



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Identification of the Active Site Residues in the nsP2 Proteinase of Sindbis Virus

ELLEN G. STRAUSS,¹ RAOUL J. DE GROOT,² RANDY LEVINSON,³ AND JAMES H. STRAUSS

Division of Biology, California Institute of Technology, Pasadena, California 91125

Received July 17, 1992; accepted August 19, 1992

The nonstructural polyproteins of Sindbis virus are processed by a virus-encoded proteinase which is located in the C-terminal domain of nsP2. Here we have performed a mutagenic analysis to identify the active site residues of this proteinase. Substitution of other amino acids for either Cys-481 or His-558 completely abolished proteolytic processing of Sindbis virus polyproteins *in vitro*. Substitutions within this domain for a second cysteine conserved among alphaviruses, for four other conserved histidines, or for a conserved serine did not affect the activity of the enzyme. These results suggest that nsP2 is a papain-like proteinase whose catalytic dyad is composed of Cys-481 and His-558. Since an asparagine residue has been implicated in the active site of papain, we changed the four conserved asparagine residues in the C-terminal half of nsP2 and found that all could be substituted without total loss of activity. Among papain-like proteinases, the residue following the catalytic histidine is alanine or glycine in the plant and animal enzymes, and the presence of Trp-559 in alphaviruses is unusual. A mutant enzyme containing Ala-559 was completely inactive, implying that Trp-559 is essential for a functional proteinase. All of these mutations were introduced into a full-length clone of Sindbis virus from which infectious RNA could be transcribed *in vitro*, and the effects of these changes on viability were tested. In all cases it was found that mutations which abolished proteolytic activity were lethal, whether or not these mutations were in the catalytic residues, indicating that proteolysis of the nonstructural polyprotein is essential for Sindbis replication. © 1992 Academic Press, Inc.

INTRODUCTION

Proteins encoded by plus-stranded RNA viruses are commonly produced as polyproteins which are post-translationally processed by one or more virally encoded proteolytic enzymes (reviewed in Strauss, 1990). Viral proteinases have been described which resemble cellular proteinases belonging to three distinct families: the aspartate proteases that resemble pepsin, the serine/cysteine proteinases that resemble trypsin or chymotrypsin, and the thiol proteases that resemble papain. The best characterized viral proteases are the aspartate proteases encoded by retroviruses. The residues making up the catalytic triad (Asp, Thr, Gly) of the HIV proteinase have been confirmed by site-specific mutagenesis (Loeb *et al.*, 1989), and the structure of the proteinase has been determined to high resolution by X-ray crystallography (Navia *et al.*, 1989). The three-dimensional folding of the viral proteinase resembles pepsin, suggesting a common evolutionary origin.

Proteinases similar to trypsin or chymotrypsin have been identified in picornaviruses; alphaviruses; plant

como-, nepo-, and potyviruses; coronaviruses; and flaviviruses (and their proposed relatives pestiviruses and hepatitis C virus.) Originally these domains were predicted to have proteolytic activity based on the presence of certain conserved amino acid residues and on the basis of protein-modeling studies (Bazan and Fletterick, 1989; Boege *et al.*, 1981; Gorbalenya *et al.*, 1989; Hahn *et al.*, 1985). The catalytic triad of trypsin is composed of a serine, a histidine, and an aspartic acid; in the flavivirus NS3 proteinase (Chambers *et al.*, 1990), the pestivirus p80 proteinase (Wiskerchen *et al.*, 1991), and the alphavirus capsid proteinase (Hahn and Strauss, 1990), such a triad has been confirmed by mutagenesis. Furthermore, the crystal structure of the Sindbis alphavirus capsid protein has been determined to 3 Å, and the C-terminal 151 amino acids are folded into a structure remarkably similar to that of chymotrypsin, with His-141, Asp-163, and Ser-215 forming the catalytic site (Choi *et al.*, 1991b). The similarity in the folding of this capsid protein to that of chymotrypsin suggests that the virus acquired a proteinase from a host at some time in the past, and that all of the viral chymotrypsin-like proteinases will in fact be similar in structure.

The poliovirus 3C proteinase and the comparable 3C-like proteinases found in plant viruses and coronaviruses are also believed to be related to chymotrypsin, but in these cases the catalytic serine has been replaced by cysteine (reviewed in Bazan and Fletterick,

¹ To whom reprint requests should be addressed.

² Present address: Academic Hospital Leiden, P.O. Box 320, NL-2300 AH Leiden, The Netherlands.

³ Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448.

1990; Goldbach, 1990; Harris *et al.*, 1990). Mutagenesis experiments and X-ray structure experiments should shortly confirm or disprove this prediction.

Recently, we proposed that the nonstructural proteinase of Sindbis alphavirus was related to papain, thus representing a third family of viral proteinases (Hardy and Strauss, 1989; Strauss and Strauss, 1990), and Gorbalenya *et al.* (1991) proposed that papain-like proteinases were present in several RNA viruses. In papain, a cysteine and a histidine form a catalytic dyad, and there is conflicting evidence as to whether an asparagine residue is also an essential component of the active site (Higaki *et al.*, 1987; Kamphuis *et al.*, 1985). Other proposed papain-like proteinases in RNA viruses include two domains within ORF1a of coronaviruses identified by protein modeling studies (Baker *et al.*, 1991; Lee *et al.*, 1991) and the HC-Pro protein of potyviruses, in which the catalytic Cys and His residues have been identified by mutagenesis (Oh and Carrington, 1989). Interestingly, in a number of cases viruses encode proteinases of more than one type. Coronaviruses, for example, encode one or two proteins with sequences characteristic of a papain-like proteinase as well as one with similarity to a 3C-like proteinase within ORF1a (Boursnell *et al.*, 1987; Lee *et al.*, 1991). Similarly, in potyviruses there is both a 3C-like proteinase (the 49-kDa protein) as well as a papain-like proteinase (the HC or helper component protein) (Oh and Carrington, 1989). Finally, as described above, the alphaviruses possess a chymotrypsin-like structural proteinase and a papain-like nonstructural proteinase.

