Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2076–2080, March 1992
Biochemistry

Differentiation of multiple domains in the herpes simplex virus 1 protease encoded by the U_L26 gene

(serine protease/protease inhibitors/dispensable domains/mutagenesis)

FENYONG LIU AND BERNARD ROIZMAN*

The Marjorie B. Kovler Viral Oncology Laboratories, The University of Chicago, 910 East 58th Street, Chicago, IL 60637

Contributed by Bernard Roizman, December 9, 1991

ABSTRACT **Previous studies have shown that the herpes simplex virus 1 gene U_L26 encodes a 635-amino acid protease that cleaves approximately 20 amino acids from the carboxyl terminus of itself and of a 329-amino acid product of the $U_L26.5$ gene. The results of studies with a variety of protease inhibitors showed that the U_L26 protease was inhibited by serine protease inhibitors but not by inhibitors of cysteine protease, aspartic acid protease, or metalloprotease. Mutations resulting in amino acid substitutions, deletions, or insertion of stop codons in the gene or of 20-amino acid stretches into the protease have delineated the dispensable domains I and IV at the amino and carboxyl domains of the gene product. The essential carboxyl-proximal domain (III) can be separated from the essential amino-proximal domain (II) by at least 20 amino acids. The amino-proximal domain is the most conserved region among varicella–zoster virus and human cytomegalovirus homologues of U_L26. Of the conserved aspartic acid, histidine, or serine amino acids in this domain, only histidine-61 and -148 could not be replaced without impairment of the proteolytic activity.**

Earlier we reported that the herpes simplex virus 1 (HSV-1) open reading frame (ORF) U_L26 encodes a protease (1). The sequence of U_L26 (2) predicts that the protease contains 635 amino acids. The substrate of the protease is the protease itself and a more abundant protein encoded by the ORF $U_L26.5$. The promoter of the $U_L26.5$ ORF is imbedded in the 5′ terminal domain of the coding sequences of U_L26 ORF. Its coding domain is identical to the coding domain of the terminal 329 amino acids of U_L26 ORF, and its product was previously designated as the infected-cell protein no. 35 (ICP35) (1, 3). The protease and its more abundant substrate are both cleaved by the protease approximately 20 amino acids from its carboxyl terminus. We report here that (*i*) the activity of the protease is consistent with that of an enzyme with serine at the active site, a serine protease; (*ii*) the protease protein contains several domains that are not required for its catalytic activity; and (*iii*) the active site is near the amino terminus of the protease.

The experimental design used in these studies is based on two observations. First, insertion of additional amino acid sequences including the IgG binding domains from staphylococcal protein A into the carboxyl terminus of the protease does not interfere with the self-cleavage of the protease but yields a readily detectable product of the reaction. Second, insertion of the sequence encoding a 20-amino acid epitope of human cytomegalovirus (CMV) monoclonal antibody in frame into the coding domains of the U_L26 and $U_L26.5$ ORFs serves two purposes. Foremost, it serves to identify specifically the products of these ORFs. Second, by separating the various domains of the protease, it serves to identify regions of the protease that must be contiguous for its catalytic function.

Table 1. List of mutations in the gene encoding the U_L26 protease

Designation	Mutation introduced into wild-type gene
	Insertion mutants (20-amino acid CMV epitope)
P	Insertion after amino acid 218
J	Insertion after amino acid 514
Q	Insertion after amino acid 615
	Construction of deletion mutants
D	Deletion of amino acids 1–220
G	Deletion of amino acids 219–615
EE	Deletion of amino acids 1–9
FF	Deletion of amino acids 1–32
AA	Insertion of stop codon after amino acid 615
BB	Insertion of stop codon after amino acid 514
CC	Insertion of stop codon after amino acid 287
DD	Insertion of stop codon after amino acid 218
MM	Insertion of stop codon after amino acid 306
NN	Deletion of amino acids 307–635
	Amino acid substitutions
GG	^{7}Gly-^{8}Asp-^{9}Arg with Ser-Arg-Thr (new *Xba* I site)*
HH	^{31}Asp-^{32}Ser-^{33}Gly with Leu-Asp-Met (new *Xba* I site)
II	His-61 → Val (new *Aat* II site)
JJ	His-148 → Ala (new *Pst* I site)
KK	Ser-215 → Ala (new *Nhe* I site)
LL	Asp-34 → Ala (new *Nhe* I site)

