Functional Profiling of Recombinant NS3 Proteases from All Four Serotypes of Dengue Virus Using Tetrapeptide and Octapeptide Substrate Libraries

Received for publication, January 18, 2005, and in revised form, May 26, 2005
Published, JBC Papers in Press, June 1, 2005, DOI 10.1074/jbc.M500588200

From the ‡Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, the ¶Novartis Institute for Tropical Diseases, 10 Biopolis Road, 05-01 Chromos, Singapore 138670, and the ¶Novartis Institute for Biomedical Research, CH-4002 Basel, Switzerland

Regulated proteolysis by the two-component NS2B/NS3 protease of dengue virus is essential for virus replication and the maturation of infectious virions. The functional similarity between the NS2B/NS3 proteases from the four genetically and antigenically distinct serotypes was addressed by characterizing the differences in their substrate specificity using tetrapeptide and octapeptide libraries in a positional scanning format, each containing 130,321 substrates. The proteases from different serotypes were shown to be functionally homologous based on the similarity of their substrate cleavage preferences. A strong preference for basic amino acid residues (Arg/Lys) at the P1 position was observed, whereas the preferences for the P2–4 sites were in the order of Arg > Thr > Gln/Asn/Lys for P2, Lys > Arg > Asn for P3, and Nle > Leu > Lys > Xaa for P4. The prime site substrate specificity was for small and polar amino acids in P1′ and P3′. In contrast, the P2′ and P4′ substrate positions showed minimal activity. The influence of the P2 and P3 amino acids on ground state binding and the P4 position for transition state stabilization was identified through single substrate kinetics with optimal and suboptimal substrate sequences. The specificities observed for dengue NS2B/NS3 have features in common with the physiological cleavage sites in the dengue polyprotein; however, all sites reveal previously unrecognized suboptimal sequences.

Dengue virus is the etiologic agent of dengue fever, dengue hemorrhagic fever, and dengue shock syndrome and is the most prevalent arthropod-transmitted infectious disease in humans. Dengue consists of four closely related but antigenically distinct viral serotypes (DEN1–4), of the genus Flavivirus (1, 2).

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

†† To whom correspondence may be addressed. E-mail: subhash.vasudevan@group.novartis.com.
‡‡ To whom correspondence may be addressed. E-mail: vasudevan@group.novartis.com.
§ Both contributed equally to this work.
** To whom correspondence may be addressed. E-mail: jharris@gnf.org.
¶¶ To whom correspondence may be addressed. E-mail: subhash.vasudevan@group.novartis.com.

The on-line version of this article (available at http://www.jbc.org) contains Supplemental Tables 1 ("Steady-state kinetic parameters of DEN2 CF40-Gly-NS3pro185 using fluorogenic hexapeptides with natural dengue cleavage sites") and II ("P4-P1 preference of dengue 1 NS3 protease").

Following primary infection, lifelong immunity develops that prevents repeated assault by the same serotype but does not provide protection from a virus of a different serotype (3). Dengue diseases are endemic in the tropics and subtropics, and the viruses are maintained in a cycle that involves humans and the Aedes aegypti mosquito. Infection with dengue viruses produces a spectrum of clinical illness ranging from a nonspecific viral syndrome to severe and fatal hemorrhagic disease (1, 2). Currently there is no antiviral drug or vaccine available against dengue viruses, and the pathogenesis of the disease is poorly understood.

As with other members of the Flaviviridae family, the genomes of the dengue viruses consist of a positive single-stranded RNA of 10–12 kb in length (4). Co-translational processing and post-translational processing of the polyprotein give rise to three structural proteins and at least seven nonstructural proteins (4). The correct processing of these proteins is essential for virus replication and requires host proteases such as signalase and furin (5) and a two-component viral protease, NS2B/NS3 (4). Previous studies have shown that the N-terminal part of NS3 contains trypsin-like protease domain (6) and that the activity of NS3 was dependent on at least 40 amino acids of NS2B (6–8).

The preferred NS3 protease-cleavage sites in the viral polyprotein have two basic amino acid residues (Arg-Arg, Arg-Lys, Lys-Arg, or occasionally Gln-Arg) at the P2 and P1 positions, followed by a Gly, an Ala, or a Ser at the P1′ position (4). The crystal structure of the DEN-2 NS3pro in the absence of NS2B has been determined at 2.1-Å resolution by Murthy et al. (9) and shows a shallow substrate binding site, indicating a lack of significant interactions beyond P2-P2′. The NS3pro domain in the absence of NS2B is an inefficient protease as demonstrated by the low turnover rate of the small chromogenic substrate N-α-benzoyl-L-Arg-p-nitroanilide (10). Although NS2B is required for efficient enzymatic activity of the NS3pro, the structure of the latter without the cofactor resembles that of the related hepatitis C NS3 protease bound to its activating peptide NS4A. The exact mechanism by which the NS2B cofactor stimulates the protease is not currently known. However, it is plausible that NS2B resembles NS4A and interacts directly with the NS3 protease domain, causing a conformational change that extends the binding pockets (10).
Substrate Specificity of Dengue NS2B-NS3 Proteases

The aim of the current study was to elucidate and compare the substrate specificity of NS3 protease from all four serotypes. We performed functional substrate profiling of the P1-P4 and P1–P4′ for the DEN1–4 protease complexes using tetrapeptide and octapeptide positional scanning peptide libraries. As a consequence, we expanded the earlier findings on DEN2 NS3 to a broader extent (P4–P4′) and discovered that its substrate preference was shared by enzymes of the other three serotypes.

