Structure of a murine norovirus NS6 protease-product complex revealed by adventitious crystallisation

Eoin N. Leen, Gabriela Baeza and Stephen Curry*

Department of Life Sciences, Imperial College, London, SW7 2AZ, United Kingdom.

*Corresponding author: s.curry@imperial.ac.uk

Murine noroviruses have emerged as a valuable tool for investigating the molecular basis of infection and pathogenesis of the closely related human noroviruses, which are the major cause of non-bacterial gastroenteritis. The replication of noroviruses relies on the proteolytic processing of a large polypeptide precursor into six non-structural proteins (NS1-2, NS3, NS4, NS5, NS6pro, NS7pol) by the virally-encoded NS6 protease. We report here the crystal structure of MNV NS6pro, which has been determined to a resolution of 1.6 Å. Adventitiously, the crystal contacts are mediated in part by the binding of the C-terminus of NS6pro within the peptide-binding cleft of a neighbouring molecule. This insertion occurs for both molecules in the asymmetric unit of the crystal in a manner that is consistent with physiologically-relevant binding, thereby providing two independent views of a protease-peptide complex. Since the NS6pro C-terminus is formed in vivo by NS6pro processing, these crystal contacts replicate the protease-product complex that is formed immediately following cleavage of the peptide bond at the NS6-NS7 junction. The observed mode of binding of the C-terminal product peptide yields new insights into the structural basis of NS6pro specificity.

Introduction

NOROVIRUSES are members of the calicivirus family of positive sense RNA viruses. In humans noroviruses cause rapid onset diarrhoea and vomiting, a condition often referred to as gastric flu. Noroviral gastroenteritis is estimated to affect 21 million people annually in the United States [1], and to be responsible for up to 200,000 deaths a year in developing nations [2]. Despite the obvious medical importance of norovirus infections, there are still no vaccines or antiviral treatments.

One of the reasons for the slow progress in drug and vaccine development has been the lack of cell culture and small animal model systems for studying the molecular mechanisms and pathology of noroviral gastroenteritis. These limitations have been partly overcome thanks to the relatively recent development of cell culture [3], and reverse genetics techniques for murine norovirus (MNV) [4-6]. MNV is closely related to human noroviruses and some strains are known to cause gastroenteritis in STAT1 (-/-) mice [7]. The virus has therefore emerged as an important surrogate model for studying human norovirus infections at the molecular level.

The Norovirus genome is approximately 7.5 kb in length and encodes three open reading frames (ORFs). ORF2 and ORF3 encode the major and minor capsid proteins respectively [8,9], while ORF1 encodes a non-structural polypeptide that is cleaved into functional units by the viral protease (NS6pro) [10-14]. There are five cleavage junctions in the norovirus polypeptide, so processing ultimately releases six cleavage products, all of which play essential roles in the intracellular replication cycle of the virus [12,14]. These include NS1-2, a protein that disrupts cellular trafficking [15]; NS3, an ATPase and putative helicase [16]; NS4, a protein that also interferes with cellular protein trafficking within infected cells [17]; NS5, also known as VPg, the genome-linked protein that has key roles in the translation and replication of the viral RNA genome [18]; NS6pro, the viral protease itself [19]; and NS7pol, an RNA-dependent RNA polymerase [20].

Although there is some variation in the sequence of the five cleavage sites recognised by norovirus NS6pro, common features are identifiable. The amino acids on the N-terminal side of the cleaved peptide bond (the P1 position in the nomenclature of Schechter and Berger [21]) are typically glutamine or glutamic acid, while the P1’ amino acids on the C-terminal side is usually glycine or alanine; the

*Corresponding author: s.curry@imperial.ac.uk
residues at the P2 and P4 positions tend to be large and hydrophobic in nature [12,14].

X-ray crystallographic analysis of the NS6 protease from three different strains of human norovirus [19,22,23] has revealed the protein to be a cysteine protease with a chymotrypsin-like fold, similar to that of the 3C proteases from picornaviruses [24,25]. The NS6pro structure is composed of two β-barrel domains tightly packed against each other; the interface between these two domains forms the peptide-binding cleft at the centre of which is the protease active site, a catalytic triad consisting of cysteine, histidine and aspartic or glutamic acid [22]. While the catalytic mechanism of NS6pro and 3Cpro enzymes has not been fully defined, it has been shown to involve all three amino acids of the triad [22,26-28] and is therefore likely to resemble the mechanism of the serine proteases.