*The sequences were replaced as follows: plasmid GG, CCGGGAGACCGGATG with CCGTCTAGAACCATG; plasmid HH, TATGACAGCGGGGAC with TATCTAGACATGGAC; plasmid II, GACCACCGC with GACGTCCGC; plasmid JJ, GCGCACGTC with GCTGCAGTC; plasmid KK, ACGCTTTCCACC with ACGCTAGCCACC; and plasmid LL, GGGGACTCGGGG with GGGGCTAGCGGC.

METHODS AND MATERIALS

Virus and Cells. HSV-1(F) is the prototype HSV-1 used in this laboratory (4). The propagation of thymidine kinase-negative baby hamster kidney (BHK) cells was described elsewhere (5).

Monoclonal Antibodies. Monoclonal antibodies H725 and CH28-2 used in these studies were obtained from Lenore Pereira and have been described (3, 6).

***In Vitro* Transcription and Translation.** Plasmid DNA templates (5 μg) linearized with *Eco*RI or *Hin*dIII were transcribed in the presence of the cap analog GppG (New England Biolabs) with phage Sp6 or T7 RNA polymerase as recommended by Promega. One microgram of the RNAs was translated for 10 min in 50-μl reaction mixtures containing

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HSV, herpes simplex virus; CMV, cytomegalovirus; ORF, open reading frame; ICP, infected cellular polypeptide.
*To whom reprint requests should be addressed.

