THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 50, pp. 38448–38458, DECEMBER 15, 2006

Insights to Substrate Binding and Processing by West Nile Virus NS3 Protease through Combined Modeling, Protease Mutagenesis, and Kinetic Studies*[S]

Received for publication, August 10, 2006, and in revised form, October 18, 2006 Published, JBC Papers in Press, October 19, 2006, DOI 10.1074/jbc.M607641200

Keith J. Chappell[‡1], Martin J. Stoermer[§1], David P. Fairlie[§2], and Paul R. Young[‡3]

From the [‡]School of Molecular and Microbial Sciences and [§]Centre for Drug Design and Development, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia

West Nile Virus is becoming a widespread pathogen, infecting people on at least four continents with no effective treatment for these infections or many of their associated pathologies. A key enzyme that is essential for viral replication is the viral protease NS2B-NS3, which is highly conserved among all flaviviruses. Using a combination of molecular fitting of substrates to the active site of the crystal structure of NS3, site-directed enzyme and cofactor mutagenesis, and kinetic studies on proteolytic processing of panels of short peptide substrates, we have identified important enzyme-substrate interactions that define substrate specificity for NS3 protease. In addition to better understanding the involvement of S2, S3, and S4 enzyme residues in substrate binding, a residue within cofactor NS2B has been found to strongly influence the preference of flavivirus proteases for lysine or arginine at P2 in substrates. Optimization of tetrapeptide substrates for enhanced protease affinity and processing efficiency has also provided important clues for developing inhibitors of West Nile Virus infection.

West Nile virus (WNV)[4] is a member of the *Flavivirus* genus, which contains many significant human pathogens, including dengue virus (Den), Japanese encephalitis virus (JE), and yellow fever virus (YF), and was first isolated in 1937 from Uganda's West Nile province. WNV has subsequently been found in regions of Africa, the Middle East, Europe, Russia, western Asia, and Australia (less severe subtype Kunjin) and most recently in North America (1). WNV is transmitted by *Culex* mosquitoes from avian reservoir hosts to vertebrate dead end hosts, including humans and horses (2). Human infection is generally asymptomatic or causes a mild febrile disease, West Nile fever. However, more recent infections of WNV have also been associated with higher rates of severe neurological disease and fatalities, particularly among the elderly (2). Since the introduction of WNV into New York in 1999, the virus has spread rapidly throughout North America, infecting over 19,000 people and causing more than 700 fatalities (see the Center for Disease Control and Prevention site on the World Wide Web at www.cdc.gov/ncidod/dvbid/westnile/index.htm). Currently there is no vaccine or antiviral therapy for the prevention or treatment of human WNV infection (1).

WNV is a small, enveloped virus with a single-stranded, positive sense 11-kb RNA genome, which encodes a single polyprotein precursor. This polyprotein must be cleaved co- and post-translationally to produce 10 functional proteins: three structural (C, prM, and E) and seven nonstructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). Translation of the viral polyprotein is membrane-associated with host proteases cleaving junctions within the lumen of the endoplasmic reticulum and the Golgi, whereas a viral protease encoded within NS3 cleaves at the junctions NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5 and also internal sites within C, NS3, and NS4A (Fig. 1) (3). Cleavage at these sites by the NS3 protease is essential for viral replication, so the protease is a potential therapeutic target (4–7).

NS3 is a multifunctional protein, the protease comprising the N-terminal third and nucleotide triphosphatase, RNA triphosphatase, and helicase components comprising the remainder (8–10). NS3 is a trypsin-like serine protease with a classical catalytic triad (His-51, Asp-75, Ser-135) (11) and is highly specific for substrates with dibasic P1 and P2 components and a small amino acid at P1′. This recognition sequence is highly conserved throughout flaviviruses (12); however, different flaviviruses prefer either Lys or Arg at P2. Although Den and YF NS3 proteases predominantly recognize Arg at P2, WNV protease recognizes Lys at P2 (Table 1). The activity of flavivirus NS3 proteases is dependent on an NS2B cofactor, with truncation studies in Den2 having shown that a central 40-amino acid hydrophilic domain is sufficient for activity (13). The flanking

* This work was supported by the National Health and Medical Research Council of Australia and the Australian Research Council. 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.

[S] The on-line version of this article (available at http://www.jbc.org) contains Tables S1 and S2.

[1] These authors contributed equally to this work.

[2] To whom correspondence may be addressed: Centre for Drug Design and Development, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia. Tel.: 617-3346-2989; Fax: 617-3346-2101; E-mail: d.fairlie@imb.uq.edu.au.

[3] To whom correspondence may be addressed: School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland 4072, Australia. Tel.: 617-3365-4646; Fax: 617-3365-4620; E-mail: p.young@uq.edu.au.

[4] The abbreviations used are: WNV, West Nile virus; Aib, 2-aminoisobutyric acid; Aoc, L-2-aminooctanoic acid; BSQV, Bussuquara virus; Cha, L-cyclohexylalanine; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; Chg, L-cyclohexylglycine; Cit, L-citrulline; hArg, L-homoarginine; JE, Japanese encephalitis virus; MVE, Murray Valley encephalitis; Nle, L-norleucine; Nva, L-norvaline; Orn, L-ornithine; *p*NA, *para*-nitroanilide; Tbg, L-*tert*-butylglycine; YF, yellow fever virus; ZIKV, Zika virus; Den, dengue virus; SLEV, St. Louis encephalitis virus; LGTV, Langat virus; TBE, tick-borne encephalitis virus.

hydrophobic domains within NS2B are likely to function in promoting membrane association of NS2B-NS3 (14).

Due to their pivotal roles in both normal physiology and disease, proteases are increasingly attracting interest as pharmaceutical targets (15). Since early successes in human immunodeficiency virus chemotherapy (human immunodeficiency virus-protease inhibitors) and in the treatment of high blood pressure (angiotensin-converting enzyme (ACE) inhibitors), a large number of new protease inhibitors have entered clinical trials (16). One reason for the drug potential of proteases is the relatively predictable way in which they recognize their substrates and inhibitors in extended (β-strand) conformations (4, 17). There are no known examples of proteolytic processing of peptide α-helices, β-sheets, or β-turns. Greater access to three-dimensional structures for proteases (over 1500 in the Protein Data Bank) has also facilitated hybrid structure/substrate-based drug design (5).

Recently reported crystal structures for NS2B/NS3 proteases of both WNV and Den2 (6) provide new structural insights to flaviviral proteases in ligand-bound conformations. An earlier homology model of the WNV protease (7), derived from the crystal structures of a highly homologous dengue NS3 protease without NS2B cofactor (18) and a less homologous hepatitis NS3 with bound NS4A cofactor (19), differs significantly from the crystal structure of WNV NS2B-NS3. This has prompted a reexamination now of some of the previous mutagenesis data (20). The reported WNV protease crystal structure shows an N-capped tetrapeptide aldehyde inhibitor bound in the substrate-binding cleft as a loop (instead of the commonly observed β-strand conformation) with the "P5" capping benzoyl residue sitting on top of the P1 residue. It therefore seemed unlikely that this ligand had the same binding mode as substrates beyond P1 and P2.

In previous kinetic studies using short hexapeptide *p*-nitroanilide substrates derived from endogenous polypeptide cleavage sites, we found no preference of WNV protease for specific residues except at P1 and P2 (7). However, more recent studies using nonnative dengue hexapeptide and decapeptide[5] and tetrapeptide and octapeptide (21) substrate sequences have suggested opportunities for enhancing substrate affinity for flaviviral proteases using nonnative or nonproteinogenic amino acids with hydrophobic (Nle, Leu) residues at P4 and a basic (Lys > Arg) residue at P3, a feature not seen in suboptimal native sequences. Using a combination of computer docking of substrates into the enzyme crystal structure, site-directed mutagenesis of the protease (Fig. 1), and kinetic studies of the processing of tetrapeptide substrates, we have focused the present study on increasing substrate affinity and processing efficiency, identifying the enzyme residues likely to be involved in binding to the P2–P4 positions of substrates, and taking early steps toward potent substrate-based nonpeptidic inhibitors by incorporating unnatural amino acids in tetrapeptide substrates.

EXPERIMENTAL PROCEDURES

para-Nitroanilide (pNA) Substrate Synthesis—*p*NA substrates were synthesized according to the general method of Abbenante *et al.* (22) and characterized by analytical high performance liquid chromatography, mass spectrometry, and NMR (see supplemental material).

Modeling of Substrates into the NS3 Crystal Structure—The crystal structure of West Nile virus NS2B/NS3 protease (Protein Data Bank code 2fp7) was prepared for docking by adding protons using InsightII (version 2000; Accelrys Inc.). Substrates were assembled using the Biopolymer and Sketcher modules within InsightII and minimized using Discover. All substrate docking experiments were conducted using GOLD version 2.1.2 (23). Hydrogen bonding and distance constraints were used to align the substrate within the active site as follows. The P1 Arg residue was positioned as observed for the corresponding residue in the aldehyde inhibitor complex (Protein Data Bank code 2fp7) (24), using a distance constraint of 3.5 ± 1.5 Å between the Arg ζ-carbon and the aromatic γ-carbon of Tyr-161 in the S1 pocket. This positions the positive charge of the arginine optimally for a π-cation interaction but also enables a charge-charge interaction with Asp^{129}. A hydrogen bond between the P2 Lys *z*-NH_3^+ and the Asn-152 side-chain carbonyl oxygen was used to anchor P2 in the shallow solvent-exposed S2 pocket, as predicted in earlier modeling work (7) and later verified experimentally (24). Hydrogen bond constraints for critical substrate backbone-enzyme interactions between NS3-Gly-153 carbonyl oxygen and substrate P3 NH, Gly-153 NH and substrate P3 carbonyls, and Gly-151 carbonyl oxygen and Arg P1 αNH were also used. No constraints on the position of P3 and P4 side chains were used. For the larger 2-naphthoyl residue in 2-naphthoyl-KKR-*p*NA, the docking (Fig. 2) produced poses with a high degree of steric clash in the vicinity of S4. In these cases, the docking poses were minimized with the enzyme backbone either fixed or tethered and the enzyme side chains and the entire substrate allowed to move using Discover (Accelrys).

Plasmid Construction—The expression plasmid pQE9 WNV CF40.Gly.NS3pro, previously generated from WNV (strain NY99-4132) (10), was used as a template for site-directed mutagenesis. The site-directed mutations in NS2B (V75A, V75F, N84A, N84D, N84E, N84L, N84S, Q86A, Q86E, Q86L, L87A, and L87F) and NS3 (T111F, T111L, V154F, V154L, I155F, M156A, I162F, A164S, A164V, and V166L) were inserted by amplifying the entire plasmid with PCR using Phusion™ polymerase (Finnzymes) together with a pair of partially overlapping primers (supplemental material). The template plasmid was digested with DpnI at 37 °C for 1 h. The PCR-amplified plasmid was then used to transform *Escherichia coli* strain XL10 Gold competent cells, which were grown in the presence of 100 μg/ml ampicillin, the expression plasmid was purified, and sequences were confirmed by automated analysis.

Enzyme Expression and Purification—The pQE9 vector was used to allow high level, inducible expression of N-terminal His_6-tagged recombinant proteins. Cultures of *E. coli* strain SG13009 transformed with the expression plasmids containing the site-directed mutations were grown in 2 × 25 ml of LB medium containing 100 μg/ml ampicillin and 25 μg/ml kanamycin at 37 °C until the A_{600} reached 0.5. Expression of the recombinant protein was induced by the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 0.3 mM and incubated for an additional 3 h at 22 °C. Cells were then

[5] L. Juliano, personal communication.

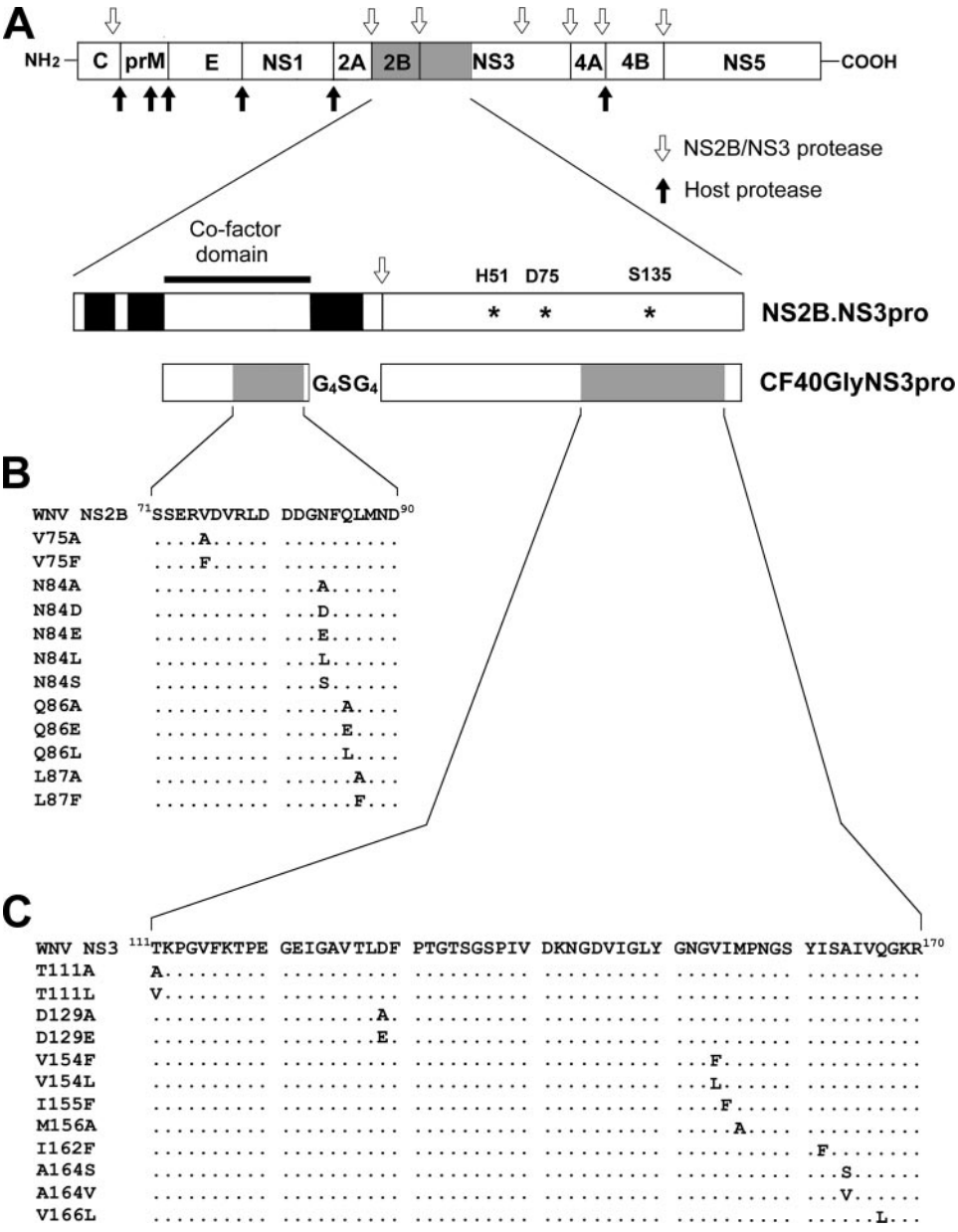

FIGURE 1. **Flavivirus polyprotein processing.** *A*, the *upper schematic* shows the sites cleaved by host proteases (*block arrows*) and virus-encoded NS2B/NS3 protease (*open arrows*). The proteolytic domain of NS3 and the NS2B cofactor are *shaded*. The *schematic below* shows this region expanded and the recombinant construct used in this study. Features of the NS2B-NS3pro complex include the cleavage site (*open arrow*), the catalytic triad (*asterisks*), the hydrophobic domains of NS2B (*solid boxes*), and the minimum 40-amino acid cofactor domain required for catalytic activity. The construct CF40.gly.NS3pro comprises the essential 40-amino acid domain of NS2B linked to the NS3 protease domain by a flexible glycine linker (Gly_4,Ser,Gly_4). *B*, sequence of wild type WNV NS2B between residues 71 and 90 (*top sequence*) and the residues mutated in this study. *C*, sequence of wild type WNV NS3 between residues 111 and 170 (*top sequence*) and the residues mutated in this study.

harvested by centrifugation at 4500 × *g* for 10 min and stored at −20 °C. For protein purification, the cell pellets were thawed and resuspended in 1 ml of lysis buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 10 mM imidazole, 5% glycerol). To prevent proteolytic cleavage of protein during lysis and purification, the following protease inhibitors were added to give the final concentrations of 1 μg/ml aprotinin, 1 μg/ml leupeptin, 1 μg/ml benzamidine, and 1 mM phenylmethylsulfonyl fluoride. Resuspended cells were lysed by sonication, and insoluble products were pelleted by centrifugation at 27,000 × *g* for 20 min. The recombinant proteases were purified by affinity chromatography using an N-terminal His_6 tag on Ni^{2+} nitrilotriacetic acid-agarose. Resin (0.5 ml) was pre-equilibrated with 10 ml of column buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 10 mM imidazole, 5% glycerol), and then the resin was removed and mixed with the supernatant of the cell lysates. These mixtures were incubated at 4 °C on a shaker for 30 min to allow the His-tagged protein to bind to the Ni^{2+} column. Resin was pelleted at low speed (100 × *g*), and the buffer was removed. The resin was washed with 3 × 5 ml of column buffer containing 50 mM imidazole, and the proteins were eluted into a single 300-μl fraction with column buffer containing 500 mM imidazole. Purification was confirmed by 12% SDS-PAGE.