In this paper we report that Cys-481 and His-558 have been identified as the catalytic dyad of the Sindbis virus nsP2 nonstructural proteinase. Furthermore, we present data to show that none of the asparagine residues in the proteinase domain of Sindbis nsP2 that are conserved among alphaviruses are absolutely required for proteolytic activity, but that Trp-559, adjacent to His-558, is essential for function.

MATERIALS AND METHODS

Site-specific mutagenesis

For mutagenesis, the *Bgl*II (nt 2288) to *Pst*I (nt 3949) fragment from Toto50, a cDNA clone containing the entire sequence of Sindbis virus from which infectious RNA can be transcribed with SP6 polymerase (Rice *et al.*, 1987), was inserted into the replicative form of M13mp18, which had been digested with *Bam*HI and *Pst*I, to form M13mp18T2.155. Uracil-containing single-stranded DNA was prepared from phage by the method of Kunkel (1985) and used as the template for oligonucleotide-directed mutagenesis (Zoller and

Smith, 1984), after removal of low-molecular-weight material using NACS PrePac cartridges (Bethesda Research Laboratories). Both degenerate oligonucleotides capable of producing multiple mutations and oligonucleotides containing a single altered codon were used. Single-stranded DNAs from mutated bacteriophages were sequenced using appropriate sequencing primers and Sequenase (United States Biochemical) according to the manufacturer's instructions. Mutated fragments 1236 nt in length were excised from M13mp18T2.155 replicative form with *Cla*I and *Pst*I and inserted into the full-length clone by a three-piece ligation with the *Pst*I(3948) → *Spe*I(5262) and *Spe*I(5262) → *Cla*I(2712) fragments of Toto50. Full-length clones were checked for the presence of the mutations by direct sequencing of the double stranded plasmid DNA as previously described (Shirako *et al.*, 1991).

Transcription and translation

Miniprep DNA (Sambrook *et al.*, 1989) was treated with RNase A, linearized with *Xho*I, and treated with Proteinase K followed by repeated phenol extraction and ethanol precipitation. Linearized DNA was resuspended in RNase-free TNE and frozen at -70° until use. RNA transcripts were synthesized with SP6 polymerase as previously described (Rice *et al.*, 1987). Approximately 10 to 50 ng of RNA were translated in 9 μ l of nuclease-treated, methionine-depleted reticulocyte lysate (Promega) supplemented with 20 μ M of an amino acid mixture minus methionine and 10–15 μ C [35 S]methionine, for 90 min at 30° . Translation mixes were diluted 1:1 with 2X Laemmli sample buffer containing 10% β -mercaptoethanol, and analyzed on 10% SDS-containing polyacrylamide gels (Laemmli, 1970). Gels were fluorographed using En 3 Hance (NEN Research Products) according to the manufacturer's instructions.

Infectivity assays

RNA transcribed *in vitro* was transfected onto confluent monolayers of BHK cells in six-well cluster dishes (Costar) using the DEAE-Dextran method described in Rice *et al.* (1987). For plaque assays the monolayers were overlaid with Eagle's medium containing 10% fetal bovine serum and 1% agarose. Plates were incubated for 2 days at 30° and stained for at least 8 hr at 30° with neutral red.

RESULTS

Translation of Sindbis RNA

During infection by Sindbis virus, the viral genome of 11.7 kb serves as a messenger for the translation of

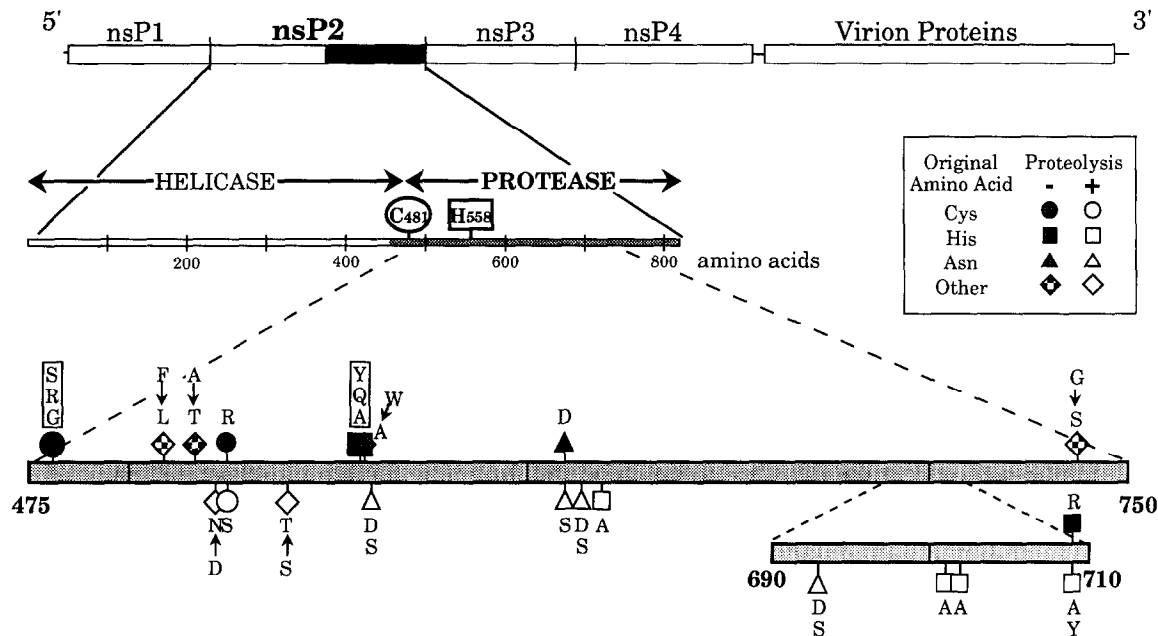


Fig. 1. Map of the Sindbis virus genome indicating location of the protease and the mutations tested. The second and third lines are expanded views of nsP2 showing all of the site-directed mutations made as well as the temperature-sensitive RNA⁻ mutants previously mapped (Hahn *et al.*, 1989). Solid-stalked symbols indicate reduced or absent proteolytic activity; open symbols have wild type enzymatic activity as determined *in vitro*. The shape of the symbol indicates the parental wild type amino acid; circles for Cys, squares for His, and triangles for Asn. Checkered diamonds indicate that the mutants are temperature-sensitive for proteolysis *in vivo* (Hardy *et al.*, 1990). These substitutions are Phe-509 → Leu, Ala-517 → Thr, Asp-522 → Asn, and Gly-736 → Ser.

