The mosquito-borne dengue viruses are widespread human pathogens causing dengue fever, dengue hemorrhagic fever, and dengue shock syndrome, placing 40% of the world’s population at risk with no effective treatment. The viral genome is a positive strand RNA that encodes a single polyprotein precursor. Processing of the polyprotein precursor into mature proteins is carried out by the host signal peptidase and by NS3 serine protease, which requires NS2B as a cofactor. We report here the crystal structure of the NS3 serine protease domain at 2.1 Å resolution. This structure of the protease combined with modeling of peptide substrates into the active site suggests identities of residues involved in substrate recognition as well as providing a structural basis for the mutational effects on enzyme activity. This structure may be useful for development of specific inhibitors as therapeutics against dengue and other flaviviral proteases.

Dengue viruses, members of the family Flaviviridae, are transmitted by mosquitoes, *Aedes aegypti* and *Aedes albopictus*, and cause severe and widespread epidemics of diseases such as dengue fever and dengue hemorrhagic fever/dengue shock syndrome (for a review, see Ref. 3). The virus is transmitted to humans directly, with approximately 2.5 billion people at risk, mostly in the southern hemisphere. Nearly 5% of the one million dengue hemorrhagic fever cases each year are fatal (2). Currently, dengue infections are endemic in all continents except Europe (for recent reviews, see Refs. 3 and 4). Management of dengue fever is largely supportive; however, severe hemorrhagic manifestation may require blood transfusions (1). No vaccine is available to protect against dengue virus infections, and thus there is considerable interest in developing new antiviral therapeutic agents to combat diseases caused by dengue viruses.

Dengue virus type 2 (Den2), the most prevalent of the four serotypes, contains a single-stranded RNA of positive polarity with a type 1 cap structure at the 5'-end and codes for a single polyprotein precursor (3,391 amino acid residues for Den2) (5) arranged in the order NH2-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-COOH. Maturation of the polyprotein precursor occurs both cotranslationally and post-translationally within the endoplasmic reticulum, yielding three structural proteins, C (core), prM (precursor to membrane), and E (envelope), which are components of the virion, and at least seven nonstructural (NS) proteins (NS1–NS5). The nonstructural proteins are thought to function in viral replication, although the components of the viral RNA replicase have not been defined precisely.

Processing at the C-prM, prM-E, E-NS1, and NS4A-NS4B junctions (6–8). Processing at the NS1-NS2A junction is mediated by a host protease within the endoplasmic reticulum which may be either signal peptidase or another unknown enzyme (9). Maturation of prM to M is mediated by another host protease located in an acidic compartment encountered during the exocytic pathway and occurs at a late step of virion morphogenesis (10). The remaining cleavages in the polyprotein precursor are mediated by the trypsin-like serine protease which recognizes the cleavage site, N-H2-terminal 180 amino acids in the case of NS2B (11). The protease domain was identified by two separate studies of Den2 NS3 with several viral and cellular serine proteases (11, 12). Besides the protease domain at the NH2 terminus, the COOH-terminal three-fourths of NS3 contains conserved motifs found in DExH family of several viral RNA-stimulated NTPases/RNA helicases. The RNA helicase activity has been implicated in unwinding of double-stranded RNA replicative intermediate. NS3 also interacts with the viral RNA-dependent RNA polymerase (NS5) (13), and these protein-protein interactions may facilitate the localization of the viral replicase complex to endoplasmic reticulum membranes where genome replication occurs (for reviews, see Refs. 14 and 15).

Analysis of polyprotein processing established that the NS3 protease, as a complex with the viral activator protein NS2B (16–18) catalyzes the cleavages at NS2A-NS2B, NS2B-NS3, NS3-NS4A, and NS4B-NS5 sites in the polyprotein which have Lys-Arg, Arg-Arg, Arg-Lys, and occasionally Gln-Arg at the P2 position, followed by a short chain amino acid Gly, Ala, or Ser at the P1 position (16, 17, 19–23). Thus, the biologically active viral protease is a heterodimeric complex of NS2B-NS3. The importance of this protease activity in viral viability is underscored by the finding that mutations that abolished protease activity when introduced in the context of an infectious cDNA clone eliminated virus recovery (16). NS2B has three hydrophobic regions flanking a conserved hydrophilic domain.
of about 40 amino acid residues. This hydrophilic region is necessary and sufficient for activation of the NS3 protease domain in vivo (18, 32) and in vitro. Although the hydrophobic regions of NS2B are dispensable for the protease activity, they are required for cotranslational membrane insertion of full-length NS2B and its efficient activation of the NS3 serine protease domain (24).