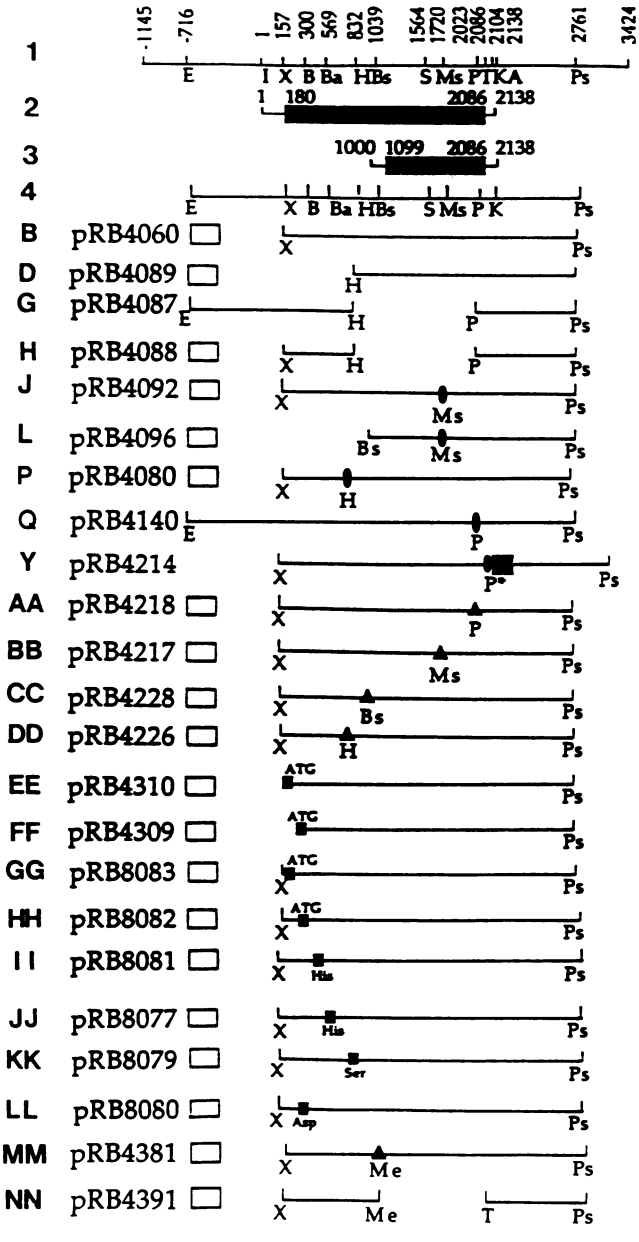

FIG. 1. The structure of the U_L26 and $U_L26.5$ ORFs tested in these studies. Line 1 shows nucleotide numbers relative to the approximate transcription initiation site of U_L26 (I represents nucleotide +1), restriction endonuclease sites of the HSV-1 *Eco*NI–*Pst* I DNA fragment containing the U_L26 and $U_L26.5$ ORFs, and the position of the translational termination codon (T) and of the single poly(A) signal (A) that serves both the U_L26 and $U_L26.5$ RNAs. Lines 2 and 3 show the nucleotide numbers for the transcription initiation, the translation initiation and termination, and the polyadenylylation (A) signal for U_L26 and $U_L26.5$ ORFs, respectively. All numbers shown are relative to the transcription initiation site of the U_L26 mRNA. The thick bar represents the coding sequences. Line 4 is a restriction endonuclease map drawn to scale with reference to lines B through NN, which are schematic representations of U_L26 and $U_L26.5$ ORFs used in studies reported here. Constructs AA, BB, CC, DD, and MM were constructed by inserting the translational stop codon into pRB4060 (3) at *Pml* I, *Mst* II, *Bss*HII, and *Hpa* I sites and at the ICP35 translation initiation site, respectively. The U_L26 ORF fused to *Bam*HI Z fragment cloned in pGEM3zf(+) as pRB4245 was mutagenized to give rise to II, JJ, KK, LL, HH, and GG with the aid of the Muta-Gene kit (Bio-Rad) in accordance with the manufacturer's recommendations. The 40-mer oligonucleotides used for this purpose were synthesized on an Applied Biosystems model 380B DNA synthesizer. Plasmids EE and FF were constructed by cleavage with *Xba* I and religation of constructs GG and HH to delete the first 10 and 33 amino acids of U_L26, respectively. Plasmid NN was constructed by deleting the DNA sequence between the ICP35 translation initiation site and its stop codon. All other constructs were described elsewhere (1, 3). Open quadrangles represent the *Bam*HI Z fragment used as the source of the $\alpha 4$ gene promoter and inserted in the proper transcriptional orientation relative to that of the U_L26 and $U_L26.5$ ORFs; filled ovals, the 20-amino acid CMV epitope described elsewhere (3); filled triangles, the inserted stop codon; filled rectangles, the DNA sequence encoding the IgG binding domain of protein A; P*, a new *Pml* I site created in conjunction with the insertion of the IgG binding domain of protein A. The new translational initiation codons by *in vitro* mutagenesis are marked "ATG." B, *Bam*HI; Ba, *Bal* I; Bs, *Bst*EII; E, *Eco*NI; H, *Hpa* I; K, *Kpn* I; Ms, *Mst* II; P, *Pml* I; Ps, *Pst* I; S, *Sal* I; X, *Xcm* I. Me, methionine translation initiation codon of the $U_L26.5$ open reading frame.

nuclease-treated rabbit reticulocyte lysate (Promega) and [^{35}S]methionine (DuPont/NEN), and the translation was then terminated either by the addition of a disruption buffer [0.05 M Tris, pH 7.0/8.5% (wt/vol) sucrose/5% (vol/vol) 2-mercaptoethanol/2% (wt/vol) sodium dodecyl sulfate] or by 1:20 dilution in phosphate-buffered saline containing cycloheximide at 100 μg/ml final concentration and various concentrations of protease inhibitors. After 6 hr of reaction in the presence of cycloheximide, the mixtures were denatured in disruption buffer, subjected to electrophoresis in polyacrylamide gels, electrically transferred to nitrocellulose sheets as described (3, 6), and finally exposed to Kodak X-Omat films.

Transfections and Superinfection of Cells. Transfections and superinfections were done as detailed (1, 3).