PROCEDURAL EXPERIMENTS

MATERIALS—Dengue virus serotype 1 (strain Hawaii), and serotype 4 (H241) were purchased from American Type Culture Collection (Manassas, VA). Dengue virus serotype 3 (strain S221/03, GenBank accession number AY037116) was obtained from a dengue patient and was a kind gift from Dr. Eng Eong Ooi (Environmental Health Institute, Singapore). The plasmids pGEM-T-(E-NS3) and pET15b-NS3NS5 containing, respectively, the NS2B/NS3 and NS3 cDNAs from Dengue virus serotype 2 (strain TSV01, GenBank accession number AY037116) were kind gifts from Dr. James Cook University, Queensland, Australia. The dengue virus NS3 protease substrate peptide Boc-Gly-Arg-Arg-AMC was purchased from Bachem (Bubendorf, Switzerland). Restriction enzymes and modifying enzymes were purchased from New England Biolabs (Beverly, MA).

Reverse Transcription PCR of DEN1, DEN3, and DEN4 cDNA Fragments Encoding NS2B/NS3—C6/36 cells were inoculated with the DEN1, DEN3, or DEN4 virus and incubated at 28 °C for 5–7 days. Cell culture media were collected and spun at 14,000 rpm to remove cell debris. Viral RNAs were obtained by extracting 2 ml of the clarified media using Trizol Reagent (Invitrogen) according to the manufacturer’s instructions. First strand cDNA synthesis for the NS2B/NS3 sequences was performed using the primers DEN1 reverse (5′-TGGTTGGAAGAATCCTATGAG-3′), DEN3 reverse (5′-TGGTTGGAAGAATCCTATGAG-3′), and DEN4 reverse (5′-GTAAGTGCCAAAAGTCGGAATC-3′) with SuperScript II (Invitrogen) at 45 °C for 1 h, followed by PCR with Phi polymerase (Stratagene) and the primers DEN1 forward (5′-GGGAGGCTGATCTGCAATG-3′), DEN3 forward (5′-GAGGCTGATCTGCAATG-3′), and DEN4 forward (5′-CCATATGCTGCTGTTT-3′) along with the corresponding reverse primers.

Preparation of CF40-Gly-NS3pro185 Expression Constructs—All DEN1–4 CF40-Gly-NS3pro185 expression constructs comprised the 40-amino acid hydrophilic core sequence of serotype-specific NS2B (cNS2B, amino acids 1394–1430) linked via a flexible Gly4SerGly4 linker to the NS2B/NS3 domains (strain TSV01, GenBank accession number AY037116) and cloned into the vector pET15b (Novagen, Madison, WI). To obtain the cNS2B sequence, PCR was carried out using the primers DEN1 reverse (5′-TGGTTGGAAGAATCCTATGAG-3′), DEN3 reverse (5′-TGGTTGGAAGAATCCTATGAG-3′), and DEN4 reverse (5′-GTAAGTGCCAAAAGTCGGAATC-3′) with SuperScript II (Invitrogen) at 45 °C for 1 h, followed by PCR with Phi polymerase (Stratagene) and the primers DEN1 forward (5′-GGGAGGCTGATCTGCAATG-3′), DEN3 forward (5′-GAGGCTGATCTGCAATG-3′), and DEN4 forward (5′-CCATATGCTGCTGTTT-3′) along with the corresponding reverse primers.

Expression and Purification of DEN 1–4 CF40-Gly-NS3pro185—Competent E. coli BL21-CodonPlus(DE3) (Stratagene) were transformed with pET15b-DEN 1–4 CF40-Gly-NS3pro185 expression vectors and grown in 500 ml Luria-Bertani broth containing ampicillin (100 μg/ml), chloramphenicol (50 μg/ml), and 0.2% (w/v) glucose at 37 °C with shaking until an 670 reached 0.5. Cells were centrifuged in a Sorvall SLA 3000 rotor at 5000 × g for 10 min and resuspended in 500 ml of Luria-Bertani media with ampicillin and chloramphenicol. Cultures were induced with 0.4 mM isopropyl β-D-thiogalactopyranoside, and growth was continued for a further 16 h at 16 °C. The resulting cells were pelleted and resuspended in 30 ml of cold buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, and 5% glycerol). Cells were passed through a cell disruptor twice at 20,000 p.s.i. (Basic Z model; Constant Systems Ltd.), and debris was removed by centrifugation at 35,000 × g for 30 min. The protein solution was filtered by 0.22-μm filter and loaded onto a 5-ml HiTrap chelating heparin (Amersham Biosciences) column, equilibrated with the lysis buffer. The purified proteins were eluted from the column with lysis buffer and a linear gradient of imidazole from 20–300 mM in the same buffer. The peak fractions were analyzed by 10% SDS-PAGE. The positive fractions were pooled, desalted, and concentrated with spin concentrators (Amicon Ultra-15 ml; Millipore, Billerica, MA) with a molecular mass cutoff of 10,000 Da. Substrate Specificity of Dengue NS2B-NS3 Proteases—For P4–P4′ substrate specificity determination, two-position fixed positions scanning tetrapeptide libraries of the NS3 protease were assayed as described previously (12–15). Assays were carried out in 384-well plates on SpectraMax Gemini EM or XS microtiter plate reader (Molecular Device). The final reaction mixtures (30 μl) contained 50 mM Tris-HCl (pH 8.5), 20% glycerol, 1 mM CHAPS, and ~150 μM total substrate. After the addition of enzymes (1–5 μl CF40-Gly-NS3pro185 proteases) to the tetrapeptide coumarin library, reaction mixtures were incubated at 37 °C, and the liberated coumarin fluorophore was monitored at a λex of 380 nm and a λem of 450 nm. Initial fluorescent velocities in relative fluorescent units per second were calculated as a fraction of the highest velocity in the library set and plotted to a two-dimensional format with DecisionSite (Spotfire). The tetrapeptide donor quencher positional scanning library was synthesized and assayed as described previously (11–13). Briefly, CF40-Gly-NS3pro185 proteases (0.5–2 μl) were incubated in 96-well plates with 10 μl reaction mixture containing the same substrate as described above with ~100 μM total substrate (16, 17). The reactions were monitored at a λem of 320 nm and a λem of 380 nm, and initial velocities were analyzed and graphed in DeltaGraph.

