

Structure determination of Murine Norovirus NS6 proteases with C-terminal extensions designed to probe protease-substrate interactions

Humberto Fernandes, Eoin N Leen, Hamlet Cromwell Jnr, Marc-Philipp Pfeil, Stephen Curry

Noroviruses are positive-sense single-stranded RNA viruses. They encode an NS6 protease that cleaves a viral polyprotein at specific sites to produce mature viral proteins. In an earlier study we obtained crystals of murine norovirus (MNV) NS6 protease in which crystal contacts were mediated by specific insertion of the C-terminus of one protein (which contains residues P5-P1 of the NS6-7 cleavage junction) into the peptide binding site of an adjacent molecule, forming an adventitious protease-product complex. We sought to reproduce this crystal form to investigate protease-substrate complexes by extending the C-terminus of NS6 construct to include residues on the C-terminal (P') side of the cleavage junction. We report the crystallization and crystal structure determination of inactive mutants of murine norovirus NS6 protease with C-terminal extensions of one, two and four residues from the N-terminus of the adjacent NS7 protein (NS6 1', NS6 2', NS6 4'). We also determined the structure of a chimeric extended NS6 protease in which the P4 P4' sequence of the NS6-7 cleavage site was replaced with the corresponding sequence from the NS2-3 cleavage junction (NS6 4' 2|3).

The constructs NS6 1' and NS6 2' yielded crystals that diffracted anisotropically. We found that, although the uncorrected data could be phased by molecular replacement, refinement of the structures stalled unless the data were ellipsoidally truncated and corrected with anisotropic B-factors. These corrections significantly improved phasing by molecular replacement and subsequent refinement.

The refined structures of all four extended NS6 proteases are very similar in structure to the mature MNV NS6 — and in one case reveal additional details of a surface loop. Although the packing arrangement observed showed some similarities to those observed in the adventitious protease-product crystals reported previously, in no case were specific protease-substrate interactions observed.

Structure determination of Murine Norovirus NS6 proteases with C-terminal extensions designed to probe protease-substrate interactions

Humberto Fernandes, Eoin N. Leen, Hamlet Cromwell Jr, Marc-Philipp Pfeil, Stephen Curry

Noroviruses are positive-sense single-stranded RNA viruses. They encode an NS6 protease that cleaves a viral polyprotein at specific sites to produce mature viral proteins. In an earlier study we obtained crystals of murine norovirus (MNV) NS6 protease in which crystal contacts were mediated by specific insertion of the C-terminus of one protein (which contains residues P5-P1 of the NS6-7 cleavage junction) into the peptide binding site of an adjacent molecule, forming an adventitious protease-product complex. We sought to reproduce this crystal form to investigate protease-substrate complexes by extending the C-terminus of NS6 construct to include residues on the C-terminal (P') side of the cleavage junction. We report the crystallization and crystal structure determination of inactive mutants of murine norovirus NS6 protease with C-terminal extensions of one, two and four residues from the N-terminus of the adjacent NS7 protein (NS6 1', NS6 2', NS6 4'). We also determined the structure of a chimeric extended NS6 protease in which the P4-P4' sequence of the NS6-7 cleavage site was replaced with the corresponding sequence from the NS2-3 cleavage junction (NS6 4' 2/3).

The constructs NS6 1' and NS6 2' yielded crystals that diffracted anisotropically. We found that, although the uncorrected data could be phased by molecular replacement, refinement of the structures stalled unless the data were ellipsoidally truncated and corrected with anisotropic B-factors. These corrections significantly improved phasing by molecular replacement and subsequent refinement.

The refined structures of all four extended NS6 proteases are very similar in structure to the mature MNV NS6 — and in one case reveal additional details of a surface loop. Although the

packing arrangement observed showed some similarities to those observed in the adventitious protease-product crystals reported previously, in no case were specific protease-substrate interactions observed.

Authors

*Humberto Fernandes**

Department of Life Sciences, Sir Ernst Chain Building, Imperial College London, SW7 2AZ, UK. Present address: Laboratory of Genomic Engineering, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5A, 02-106 Warsaw, Poland

Eoin N. Leen

Department of Life Sciences, Sir Ernst Chain Building, Imperial College London, SW7 2AZ, UK.

Hamlet Cromwell Jr

Department of Life Sciences, Sir Ernst Chain Building, Imperial College London, SW7 2AZ, UK.