the nonstructural proteins that form the viral replicase (Strauss and Strauss, 1986). Two polyproteins are produced upon translation. The first terminates at an in-frame opal codon at nucleotides 5748–5750 and this polyprotein of 1896 amino acids is subsequently cleaved into nonstructural proteins nsP1, nsP2, and nsP3, numbered in order from 5' to 3' along the viral genome. The second polyprotein of 2513 amino acids is produced by readthrough of the opal codon and contains in addition the sequences of nsP4; cleavage of this longer polyprotein produces nsP1, nsP2, nsP3, and nsP4, as well as nsP34, which is also thought to be functionally important (de Groot *et al.*, 1990). The genome organization and cleavage sites in the polyproteins are schematically illustrated in Fig. 1.

When Sindbis RNA, whether derived from the virion or from SP6 transcripts, is translated in a cell-free system, cleavage occurs at all three sites in the polyproteins to produce mature nsP1, nsP2, nsP3, and nsP4 and various intermediates (de Groot *et al.*, 1990, 1991; Hardy and Strauss, 1989). In addition some unprocessed P123 and P1234 usually remain. The polypeptides produced in such a translation are shown in Fig. 2A (labeled "Wild Type"). The proteinase that cleaves all three sites in the polyproteins has been mapped by deletion studies to the C-terminal half of nsP2; only deletions that invaded the domain between amino

acids 475 and 728 abolished all activity of the enzyme, and both nsP2 and polyproteins containing nsP2 were found to be proteolytically active (de Groot *et al.*, 1990; Hardy and Strauss, 1989). Hardy and Strauss (1989) proposed that the enzyme is a papain-like proteinase in which Cys-481 and one of the five conserved histidines in this region form the active site of the enzyme.

Cys-481 is required for cleavage

To confirm that Cys-481 is required for activity of the nsP2 proteinase, we changed this residue to serine, glycine, or arginine. We also examined the effect of changing Cys-525, one of the two remaining conserved cysteine residues in the C-terminal half of nsP2, to serine or arginine, as well as changing Ser-535, which is found in a domain of limited similarity to the active site serine of serine proteinases, to threonine. The mutations were introduced into a full-length cDNA construct, and RNA was transcribed with SP6 RNA polymerase and translated in rabbit reticulocyte lysates. Representative results with the substitutions for Cys-481 and Cys-525 are shown in Fig. 2, results for Ser-535 are shown in Fig. 3, and a summary of these results is given in Table 1. In the case of every mutant tested, multiple independent clones were examined to guard against the possibility that extraneous changes

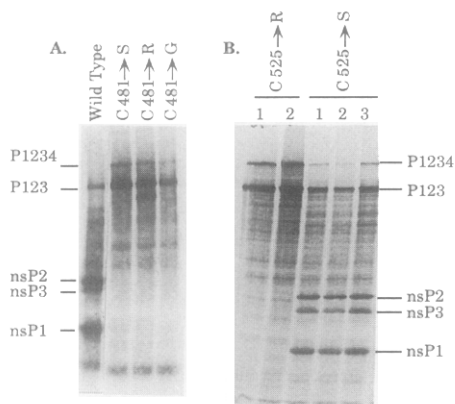


FIG. 2. Cleavage by Sindbis nsP2 with mutations in conserved cysteines. Sindbis cDNA constructs were made that had various substitutions for Cys-481 or Cys-525 of nsP2. RNA was transcribed from linearized full-length cDNA constructs with SP6 polymerase and translated in a reticulocyte lysate in the presence of [³⁵S]methionine for 90 min at 30°. The products were analyzed on 10% SDS-containing polyacrylamide gels. The positions of the mature processed proteins nsP1, nsP2, and nsP3 are indicated, as are those of the precursors P123 and P1234. Note that nsP4 is rapidly degraded under conditions of *in vitro* translation and requires special treatment to be visualized (de Groot *et al.*, 1991). (A) substitutions at Cys-481 as indicated. Wild Type, RNA transcribed from the wild type construct. (B) substitutions at Cys-525 as indicated. Two independent constructs of Cys-525 → Arg and three independent constructs of Cys-525 → Ser are shown to indicate the reproducibility of the phenotype observed.

other than the mutation of interest might have affected the results.

Change of Cys-481 to serine, glycine, or arginine resulted in abolishment of proteolytic activity. However, at least one substitution for Cys-525 and for Ser-535 resulted in an active enzyme. Taken together, these results are consistent with the hypothesis that Cys-481 is in fact a component of the active site of the nsP2 proteinase.

His-558 is also required for cleavage

All five conserved histidines in the proteinase region, His-558, His-619, His-701, His-702, and His-709, were site specifically changed to alanine. In addition, His-558 was also changed to glutamine and tyrosine, and His-709 was also changed to tyrosine and arginine. Results for a representative sample of these mutants are shown in Fig. 3, and the results for all of the His mutants are summarized in Table 1. All three changes in His-558 totally abolished proteolytic activity. In contrast, change of any of the other four histidine residues to alanine, or change of His-709 to tyrosine, resulted in an active protease. Thus His-558 is uniquely required for proteolytic activity, and we postu-

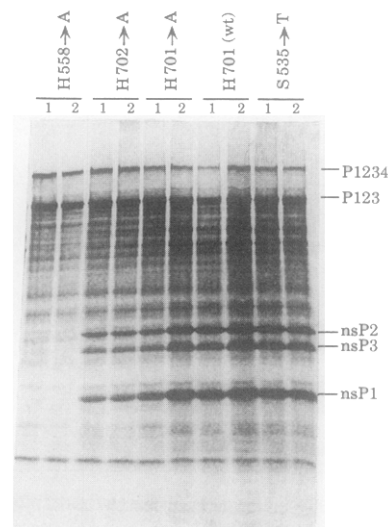


FIG. 3. Cleavage by Sindbis nsP2 containing various mutations in conserved histidines or in a conserved serine. Acrylamide gel electrophoresis of translations of Sindbis constructs with various substitutions for conserved histidine residues or a conserved serine in the proteolytic domain of nsP2. The same methods and conventions are used as in Fig. 2.

late that this residue forms part of the active site of the nsP2 proteinase.

