Identification of the Herpes Simplex Virus-1 Protease Cleavage Sites by Direct Sequence Analysis of Autoproteolytic Cleavage Products*

Carolyn L. Dilanni‡§, Diana A. Drier‡, Ingrid C. Deckman‡, Patrick J. McCann III‡, Fenyong Liu‡, Bernard Roizman‡, Richard J. Colonna‡, and Michael G. Cordingly‡

From the ‡Department of Virology, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543-4000 and the §Marjorie B. Kooler Viral Oncology Laboratories, University of Chicago, Chicago, Illinois 60637

Herpes simplex virus type-1 (HSV-1) encodes a protease responsible for proteolytic processing of the virus assembly protein, ICP35 (infected cell protein 35). The coding region of ICP35 is contained within the gene that encodes the protease, and ICP35 shares amino acid identity with the carboxyl-terminal 329 amino acids of the protease. The HSV-1 protease was expressed in Escherichia coli as a fusion protein containing a unique epitope and the protein A Fc binding domain at its carboxyl terminus. The fusion protease underwent autoproteolytic cleavage at two distinct sites. The size of the cleavage products containing the carboxyl-terminal epitope mapped one cleavage site near the carboxyl terminus of the protease corresponding to the proteolytic processing site of ICP35, and the second site proximal to the amino terminus consistent with previous data. The carboxyl-terminal autoproteolytic cleavage products were partially purified on an IgG affinity column by virtue of the protein A Fc binding domain and subjected to direct amino-terminal sequence analysis. Protein sequencing revealed that cleavage occurs between the Ala and Ser residues at amino acids 610/611 and 247/248 of the HSV-1 protease. The flanking sequences share homology with each other and are highly conserved in homologous proteases of other herpes viruses.

Infected cell protein 35 (ICP35) of HSV-1 is a family of proteins that is a major component of viral B-capsids found in the nucleus of infected cells but is absent from mature virions (1–3). Consequently, ICP35 and its homologues in other herpesviruses have been called virus assembly proteins and are proposed to play a role analogous to that of the carboxyl-terminal (5). An HSV-1 ts mutant that is defective in the processing of ICP35 at the nonpermissive temperature also fails to package viral DNA, and only aberrant empty capsids accumulate (6). The proteolytic processing of the virus assembly protein therefore appears to be critical for viral particle maturation. The protease responsible for these processing events is encoded in the 635-amino-acid open reading frame of UL26, which is 3′ coterminal with UL26.5, the gene encoding ICP35 (7). The 329-amino-acid open reading frame encoding ICP35 is in frame with that of the protease (7) and as a consequence, the protease shares its carboxyl-terminal domain with ICP35 and is capable of autoproteolytic processing of its own carboxyl terminus at a site identical to that of its substrate. We have recently demonstrated that Escherichia coli-expressed HSV-1 protease undergoes autoproteolytic processing at the carboxyl-terminal site that it shares with ICP35 and at an additional site, proximal to the amino terminus, that is unique to the protease (8).

In this report we identify the two proteolytic processing sites within the protease by direct amino acid sequencing of the isolated proteolytic products. This was achieved by constructing a HSV-1 protease fusion gene (PraF) encoding the complete HSV-1 protease fused at its carboxyl terminus with a 20-amino-acid epitope that is recognized by monoclonal antibody CH28–2 (5) and 253 amino acids from the Fc binding domain of staphylococcal protein A. The resulting carboxyl-terminal products are easily identified by the unique epitope and purified by affinity chromatography exploiting the Fc binding domain of protein A. This construct has been used previously in studies to characterize the proteolytic processing of the UL26 gene product expressed in reticulocyte lysates (5). In addition, a mutant HSV-1 protease containing substitution of the His at position 148 with an Ala (H148A), which results in a protein that does not undergo autoproteolytic processing (9), was utilized to verify that the cleavage products generated in this construct arise from autoproteolytic processing.

MATERIALS AND METHODS

Constructs—The protease fusion protein was expressed in E. coli in the vector pET11d (Novagen) under the control of the inducible T7 promoter using established protocols (10). The construct was created by inserting a 2745-base pair XbaI-EcoRI fragment of pBH4214 into the NheI-EcoRI sites of the vector. The resulting vector, pTT4214PRT, encodes a fusion protein containing 2 amino acids of the vector, the upstream 7 amino acids of the UL26 gene, followed by the start methionine and coding region of HSV-1 protease, the 20-amino-acid epitope recognized by monoclonal antibody CH28–2 (5), and the Fc binding domain of staphylococcal protein A. The mutant fusion protein H148A was cloned from pRR8077 (9) utilizing an identical strategy.

Antibodies—Monoclonal antibody CH28–2 reacts specifically with a 20-amino-acid CMV epitope and was obtained from B. Roizman and L. Pereira (5).

