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 4072, Australia.

Running title: Substrate processing by the West Nile virus protease

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. Optimisation 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) 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 between avian reservoir hosts and 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 (3). Currently there is no vaccine or antiviral therapy for the prevention or treatment of human WNV infection (7).

WNV is a small, enveloped virus with a single stranded, positive sense 11kb RNA genome, which encodes a single polyprotein precursor. This polyprotein must be cleaved co- and post-translationally to produce ten functional proteins; three structural (C, prM and E) and 7 non-structural (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, while 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 (Figure 1) (4). Cleavage at these sites by the NS3 protease is essential for viral replication and so the protease is a potential therapeutic target (15-18).
NS3 is a multifunctional protein, the protease comprising the N-terminal third and nucleotide triphosphatase, RNA triphosphatase and helicase components the remainder (5-7). NS3 is a trypsin-like serine protease with a classical catalytic triad (His-51, Asp-75, Ser-135) (8), 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 flaviruses (9), however different flaviviruses prefer either Lys or Arg at P2. While Den and YF NS3 proteases predominantly recognise Arg at P2, WNV protease recognises 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 (10). The flanking hydrophobic domains within NS2B are likely to function in promoting membrane association of NS2B-NS3 (11).

Due to their pivotal roles in both normal physiology and disease, proteases are increasingly attracting interest as pharmaceutical targets (12). Since early successes in HIV chemotheraphy (HIV-protease inhibitors) and in the treatment of high blood pressure (ACE inhibitors), a large number of new protease inhibitors have entered clinical trials (13). One reason for the druggability of proteases is the relatively predictable way in which they recognise their substrates and inhibitors in extended (β-strand) conformations (14, 15). 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 PDB) has also facilitated hybrid structure/substrate based drug design (16).

Recently reported crystal structures for NS2B/NS3 proteases of both WNV and Den2 (17) provide new structural insights to flaviviral proteases in ligand-bound conformations. An earlier homology model of the WNV protease (18), derived from the crystal structures of a highly homologous Dengue NS3 protease without NS2B cofactor (19) and a less homologous Hepatitis NS3 with bound NS4A cofactor (20), differs significantly from the crystal structure of WNV NS2B-NS3. This has prompted a re-examination now of some of the previous mutagenesis data (21). 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 (18). However more recent studies using non-native Dengue hexapeptide and decapeptide (22) and tetrapeptide and octapeptide (23) substrate sequences have suggested opportunities for enhancing substrate affinity for Flaviviral proteases using non-native or non-proteinogenic amino acids with hydrophobic (Nle, Leu) residues at P4 and a basic (Lys > Arg) residue at P3, a feature not seen in sub-optimal native sequences. Using a combination of computer docking of substrates into the enzyme crystal structure, site-directed mutagenesis of the protease (Figure 1), and kinetic studies of the processing of tetrapeptide substrates, we have focussed 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 towards potent substrate-based nonpeptidic inhibitors by incorporating unnatural amino acids in tetrapeptide substrates.

Experimental procedures

pNA Substrate Synthesis - pNA substrates were synthesised according to the general method of Abbenante (26) and characterised by analytical HPLC, mass spectrometry and NMR (see Supplementary Material).

Modelling of substrates into the NS3 crystal structure - The crystal structure of West Nile virus NS2B/NS3 protease (pdb code 2fp7) was prepared for docking by adding protons using InsightII (v.2000©, 2002, Accelrys Inc.). Substrates were assembled using the Biopolymer and Sketcher modules within InsightII and minimised using Discover. All substrate docking experiments were...
conducted using GOLD v2.1.2 (27). 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 (pdb: 2fp7) (28), using a distance constraint of 3.5±1.5 Å between the Arg zeta carbon and the aromatic gamma carbon of Tyr161 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 Asp129. A hydrogen bond between the P2 Lys z-NH$_3^+$ and the Asn152 side-chain carbonyl oxygen was used to anchor P2 in the shallow solvent-exposed S2 pocket, as predicted in earlier modelling work (18) and later verified experimentally (28). Hydrogen bond constraints for critical substrate backbone-enzyme interactions between NS3-Gly153 carbonyl oxygen and substrate P3 NH, Gly153 NH and substrate P3 carbonyls, and Gly151 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-position of P3 and P4 side-chains were used. For 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 minimised with the enzyme backbone either fixed or tethered and the enzyme side-chains and the entire substrate allowed to move using Discover (Accelrys Inc.).

**Plasmid Construction** - The expression plasmid pQE9 WNV CF40.Gly.NS3pro, previously generated from WNV (strain NY99-4132) (18), 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, L87F) and NS3 (T111F, T111L, V154F, V154L, I155F, M156A, I156F, A164S, A164V, V166L) were inserted by amplifying the entire plasmid with PCR using Phusion™ polymerase (Finnzymes) together with a pair of partially overlapping primers (Supplementary Material). The template plasmid was digested with DpnI at 37°C for 1 h. The PCR amplified plasmid was then used to transform E. coli strain XL10 Gold competent cells, which were grown in the presence of 100 µg/mL Ampicillin, the expression plasmid was purified, and sequences confirmed by automated analysis.