RESULTS

Delineation of the Functional Domains of the U_L26 Protein. Three sets of mutants of the U_L26 protease (Table 1) were tested. Set 1 was prepared for mapping of the U_L26 and $U_L26.5$ ORFs (3) and consists of three U_L26 genes into which were inserted in frame at various sites DNA sequences encoding a 20-amino acid CMV epitope. The second set consisted of 10 U_L26 gene constructs with either stop codons or deletions spanning various regions of the gene (Table 1). Set 3 consists of six U_L26 genes containing substitutions in predicted amino acid sequences starting in the region of amino acids 7–215. In each of the experiments described below, the U_L26 gene in each of these plasmids was expressed from an $\alpha 4$ gene promoter. In all but two experiments, the target of the protease was the $U_L26.5$ gene that was cloned in plasmid L and contained in frame the CMV epitope insert. The exceptions, clones P and J, contained the U_L26 gene containing the CMV epitope either in the coding sequences of both the U_L26 and $U_L26.5$ gene (plasmid J) or only in the U_L26 coding domain (plasmid P). Fig. 1 shows the structure of the ORFs and plasmid constructs.

In the typical experiment described below, the baby hamster kidney cells were transfected with plasmid L and one of the plasmids encoding a protease and were superinfected with HSV-1(F) at 39°C. The cell lysates were electrophoretically separated in denaturing gels, transferred to a nitrocellulose sheet, and treated with the HSV monoclonal antibody H725, which reacts with all $U_L26.5$ gene products, and CMV monoclonal antibody CH28-2, which reacts only with the $U_L26.5$ products that contain the CMV epitope. Because HSV-1(F) contains a temperature-sensitive lesion in the $\alpha 4$ gene specifying the major HSV-1 regulatory protein, it does not express its own U_L26 protease or substrate at 39°C. However, the α-gene transinducing factor (virion protein VP16) is functional at higher temperatures (7, 8) and transactivates the expression of the genes specifying both the