As a word of caution on this method of determining catalytic site residues of a proteinase, note that two other substitutions, presumably not in the catalytic

TABLE 1
MUTAGENESIS OF CONSERVED CYS AND HIS RESIDUES

Original amino acid	Mutant amino acid	Processing	Viability (plaque size) ^a
Cys-481	Ser	None	Lethal
	Arg	None	Lethal
	Gly	None	Lethal
Cys-525	Ser	Normal	Large
	Arg	None	Lethal
Ser-535	Thr	Normal	Not detn.
His-558	Ala	None	Lethal
	Gln	None	Lethal
	Tyr	None	Lethal
His-619	Ala	Normal	Not. detn.
His-701	Ala	Normal	Large
His-702	Ala	Normal	Large
His-709	Ala	Normal	Large
	Tyr	Normal	Large
	Arg	None	Lethal

^a "Lethal" indicates that no plaques were found after transfection. Wild type constructs in this assay give "Large" plaques. "Not detn." means that the viability was not tested.

TABLE 2
MUTAGENESIS OF OTHER AMINO ACIDS

Original amino acid	Mutant amino acid	Processing	Viability (plaque size) ^a
Asn 561	Asp	Normal	Small
	Ser	Normal	Mixed sizes
Asn 609	Asp	Almost none	Minute
	Ser	Reduced	Large
Asn 614	Asp	Enhanced	Lethal
	Ser	Normal	Medium
Asn 693	Asp	Variable	Medium
	Ser	Variable	Lethal
Trp-559	Ala	None	Lethal

^a "Lethal" indicates that no plaques were found after transfection. Wild type constructs in this assay give "Large" plaques.

site, also abolished the enzymatic activity: His-709 → Arg and Cys-525 → Arg. It appears that the large, positively charged arginine residue perturbs the structure such that the protein is no longer proteolytically active. To rule out such structural alterations, in most cases we have tested more than one substitution for the wild type residue.

Asparagine is not required for proteolysis

It has been suggested that in addition a cysteine and a histidine an asparagine residue is also important for catalysis by papain (Garavito *et al.*, 1977; Higaki *et al.*, 1987). To investigate the possible role of an asparagine residue in the catalytic activity of the nsP2 proteinase, we mutagenized all four asparagine residues within the catalytic domain that are conserved among alphaviruses, namely Asn-561, Asn-609, Asn-614, and Asn-693. In each case we changed the asparagine to both aspartic acid and to serine, and the results are summarized in Table 2. Substitution of either serine or aspartic acid for Asn-561 and Asn-614 produced enzymes with near wild type activity. Substitutions at Asn-693 gave variable amounts of processing, which depended upon the particular preparation of reticulocyte lysate used. Some experiments showed almost wild type processing (data not shown), while in the translation shown in Fig. 4 Asp-693 and Ser-693 both show markedly reduced processing. Similar variability in processing during *in vitro* translation in reticulocyte lysates has been noted for constructs containing the NS3 proteinase of flaviviruses (Preugschat *et al.*, 1991). The reason for such variability is not clear, but may be due to variation in the translation efficiency or in the fidelity of folding of the translated proteins.

Asn-609 → Ser gave reduced and variable processing; in the translation shown in Fig. 4 normal products are made but in very reduced amounts. For Asn-609 → Asp certain constructs seemed completely inactive, while others gave very incomplete processing, the extent of which was again dependent upon the reticulocyte preparation used. The severely reduced activity of the enzyme containing the Asn-609 → Asp mutation was not due to some additional lesion elsewhere in the protein. As a control a restriction fragment of 375 nucleotides containing this change from the mutant construct was excised, sequenced in its entirety, and reinserted into the Toto background, and the resulting construct was also nearly inactive. Presumably the presence of a charged residue at this position alters the overall conformation of the protein, leading to reduced activity. Multiple independent isolates of this mutation all had similar phenotypes.

We conclude that none of the conserved asparagines is absolutely required for proteolysis by the nsP2 proteinase, although substitutions for Asn-609 markedly reduce the activity of the enzyme, apparently without affecting the specificity for the cleavage sites. These results suggest that an asparagine residue is not a component of the active site of the enzyme.

Role of Trp-559 in proteolysis

In the case of most papain-family proteinases, the amino acid C-terminal to the catalytic histidine is small. In plant and animal proteinases, glycine or alanine is present at this position, whereas in other viral proteinases cysteine, serine, or valine is found. However, all of the alphaviruses have tryptophan in this position. To determine whether Trp-559 was essential for proteoly-

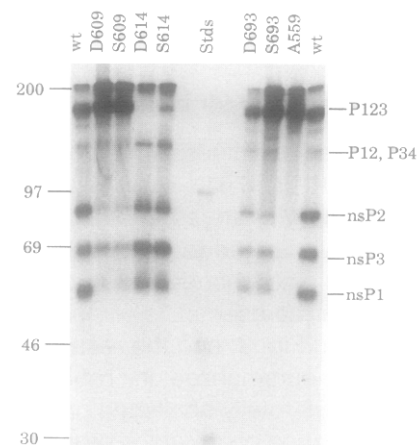


FIG. 4. Translation of constructs with substitutions in conserved asparagine residues or in Trp-559. Methods and conventions are as in Fig. 2.

sis, we examined the mutant Trp-559 → Ala. When this construct was transcribed and translated *in vitro*, it was totally inactive. We conclude that Trp-559 is required for activity of the nsP2 protease.