**Enzyme Expression and Purification** - The pQE9 vector was used to allow high level, inducible expression of N-terminal hexa-His tagged recombinant proteins. Cultures of *E. coli* strain SG13009 transformed with the expression plasmids containing the site-directed mutations were grown in 2 x 25 mL of LB medium containing 100 µg/mL Ampicillin and 25 µg/mL Kanamycin at 37°C until the A600$_{nm}$ reached 0.5. Expression of the recombinant protein was induced by addition of isopropyl-β-D-thiogalactopyranose (IPTG) to the final concentration of 0.3 mM and incubated for an additional 3 h at 22°C. Cells were then harvested by centrifugation at 4,500 × g for 10 min and stored at -20°C. For protein purification, the cell pellets were thawed and resuspended in 1mL 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, protease inhibitors (aprotinin, 1 µg/mL; leupeptin, 1 µg/mL; benzamidine; 1 mM PMSF) were added to give a final concentration of 1 µg/mL. 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 hexa-His 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 removed. The resin was washed with 3 x 5 mL 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-pNA) and hexapeptide (Ac-LQYTKR-pNA) substrates corresponding to P6-P1 of cleavage sites in the endogenous substrates, but with a chromogenic para-nitroanilide (pNA) group at the P1’ position.
Cleavage of pNA 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 50mM Glycine-NaOH pH 9.5, 30% glycerol and 1mM CHAPS) (18). Eight different substrate concentrations, each in triplicate, were used for determining kinetic constants. After pre-incubation in separate wells (10 min, 37°C), catalysis was initiated by mixing substrate with enzyme-buffer solution by automatic shaking for 5s. The optical density was measured at 405 nm every 11-30s for 210s-30mins (depending on activity) in a SpectraMax 250 reader and the average change in mOD/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 mOD/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 kcat, K_m and kcat/K_m were calculated 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 (14, 15), enzyme contacts with substrate beyond P1 and P2 positions were not possible to deduce 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-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 modelling experiments using GOLD to dock tetrapeptide pNA substrates into the crystal structure of the WNV NS2B/NS3 protease (pdb code: 2fp7). The tetrapeptide substrate Ac-LKRR-pNA spanning P4-P1 had been previously identified (23) as optimal for Dengue protease and, as we knew that WNV elicited a preference for Lys over Arg at P2, our docking studies began with the tetrapeptide substrate Ac-LKKR-pNA.

Docking flexible molecules such as peptides into rigid solid state structures of proteases is notoriously difficult due to inadequate sampling of conformational space, insufficiently minimised docked poses (29), 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 (14, 15).

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 (Figure 2A) 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 towards S1. This hydrophobic region is perhaps the reason that the four endogenous cleavage sequences contain a range of residues at P3 (Arg, Thr, Gly, 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. Ile155), as well as being able to accommodate both charged and polar side-chain termini directed outwards into solvent. Adjacent to Asn152, hypothesised to be the S2 hydrogen-bond acceptor of the P2 Lys side-chain (18), is a hydrophobic surface patch consisting of Val154, Met156 and cofactor residue Leu87. 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 Gln86 is observed in
the crystal structure to participate in a hydrogen
bond with the P3 Lys-NH\textsubscript{3}\textsuperscript{+} of an aldehyde
inhibitor bound in an unusual conformation but, as
this residue is poorly conserved amongst the
flaviviruses, the sidechain is probably not
important for substrate binding. Further away, the
cofactor residue Val75 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
Asn84 which appears to make a hydrogen bond
within the Flavivirus genus as either a polar or
negatively charged residue (Asn, Ser, Thr, Asp,
Glu).

Docking of 2-naphthoyl-\textit{p}NA generally resulted in two docking poses. One had
the bulky naphthoyl residue in the S1 pocket,
resulting in a non-extended conformation
reminiscent of the turn-like 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 (29). 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 minimisation to investigate possible
induced fit binding modes. Figure 2B shows one
minimised docking pose where the small S4
pocket has been enlarged by a subtle movement of
the side-chains of Ile155 and Val154 in NS3 and
Val75, Val77 and Leu87 of NS2B. After these
movements, the now slightly deeper S4 pocket is
additionally defined by two residues (Phe116 from
NS3, and Phe85 from NS2B) that make favourable
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 2.5-2.8Å 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
modelling 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 synthesised a library
of chromogenic \textit{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 Asn84 of the NS2B cofactor is within
hydrogen bonding distance of the P2 Lys of the
bound ligand. Asn84 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-N84
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 (SLE) 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 Yellow Fever (YF) a Glu at this
position in the cofactor always corresponds with
Arg at P2 in the native cleavage site. In Japanese
Encephalitis (JE), Murray Valley Encephalitis
(MVE), Zika virus (ZIKV) and Bussuquara virus
(BSQV) an Asp is matched by a P2 Lys
predominately 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 SLE
proteases. Despite Asn at the homologous position
in the cofactor for TBE and LGTV proteases, in all
cases they recognise Arg at P2 in the cleavage
sites of their native substrates, compared to Lys at P2 in substrates for SLE and WNV proteases. The tick-borne flaviviruses form a 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 Val154 and Met156 of NS3, and Leu87 of NS2B. We therefore prepared substrates based on Ac-XKKR-pNA, where X was a hydrophobic unnatural amino acid (Table 2, Figure 3; entries 1-9). Most of the examined substrates had comparable K_m and k_cat values to 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 (4, 6, 7). 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⁻¹s⁻¹). Taken together these results confirm the hypothesis that hydrophobic interactions are the key to interactions between P4 of substrates and S4 of enzyme.