Steady-state Kinetics of Fluorogenic and Chromogenic Peptide Substrates—Five fluorogenic tetrapeptide substrates with the 7-amino-3-carbamoylmethyl-4-methyl coumarin (ACM) leaving group (Bz-Nle-Lys-Arg-ACM, Bz-Nle-Lys-Thr-Arg-ACM, Bz-Nle-Lys-Arg-ACM, Bz-Thr-Lys-Arg-ACM, and Bz-Thr-Arg-ACM) were synthesized using standard Fmoc (9-fluorenylmethoxycarbonyl) solid phase peptide synthesis techniques (14, 15). The thiopeptide substrate, Bz-Nle-Lys-Arg-SBzl, was purchased from Peptron Technologies. After high performance liquid chromatography purification, the concentration of aliquots of each fluorogenic substrate was determined using total hydrolysis with trypsin, and the released ACM fluorophore was read at a λem of 380 nm and a λem of 450 nm. The concentration of each substrate was then calculated with standard ACM solutions. The concentration of the SBzl substrate was also determined with purified CF40-Gly-NS3pro185 proteases. A 10 μM solution of the substrate was monitored spectrophotometrically at 324 nm in the presence of 0.5 mM 4,4′-dithiodipyridine, and concentration was determined using the extinction coefficient of 19,800 M–1 cm–1 for the SBzl-thiopyridine conjugate. Active site titration for purified CF40-Gly-NS3pro185 proteases was performed by inhibition with freshly reconstituted apotrinin (18, 19). For kinetic studies, CF40-Gly-NS3pro185 proteases were incubated with various concentrations of individual ACM, AMC, or SBzl.
peptide substrates at 37 °C. The proteolytic reaction was monitored as an increase in fluorescence at a λex of 380 nm and a λem of 450 nm for the ACMC and AMC substrates or an increase in absorbance at 324 nm in the presence of 0.5 mM 4′-dithiodipyridine for the SBd substrate. Typical reaction mixtures (100 μl) contained 50 mM Tris-HCl, pH 8.5, 20% glycerol, 1 mM CHAPS, 10 mM enzyme, and fluorogenic/chromogenic peptide substrates ranging from 0.5 μM to 1 mM. Initial fluorescence or absorbance velocities (relative fluorescence units per minute or relative absorbance units per minute) were converted to M -1s -1 from a standard ACMC or AMC calibration curve or to an extinction coefficient of 19,800 M -1cm -1 for the SBd-thiogalactopyranose conjugate. Progression curves were fitted into a Michaelis-Menten equation by nonlinear regression using GraphPad Prism. Steady-state kinetic constants of each curve were determined from duplicate measurements and reported as mean ± S.E.

Model of Substrate Binding to the NS3pro Structure—The P4-P4′ octapeptide (Nle-Lys-Arg-Arg-Ser-Gly-Ser-Gly) was fitted to the active site of the enzyme using the crystal structure of the dengue NS3 protease complex with the mung bean Bowman-Birk inhibitor (Protein Data Bank code 1DF9) (20) as a guide. The side chains of residues 44–51 of the inhibitor, representing P4-P4′, were mutated to the sequence of the octapeptide and manually fitted to dengue NS3 protease with the molecular modeling program Maestro (Schrodinger LLC, Portland, OR), seeking to maximize electrostatic interactions, hydrogen bond formation, and hydrophobic interactions. The main chain coordinates were not moved, nor were any atoms of the protein altered.

RESULTS

In Vitro Expression and Purification of DEN1–4 CF40-Gly-NS3pro185—Clum et al. showed that the expression of the core hydrophilic domain of the DEN2 New Guinea C strain NS2B (cNS2B; 40 amino acids) as an N-terminal fusion was sufficient for activating the NS3 protease domain (21). The introduction of a flexible, protease-resistant, nine-amino acid linker (GlySerGlyGly) generated a soluble and catalytically active protease complex (11). In the study presented here, similar constructs were expressed based on this strategy by cloning the chimeric CF40-Gly-NS3pro185 cDNAs derived from dengue serotypes 1–4 into the pET15b vector from Novagen (see “Experimental Procedures”). Expression of the recombinant DEN2 CF40-Gly-NS3pro185 protease as an N-terminal His fusion protein in E. coli followed by affinity purification led to high yields (typically 15–20 mg from a 1-liter culture) of soluble protein, of which >95% were full-length (Fig. 1). The identity of CF40-Gly-NS3pro185 was confirmed with Western analyses using anti-His and anti-NS3 antibodies (Fig. 1). The minor lower band detected by the anti-NS3 antibody (Fig. 1B, asterisk), but not by the anti-His antibody (Fig. 1C), is presumably the heterodimeric protein without the hexahistidine tag. CF40-Gly-NS3pro185 proteases for DEN1, DEN3, and DEN4 were similarly expressed and purified, except that the yield of the DEN4-derived enzyme was 10-fold lower (data not shown). Active site titration with apronin was used to accurately assess the active enzyme concentration (11). Apronin binds to the four CF40-Gly-NS3pro185 proteases with high affinity (Kd 79, 25, 88, and 6.4 μM for DEN1–4 CF40-Gly-NS3pro185, respectively).