The covalently linked trypsin-like serine protease domain and the RNA-stimulated NTPase/RNA helicase domain are also found in hepatitis C virus (HCV), a new member of Flaviviridae (25, 26), although HCV does not depend on arthropod vectors for transmission (27). Moreover, unlike the arthropod-borne flaviviral NS3 protease domain, which is activated by the hydrophilic region of NS2B, the protease domain of HCV is activated by a 19-residue NH2-terminal hydrophobic region of NS4A protein (28–33).

Serine proteases are perhaps the best studied (34–40) of the four classes (serine, aspartic, metallo, and cysteine) of proteases (for a review, see Ref. 41 and the references therein). Pioneering crystallographic, biochemical, and molecular biological studies have extensively documented the existence of a common catalytic apparatus (Asp-His-Ser triad) participating in a conserved mechanism of catalysis among serine proteases (35, 37–39) (for a review, see Ref. 41). The basic mechanism consists of a charge relay system that transfers the unfavorable negative charge on the buried carboxyl via the histidine to the serine, resulting in transfer of the Ser O\(^{-}\) negative charge on the residue in a conserved mechanism of catalysis among serine proteases (40–43). Selectivity of proteases for particular substrates is a result of differences; for example, the structural zinc binding site and components of the catalytic triad have also been identified (41). The positions of the two samarium ions in the asymmetric unit were determined from the Harker section of a Bijvoet differences Patterson map and the relative geometry, thus reducing systematic errors in Bijvoet differences. Similar geometry, thus reducing systematic errors in Bijvoet differences, was observed in the asymmetric unit of a Bijvoet differences Patterson map at 2.7 Å resolution, which provides a structural basis for understanding the structure of the Den2 protease domain at 2.1 Å resolution, which provides a structural basis for understanding the structure of the Den2 protease domain of HCV in the asymmetric unit. A native data set to 2.1 Å was derived from 59,122 measurements made on three crystals (an average redundancy of 4.6) on a Siemens X1000 area detector system at room temperature and processed with XDS (46). Data measurement statistics are presented in Table I. The structure was determined using a single two-site sarramum derivative, obtained by soaking crystals in a solution containing 2.5 mM samarium acetylacetone at pH 7.8 over 2 days. The strong Bijvoet signal from samarium in conjunction with isomorphous differences (Table I) was used to obtain an excellent set of SIRAS phases. All derivative data (80% complete to 2.7 Å) used in phasing were measured from one crystal rotated around the crystallographic b direction to map hkl and b-kl reflections on to the detector simultaneously and in similar geometry, thus reducing systematic errors in Bijvoet differences. Systematic errors in both Bijvoet and isomorphous differences were reduced further by anisotropic local scaling (47). Positions of the two samarium ions in the asymmetric unit were determined from the Harker section of a Bijvoet differences Patterson map and the relative y translation between them derived from analysis of cross-vectors. The positions were confirmed from isomorphous differences Patterson maps. Correct chirality was established by inspection of electron density maps. SIRAS phasing and refinement were done using MLPHARE (48), part of the CCP4 package (49) with reflections to 2.7 Å, the limit of usable data from the derivative. Phasing statistics are given in Table II.

Significant improvement to the electron density map calculated with weighted SIRAS phases was obtained using DM (50), and a model of approximately 40% of the protein was built into this map using Bones and O (51). At this stage similarities to HCV NS3 protease became apparent, and the rest of the structure was built into SIGMAA (45) weighted 2Fo-Fc maps calculated using SIRAS phases combined with those from the partial model. The initial model, consisting of residues
Structure of the Dengue Virus Serine Protease