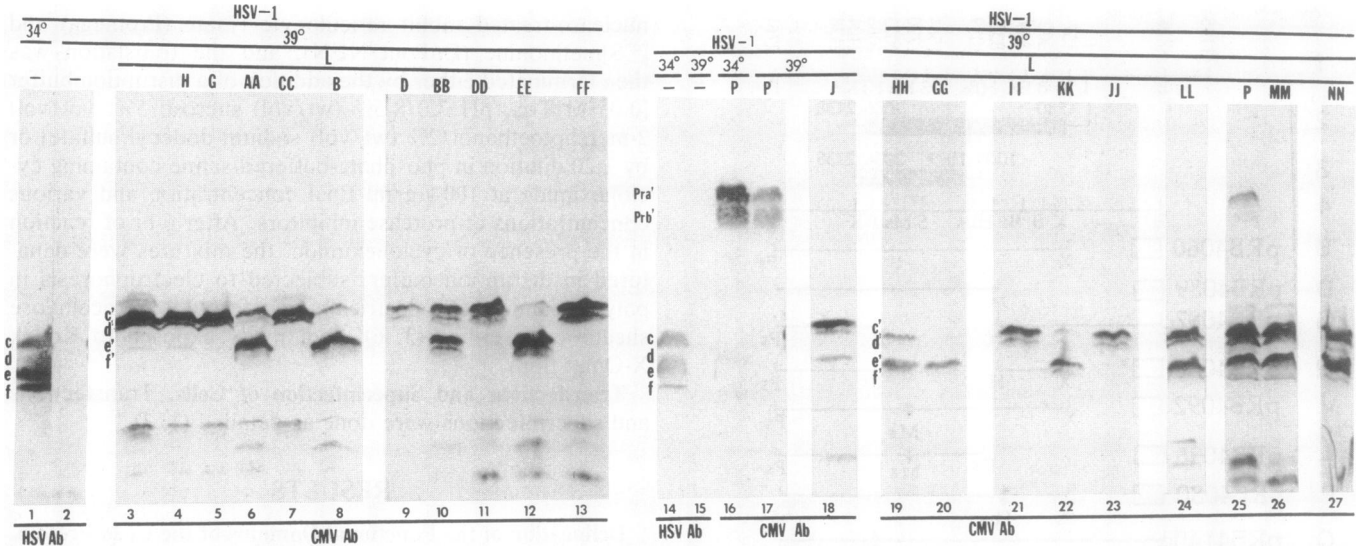

FIG. 2. Photographs of polypeptides from lysates of cells that were transfected with plasmid constructs and superinfected with HSV-1(F) at 34°C (34°) or at 39°C (39°); the polypeptides were electrophoretically separated in polyacrylamide gels, electrically transferred to a nitrocellulose sheet, and treated first with monoclonal antibody H725 to HSV-1 ICP35 (lanes HSV Ab) or CH28-2 to the CMV epitope (lanes CMV Ab) and then with goat anti-mouse IgG antibody coupled to peroxidase as described (1). As previously described (1, 6), the ICP35 bands e and f are the products of the cleavage of the proteins in bands c and d. The proteins migrating in bands c', d', e', and f' contain the CMV epitope and therefore migrate more slowly than the authentic proteins in bands c, d, e, and f, respectively. The procedures for denaturation, electrophoretic separation in denaturing polyacrylamide gels, electric transfer to nitrocellulose, and reaction with antibody were as previously described (1).

protease (U_L26) and the substrate ($U_L26.5$). The experiment produced the following five results as seen in Fig. 2.

(*i*) As expected from previous studies (1), the *$U_L26.5$* gene resident in the viral genome yielded protein bands reactive with the HSV monoclonal antibody at 34°C (lanes 1 and 14) but not at 39°C (lanes 2 and 15). Furthermore, the presence of the products of proteolytic cleavage (bands e and f) indicate that the viral protease encoded by *U_L26* was also made at 34°C.

(*ii*) The U_L26 protease encoded in the viral genome was not expressed at 39°C (lane 3). Thus, in the absence of the plasmid encoding the protease, the substrate encoded by plasmid L was made (bands c and d) but not cleaved to yield bands e and f, indicating that the protease encoded in the viral genome was not expressed at the nonpermissive temperature.

(*iii*) Only the precursor forms of ICP35, c and d derived from plasmid L, accumulated in cells transfected with the mutated *U_L26* genes in plasmids H (lane 4), G (lane 5), CC (lane 7), D (lane 9), DD (lane 11), FF (lane 13), II (lane 21), and JJ (lane 23). In these plasmids, the protease activity of the gene product was inactivated.

(*iv*) Both the precursor and product forms of ICP34.5 derived from plasmid L accumulated in cells transfected with the mutated *U_L26* genes AA (lane 6), B (lane 8), BB (lane 10), EE (lane 12), GG (lane 20), HH (lane 19), KK (lane 22), P (lane 25), MM (lane 26), and NN (lane 27).

(*v*) As noted above, in plasmid P the 20-amino acid epitope was inserted after amino acid 218—i.e., upstream from the coding domain of the substrate protein encoded by *$U_L26.5$*. The protease encoded by plasmid P cleaved itself and the ICP35 substrate (lanes 16, 17, and 25). Plasmid J contained the CMV insert after the amino acid 514 (Fig. 1). In the assays, reported previously and repeated here (lane 18), it cleaved the product of the U_L26 encoded in plasmid J itself. The only cleavage product detected in this assay was band e. Because the epitope was also inserted into the $U_L26.5$ protein, it is conceivable that the inserted 20-amino acid epitope interfered with and diminished the efficiency of the cleavage. We have previously reported that the protease encoded by plasmid Q (Fig. 1 and Table 1) encoded a protease that cleaved the product of the *$U_L26.5$* genes specified by other plasmids but not by the gene encoded in its own domain because the epitope inserted after amino acid 615 interfered with the cleavage (1).