Characterization of Enzymatic Activity of DEN1–4 CF40-Gly-NS3pro185 on a Fluorogenic Tripeptide Substrate—The activities of the four CF40-Gly-NS3pro185 enzymes were characterized using the fluorogenic peptide Boc-GRR-AMC, which had been shown previously to be cleaved by the DEN2-NGC cNS2B/NS3 protease complex (10). The Km, Vmax, and kcat/Km or substrate specificity for DEN2 CF40-Gly-NS3pro185 using Boc-GRR-AMC were determined by varying the substrate concentration from 1000 to 10 μM using 12 serial dilutions (see “Experimental Procedures”). The steady-state kinetic parameters obtained were kcat = 0.13 ± 0.02 s -1, K m = 150 ± 15 μM, and kcat/Km = 840 ± 100 M -1 s -1 (Table I). The activities of the purified DEN1, DEN3, and DEN4 CF40-Gly-NS3pro185 were characterized using the same Boc-GRR-AMC fluorogenic peptide. All four proteases exhibited comparable K m values, but the kcat and kcat/Km values showed greater variations, with the values for DEN1 protease being lower and, hence, the least active (Table I). This observation was consistent also for a DEN1 protease cloned from a clinical isolate obtained during the Dengue fever outbreak in the Indonesian city of Jakarta (data not shown).

Profiling of P4-P1 Specificities of DEN1–4 CF40-Gly-NS3pro185—Sequence analysis of the NS3 proteases from all four distinct serotypes indicated that they share greater than 60% identity in their primary sequences (Fig. 2). To explore the substrate structure-activity relationship, the P4-P1 substrate specificities of recombinant NS3 proteases from DEN1–4 were examined using tetrapeptide positional scanning synthetic combinatorial libraries of the general structure Ac-XXXX-7-amino-4-carbamoylmethyl coumarin (Fig. 3) (14, 15, 17). The tetrapeptide substrates were synthesized and assayed as mixtures of peptides in a positional scanning format where two positions were fixed with a specific amino acid and two positions were randomized with 19 amino acids (X represents all
natural amino acids with the exception of Cys and Met and the inclusion of Nle). Specifically, the tetrapeptide substrates used in this study represented each combination of the P1 position fixed as a specific amino acid with either a fixed P2, a fixed P3, or a fixed P4 position for a total of 1083 wells (361 wells for inclusion of Nle). Specifically, the tetrapeptide substrates used in the library was 130,321 (19 randomized amino acids × 19 randomized amino acids × 1 fixed amino acid × 1 fixed amino acid). The total number of substrates in the library was 130,321 (19 × 19 × 19 × 19). Cleavage of the peptide-7-amino-4-carbamoylmethyl coumarin bond results in an increase in fluorescence that can be directly monitored. The total concentration of substrates in each well was ~150 or ~0.4 μM for each substrate. The relative rates for the mixture of substrates are represented in a two-dimensional matrix, with each square in the matrix representing both the identity of the two fixed amino acids (x-axis and y-axis) and the relative activity as indicated on a gray scale in which white represents no activity and black represents the highest activity (Fig. 3B). The activity of the enzyme across all three sub-libraries (P1 × P2, P1 × P3, and P1 × P4) was normalized to the highest activity as indicated by the white-to-black scale below each of the two-dimensional graphs. The enzymatic activity was also represented in histogram form (Fig. 3B), where the P1 position is fixed as arginine, the x-axis represents the P2, P3, and P4 fixed positions, and the y-axis represents the normalized hydrolysis rates in relative fluorescence units per second. The substrate specificity at each subsite in the tetrapeptide can be determined by the highest hydrolysis rate (in reflective fluorescent units per second) observed in the individual P2, P3, and P4 sub-libraries.

The substrate specificities for the four NS3 proteases from DEN1–4 were found to be very similar (Fig. 3B). Because the whole library is assayed at a constant substrate concentration, the relative importance of the amino acid at a particular position can be compared with the intensities of corresponding signals across the library. An exclusive preference for basic amino acid residues (either Arg or Lys) at the P1 position was observed in P1-P2, P1-P3, or P1-P4 fixed libraries (Fig. 3B, left). After plotting the activities from wells containing Arg at P1, P2–4 preferences were readily observed in the order of Arg > Thr > Gln/Asn/Lys for P2, Lys > Arg > Asn for P3, and Nle > Leu > Lys > Xaa for P4 (Fig. 3B, right).

**Steady-state Kinetic Constants for the Hydrolysis of Optimal Substrates by Dengue NS3 Proteases**—For serine proteases, the well established catalytic mechanism involves a two-step process of acylation and deacylation as shown in Scheme 1.

During the acylation step the catalytic serine acts as a nucleophile to attack the P1 carbonyl of the substrate, forming an acyl-enzyme intermediate where the non-prime portion of the peptide substrate remains covalently bound to the enzyme and the prime site segment of the substrate dissociates from the enzyme. In the subsequent deacylation step, water acts as the nucleophile to form the new C terminus of the cleaved substrate with the ensuing regeneration of the catalytic serine (22). A consequence of this catalytic mechanism, as shown in Equations 1 and 2,

\[ k_{cat} = \frac{k_3 \times k_5}{k_3 + k_5} \]  
\[ K_m = \frac{k_5}{k_3 + k_5} \]

is that the macroscopic steady state constants \( k_{cat} \) and \( K_m \) are related to the acylation (\( k_3 \)) and deacylation (\( k_5 \)) rate constants.