TABLE I
Data measurement, phasing, and refinement statistics

| Resolution (Å) | % > 2σ(I) | (I/σ(I)) | R-merge<sup>a</sup> (R-iso) | R-Bijvoet<sup>b</sup> (Centric) | R-Cullis<sup>b</sup> (R-Cullis) | PoM<sup>c</sup> (R-Cullis) | Phasing power<sup>d</sup> | R-cryst<sup>c</sup> (R-free) |
|----------------|-----------|----------|-----------------------------|-------------------------------|-----------------------------|-------------------------|----------------------|-------------------------|
| 4.3            | 97.7      | 20.4     | 0.031 (0.282)               | 0.097 (0.030)                 | 0.71 (0.68)                 | 2.86                    | 1.98 (0.210)         |
| 3.4            | 91.1      | 18.8     | 0.034 (0.372)               | 0.157 (0.043)                 | 0.65 (0.55)                 | 2.69                    | 1.81 (0.219)         |
| 2.9            | 89.9      | 15.7     | 0.027 (0.378)               | 0.208 (0.048)                 | 0.62 (0.68)                 | 2.30                    | 1.75 (0.216)         |
| 2.7            | 86.3      | 10.2     | 0.043 (0.434)               | 0.191 (0.040)                 | 0.56 (0.71)                 | 1.75                    | 1.94 (0.242)         |
| 2.5            | 83.8      | 8.1      | 0.048                       | 0.181 (0.040)                 | 1.78                        | 1.91 (0.240)         |
| 2.3            | 81.3      | 8.2      | 0.101                       | 0.202 (0.237)                 | 1.75                        | 1.94 (0.242)         |
| 2.2            | 80.9      | 6.3      | 0.157                       | 0.239 (0.258)                 | 1.75                        | 1.94 (0.242)         |
| 2.1            | 79.5      | 6.3      | 0.168                       | 0.259 (0.301)                 | 1.75                        | 1.94 (0.242)         |
| All            | 84.8      | 10.2     | 0.065 (0.353)               | 0.170 (0.045)                 | 0.64 (0.65)                 | 2.40                    | 1.96 (0.228)         |

<sup>a</sup> R-merge = Σ||F<sub>o</sub>|| - |I|2/Σ|F<sub>o</sub>|; values are for native data, they were comparable for Samarium data; R-iso = Σ|F<sub>o</sub>| - |I|2/Σ|F<sub>o</sub>|

<sup>b</sup> R-Bijvoet = Σ|F<sub>o</sub>| - |F<sub>call</sub>|/Σ|F<sub>o</sub>| + |F<sub>call</sub>|; values for h0l reflections in parentheses.

<sup>c</sup> R-Cullis = Σ|F<sub>o</sub>| - |F<sub>calc</sub>|/Σ|F<sub>o</sub>| and |F<sub>calc</sub>|, defined for centric reflections; PoM, mean figure of merit after density modification.

<sup>d</sup> Phasing power = r.m.s. isomorphous difference/r.m.s. residual lack of closure.

<sup>e</sup> R-cryst = Σ|F<sub>o</sub>| - |F<sub>calc</sub>|/Σ|F<sub>o</sub>| R-free, for 10% of reflections (1,283) chosen at random over 8–2.1Å.

Results and Discussion

Structure Description—A side-by-side stereo representation of the electron density map around the catalytic triad is shown in Fig. 1. Fig. 2a shows the overall folding along with the conserved orientation of the catalytic triad. Comparisons with the structures of HCV protease domain, either as a heterodimeric complex with the activating NS5A peptide (Protein Data Base no. 1jxp (56); see also Ref. 45) or in the absence of NS4A (Protein Data Base no. 1alq (44)) show similarities expected from sequence comparisons as well as significant differences. Although both the amino- and carboxyl-terminal β barrels are six-stranded in both Den2 and HCV proteases, the strands in the NH<sub>2</sub>-terminal domain of Den2 protease are shorter, and the barrel is significantly more deformed (Fig. 2a).

In the carboxyl-terminal domain, all six strands are more strongly conserved and are of comparable length. In general, the carboxyl-terminal domain carries most of the residues that contribute to substrate specificity and is highly conserved among serine proteases (34), and Den2 protease is not an exception. Superposition (57) of 68 α carbon atoms of the COOH-terminal domain of Den2 protease (residues 96–167) and either HCV protease structure (residues 96–186), omitting strands B2 and C2 in both, yields a r.m.s. deviation of 0.9 Å compared with 0.6 Å for the two structures of HCV protease. Alignment of pairs of structures using this transformation shows that the conformation of Den2 protease has a greater resemblance to that of the HCV protease-NS4A complex in this region. The first 30 residues in the two HCV protease structures have very different conformations (Fig. 2b). In the absence of NS4A these residues make β sheet interactions with symmetry-related molecules. Upon interaction with NS4A peptide, these 30 residues form two β strands (A0 and A1) and a helix α0. However, the structure of the Den2 protease in the absence of NS2B peptide more closely resembles that of the HCV protease-NS4A complex rather than the HCV protease domain alone (r.m.s. deviation 0.4 Å) (Fig. 2b). A second major difference between the two HCV structures occurs near helix Ha and strands D1, E1, and F1. Here again, the conformation of Den2 protease is much closer to that of the HCV protease-NS4A complex (Fig. 2, a and b).