Enzymatic Characterization and Substrate Analysis—Purified recombinant protease, WNV CF40.Gly. NS3pro, and site-directed mutants were assayed against tetrapeptide (Ac-LKKR-*p*NA) and hexapeptide (Ac-LQYTKR-*p*NA) substrates corresponding to P6–P1 of cleavage sites in the endogenous substrates but with a chromogenic *p*NA group at the P1′ position. Cleavage of *p*NA from the peptides by WNV protease produced a yellow color that allowed monitoring at 405 nm. The assay was conducted in a 96-well plate, with a final reaction volume of 200 μl containing 0.25 or 0.5 μM recombinant protease, using optimized conditions (final concentration of 50 mM glycine-NaOH, pH 9.5, 30% glycerol, and 1 mM CHAPS) (7). Eight different substrate concentrations, each in triplicate, were used for determining kinetic constants. After preincubation in separate wells (10 min, 37 °C), catalysis was initiated by mixing substrate with enzyme-buffer solution by automatic shaking for 5 s. The optical density was measured at 405 nm every 11–30 s for 210 s to 30 min (depending on activity) in a SpectraMax 250 reader, and the average change in millioptical density/min was calculated. For low substrate concentration, where there was a visible loss in activity over time, only the first five points were used to calculate the average change in millioptical density/min. Kinetic parameters were calculated from weighted nonlinear regression of the initial velocities as a function of the eight substrate concentrations using Graphpad Prism 4® software. The parameters k_{cat}, K_m, and k_{cat}/K_m were cal-

culated assuming Michaelis-Menten equilibrium kinetics, $v = V_{max}[S]/([S] + K_m)$. Triplicate measurements were taken for each data point, and means ± S.E. are reported.

RESULTS

Predicted Substrate-Enzyme Interactions—The reported crystal structure of WNV NS3 protease in association with the cofactor domain of NS2B has increased the understanding of the mechanism of substrate binding and protease activity. Since the enzyme was also bound to a tetrapeptide-aldehyde inhibitor, it was possible to observe enzyme residues that interact with P1 and P2 side chains. However, because the inhibitor was not bound in the extended β-strand conformation typical of substrate-protease interactions (4, 17), enzyme contacts with substrate beyond P1 and P2 positions could not be deduced from that crystal structure. In this work, we sought to dissect critical substrate interactions with enzyme at S2, S3, and S4. Cleavage sites for WNV protease vary considerably between P6 and P3 in native polypeptide substrates (*e.g.* DPNRKR ↓ GW (NS2A-NS2B), LQYTKR ↓ GG (NS2B-NS3), FASGKR ↓ SQ (NS3-NS4A), KPGLKR ↓ GG (NS4A-NS5)), with polar, acidic, basic, and hydrophobic residues all tolerated at P6–P3 positions.

To predict protease residues that are important for substrate binding, we conducted molecular modeling experiments using GOLD to dock tetrapeptide *p*NA substrates into the crystal structure of the WNV NS2B/NS3 protease (Protein Data Bank code 2fp7). The tetrapeptide substrate Ac-LKRR-pNA spanning P4–P1 had been previously identified (21) as optimal for dengue protease and, since we knew that WNV elicited a preference for Lys over Arg at P2, our docking studies began with the tetrapeptide substrate Ac-LKKR-*p*NA. Docking flexible molecules such as peptides into rigid solid state structures of proteases is notoriously difficult due to inadequate sampling of conformational space and insufficiently minimized docked poses (25) and because GOLD does not allow for cooperative interactions or enzyme flexibility. To generate more valid docking results, hydrogen bonding and distance constraints were used to restrict the substrate to β-strand-like conformations that are more biologically relevant (4, 5).

The crystal structure of WNV protease suggests that in addition to substrate-binding residues within NS3, residues within the NS2B cofactor also interact with substrate. Molecular docking (Fig. 2*A*) suggests that the P3 Lys side chain does not occupy a well defined S3 pocket in the enzyme but instead is largely solvent-exposed and binds in a shallow groove extending toward S1. This hydrophobic region is perhaps the reason that the four endogenous cleavage sequences contain a range of residues at P3 (Arg, Thr, Gly, and Leu). In all cases, there are hydrophobic elements in proximity to the main chain that are able to interact with the hydrophobic wall of the groove (*e.g.* Ile-155) as well as being able to accommodate both charged and polar side-chain termini directed outward into solvent. Asn-152 is hypothesized to be the S2 hydrogen bond acceptor of the P2 Lys side chain (7). On the opposite side of the substrate binding cleft to S2 is a hydrophobic surface patch consisting of Val-154, Met-156, and cofactor residue Leu-87. The hydrophobicity of these residues is highly conserved within the *Flavivirus* genus, and they most likely constitute one side of the shallow S4 pocket. The cofactor residue Gln-86 is observed in the crystal structure to participate in a hydrogen bond with the P3 Lys-NH_3^+ of an aldehyde inhibitor bound in an unusual conformation, but, since this residue is poorly conserved among the flaviviruses, the side chain is probably not important for substrate binding. Farther away, the cofactor residue Val-75 may make interactions with P5 or P6. However, it is difficult to predict how substrate might extend into and bind at this position. Also of interest is the cofactor residue Asn-84, which appears to make a hydrogen bond with the P2 Lys. This residue is semi-conserved within the *Flavivirus* genus as either a polar or negatively charged residue (Asn, Ser, Thr, Asp, and Glu).

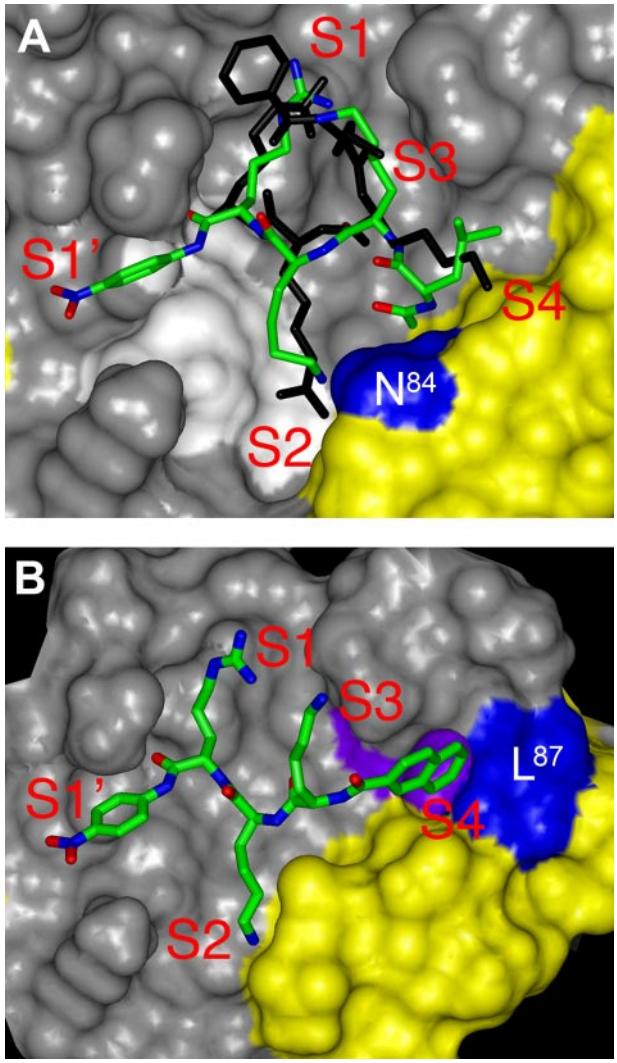

FIGURE 2. **Docking of substrates into the WNV NS2B/NS3 crystal structure.** *A*, active site of WNV NS2B/NS3 crystal structure (Protein Data Bank code 2fp7). The Connolly surface of NS3 is in *gray*, with the catalytic triad in *white* and the cofactor NS2B in *yellow*; asparagine 84 of the cofactor is shown in *blue*. The docked extended conformation of Ac-LKKR-*p*NA is represented by the *green stick structure* with nitrogen and oxygen atoms *colored blue* and *red*, respectively. The crystallographically determined tetrapeptide inhibitor benzoyl-(Nle)KRR-H is shown in *black*. *B*, docked and minimized extended conformation of naphthoyl-capped tripeptide substrate 2-naphthoyl-KKR-*p*NA. The Connolly surface of NS3 is again in *gray*, with the cofactor NS2B in *yellow*. Leucine 87 of the cofactor NS2B is shown in *blue*, and valine 154 of NS3 is shown in *purple*.

TABLE 1

Flavivirus cleavage sequences and cofactor homology

Mosquito-borne flaviviruses include West Nile virus (WNV), St. Louis encephalitis (SLEV), dengue virus subtypes 1–4 (Den1, Den2, Den3, Den4), yellow fever virus (YF), Japanese encephalitis virus (JE), Murray Valley encephalitis (MVE), Zika virus (ZIKV), and Bussuquara virus (BSQV). Tick-borne flaviviruses include Langat virus (LGTV) and tick-borne encephalitis virus (TBE). The first column shows the alignment of the region of NS2B involved in substrate binding, whereas the next four columns show the native flavivirus cleavage sequences (P4–P1′). The final column shows the degree of homology between the various flaviviruses and the WNV NS2B 40-amino acid cofactor domain and the NS3 protease domain. Residues shown in green and yellow designate homology. The residue shown in boldface type and designated by the asterisk is believed to interact with P2. A P2 arginine residue is shown in blue, and a P2 lysine is shown in red.

	NS2B	2A/2B	2B/3	3/4A	4B/5	Homology to WNV	
WNV	^{83}G**N**FQL^{87}	^{P4}NRKR$\downarrow$G$^{P1'}$	^{P4}YTKR$\downarrow$G$^{P1'}$	^{P4}SGKR$\downarrow$S$^{P1'}$	^{P4}GLKR$\downarrow$G$^{P1'}$		
SLEV	^{83}G**N**FKL^{87}	^{P4}NGKR$\downarrow$S$^{P1'}$	^{P4}HSKR$\downarrow$G$^{P1'}$	^{P4}AGKR$\downarrow$S$^{P1'}$	^{P4}KGKR$\downarrow$G$^{P1'}$	76.3%	
Den 1	^{82}G**T**MKI^{86}	^{P4}WGRK$\downarrow$S$^{P1'}$	^{P4}KKQR$\downarrow$S$^{P1'}$	^{P4}AGRF$\downarrow$S$^{P1'}$	^{P4}GGRR$\downarrow$G$^{P1'}$	67.4%	
Den 2	^{82}G**S**MSI^{86}	^{P4}SKKR$\downarrow$S$^{P1'}$	^{P4}KKQR$\downarrow$A$^{P1'}$	^{P4}AGFK$\downarrow$S$^{P1'}$	^{P4}NTRR$\downarrow$G$^{P1'}$	66.5%	
Den 3	^{82}G**T**MRI^{86}	^{P4}LKRR$\downarrow$S$^{P1'}$	^{P4}QTQR$\downarrow$S$^{P1'}$	^{P4}AGFK$\downarrow$S$^{P1'}$	^{P4}TGKR$\downarrow$G$^{P1'}$	63.9%	
Den 4	^{82}G**S**FSI^{86}	^{P4}ASRR$\downarrow$S$^{P1'}$	^{P4}KTQR$\downarrow$S$^{P1'}$	^{P4}SGFK$\downarrow$S$^{P1'}$	^{P4}TPRR$\downarrow$G$^{P1'}$	67.9%	Mosquito-borne Flaviviruses
YF	^{82}G**E**FKL^{86}	^{P4}FGRR$\downarrow$S$^{P1'}$	^{P4}GARR$\downarrow$S$^{P1'}$	^{P4}EGFR$\downarrow$G$^{P1'}$	^{P4}TGRR$\downarrow$G$^{P1'}$	64.8%	
JE	^{83}G**D**FHL^{87}	^{P4}NKKR$\downarrow$G$^{P1'}$	^{P4}TTKR$\downarrow$G$^{P1'}$	^{P4}AGKR$\downarrow$S$^{P1'}$	^{P4}SLKR$\downarrow$G$^{P1'}$	83.9%	
MVE	^{83}G**D**FHL^{87}	^{P4}NKKR$\downarrow$G$^{P1'}$	^{P4}YTKR$\downarrow$G$^{P1'}$	^{P4}AGKR$\downarrow$S$^{P1'}$	^{P4}AFKR$\downarrow$G$^{P1'}$	85.3%	
ZIKV	^{82}G**D**FSL^{86}	^{P4}SGKR$\downarrow$S$^{P1'}$	^{P4}TGKR$\downarrow$S$^{P1'}$	^{P4}AGKR$\downarrow$G$^{P1'}$	^{P4}VKRR$\downarrow$G$^{P1'}$	75.5%	
BSQV	^{83}G**D**FSL^{87}	^{P4}HGKR$\downarrow$S$^{P1'}$	^{P4}SGRR$\downarrow$A$^{P1'}$	^{P4}GGFR$\downarrow$S$^{P1'}$	^{P4}TGKR$\downarrow$G$^{P1'}$	67.8%	
LGTV	^{81}G**N**LHL^{85}	^{P4}RGRR$\downarrow$S$^{P1'}$	^{P4}SPRR$\downarrow$T$^{P1'}$	^{P4}SGRF$\downarrow$S$^{P1'}$	^{P4}GTRR$\downarrow$G$^{P1'}$	49.4%	Tick-borne Flaviviruses
TBE	^{81}G**N**FHL^{85}	^{P4}RGRR$\downarrow$S$^{P1'}$	^{P4}SSRR$\downarrow$S$^{P1'}$	^{P4}SGFR$\downarrow$S$^{P1'}$	^{P4}GGRR$\downarrow$G$^{P1'}$	48.8%	

*

Docking of 2-naphthoyl-KKR-*p*NA generally resulted in two docking poses. One had the bulky naphthoyl residue in the S1 pocket, resulting in a nonextended conformation reminiscent of the turnlike structure of the inhibitor aldehyde in the published crystal structure. The other had the aromatic residues interacting with hydrophobic enzyme residues at S4 but with some resultant steric clashes. Docking with GOLD used an explicit rigid protein method that is frequently inadequate for proteases, since they often display a high degree of active site plasticity (25). In this case, we took the docked poses where the 2-naphthoyl residue occupied the conventional extended conformation in the S4 site and used a combination of molecular dynamics and energy minimization to investigate possible induced fit binding modes. Fig. 2*B* shows one minimized docking pose where the small S4 pocket has been enlarged by a subtle movement of the side chains of Ile-155 and Val-154 in NS3 and Val-75, Val-77, and Leu-87 of NS2B. After these movements, the now slightly deeper S4 pocket is additionally defined by two residues (Phe-116 from NS3 and Phe-85 from NS2B) that make favorable aromatic-aromatic interactions with the naphthoyl ring.