---

**Table 1**

| Serotype | Substrate | \( k_{cat} \) | \( K_m \) |
|----------|-----------|----------------|------------|
| DEN1     | Boc-GRR-AMC | 0.03 ± 0.01 | 120 ± 30 |
|          | Bz-nKRR-AMC | 0.32 ± 0.01 | 6.2 ± 0.8 |
|          | Bz-nKTR-AMC | 0.97 ± 0.07 | 202 ± 29 |
|          | Bz-nTRR-AMC | 0.28 ± 0.02 | 71 ± 15  |
|          | Bz-TKRR-AMC | 0.05 ± 0.01 | 6.6 ± 1.5 |
|          | Bz-TTRR-AMC | 0.21 ± 0.03 | 360 ± 77 |
| DEN2     | Boc-GRR-AMC | 0.13 ± 0.02 | 150 ± 15  |
|          | Bz-nKRR-AMC | 1.4 ± 0.1   | 12 ± 2   |
|          | Bz-nKTR-AMC | 1.4 ± 0.1   | 34 ± 10  |
|          | Bz-nTRR-AMC | 0.76 ± 0.03 | 46 ± 6   |
|          | Bz-TKRR-AMC | 0.20 ± 0.01 | 11 ± 1   |
|          | Bz-TTRR-AMC | 0.17 ± 0.01 | 76 ± 9   |
| DEN3     | Boc-GRR-AMC | 0.09 ± 0.02 | 125 ± 16  |
|          | Bz-nKRR-AMC | 0.61 ± 0.02 | 12 ± 1   |
|          | Bz-nKTR-AMC | 1.1 ± 0.1   | 180 ± 20 |
|          | Bz-nTRR-AMC | 0.58 ± 0.02 | 53 ± 5   |
|          | Bz-TKRR-AMC | 0.10 ± 0.01 | 13 ± 3   |
|          | Bz-TTRR-AMC | 0.33 ± 0.02 | 220 ± 23 |
| DEN4     | Boc-GRR-AMC | 0.07 ± 0.01 | 180 ± 40  |
|          | Bz-nKRR-AMC | 2.5 ± 0.1   | 73 ± 0.9 |
|          | Bz-nKTR-AMC | 4.9 ± 0.2   | 52 ± 5   |
|          | Bz-nTRR-AMC | 1.1 ± 0.02  | 72 ± 4   |
|          | Bz-TKRR-AMC | 0.43 ± 0.01 | 8.6 ± 0.7 |
|          | Bz-TTRR-AMC | 0.48 ± 0.07 | 210 ± 60 |
|          | Bz-nKRR-AMC | 2.9 ± 0.1   | 8.6 ± 1.8 |
|          | Bz-nKTR-AMC | 300 ± 25    | 6.0 ± 2.9 |

\( ^* \) Single letter amino acid abbreviations are used in the middle sequences. The lowercase “n” stands for norleucine.
When acylation is rate-determining, $k_5 \gg k_3$, the steady state kinetic constants simplify to $k_{\text{cat}} = k_3$ and $K_m = K_d$. For most serine proteases, acylation is rate-determining for amide bond hydrolysis, and deacylation is rate-determining for ester bond hydrolysis (22–24). To determine whether the acylation step is rate-determining for dengue NS3 protease, the steady-state kinetic constants were determined for two substrates that contained the same peptide sequence but with different leaving groups, namely Bz-Nle-Lys-Arg-Arg-ACMC, representing amide bond hydrolysis, and Bz-Nle-Lys-Arg-Arg-SBzl, representing ester bond hydrolysis. Numerical data are available in the supplemental material, which can be found in the on-line version of this article.
ing this ester bond hydrolysis. If deacylation were rate determining for both substrates, then the catalytic rate constants would be largely indistinguishable because the catalytic rate would be dependent on deacylation of the acyl-enzyme intermediate, and the acyl-enzyme intermediate formed by both substrates would be identical. If acylation is rate-determining for one or both of the substrates, then the catalytic rates would be significantly different and would depend on the relative reactivities of the leaving groups in the original complex. Indeed, for dengue NS3 protease (Table I, DEN4) the $k_{cat}$ for the Bz-Nle-Lys-Arg-SBzl substrate is ~1000-fold greater than that of the Bz-Nle-Lys-Arg-ACMC substrate, $k_{cat,SBzl} = 300 \text{ s}^{-1}$ versus $k_{cat,ACMC} = 2.8 \text{ s}^{-1}$. This observation is consistent with acylation being the rate-determining step for amide bond hydrolysis by dengue NS3 protease.

To validate the substrate specificities identified from the non-prime site substrate profiling, we synthesized a series of individual substrates and examined the contribution of the P2-P4 positions by comparing their steady-state kinetic parameters ($K_m$, $k_{cat}$, and $k_{cat}/K_m$; Table I). Bz-Nle-Lys-Arg-ACMC contains the combination of optimal amino acids at P4-P1, whereas the other substrates (Bz-Nle-Lys-Thr-Arg-ACMC, Bz-Nle-Thr-Arg-ACMC, Bz-Thr-Lys-Arg-ACMC, and Bz-Thr-Thr-Arg-ACMC) each contain one or two substitutions in P2-P4, with the corresponding optimal residue replaced by the suboptimal Thr residue. The $k_{cat}/K_m$ values of Bz-Nle-Lys-Arg-ACMC for the four NS3 enzymes are 75–1000-fold higher than that of Boc-Gly-Arg-Arg-AMC, a widely used NS3 substrate (Table I). The increase in activity is due to both a decrease in $K_m$ and an increase in $k_{cat}$. This drastic difference is not likely caused by the use of a modified fluorophore, because Bz-Nle-Lys-Arg-AMC and Bz-Nle-Lys-Arg-ACMC have comparable $K_m$ and $k_{cat}$ values (Table I, DEN4).