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 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 efficacy by either 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 non-proteinogenic 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 the changes made in this panel of substrates (Figure 4, 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 homoaarginine, 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 homoaarginine, 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 mainchain (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 neighbouring 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 (23), 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 homoaarginine 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 two-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, V166L) and 12 additional recombinant proteases incorporating mutations in cofactor NS2B (V75A, V75F, N84A, N84D, N84E, N84L, N84S, Q86A, Q86E, Q86L, L87A, 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 which can be purified in high concentration (18). The purified mutant proteases were run on 12% SDS-PAGE (not shown) to confirm expression and purification of the desired 27kDa product (18).

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 optimised 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 further 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 2Naph-KKR-$p$NA (Table 4). Docking of the substrate 2Naph-KKR-$p$NA into the crystal structure of WNV protease yielded two possible substrate conformations. The $\beta$-strand substrate conformation orients the 2-naphthoyl group in a hydrophobic S4 subsite in the enzyme, bounded by NS3 residues V154, and I155 and the NS2B cofactor residues V75, V77 and L87, while 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, 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 3-fold decrease in $k_{cat}$. While 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 Phe87 may make a $\pi-\pi$ interaction that stabilises 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/E/S mutant proteases were tested against the tetrapeptide substrates Ac-LKKR-$p$NA and Ac-LKRR-$p$NA (Table 5). The comparative enzyme kinetic parameters show that the N84S mutation resulted in a four-fold decrease in $K_m$ when P2 was Lys, and a two-fold increase in $K_m$ when P2 was Arg. This supports observations that Asn84 (native to WNV) binds more strongly to a P2 Lys, while Ser84 (native to Den) binds more strongly to a P2 Arg. Against both substrates the N84S mutation caused a decrease in catalytic efficiency, approximately two-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 two-fold higher $K_m$ for Ac-LKRR-pNA and a 2-3 fold lower $K_m$ for Ac-LKRR-pNA, while the $k_{cat}$ was unaffected for the former substrate but about six-fold lower for the latter substrate. The overall effect was a two-fold decrease in catalytic efficiency for both substrates following this enzyme mutation. Comparison between the substrates reveals an almost four-fold higher $k_{cat}$ and a two-fold higher catalytic efficiency, $k_{cat}/K_m$ for Ac-LKRR-pNA, compared to Ac-LKRR-pNA indicating a preference for a P2 Lys (predominantly present in native cleavage sequences for JE, MVE, ZIKV, 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 trend was not the same when comparing substrate affinity as both mutants had slightly higher affinity for the Ac-LKRR-pNA substrate.

The N84S mutant protease was also tested against the truncated substrates 2Naph-KRR-pNA and 2Naph-K(hR)R-pNA in order to examine whether a P2 homoarginine residue could possibly be recognised with similar efficiency by either Asn84 or Ser84. Both the wild type protease and the N84S mutant had very similar $K_m$ values when assayed against the 2Naph-K(hR)R-pNA suggesting that a P2 homoarginine can bind to Asn84 or Ser84 with similar affinity, however the catalytic efficiency was decreased by over 20-fold. Taken together, these results support our earlier work (18, 30) 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 optimisation. Such approaches for Hepatitis C virus NS3 protease have led to increases in $IC_{50}$ values of over 1000-fold (31, 32). In this study we began with a tetrapeptide substrate and systematically replaced P4, P3 and P2 amino acids with peptidic and non-peptidic 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 the crystal structure.