Regarding the orientation of the substrate in the active site of the enzyme, the model suggests that the carbonyl carbon of the substrate scissile amide was 2.5–2.8 Å from the catalytic serine hydroxyl and in an orientation reminiscent of a Michaelis complex, despite no explicit restraints being used to fix it.

Following these substrate-docking modeling experiments, we prepared a number of site-directed mutants with residue substitutions in both NS3 and NS2B to test the predicted interactions. In parallel, we synthesized a library of chromogenic *p*NA substrates, designed around the optimal substrate used for our docking studies, and examined their processing kinetics by both wild type and mutant NS2B/NS3 West Nile Virus proteases.

Cofactor-Substrate Correlations—The crystal structure of WNV NS2B/NS3 protease revealed that Asn-84 of the NS2B cofactor is within hydrogen bonding distance of the P2 Lys of the bound ligand. Asn-84 is located within a highly conserved region of the cofactor, the Gly residue on the N-terminal side is completely conserved, and the third residue on the C-terminal side is a highly conserved hydrophobic Leu or Ile. Although the residue homologous with NS2B Asn-84 in other flaviviruses is variable, it is always polar or negatively charged (Table 1). This is of particular interest, because there appears to be an association between this residue and either Lys or Arg at P2 in native cleavage sequences. An Asn residue is at this position in WNV and St. Louis encephalitis (SLEV) proteases, corresponding to a preferred Lys at P2 of native substrates. However, in the proteases of all four serotypes of dengue, there is either Ser or Thr at this position matched by Arg at P2 in native substrates. The presence of Gln at P2 for one of the crucial cleavage sites (NS2B-NS3) suggests a requirement for a hydrogen-bonding pair and not a charge-charge interaction pair.

Specific partnering between cofactor and substrate is also seen when a negatively charged residue is present in the homologous position of the cofactor. In YF, a Glu at this position in the cofactor always corresponds with Arg at P2 in the native cleavage site. In JE, Murray Valley encephalitis (MVE), Zika virus (ZIKV), and Bussuquara virus (BSQV), an Asp is matched by a P2 Lys predominantly in the native cleavage site (Table 1). On the other hand, proteases of flaviviruses that are transmitted by ticks, tick-borne encephalitis (TBE) and Langat virus (LGTV), do not show the same preference as WNV and St. Louis encephalitis proteases. Despite Asn at the homologous position in the cofactor for TBE and LGTV proteases, in all cases they recognize Arg at P2 in the cleavage sites of their native substrates, compared with Lys at P2 in substrates for SLEV and WNV proteases. The tick-borne flaviviruses form a

TABLE 2

Enzyme kinetics of substrates modified at P4, P3, and P2

2-Naph, 2-naphthoyl; 3Pya, 3-pyridylalanine; 4Pya, 4-pyridylalanine; nPhe, *p*-nitrophenylalanine; aPhe, *p*-aminophenylalanine. Enzyme kinetic data were obtained using the *in vitro* enzyme assay, and each data point represents the mean of triplicate measurements ± S.E.

Substrate	K_m	k_{cat}	k_{cat}/K_m
	μM	s^{-1}	$M^{-1}\,s^{-1}$
P4 variants			
1. Ac-Leu-Lys-Lys-Arg-*p*NA	54.6 ± 5.1	1.64 ± 0.04	30,117 ± 2248
2. Ac-Nle-Lys-Lys-Arg-*p*NA	80.7 ± 6.7	2.35 ± 0.06	29,091 ± 1831
3. Ac-Nva-Lys-Lys-Arg-*p*NA	67.3 ± 6.8	2.36 ± 0.08	34,997 ± 2696
4. Ac-Chg-Lys-Lys-Arg-*p*NA	81.9 ± 8.8	1.78 ± 0.06	21,735 ± 1799
5. Ac-Cha-Lys-Lys-Arg-*p*NA	69.9 ± 8.1	2.28 ± 0.04	32,599 ± 2920
6. Ac-Aib-Lys-Lys-Arg-*p*NA	79.7 ± 6.2	1.60 ± 0.04	20,050 ± 1186
7. Ac-Tbg-Lys-Lys-Arg-*p*NA	134.2 ± 8.4	1.74 ± 0.04	12961 ± 590
8. Ac-Aoc-Lys-Lys-Arg-*p*NA	75.5 ± 10.5	2.51 ± 0.11	33,188 ± 3582
9. 2-Naph-Lys-Lys-Arg-*p*NA	25.4 ± 4.4	1.08 ± 0.06	42,603 ± 5472
P3 variants			
10. Ac-Leu-3Pya-Lys-Arg-*p*NA	88.7 ± 9.4	0.50 ± 0.02	5601 ± 419
11. Ac-Leu-4Pya-Lys-Arg-*p*NA	254.8 ± 19.5	0.90 ± 0.03	3536 ± 178
12. Ac-Leu-hArg-Lys-Arg-*p*NA	62.5 ± 5.8	1.18 ± 0.03	18,960 ± 1374
13. Ac-Leu-nPhe-Lys-Arg-*p*NA	254.1 ± 21.0	0.72 ± 0.03	2817 ± 151
14. Ac-Leu-Cit-Lys-Arg-*p*NA	208.6 ± 16.7	1.44 ± 0.05	6904 ± 372
15. Ac-Leu-Orn-Lys-Arg-*p*NA	105.5 ± 7.7	1.79 ± 0.05	16,999 ± 841
16. Ac-Leu-aPhe-Lys-Arg-*p*NA	141.3 ± 13.8	0.75 ± 0.03	5306 ± 369
17. 2-Naph-Cit-Lys-Arg-*p*NA	21 ± 4.0	0.32 ± 0.02	15,148 ± 2168
18. 2-Naph-Orn-Lys-Arg-*p*NA	38 ± 5.7	0.98 ± 0.04	25,743 ± 3184
19. 2-Naph-3Pya-Lys-Arg-*p*NA	65.8 ± 5.1	0.96 ± 0.03	14617 ± 731
20. 2-Naph-hArg-Lys-Arg-*p*NA	48 ± 11	0.11 ± 0.01	2294 ± 362
P2 variants			
21. 2-Naph-Lys-Cit-Arg-*p*NA	3423 ± 1306	0.43 ± 0.09	127 ± 23
22. 2-Naph-Lys-Orn-Arg-*p*NA	36.8 ± 3.8	1.58 ± 0.05	42,889 ± 3152
23. 2-Naph-Lys-Arg-Arg-*p*NA	43.4 ± 4.6	0.89 ± 0.04	20,382 ± 1338
24. 2-Naph-Lys-hArg-Arg-*p*NA	26.7 ± 3.7	0.59 ± 0.03	22,066 ± 1953
25. 2-Naph-Lys-AcLys-Arg-*p*NA	2303 ± 544	0.30 ± 0.03	131 ± 17
26. 2-Naph-Lys-aPhe-Arg-*p*NA	No activity		
27. 2-Naph-Lys-nLeu-Arg-*p*NA	No activity		
28. 2-Naph-Lys-Phe-Arg-*p*NA	No activity		
29. 2-Naph-Orn-Orn-Arg-*p*NA	22.9 ± 3.4	0.27 ± 0.01	11,899 ± 1367
30. 2-Naph-Cit-Orn-Arg-*p*NA	100.7 ± 8.4	0.57 ± 0.02	5612 ± 260

cluster distinct from the mosquito-borne flaviviruses, and there are significant amino acid changes between the two subgroups. Therefore it is likely that there are other significant differences within the substrate binding cleft that may account for the preference of a P2 Arg in substrates. To investigate the role of NS2B-84, various mutant proteases of WNV were produced with this residue altered to homologous residues from other flaviviruses (see below).

Substrate Design and Kinetic Processing—Kinetic parameters for proteolytic processing of modified substrates by WNV NS2B-NS3 protease are shown in Table 2. Initially, our hypothesis was that hydrophobic residues at P4 could interact with the surface-exposed hydrophobic region bounded by Val-154 and Met-156 of NS3 and Leu-87 of NS2B. We therefore prepared substrates based on Ac-*X*KKR-*p*NA, where *X* represents a hydrophobic unnatural amino acid (Table 2, entries 1–9, and Fig. 3). Most of the examined substrates had K_m and k_{cat} values comparable with those of Ac-LKKR-pNA, with a β-branched residue *t*-butylglycine that induces steric crowding on the peptide main chain being most detrimental to binding and processing. Substrate processing was least efficient for this and similar β-branched residues like aminoisobutyric acid and cyclohexylglycine (Table 2, entries 3, 9, 10). The substrate containing 2-amino-octanoic acid Aoc gave the highest k_{cat} of any substrate in this study, but this was tempered by a slightly lower affinity (K_m). Replacing the flexible Ac-P4 moiety with a rigid 2-naphthoyl group resulted in the highest affinity substrate (K_m = 25 μM) with a slightly lower k_{cat} but the most efficient processing (k_{cat}/K_m = 42,603 $M^{-1}\,s^{-1}$). Taken together, these results confirm the hypothesis that hydrophobic interactions are the key to interactions between P4 of substrates and S4 of enzyme.

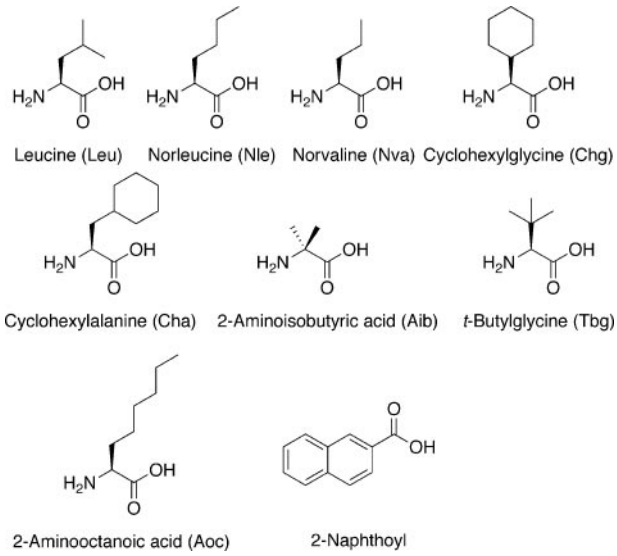

FIGURE 3. **Natural and unnatural hydrophobic amino acids used in P4 mutant substrate synthesis.**

The observation that substrates bearing the bulky 2-naphthoyl group at P4 possessed enhanced K_m but lower k_{cat} values may be attributable to an induced fit at P4/S4. As found in the

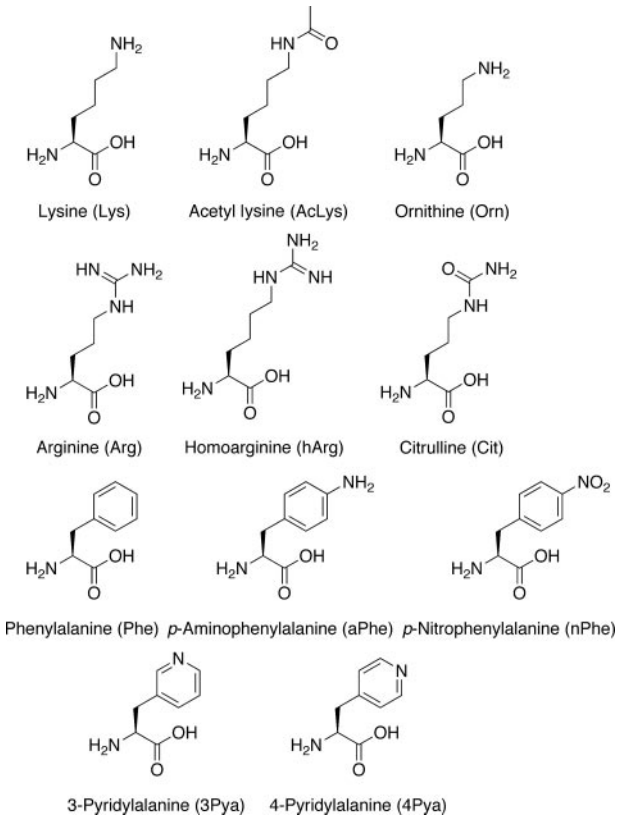

FIGURE 4. **Natural and unnatural amino acids used in P1, P2, and P3 mutant substrate synthesis.**

docking studies on this class of substrate, allowing the enzyme side chains to move produced better binding modes at this site. However, this requirement may contribute to lower overall catalytic efficiency by causing a repositioning of the scissile bond relative to the catalytic machinery, or alternatively, if the enzyme moves back into a "native" conformation after substrate hydrolysis, it may affect the departure of the N-terminal cleavage fragment from the active site (k_{off}).

We next examined whether nonproteinogenic amino acids could be incorporated at P3 (Table 2, entries 10–20). In light of the finding that 2-naphthoyl was a higher affinity replacement for Ac-Leu, a number of P3 mutants were examined with both Ac-Leu and 2-naphthoyl at P4. Interestingly, all of the changes made in this panel of substrates (Fig. 4 and Table 2, entries 10–19) resulted in lower overall k_{cat}/K_m values, primarily due to large reductions in k_{cat}. Substitution of 2-naphthoyl for Ac-Leu at P4 contributed a 10-fold improvement in K_m for substrates containing citrulline at P3, 3-fold for ornithine, 2-fold for homoarginine, but negligible improvement for 3-pyridylalanine. These gains, however, came at a cost of a 4-fold reduction in k_{cat} for citrulline, 2-fold for ornithine, 10-fold for homoarginine, but surprisingly a 2-fold gain in k_{cat} for 3-pyridylalanine. It therefore appears that a protonated amine is not absolutely required for affinity at P3, since the citrulline mutant substrate showed the best K_m. However, this required the presence of 2-naphthoyl at P4 for a cooperative effect. The effect of lengthening the spacer between the positive charged Arg and the main chain (hArg) was minor when P4 was Ac-Leu but detrimental when 2-naphthoyl capped the substrate. This clearly indicates that the naphthoyl substituent, possibly due to induced fit effects, had a major effect on the mode of binding of the substrates at the neighboring S3 enzyme subsite.