Marc-Philipp Pfeil

Department of Life Sciences, Sir Ernst Chain Building, Imperial College London, SW7 2AZ, UK. Present address: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Stephen Curry

Department of Life Sciences, Sir Ernst Chain Building, Imperial College London, SW7 2AZ, UK.

*Corresponding author.

Tel/Fax: +48 22 592 1310

E-mail: hfernandes@ibb.waw.pl

38 Introduction

39 Noroviruses are responsible for over half of the outbreaks of gastroenteritis worldwide
40 (Karst, 2010). They belong to the *Caliciviridae*, a family of positive-sense, single-stranded
41 RNA, non-enveloped viruses with a ~7.5 kb genome that generally contains three open
42 reading frames (ORF) (Jiang et al., 1993; Lambden et al., 1993; Glass et al., 2000); a novel
43 fourth ORF was recently identified in Murine norovirus (MNV) (McFadden et al., 2011).
44 Whereas ORF2, ORF3 and ORF4 each encode single proteins, ORF1 codes for a single large
45 polyprotein (~ 190 kDa) that is processed by the virally-encoded protease at five specific sites
46 to release the six non-structural proteins (NS1/2 – NS7) — and an array of functional
47 precursors — that are required for virus replication (Belliot et al., 2003; Sosnovtsev et al.,
48 2006; Muhaxhiri et al., 2013). The viral NS6 protease present within the C-terminal half of
49 the polyprotein performs all the processing, including its own autocatalytic release from the
50 precursor (Liu, Clarke & Lambden, 1996; Belliot et al., 2003; Sosnovtsev et al., 2006;
51 Scheffler et al., 2007).

52 Crystal structures have been determined for the NS6 proteases from several norovirus strains
53 (Chiba virus, Murine norovirus, Norwalk virus, Southampton norovirus) (Nakamura et al.,
54 2005; Zeitler, Estes & Venkataram Prasad, 2006; Hussey et al., 2011; Kim et al., 2012; Leen,
55 Baeza & Curry, 2012; Muhaxhiri et al., 2013). The solution structure and dynamics of
56 Norwalk virus NS6 have also been analysed by NMR spectroscopy (Takahashi et al., 2013).
57 Norovirus NS6 is a cysteine protease with a chymotrypsin-like fold: two β -barrel domains
58 separated by a cleft that contains a Cys-His-Asp/Glu catalytic triad similar in arrangement to
59 the Ser-His-Asp triad characteristic of serine proteases (Allaire et al., 1994; Matthews et al.,
60 1994). Calicivirus NS6 is related in sequence and structure to the picornavirus 3C proteases,

61 which have the same role in polyprotein processing for these single-stranded RNA viruses
62 (Leen, Baeza & Curry, 2012).

63 Previously our lab determined the structure of full-length Murine norovirus NS6 (residues 1-
64 183) (Leen, Baeza & Curry, 2012). Strikingly, adventitious crystal contacts placed the C-
65 terminus of one molecule in the active site of another, thereby generating the structure of a
66 protease-product complex that plausibly represents the final step of the *trans* cleavage by
67 NS6 at the NS6-NS7 junction. An almost identical packing arrangement was obtained in a
68 different space-group with crystals of Norwalk virus NS6 protease (Muhaxhiri et al., 2013).
69 These structures, together with the structures of di-, tri- and penta-peptidyl substrate
70 analogues bound to human norovirus NS6 structures (Hussey et al., 2011; Kim et al., 2012;
71 Muhaxhiri et al., 2013) revealed details of the specific contacts made by the P5-P1 residues
72 of peptidyl products of the protease. However, to date there are no structural data on
73 protease-substrate complexes that might uncover details of the interactions made by the
74 protease with the C-terminal, prime side of noroviral peptide cleavage junctions (residues
75 P1'-P4').

76 We reasoned that catalytically inactive MNV NS6 constructs extended at the C-terminus
77 might be able to crystallise with the same packing contacts as we had observed for the full-
78 length protein, and that this could therefore give us a convenient way to investigate the
79 structures of NS6-substrate complexes. To this end we extended an inactive MNV NS6
80 construct (which incorporates a C139A mutation to knock out the active site nucleophile) by
81 adding residues from NS7, which is immediately downstream in the polyprotein. We made
82 three constructs, NS6 1', NS6 2' and NS6 4', which were extended by 1, 2 and 4 residues
83 respectively, generating proteases that contain substrates that correspond to P4-P1', P4-P2'
84 and P4-P4' of the NS6-NS7 cleavage junction. To investigate the structural variation in

85 substrate recognition we also made an NS6 chimera which interchanged the residues P4-P4'
86 of the NS6-7 cleavage junction with the sequence of the NS2-3 junction from MNV (NS6 4'
87 2|3).