There are significant differences in the mode of catalysis by Den2 and HCV proteases. For example, the substrate cleavage
specificities of HCV and Den2 proteases are different; HCV protease prefers a Cys residue at the P1 position of the substrates for cleavages at the NS4A-NS4B, NS4B-NS5A, and NS5A-NS5B sites but a Thr residue at this position for the intramolecular cleavage (NS3-NS4A site) (28–30, 58, 59) in contrast to a basic residue (Lys or Arg) for all of the cleavages by Den2 protease (14) (see Table II). The activating peptide, NS4A, enhances the efficiency of HCV protease activity in vitro to a different extent: 100-fold for the cleavage of NS4B-NS5A site but only 11- and 3-fold for the cleavages of NS4A-NS4B and NS5A-NS5B sites, respectively (60). However, the Den2 protease requires NS2B for cleavages of all protease-sensitive sites at least within the sensitivity of currently available assays. Moreover, the region of NS2B required for activation of Den2 protease is a hydrophilic domain of 40 residues in contrast to the HCV NS4A activator peptide, which consists of hydrophobic residues. Thus the mode of activation of Den2 and HCV protease domains by their respective activator peptides and their activated states are likely to be different.

Other significant (>2 Å) deviations of Den2 from the fold of both HCV protease structures occur in strand F1 (Fig. 2, a and b) where there is a two-residue insertion in HCV protease relative to Den2 protease, strand D2, the site of a three-residue insertion in HCV protease, and strand C2 where Den2 protease has a seven-residue insertion relative to HCV protease. Strand C2, a long 13-residue strand in HCV protease, is split into two shorter strands of three residues each interrupted by a four-residue coil segment in Den2 protease. Among the conserved
Den2 virus, New Guinea-C strain; data from Ref. 2; data for Japanese encephalitis (JE), Kunjin virus (KUN), Murray Valley encephalitis (MVE), yellow fever virus (YF), West Nile virus (WN), and tick-borne encephalitis (TBE) are from Ref. 3 and references therein. Residues P6 through P5' are shown for each cleavage site.

| Virus | NS2A-NS2B | NS2B-NS3 | NS3A-NS4A | NS4B-NS5 |
|-------|-----------|----------|-----------|----------|
| Den2  | RTNKR | SFPLNE | EVKQR | AGVLND | FAGKR | SITLNL | TTNTR | GTGNIQ |
| Den4  | KGASRR | SFPLNE | QVKQR | SGALND | FAGKR | SITLID | AQTPRR | GTGTTG |
| JE    | NPNKR | GPATE | LKTQR | GGVFND | FAGKR | DAISSF | KPSLR | GRPGGR |
| KUN   | DFNKR | GPATE | LQTQR | GGVLND | FAGKR | SQIGFI | KPGKLK | GOAKGR |
| MVE   | NPNKR | GPATE | LKTVR | GGVFND | FAGKR | SAI GF | KPAF KR | GRAGGR |
| YF    | RIFGR | STVNE | VRGAR | SGGDLN | FAGGR | GAAEVL | MTKGR | GSANGK |
| WN    | DFNKR | GPATE | LQTQR | GGVLND | FAGKR | SQIGLV | KPGKLK | GOAKGR |
| TBE   | AHRGR | SFSEPL | LRSSR | SDLVFS | YASGRG | SFGDVL | ASSGGR | GSEGGD |

Sequences around the in vivo cleavage sites of Den2 protease and seven other flaviviral NS3 proteases (14) are given in Table II. The substrate binding cleft of the Den2 protease is not very extensive (Figs. 2 and 3, a and b) and does not appear capable of providing specific interactions, in the absence of NS2B activating peptide, with side chains beyond P2 and P1. This observation is consistent with heterogeneity of residues beyond these sites seen (14) in flaviviral proteases (Table II). Most viral proteases are considerably smaller than their cellular counterparts; several loops that are present in cellular enzymes that provide specific interactions to P3, P4, and P5 side chains of substrates are absent in the viral proteases (45). It is possible that heterodimerization with NS2B peptide could generate additional specific interactions for side chains beyond P2 and P1, both by altering the protease conformation into an activated state and by interacting directly with the substrate as has been suggested for other two-component protease (45).