For the study of P2 mutant substrates (Table 2, entries 21–30), 2-naphthoyl was invariant at P4, and P3 was initially Lys. As for dengue protease (21), the presence of a basic amine at this position was found to be essential, with aromatic amines, capped amines, and hydrophobic residues not being tolerated. Only ornithine, with its high affinity and good k_{cat}, was a satisfactory replacement for Lys, although homoarginine with the longer side chain before the positive charge, had comparable activity, with an increase in K_m and a decrease in k_{cat}. The effects of cooperative changes were then examined using ornithine at P2 and either citrulline or ornithine at P3. As predicted, there was no cooperative gain in substrate fitness by these double mutations. The Orn-Orn mutant possessed a good K_m but had a further reduction in k_{cat}, with Cit-Orn conferring inferior K_m and k_{cat}. It is interesting to note that WNV protease processed substrates with a P2 Lys over 2-fold more efficiently than those with a P2 Arg.

Kinetic Studies of Mutant WNV NS3 and NS2B—To test the importance of specific enzyme residues on catalytic activity, we constructed 12 recombinant proteases, incorporating mutations of NS3 (T111F, T111L, D129A, D129E, V154F, V154L, I155F, M156A, I162F, A164S, A164V, and V166L) and 12 additional recombinant proteases incorporating mutations in cofactor NS2B (V75A, V75F, N84A, N84D, N84E, N84L, N84S, Q86A, Q86E, Q86L, L87A, and L87F). These 24 mutants were generated by site-directed mutagenesis of a parent construct WNCF40GlyNS3pro. We have previously shown that expression of recombinant protease WNCF40GlyNS3pro yields a soluble protein that can be purified in high concentration (7). The purified mutant proteases were run on 12% SDS-PAGE (not shown) to confirm expression and purification of the desired 27-kDa product (7).

We first evaluated the kinetics of substrate processing by WNV protease mutants with tetrapeptide and hexapeptide substrates. The 24 mutant proteases were tested against the optimized tetrapeptide substrate Ac-LKKR-*p*NA, and 10 were also tested against the hexapeptide substrate Ac-LQYTKR-*p*NA based on the native NS2B/NS3 cleavage site. Kinetic parameters (Table 3) for these mutant proteases showed that most had impaired function compared with wild type enzyme, with only the conservative NS2B-L87F and polar charged NS2B-Q86E mutants retaining comparable catalytic efficiency. In particular, it is notable that NS2B-Q86 participates in a hydrogen bond with the charged P3 Lys of inhibitor in the crystal structure, suggesting that the Q86E mutant might enhance this interaction. However, K_m is not significantly higher than the wild type, and the overall reduction in processing efficiency is due to a reduction in k_{cat}. This may be due to a reduction in k_{off} or alternatively due to repositioning of the scissile bond farther from the catalytic machinery. All WNV protease mutants also exhibited variable decreases in substrate affinity. The effects and their implications for substrate-binding interactions are discussed below.

We also examined the kinetics of processing by WNV protease mutants on the truncated N-capped tripeptide substrate

TABLE 3

Site-directed mutant enzyme kinetics

Enzyme kinetics were obtained using the *in vitro* enzyme assay and Ac-LKKR-*p*NA and Ac-LQYTKR-*p*NA as substrates with each data point representing the mean of triplicate measurements ± S.E.

Mutant	AcLKKR-*p*NA			AcLQYTKR-*p*NA		
	K_m	k_{cat}	k_{cat}/K_m	K_m	k_{cat}	k_{cat}/K_m
	μM	s^{-1}	$M^{-1}\,s^{-1}$	μM	s^{-1}	$M^{-1}\,s^{-1}$
Wild type	54.6 ± 5.1	1.644 ± 0.041	30,117 ± 2248	199 ± 22	0.266 ± 0.007	1332 ± 118
NS3						
T111F	169 ± 13	0.198 ± 0.005	1176 ± 65	376 ± 18	0.051 ± 0.001	135 ± 4
T111L	214 ± 27	0.060 ± 0.003	279 ± 25	617 ± 65	0.018 ± 0.001	30 ± 2
D129A	635 ± 49	0.048 ± 0.002	75 ± 4			
D129E	432 ± 18	0.036 ± 0.001	83 ± 2			
V154F	263 ± 39	0.057 ± 0.003	216 ± 21			
V154L	190 ± 13	0.800 ± 0.018	4223 ± 210	344 ± 28	0.245 ± 0.007	712 ± 40
I155F	207 ± 17	0.733 ± 0.021	3532 ± 196			
M156A	164 ± 10	0.778 ± 0.015	4741 ± 210	879 ± 53	0.314 ± 0.010	357 ± 11
I162F	268 ± 28	0.164 ± 0.006	612 ± 43	416 ± 40	0.040 ± 0.002	95 ± 6
A164S	221 ± 23	0.260 ± 0.009	1178 ± 85			
A164V	331 ± 15	0.181 ± 0.003	546 ± 16			
V166L	96 ± 6	0.638 ± 0.009	6676 ± 309	232 ± 13	0.297 ± 0.005	1282 ± 52
NS2B						
V75A	199 ± 10	0.173 ± 0.003	866 ± 30	5318 ± 507	0.164 ± 0.011	31 ± 1
V75F	272 ± 42	0.246 ± 0.017	905 ± 84	809 ± 105	0.091 ± 0.006	113 ± 8
N84A	376 ± 44	0.126 ± 0.007	336 ± 22			
N84D	117 ± 5	1.532 ± 0.019	13,109 ± 393			
N84E	79.3 ± 4.4	3.027 ± 0.047	38,188 ± 1660			
N84L	220 ± 11	0.34 ± 0.006	1543 ± 52			
N84S	205 ± 16	0.682 ± 0.019	3333 ± 183			
Q86A	258 ± 18	0.367 ± 0.011	1424 ± 61	1203 ± 182	0.127 ± 0.011	105 ± 7
Q86E	77.7 ± 7.3	0.868 ± 0.018	11,182 ± 863			
Q86L	148 ± 12	0.973 ± 0.024	6586 ± 381			
L87A	434 ± 39	0.070 ± 0.003	161 ± 8			
L87F	144 ± 12	1.342 ± 0.033	9345 ± 589	764 ± 53	0.228 ± 0.008	299 ± 11

TABLE 4

Site-directed mutant enzyme kinetics against 2-naphthoyl-Lys-Lys-Arg-*p*NA

Enzyme kinetics were obtained using the *in vitro* enzyme assay against the substrate 2-naphthoyl-KKR-*p*NA, with each data point representing the mean of triplicate measurements ± S.E.

Mutant	2-naphthoyl-KKR-*p*NA		
	K_m	k_{cat}	k_{cat}/K_m
	μM	s^{-1}	$M^{-1}\,s^{-1}$
Wild type	25.4 ± 4.4	1.083 ± 0.055	42,603 ± 5472
NS3-V154F	97.7 ± 6.9	0.0431 ± 0.0013	442 ± 19
NS2B-L87A	88.4 ± 16.0	0.0478 ± 0.0030	541 ± 70
NS2B-L87F	36.9 ± 3.6	0.288 ± 0.008	7796 ± 580

2-naphthoyl-KKR-*p*NA (Table 4). Docking of the substrate 2-naphthoyl-KKR-*p*NA into the crystal structure of WNV protease yielded two possible substrate conformations. The β-strand substrate conformation orients the 2-naphthoyl group in a hydrophobic S4 subsite in the enzyme, bounded by NS3 residues Val-154 and Ile-155 and the NS2B cofactor residues Val-75, Val-77, and Leu-87, whereas a turn-like conformation positions the 2-naphthoyl substituent over the P1 arginine of the substrate. To identify which of these is more likely, we tested one relevant NS3 mutant (V154F) and two separate relevant NS2B mutants (L87A and L87F). The V154F and L87A mutations caused similarly dramatic reductions in k_{cat} and minor increases in K_m, whereas L87F resulted in a similar K_m and only a 3-fold decrease in k_{cat}. Although these results show that the V154F and L87A mutations have a large effect on processing of substrates with a naphthoyl at P4, those substrates are still relatively efficiently processed by the L87F mutant protease, suggesting that Phe-87 may make a π-π interaction that stabilizes naphthoyl binding at S4.

To investigate the hypothesis that the residue at NS2B-84 contributes to substrate binding and provides flavivirus NS3/NS2B proteases with specificity for either Lys or Arg at P2, the wild type WNV recombinant protease and the NS2B-N84D/N84E/N84S mutant proteases were tested against the tetrapeptide substrates Ac-LK**K**R-*p*NA and Ac-LK**R**R-*p*NA (Table 5). The comparative enzyme kinetic parameters show that the N84S mutation resulted in a 4-fold decrease in K_m when P2 was Lys and a 2-fold increase in K_m when P2 was Arg. This supports observations that Asn-84 (native to WNV) binds more strongly to a P2 Lys, whereas Ser-84 (native to Den) binds more strongly to a P2 Arg. Against both substrates the N84S mutation caused a decrease in catalytic efficiency, ~2-fold against Ac-LKKR-*p*NA and as much as 15-fold against Ac-LKRR-*p*NA, translating in both cases to about an 8-fold decrease in catalytic efficiency (k_{cat}/K_m) *versus* wild type enzyme.

Similarly, the substitution of N84D (native to JE, MVE, ZIKV, and BSQV) produced a mutant protease with a 2-fold higher K_m for Ac-LKKR-*p*NA and a 2–3-fold lower K_m for Ac-LKRR-*p*NA, whereas the k_{cat} was unaffected for the former substrate but about 6-fold lower for the latter substrate. The overall effect was a 2-fold decrease in catalytic efficiency for both substrates following this enzyme mutation. Comparison between the substrates reveals an almost 4-fold higher k_{cat} and a 2-fold higher catalytic efficiency, k_{cat}/K_m for Ac-LKKR-*p*NA compared with Ac-LKRR-*p*NA, indicating a preference for a P2 Lys (predominantly present in native cleavage sequences for JE, MVE, ZIKV, and BSQV).

The N84E mutation (native to YF) produced a mutant protease with a slightly higher catalytic efficiency for a P2 Arg (present in YF native cleavage sequences). However, the

TABLE 5

Comparison of wild type, NS2B-N84S, NS2B-N84D, and NS2B-N84E enzyme kinetics

Enzyme kinetics were obtained using the *in vitro* enzyme assay against the Ac-LKKR-*p*NA, Ac-LKRR-*p*NA, 2-naphthoyl-KRR-*p*NA, and 2-naphthoyl-KhRR-*p*NA substrates. Each data point represents the mean of triplicate measurements ± S.E.

Mutant	Ac-LKKR-*p*NA			Ac-LKRR-*p*NA		
	K_m	k_{cat}	k_{cat}/K_m	K_m	k_{cat}	k_{cat}/K_m
	μM	s^{-1}	$M^{-1}\,s^{-1}$	μM	s^{-1}	$M^{-1}\,s^{-1}$
Wild type	54.6 ± 5.1	1.644 ± 0.041	30,117 ± 2248	195 ± 13	2.494 ± 0.053	12,783 ± 596
NS2B-N84S	205 ± 16	0.682 ± 0.019	3333 ± 183	100 ± 8	0.158 ± 0.003	1577 ± 104
NS2B-N84D	117 ± 5	1.532 ± 0.019	13,109 ± 393	69.8 ± 3.3	0.398 ± 0.005	5698 ± 215
NS2B-N84E	79.3 ± 4.4	3.027 ± 0.047	38,188 ± 1660	66.4 ± 4.9	2.707 ± 0.053	40,796 ± 2417
Mutant	**2-naphthoyl-KRR-*p*NA**			**2-naphthoyl-KhRR-*p*NA**		
	K_m	k_{cat}	k_{cat}/K_m	K_m	k_{cat}	k_{cat}/K_m
	μM	s^{-1}	$M^{-1}\,s^{-1}$	μM	s^{-1}	$M^{-1}\,s^{-1}$
Wild type	43.4 ± 4.6	0.885 ± 0.040	20,382 ± 1338	26.7 ± 3.7	0.59 ± 0.033	22,066 ± 1953
NS2B-N84S	19.7 ± 5.0	0.151 ± 0.040	7695 ± 1804	29.5 ± 4.1	0.027 ± 0.001	904 ± 93

trend was not the same when comparing substrate affinity, since both mutants had slightly higher affinity for the Ac-LKRR-*p*NA substrate.

The N84S mutant protease was also tested against the truncated substrates 2-naphthoyl-KRR-*p*NA and 2-naphthoyl-K(hR)R-*p*NA in order to examine whether a P2 homoarginine residue could possibly be recognized with similar efficiency by either Asn-84 or Ser-84. Both the wild type protease and the N84S mutant had very similar K_m values when assayed against the 2-naphthoyl-K(hR)R-*p*NA, suggesting that a P2 homoarginine can bind to Asn-84 or Ser-84 with similar affinity; however, the catalytic efficiency was decreased by over 20-fold. Taken together, these results support our earlier work (7, 26), which suggested that these WNV constructs of the NS2B/NS3 protease are more efficient enzymes than their dengue counterparts at processing their native sequences.

DISCUSSION

Substrate Modifications—One of the most common approaches to inhibitor development is substrate-based and begins with substrate optimization. Such approaches for hepatitis C virus NS3 protease have led to increases in IC_{50} values of over 1000-fold (27, 28). In this study, we began with a tetrapeptide substrate and systematically replaced P4, P3, and P2 amino acids with peptidic and nonpeptidic side chains designed to map the individual substrate-binding subsites in the protease, to test the importance of specific enzyme residues that line the substrate-binding groove, and to enhance substrate affinity and catalytic efficiency. These solution studies have the potential to complement structural information provided from solid state crystal structures of the enzyme, especially providing new information on the binding of substrate in an extended conformation and induced fit effects, information that is not available from crystal structures.

A significant improvement in ligand affinity came from replacements of the "P5" acetyl cap and P4 hydrophobic Leu with the more rigid, nonpeptidic, 2-naphthoyl group. This change enhanced substrate affinity, with K_m increasing over 2-fold. Docking studies suggested that the bulkier 2-naphthoyl group may insert deeper into the S4 subsite, with cooperative movement of both NS3 and NS2B amino acids for optimal ligand binding. It is likely that further modifications of this nonpeptidic P4 cap can produce even higher affinity substrate/inhibitor ligands. Incorporating a citrulline into P3 also improved substrate affinity but reduced k_{cat}. However, when next to a P4 Leu in the substrate, a P3 citrulline had a 10-fold lower K_m but a relatively high k_{cat}. From a drug design standpoint, the observation that citrulline-containing peptides had higher affinity than basic amines when 2-naphthoyl was present with poorer substrate turnover (k_{cat}), while having good turnover in the presence of Ac-Leu at P4, suggests that a double mutation to a nonpeptidic entry with no positive charges may be useful for the preparation of potent inhibitors. Removal of this positive charge increases drug-like characteristics, such as membrane permeability. Replacing positive charges at P2 and P1 has so far not been effective in flavivirus protease inhibitors.

The finding that WNV protease was twice as active against substrates with Lys rather than Arg at P2 contrasts with the four serotypes of dengue protease (21) that were more active against substrates with Arg instead of Lys at P2. This confirms predictions from native cleavage sequences and constitutes the first recognized difference in substrate preference between related flavivirus proteases. It was previously assumed, based on the high level of homology within NS3, that an inhibitor developed against one flaviviral protease could be active against multiple flaviviruses. This may still be possible if a replacement at P2 can be found that has a high affinity against multiple flavivirus proteases. The WNV protease had a substrate affinity (K_m) for a P2 homoarginine that was comparable with lysine, and a P2 ornithine also gave a relatively high K_m and k_{cat}. Neither of these residues have yet been tested in dengue protease substrates.