Infectivity of Sindbis virus containing mutated proteinases

We have tested a number of these constructs for viability, since the mutations had been inserted into a full-length clone of Sindbis virus from which infectious RNA can be transcribed *in vitro*. BHK cells were transfected with this RNA to look for the formation of plaques and the results of these experiments are shown in the last columns of Tables 1 and 2. In all cases in which proteolysis was abolished, no plaques could be detected after transfection, indicating that proteolytic processing of the nonstructural polyprotein is essential for Sindbis replication. These lethal mutations include all substitutions tested for the catalytic Cys-481 and His-558, as well as for Arg-709, Arg-525, and Ala-559.

The effects of the asparagine substitutions are not as straightforward. Asp-561 and Ser-561 gave normal processing during *in vitro* translation but produced small plaques and mixed plaque sizes respectively after transfection (Table 2). On the other hand, the Ser-609 virus, which showed very little processing, gave wild type large plaques and the Asp-609 variant, which scarcely processed at all, gave minute plaques. Still more interesting are the results with the substitutions at Asn-614. Ser-614 gave processing similar to wild type and made slightly smaller plaques than the wild type. Asp-614 gave normal amounts of the fully processed proteins nsP1, nsP2, and nsP3 and seemed to process more effectively, since no uncleaved P123 was seen after *in vitro* translation (Fig. 4), but surprisingly, Asp-614 is lethal. Ser-693, which produced variable processing results, is also lethal.

DISCUSSION

We have shown that Cys-481 and His-558 are required for proteolytic activity by the Sindbis virus nsP2 proteinase. Any change in either of these residues abolished all activity by the enzyme and we propose that these two residues form the catalytic dyad of a papain-like proteinase. We also found that Trp-559 is required for proteolytic activity. All of the mutations which we made in this study, as well as temperature-sensitive mutations mapped in the proteinase domain of nsP2 (Hahn *et al.*, 1989) are shown diagrammatically in Fig. 1. It is notable that in some cases substitution with an arginine residue killed the enzymatic activity, whereas substitution of other residues did not.

A great deal has been learned about the thiol papain-like proteinases over the last few years. These proteinases include the prototype papain from papaya and other plant enzymes such as actinidin from kiwi-fruit and bromelain from pineapple as well as animal lysosomal enzymes such as the various cathepsins and calpain (reviewed in Baker and Drenth, 1987; Brocklehurst, 1987). The crystal structure of papain has been determined to 1.65 Å resolution (Kamphuis *et al.*, 1984), but structural information is lacking for other enzymes. Because the amino acid similarities around the active site residues or throughout the rest of the proteins are limited and even the spacing between the active cysteine and histidine is quite variable, it remains to be determined what the relationships are between the plant, animal, and viral enzymes. In particular, it is not clear whether the folding of the viral proteins resembles that of the plant or animal proteins; similar folding would be strongly indicative of the descent of animal, viral, and plant proteins from a common ancestral enzyme.

In addition to the catalytic Cys-25 and His-159 residues of papain, other amino acid residues have been suggested to be important for catalysis. One such residue is Asn-175 which occupies precisely the same position in a possible catalytic triad of papain as Asp-102 in the catalytic triad of trypsin (Garavito *et al.*, 1977; Higaki *et al.*, 1987). Furthermore, there is considerable sequence identity among papain-like enzymes around this amino acid. In order to examine whether an asparagine might be required for activity of the Sindbis nsP2 proteinase, all four asparagine residues within the proteinase domain that are conserved among alphaviruses were subjected to mutagenesis. In every case at least one substitution was partially tolerated, and we conclude that no asparagine residue is absolutely essential for proteolysis. This is not to say that changes at various asparagine residues were without effect: changes at both Asn-609 and Asn-693 reduced the extent of proteolysis, and Asp-614 appeared to enhance proteolysis.

If the role of Asn-175 in papain were to stabilize the active site histidine in a tautomericly favorable conformation through its hydrogen bonding to a nitrogen in the imidazole ring of His-159, it is conceivable that Trp-559 adjacent to His-558 could fulfill a related role in alphaviruses. As described above, the presence of a tryptophan following the active site histidine is exceptional among all of the papain-like enzymes described to date. In this model Trp-559 might interact with Trp-482 adjacent to the catalytic cysteine to maintain His-558 and Cys-481 in an active conformation. In any event, an important role for this tryptophan residue is