The recognition of dibasic residues at the P1 and P2 sites by dengue NS3 protease is considered the key specificity characteristic of flaviviral NS3 enzymes and has been reported by a number of groups (7, 8, 25). The data presented here demonstrate that P2 is very sensitive to substitution and supports the role of P2 in substrate ground state binding in view of the markedly increased $K_m$ of the suboptimal substrate Bz-Nle-Thr-Arg-ACMC when compared with Bz-Nle-Lys-Arg-ACMC (Table I).

The single substrate kinetic data also indicate that non-prime subsites beyond P2 also contribute significantly to substrate binding and turnover. Specifically, a suboptimal P3 substitution (Bz-Nle-Thr-Arg-Arg-ACMC) causes an increase in $K_m$ up to 10-fold but has little influence on $k_{cat}$. The P4 substitution with a suboptimal amino acid, Bz-Thr-Lys-Arg-ACMC, maintains a similar $K_m$ but displays a 6-fold decrease in $k_{cat}$. The role of P4 in catalysis can also be observed when comparing the suboptimal substrates Bz-Nle-Thr-Arg-Arg-ACMC and Bz-Thr-Thr-Arg-ACMC. Not surprisingly, changing both P3 and P4 (Bz-Thr-Thr-Arg-Arg-ACMC) to suboptimal residues affects both $K_m$ and $k_{cat}$ and leads to the loss of $k_{cat}/K_m$ by ~34–168-fold.

**Profiling of P1’-P4’ Specificities of DEN1–4 NS3 Protease**—A number of observations have suggested the presence of prime site substrate specificity in dengue NS3 proteases. Murthy et al. (9) first reported the crystal structure of the apo NS3 serine protease domain at 2.1 Å. This structure revealed a restricted substrate binding cleft with few predicted interactions beyond P2-P2’. Defined interactions with the prime site pockets was recently observed in the structure of NS3 protease complexed with a Bowman-Birk inhibitor that has P1’-Ser and P3’-Pro in both active-site loops (20). Further direct evidence stems from a mutagenesis study on the S2’ pocket where a single Gly-133 to Ala substitution strongly reduced the auto-processing of NS2B-NS3 (27).

To further elaborate the prime site substrate specificity of NS3, a focused P1’-P4’ octapeptide donor-querencer library was synthesized in a positional scanning format (Fig. 4A). The P1’-P4’ region of the donor-querencer positional scanning library contained a tetrapeptide sequence with one position fixed as a specific amino acid and three positions randomized as 19 amino acids for a total of 130,321 substrates (19 × 19 × 19 × 19) in mixtures of 6,859 per well (19 randomized amino acids × 19 randomized amino acids × 19 fixed amino acid). Cleavage of the peptide between the fluorescence donor methoxycoumarin group and the quencher dinitrophenyl group results in an increase in fluorescence. The library was designed to bias for cleavage between the P1 and P1’ positions by occupying the P4-P1 sites with the sequence Nle-Lys-Arg-Arg, the non-prime specificity determined from the tetrapeptide positional scanning library (Fig. 3). The dependence of the hydrolysis rate on the identity of the amino acid in the prime site position is represented in Fig. 4B, where the x-axis represents the amino acid in the fixed position and the y-axis represents the relative rate of hydrolysis in relative fluorescent units per second.

Dengue NS3 proteases from all four serotypes displayed similar prime site substrates specificity as observed in the donor-querencer positional scanning substrate library (Fig. 4). In particular, P1’ and P3’ sites exhibited specificity for small and polar amino acids such as serine, whereas the P2’ and P4’ substrate positions showed minimal activity when compared with the P1’ and P3’ positions.

**Correlation of Substrate Specificities with Natural Cleavage Sites**—The NS3 protease is responsible for the cleavage of at least 2A2B, 2B/3, 3/4A, and 4B/5 boundaries of the virus encoded polyprotein (Fig. 5B) (6, 7, 25, 28, 29). These sites are highly conserved; dibasic residues at P1 and P2 are followed by a small or polar residue at P1’ (Ser, Gly, or Ala). The only exception is that the N2B/3 sites of all four dengue serotypes contain a Gln residue at the P2 position. The dibasic cleavage patterns are evident from the substrate profiling experiment (summarized in Fig. 5A). At the P3 position, the four sites of DEN1–3 contain optimal Lys/Arg or near optimal Gly, whereas three of DEN4 sites are occupied with unfavorable Ser/Thr/Pro. On the prime site, the strong preference for Ser at P3’ in the profiling study was not reflected in the native sites except for the 3/4A linker in DEN1.

**Octapeptide Substrate Docking into the NS3 Active Site**—The structure of the dengue NS3 protease complexed with Bowman-Birk inhibitor was used to dock the optimized octapeptide (Nle-Lys-Arg-Ser-Gly-Ser-Gly) by mutating the corresponding P4-P4’ residues in the bound inhibitor. The schematic of the structure with the potential interacting side-chains is shown in Fig. 6. Despite the strong preference for a positive charge at P1 and P2 positions in the substrate, the corresponding substrate binding pockets do not appear to contain any negative charges. In the S1 pocket, it has been recently suggested that the Tyr-150 may be involved in r-interaction with the side-chain of P1-Arg (30). The residues that line the P1’-pocket are mostly conserved in all the NS3 sequences (data not shown) except at position 115, where Leu, Thr, or Ile can be found. The predicted S2 pocket is lined by the side chains of Gln-35 and Asn-152. All three residues are completely conserved, and it is conceivable that the P2 Arg side chain may hydrogen bond with Gln-35. In the S3 pocket the side chains of Leu-128, Asp-129, Pro-132, and Val-155 are completely conserved in all NS3 sequences with the exception of Arg-157,
Fig. 4. P1'-P4' substrate specificity of NS3 proteases. A, representative structure of the donor-quencher peptide substrate in a positional scanning library format. P1'-P4' represents positions in the peptide that are either fixed as a specific amino acid or are an equimolar mixture of...
these serotypes can infect human and induce host immunity to the corresponding serotype. However, dengue hemorrhagic fever or dengue shock syndrome is thought to be mostly associated with secondary infections and therefore remains a factor that poses challenges to the development of a safe vaccine (31). The encoded dengue viral protease, NS3, is an attractive therapeutic target, as it is essential for the formation of the virally encoded non-structural proteins (32, 33).