Correlation of Mutagenesis Results to Architecture of the Substrate-binding Cleft—Ile-155 in the crystal structure lies just beyond the rim of the S1 pocket, with docking of a substrate in the extended β-strand conformation suggesting that Ile-155 may contribute hydrophobic contacts to the methylene region of the substrate P3 Lys side chain. Mutation of this residue to a bulkier Phe could potentially affect binding in three ways. First, steric clashes may reduce substrate affinity, seen in the 4-fold reduction in K_m and a halving of k_{cat}. Second, extra bulk at the lip of the S1 pocket may interfere with entry of the critical P1 Arg of substrates into the S1 pocket, partially capping it. Third, it is possible that the aromatic Phe side chain participates in a π-cation interaction with the charged Lys side chain at P3. Experimental

evidence suggests that this interaction, if present at all, only slightly increases affinity.

Hydrophobic residues in NS3 (Val-154 and Met-156) and NS2B (Leu-87) in the crystal structure form a hydrophobic patch that probably constitutes S4. The hydrophobic character of these residues are conserved throughout the *Flavivirus* genus. Both WNV and dengue NS3 proteases have a preference for hydrophobic P4 residues. Substitution of each of these residues gave large losses in substrate affinity, NS3-M156A resulting in a 3-fold decrease in K_m, NS2B-L87A resulting in an 8-fold reduction, and even the conservative substitution NS3-V154L resulting in a 3-fold decrease in K_m. Substitution at this site with a bulky Phe had a large effect on substrate affinity; NS3-V154F gave a 5-fold decrease in K_m, whereas NS2B-L87F caused a 3-fold decrease. Each of these substitutions caused a large loss of catalytic activity except NS2B-L87F, which maintained a similar catalytic turnover. Modeling suggests that the L87F mutant directs the Phe side chain away from S4, whereas in the L87A mutant the small side chain remains directed into the pocket. This suggests that the S4 pocket is quite flexible and capable of induced fitting to various P4 substituents, but Leu may be optimal for efficient association of NS2B to NS3. Each of these mutants was also tested against the hexapeptide substrate Ac-LQYTKR-*p*NA and produced similar effects on substrate affinity and catalytic turnover, reflecting similar binding to the P4 Tyr as to the P4 Leu.

The Gln at NS2B-86 is within hydrogen bonding distance to the main chain between the P3 and P4 substrate residues. Gln-86 is relatively poorly conserved within the *Flavivirus* genus but is predominantly polar or positively charged (Glu, Ser, Lys, Arg, and His; Table 3), suggesting that it may make a hydrogen bond with the oxygen atom of the P4 backbone carbonyl. Substitution of this residue for Ala or Leu produced 5- and 3-fold losses of substrate affinity, respectively. However, substitution with a negatively charged Glu produced only a minor affect on substrate affinity, suggesting that this residue may participate in hydrogen bonding with the adjacent NS2B Asn-84. The Q86A and Q86L mutants would disrupt this interaction and destabilize the region, whereas Q86E would maintain it.

The NS2B Val-75 is located in a position that could potentially interact with either P5 or P6. A hydrophobic residue is completely conserved in this position within the *Flavivirus* genus. Substitution with Ala resulted in a 4-fold loss in K_m when assayed against the tetrapeptide Ac-LKKR-*p*NA but a large 27-fold loss in K_m when assayed against the hexapeptide Ac-LQYTKR-*p*NA, supporting the prediction that NS2B Val-75 makes an interaction with P5 or P6. Substitution with an aromatic Phe produced a comparable effect on both tetrapeptide and hexapeptide substrates, 5- and 4-fold losses in K_m and 7- and 3-fold losses in k_{cat}, respectively. It is possible that NS2B Val-75 is involved in a hydrophobic interaction between the cofactor and NS3 and contributes to the loss in activity and affinity.

There is also an interesting correlation between the mutagenesis results and the binding site for the 2-naphthoyl substituent. Although the docking of the substrate 2-naphthoyl-KKR-*p*NA into the crystal structure of WNV protease yielded two possible conformations for the substrate, kinetic analysis of this substrate against mutant proteases has suggested the β-strand substrate conformation to be most likely. This conformation orients the bulky 2-naphthoyl group by induced fit into a slightly enlarged hydrophobic P4 subsite. Since NS3-V154F and NS2B-L87A mutants retained good affinity for the substrate 2-naphthoyl-KKR-*p*NA but dramatically affect k_{cat}, there are possibly favorable hydrophobic, aromatic-aromatic or π-π interactions occurring. The decrease in k_{cat} is presumably due to the slow release (k_{off}) of the N-terminal cleavage product from the enzyme-product complex. It is unlikely that this naphthoyl residue is exerting its effect on k_{cat} by displacing the scissile amide bond from its optimal location due to its significant distance from the catalytic serine residue. When the important amine/basic side chains of the substrates are shortened or lengthened (Table 2, entries 10–30), modeling studies suggest that this causes a displacement of the scissile bond, the likely cause of the marked drop in k_{cat} for these substrates.

Substrate Specificity and Cofactor Residue Asn-84—The cofactor residue at NS2B-84 is associated with a preference for Lys or Arg at the substrate P2 position. For flaviviral proteases in which Asn or Asp is at NS2B-84, Lys is predominantly present in P2 of the native substrate. When Ser, Thr, or Glu is at NS2B-84, Arg is preferred at P2 (Table 1).

Various mutations of Asn-84 were tested to further analyze the effect of this residue on substrate specificity. N84S decreased affinity of a P2 Lys substrate by 4-fold but increased affinity of the Arg analogue 2-fold. This came at the expense of catalytic activity, which decreased by 2- or 20-fold, depending on the substrate. The altered substrate affinity may prevent dissociation of the cleavage products that inhibit activity, or alternatively substrate may bind in a slightly different position in the mutant enzyme, altering the position of the scissile bond and affecting catalytic activity.

N84D (native to JE, MVE, ZIKV, and BSQV) gave a higher catalytic efficiency against substrates containing a P2 Lys, whereas the N84E (native to YF) gave higher catalytic efficiency against substrates containing P2 Arg. Because there are subtle differences in the sizes and shapes of substrate binding pockets in flavivirus proteases, substitution of a single residue is unlikely to account for the divergence of specificity, and it is likely to be these additional differences that add to the specificity for either Lys or Arg at P2 in substrates.

Although it is unlikely that NS2B-84 contributes the only difference between the active sites, it does appear to be a major factor that will need to be addressed in order to develop a broad spectrum flavivirus protease inhibitor. As an initial attempt to identify compounds that can bind equally well to either Asn or Ser at NS2B-84, we found that the tetrapeptide substrate, 2-naphthoyl-KhRR-*p*NA, with a P2 homoarginine residue could bind with similar affinity to either the wild type WNV protease or the NS2B-N84S mutant (K_m = 26.7 ± 3.7 and 29.5 ± 4.1, respectively). This residue showed a high affinity that was similar to that for substrate with a P2 lysine binding to wild type enzyme (K_m = 25.4 ± 4.4), better than that for substrate with a P2 arginine binding to wild type enzyme (K_m = 43.4 ± 4.6) and only slightly worse than for a substrate with a P2

arginine binding to the mutant N84S protease (K_m = 19.7 ± 5.0). Although this substrate has yet to be tested against Den NS2B/NS3 protease, these results indicate that the P2 homoarginine is likely to bind well to the P2 pocket of both Den and WNV, and therefore it may be a good candidate for incorporation into broad spectrum flavivirus protease inhibitors.

Mutations outside the Substrate-binding Cleft—Some mutations were also made in the vicinity of the putative P3 and P4 substrate-binding sites as defined by the homology model. Within the resolution of the WNV NS2B/NS3 crystal structure, these mutated residues were some distance from the active site. Thr-111 is on the opposite side in the protease to the substrate-binding cleft but is within hydrogen bonding distance of the carbon backbone of the NS2B residue Thr-69. The T111L mutation caused a 4-fold reduction in K_m and a 27-fold reduction in k_{cat}, much greater than a T111F mutation (3-fold reduction in K_m, 8-fold reduction in k_{cat}). This suggests that the hydrogen bond contributed by Thr-111 is important for cofactor binding, and in the absence of this, the cofactor binds less efficiently to protease or in a way that affects catalytic activity and substrate affinity.

The NS3-I162F mutant produced a large reduction in K_m (5-fold) and k_{cat} (10-fold) against the tetrapeptide substrate but a lesser reduction against the hexapeptide substrate for both K_m (2-fold) and k_{cat} (5-fold). Ile-162 is located below the Val-154 and Met-156 residues, which are proposed to constitute the hydrophobic S4 pocket, and disruption of these residues by the I162F mutation is likely to affect substrate binding at S4. Both Ala-164 and Val-166 are in the center of the enzyme, and their substitution may affect folding of the enzyme. Mutation of A164S/A164V substantially reduced K_m (3- and 6-fold, respectively) and k_{cat} (6- and 9-fold, respectively). The more conservative substitution of V166L had only a minor effect on processing of both tetrapeptide and hexapeptide substrates. The largest effect was the 2.5-fold reduction in k_{cat} for Ac-LKKR-*p*NA.

CONCLUSION

This study has provided valuable new information about the architecture of the flavivirus NS3 protease in solution and about interactions made between peptide substrate analogues and the substrate-binding cleft of the enzyme. Together with solid state crystal structures of the WNV NS2B/NS3 protease, this new information helps to provide a firmer basis for rational drug design. The changes to tetrapeptide substrates described herein substantially increased enzyme affinity and provide important clues toward development of substrate-based inhibitors of flavivirus proteases.

REFERENCES

1. van der Meulen, K. M., Pensaert, M. B., and Nauwynck, H. J. (2005) *Arch Virol.* **150,** 637–657
2. Hayes, C. G. (2001) *Ann. N. Y. Acad. Sci.* **951,** 25–37
3. Brinton, M. A. (2002) *Annu. Rev. Microbiol.* **56,** 371–402
4. Tyndall, J. D., Nall, T., and Fairlie, D. P. (2005) *Chem. Rev.* **105,** 973–99
5. Loughlin, W. A., Tyndall, J. D. A., Glenn, M. P., and Fairlie, D. P. (2004) *Chem. Rev.* **104,** 6085–6117
6. D'Arcy, A., Chaillet, M., Schiering, N., Villard, F., Pheng Lim, S., Lefeuvre, P., and Erbel, P. (2006) *Acta Crystallogr. Sect. F* **62,** 157–162
7. Nall, T. A., Chappell, K. J., Stoermer, M. J., Fang, N. X., Tyndall, J. D., Young, P. R., and Fairlie, D. P. (2004) *J. Biol. Chem.* **279,** 48535–48542
8. Gorbalenya, A. E., Donchenko, A. P., Koonin, E. V., and Blinov, V. M. (1989) *Nucleic Acids Res.* **17,** 3889–3897
9. Wengler, G. (1991) *Virology* **184,** 707–715
10. Li, H., Clum, S., You, S., Ebner, K. E., and Padmanabhan, R. (1999) *J. Virol.* **73,** 3108–3116
11. Bazan, J. F., and Fletterick, R. J. (1989) *Virology* **171,** 637–639
12. Chambers, T. J., Hahn, C. S., Galler, R., and Rice, C. M. (1990) *Annu. Rev. Microbiol.* **44,** 649–688
13. Falgout, B., Miller, R. H., and Lai, C. J. (1993) *J. Virol.* **67,** 2034–42
14. Clum, S., Ebner, K. E., and Padmanabhan, R. (1997) *J. Biol. Chem.* **272,** 30715–23
15. Leung, D., Abbenante, G., and Fairlie, D. P. (2000) *J. Med. Chem.* **43,** 305–341
16. Abbenante, G., and Fairlie, D. P. (2005) *Med. Chem.* **1,** 71–104
17. Fairlie, D. P., Tyndall, J. D. A., Reid, R. C., Wong, A. K., Abbenante, G., Scanlon, M. J., March, D. R., Bergman, D. A., Chai, C. L. L., and Burkett, B. A. (2000) *J. Med. Chem.* **43,** 1271–1281
18. Murthy, H. M., Judge, K., DeLucas, L., and Padmanabhan, R. (2000) *J. Mol. Biol.* **301,** 759–767
19. Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996) *Cell* **87,** 343–55
20. Chappell, K. J., Nall, T. A., Stoermer, M. J., Fang, N. X., Tyndall, J. D., Fairlie, D. P., and Young, P. R. (2005) *J. Biol. Chem.* **280,** 2896–2903
21. Li, J., Lim, S. P., Beer, D., Patel, V., Wen, D., Tumanut, C., Tully, D. C., Williams, J. A., Jiricek, J., Priestle, J. P., Harris, J. L., and Vasudevan, S. G. (2005) *J. Biol. Chem.* **280,** 28766–28774
22. Abbenante, G., Leung, D., Bond, T., and Fairlie, D. P. (2001) *LIPS* **7,** 347–351
23. Jones, G., Willett, P., and Glen, R. C. (1995) *J. Mol. Biol.* **245,** 43–53
24. Erbel, P., Schiering, N., D'Arcy, A., Renatus, M., Kroemer, M., Lim, S. P., Yin, Z., Keller, T. H., Vasudevan, S. G., and Hommel, U. (2006) *Nat. Struct. Mol. Biol.* **13,** 372–373
25. Stoermer, M. J. (2006) *Med. Chem.* **2,** 89–112
26. 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
27. Ingallinella, P., Altamura, S., Bianchi, E., Taliani, M., Ingenito, R., Cortese, R., De Francesco, R., Steinkuhler, C., and Pessi, A. (1998) *Biochemistry* **37,** 8906–8914
28. Llinas-Brunet, M., Bailey, M. D., Ghiro, E., Gorys, V., Halmos, T., Poirier, M., Rancourt, J., and Goudreau, N. (2004) *J. Med. Chem.* **47,** 6584–6594

SUPPLEMENTARY MATERIALS

WEST NILE VIRUS NS3 PROTEASE: INSIGHTS INTO SUBSTRATE BINDING AND PROCESSING THROUGH COMBINED MODELLING, PROTEASE MUTAGENESIS, AND KINETIC STUDIES.

Keith J. Chappell[†]#, Martin J. Stoermer[§]#, David P. Fairlie[§], and Paul R. Young[†]
From the [†]School of Molecular and Microbial Sciences and [§]Centre for Drug Design and Development, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia, 4072.

Running title: Substrate processing by the West Nile virus protease

Address correspondence to:
Paul R Young, School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland, Australia 4072, Tel. 617 33654646; Fax. 617 3365 4620; E-Mail: p.young@uq.edu.au and/or
David P. Fairlie, Centre for Drug Design and Development, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia, 4072, Tel. 617 3346 2989; Fax. 617 3346 2101; E-Mail: d.fairlie@imb.uq.edu.au
# These authors contributed equally to this work.