In one of these crystallographic studies, the structure of human norovirus NS6pro was determined in complex with a peptide-like inhibitor [23]. In this structure the N-terminal half of the peptide (positions P5 to P1) was covalently attached — by the action of a C-terminal Michael acceptor — to the cysteine nucleophile in the active site of the enzyme. Although the mode of binding was similar to that observed for other chymotrypsin-like proteases, the chemical modifications of the peptide that were made to add the Michael acceptor to the inhibitor resulted in ionic clashes that distorted the protease active site. In this report we extend the structural work on noroviral proteases by describing the X-ray structure of full-length MNV NS6pro. Fortuitously the structure has been solved in a crystal form in which the C-terminus of one protease extends into the active site of another. This provides new insights into peptide recognition by MNV NS6pro.

Results and discussion

To prevent autolysis of the protease in the high concentrations used for crystallisation trials MNV NS6pro was inactivated by mutation of the active site nucleophile to alanine (C139A). This mutant protease was expressed in E. coli and purified using affinity and size exclusion chromatography as described in the Materials and Methods. The protein crystallised by hanging drop vapour diffusion in 20% w/v PEG 3350, 0.2 M KSCN, 0.1 M Bis-Tris propane pH 7.5 to yield crystals that belong to space-group C221 and have two molecules in the asymmetric unit. They diffracted X-rays to 1.6 A resolution, producing an electron density map of extremely good quality for residues 1-123 and 132-183 in both molecules of the asymmetric unit. However, electron density for the solvent-exposed loop formed by residues 124-131 was weak and so was not included in the NS6pro model. The evident flexibility of this loop is consistent with previous crystallographic analyses of human norovirus NS6 proteases which found the loop to be either disordered, as in the Norwalk virus NS6pro structure [22], or to adopt very dissimilar conformations, as in the structures of Chiba virus (CV) and Southampton virus (SV) NS6pro [19,23]. The final refined crystal structure of MNV NS6pro has an Rfree value of 20.8% and very good stereochemistry (full data collection and refinement statistics are given in Table 1).

The overall structure of the MNV NS6 protease is nearly identical to Norwalk virus, SV and CV NS6pro (Fig. 1), which comes as no surprise given the high degree of amino acid identity (~60%) between these proteins. As described previously, the structure is an abbreviated form of the chymotrypsin-like 3C proteases from picornaviruses [19,22], which comprise two β-barrel domains. Specifically, in norovirus NS6 proteases, the β-strands on one side of the N-terminal β-barrel are so much shorter that the domain is better considered as a single β-sheet decorated by loops (Fig. 1A). The active site triad of H30, D54, and C139...
(A139 in our mutated protein) is located in the centre of the peptide-binding cleft formed at the interface of these two domains.

In the norovirus NS6pro crystal structures published to date it has already been observed that, whereas the N-terminal helix found in picornavirus 3C proteases is retained in NS6pro, the C-terminal helix in 3Cpro is not found in the norovirus proteases [19,22,23]. Instead the C-terminal polypeptide of NS6pro appears to be largely unstructured; it is either disordered or adopts very different conformations that appear to be determined by crystal contacts (Fig. 1B). Strikingly, in our structure the flexibility of the C-terminal polypeptide has actually allowed it to associate with the peptide-binding sites of a neighbouring molecule in the crystal — a happy accident that has revealed the structure of an MNV NS6pro-peptide complex (Fig. 2A). The interaction occurs for both molecules of the asymmetric unit, although

| Table 1: Crystallographic data collection and model refinement statistics for MNV NS6pro |
|---------------------------------|-----------------|
| Data collection                 |                 |
| Space-group                     | C222_1          |
| a, b, c (Å)                     | 50.9, 70.4, 192.3 |
| α, β, γ (°)                     | 90              |
| Resolution range (Å)            | 48.03-1.58      |
| No. of independent reflections  | 44,988          |
| Multiplicity^1                  | 12.4 (8.4)      |
|Completeness (%)                 | 94.2 (65.3)     |
|I/σ₁                            | 29.8 (3.2)      |
|Rmerge (%)^2                     | 4.0 (55.0)      |