Accordingly, molecular modeling was limited to building a tetrapeptide into the substrate binding cleft. Complexes with substrates were modeled using well known principles of serine-protease-substrate interactions (64). Rigid-body transformations (57) and small manual adjustments were used throughout. Using the relevant tetrapeptide section of the main chain from a porcine trypsin-inhibitor complex (65) (Protein Data Base No. 1mct), the scissile carbonyl was positioned within hydrogen bonding distance of a nitrogen atoms of Gly-133 (2.71 Å) and Ser-135 (2.8 Å). Simultaneously the peptide nitrogen was placed at 3.1 Å from the carbonyl oxygen of Gly-151. Gly-133 and Ser-135 were identified as the most likely to form the oxyanion hole by comparison with structures of other serine proteases. Further small translations established hydrogen bonding interactions between the main chain of P1 and P2 residues with appropriate main chain atoms of Gly-153 and Asn-152 to generate the short section of β sheet common to serine protease-inhibitor interactions (64). Side chains were built in their most probable conformations using O to generate the desired sequence at each of the viral polypeptide junctions (Table II).

Table III presents a summary of results obtained. There are three cleavage sites (NS2A-NS2B, NS2B-NS3, and Ser-163) within the S1 pocket (Fig. 3b), in addition to the catalytic Ser-135, which are the most probable sites of providing specific stabilizing interactions. These interactions include hydrogen bonding interactions with an Arg residue at P1. In addition, Leu-115 provides nonspecific van der Waals interactions. These interactions can also accommodate a Lys side chain as found in the NS3-NS4A cleavage site; see Table II) and Ile-36 by well. All of the residues that form the S1 pocket except Ser-131 and Leu-115 are absolutely conserved (14) among all of the flaviviruses listed in Table III. Primary specific interaction of the Den2 protease with a lysyl (NS2A-NS2B site), glutaminyl (NS2B-NS3 site) and asparaginyl (NS3-NS4A and NS4B-NS5 sites) main chain atom is provided by Asn-152. The catalytic His residue, as well as Tyr-150 and Gly-153 make additional hydrogen bonding or van der Waals interactions. These residues are also strongly conserved among the listed flaviviruses (Table III). The Oε1 atom of Asn-152 forms a salt bridge/hydrogen bond with Ne of the P2 Arg in the modeled complex (Fig. 3b); alternate interactions are possible by rearrangement of the side chains of both the protease and the P2 residue of the substrate to accommodate a Lys or Gln (Table II). A serine side chain at P1’ll fits into the S1’ pocket formed by the catalytic His-51 and Ser-135 and residues Gly-35, Ile-36, and Val-52. Of these Ile-36 interacts only through its main chain atoms and is not conserved (Table III). A Ser at P2’ is also stabilized exclusively by interaction with main chain atoms of the residues listed (Table III). A larger side chain at P2’, such as a Trp (as in the NS2A-NS2B cleavage site; see Table II) or an Arg (occurring in an internal cleavage site within Den2 protease) makes additional van der Waals interactions with side chain atoms of residues defining the S2’ pocket as well as with Asp-129 and Phe-130. It is possible that the paucity of specific interactions at sites other than P2 and P2’ is caused by the protease being not complexed with NS2B. Heterodimerization with NS2B could provide additional interactions beyond the P2 and P2’ positions of the substrate side chains directly as well as indirectly through altered positions of residues in the protease.

**Correlation with Mutational Studies—**Sequence comparisons (12, 66) identified four sets of residues (labeled as conserved regions 1–4) which were conserved between flaviviral and cellular proteases. Conserved region 1 includes strands C1 and D1 (Fig. 2a) along with catalytic His-51; conserved region 2 comprises of strand F1 and catalytic Asp-75; conserved region
3 consists of strand D2 and includes the catalytic Ser-135 together with several residues expected to contribute to substrate binding; and conserved region 4 includes strand E2 with additional residues involved in substrate recognition. Residues Asp-129, Phe-130, Tyr-150, Asn-152, and Gly-153 were judged to be important for determination of substrate specificity in Den2 protease as these are analogs of residues 189, 190, 213, 215, and 216 in trypsin (chymotrypsin numbering) which determine interactions with side chains of the substrate in that enzyme. Asp-129, in particular, was identified as the residue that lies at the bottom of the S1 site and provides charge neutralization for the substrate Arg or Lys side chain (12, 66).

A wide ranging site-specific mutagenesis study reported recently (42) measured change in catalytic activity of the enzyme resulting from each of a number of mutations in the protease domain. Residues, mutations of which resulted in significant alteration of enzyme activity (42), are plotted in Fig. 3. In conserved region 3, Asp-129 and Phe-130 are residues that were expected to form part of the S1 pocket, with Asp-129 providing charge neutralization of a Lys or Arg residue at P1, in analogy with trypsin. Residues near the guanidine moiety of P1 arginine are shown as sticks, colored cyan, and labeled. Asn-152 is shown in yellow and can be seen interacting with Ne of the P2 arginine. Note that protease residues are shown in their conformations in the native structure to illustrate the multiple possibilities for interaction, but they are capable of interaction with the P1 residue by side chain rotations. The figure was made with GRASP (71).
In each pocket residues are listed in the order; abbreviations are as in Table II. Den2 residues (Ref. 2) are numbered. Identities of residues in Den2 are derived from substrate modeling (see text). Corresponding residues for other viruses are from sequence alignments from Ref. 3.

