites and testing their efficacy in protease reactions.

Of those combinations tested, our functional EMC cleavage sites all possessed a Gln or Glu residue in the P1 position, followed by a small aliphatic amino acid, Gly, Ser, Ala, or Cys, in the P1 position. The latter four residues, of which Cys was the least active, represent the physically smallest of all the amino acids. Larger hydrophobic (Val and Ile), charged (Glu), branched (Thr), or aromatic (Tyr) residues were not active in the P1 position, even though two of the tested combinations (Gln-Thr and Gln-Ile) do occur naturally within the related foot-and-mouth disease virus polyprotein (14). We are currently supporting efforts towards the resolution of the EMC 3C crystal structure and hope that our mutant processing data will provide a useful perspective when modeling active and inactive substrates into the catalytic site of the enzyme.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI-17331 from the National Institutes of Health to A.C.P. During the course of this work, G.D.P. was partially supported by American Cancer Society grant MV-9 to Roland R. Rueckert.

LITERATURE CITED