| Model refinement               |                 |
| No. of Non-hydrogen atoms/      | 2611/290        |
| waters                         |                 |
|R_work (%)^3                    | 18.4 (26.6)     |
|R_free (%)^4                    | 20.8 (30.3)     |
|Average B-factor                | 32.1            |
|RMS deviations - Bonds (Å)^5     | 0.006           |
|RMS deviations - Angles (°)     | 0.980           |
|Ramachandran plot               |                 |
|Favoured                        | 98.3            |
|Allowed                         | 1.7             |
|PDB Accession Code              | 4ash            |

^1Values for highest resolution shell given in parentheses
^2Rmerge = 100 × Σj(I(hkl) − <I(hkl)>)/ΣjΣI(hkl), where I(hkl) and <I(hkl)> are the intensity of measurement j and the mean intensity for the reflection with indices hkl, respectively.
^3R_work = 100 × Σhkl|Fobs| − |Fcalc||Σhkl|Fobs|.
^4R_free is the R_model calculated using a randomly selected 5% sample of reflection data that were omitted from the refinement.
^5RMS, root-mean-square; deviations are from the ideal geometry defined by the Engh and Huber parameters [45].

(A139 in our mutated protein) is located in the centre of the peptide-binding cleft formed at the interface of these two domains.

In the norovirus NS6pro crystal structures published to date it has already been observed that, whereas the N-terminal helix found in picornavirus 3C proteases is retained in NS6pro, the C-terminal helix in 3Cpro is not found in the norovirus proteases [19,22,23]. Instead the C-terminal polypeptide of NS6pro appears to be largely unstructured; it is either disordered or adopts very different conformations that appear to be determined by crystal contacts (Fig. 1B). Strikingly, in our structure the flexibility of the C-terminal polypeptide has actually allowed it to associate with the peptide-binding sites of a neighbouring molecule in the crystal — a happy accident that has revealed the structure of an MNV NS6pro-peptide complex (Fig. 2A). The interaction occurs for both molecules of the asymmetric unit, although
in each case the C-terminus extends from the body of the protease in a different direction towards its neighbour (Fig. 2B). For chain A, the interaction is symmetric: it donates its C-terminus to a neighbouring A-chain that is related by crystallographic two-fold symmetry and binds the C-terminus received from this molecule. This pair of molecules thus forms a closed complex. In contrast, chain B donates its C-terminus to a neighbouring B-chain but binds the C-terminus of a third molecule (also a B-chain). Despite these differences, the interactions made with the neighbouring protease by the residues corresponding to the P4-P1 amino acids of the donated C-terminus are in each case essentially identical (Fig. 2C).

Superposition of the crystal structure of Southampton Virus NS6pro with a covalently-attached peptide-like inhibitor shows that the positions of the P4-P1 amino acids are very similar to our MNV NS6pro ‘co-crystal’ structure [23] (Fig. 3A, C). The positions of the P4-P1 amino acids observed in the peptide-binding site of MNV NS6pro are also very similar to the equivalent residues in the co-crystal structure of FMDV 3Cpro:peptide complex [29] (not shown). This indicates that, although the formation of an MNV protease-peptide complex is an accident of crystallisation, crystal-packing constraints do not prevent the C-terminal residues from adopting a physiologically-relevant conformation in the binding site of a neighbouring molecule.

The adventitious insertion of the NS6pro C-terminus into the peptide-binding cleft of neighbouring proteases in our crystals has provided us with a detailed view of a structure that corresponds to the protease-product complex of MNV NS6pro that forms following cleavage of the NS6-NS7 junction in the viral polyprotein. All of the amino acids on the P-side of the cleavage junction are visible in the electron density map and are included in the model. The primary contacts with the protease are made by the carboxylate group of P1-Gln and side-chains of P1-Gln, P2-Phe and P4-Leu. Although the P5 and P6 amino acids make some contact with the protease, their positions differ for molecules A and B in the asymmetric unit. These differences are likely to be due to differences in crystal packing, so our structure does not give a clear indication that they are important for peptide recognition.