### Table III

| Specificity pocket and virus | Residues |
|-----------------------------|----------|
| S2                          | DEN2     |
|                             | H51      |
|                             | D75      |
|                             | G151     |
|                             | N152     |
|                             | G153     |
|                             | DEN4     |
|                             | H        |
|                             | D        |
|                             | G        |
|                             | N        |
|                             | G        |
|                             | KUN      |
|                             | H        |
|                             | D        |
|                             | G        |
|                             | N        |
|                             | YF       |
|                             | H        |
|                             | G        |
|                             | N        |
|                             | WN       |
|                             | H        |
|                             | D        |
|                             | G        |
|                             | N        |
|                             | TBE      |
|                             | H        |
|                             | D        |
|                             | G        |
|                             | N        |
| S1                          | DEN2     |
|                             | L115     |
|                             | S131     |
|                             | S135     |
|                             | G136     |
|                             | Y150     |
|                             | S163     |
|                             | DEN4     |
|                             | L        |
|                             | K        |
|                             | S        |
|                             | G        |
|                             | Y        |
|                             | S        |
|                             | JE       |
|                             | V        |
|                             | P        |
|                             | S        |
|                             | G        |
|                             | Y        |
|                             | S        |
|                             | KUN      |
|                             | V        |
|                             | P        |
|                             | S        |
|                             | G        |
|                             | Y        |
|                             | S        |
|                             | YF       |
|                             | L        |
|                             | P        |
|                             | S        |
|                             | G        |
|                             | Y        |
|                             | S        |
|                             | WN       |
|                             | V        |
|                             | P        |
|                             | S        |
|                             | G        |
|                             | Y        |
|                             | S        |
|                             | TBE      |
|                             | E        |
|                             | A        |
|                             | S        |
|                             | G        |
|                             | Y        |
| S1'                         | DEN2     |
|                             | Q35      |
|                             | I36      |
|                             | H51      |
|                             | V52      |
|                             | S135     |
|                             | DEN4     |
|                             | Q        |
|                             | V        |
|                             | H        |
|                             | V        |
|                             | S        |
|                             | JE       |
|                             | Q        |
|                             | A        |
|                             | H        |
|                             | T        |
|                             | S        |
|                             | KUN      |
|                             | Q        |
|                             | A        |
|                             | H        |
|                             | T        |
|                             | S        |
|                             | YF       |
|                             | Q        |
|                             | A        |
|                             | T        |
|                             | G        |
|                             | S        |
|                             | WN       |
|                             | Q        |
|                             | A        |
|                             | R        |
|                             | S        |
|                             | T        |
|                             | S        |
|                             | TBE      |
|                             | Q        |
|                             | V        |
|                             | K        |
|                             | G        |
| Hydrophobic cluster         | DEN2     |
|                             | L115     |
|                             | F116     |
|                             | V126     |
|                             | A160     |
|                             | Y161     |
|                             | DEN4     |
|                             | L        |
|                             | F        |
|                             | V        |
|                             | A        |
|                             | Y        |
|                             | JE       |
|                             | V        |
|                             | T        |
|                             | A        |
|                             | Y        |
|                             | KUN      |
|                             | V        |
|                             | E        |
|                             | A        |
|                             | R        |
|                             | Y        |
|                             | DN       |
|                             | V        |
|                             | F        |
|                             | T        |
|                             | S        |
|                             | TBE      |
|                             | E        |
|                             | L        |
|                             | V        |
|                             | V        |

3, b and c) from the S1 pocket and cannot interact with a P1 Arg or Lys side chain without substantial main chain motion. Phel-130, likewise is too far from the P1 side chain, but both Asp-129 and Phe-130 could provide some van der Waals contacts for a large side chain at P2’, such as a Trp found at the NS2A-NS2B cleavage site. Ser-135 is an essential residue and cannot tolerate replacement by Ala; however, a replacement of Ser-135 by a cysteine also renders the Den2 protease nearly inactive. Cysteine proteases, in which a Cys plays the role of the active site serine in serine proteases, have structural requirements that are similar to those of serine proteases. Mechanistic details of substrate hydrolysis for the two classes of proteases are also very similar (67). Thus it would be of interest to determine the underlying cause of the severe loss of activity in the S135C mutant Den2 protease. Gly-133 is in the loop that carries the catalytic Ser-135 (Fig. 3d) and is part of a β turn that positions strand D2 to interact with strand A2 to form the core β barrel (Fig. 2a). Replacement of this residue might destabilize the barrel as well as alter the position of Ser-135 significantly. Other residues in conserved region 3, residues 139–144 (Fig. 4 of Ref. 42), do not seem to play an essential role in the modeled substrate-enzyme complex. This observation is consistent with the mutational effects of these residues (42).