Plant Proteinases		25	159
Papain	(11)	A V T P V K N Q G S C G S C W A F S A V V T I E G I	(118) K V D H A V A A V G Y N
Actinidin	(11)	. . V D I . S . . E . . G C I A . V . . .	(121) A . . H . I V I . . . G
Bromelain	(12)	. . . S N P . . A C . . . A I A . V . S .	(116) S L N H . . T . I . . G
CP1	(123) Q . . . C . S . T T G N V . . Q	(128) S L . H G I L I . . . S
CP2	(132) I . D . . Q . . . C . S . T T G S T . . A	(123) E L . H G . L V . . . G
		*	*
Animal Proteinases			
Cathepsin H	(12)	V . S A . . . C . T . . T T G A L . S A	(124) . . N H . . L G
Cathepsin L	(11)	C Q . . . C S G C L . . Q	(122) D L . H G . L V . . . G
Cathepsin B	(15)	T I K E I R D C . . . G . . E A . S D R	(152) M G G H . I R I L . W G
Calpain	(94)	. T R T D I C . . A L . D C . L L A . I G S L T L N	(141) V K G H . Y S V T A P K
		*	*
Viral Proteinases			
Coronaviruses			
IBV	[1235]	R D N F L I L E W R D . N C . I S . . I . L L Q A A	(147) N S G H C Y T Q A A G Q
MHV-1	[1695]	C G N Y F A F K Q . N N N C Y I N V . C L M L Q H L	(141) S . A H . Y T H . K C K
MHV-2	[1099]	C G F Y S P A I E R T N C . L R . T L I V M Q S L	(133) N D C H S M . V . D G K
		*	*
Togaviruses			
RUB	[1137]	R A S T R G G E L D P N T C . L R A . A N V A Q A A	(105) P T G H F . C . . G G
SIN	(467)	T P R A N P F S C K T N V C . . K A L E P I I L A T A	(63) P . A H W D N S P . T R
SF	(464)	. A P V D A F . N K A N V C . . K . L . P V L D T A	(56) N H W D N R P . G R
ONN	(464)	P D P T D V F . N K A N V C . . K A L . P V L K T A	(55) N H W D N S P S P .
RR	(464)	S T A V D P F . N K A K V C . . K C L . Q V L Q T A	(55) N H W D N R P . G R
VEE	(463)	Q M A F D T F . N K A N V C . . K V C L . P I L D T A	(56) N H W D N R P . G K
		*	*
Potyviruses			
TEV	[635]	L N E E K M Y I A N E G Y C Y M N I F F A L L V N V	(57) . T M H V L D S Y . S R
PPV	[638]	. K G G A M F I A K A G Y C Y I N I F L A M L I N	(57) . I F H V . D E F . S L
TVMV	[585]	E I S N L M Y I A K E G Y C Y I N I F L A M L V N V	(57) . T I H V . D S Y . S L
PVY	[612]	G D S E M L Y I A K Q G Y C Y I N V F L A M L I N	(57) Q T C H V . D S F . S Q
		*	*
Other Plant viruses			
BaYMV	(129)	V Q . F I A F D F A H G Y C Y L S L F I P L S F R .	(56) L Q F H V S D . R . L
HAV	(148)	S R N G S L A . F G Q G Y C Y L S A I . D S A R W R	(37) H . Y H V . V D A E . Q
		*	*

Fig. 5. Aligned amino acid sequences of papain-like proteinases in the vicinity of the conserved Cys and His catalytic residues. Alignment of the sequences in according to Gorbalenya *et al.* (1991). Sequences are from the following sources: papain (Mitchel *et al.*, 1970); actinidin (Carne and Moore, 1978); bromelain, stem bromelain (Ritonja *et al.*, 1989); CP1 and CP2, cysteine proteinases I and II from *Dictyostelium discoideum* (Pears *et al.*, 1985); cathepsin H and cathepsin L, human cathepsins (Wiederanders *et al.*, 1991); cathepsin B, human cathepsin B (Ritonja *et al.*, 1985); calpain, chicken calpain (Ohno *et al.*, 1984); IBV, avian infectious bronchitis virus (Bournnell *et al.*, 1987); MHV1 and MHV2, two copies of a papain-like proteinase in the long ORF 1 of mouse hepatitis virus (Lee *et al.*, 1991); RUB, rubella virus (Dominguez *et al.*, 1990); SIN, Sindbis virus (Strauss *et al.*, 1984); SF, Semliki Forest virus (Takkinen, 1986); ONN, O'Nyong-nyong virus (Levinson *et al.*, 1990); RR, Ross River virus (Faragher *et al.*, 1988); TEV, tobacco etch virus, PPV, plum pox virus, TVMV, tobacco vein mosaic virus, and PVY, potato virus Y (Oh and Carrington, 1989); BaYMV, barley yellow mosaic virus (Kashiwazaki *et al.*, 1991); HAV, hypovirulence-associated virus of chestnut blight fungus (Choi *et al.*, 1991a). Numbers preceding the first amino acid are either the number of residues to the known N-terminus of the protein (parentheses), or the number of residues to the beginning of the ORF encoding the proteinase (square brackets). Residues which are the same as those of papain are indicated with a dot. Within the alphaviruses and the potyviruses, residues which are conserved among four of five and three of four viruses, respectively, are boxed.

suggested by the fact that change of this residue to alanine resulted in loss of enzymatic activity.

Other residues that have been proposed to be important for catalysis in papain-like enzymes include Gly-23, the two residues N-terminal to the catalytic Cys-25, and Gln-19, all of which are invariant in the plant and animal enzymes. It was suggested that these residues were involved in substrate binding and that a longer side chain at position 23 might block substrate binding by approaching too closely the side chain of Gln-19 (Baker and Drenth, 1987). In alphaviruses the residue corresponding to position 23 of papain is asparagine, whereas that at position 19 is Ser in Sindbis virus and Gln in other alphaviruses (Fig. 5). No firm conclusions can be drawn in the absence of crystal

structures, because the substrate specificity of alphavirus nsP2 is very different from that of papain or cathepsin. Papain has a strong preference for bulky hydrophobic amino acids, phenylalanine, tyrosine, valine, or leucine in the P2 position (the penultimate amino acid relative to the site of cleavage), while the animal cathepsins require nonpolar side chains at this position. The P2 position of alphavirus cleavage sites is always glycine. When this glycine is changed to valine or to glutamic acid, the site becomes noncleavable (Shirako and Strauss, 1990). Comparisons of this nature assume that the overall folding of the viral proteinases is similar to that of papain, but confirmation of this will require three-dimensional crystal structures of the viral counterparts.