The *U_L26* Gene Encodes a Serine Protease. Into the *U_L26* ORF used in these studies were inserted between the terminal amino acid and the stop codon the sequence encoding the 20-amino acid CMV epitope described elsewhere (1) and the 256-amino acid IgG binding domain of protein A (plasmid Y; Fig. 1). Transcripts of the coding domain of plasmid Y by the Sp6 RNA polymerase were translated in a nuclease-treated rabbit reticulocyte lysate in the presence of [^{35}S]methionine for 10 min. The cycloheximide was added to stop further translation, and as previously described (1), the translation product of the Y plasmid was incubated for another 6 hr to allow self-cleavage in the presence of protease inhibitors. The products of the reaction were then electrophoretically separated on denaturing polyacrylamide gels with the following four results as seen in Fig. 3.

(*i*) The products of the 10-min translation formed a single labeled polypeptide band containing the uncleaved protease (Pra) (lanes 15 and 28).

(*ii*) After 6 hr of reaction in the presence of cycloheximide (100 μg/ml) but in the absence of protease inhibitors, the translation mixture formed three bands corresponding to the intact translation product (Pra) and the amino-terminal (Prb) and carboxyl-terminal (PA) portions of the cleavage products of the translation (lanes 9, 10, 21, 22, and 33).

(*iii*) The amounts of cleavage products, Prb and PA, were reduced in translation mixtures treated in the presence of cycloheximide and lower concentrations of the serine protease inhibitors as shown in Fig. 3, lanes DFP, TPCK, TLCK, PMSF, and chymostatin. At the highest concentrations tested, the digestion products were not detected (lanes 1, 29, 34, 39, and 44). We should note that the inhibitory concentrations were in the millimolar range and that in general the protease appeared to be relatively resistant to the inhibitors tested.

(*iv*) The cleavage of the translation product Pra was not affected by the cysteine protease inhibitors iodoacetic acid (Sigma) and cystatin (Boehringer Mannheim; lanes 5–9 and