Protein Purification—Bacterial cells from a 1-liter culture ex-
pressing the fusion protein were suspended in 50 ml of 10 mM sodium phosphate (pH 7.2), 500 mM NaCl and lysed by pulsed sonication on ice. Cell debris was pelleted by centrifugation at 28,400 \times g for 40 min. The resulting supernatant was incubated with 15 ml of rabbit IgG-agarose (Sigma) by rocking for 18 h at 4°C. Elution of the NaF and C,F species was achieved by washing the column with 100 mM glycine (pH 2.4) and 156 mM NaCl. Concentration of this material was achieved by dialysis against 5 mM Hepes (pH 7.0), 0.5 mM EDTA, 1 mM dithiothreitol and subsequent lyophilization. Further purification of NaF was accomplished by HPLC on a reverse phase C4 column (4.6 mm (inner diameter) \times 25 cm, Vydac) eluting with a 0-100%/100-min gradient of acetonitrile in 0.1% trifluoroacetic acid.

Sequence Analysis—Concentrated fractions containing C,F and NaF were resolved by SDS-PAGE and transferred to Immobilon (Millipore) for direct sequence analysis. In order to minimize chemical blocking of the amino terminus, 12.5% acrylamide resolving gels were polymerized for 18 h at 4°C and prerun using SDS-PAGE running buffer containing 0.1 mM thioglycolic acid (11) for 1 h at 8 mA at 25°C. Samples were electrophoresed at 30 mA (constant current) for 8 h at room temperature in SDS-PAGE running buffer without thioglycolic acid and then transferred (50 V) to Immobilon (Waters) in Laemmli buffer with 20% methanol for 18 h at 30 V, 4°C. Sequence analyses were performed by the Edman degradation method.

RESULTS

DNA encoding the wild-type HSV-1 fusion protease or the H148A mutant fusion protease was cloned into a T7 vector for expression in E. coli (see "Materials and Methods"). The HSV-1 protease fusion protein and its predicted cleavage products are illustrated in Fig. 1. The expression of PraF and H148A proteins was monitored by immunoblotting with monoclonal antibody directed against the unique carboxy-terminal epitope (Fig. 2). In each case, full-length protein was detected within 1 h of induction (Fig. 2); however, only the wild-type protease underwent autoproteolysis to generate 80- and 33-kDa species, NaF and C,F; respectively, that accumulate to high levels at later time points. While it cannot be ruled out that some NaF is generated by the mutant protease, it is most likely that these minor species represent degradation products consistently observed with this mutant protein.2 Additional immunoreactive species only accumulated to a low level following induction of either form of the protease (Fig. 2).

The carboxyl-terminal cleavage products, NaF and C,F, were generated by autoproteolysis of the full-length precursor protein at each of the two predicted cleavage sites within the protease domain of the fusion protein, and each contains the C,F binding domain of protein A at the carboxyl terminus (Fig. 1). This domain facilitated simple purification of the carboxyl-terminal cleavage products by affinity chromatography on an IgG column. E. coli cells expressing the fusion protein were harvested 2–5 h after induction to ensure the presence of the carboxyl-terminal cleavage products. The cells were lysed by sonication, cell debris was removed by centrifugation, and the resulting soluble supernatant was subjected to affinity chromatography on an IgG-agarose column. NaF and C,F, were identified in the column fractions by SDS-PAGE, immunoblotting with the monoclonal antibody directed against the unique carboxy-terminal epitope, and staining with Coomassie Brilliant Blue. Ninety-seven percent of the total protein flowed through the column while NaF and C,F, were bound, and eluted at low pH (see "Materials and Methods"). Coomassie staining of the proteins in eluted fractions (Fig. 3A, lane 2) revealed that C,F could be sufficiently resolved by gel electrophoresis to enable direct sequencing. In contrast, the lower yield of NaF and existence of several contaminating species with similar molecular weights made further purification of this species imperative. IgG-purified material containing NaF (Fig. 3A, lane 2) was therefore subjected to reverse phase HPLC (Fig. 3B). Fractions containing NaF were effectively resolved from contaminating species as a discrete peak of protein (Fig. 3B, fraction 58) in which NaF was the major staining species (Fig. 3A, lane 3). This material served as a source of NaF for sequence analysis.