The co-crystal structure of SV NS6pro with a P5-P1 (EFQLQ) peptide-based inhibitor covalently attached to the active site nucleophile [23] reveals a conformation for the bound peptide over the region P1-P4 that is very similar to the MNV NS6pro-peptide complex, although there are some differences in detail caused by sequence variation and a notable structural alteration at the active site due to the chemical modification of the peptide inhibitor needed to introduce the reactive Michaelis acceptor (see below).
P1 residue recognition by MNV NS6\textsuperscript{pro} is mediated primarily by a specificity triad of residues composed of H157, T134 and Y143. The side-chain carboxyl of the P1-Gln in the peptide (Q183) forms hydrogen bonds with the imidazole ring of H157 and the T134 hydroxyl group (Fig. 3A); in turn, H157 is stabilised by a hydrogen bond from the hydroxyl group of Y143. This mode of interaction could be replicated by a P1-Glu residue and accounts for the ability of NS6\textsuperscript{pro} to cleave substrates with P1-Gln or P1-Glu. The specificity triad, which is also found in human norovirus NS6\textsuperscript{pro} structures, is also conserved in picornaviral 3C\textsuperscript{pro} enzymes from poliovirus, human rhinovirus, coxsackie A virus (CAY) and FMDV [24,30-32] and explains the ability to cleave after P1-Gln or P1-Glu, although only FMDV is exceptional in not having a strong preference for P1-Gln [29,33].

The P1-Gln of the peptide 'product' is also stabilised by interactions made by its carboxylate group with the main-chain amides of G137 and A139 (the oxanion hole), and with the ε amine group of H30 (Fig. 3A). An essentially identical set of interactions was observed to stabilise the carboxylate group of the P1 residue in the co-crystal structure of the picornavirus CAY 3C\textsuperscript{pro} in complex with a peptide corresponding to the N-terminal half of the cleavage product [30] (Fig. 3F), and in the crystal structure of the Tobacco Etch virus 3C-like Nla protease where, as in our structure, the C-terminus of the protease is observed to be bound in the active site (although in the case of Nla crystals it is not known whether this is a cis or trans insertion) [34]. In contrast, this mode of interaction did not occur in the complex of SV NS6\textsuperscript{pro} with a peptide-like inhibitor because of the non-standard nature of the C-terminus of the SV peptide [23]; in that case the carboxylate group at the end of the peptide is separated from the Cα of the P1 residue by an additional C-C bond, which has the effect of distorting the geometry of the SV NS6\textsuperscript{pro} site (Fig. 3C). In particular, the side chain of H30 in the catalytic triad of SV NS6\textsuperscript{pro} structure is almost identical to that of the unliganded Norwalk virus protease, albeit allowing for the fact that the general acid in Norwalk virus NS6\textsuperscript{pro} catalytic triad is a Glu but an Asp in MNV (Fig. 1A, B) [19,22].

The side-chain of F182 at the P2 in the MNV NS6\textsuperscript{pro} structure is sandwiched between H30 and I109 at the entrance to the hydrophobic S2 pocket (Fig 3A,B) but also contacts V114, V158, A159 and D54. The largely apolar nature of the S2 pocket accounts for the general selectivity for hydrophobic amino acids at the P2 amino acid in norovirus cleavage junctions [14] (Fig. 4). An exception to this is the NS4/NS5 junction of the MNV polyprotein, which has a P2-Ser. It seems unlikely that at the side-chain hydroxyl of this amino acid could interact specifically with the S2 pocket, suggesting that the NS4/NS5 junction may be a sub-optimal substrate for the protease. Intriguingly, the S2 pocket is more enclosed in human norovirus NS6\textsuperscript{pro} structures because the Q110-G111-R112 sequence at the top of the loop that forms one flank of the pocket is replaced in MNV NS6\textsuperscript{pro} by the G110-S111-A112 sequence, which has much smaller amino acids (Fig. 3A-D).