**Mutations in Conserved Region 4—**Tyr-150 and Ser-131 (conserved region 3), which were mutated to a number of other residues (42), are close enough to the basic side chains of either Arg or Lys at P1 to form a salt bridge or hydrogen bond with small side chain rotations. In contrast to conclusions from sequence comparisons, it is Tyr-150 rather than Asp-129 which appears to provide primary charge stabilization within the S1 pocket. In addition, the side chain of Ser-163, which is also highly conserved, could stabilize a P1 Arg or Lys in a different conformation after side chain rotation (Fig. 3c). Tyr-150, in addition, could provide stabilization for a positively charged side chain through its aromatic electron cloud, a mode of interaction which has been suggested earlier (68) and proposed for the other serine proteases (44, 45). Thus replacement of Tyr-150 by Phe will not reduce activity significantly because although it would abolish the salt bridge/hydrogen bond to the Tyr-150-κOH, it would not abrogate aromatic stabilization and the potential for interaction with Ser-131 or Ser-163. However, replacement of Tyr-150 with Ala, Val, or His would be expected to decrease activity because fewer modes of stabilization of the P1 side chain would be available, which is consistent with observation (42). It would be interesting to see the effects of a double or triple mutant of residues Tyr-150, Ser-131, and Ser-163. Gly-148 is part of the strand E2 packing against the strand F2 in the COOH-terminal β barrel (Fig. 2a), and a mutation is likely to cause destabilization of the barrel, which was indeed observed (42).

The S2 pocket of Den2 protease has Asn-152 as one of the conserved residues (see Table II), which provide hydrogen bonding interactions for its Asn. Asn-152 residues at P2. A number of hydrophobic core residues in the barrel, which are electrostatically favorable environment for a basic P2 side chain. If Asn-152 does indeed provide primary interaction with the P2 side chain, its replacement with Ala, as was done (42), would negate this interaction. Its replacement with a Lys would also be expected to decrease activity because a basic side chain at P2 would be repelled. A Lys would still be capable of providing an interaction for a P2 Gln, and perhaps this mutant enzyme will have activity at the NS2B-NS3 (Table II) junction but not at other sites. It is remarkable to see the observed decrease in activity at the NS2B-NS3 junction, because this would appear to come at a critical position for a positively charged substrate. The structure of the heterodimeric NS2B-NS3 protease, the species that is probably present under the assay conditions of that study (42), might suggest an appropriate rationalization.

**Mutations of Other Residues**—It is clear why Den2 protease loses activity if Gly-153 is replaced (42). Gly-153 is part of the strand E2 which forms one half of the sheet E2-F2 (Fig. 2a). It is packed against Ser-163 in the opposite strand and is almost completely inaccessible to solvent. Substitution of the α hydrogen on Gly-153 by any other residue will undoubtedly destabilize this sheet and cause main chain changes in surrounding residues. Modeling studies indicate that Gly-153 is likely to be one of the residues that form the oxyanion hole, donating its NH to the carbonyl atom of the P1 residue to stabilize and orient the substrate in the Michaelis complex. Any changes in its position would, therefore, severely curtail enzyme activity. In addition, Gly-153 is in the vicinity of Tyr-150 and Asn-152, both of which are part of the strand E2. Changes in this part of the main chain will likely alter the positions of Tyr-150 and Asn-152, further reducing the ability to stabilize the basic P1 side chain. Effects of mutations at Val-126 appear to be related to the packing environment of this residue. Val-126 participates in a hydrophobic cluster that probably defines the conformation of this loop and which positions Ser-135 appropriately for catalysis. The main chain beyond Val-126 descends into the active site cleft through a series of stacked β turns (Fig. 3d). None of these turns is internally hydrogen bonded, thus deriving their stability from van der Waals interactions in the vicinity and a hydrogen bond between Asp-129 and Ser-127 side chains (Fig. 2a).
3d). The side chains of Val-126 along with those of Leu-115 from the loop between the strands B2 and C2 (Fig. 2a) and Ala-160 and Tyr-161 from the strand F2 participate in this cluster (Fig. 3d and Table III). Replacement of Val-126 with smaller side chains will create a packing fault, very likely altering the conformation of the loop, making it more mobile and therefore making it more difficult to position Ser-135 optimally for catalysis. Explanation of severe loss of activity upon deletion of residues amino-terminal to residue 167 in the NS3 domain also follows from the observation that residues Ala-160 and Ser-163 probably contribute critical interactions to stabilize the enzyme-substrate complex.