Although efforts are under way to develop protease inhibitors against dengue viral infection, the question remains open if a pan-serotype inhibitor can be developed or if multiple inhibitors will have to be designed for individual serotypes. Based on the sequence analysis, the NS3 proteases from the four serotypes share >60% identity in the protease domains (Fig. 2). To reveal the functional similarity between the four dengue NS2B/NS3 proteases, we report the bacterial expression and purification of highly active chimeric single chain NS2B/NS3 proteases from all four serotypes. With combinatorial peptide substrate libraries it was demonstrated for the first time that all four enzymes exhibit very similar substrate specificities at both non-prime sites (Fig. 3) and prime sites (Fig. 4). Individual substrate kinetics further confirmed the similar preference and sensitivity to replacement at P4-P1 positions by all four NS3 proteases (Table I). These results suggest that the four NS3 proteases share very similar, if not identical, peptide substrate structure activity relationships and imply that it is possible to develop a single inhibitory agent targeting all four dengue NS3 proteases.

The other open question pertains to the extent of the enzyme substrate binding site. The crystal structure of the DEN2 NS3pro in the absence of NS2B shows a shallow substrate binding site, indicating that significant interactions are restricted to P2-P2' (9). The substrate profiling study presented here clearly supports the importance of P2-P2' in substrate binding as indicated by the strong dibasic preference at P1 and P2 as well as the small amino acids at P1' (Figs. 3 and 4). The critical role of P2-P2' is also reflected in changes of kinetic parameters upon single residue substitution (Table I). For ex-
ample, Bz-Nle-Lys-Arg-Arg-ACMC binds to the enzyme 3–30-fold tighter as compared with its P2 suboptimal counterpart, Bz-Nle-Lys-Thr-Arg-ACMC (3–30-fold increase in \( K_{m} \)). More importantly, the profiling and kinetic data reveal for the first time that the P3 position also contributes to the substrate ground state binding. Substitution of optimal P3 Lys to Thr leads to a increase in \( K_{m} \) by 4–10-fold for all four NS3 proteases. Besides P3, a clear preference for Ser at P3’ also suggests a strong enzyme-substrate interaction beyond P2-P2’. The discrepancy between this observation and the crystal structure could be explained by the difference of enzyme source. The enzymes used in this study are composed of the protease domain of NS3 tethered to the central 40 amino acids of hydrophilic NS2B element that confers full proteolytic activity to NS3 and, therefore, more closely resembles its native conformation (11). In contrast, the crystal structure is derived from the NS3pro domain in the absence of NS2B (9, 20). Although the structure without the cofactor resembles that of the related hepatitis C NS3 protease bound to its activating peptide, NS4A, the exact mechanism by which the NS2B cofactor stimulates the protease activity is not currently known. Yusof et al. (10) experimentally compared the kinetic properties of NS3 and NS2B/NS3, and their results suggested that NS2B generates additional specific interactions with the P2 and P3 residues of the substrates. It is possible that NS2B activates the NS3 protease directly or interacts with the NS3 protease domain, causing a conformational change in the substrate binding pockets (10). Although substitution at P4 from suboptimal Thr to optimal Nle does not significantly alter the \( K_{cat} \), it does increase \( k_{cat}/K_{m} \) by 6-fold. This result again supports the idea that the substrate specificity of NS3 protease is influenced by more extensive contact than has been reported previously.

Several groups have characterized the enzymatic properties of NS2B/NS3 proteases with synthetic peptide substrates bearing endogenous dengue cleavage sites (11, 34, 35). The best substrate from both studies, Ac-Thr-Thr-Ser-Thr-Arg-Arg-

amide, is the extended substrate specificity of dengue serine proteases from all four distinct serotypes by using a combination of synthetic positional scanning combinatorial libraries and single substrate kinetics. This study represents the first observation on the conserved and extended substrate specificities among the four dengue NS3 proteases. The data provided here should facilitate the development of dengue NS3 protease inhibitors with detailed peptide substrate structure-activity relationships and greatly improve protease activity detection agents.

Acknowledgments—We thank Feng Chi Shyang, Phong Wai Yee, and Andy Yip for assistance in cloning the DEN1–4 protease expression constructs and Drs. Mark Schreiber and Ooi Eng Eong for bioinformatics assistance and for providing the DEN3 virus stock, respectively.