Experimental procedures

General Synthetic methods - Fmoc-protected amino acids and resins were obtained from Auspep, Novabiochem, ChemImpex, NeoMPS, and PepChem. TFA, piperidine, DIPEA and DMF (peptide synthesis grade) were purchased from Auspep. HCTU was purchased from Iris Biotech. Oxone™ was purchased from Sigma-Aldrich. All other materials were reagent grade unless otherwise stated. Crude peptides were purified by reversed phase HPLC on Vydac C18, 10-15 μm, 300 Å, 50 x 250 mm) and Phenomenex Luna C18 (2.2 x 25 cm) columns, using a gradient mixture of solvents (A) 0.1%TFA/water and (B) 0.1%TFA/10%water/90%acetonitrile. Analytical HPLC was performed on a Waters system equipped with a 717 plus autosampler, 660 controller and a 996 photodiode array detector, using reverse-phase Phenomenex Luna C8 or C18, 5 μm, 100 Å, 250 x 4.6 mm columns. Purified peptides were characterised by analytical HPLC (linear gradient 0-100%B over 30 mins), mass spectrometry and NMR. The molecular weight of the peptides (1 mg/mL) was determined by electrospray mass spectrometry on a Micromass LCT mass spectrometer. ^{1}H NMR spectra were recorded on samples containing 4 mM peptide in DMSO-d_6 (550 μL) on Bruker Avance 600 spectrometer at 298 K. Proton resonance assignments were determined by TOCSY (80 ms mixing time), DQF-COSY, and NOESY (300 ms mixing time) spectra using the sequential assignment method (*S1*). All spectra were processed on Silicon Graphics R10000 or R12000 workstations using XWINNMR 2.1 (Bruker) under Irix or Topspin (Bruker) running under Windows (*S2*).

pNA Substrate Synthesis - *p*NA substrates were synthesised according to the general method of Abbenante (*S3*). Modifications to the general method were necessary for substrates containing *p*-aminoPhe, citrulline, pyridylalanine and 2-naphthoyl residues. *p*-Aminophe containing *p*-aminoanilides were cleaved from resin using 1% TFA in DCM and were oxidised to the corresponding *p*-nitroanilides using Oxone™ in water/acetonitrile mixtures with the side-chain protecting groups in place. Subsequent deprotection with TFA/water/TIPS generated the crude peptides. Citrulline, pyridylalanine and 2-Naphthoyl-containing substrates were observed to slowly oxidise with excess Oxone in solution.

Citrulline derivatives gave ornithine-containing by-products, pyridylalanines were converted to their corresponding N-oxides, and substrates capped with 2-naphthoic acid were observed to undergo partial oxidation of the naphthalene B ring to 1,4-naphthoquinone-6-carboxamide capped substrates. In all of these cases, reduction in the number of equivalents of Oxone™ to 2 and careful monitoring of the oxidation reaction by mass spectrometry reduced the incidence of side-products.

Ac(Nva)KKR-pNA. This substrate was synthesised on *p*-phenylenediamine-derivatised TCP resin *(28)* using Fmoc protocols. Washed and dried peptide on resin (0.25 mmol based on original resin loading) was cleaved using trifluoroacetic acid (9.5 mL), water (250 μL) and triisopropylsilane (250 μL) for 1 hour and the mixture was filtered, washing the resin with TFA (3x2 mL). The combined filtrates were evaporated under a stream of dry nitrogen gas, and the residue was partitioned between ether (20 mL) and water (20 mL). The aqueous extract containing the peptide *p*-aminoanilide was oxidised with Oxone™ (1.5 mmol) at room temperature with stirring overnight. The mixture was filtered through a 0.45 μM filter and purified by preparative reverse-phase HPLC (100% solvent A for 10 minutes to desalt the sample followed by a 0-100% B gradient over 20 minutes) followed by freeze drying of pure fractions to yield the substrate tris-trifluoroacetate salt as a white powder (161 mg, 62%). (C8 (analytical) R_t = 17.6 min.). ^{1}H nmr (DMSO-d_6) δ 10.67 (s, 1H, *p*NA-NH), 8.26 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.02 (d, 1H, NvaαNH), 8.00 (d, 1H, LysαNH), 7.86 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.70 (m, 6H, 2xLyszNH_3), 7.62 (t, 1H, ArgεNH), 7.5-6.5 (4H, m, 2xArgNH_2), 4.36 (m, 1H, ArgαCH), 4.25 (m, 1H, LysαCH), 4.22 (m, 1H,LysαCH), 4.17 (m, 1H, NvaαCH), 3.12 (m, 2H, ArgδCH_2), 2.74 (m, 4H, 2xLysεCH_2), 1.85 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.66 (m, 3H, ArgβCH + 2xLysβCH), 1.61-1.42 (m, 10H, 2xLysβCH + 2xLysδCH_2+ NvaβCH_2+ ArgγCH_2), 1.38-1.21 (m, 6H, 2xLysγCH_2 + NvaγCH_2), 0.85 (t, 3H, NvaδCH_3). ESMS [M+H] = 692.5, [M+2H]/2 = 346.7.

In a similar manner the following substrates were prepared.

Ac(Nle)KKR-pNA. (C18 R_t = 17.7 min.): ^{1}H nmr (DMSO-d_6) δ 10.67 (s, 1H, *p*NA-NH), 8.25 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.03 (d, 1H, NleαNH), 7.99 (d, 1H, LysαNH), 7.86 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.70 (m, 6H, 2xLyszNH_3), 7.63 (t, 1H, ArgεNH), 7.6-6.7 (4H, m, 2xArgNH_2), 4.36 (m, 1H, ArgαCH), 4.25 (m, 1H, LysαCH), 4.22 (m, 1H,LysαCH), 4.15 (m, 1H, NleαCH), 3.12 (m, 2H, ArgδCH_2), 2.74 (m, 4H, 2xLysεCH_2), 1.85 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.66 (m, 3H, ArgβCH + 2xLysβCH), 1.63-1.42 (m, 10H, 2xLysβCH + 2xLysδCH_2+ NleβCH_2+ ArgγCH_2), 1.37-1.18 (m, 8H, 2xLysγCH_2 + NleγCH_2 + NleδCH_2), 0.84 (t, 3H, NvaεCH_3). ESMS [M+H] = 706.6, [M+2H]/2 = 353.8.

AcLKKR-pNA. (C18 R_t = 17.0 min.): ^{1}H nmr (DMSO-d_6) δ 10.67 (s, 1H, *p*NA-NH), 8.25 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.03 (d, 1H, LeuαNH), 7.99 (d, 1H, LysαNH), 7.86 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.71 (m, 6H, 2xLyszNH_3), 7.63 (t, 1H, ArgεNH), 7.5-6.5 (4H, m, 2xArgNH_2), 4.37 (m, 1H, ArgαCH), 4.25 (m, 1H, LysαCH), 4.23 (m, 1H, LeuαCH), 4.22 (m, 1H,LysαCH), 3.13 (m, 2H, ArgδCH_2), 2.74 (m, 4H, 2xLysεCH_2), 1.85 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.72-1.22 (m, 18H, 8xCH_2 + ArgβCH + LeuγCH), 0.88 (t, 3H, LeuδCH_3), 0.83 (t, 3H, LeuδCH_3). ESMS [M+H] = 706.4, [M+2H]/2 = 353.9.

AcLKRR-pNA. (C18 R_t = 17.5 min.): ^{1}H nmr (DMSO-d_6) δ 10.67 (s, 1H, *p*NA-NH), 8.28 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.04 (d, 1H, LeuαNH), 8.01 (d, 1H, LysαNH), 7.89 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.71 (m, 3H, LyszNH_3), 7.65 (t, 1H, ArgεNH), 7.59 (t, 1H, ArgεNH), 7.52-6.53 (4H, m, 2xArgNH_2), 4.36 (m, 1H, ArgαCH), 4.27 (m, 1H, ArgαCH), 4.22 (m, 1H, LeuαCH), 4.21 (m, 1H,LysαCH), 3.12 (m, 2H, ArgδCH_2), 3.09 (m, 2H, ArgδCH_2), 2.73 (m, 2H, 2xLysεCH_2), 1.85 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.72-1.22 (m, 16H, 7xCH_2 + ArgβCH + LeuγCH), 0.87 (t, 3H, LeuδCH_3), 0.83 (t, 3H, LeuδCH_3). ESMS [M+H] = 734.5, [M+2H]/2 = 368.0.

Ac(Chg)KKR-*p*NA. (C18 R_t = 18.2 min.): ^{1}H nmr (DMSO-d_6) δ 10.66 (s, 1H, *p*NA-NH), 8.26 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.03 (d, 1H, LysαNH), 7.91 (d, 1H, ChgαNH), 7.85 (d, 2H, *p*NA-ArH), 7.81 (d, 1H, LysαNH), 7.69 (m, 6H, 2xLyszNH_3), 7.60 (t, 1H, ArgεNH), 7.5-6.6 (4H, m,

2xArgNH_2), 4.37 (m, 1H, ArgαCH), 4.26 (m, 1H, LysαCH), 4.21 (m, 1H,LysαCH), 4.07 (m, 1H, ChgαCH), 3.12 (m, 2H, ArgδCH_2), 2.73 (m, 4H, 2xLysεCH_2), 1.86 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.72-0.90 (m, 26H, 12xCH_2 + ArgβCH + ChgβCH). ESMS [M+H] = 732.5, [M+2H]/2 = 366.8.

Ac(Cha)KKR-pNA. (C18 R_t = 19.0 min.): ^{1}H nmr (DMSO-d_6) δ 10.65 (s, 1H, *p*NA-NH), 8.24 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.03 (d, 1H, ChaαNH), 7.97 (d, 1H, LysαNH), 7.86 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.70 (m, 6H, 2xLyszNH_3), 7.63 (t, 1H, ArgεNH), 7.5-6.6 (4H, m, 2xArgNH_2), 4.36 (m, 1H, ArgαCH), 4.23 (m, 3H, 2xLysαCH + ChaαCH), 3.12 (m, 2H, ArgδCH_2), 2.74 (m, 4H, 2xLysεCH_2), 1.85 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.72-1.22 (m, 22H, ArgβCH + ArgγCH_2 + 2xLysβCH_2 + 2xLysγCH_2+ 2xLysδCH_2+ ChaβCH_2 + ChaγCH + 3xChaCH_2), 1.16 (m, 1H, ChaCH), 1.10 (m, 1H, ChaCH), 0.87 (m, 1H, ChaCH), 0.81 (m, 1H, ChaCH). ESMS [M+H] = 746.5, [M+2H]/2 = 373.8.

Ac(Aib)KKR-pNA. (C18 R_t = 17.0 min.): ^{1}H nmr (DMSO-d_6) δ 10.46 (s, 1H, *p*NA-NH), 8.37 (s, 1H, AibαNH), 8.25 (d, 2H, *p*NA-ArH), 8.01 (br d, 1H, ArgαNH), 7.95 (d, 1H, LysαNH), 7.88 (d, 2H, *p*NA-ArH), 7.77 (d, 1H, LysαNH), 7.71 (m, 6H, 2xLyszNH_3), 7.62 (br s 1H, ArgεNH), 7.5-6.7 (4H, m, 2xArgNH_2), 4.34 (m, 1H, ArgαCH), 4.15 (m, 1H, LysαCH), 4.08 (m, 1H,LysαCH), 3.10 (m, 2H, ArgδCH_2), 2.74 (m, 4H, 2xLysεCH_2), 1.86 (s, 3H, Ac-CH_3), 1.78 (m, 1H, ArgβCH), 1.73-1.58 (m, 3H, ArgβCH + 2xLysβCH), 1.58-1.41 (m, 8H, 2xLysβCH + 2xLysδCH_2 + ArgγCH_2), 1.40-1.20 (m, 4H, 2xLysγCH_2), 1.32 (s, 3H, AibβCH_3), 1.30 (s, 3H, AibβCH_3). ESMS [M+H] = 678.5, [M+2H]/2 = 339.7.

Ac(Tbg)KKR-pNA. (C8 R_t = 17.3 min.): ^{1}H nmr (DMSO-d_6) δ 10.68 (s, 1H, *p*NA-NH), 8.27 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.04 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.84 (d, 1H, LysαNH), 7.78 (d, 1H, TbgαNH), 7.68 (m, 6H, 2xLyszNH_3), 7.58 (t, 1H, ArgεNH), 7.5-6.6 (4H, m, 2xArgNH_2), 4.37 (m, 1H, ArgαCH), 4.27 (m, 1H, LysαCH), 4.21 (m, 1H,LysαCH), 4.17 (d, 1H, TbgαCH), 3.12 (m, 2H, ArgδCH_2), 2.73 (m, 4H, 2xLysεCH_2), 1.90 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.65 (m, 3H, ArgβCH + 2xLysβCH), 1.60-1.43 (m, 8H, 2xLysβCH + 2xLysδCH_2 + ArgγCH_2), 1.31 (m, 4H, 2xLysγCH_2), 0.92 (s, 9H, 3xTbgγCH_3). ESMS [M+H] = 706.4, [M+2H]/2 = 353.7.

Ac(Aoc)KKR-pNA. (C18 R_t = 19.3 min.): ^{1}H nmr (DMSO-d_6) δ 10.69 (s, 1H, *p*NA-NH), 8.28 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.05 (d, 1H, AocαNH), 8.01 (d, 1H, LysαNH), 7.90 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.71 (m, 6H, 2xLyszNH_3), 7.65 (t, 1H, ArgεNH), 7.6-6.7 (4H, m, 2xArgNH_2), 4.34 (m, 1H, ArgαCH), 4.24 (m, 1H, LysαCH), 4.21 (m, 1H,LysαCH), 4.14 (m, 1H, AocαCH), 3.12 (m, 2H, ArgδCH_2), 2.73 (m, 4H, 2xLysεCH_2), 1.84 (s, 3H, Ac-CH_3), 1.76 (m, 1H, ArgβCH), 1.66 (m, 3H, ArgβCH + 2xLysβCH), 1.60-1.42 (m, 12H, 2xLysβCH + 2xLysδCH_2 + ArgγCH_2+ AocβCH_2+ AocγCH_2), 1.37-1.16 (m, 6H, 2xLysγCH_2 + AocδCH_2+ AocεCH_2+ AoczCH_2), 0.84 (t, 3H, AocCH_3). ESMS [M+H] = 734.6, [M+2H]/2 = 367.8.

2-Naphthoyl-KKR-pNA. (C18 R_t = 19.1 min.): ^{1}H nmr (DMSO-d_6) δ 10.70 (s, 1H, *p*NA-NH), 8.65 (d, 1H, LysαNH), 8.51 (s, 1H, ArH), 8.29 (br d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.09 (br d, 1H, LysαNH), 8.05-7.94 (m, 4H, ArH), 7.84 (d, 2H, *p*NA-ArH), 7.70 (m, 6H, 2xLyszNH_3), 7.66-7.58 (m, 3H, ArH + ArgεNH), 7.6-6.7 (4H, m, 2xArgNH_2), 4.50 (m, 1H, LysαCH), 4.37 (m, 1H, ArgαCH), 4.33 (m, 1H, LysαCH), 3.12 (m, 2H, ArgδCH_2), 2.79 (m, 2H, LysεCH_2), 2.74 (m, 2H, LysεCH_2), 1.85-1.26 (m, 16H, 8xCH_2). ESMS [M+H] = 705.5, [M+2H]/2 = 353.2.

2-Naphthoyl-(Cit)KR-pNA. (C18 R_t = 20.5 min.): ^{1}H nmr (DMSO-d_6) δ 10.70 (s, 1H, *p*NA-NH), 8.67 (d, 1H, CitαNH), 8.51 (s, 1H, ArH), 8.28 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.12 (d, 1H, LysαNH), 8.06-7.94 (m, 4H, ArH), 7.85 (d, 2H, *p*NA-ArH), 7.67 (m, 6H, 2xLyszNH_3), 7.65-7.58 (m, 2H, ArH), 7.56 (t, 1H, ArgεNH), 7.5-6.7 (4H, m, 2xArgNH_2), 6.03 (t, 1H, CitεNH), 5.44 (br s, 2H, CitNH_2), 4.50 (m, 1H, CitαCH), 4.39 (m, 1H, ArgαCH), 4.33 (m, 1H, LysαCH), 3.12 (m, 2H, ArgδCH_2), 3.01 (m, 2H, CitδCH_2), 2.75 (m, 2H, LysεCH_2), 1.83-1.29 (m, 14H, 7xCH_2). ESMS [M+H] = 734.5, [M+2H]/2 = 367.8.