A significant improvement in ligand affinity came from replacements of the “P5” acetyl cap and P4 hydrophobic Leu with the more rigid, non-peptidic, 2-naphthoyl group. This change enhanced substrate affinity, with $K_m$ increasing over two-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 non-peptidic 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 ($k_{cat}$), whilst having good turnover in the presence of Ac-Leu at P4, suggests that a double mutation to a non-peptidic 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 not so far 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 (23) which were more active against substrates with Arg instead of Lys at P2. This confirms predictions from native cleavage sequences and constitutes the first recognised difference in substrate preference between related flaviviruses. 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 which has a high affinity against multiple flavivirus proteases. The WNV protease had a substrate affinity ($K_m$) for a P2 homoarginine that was comparable to 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 - Ile155 in the crystal structure lies just beyond the rim of the S1 pocket, with docking of a substrate in the extended $\beta$-strand conformation suggesting that Ile155 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. Firstly, steric clashes may reduce substrate affinity, seen in the four-fold reduction in $K_m$ and a halving of $k_{cat}$. Secondly, 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. Thirdly, it is possible that the aromatic Phe side-chain participates in a $\pi$-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 (Val154, Met156) and NS2B (Leu87) in the crystal structure form a hydrophobic patch that likely constitutes S4. The hydrophobic character of these residues is 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 an 8-fold reduction, and even the conservative substitution NS3-V154L resulted 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 $K_m$ while 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. Modelling 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-pNA 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. Q86 is relatively poorly conserved within the Flavivirus genus but is predominantly polar or positively charged (Glu, Ser, Lys, Arg, 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-fold 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-N84. The Q86A/L mutants would disrupt this interaction and destabilise the region, whilst Q86E would maintain it.

The NS2B-V75 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-pNA but a large 27-fold loss in $K_m$ when assayed against the hexapeptide Ac-LQYTKR-pNA, supporting the prediction that NS2B-V75 makes an interaction with P5 or P6. Substitution with an aromatic Phe produced a comparable effect on both tetrapeptide and hexapeptide substrates, 5-fold and 4-fold losses in $K_m$ and 7-fold and 3-fold losses in $k_{cat}$, respectively. It is possible that NS2B-V75 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. While, the docking of the substrate 2Naph-KKR-pNA 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 $\beta$-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-pNA but dramatically affect $k_{cat}$, there are possibly favourable hydrophobic, aromatic-aromatic or $\pi-\pi$ 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 sidechains of the substrates are shortened or lengthened (Table 2, entries 10-30), modelling 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 N84 - 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 N84 were tested to further analyse 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 which 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, while 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 which 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-Naph-KhRR-pNA, with a P2 homoarginine residue could bind with similar affinity to either the wild type WNV protease or the NS2B-N84S mutant ($K_m = 26.7\pm3.7$ and $29.5\pm4.1$, respectively). This residue showed a high affinity that was similar to that for substrate with a P2 lysine binding to wild type enzyme ($K_m = 25.4 \pm 4.4$), better than that for substrate with a P2 arginine binding to wild type enzyme ($K_m = 43.4 \pm 4.6$) and only slightly worse than for a substrate with a P2 arginine binding to the mutant N84S protease ($K_m = 19.7 \pm 5.0$). Although this substrate is yet to be tested against Den NS2B/NS3protease, 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 in 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. Thr111 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 T69. T69L mutation caused a 4-fold reduction in $K_m$ and a 27-fold reduction in $k_{cat}$, much greater than a T69F mutation (3-fold reduction in $K_m$, 8-fold reduction in $k_{cat}$). This suggests that the hydrogen bond contributed by Thr111 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). Ile162 is located below the Val154 and Met156 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 Ala164 and Val166 are in the centre of the enzyme and their substitution may affect folding of the enzyme. Mutation of A164S/V substantially reduced $K_m$ (3-fold and 6-fold respectively) and $k_{cat}$ (6-fold 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-pNA.

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 towards development of substrate-based inhibitors of flavivirus proteases.

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

The abbreviations used are - AcK, , L-o-acetyllysine; Aib, 2-aminoisobutyric acid; Aoc, L-2-aminoocanoic acid; BSQV, Bussaquara virus; Cha, L-cyclohexylalanine; CHAPS, 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate; Chg, L-cyclohexylglycine; Cit, L-citrulline; DCM, dichloromethane; DQF-COSY, double-quantum filtered correlation spectroscopy; DIPEA, $N,N$-dissopropylethylamine; DMF, $N,N$-dimethylformamide; ESMS, electrospray mass spectrometry; hArg, L-homoarginine; HCV, Hepatitis C virus; HCTU, 1H-Benzotriazolium 1-[bis(dimethylamino)methylene]-5chloro-,hexafluorophosphonate (1-),3-oxide; FMOC, fluorenylmethyloxycarbonyl; HPLC, High performance liquid chromatography; IPTG, isopropyl -D-thiogalactopyranose; JE, Japanese Encephalitis virus; LGTV, Langat virus; MVE, Murray Valley Encephalitis; NH$_2$F, L-4-aminophenylalanine; Nle, L-norleucine; NO$_2$F, L-4-nitrophenylalanine; NMR, Nuclear Magnetic Resonance Spectroscopy; NOESY, Nuclear Overhauser Enhancement spectroscopy; Nva, L-norvaline; OD, optical density; Orn, L-ornithine; PAGE, Polyacrylamide Gel Electrophoresis; PCR, Polymerase Chain Reaction; PMSF, phenyl methyl sulfonyl fluoride; pNA, para-nitroanilide; 3-PyA, L-3-pyridylalanine; 4-PyA, L-4-pyridylalanine; R$_t$, retention time; SDS, Sodium Dodecyl Sulfate; TBE, Tick Borne Encephalitis; Tbg, L-tert-butylglycine; TCP, trityl chloride polystyrene; TFA, trifluoroacetic acid; TIPS, triisopropylsilane; TOCSY, total correlation spectroscopy; WNV, West Nile virus; YF, Yellow Fever virus; ZIKV, Zika virus.