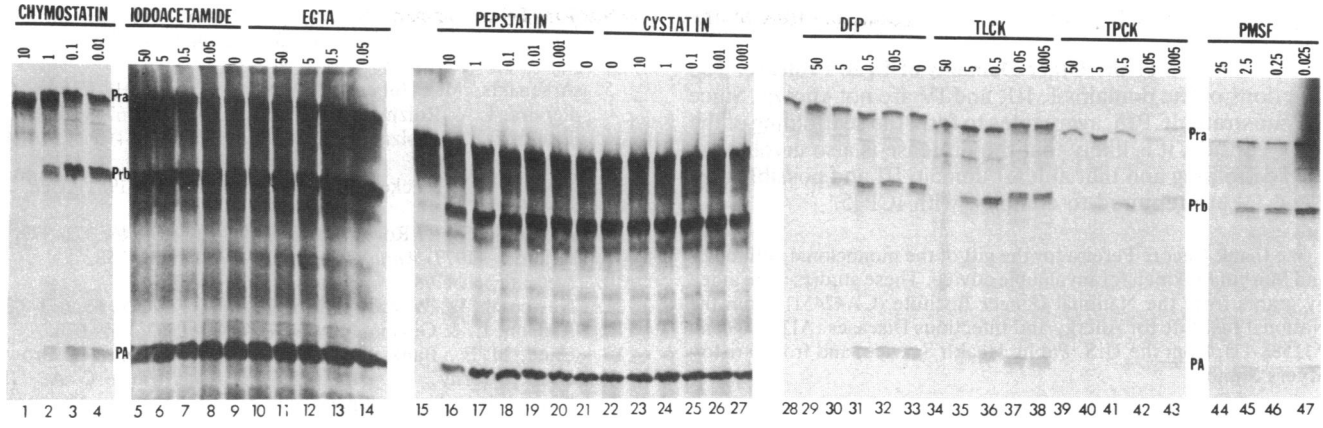

FIG. 3. Autoradiographic images of the electrophoretically separated polypeptides translated *in vitro* in a nuclease-treated rabbit reticulocyte lysate from the synthetic RNAs transcribed *in vitro* off the U_L26 ORF cloned in plasmid construct Y. The lanes shown represent portions denatured for electrophoresis immediately after the 10-min synthesis (lanes 15 and 28) or after an additional 6 hr of reaction in the presence of cycloheximide (100 μg/ml) alone or with the protease inhibitors shown (lanes 1–4, 16–27, and 29–47), with millimolar concentrations at the top of each lane. All of the protease inhibitors were dissolved in dimethyl sulfoxide prior to use. DFP, diisopropyl fluorophosphate (Sigma); TPCK, L-1-tosylamido-2-phenylethyl chloromethyl ketone (Sigma); TLCK, *N*-α-*p*-tosyl-L-lysine chloromethyl ketone (Sigma); PMSF, phenylmethylsulfonyl fluoride (Sigma); chymostatin was from Boehringer Mannheim.

22–27); by EGTA, a chelator and inhibitor of metalloprotease (lanes 12–14); or by the aspartic acid protease inhibitor pepstatin (Boehringer Mannheim; lanes 15–21).

These results are consistent with the hypothesis that U_L26 gene product is a serine protease.

DISCUSSION

The Domains of the U_L26 Protease. The results shown in Fig. 2 and summarized in Fig. 4 indicate that the U_L26 protease consists of four domains of which two are dispensable and two are not. The dispensable domains I and IV extend from amino acid 1 through amino acid 9 but not to amino acid 32 and from the carboxyl terminus (amino acid 635) to at least amino acid 307 but not to amino acid 287, respectively. The domain III is less precisely defined; it extends from at least amino acid 218 to at most amino acid 306. This domain can be displaced by at least 20 amino acids (CMV epitope insertion after amino acid 218) relative to the amino-terminal portion of the protease. Domain II is also not dispensable and is imprecisely defined as being located between amino acids 10 and 218.

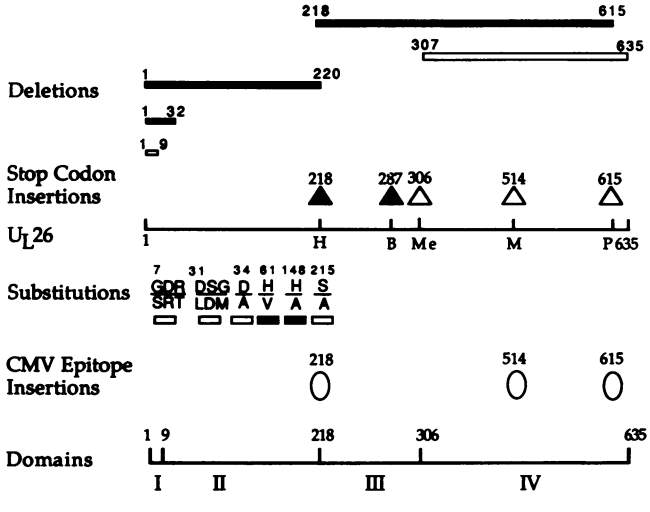

FIG. 4. Schematic representation of the results of the mutagenesis studies. The numbers refer to the amino acid numbers predicted from the nucleotide sequence of the U_L26 ORF as reported by McGeoch *et al.* (2). For each stop codon insertion, the amino acid shown immediately precedes the site of insertion. The amino acids are identified by single letter code. Open symbols indicate that the protease was functional. Closed symbols indicate the protease was inactivated by mutagenesis. The line at the bottom of the figure identifies the domains of the protease described in the text. B, *Bst*EII; H, *Hpa* I; M, *Mst* II; P, *Pml* I. Me, methionine translation initiation codon of the $U_L26.5$ gene.

The Catalytic Domain of the U_L26 Protease. The studies with protease inhibitors are consistent with the hypothesis that the enzyme encoded by the U_L26 ORF belongs to either the chymotrypsin or subtilisin superfamilies of serine proteases (9, 10). A shared property of the two serine protease superfamilies are active sites containing histidine, aspartic acid, and serine amino acids.