AcL(Cit)KR-pNA. (C18 R_t = 17.9 min.): ^{1}H nmr (DMSO-d_6) δ 10.67 (s, 1H, *p*NA-NH), 8.24 (m, 3H, *p*NA-ArH + ArgαNH), 8.01 (d, 1H, LeuαNH), 7.94 (d, 1H, CitαNH), 7.91 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.67 (m, 3H, LyszNH_3), 7.55 (t, 1H, ArgεNH), 7.5-6.6 (4H, m, 2xArgNH_2), 5.98 (t, 1H, CitεNH), 5.43 (br s, 2H, CitNH_2), 4.38 (m, 1H, ArgαCH), 4.24 (m, 3H, LysαCH + CitαCH + LeuαCH), 3.12 (m, 2H, ArgδCH_2), 2.94 (m, 2H, CitδCH_2), 2.74 (m, 2H, LysεCH_2), 1.85 (s, 3H, Ac-CH_3), 1.76 (m, 1H, ArgβCH), 1.70-1.23 (m, 16H, ArgβCH + ArgγCH_2 + LysβCH_2 + LysγCH_2 + LysδCH_2 + CitβCH_2 + CitγCH_2+ LeuβCH_2 + LeuγCH), 0.87 (d, 3H, LeuδCH_3), 0.83 (d, 3H, LeuδCH_3). ESMS [M+H] = 735.3, [M+2H]/2 = 368.2.

2-Naphthoyl-(Orn)KR-pNA. (C18 R_t = 19.3 min.): ^{1}H nmr (DMSO-d_6) δ 10.70 (s, 1H, *p*NA-NH), 8.70 (d, 1H, OrnαNH), 8.52 (s, 1H, ArH), 8.31 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.11 (d, 1H, LysαNH), 8.04-7.95 (m, 4H, ArH), 7.84 (d, 2H, *p*NA-ArH), 7.71 (m, 6H, LyszNH_3 + OrnεNH_3), 7.66-7.59 (m, 3H, ArH + ArgεNH), 7.5-6.8 (4H, m, 2xArgNH_2), 4.55 (m, 1H, OrnαCH), 4.37 (m, 1H, ArgαCH), 4.34 (m, 1H, LysαCH), 3.12 (m, 2H, ArgδCH_2), 2.82 (m, 2H, OrnδCH_2), 2.74 (m, 2H, LysεCH_2), 1.90-1.20 (m, 14H, 7xCH_2). ESMS [M+H] = 691.5, [M+2H]/2 = 346.3.

AcL(Orn)KR-pNA. (C18 R_t = 17.2 min.): ^{1}H nmr (DMSO-d_6) δ 10.70 (s, 1H, *p*NA-NH), 8.27 (d, 1H, ArgαNH), 8.23 (m, 2H, *p*NA-ArH), 8.06 (d, 1H, OrnαNH), 8.04 (d, 1H, LeuαNH), 7.89 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.76 (m, 6H, OrnεNH_3 + LyszNH_3), 7.73 (t, 1H, ArgεNH), 7.6-6.8 (4H, m, 2xArgNH_2), 4.36 (m, 1H, ArgαCH), 4.25 (m, 3H, LysαCH + OrnαCH + LeuαCH), 3.12 (m, 2H, ArgδCH_2), 2.77 (m, 2H, OrnδCH_2), 2.73 (m, 2H, LysεCH_2), 1.85 (s, 3H, Ac-CH_3), 1.81-1.45 (m, 13H, 6xCH_2 + LeuγCH), 1.42 (m, 2H, LeuβCH_2), 1.31 (m, 2H, LysγCH_2), 0.87 (d, 3H, LeuδCH_3), 0.83 (d, 3H, LeuδCH_3). ESMS [M+H] = 692.5, [M+2H]/2 = 346.7

2-Naphthoyl-(3PyA)KR-pNA. (C18 R_t = 18.9 min.): ^{1}H nmr (DMSO-d_6) δ 10.74 (s, 1H, *p*NA-NH), 8.78 (d, 1H, 3PyAαNH), 8.62 (s, 1H, ArH), 8.45 (br d, 1H, ArH), 8.39 (s, 1H, ArH), 8.38 (d, 1H, ArgαNH), 8.30 (d, 1H, LysαNH), 8.23 (d, 2H, *p*NA-ArH), 8.02-7.96 (m, 3H, ArH), 7.91 (m, 1H, ArH), 7.84 (m, 3H, ArH), 7.70-7.57 (m, 5H, ArH + LyszNH_3), 7.53 (t, 1H, ArgεNH), 7.5-6.6 (4H, m, 2xArgNH_2), 7.41 (m, 3H, ArH), 4.84 (m, 1H, 3PyAαCH), 4.42 (m, 1H, ArgαCH), 4.38 (m, 1H, LysαCH), 3.21 (dd, 1H, 3PyAβCH), 3.12 (m, 2H, ArgδCH_2), 3.08 (dd, 1H, 3PyAβCH), 2.74 (m, 2H, LysεCH_2), 1.78 (m, 1H, ArgβCH), 1.71 (m, 1H, LysβCH), 1.66 (m, 1H, ArgβCH), 1.63-1.45 (m, 5H, LysβCH + LysδCH_2 + ArgγCH_2), 1.36 (m, 2H, LysγCH_2). ESMS [M+H] = 725.2, [M+2H]/2 = 363.2.

AcL(3PyA)KR-pNA. (C18 R_t = 16.6 min.): ^{1}H nmr (DMSO-d_6) δ 10.72 (s, 1H, *p*NA-NH), 8.50 (br d, 1H, ArH), 8.48 (br s, 1H, ArH), 8.36 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.06 (m, 2H, LysαNH), 7.99 (d, 1H, LeuαNH), 7.95 (d, 1H, 3PyAαNH), 7.85 (d, 2H, *p*NA-ArH), 7.81 (br d, 1H, ArH), 7.69 (m, 3H, LyszNH_3), 7.61 (t, 1H, ArgεNH), 7.5-6.6 (4H, m, 2xArgNH_2), 7.43 (m, 1H, ArH), 4.58 (m, 1H, 3PyAαCH), 4.42 (m, 1H, ArgαCH), 4.30 (m, 1H, LysαCH), 4.14 (m, 1H, LeuαCH), 3.14 (m, 2H, ArgδCH_2), 3.11 (m, 1H, 3PyAβCH), 2.88 (m, 1H, 4PyAβCH), 2.74 (m, 4H, 2xLysεCH_2), 1.81 (s, 3H, Ac-CH_3), 1.78 (m, 1H, ArgβCH), 1.66 (m, 2H, ArgβCH + LysβCH), 1.62-1.43 (m, 6H, LysβCH + LysδCH_2 + LeuγCH + ArgγCH_2), 1.38-1.24 (m, 4H, LysγCH_2 + LeuβCH_2), 0.83 (d, 3H, LeuδCH_3), 0.78 (d, 3H, LeuδCH_3). ESMS [M+H] = 726.4, [M+2H]/2 = 363.7.

AcL(4PyA)KR-pNA. (C18 R_t = 17.2 min.): ^{1}H nmr (DMSO-d_6) δ 10.71 (s, 1H, *p*NA-NH), 8.53 (br s, 2H, ArH), 8.35 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.02 (m, 2H, 4PyAαNH + LysαNH), 7.97 (d, 1H, LeuαNH), 7.85 (d, 2H, *p*NA-ArH), 7.66 (m, 3H, LyszNH_3), 7.55 (t, 1H, ArgεNH), 7.5-6.6 (4H, m, 2xArgNH_2), 7.41 (br s, 2H, ArH), 4.62 (m, 1H, 4PyAαCH), 4.41 (m, 1H, ArgαCH), 4.30 (m, 1H, LysαCH), 4.14 (m, 1H, LeuαCH), 3.13 (m, 3H, 4PyAβCH + ArgδCH_2), 2.91 (m, 1H, 4PyAβCH), 2.74 (m, 4H, 2xLysεCH_2), 1.83 (s, 3H, Ac-CH_3), 1.77 (m, 1H, ArgβCH), 1.65 (m, 2H, ArgβCH + LysβCH), 1.60-1.42 (m, 6H, LysβCH + LysδCH_2 + LeuγCH + ArgγCH_2), 1.38-1.22 (m, 4H, LysγCH_2 + LeuβCH_2), 0.83 (d, 3H, LeuδCH_3), 0.77 (d, 3H, LeuδCH_3). ESMS [M+H] = 726.8, [M+2H]/2 = 363.7.

2-Naphthoyl-(hArg)KR-pNA. (C18 R_t = 20.3 min.): ^{1}H nmr (DMSO-d_6) δ 10.65 (s, 1H, *p*NA-NH), 8.62 (d, 1H, hArgαNH), 8.50 (s, 1H, ArH), 8.25 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.06-

7.94 (m, 5H, ArH + LysαNH), 7.84 (d, 2H, *p*NA-ArH), 7.71-7.59 (m, 5H, LyszNH$_3$ + ArH), 7.52 (br t, 1H, ArgεNH), 7.48 (br t, 1H, hArgzNH), 7.5-6.5 (8H, m, 2xArgNH$_2$ + 2xhArgNH$_2$), 4.50 (m, 1H, hArgαCH), 4.39 (m, 1H, ArgαCH), 4.34 (m, 1H, LysαCH), 3.11 (m, 4H, ArgδCH$_2$ +hArgεCH$_2$), 2.73 (m, 2H, LysεCH$_2$), 1.87-1.25 (m, 16H, 8xCH$_2$). ESMS [M+H] = 747.2, [M+2H]/2 = 374.3.

AcL(hArg)KR-pNA. (C18 R$_t$ = 17.3 min.): ^{1}H nmr (DMSO-d$_6$) δ 10.68 (s, 1H, *p*NA-NH), 8.26 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.05 (d, 1H, LeuαNH), 7.99 (d, 1H, hArgαNH), 7.87 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.75 (m, 3H, LyszNH$_3$), 7.71 (t, 1H, ArgεNH), 7.61 (t, 1H, hArgzNH), 7.6-6.7 (4H, m, 2xArgNH$_2$), 4.36 (m, 1H, ArgαCH), 4.24(m, 3H, LysαCH + hArgαCH + LeuαCH), 3.12 (m, 2H, ArgδCH$_2$), 3.05 (m, 2H, hArgεCH$_2$), 2.73 (m, 4H, 2xLysεCH$_2$), 1.85 (s, 3H, Ac-CH$_3$), 1.77 (m, 1H, Argβ SMS [M+H] = 748.3, [M+2H]/2 = 374.7.

AcL(NO$_2$F)KR-pNA. (C18 R$_t$ = 20.7 min.): ^{1}H nmr (DMSO-d$_6$) δ 10.71 (s, 1H, *p*NA-NH), 8.35 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.10 (d, 2H, ArH), 8.07 (d, 1H, LysαNH), 7.96 (d, 1H, NO$_2$PheαNH), 7.95 (d, 1H, LeuαNH), 7.85 (d, 2H, *p*NA-ArH), 7.70 (m, 3H, LyszNH$_3$), 7.60 (t, 1H, ArgεNH), 7.6-6.5 (4H, m, 2xArgNH$_2$), 7.49 (d, 2H, ArH), 4.61 (m, 1H, NO$_2$PheαCH), 4.41 (m, 1H, ArgαCH), 4.30 (m, 1H,LysαCH), 4.14 (m, 1H, LeuαCH), 3.19 (dd, 1H, NO$_2$PheβCH), 3.13 (m, 2H, ArgδCH$_2$), 2.94 (dd, 1H, NO$_2$PheβCH), 2.74 (m, 2H, LysεCH$_2$), 1.81 (s, 3H, Ac-CH$_3$), 1.78 (m, 1H, ArgβCH), 1.67 (m, 2H, ArgβCH + LysβCH), 1.61-1.46 (m, 5H, LysβCH + ArgγCH$_2$ + LysδCH$_2$), 1.44 (m, 1H, LeuγCH), 1.32 (m, 2H, LysγCH$_2$), 1.26 (m, 2H, LeuβCH$_2$), 0.81 (d, 3H, LeuδCH$_3$), 0.75 (d, 3H, LeuδCH$_3$). ESMS [M+H] = 770.4, [M+2H]/2 = 385.7.

AcL(NH$_2$F)KR-pNA. (C18 R$_t$ = 17.2 min.): ^{1}H nmr (DMSO-d$_6$) δ 10.68 (s, 1H, *p*NA-NH), 8.28 (d, 1H, ArgαNH), 8.24 (m, 3H, *p*NA-ArH), 7.99 (d, 1H, LeuαNH), 7.93 (d, 1H, LysαNH), 7.85 (d, 2H, *p*NA-ArH), 7.81 (d, 1H, NH$_2$PheαNH), 7.67 (m, 3H, LyszNH$_3$), 7.55 (t, 1H, ArgεNH), 7.5-6.6 (4H, m, 2xArgNH$_2$), 7.00 (br s, 2H, ArH), 6.69 (br s, 2H, ArH), 4.40 (m, 2H, ArgαCH + NH$_2$PheαCH), 4.28 (m, 1H, LysαCH), 4.17 (m, 1H, LeuαCH), 3.13 (m, 2H, ArgδCH$_2$), 2.92 (br d, 1H, NH$_2$PheβCH$_2$), 2.74 (m, 2H, LysεCH$_2$), 2.70 (br d, 1H, NH$_2$PheβCH$_2$), 1.82 (s, 3H, Ac-CH$_3$), 1.78 (m, 1H, ArgβCH), 1.67 (m, 2H, ArgβCH + LysβCH), 1.61-1.45 (m, 6H, ArgγCH$_2$ + LysβCH+ LysδCH$_2$ + LeuγCH), 1.37-1.26 (m, 4H, LeuβCH$_2$ + LysγCH$_2$), 0.84 (d, 3H, LeuδCH$_3$), 0.79 (d, 3H, LeuδCH$_3$). ESMS [M+H] = 740.4, [M+2H]/2 = 370.8.

2-Naphthoyl-K(Cit)R-pNA. (C18 R$_t$ = 19.4 min.): ^{1}H nmr (DMSO-d$_6$) δ 10.64 (s, 1H, *p*NA-NH), 8.67 (d, 1H, LysαNH), 8.51 (s, 1H, ArH), 8.25 (d, 1H, ArgαNH), 8.22 (d, 2H, *p*NA-ArH), 8.07 (br d, 1H, CitαNH), 8.05-7.95 (m, 4H, ArH), 7.84 (d, 2H, *p*NA-ArH), 7.70 (m, 3H, LyszNH$_3$), 7.65-7.56 (m, 3H, ArH + ArgεNH), 7.5-6.6 (4H, m, 2xArgNH$_2$), 6.01 (t, 1H, CitεNH), 5.43 (br s , 2H, CitNH$_2$), 4.50 (m, 1H, LysαCH), 4.35 (m, 1H, ArgαCH), 4.33 (m, 1H, CitαCH), 3.11 (m, 2H, ArgδCH$_2$), 2.97 (m, 2H, CitδCH$_2$), 2.79 (m, 2H, LysεCH$_2$), 1.87-1.33 (m, 14H, 7xCH$_2$). ESMS [M+H] = 734.1, [M+2H]/2 = 367.8.