REFERENCES

1. Gubler, D. J. (1998) Clin. Microbiol. Rev. 11, 480–496
2. Monath, T. P. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 2395–2400
3. Halstead, S. B., Rojanasuphot, S., and Sangkawibha, N. (1983) Am. J. Trop. Med. Hyg. 32, 154–156
4. Chambers, T. J., Hahn, C. S., Galler, R., and Rice, C. M. (1990) Annu. Rev. Microbiol. 44, 649–688
5. Stadler, K. A., Aloni, S. L., Schalich, J., and Heinz, F. X. (1997) J. Virol. 71, 8475–8481
6. Preuschgat, F., Yao, C. W., and Strauss, J. H. (1990) J. Virol. 64, 4364–4374
7. Falgout, B., Petelth, M., Zhang, Y. M., and Lai, C. J. (1991) J. Virol. 65, 2467–2475
8. Zhang, L., Mohan, P. M., and Padmanabhan, R. (1992) J. Virol. 66, 7549–7554
9. Murthy, H. M., Chun, S., and Padmanabhan, R. (1999) J. Biol. Chem. 274, 5573–5580
10. Yusof, B., Chun, S., Wetzel, M., Murthy, H. M., and Padmanabhan, R. (2000) J. Biol. Chem. 275, 9963–9969
11. Leung, D., Schroder, K., White, H., Fang, N. X., Stoermer, M. J., Abbenante, G., Martin, J. L., Young, P. R., and Fairlie, D. P. (2001) J. Biol. Chem. 276, 45762–45771
12. Pinilla, C., Appel, J. R., Blanc, P., and Houghten, R. A. (1992) BioTechniques 13, 901–902, 904–905
13. Rano, T. A., Timkey, T., Peterson, E. P., Rotonda, J., Nicholson, D. W., Becker, J. W., Chapman, K. T., and Thornberry, N. A. (1997) Chem. Biol. 4, 149–155
14. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 7754–7759
15. Harris, J. L., Alper, P. B., Li, J., Rechtsteiner, M., and Backes, B. J. (2001) Chem. Biol. 8, 1131–1141
16. Petrasen, H. M., Williams, J. A., Li, J., Tumanut, C., Ek, J., Nakai, T., Masick, B., Backes, B. J., and Harris, J. L. (2005) Bioorg. Med. Chem. Lett. 15, 3162–3166
17. Shipway, A., Danahay, H., Williams, J. A., Tully, D. C., Backes, B. J., and Harris, J. L. (2004) Biochem. Biophys. Res. Commun. 324, 953–963
18. Copeland, R. A. (2000) Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis, 2nd Ed., pp. 313–317, Wiley-VCH, Inc., New York
19. Kuzmiz, P., Sideris, S., Cregar, L. M., Elrod, K. C., Rice, K. D., and Jane, J. W. (2000) Anal. Biochem. 281, 62–67
20. Murthy, H. M., Judge, K., Delucas, L., and Padmanabhan, R. (2000) J. Mol. Biol. 301, 759–767
21. Chun, S., Elsner, K. E., and Padmanabhan, R. (1997) J. Biol. Chem. 272, 30715–30723
22. Zerene, B., and Bender, M. L. (1964) J. Am. Chem. Soc. 86, 3669–3674
23. Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) Science 255, 1249–1253
24. Harris, J. L., Peterson, E. P., Hudig, D., Thornberry, N. A., and Craik, C. S. (1990) J. Biol. Chem. 263, 27364–27373
25. Cahour, A., Falgout, B., and Lai, C. J. (1992) J. Virol. 66, 1535–1542
26. Wiegers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H., and Krauslich, H. G. (1998) J. Virol. 72, 2846–2854
27. Valle, R. P., and Falgout, B. (1998) J. Virol. 72, 624–632
28. Preuschgat, F., and Straus, J. H. (1991) Virology 185, 689–697
29. Arias, C. F., Preuschgat, F., and Straus, J. H. (1993) Virology 193, 888–899
30. Nall, T. A., Chappell, K. J., Stoermer, M. J., Fang, N. X., Tyndall, J. D. A., Young, P. R., and Fairlie, D. P. (2004) J. Biol. Chem. 279, 48535–48542
31. Rothman, A. L. (2004) J. Clin. Investig. 113, 946–951
32. Matusan, A. E., Kelley, P. G., Pryor, M. J., Whisstock, J. C., Davidson, A. D., and Wright, P. J. (2001) J. Gen. Virol. 82, 1647–1656
33. Matusan, A. E., Pryor, M. J., Davidson, A. D., and Wright, P. J. (2001) J. Virol. 75, 9633–9643
34. Khumthong, R., Angushanomasombat, C., Panyim, S., and Katzenmeier, G. (2000) J. Biochem. Mol. Biol. 35, 206–212
35. Khumthong, R., Niyomatjantakun, P., Chanthaphap, S., Angushanomasombat, C., Panyim, S., and Katzenmeier, G. (2003) Protein Pept. Lett. 10, 19–26
36. Erickson-Vitanen, S., Manfredi, J., Vitanen, P., Tribe, D. E., Treh, B., Gunstone, C. A., III, Leeb, D. D., and Swanstrom, R. (1989) AIDS Res. Hum. Retroviruses 5, 577–591
37. Pettit, S. C., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Lanterm, P. V., Klein, C. A., and Swanstrom, R. (1994) J. Virol. 68, 8017–8027
Functional Profiling of Recombinant NS3 Proteases from All Four Serotypes of Dengue Virus Using Tetrapeptide and Octapeptide Substrate Libraries

Jun Li, Siew Pheng Lim, David Beer, Viral Patel, Daying Wen, Christine Tumanut, David C. Tully, Jennifer A. Williams, Jan Jiricek, John P. Priestle, Jennifer L. Harris and Subhash G. Vasudevan

J. Biol. Chem. 2005, 280:28766-28774. doi: 10.1074/jbc.M500588200 originally published online June 1, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M500588200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2005/06/14/M500588200.DC1

This article cites 36 references, 20 of which can be accessed free at
http://www.jbc.org/content/280/31/28766.full.html#ref-list-1