88 Proteins expressed from all four constructs were crystallised and their structures determined
89 at high resolution by X-ray crystallography. The diffraction patterns from crystals of NS6 1'
90 and NS6 2' exhibited marked anisotropy, which stalled the crystallographic refinement at
91 high R-factors. However, this was overcome by the successful application of an anisotropic
92 data correction procedure (Sawaya, 2014).

93 The modified MNV NS6 constructs all crystallized with packing arrangements that were
94 distinct from that observed in crystals we previously obtained with the mature NS6 protein
95 (Leen, Baeza & Curry, 2012; Muhaxhiri et al., 2013). Unfortunately, in no case did the
96 packing arrangements involve insertion of the cleavage junction in the extended C-terminus
97 into the active site of an adjacent protein in a way that allowed the formation of specific
98 protease-substrate interactions. The four new structures reported here therefore do not
99 provide any new information on the mode of binding of the P1'-P4' residues of junctions
100 cleaved by MNV NS6. However they do confirm the MNV NS6 protease structure and, in at
101 least one case, reveal the structure for the loop connecting β -strands cII and dII that was
102 disordered in the previously reported structure (Leen, Baeza & Curry, 2012).

103

104 **Materials and methods**

105 **Cloning and purification of MNV NS6 variants.** C-terminally extended constructs of the
106 murine norovirus NS6 protease (UniProt accession no. Q80J95; residues 995-1177),
107 inactivated by mutation of the active site Cys 139 to Ala (C139A) were generated by the

polymerase chain reaction using the inactive full length MNV NS6 as a template (Leen,
Baeza & Curry, 2012). The same forward primer was used throughout (5'-
CATCATGGATCCGCCCCAGTCTCCATCTGG). To create constructs extended by one
(Gly; NS6 1'), two (Gly-Pro; NS6 2') or four residues (Gly-Pro-Pro-Met; NS6 4') beyond the
natural C-terminus of NS6, we used the reverse primers, 5'-
ATGATGAAGCTTAGCCCTGGAAGTCCAGAGCCTCAA, 5'-
CATCATAAGCTTACGGGCCCTGGAAGTCCAGAGCCTCAAGTGTGGGTTCTCCGTG
AGT and 5'-
CATCATAAGCTTACATCGGCGGGCCCTGGAAGTCCAGAGCCTCAAGTG
respectively. To generate the chimera with the P4-P4' cleavage site of NS2-3 (NS6 4' 2|3) the
reverse primer 5'-
TACTACAAGCTTAATCAAACGGGCCTTCCGCCTGCCAAGCCTCAAGTGTGGGTTCT
TCCGTGAGT was used. PCR products were digested with BamHI and HindIII and ligated
into the pETM-11 vector as described previously for full-length NS6 (Leen, Baeza & Curry,
2012). The expressed MNV NS6 variants thus contain a thrombin-cleavable N-terminal His₆
tag; processing by thrombin leaves a Gly-Ser di-peptide preceding the Ala1 residue at the N-
terminus of our constructs. All plasmid insert sequences were confirmed by DNA sequencing
(Eurofins MWG Operon). NS6 1' and NS6 2' constructs were transformed into *E. coli* BL21-
CodonPlus (Stratagene) and grown in lysogeny broth (LB) supplemented with 25 µg/mL
kanamycin and 35 µg/mL chloramphenicol, while NS6 4' and NS6 4' 2|3 constructs were
transformed into *E. coli* BL21(DE3) pLysS (Promega) and grown in LB supplemented with
25 µg/mL kanamycin and 35 µg/mL chloramphenicol. For large-scale protein expression, 1 L
of LB was inoculated with overnight cultures. The cultures were incubated at 37°C with
shaking at 220 RPM and protein expression was induced for 3 to 4 hours by addition of 1
mM isopropyl β-D-1-thiogalactopyranoside (IPTG) once the cultures had attained an OD at

600 nm of 1.0. Cells were harvested by centrifugation at 4000 g for 10 min, and pellets frozen at -80°C.