2-Naphthoyl-K(Orn)R-pNA. (C18 R$_t$ = 18.3 min.): ^{1}H nmr (DMSO-d$_6$) δ 10.73 (s, 1H, *p*NA-NH), 8.67 (d, 1H, LysαNH), 8.52 (s, 1H, ArH), 8.31 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.18 (br d, 1H, OrnαNH), 8.05-7.95 (m, 4H, ArH), 7.85 (d, 2H, *p*NA-ArH), 7.79 (m, 6H, OrnzNH$_3$ + LyszNH$_3$), 7.74 (t, 1H, ArgεNH), 7.65-7.58 (m, 2H, ArH), 7.6-6.8 (4H, m, 2xArgNH$_2$), 4.51 (m, 1H, LysαCH), 4.38 (m, 2H, ArgαCH + OrnαCH), 3.12 (m, 2H, ArgδCH$_2$), 2.81 (m, 4H, OrnδCH$_2$ + LysεCH$_2$), 1.86-1.34 (m, 14H, 7xCH$_2$). ESMS [M+H] = 691.2, [M+2H]/2 = 346.2.

2-Naphthoyl-KRR-pNA. (C18 R$_t$ = 18.4 min.): ^{1}H nmr (DMSO-d$_6$) δ 10.71 (s, 1H, *p*NA-NH), 8.68 (d, 1H, LysαNH), 8.51 (s, 1H, ArH), 8.31 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.12 (br d, 1H, LysαNH), 8.04-7.95 (m, 4H, ArH), 7.84 (d, 2H, *p*NA-ArH), 7.74 (m, 3H, LyszNH$_3$), 7.68 (t, 1H, ArgεNH), 7.65-7.59 (m, 2H, ArH), 7.6-6.7 (8H, m, 4xArgNH$_2$), 4.50 (m, 1H, LysαCH), 4.36 (m, 2H, 2xArgαCH), 3.11 (m, 4H, 2xArgδCH$_2$), 2.79 (m, 2H, LysεCH$_2$), 1.84-1.35 (m, 14H, 7xCH$_2$). ESMS [M+H] = 733.2, [M+2H]/2 = 367.2.

2-Naphthoyl-K(hArg)R-pNA. (C18 R_t = 18.6 min.): ^{1}H nmr (DMSO-d_6) δ 10.70 (s, 1H, *p*NA-NH), 8.66 (d, 1H, LysαNH), 8.51 (s, 1H, ArH), 8.29 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.09 (d, 1H, hArgαNH), 8.05-7.94 (m, 4H, ArH), 7.85 (d, 2H, *p*NA-ArH), 7.75 (m, 3H, LyszNH_3), 7.69(t, 1H, ArgεNH), 7.66-7.59 (m, 2H, ArH), 7.6-6.7 (8H, m, 2xArgNH_2 + 2xhArgNH_2), 4.50 (m, 1H, LysαCH), 4.37 (m, 1H, ArgαCH), 4.33 (m, 1H, hArgαCH), 3.12 (m, 2H, ArgδCH_2), 3.05 (m, 2H, hArgδCH_2), 2.79 (m, 2H, LysεCH_2), , 1.85-1.26 (m, 16H, 8xCH_2). ESMS [M+H] = 747.2, [M+2H]/2 = 374.3.

2-Naphthoyl-LK(AcK)R-pNA. (C18 R_t = 20.5 min.): ^{1}H nmr (DMSO-d_6) δ 10.64 (s, 1H, *p*NA-NH), 8.67 (d, 1H, LysαNH), 8.51 (s, 1H, ArH), 8.26 (d, 1H, ArgαNH), 8.22 (d, 2H, *p*NA-ArH), 8.05-7.94 (m, 5H, ArH + LysαNH), 7.84 (d, 2H, *p*NA-ArH), 7.79 (t, 1H, LyzzNH), 7.70 (m, 3H, LyszNH_3), 7.65-7.56 (m, 3H, ArH + ArgεNH), 7.5-6.7 (4H, m, 2xArgNH_2), 4.50 (m, 1H, LysαCH), 4.35 (m, 1H, ArgαCH), 4.29 (m, 1H, LysαCH), 3.11 (m, 2H, ArgδCH_2), 2.98 (m, 2H, LysεCH_2), 2.79 (m, 2H, LysεCH_2), 1.86-1.22 (m, 16H, 8xCH_2), 1.77 (s, 3H, Ac-CH_3). ESMS [M+H] = 747.2, [M+2H]/2 = 374.3.

2-Naphthoyl-K(NH_2F)R-pNA. (C18 R_t = 19.6 min.): ^{1}H nmr (DMSO-d_6) δ 10.66 (s, 1H, *p*NA-NH), 8.65 (d, 1H, LysαNH), 8.49 (s, 1H, ArH), 8.35 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.05-7.98 (m, 3H, ArH), 7.94 (m,2H, ArfH + NH_2FαNH), 7.86 (d, 2H, *p*NA-ArH), 7.69 (m, 3H, LyszNH_3), 7.66-7.53 (m, 3H, ArH + ArgεNH), 7.5-6.6 (4H, m, 2xArgNH_2), 7.05 (m, 2H, ArH), 6.67 (m, 2H, ArH), 4.54 (m, 1H, NH_2FαCH), 4.43 (m, 1H, LysαCH), 4.38 (m, 1H, ArgαCH), 3.11 (m, 2H, ArgδCH_2), 2.97 (dd, 1H, NH_2FβCH), 2.77 (m, 3H, LysεCH_2 + NH_2FβCH), 1.85-1.26 (m, 10H, 5xCH_2). ESMS [M+H] = 739.3, [M+2H]/2 = 370.4.

2-Naphthoyl-K(Nle)R-pNA. (C18 R_t = 22.7 min.): ^{1}H nmr (DMSO-d_6) δ 10.64 (s, 1H, *p*NA-NH), 8.68 (d, 1H, LysαNH), 8.51 (s, 1H, ArH), 8.27 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.05-7.95 (m, 5H, ArH + NleαNH), 7.84 (d, 2H, *p*NA-ArH), 7.71 (m, 3H, LyszNH_3), 7.65-7.58 (m, 3H, ArH + ArgεNH), 7.6-6.7 (4H, m, 2xArgNH_2), 4.50 (m, 1H, LysαCH), 4.36 (m, 1H, ArgαCH), 4.31 (m, 1H, NleαCH), 3.11 (m, 2H, ArgδCH_2), 2.79 (m, 2H, LysεCH_2), 1.87-1.20 (m, 16H, 8xCH_2), 0.81 (t, 3H, NleεCH_3. ESMS [M+H] = 690.4, [M+2H]/2 = 345.9.

2-Naphthoyl-KFR-pNA. (C18 R_t = 22.8 min.): ^{1}H nmr (DMSO-d_6) δ 10.66 (s, 1H, *p*NA-NH), 8.65 (d, 1H, LysαNH), 8.48 (s, 1H, ArH), 8.41 (d, 1H, ArgαNH), 8.24 (d, 2H, *p*NA-ArH), 8.05-7.97 (m, 4H, ArH + PheαNH), 7.95 (dd, 1H, ArH), 7.86 (d, 2H, *p*NA-ArH), 7.69 (m, 6H, 2xLyszNH_3), 7.66-7.57 (m, 3H, ArH + ArgεNH), 7.5-6.7 (4H, m, 2xArgNH_2), 7.23 (br d, 2H, ArH), 7.14 (br t, 2H, ArH), 7.10 (br t, 1H, ArH), 4.62 (m, 1H, PheαCH), 4.41 (m, 1H, LysαCH), 4.39 (m, 1H, ArgαCH), 3.12 (m, 2H, ArgδCH_2), 3.08 (dd, 1H, PheβCH), 2.86 (dd, 1H, PheβCH), 2.75 (m, 2H, LysεCH_2), 1.84-1.20 (m, 10H, 5xCH_2). ESMS [M+H] = 724.2, [M+2H]/2 = 362.6.

2-Naphthoyl-(Orn)(Orn)R-pNA. (C18 R_t = 18.2 min.): ^{1}H nmr (DMSO-d_6) δ 10.73 (s, 1H, *p*NA-NH), 8.69 (d, 1H, OrnαNH), 8.52 (s, 1H, ArH), 8.34 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.20 (br d, 1H, OrnαNH), 8.04-7.96 (m, 4H, ArH), 7.84 (d, 2H, *p*NA-ArH), 7.73 (m, 6H, 2xOrnεNH_3), 7.66-7.59 (m, 3H, ArH + ArgεNH), 7.5-6.7 (4H, m, 2xArgNH_2), 4.57 (m, 1H, OrnαCH), 4.38 (m, 2H, ArgαCH+ OrnαCH), 3.12 (m, 2H, ArgδCH_2), 2.90-2.74 (m, 4H, 2xOrnδCH_2), 1.90-1.43 (m, 12H, 6xCH_2). ESMS [M+H] = 677.1, [M+2H]/2 = 339.3.

2-Naphthoyl-(Cit)(Orn)R-pNA. (C18 R_t = 19.4 min.): ^{1}H nmr (DMSO-d_6) δ 10.71 (s, 1H, *p*NA-NH), 8.66 (d, 1H, CitαNH), 8.51 (s, 1H, ArH), 8.31 (d, 1H, ArgαNH), 8.23 (d, 2H, *p*NA-ArH), 8.18 (br d, 1H, OrnαNH), 8.04-7.95 (m, 4H, ArH), 7.84 (d, 2H, *p*NA-ArH), 7.70 (m, 3H, OrnεNH_3), 7.65-7.59 (m, 2H, ArH), 7.5-6.6 (4H, m, 2xArgNH_2), 6.07 (t, 1H, CitεNH), 5.49 (br s, 2H, CitNH_2), 4.51 (m, 1H, CitαCH), 4.40 (m, 1H, ArgαCH), 4.37 (m, 1H, OrnαCH), 3.11 (m, 2H, ArgδCH_2), 3.01(m, 2H, CitδCH_2), 2.80 (m, 2H, OrnδCH_2), 1.84-1.41 (m, 12H, 6xCH_2). ESMS [M+H] = 720.6, [M+2H]/2 = 360.6.

Table S1. Primers used for site-directed mutagenesis

Primer	**Sequence**
WNNS3-111.R	5’-CCGGACGTTCTTAACGTTCTTGCCAGG-3’
WNNS3-T111L.F	5’-TTAAGAACGTCCGGCTGAAACCAGGGGTG-3’
WNNS3-T111F.F	5’-TTAAGAACGTCCGGTTCAAACCAGGGGTG-3’
WNNS3-154,155.R	5’-TCCATTGCCATAAAGCCCAATCACATCACC-3’
WNNS3-V154F.F	5’-CTTTATGGCAATGGATTCATAATGCCCAAC-3’
WNNS3-V154L.F	5’-CTTTATGGCAATGGACTCATAATGCCCAAC-3’
WNNS3-I155F.F	5’-CTTTATGGCAATGGAGTCTTCATGCCCAACGGC-3’
WNNS3-M156A.F	5’-ATAGCGCCCAACGGCTCATACATAAGC-3’
WNNS3-M156A.R	5’-GTTGGGCGCTATGACTCCATTGCCATA-3’
WNNS3-162,164,166.R	5’-GTATGAGCCGTTGGGCATTATGACTCCATT-3’
WNNS3-I162F.F	5’-CCAACGGCTCATACTTCAGCGCGATAGTG-3’
WNNS3-A164S.F	5’-CCAACGGCTCATACATAAGCTCGATAGTGCAG-3’
WNNS3-A164V.F	5’-CCAACGGCTCATACATAAGCGTGATAGTGCAG-3’
WNNS3-V166L.F	5’-CCAACGGCTCATACATAAGCGCGATACTGCAGGGTGAA-3’
WNCF40-V75A.F	5’-AGAGCTGATGTGCGGCTTGATGATGATG-3’
WNCF40-V75A.R	5’-CGCACATCAGCTCTTTCGCTTGAGCCT-3’
WNCF40-75.R	5’-TCTTTCGCTGGAGCCTGTAATTTCTGCAT-3’
WNCF40-V75L.F	5’-TCGAGCGAAAGACTTGATGTGCGGC-3’
WNCF40-V75F.F	5’-TCGAGCGAAAGATTTGATGTGCGGC-3’
WNCF40-N84A.F	5’-GGAGCCTTCCAGCTCATGAATGATCCA-3’
WNCF40-N84A.R	5’-CTGGAAGGCTCCATCATCATCAAGCCG-3’
WNCF40-84.R	5’-TCCATCATCATCAAGCCGCACATCAAC-3’
WNCF40-N84D.F	5’-CTTGATGATGATGGAGATTTCCAGCTCATG-3’
WNCF40-N84E.F	5’-CTTGATGATGATGGAGAATTCCAGCTCATG-3’
WNCF40-N84L.F	5’-CTTGATGATGATGGACTCTTCCAGCTCATG-3’
WNCF40-N84S.F	5’-CTTGATGATGATGGATCATTCCAGCTCATG-3’
WNCF40-Q86A.F	5’-TTCGCGCTCATGAATGATCCAGGTGCT-3’
WNCF40-Q86A.R	5’-CATGAGCGCGAAGTTTCCATCATCATCA-3’
WNCF40-86,87.R	5’-GAAGTTTCCATCATCATCAAGCCGCACAT-3’
WNCF40-Q86L.F	5’-GATGATGGAAACTTCCTGCTCATGAATGATCC-3’
WNCF40-Q86E.F	5’-GATGATGGAAACTTCGAGCTCATGAATGATCC-3’
WNCF40-L87F.F	5’-GATGATGGAAACTTCCAGTTTATGAACGATCC-3’
WNCF40-L87A.F	5’-CAGGCCATGAATGATCCAGGTGCTCCT-3’
WNCF40-L87A.R	5’-CATTCATGGCCTGGAAGTTTCCATCATC-3’

Table S2. List of interacting residues from NS2B/NS 3 WNV-Pr predicted by molecular docking studies to interact with peptide substrate residues

Substrate Residue	NS3 residues	NS2B Residues
P1'	Ala 36	
	His 51	
	Thr 52	
	Thr 132	
	Ser 135	
P1	His 51	
	Asp 129	
	Tyr 130	
	Pro 131	
	Thr 132	
	Ser 135	
	Tyr 150	
	Gly 151	
	Tyr 161	
P2	His 51	Gly 83
	Asp 75	Asn 84
	Gly 151	
	Asn 152	
P3	Gly 153	
	Val 154	
	Ile 155	
	Tyr 161	
P3	Phe 116	Phe 85
	Gly 153	Gln 86
	Val 174	Leu 87
	Ile 175	

References

S1. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* , Wiley-Interscience, New York.

S2. Xwinnmr v2.6 © 2004 + Topspin v1.0 © 2005. Bruker BioSpin GmbH, Rheinstetten, Federal Republic of Germany.

S3. Abbenante, G., Leung, D., Bond, T., and Fairlie, D. P. (2001) *LIPS* 7, 347-351.

Insights to Substrate Binding and Processing by West Nile Virus NS3 Protease through Combined Modeling, Protease Mutagenesis, and Kinetic Studies

Keith J. Chappell, Martin J. Stoermer, David P. Fairlie and Paul R. Young

J. Biol. Chem. 2006, 281:38448-38458.
doi: 10.1074/jbc.M607641200 originally published online October 19, 2006