The substrate of the protease, ICP35, has been reported to play a role as a scaffolding protein in the assembly of the HSV capsid (11). The sequence of events in the replication of other herpesviruses is similar, and homologues of ICP35 have been reported (12). Of particular interest was the question of whether homologues of the U_L26 ORF in other herpesviruses contained conserved histidine, aspartic acid, and serine amino acids play a role in the proteolytic activity of U_L26 protease. Nucleotide sequence comparisons indicate that the ORF 33 of varicella–zoster virus and the CMV U_L80 ORF of human CMV encode homologues of the U_L26 ORF of HSV-1 (2, 13, 14). The amino acid sequence comparison between HSV U_L26, CMV U_L80, and varicella–zoster virus gene *33* protein (not shown) indicated the amino terminus is the most conserved domain of the U_L26 protease. The conclusion that the protease maps in the amino-proximal domain of the U_L26 ORF is consistent with the observation that ICP35, the product of $U_L26.5$ ORF, is devoid of enzymatic activity; therefore, we focused on the conserved amino acids in the amino-proximal domain of U_L26. The substitutions in the amino acids encoded in plasmids GG, II, JJ, KK, and LL probed Asp-31, Ser-32, Asp-34, His-61, His-148, and Ser-215. The results indicate that the only amino acids whose substitutions abolished enzymatic activity were the conserved histidines at positions 61 and 148. In anticipation of more defined mapping studies, the catalytic domain of the protease most likely maps in domain II of the U_L26 protease.

The studies reported here showed that at least portions of domains II and III are both essential and sufficient for proteolytic activity. Of particular interest is the observation that domains II and III do not contain the coding sequences of the ICP35 substrate and that they can be separated by at

least 20 amino acids without affecting the function of the protease.

The Function of the Other Domains of U_L26 Protease. The functions of the domains I, III, and IV are not known. Since the substrate, ICP35, aggregates to form the scaffolding of the HSV capsids, it is likely that the protease is also involved in the scaffolding and that at least domain III and possibly also I and IV are required to complex with ICP35.

We thank Lenore Pereira for the gift of the monoclonal antibodies and Marvin Makinen for invaluable advice. These studies were aided by grants from the National Cancer Institute (CA47451), from the National Institute for Allergy and Infectious Diseases (AI124009 and AI1588-11), from the U.S. Public Health Service, and from Bristol-Myers Squibb.

1. Liu, F. & Roizman, B. (1991) *J. Virol.* **65,** 5149–5156.
2. McGeoch, D. J., Dalrymple, M. A., Davison, A. J., Dolan, A., Frame, M. C., McNab, D., Perry, L. J., Scoot, J. E. & Taylor, P. (1988) *J. Gen. Virol.* **69,** 1531–1574.
3. Liu, F. & Roizman, B. (1991) *J. Virol.* **65,** 206–212.
4. Ejercito, P. M., Kieff, E. D. & Roizman, B. (1968) *J. Gen. Virol.* **67,** 1759–1816.
5. Arsenakis, M., Hubenthal-Voss, J., Campadelli-Fiume, G., Pereira, L. & Roizman, B. (1986) *J. Virol.* **60,** 674–682.
6. Braun, D. K., Roizman, B. & Pereira, L. (1984) *J. Virol.* **49,** 142–153.
7. Post, L. E., Mackem, S. & Roizman, B. (1981) *Cell* **24,** 555–565.
8. Batterson, W. & Roizman, B. (1983) *J. Virol.* **46,** 371–377.
9. Kraut, J. (1977) *Annu. Rev. Biochem.* **46,** 331–358.
10. Neurath, H. (1983) *Science* **224,** 350–357.
11. Newcomb, W. W. & Brown, J. C. (1991) *J. Virol.* **65,** 613–620.
12. Robson, L. & Gibson, W. (1989) *J. Virol.* **63,** 669–676.
13. Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., III, Kouzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, K. M. & Barrell, B. G. (1990) *Curr. Topics Microbiol. Immunol.* **154,** 127–169.
14. Davison, A. J. & Scott, J. E. (1986) *J. Gen. Virol.* **67,** 1759–1816.