The preparations of C,F and NaF were concentrated, subjected to electrophoresis, and then transferred to Immobilon. C,F and NaF were identified by Coomassie staining, excised, and subjected to amino-terminal sequence analysis. The resulting amino-terminal sequence of the C,F cleavage fragment was compared with the predicted amino acid sequence of the protease as deduced from the nucleotide sequence of the UL26 gene (12) and found to align with sequences from amino acids 611 to 622 (Table 1). Only 25 pmol of Ser was detected in

---

2 I. C. Deckman and P. McCann, unpublished data.

---
Identification of the HSV-1 Protease Cleavage Sites

A

2050 Identification of the HSV-1 Protease Cleavage Sites

A

28-

1-

i

0.0

y-y

10 20

30

40 50

70 80

90 100

Function Number

B

FIG. 3. Purification of C,F and NaF. A, Coomassie-stained proteins following purification by affinity chromatography on IgG-agarose (lane 2) and subsequently by C, reverse phase HPLC column (lane 3); lane 1 contains molecular weight markers. B, chromatogram of absorbance at 280 nm of the C, reverse phase HPLC purification of NaF. Fraction 58 from the C, purification, which contains primarily NaF, is shown in lane 3.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>CoF Amino acid</th>
<th>Predicted sequence</th>
<th>NaF Amino acid</th>
<th>Predicted sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ser</td>
<td>Ser-611</td>
<td>Ser</td>
<td>Ser-248</td>
</tr>
<tr>
<td>2</td>
<td>Ser</td>
<td>Ser-612</td>
<td>Glu</td>
<td>Glu-249</td>
</tr>
<tr>
<td>3</td>
<td>Ala</td>
<td>Ala-613</td>
<td>Lys</td>
<td>Lys-250</td>
</tr>
<tr>
<td>4</td>
<td>Ala</td>
<td>Ala-614</td>
<td>Phe</td>
<td>Phe-251</td>
</tr>
<tr>
<td>5</td>
<td>His</td>
<td>His-615</td>
<td>Lys</td>
<td>Lys-252</td>
</tr>
<tr>
<td>6</td>
<td>Val and Tyr</td>
<td>Val-616</td>
<td>Met</td>
<td>Met-253</td>
</tr>
<tr>
<td>7</td>
<td>Asp</td>
<td>Asp-617</td>
<td>Trp</td>
<td>Trp-254</td>
</tr>
<tr>
<td>8</td>
<td>Val and Tyr</td>
<td>Val-618</td>
<td>Gly</td>
<td>Gly-255</td>
</tr>
<tr>
<td>9</td>
<td>Asp</td>
<td>Asp-619</td>
<td>Ala</td>
<td>Ala-256</td>
</tr>
<tr>
<td>10</td>
<td>Ala</td>
<td>Thr-620</td>
<td>Glu</td>
<td>Glu-257</td>
</tr>
<tr>
<td>11</td>
<td>Ala</td>
<td>Ala-621</td>
<td>Pro</td>
<td>Pro-258</td>
</tr>
<tr>
<td>12</td>
<td>Arg</td>
<td>Arg-622</td>
<td>Val</td>
<td>Val-259</td>
</tr>
</tbody>
</table>

TABLE I
Amino-terminal sequencing of CoF and NaF by Edman degradation

Each sequence is the result of two determinations.

The sequence determined for the amino-terminal 12 amino acids of NaF unambiguously matched the predicted amino acid sequence of the protease from amino acid 248 to 260 (Table I), indicating conclusively that the amino-terminal cleavage occurred between Ala-247 and Ser-248.

DISCUSSION

The HSV-1 protease has been expressed in E. coli as a fusion protein that undergoes autoproteolytic cleavage to generate characteristic cleavage products. A mutant protease in which His-148 was changed to Ala failed to undergo proteolytic processing, confirming that the protease itself is responsible for the observed cleavages. The size and immunoreactivity of the cleavage products that accumulated are consistent with the previous observation (8) that two processing sites exist in the HSV-1 protease, one close to the carboxyl terminus and the other proximal to the amino terminus (Fig. 1). Amino-terminal sequence analysis of the carboxyl-terminal proteolytic cleavage products, C,F and NaF, mapped the location of the carboxyl- and amino-terminal cleavages, respectively. The carboxyl-terminal cleavage occurred between Ala-610 and Ser-611. 25 amino acids from the carboxyl terminus of the protease, and the amino-terminal cleavage occurred between Ala-247 and Ser-248. The sequences surrounding the two cleavage sites share homology with one another and are conserved at analogous positions within the homologues of the protease in other herpesviruses (Table II).

Examination of the amino acid sequences flanking the processing sites of herpesvirus proteases reveals that a consensus sequence can be derived (Table II) in which P3, P1, and P1' are strongly conserved in both the amino- and carboxyl-terminal cleavage sites. These data strongly suggest that the cleavage sites for all of the herpesvirus proteases will occur between the conserved Ala and Ser residues (P1/P1'). Downstream residues do not appear to be conserved except that P2' is usually small. Interestingly, the P4 position at the carboxyl-terminal site is conserved as a small aliphatic residue, while the equivalent position at the amino-terminal site is occupied by a highly conserved Tyr in all of the sequences.