REFERENCES

FIGURE LEGENDS

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 40aa cofactor domain required for catalytic activity. The construct CF40.gly.NS3pro comprises the essential 40aa domain of NS2B linked to the NS3 protease domain by a flexible glycine linker (Gly4,Ser,Gly4). 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.

Figure 2. Docking of substrates into the WNV NS2B/NS3 crystal structure. A. Active site of WNV NS2B/NS3 crystal structure (pdb code 2fp7). The Connolly surface of the enzyme is in grey, with the catalytic triad in white. The docked extended conformation of Ac-LKKR-pNA is represented by the green stick structure with nitrogen and oxygen atoms coloured blue and red respectively. The crystallographically determined pentapeptide inhibitor Bz-(Nle)KRR-H is shown in black. B. Docked and minimised extended conformation of naphthoyl-capped tripeptide substrate 2-Naphthoyl-KKR-pNA. Leucine 87 of the cofactor NS2B is shown in yellow; valine 154 of NS3 is shown in purple.

Figure 3. Natural and unnatural hydrophobic amino acids used in P4-mutant substrate synthesis.

Figure 4. Natural and unnatural amino acids used in (P1, P2, P3)-mutant substrate synthesis.
Table 1. Flavivirus Cleavage Sequences and Cofactor Homology.

<table>
<thead>
<tr>
<th></th>
<th>NS2B</th>
<th>2A/2B</th>
<th>2B/3</th>
<th>3/4A</th>
<th>4B/5</th>
<th>Homology to WNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>WNV</td>
<td>G</td>
<td>T</td>
<td>Y</td>
<td>T</td>
<td>E</td>
<td>76.3%</td>
</tr>
<tr>
<td>SLEV</td>
<td>G</td>
<td>N</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>67.4%</td>
</tr>
<tr>
<td>Den 1</td>
<td>T</td>
<td>M</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>67.4%</td>
</tr>
<tr>
<td>Den 2</td>
<td>G</td>
<td>S</td>
<td>N</td>
<td>K</td>
<td>K</td>
<td>66.5%</td>
</tr>
<tr>
<td>Den 3</td>
<td>G</td>
<td>T</td>
<td>R</td>
<td>Q</td>
<td>R</td>
<td>67.9%</td>
</tr>
<tr>
<td>Den 4</td>
<td>G</td>
<td>S</td>
<td>S</td>
<td>K</td>
<td>K</td>
<td>67.9%</td>
</tr>
<tr>
<td>YF</td>
<td>G</td>
<td>E</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>64.8%</td>
</tr>
<tr>
<td>JE</td>
<td>G</td>
<td>D</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>83.9%</td>
</tr>
<tr>
<td>MVE</td>
<td>G</td>
<td>D</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>85.3%</td>
</tr>
<tr>
<td>ZIKV</td>
<td>G</td>
<td>D</td>
<td>S</td>
<td>E</td>
<td>T</td>
<td>75.5%</td>
</tr>
<tr>
<td>BSQV</td>
<td>G</td>
<td>D</td>
<td>S</td>
<td>E</td>
<td>T</td>
<td>67.8%</td>
</tr>
<tr>
<td>LGTV</td>
<td>G</td>
<td>N</td>
<td>L</td>
<td>R</td>
<td>R</td>
<td>49.4%</td>
</tr>
<tr>
<td>TBE</td>
<td>G</td>
<td>N</td>
<td>M</td>
<td>R</td>
<td>R</td>
<td>48.8%</td>
</tr>
</tbody>
</table>

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 virus (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 while 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 40aa cofactor domain and the NS3 protease domain. Residues shown in green and yellow designate homology. The residue shown in bold 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.
Table 2. Enzyme kinetics of substrates modified at P4, P3, and P2.