E. coli pellets containing over-expressed NS6 1' and NS6 2' were re-suspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 1 mM dithiothreitol (DTT)) supplemented with 0.5 mM PMSF, 2 mg/ml lysozyme, and 0.1% (v/v) Triton X-100 and the cells disrupted by sonication. Cell debris was removed by centrifugation of the bacteria lysate at 29000 g for 1 h. Clarified lysates were incubated with 1 mg/mL protamine sulphate for 20 min at 4°C to precipitate nucleic acid contaminants, which were removed by re-centrifugation at 29000 g for 20 min. The lysate was loaded onto TALON metal affinity resin (Clontech) in a gravity flow column. After washing with 25 volumes of lysis buffer, 25 volumes of lysis buffer containing 5 mM imidazole, and finally a further 25 volumes of lysis buffer containing 10 mM imidazole, the MNV NS6 proteins were eluted with a step gradient to 100 mM imidazole. The His₆ tag was removed during overnight dialysis in lysis buffer in the presence 2 mM CaCl₂, and 10 units of thrombin (Sigma-Aldrich) per mg of MNV NS6 protein. Following dialysis, uncleaved material and cleaved tags were removed in a second round of TALON purification. As a final polishing step, the unbound fractions containing cleaved MNV NS6 variants were pooled, concentrated and loaded onto a Superdex-75 size-exclusion column (mounted on AKTA FPLC, GE Healthcare) equilibrated with 50 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT (SEC buffer).

In the case of NS6 4' and NS6 4' 2/3 constructs the above protocol (and additives) were used but the lysis buffer was composed instead of 50 mM HEPES pH 6.5, 300 mM NaCl, 1 mM DTT. Size-exclusion chromatography was performed in 25 mM Tris pH 8.0, 200 mM NaCl, 5 mM DTT.

Crystallisation of MNV NS6 variants. All crystallisation experiments were performed with a protein concentrations of 10-14.5 mg/mL in the SEC buffer used for the final purification step. Efforts to use the original crystallisation conditions to reproduce the MNV NS6 protein crystals with the packing arrangement that would place the C-terminus of one molecule in the active site of a neighbour (Leen, Baeza & Curry, 2012) failed for the four extended constructs generated in this study. Seeding equilibrated drops with fragments from extant NS6 crystals generated very small crystals that appeared similar in habit to the original crystals of MNV NS6 but they did not diffract appreciably.

We therefore widened the search for crystallisation conditions, carrying out screening on a sub-microlitre scale by sitting-drop vapour-diffusion. 100 nL protein solution at 14 mg/mL was mixed with 100 nL reservoir solution and equilibrated against 90 μ L of reservoir solution using a Mosquito automated pipetting system (TTP LabTech) and the following commercial crystallisation screens: Wizard 1 & 2 (Emerald Biosystems), PACT Premier, JCSG+, Morpheus and PGAScreen (Molecular dimensions). Larger crystals were grown from initial hits by sitting-drop vapour diffusion at 18°C in CompactClover Plates EBS-XPT (Jena Bioscience).

Useable crystals of MNV NS6 1' were obtained by mixing 1 μ L of protein with 1 μ L mother liquor consisting of 10% (v/v) poly-ethylene glycol (PEG) 10000, 20% (v/v) ethylene glycol, 0.1 M MES/Imidazole pH 6.3. Crystals with cuboid shape appeared between 2 and 5 days and grew to full size (80 x 50 x 50 μ m) in about 15 days and were flash-cooled in liquid nitrogen prior to data collection without any additional cryo-protectant.

Crystals of MNV NS6 2' were obtained in 15% (v/v) PEG 3350, 0.1 M glycine, 0.1 M Na-citrate pH 7.0. They grew as long rods in 1-2 days and achieved full size (400 x 30 x 30 μ m)

in 5-10 days. The crystals were cryo-protected in mother liquor supplemented with 15% (v/v) PEG 200 and flash-cooled in liquid nitrogen.

Crystals of NS6 4' and NS6 4' 2/3 grew from 0.2 M KSCN, 0.1 M Bis-Tris propane pH 6.5-7.5, 20% w/v PEG 3350. They were cryo-protected adjusting the mother liquor solution to a final concentration of 30% (v/v) PEG 3350.

X-Ray data collection and processing. X-Ray diffraction data from crystals of NS6 1' and NS6 2' were collected on a Pilatus 6M-F detector at the I03 beamline at the Diamond Light Source (Didcot, UK). For the MNV NS6 1' crystals a 2.3 Å data set of 200 frames was collected with an oscillation width of 1° per frame. Diffraction images were integrated and scaled using the CCP4 program suite (Winn et al., 2011). Data-collection statistics are summarized in Table 1.