Finally, there is the question of the viability of the various mutants. It is not surprising that mutations which kill the proteinase are lethal, such as the substitutions at Cys-481, His-558 and Trp-559, or that some substitutions which gave severely reduced processing were lethal or gave minute plaques. It was more surprising to discover that at position 614 the change from Asn to Asp-614 gave enhanced cleavage and apparently normal cleavage products but was lethal for viral replication, while the change to Ser-614 gave normal processing and normal virus production. Note that in Fig. 4 that the most obvious difference in the translation/processing pattern of Asp-614 was the lack of any residual P123. With the wild type, the amount of residual P123 varied with the batch of reticulocyte lysate, and there was a significant amount in the experiment shown in Fig. 4. The lethality of this particular substitution may indicate a separate and necessary role for P123 (or for some other processing intermediate) in alphavirus replication. Similarly, it is of interest to compare the results for the two substitutions at Asn-693. Both gave variable and somewhat reduced proteolysis, but in one case (Ser-693) the change was also lethal. This may mean that slight changes in the rate of proteolysis, which may not be detectable in the cell-free translation system used, can still interfere with efficient virus replication. It is also possible that some of these changes affect the overall conformation of nsP2 and polyproteins containing nsP2. Such an altered conformation could affect the helicase activity or impair some as yet unknown function of nsP2 other than proteolysis. The fact that many changes are poorly tolerated illustrates the fact that the alphaviruses, as isolated in nature, appear to be optimized for replication.

ACKNOWLEDGMENTS

This work was supported by Grants AI 10793 and AI 20612 from the National Institutes of Health. R.J. de G. was supported by a fellowship from European Molecular Biology Organization (ALTF 280-1988).

REFERENCES

- BAKER, E. N., and DRENTH, J. (1987). The thiol proteases: Structure and mechanism. In "Biological Macromolecules and Assemblies: The Active Sites of Enzymes" (F. A. Jornak and A. McPherson, Eds.), Vol. 3, pp. 314-367. Wiley, New York.
- BAKER, S. C., SHIEH, C.-K., SOE, L. H., CHANG, M.-F., VANNIER, D. M., and LAI, M. M. C. (1991). Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. *J. Virol.* **63**, 3693-3699.
- BAZAN, J. F., and FLETTERICK, R. J. (1989). Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology* **171**, 637-639.
- BAZAN, J. F., and FLETTERICK, R. J. (1990). Structural and catalytic models of trypsin-like viral proteases. *Sem. Virol.* **1**, 311-322.
- BOEGE, U., WENGLER, G., WENGLER, G., and WITTMAN-LIEBOLD, B. (1981). Primary structure of the core proteins of the alphaviruses Semliki Forest virus and Sindbis virus. *Virology* **113**, 293-303.
- BOURNELL, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. F., TOMLEY, F. M., and BINNS, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* **68**, 57-77.
- BROCKLEHURST, K. (1987). Acyl group transfer-cysteine proteinases. In "Enzyme Mechanisms" (M. I. Page and A. Williams, Eds.), pp. 140-158. The Royal Society of Chemistry, London.
- CARNE, A., and MOORE, C. H. (1978). The amino acid sequence of the tryptic peptides of Actinidin, a proteolytic enzyme from the fruit of *Actinidia chinensis*. *Biochem. J.* **173**, 73-83.
- CHAMBERS, T. J., WEIR, R. C., GRAKOUI, A., MCCOURT, D. W., BAZAN, J. F., FLETTERICK, R. J., and RICE, C. M. (1990). Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. *Proc. Natl. Acad. Sci. USA* **87**, 8898-8902.
- CHOI, G. H., PAWLKY, D. M., and NUSS, D. L. (1991a). The autocatalytic protease p29 encoded by a hypovirulence-associated virus of the chestnut blight fungus resembles the potyvirus-encoded protease HC-Pro. *Virology* **183**, 747-752.
- CHOI, H.-K., TONG, L., MINOR, W., DUMAS, P., BOEGE, U., ROSSMANN, M. G., and WENGLER, G. (1991b). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* **354**, 37-43.
- DE GROOT, R. J., HARDY, W. R., SHIRAKO, Y., and STRAUSS, J. H. (1990). Cleavage-site preferences of Sindbis virus polyproteins containing the nonstructural proteinase: Evidence for temporal regulation of polyprotein processing in vivo. *EMBO J.* **9**, 2631-2638.
- DE GROOT, R. J., RÜMENAPF, T., KUHN, R. J., STRAUSS, E. G., and STRAUSS, J. H. (1991). Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **88**, 8967-8971.
- DOMINGUEZ, G., WANG, C.-Y., and FREY, T. K. (1990). Sequence of the genome RNA of rubella virus: Evidence for genetic rearrangement during togavirus evolution. *Virology* **177**, 225-238.
- FARAGHER, S. G., MEEK, A. D. J., RICE, C. M., and DALGARNO, L. (1988). Genome sequences of a mouse-avirulent and a mouse-irulent strain of Ross River virus. *Virology* **163**, 509-526.
- GARAVITO, R. M., ROSSMANN, M. G., ARGOS, P., and EVENTOFF, W. (1977). Convergence of active site geometries. *Biochemistry* **16**, 5065-5071.
- GOLDBACH, R. (1990). Plant viral proteinases. *Sem. Virol.* **1**, 335-346.
- GORBALENYA, A. E., DONCHENKO, A. P., BLINOV, V. M., and KOONIN, E. V. (1989). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases: A distinct protein superfamily with a common structural fold. *FEBS Lett.* **243**, 103-114.
- GORBALENYA, A. E., KOONIN, E. V., and LAI, M. M. C. (1991). Putative papain-related thiol proteases of positive-strand RNA viruses. *FEBS Lett.* **288**, 201-205.
- HAHN, C. S., STRAUSS, E. G., and STRAUSS, J. H. (1985). Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid autoprotease. *Proc. Natl. Acad. Sci. USA* **82**, 4648-4652.
- HAHN, C. S., and STRAUSS, J. H. (1990). Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. *J. Virol.* **64**, 3069-3073.
- HAHN, Y. S., STRAUSS, E. G., and STRAUSS, J. H. (1989). Mapping of RNA temperature-sensitive mutants of Sindbis virus: Assignment of complementation groups A, B, and G to nonstructural proteins. *J. Virol.* **63**, 3142-3150.
- HARDY, W. R., HAHN, Y., S. DE GROOT, R. J., STRAUSS, E. G., and STRAUSS, J. H. (1990). Synthesis and processing of the nonstruc-