<table>
<thead>
<tr>
<th>Substratea</th>
<th>Km (μM)</th>
<th>kcat (s⁻¹)</th>
<th>kcat/Km (M⁻¹s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P4 variants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Ac-Leu-Lys-Lys-Arg-pNA</td>
<td>54.6 ± 5.1</td>
<td>1.64 ± 0.04</td>
<td>30117 ± 2248</td>
</tr>
<tr>
<td>2 Ac-Nle-Lys-Lys-Arg-pNA</td>
<td>80.7 ± 6.7</td>
<td>2.35 ± 0.06</td>
<td>29091 ± 1831</td>
</tr>
<tr>
<td>3 Ac-Nva-Lys-Lys-Arg-pNA</td>
<td>67.3 ± 6.8</td>
<td>2.36 ± 0.08</td>
<td>34997 ± 2696</td>
</tr>
<tr>
<td>4 Ac-Chg-Lys-Lys-Arg-pNA</td>
<td>81.9 ± 8.8</td>
<td>1.78 ± 0.06</td>
<td>21735 ± 1799</td>
</tr>
<tr>
<td>5 Ac-Cha-Lys-Lys-Arg-pNA</td>
<td>69.9 ± 8.1</td>
<td>2.28 ± 0.04</td>
<td>32599 ± 2920</td>
</tr>
<tr>
<td>6 Ac-Aib-Lys-Lys-Arg-pNA</td>
<td>79.7 ± 6.2</td>
<td>1.60 ± 0.04</td>
<td>20050 ± 1186</td>
</tr>
<tr>
<td>7 Ac-Tbg-Lys-Lys-Arg-pNA</td>
<td>134.2 ± 8.4</td>
<td>1.74 ± 0.04</td>
<td>12961 ± 590</td>
</tr>
<tr>
<td>8 Ac-Aoc-Lys-Lys-Arg-pNA</td>
<td>75.5 ± 10.5</td>
<td>2.51 ± 0.11</td>
<td>33188 ± 3582</td>
</tr>
<tr>
<td>9 2-Naph-Lys-Lys-Arg-pNA</td>
<td>25.4 ± 4.4</td>
<td>1.08 ± 0.06</td>
<td>42603 ± 5472</td>
</tr>
<tr>
<td><strong>P3 variants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Ac-Leu-3Pya-Lys-Arg-pNA</td>
<td>88.7 ± 9.4</td>
<td>0.50 ± 0.02</td>
<td>5601 ± 419</td>
</tr>
<tr>
<td>11 Ac-Leu-4Pya-Lys-Arg-pNA</td>
<td>254.8 ± 19.5</td>
<td>0.90 ± 0.03</td>
<td>3536 ± 178</td>
</tr>
<tr>
<td>12 Ac-Leu-hArg-Lys-Arg-pNA</td>
<td>62.5 ± 5.8</td>
<td>1.18 ± 0.03</td>
<td>18960 ± 1374</td>
</tr>
<tr>
<td>13 Ac-Leu-nPhe-Lys-Arg-pNA</td>
<td>254.1 ± 21.0</td>
<td>0.72 ± 0.03</td>
<td>2817 ± 151</td>
</tr>
<tr>
<td>14 Ac-Leu-Cit-Lys-Arg-pNA</td>
<td>208.6 ± 16.7</td>
<td>1.44 ± 0.05</td>
<td>6904 ± 372</td>
</tr>
<tr>
<td>15 Ac-Leu-Orn-Lys-Arg-pNA</td>
<td>105.5 ± 7.7</td>
<td>1.79 ± 0.05</td>
<td>16999 ± 841</td>
</tr>
<tr>
<td>16 Ac-Leu-aPhe-Lys-Arg-pNA</td>
<td>141.3 ± 13.8</td>
<td>0.75 ± 0.03</td>
<td>5306 ± 369</td>
</tr>
<tr>
<td>17 2-Naph-Cit-Lys-Arg-pNA</td>
<td>21 ± 4.0</td>
<td>0.32 ± 0.02</td>
<td>15148 ± 2168</td>
</tr>
<tr>
<td>18 2-Naph-Orn-Lys-Arg-pNA</td>
<td>38 ± 5.7</td>
<td>0.98 ± 0.04</td>
<td>25743 ± 3184</td>
</tr>
<tr>
<td>19 2-Naph-3Pya-Lys-Arg-pNA</td>
<td>65.8 ± 5.1</td>
<td>0.96 ± 0.03</td>
<td>14617 ± 731</td>
</tr>
<tr>
<td>20 2-Naph-hArg-Lys-Arg-pNA</td>
<td>48 ± 11</td>
<td>0.11 ± 0.01</td>
<td>2294 ± 362</td>
</tr>
<tr>
<td><strong>P2 variants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 2-Naph-Lys-Cit-Arg-pNA</td>
<td>3423 ± 1306</td>
<td>0.43 ± 0.09</td>
<td>127 ± 23</td>
</tr>
<tr>
<td>22 2-Naph-Lys-Orn-Arg-pNA</td>
<td>36.8 ± 3.8</td>
<td>1.58 ± 0.05</td>
<td>42889 ± 3152</td>
</tr>
<tr>
<td>23 2-Naph-Lys-Arg-Arg-pNA</td>
<td>43.4 ± 4.6</td>
<td>0.89 ± 0.04</td>
<td>20382 ± 1338</td>
</tr>
<tr>
<td>24 2-Naph-Lys-hArg-Arg-pNA</td>
<td>26.7 ± 3.7</td>
<td>0.59 ± 0.03</td>
<td>22066 ± 1953</td>
</tr>
<tr>
<td>25 2-Naph-Lys-AcLys-Arg-pNA</td>
<td>2303 ± 544</td>
<td>0.30 ± 0.03</td>
<td>131 ± 17</td>
</tr>
<tr>
<td>26 2-Naph-Lys-aPhe-Arg-pNA</td>
<td>No activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 2-Naph-Lys-nLeu-Arg-pNA</td>
<td>No activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 2-Naph-Lys-Phe-Arg-pNA</td>
<td>No activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29 2-Naph-Orn-Orn-Arg-pNA</td>
<td>No activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 2-Naph-Lys-Cit-Arg-pNA</td>
<td>100.7 ± 8.4</td>
<td>0.57 ± 0.02</td>
<td>5612 ± 260</td>
</tr>
</tbody>
</table>