It is clear from the structure and substrate modeling experiments that Den2 protease in the absence of interaction with NS2B peptide is not likely to provide specific interactions to substrate side chains beyond P2 on the carboxyl side of the scissile bond and P2 on the amino-terminal side. Such a limited substrate binding cleft is likely to make the design of specific inhibitors somewhat formidable. However, the structure of the protease domain provides an excellent starting point for structural studies of binary complexes with inhibitors and ternary complexes with inhibitors and activating NS2B polypeptide. These analyses will provide additional information on both substrate interactions as well as molecular mechanism of activation. If NS2B extends the binding site of the protease to provide additional specific interactions with NS2B, this information can be used to design future-resistant, NS2B-specific inhibitors (69). Structural and biochemical studies of the protease complex with NS2B could be an alternate target aimed at preventing full activation of the enzyme, a strategy that has been suggested for the HCV protease (45). Based on molecular calculations presented in this study, interactions that have not been suggested to be important for binding of the substrate has also been conserved among other serine protease domains of virus-borne flaviviruses (some shown in Table III). This observation suggests that conclusions drawn about substrate binding with the Den2 protease are applicable to other serine proteases. Results from these studies will be of interest to protein structure and design of antiviral therapeutics. This structure-based approach offers considerable promise toward treatment of a wide range of life-threatening diseases caused by flaviviruses.

Acknowledgments—We thank Drs. Abdel-Meguid, Babu, Falgout, and Valle for a critical reading of the manuscript. We also thank Dr. Murali for the use of computing facilities at the University of Pennsylvania.

REFERENCES

1. Kauerl, I., Robinson, M. J., and Kuhnle, U. (1997) J. Pediatr. 131, 516–524
2. Guha, D. J., and Clark, G. G. (1995) Emerg. Infect. Dis. 1, 55–57
3. Panheiro, F. P., and Corber, S. J. (1997) World Health Stat. Q. 50, 161–169
4. Rico Hesse, R., Harrison, L. M., Salas, R. A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M. T., Nogueira, R. M., and da Rosa, A. T. (1997) Emerg. Infect. Dis. 3, 399–406
5. Preugschat, F., Yao, C. W., and Strauss, J. H. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 7232–7243
6. Chambers, T. J., Weir, R. C., Grakoui, A., McCourt, D. W., Bazan, J. F., Kautner, I., Robinson, M. J., and Kuhnle, U. (1997) J. Virol. 71, 5045–5055
7. Chambers, T. J., Grakoui, A., and Rice, C. M. (1993) J. Virol. 67, 4364–4374
8. Svitkin, Y. V., Lyapustin, V. N., Lashkevich, V. A., and Agol, V. I. (1984) Adv. Virus Res. 36, 197–211
9. Chambers, T. J., Nestorowicz, A., Amberg, S. M., and Rice, C. M. (1993) J. Virol. 67, 6797–6807
10. Virology 8, Nowak, T., Farber, P. M., Wengler, G., and Wengler, G. (1989) Virology 171, 637–639
11. Kauferl, I., Robinson, M. J., and Kuhnle, U. (1997) J. Virol. 71, 4364–4374
12. Chambers, T. J., Weir, R. C., Grakoui, A., McCourt, D. W., Bazan, J. F., Kautner, I., Robinson, M. J., and Kuhnle, U. (1997) J. Virol. 71, 5045–5055
13. Chambers, T. J., Grakoui, A., and Rice, C. M. (1993) J. Virol. 67, 6042–6050
14. Chambers, T. J., Nestorowicz, A., Amberg, S. M., and Rice, C. M. (1993) J. Virol. 67, 573–580
15. Chambers, T. J., Grakoui, A., and Rice, C. M. (1993) J. Virol. 67, 4364–4374
16. Chambers, T. J., Weir, R. C., Grakoui, A., McCourt, D. W., Bazan, J. F., Kautner, I., Robinson, M. J., and Kuhnle, U. (1997) J. Virol. 71, 5045–5055
17. Chambers, T. J., Weir, R. C., Grakoui, A., McCourt, D. W., Bazan, J. F., Fletterick, R. J., and Rice, C. M. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 8888–8902