The side-chain of P3-Glu (E181) in the MNV NS6\textsuperscript{pro}-peptide complex makes no specific interactions with the protease, though it's position is stabilised by β-strand-like interactions between its main-chain amide and carbonyl groups and the main-chain surrounding A160 (Fig. 3A), a mode of binding that is common to chymotrypsin-like proteases. The absence of any interaction with the P3 side-chain explains the diversity of residues (Q|G|H|N|E) observed at this position in MNV cleavage junctions [14], a feature that is also shared by picornavirus 3C\textsuperscript{pro} cleavage junctions [29,30].

In contrast, at the P4 position in the 'peptide' the side-chain of L180 occupies the hydrophobic S4 specificity pocket, which is made up of T166, T161, I109, A159, 1168 and 1107. This mode of interaction is very similar to the binding of apolar P4 amino acids observed in structures of SV NS6\textsuperscript{pro} [23] and picornaviral 3C proteases [29,35], reflecting the general — but not universal — conservation of hydrophobic functional groups at P4 in specific substrates for these enzymes [36]. In MNV cleavage junctions there is a strong preference for large hydrophobic residues at P4 (W|Y|F|L), all of which could be accommodated in the deep, apolar S4 pocket.
Beyond the P4 position there appears to be little specific interaction between the peptide and the protease. In MNV NS6<sup>pro</sup> the side chain of the P5 residue (A179) does not contact the protease. This finding echoes the lack of observable contact between FMDV 3C<sup>pro</sup> and the P5 amino acid of its peptide substrate [29]. Moreover, the high diversity of amino acid at P5 in MNV cleavage junctions (E|D|W|K|A) is consistent with the notion that this residue is not a determinant of specificity (Fig. 4). Although the P5-Glu in the co-crystal structure of SV NS6<sup>pro</sup> with a peptide-like inhibitor was observed to interact with the main chain amide of Q109 via a water molecule (Fig 3C), this could be an artefact of crystallisation [23]. While it remains true that efficient catalysis by NS6<sup>pro</sup> requires substrates that have at least five residues on the P-side of the scissile bond — since removal of the P5 residue from a synthetic substrate for SV NS6<sup>pro</sup> resulted in a nine-fold decrease in the k<sub>cat</sub>/K<sub>M</sub> specificity constant [23] — there is no evidence that residues beyond P4 contribute to specificity.

It is of interest to note that, despite the observation in the crystal structure of numerous specific contacts between the P4-P1 residues and the protease, comparative analysis of the proteolysis of mutated NS6<sup>pro</sup> cleavage sites in the polyprotein has tend to suggest that there are relatively few strong determinants of cleavage specificity [28,37]. Although these studies reported tolerance of substitutions at positions P4, P2, P1 and P1', this interpretation is based on analyses of self-processing of polypeptide precursors at a single time point following translation in rabbit reticulocyte lysates [28] or following over-expression in E. coli [37]. Such techniques are lacking in sensitivity. For example, our previous work with FMDV 3C<sup>pro</sup> showed that analysis of polypeptide cleavage rates by over-expression in E. coli could not distinguish from wild-type a mutant that in vitro assays (using purified protein and synthetic peptide substrates) had only 1% of wild-type activity [26]. We suspect therefore, that consistent with the crystal structure reported here (Fig. 3) and broad patterns of sequence conservation (Fig. 4), NS6<sup>pro</sup> does indeed discriminate between amino acids at the P1, P2 and P4 positions. The contribution to specificity of residues on the P’ side of the scissile bond remains to be fully investigated.

In summary, crystallisation of MNV NS6<sup>pro</sup> has provided an unexpected insight into the nature of protease-peptide interactions because of the insertion of the C-terminus of the protein (a product of NS6<sup>pro</sup> cleavage) into the peptide binding cleft of neighbouring molecules in the crystal. Crystallisation has in effect captured a snapshot of the covalent intermediate that forms during catalysis. The protease-peptide interactions revealed in this structure provide a basis for the engineering of specific NS6<sup>pro</sup> inhibitors.

It has occurred to us that by extending the C-terminus of MNV NS6<sup>pro</sup> to incorporate the sequence of the NS6-NS7 cleavage junction, or to modify the sequence to that of other cleavage junctions, it may be possible to exploit the same crystal form to obtain structure of NS6<sup>pro</sup>-substrate complexes. This would permit examination of protease-peptide interactions on the P’ side of the cleavage junction and work towards this goal is now in progress in our laboratory.