a Nle, norleucine; Nva, norvaline; Chg, cyclohexylglycine; Cha, cyclohexylalanine; Aib, aminoisobutyric acid; Tbg, tert-butylglycine; Aoc, 2-aminoctanoic acid; 2-Naph, 2-naphthyl; 3Pya, 3-pyridylalanine; 4Pya, 4-pyridylalanine; hArg, homoarginine; nPhe, p-nitrophenylalanine; Cit, citrulline; Orn, ornithine; 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.
### Table 3. Site-directed mutant enzyme kinetics

<table>
<thead>
<tr>
<th>Mutant</th>
<th>AcLKKR-pNA</th>
<th></th>
<th></th>
<th>AcLQYTKR-pNA</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_m$ (μM)</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>$k_{cat}/K_m$ (M$^{-1}$s$^{-1}$)</td>
<td></td>
<td>$K_m$ (μM)</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
</tr>
<tr>
<td>Wild Type</td>
<td>54.6 ± 5.1</td>
<td>1.644 ± 0.041</td>
<td>30117 ± 2248</td>
<td>199 ± 22</td>
<td>0.266 ± 0.007</td>
<td>1332 ± 118</td>
</tr>
<tr>
<td>NS3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T111F</td>
<td>169 ± 13</td>
<td>0.198 ± 0.005</td>
<td>1176 ± 65</td>
<td>376 ± 18</td>
<td>0.051 ± 0.001</td>
<td>135 ± 4</td>
</tr>
<tr>
<td>T111L</td>
<td>214 ± 27</td>
<td>0.060 ± 0.003</td>
<td>279 ± 25</td>
<td>617 ± 65</td>
<td>0.018 ± 0.001</td>
<td>30 ± 2</td>
</tr>
<tr>
<td>D129A</td>
<td>635 ± 49</td>
<td>0.048 ± 0.002</td>
<td>75 ± 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D129E</td>
<td>432 ± 18</td>
<td>0.036 ± 0.001</td>
<td>83 ± 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V154F</td>
<td>263 ± 39</td>
<td>0.057 ± 0.003</td>
<td>216 ± 21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V154L</td>
<td>190 ± 13</td>
<td>0.800 ± 0.018</td>
<td>4223 ± 210</td>
<td>344 ±28</td>
<td>0.245 ± 0.007</td>
<td>712 ± 40</td>
</tr>
<tr>
<td>I155F</td>
<td>207 ± 17</td>
<td>0.733 ± 0.021</td>
<td>3532 ± 196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M156A</td>
<td>164 ± 10</td>
<td>0.778 ± 0.015</td>
<td>4741 ± 210</td>
<td>879 ± 53</td>
<td>0.314 ± 0.010</td>
<td>357 ± 11</td>
</tr>
<tr>
<td>I162F</td>
<td>268 ± 28</td>
<td>0.164 ± 0.006</td>
<td>612 ± 43</td>
<td>416 ± 40</td>
<td>0.040 ± 0.002</td>
<td>95 ± 6</td>
</tr>
<tr>
<td>A164S</td>
<td>221 ± 23</td>
<td>0.260 ± 0.009</td>
<td>1178 ± 85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A164V</td>
<td>331 ± 15</td>
<td>0.181 ± 0.003</td>
<td>546 ± 16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V166L</td>
<td>96 ± 6</td>
<td>0.638 ± 0.009</td>
<td>6676 ± 309</td>
<td>232 ± 13</td>
<td>0.297 ± 0.005</td>
<td>1282 ± 52</td>
</tr>
<tr>
<td>NS2B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V75A</td>
<td>199 ± 10</td>
<td>0.173 ± 0.003</td>
<td>866 ± 30</td>
<td>5318 ± 507</td>
<td>0.164 ± 0.011</td>
<td>31 ± 1</td>
</tr>
<tr>
<td>V75F</td>
<td>272 ± 42</td>
<td>0.246 ± 0.017</td>
<td>905 ± 84</td>
<td>809 ± 105</td>
<td>0.091 ± 0.006</td>
<td>113 ± 8</td>
</tr>
<tr>
<td>N84A</td>
<td>376 ± 44</td>
<td>0.126 ± 0.007</td>
<td>336 ± 22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N84D</td>
<td>117 ± 5</td>
<td>1.532 ± 0.019</td>
<td>13109 ± 393</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N84E</td>
<td>79.3 ± 4.4</td>
<td>3.027 ± 0.047</td>
<td>38188 ± 1660</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N84L</td>
<td>220 ± 11</td>
<td>0.34 ± 0.006</td>
<td>1543 ± 52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N84S</td>
<td>205 ± 16</td>
<td>0.682 ± 0.019</td>
<td>3333 ± 183</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q86A</td>
<td>258 ± 18</td>
<td>0.367 ± 0.011</td>
<td>1424 ± 61</td>
<td>1203 ± 182</td>
<td>0.127 ± 0.011</td>
<td>105 ± 7</td>
</tr>
<tr>
<td>Q86E</td>
<td>77.7 ± 7.3</td>
<td>0.868 ± 0.018</td>
<td>11182 ± 863</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q86L</td>
<td>148 ± 12</td>
<td>0.973 ± 0.024</td>
<td>6586 ± 381</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L87A</td>
<td>434 ± 39</td>
<td>0.070 ± 0.003</td>
<td>161 ± 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L87F</td>
<td>144 ± 12</td>
<td>1.342 ± 0.033</td>
<td>9345 ± 589</td>
<td>764 ± 53</td>
<td>0.228 ± 0.008</td>
<td>299 ± 11</td>
</tr>
</tbody>
</table>