- tural polyproteins of several temperature-sensitive mutants of Sindbis virus. *Virology* **177**, 199–208.
- HARDY, W. R., and STRAUSS, J. H. (1989). Processing the nonstructural proteins of Sindbis virus: Nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and *trans*. *J. Virol.* **63**, 4653–4664.
- HARRIS, K. S., HELLEN, C. U. T., and WIMMER, E. (1990). Proteolytic processing in the replication of picornaviruses. *Sem. Virol.* **1**, 323–333.
- HIGAKI, J. N., GIBSON, B. W., and CRAIK, C. S. (1987). Evolution of catalysis in the serine proteases. *CSHSQB* **52**, 615–621.
- KAMPHUIS, I. G., DRENTH, J., and BAKER, E. N. (1985). Thiol proteases: Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *J. Mol. Biol.* **182**, 317–329.
- KAMPHUIS, I. G., KALK, K. H., SWARTE, M. B. A., and DRENTH, J. (1984). Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.* **179**, 233–256.
- KASHIWAZAKI, S., MINOBE, Y., and HIBINO, H. (1991). Nucleotide sequence of barley yellow mosaic virus RNA 2. *J. Gen. Virol.* **72**, 995–999.
- KUNKEL, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488–492.
- LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680–685.
- LEE, H.-J., SHIEH, C.-K., GORBALENYA, A. E., KOONIN, E. V., LA MONICA, N., TULER, J., BAGDZHADZHAN, A., and LAI, M. M. C. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180**, 567–582.
- LEVINSON, R., STRAUSS, J. H., and STRAUSS, E. G. (1990). Determination of the complete nucleotide sequence of the genomic RNA of O'Nyong-nyong virus and its use in the construction of phylogenetic trees. *Virology* **175**, 110–123.
- LOEB, D. D., HUTCHINSON, C. A. I., EDGELL, N. H., FARMERIE, W. G., and SWANSTROM, R. (1989). Mutational analysis of human immunodeficiency virus type I protease suggests functional homology with aspartic proteinases. *J. Virol.* **63**, 111–121.
- MITCHEL, R. E. J., CHAIKEN, I. M., and SMITH, E. L. (1970). The complete amino acid sequence of papain. *J. Biol. Chem.* **245**, 3485–3492.
- NAVIA, M. A., FITZGERALD, P. M. D., MCKEEVER, B. M., LEU, C.-T., HEIMBACH, J. C., HERBER, W. K., SIGAL, I. S., DARKE, P. L., and SPRINGER, J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**, 615–620.
- OH, C.-S., and CARRINGTON, J. C. (1989). Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis. *Virology* **173**, 692–699.
- OHNO, S., EMORI, Y., IMAJOH, S., KAWASAKI, H., KISARAGI, M., and SUZUKI, K. (1984). Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein. *Nature* **312**, 566–570.
- PEARS, C. J., MAHBUBANI, H. M., and WILLIAMS, J. G. (1985). Characterization of two highly diverged but developmentally co-regulated cysteine proteinase genes in *Dictyostelium discoideum*. *Nucleic Acids Res.* **13**, 8853–8866.
- PREUGSCHAT, F., LENCHES, E. M., and STRAUSS, J. H. (1991). Flavivirus enzyme-substrate interactions studied with chimeric proteins: Identification of an intragenic locus important for substrate recognition. *J. Virol.* **65**, 4749–4758.
- RICE, C. M., LEVIS, R., STRAUSS, J. H., and HUANG, H. V. (1987). Production of infectious RNA transcripts from Sindbis virus cDNA clones: Mapping of lethal mutations, rescue of a temperature sensitive marker, and in vitro mutagenesis to generate defined mutants. *J. Virol.* **61**, 3809–3819.
- RITONJA, A., POPOVIC, T., TURK, V., WIEDENMANN, K., and MACHLEIDT, W. (1985). Amino acid sequence of human liver cathepsin B. *FEBS Lett.* **181**, 169–172.
- RITONJA, A., ROWAN, A. D., BUTTLE, D. J., RAWLINGS, N. D., TURK, V., and BARRETT, A. J. (1989). Stem bromelain: Amino acid sequence and implications for weak binding of cystatin. *FEBS Lett.* **247**, 419–424.
- SAMBROOK, J., FRITSCH, E. F., and MANIATIS, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- SHIRAKO, Y., NIKLASSON, B., DALRYMPLE, J. M., STRAUSS, E. G., and STRAUSS, J. H. (1991). Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. *Virology* **182**, 753–764.
- SHIRAKO, Y., and STRAUSS, J. H. (1990). Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123. *Virology* **177**, 54–64.
- STRAUSS, E. G., RICE, C. M., and STRAUSS, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133**, 92–110.
- STRAUSS, E. G., and STRAUSS, J. H. (1986). Structure and replication of the alphavirus genome. In "The Togaviridae and Flaviviridae" (S. Schlesinger and M. J. Schlesinger, Eds.), pp. 35–90. Plenum, New York.
- STRAUSS, J. H. (Ed.) (1990). Viral proteinases. *Sem. Virol.* **1**, 307–384.
- STRAUSS, J. H., and STRAUSS, E. G. (1990). Alphavirus proteinases. *Sem. Virol.* **1**, 347–356.
- TAKKINEN, K. (1986). Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus. *Nucleic Acids Res.* **14**, 5667–5682.
- WIEDERANDERS, B., BROEMMA, D., KIRSCHKE, H., KALKINEN, N., RINNE, A., PAQUETTE, T., and TOOTHMAN, P. (1991). Primary structure of bovine cathepsin S. *FEBS Lett.* **286**, 189–192.
- WISKERCHEN, M., BELZER, S. K., and COLLETT, M. S. (1991). Pestivirus gene expression: The first protein product of the bovine viral diarrhoea virus large open reading frame, P20, possesses proteolytic activity. *J. Virol.* **65**, 4508–4514.
- ZOLLER, M. J., and SMITH, M. (1984). Laboratory Methods: Oligonucleotide-directed mutagenesis: A simple method using two oligonucleotide primers and a single-stranded DNA template. *DNA* **3**, 477–488.