### Materials and Methods

cDNA for MNV1 strain CW1 NS6<sup>pro</sup> (a kind gift from Dr Ian Goodfellow — GenBank accession number YP724460) was amplified by PCR and ligated into the expression vector pETM11 [38], which contains a thrombin cleavable N-terminal 6x polyhistidine tag. QuikChange mutagenesis (Stratagene) was used to introduce a C139A mutation into NS6<sup>pro</sup> to knock out the active site nucleophile of the enzyme. The primers used in cloning and mutagenesis are given in Table 2. The protein was expressed in BL21 (DE3) pLysS Escherichia coli cells by induction with 1 mM IPTG for 3 hours at 37°C. The protein was purified on TALON resin (Clontech); the polyhistidine tag was removed by overnight incubation using approximately 10 μU of thrombin (Sigma-Aldrich) per milligram of purified protein. The cleaved protein was further purified by size exclusion chromatography using a Hi-Load 16/60 Superdex 75 column (GE healthcare). Peak fractions were concentrated to 750 μM in 25 mM Tris.HCl, pH 8.0 containing 200 mM NaCl and 5 mM Dithiothreitol. The purified protein contained 185 residues comprising a non-native GS sequence at the N-terminus, followed by all 183 residues of MNV NS6<sup>pro</sup>.

Crystals were obtained using the sitting drop vapour diffusion method by mixing 2 μL of protein solution with 1 μL of reservoir solution. This was allowed to equilibrate over 125 μL reservoir solution at room temperature. The reservoir solution was composed of 0.2 M KSCN, 0.1 M BisTris propane pH 7.5, 20% w/v poly-ethylene glycol glycol (PEG) 3350 (condition 64 in the Qiagen PACT suite of crystallisation buffers). Crystals were transferred into a cryobuffer containing 0.2 M KSCN, 0.1 M Bis-Tris propane pH 7.5, 30% w/v PEG 3350 prior to freezing in liquid nitrogen. (MNV NS6<sup>pro</sup> crystals were also obtained in 30% w/v PEG 4000, 0.2 M MgCl<sub>2</sub>, 0.1 M Tris.HCl pH 8.5; 10% w/v PEG 8000, 8% w/v ethylene glycol, 0.1 M HEPES, pH 7.5; 20% w/v PEG 3350, 0.2 M sodium iodide, 0.1 M Bis-Tris-propane pH 6.5; and 20% w/v PEG 3350, 0.2 M sodium iodide, 0.1 M Bis-Tris-propane pH 8.5, but have not been tested for diffraction.)

X-ray diffraction data were collected at the Diamond Light Source on the I04-1 beamline (wavelength 0.917 Å) using a PILATUS 2M detector. Data were indexed and scaled using the XDS software package [39] and phased by molecular replacement with Phaser using a weighted ensemble of the three known human norovirus protease structures as a search model [19,22,23,40]. Refinement and model manipulation were performed using PHENIX and Coot respectively [41,42]. Structure validation was performed using MolProbity [43].

### Table 2: Cloning and C139A mutagenic primers. Restriction sites used in cloning are underlined. Mutations introduced using QuikChange mutagenesis are in boldface.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNV Vpg 1-183 F</td>
<td>CATCATGGATCCGCCCCAGTCTCCATCTTG</td>
</tr>
<tr>
<td>MNV Vpg 1-183 R</td>
<td>CATCATAGCTTACCTGAACTCAGAGCTTC</td>
</tr>
<tr>
<td>C139A QuikChange F</td>
<td>CTCGGGACATCCCAAGGCGAGCCAGGTCGAGC</td>
</tr>
<tr>
<td>C139A QuikChange R</td>
<td>CTTATAAACATAGGGACAGCCGTCGGTCCCAG</td>
</tr>
</tbody>
</table>

PlO S ONE (2012) 7(6): e38723

Leen et al.

Structure of a Murine Norovirus NS6 Protease-Product Complex
Acknowledgements
We thank Dr Ian Goodfellow for kindly providing cDNA for MNV1 strain CW1. We are grateful to the staff of beam-line I04-1 at the Diamond Light Source, Didcot, UK for assistance with X-ray data collection.

References