Enzyme kinetics were obtained using the in vitro enzyme assay and Ac-LKKR-pNA and Ac-LQYTKR-pNA as substrates with each data point representing the mean of triplicate measurements ± S.E.
Table 4. Site-directed mutant enzyme kinetics against 2-Naph-Lys-Lys-Arg-pNA

<table>
<thead>
<tr>
<th>Mutant</th>
<th>2-Naph-KKR-pNA</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kₘ (μM)</td>
<td>kₖₛₙ (s⁻¹)</td>
<td>kₖₛₙ/Kₘ (M⁻¹s⁻¹)</td>
</tr>
<tr>
<td>Wild Type</td>
<td>25.4 ± 4.4</td>
<td>1.083 ± 0.055</td>
<td>42603 ± 5472</td>
</tr>
<tr>
<td>NS3-V154F</td>
<td>97.7 ± 6.9</td>
<td>0.0431 ±0.0013</td>
<td>442 ± 19</td>
</tr>
<tr>
<td>NS2B-L87A</td>
<td>88.4 ± 16.0</td>
<td>0.0478 ±0.0030</td>
<td>541 ± 70</td>
</tr>
<tr>
<td>NS2B-L87F</td>
<td>36.9 ± 3.6</td>
<td>0.288 ± 0.008</td>
<td>7796 ± 580</td>
</tr>
</tbody>
</table>

Enzyme kinetics were obtained using the in vitro enzyme assay against the substrate 2-Naph-KKR-pNA with each data point representing the mean of triplicate measurements ± S.E.
Table 5. Comparison of wild type, NS2B-N84S, NS2B-N84D, and NS2B-N84E enzyme kinetics

<table>
<thead>
<tr>
<th>Mutant</th>
<th>AcLKKR-pNA</th>
<th>AcLKRR-pNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Km (μM)</td>
<td>kcat (s⁻¹)</td>
</tr>
<tr>
<td>Wild Type</td>
<td>54.6 ± 5.1</td>
<td>1.644 ± 0.041</td>
</tr>
<tr>
<td>NS2B-N84S</td>
<td>205 ± 16</td>
<td>0.682 ± 0.019</td>
</tr>
<tr>
<td>NS2B-N84D</td>
<td>117 ± 5</td>
<td>1.532 ± 0.019</td>
</tr>
<tr>
<td>NS2B-N84E</td>
<td>79.3 ± 4.4</td>
<td>3.027 ± 0.047</td>
</tr>
</tbody>
</table>

Enzyme kinetics were obtained using the in vitro enzyme assay against the Ac-LKKR-pNA, Ac-LKRR-pNA, 2Naph-KRR-pNA and 2Naph-KhRR-pNA substrates. Each data point represents the mean of triplicate measurements ± S.E.
Figure 1
Figure 2
Figure 3

Leucine (Leu)  Norleucine (Nle)  Norvaline (Nva)  Cyclohexylglycine (Chg)

Cyclohexylalanine (Cha)  2-Aminoisobutyric acid (Aib)  t-Butylglycine (Tbg)

2-Aminooctanoic acid (Aoc)  2-Naphthoyl
Figure 4

- Lysine (Lys)
- Acetyl lysine (AcLys)
- Ornithine (Orn)
- Arginine (Arg)
- Homoarginine (hArg)
- Citrulline (Cit)
- Phenylalanine (Phe)
- p-Aminophenylalanine (aPhe)
- p-Nitrophenylalanine (nPhe)
- 3-Pyridylalanine (3Pya)
- 4-Pyridylalanine (4Pya)
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

J. Biol. Chem. published online October 19, 2006

Access the most updated version of this article at doi: 10.1074/jbc.M607641200

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

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

Supplemental material:
http://www.jbc.org/content/suppl/2006/10/19/M607641200.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2006/10/19/jbc.M607641200.citation.full.html#ref-list-1