3 I. C. Deckman and P. J. McCann III, unpublished data.
examined. This most likely reflects differences in the structural requirements for the two cleavage sites. Structural features outside P1 and P1′ are clearly important for cleavage site recognition since cleavage is not observed at the five other naturally occurring Ala/Ser sites within the HSV-1 protease.

The carboxyl-terminal cleavage site of the HSV-1 protease is also present at the carboxy terminus of ICP35, which shares sequence with the protease and is a substrate in infected cells (5, 7). Proteolytic processing of the ICP35 family of virus assembly proteins (ICP35c,d) to lower molecular weight forms (ICP35e,f) is essential for successful virus maturation (6). The carboxyl-terminal cleavage event itself or the newly cleaved ICP35 must play a critical role in scaffold assembly or subsequent steps in virus maturation. Interestingly, the CMV assembly protein (15), which is the homologue of ICP35, undergoes an analogous processing event near its carboxyl terminus (16, 17). Recent studies utilizing mass spectroscopy of the fragmented assembly protein, purified from immature CMV particles, have located the cleavage event in this protein between Ala-557 and Ser-558 (18). As indicated in Table II, this region shares a high degree of homology with the carboxyl-terminal HSV-1 protease cleavage site. Based upon this sequence homology, Welch and colleagues (18) proposed the existence of an amino-terminal cleavage site within the CMV protease. Our data provides direct experimental evidence that the HSV-1 protease is proteolytically cleaved at this site and suggests that this is a common feature in the maturation processing of herpesvirus proteases. The role of the amino-terminal cleavage, which occurs within sequences of the protease not shared with ICP35, is not clear. It has been suggested that the amino-terminal product of cleavage at an equivalent site in the CMV protease results in release of active protease (18). This species is absent in the HSV-1 protease (amino acids 1-247) and the corresponding carboxyl-terminal fragment processed at the carboxyl-terminal cleavage site (amino acids 248-610) are incorporated into viral B-capsids along with processed forms of the assembly protein, ICP35e,f. A role for this second cleavage in the assembly or maturation of B-capsids and subsequent encapsidation of viral DNA therefore cannot be excluded.

Additional studies of the protease and its activities both in vivo and in vitro will lead to elucidation of its precise role in virus maturation. The obligate requirement for proteolytic processing of the herpesvirus assembly proteins and the identification of the protease as virus-encoded reveal that this protease is an exciting new antiviral target.

Acknowledgments.—We thank Dr. M. Flocco at Princeton University for the sequence analysis. We thank Dr. S. P. Weinheimer for many helpful discussions.

REFERENCES


* S. P. Weinheimer and D. O’Boyle, personal communication.

---

**Table II**

Sequence homology between herpes virus proteases

<table>
<thead>
<tr>
<th>Protein</th>
<th>Amino Acids</th>
<th>Sequence Corresponding to HSV-1 Protease Terminus</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV-1</td>
<td>Ala Gly Ala Leu Val Asn Ala Ser Ser Ala Ala His Val Asp 617</td>
<td></td>
</tr>
<tr>
<td>VZV</td>
<td>Asp Val Asn Ala Val Glu Ala Ser Ser Lys Ala Pro Leu Ile 585</td>
<td></td>
</tr>
<tr>
<td>CMV</td>
<td>Gln Ala Gly Val Val Asn Ala Ser Cys Arg Leu Ala Thr Ala 688</td>
<td></td>
</tr>
<tr>
<td>EBV</td>
<td>Gly Lys Leu Val Glu Gin Ser Ser Ala Ser Gly Val Ala Gin 575</td>
<td></td>
</tr>
<tr>
<td>ILTV</td>
<td>Ala Arg Glu Thr Val Asp Ala Ser Met Pro Lys Arg Leu Lys 595</td>
<td></td>
</tr>
</tbody>
</table>

**Amino-terminal cleavage**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Amino Acids</th>
<th>Sequence Corresponding to HSV-1 Protease Terminus</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV-1</td>
<td>Gly His Thr Tyr Leu Gin Ala Ser Glu Lys Phe Lys Met Trp 254</td>
<td></td>
</tr>
<tr>
<td>VZV</td>
<td>Gly His Val Tyr Leu Gin Ala Ser Thr Gly Tyr Gly Leu Ala 243</td>
<td></td>
</tr>
<tr>
<td>CMV</td>
<td>Arg Glu Ser Tyr Val Lys Ala Ser Val Ser Pro Glu Ala Arg 301</td>
<td></td>
</tr>
<tr>
<td>EBV</td>
<td>Ala Glu Ser Tyr Leu Lys Ala Ser Asp Ala Pro Asp Leu Gin 242</td>
<td></td>
</tr>
<tr>
<td>ILTV</td>
<td>Asn Pro Lys Tyr Leu Gin Ala Asn Glu Val Ile Thr Ile Gly 276</td>
<td></td>
</tr>
</tbody>
</table>