For the MNV NS6 2' crystals, 720 frames with a 0.5° oscillation were collected. A 3.1 Å dataset was integrated and scaled as described above; see Table 1 for data collection statistics.

MNV NS6 4' and NS6 4' 2/3 crystals data was collected in-house using a Rigaku MicroMax-007 HF-M X-ray generator and a Saturn 944+ CCD detector. Data sets of 316 and 344 0.5° oscillation frames were collected for NS6 4' and NS6 4' 2/3 crystals at 2.47 and 2.42 Å respectively. The data were processed and scaled as described above.

All four datasets were submitted to the UCLA MBI Diffraction Anisotropic Server (<http://services.mbi.ucla.edu/anisoscale>) for anisotropic analysis. Following the server indication of severe anisotropy diffraction of the MNV NS6 1' and NS6 2' crystals both data sets were truncated/scaled using the server default values, in particular using a 3.0 cut-off for F/sigma (Strong et al., 2006; Sawaya, 2014).

Phasing, model building and refinement. Molecular-replacement phasing was performed in Phaser (McCoy et al., 2007) using the crystal structure of full-length MNV NS6 (PDB entry 4ASH) (Leen, Baeza & Curry, 2012) pruned of double conformations and the initial 6 and terminal 11 residues to avoid biasing the conformations of the termini.

The MNV NS6 1' and NS6 2' structures obtained from molecular replacement were subjected to restrained refinement using REFMAC (Murshudov et al., 2011), in the CCP4 program suite (Collaborative Computational Project, 1994). Molecular replacement solutions of NS6 4' and NS6 4' 2|3 were refined using Phenix refine (Adams et al., 2010). All manual model adjustments were made in Coot (Emsley & Cowtan, 2004).

Results and discussion

Structure determination

Datasets collected for MNV NS6 1' and NS6 2' were initially scaled and truncated — at 2.3 Å and 3.1 Å respectively — with the assumption in each case that the diffraction was isotropic even though some anisotropy was evident in the diffraction images. This approach resulted in data sets with relatively high values of R_{merge} and low signal-to-noise ratios $[I/\sigma(I)]$ for the highest-resolution shells of data (Table 1). The MNV NS6 1' crystals were determined to be composed of four molecules in the asymmetric unit of the $C2$ unit cell, which is consistent with a solvent content of 50.3% (Matthews, 1968). MNV NS6 2' crystals belong to space-group $P6_122$ and were determined to contain two molecules in the asymmetric unit, with slightly higher solvent content of 55.3%.

Initial structural factors (processed isotropically) were used for molecular replacement, and in the case of MNV NS6 1' produced four possible solutions with a highest LLG of 2723 and

TFZ of 19.7. However, initial refinement of the model obtained by molecular replacement stalled at relatively high values of R_{work} (~33%) and R_{free} (~39%). A similar problem was encountered with initial attempts to refine the MNV NS6 2' model.

At this stage we re-visited the processed data to determine if the anisotropy was at the root of the refinement problems. Plots of F/σ against resolution for each of the 3 principal axes revealed the severity of the anisotropy of the diffraction from both crystals (Fig. 1a, b); when truncated at $F/\sigma < 3.0$ along each axis, the spread of data was ellipsoidal in appearance. The server identified the c^* axis as stronger diffracting than the a^* and b^* directions in each case, and detected anisotropic ΔB values of 68.6 \AA^2 and 51.0 \AA^2 for MNV NS6 1' and NS6 2', respectively. Anisotropic ΔB reports the directionality dependence of the intensity falloff with resolution (<http://services.mbi.ucla.edu/anisoscale/>). Following the F/σ analysis, the MNV NS6 1' data were truncated to 3.1 \AA , 2.3 \AA , and 2.1 \AA along a^* , b^* and c^* , respectively. To create a nominally isotropic data set, B -factor corrections of 39.2 , -9.8 and -29.4 \AA^2 were applied to the observed structure factors along the a^* , b^* and c^* directions respectively. The reduction in the anisotropy of the corrected data in the a^*c^* ($h0l$) plane can be seen by comparing Figures 1c and 1d.

Following anisotropy analysis the MNV NS6 2' data was truncated to 3.2 \AA , 3.2 \AA , and 2.7 \AA along a^* , b^* and c^* directions. To generate the nominally isotropic data set, B -factor correction of 17.0 , 17.0 and -34.0 \AA^2 along the same axes were applied to the observed structure factors. The anisotropic truncation of the data, with the new limit of 2.7 \AA along the c^* axis, resulted in a dramatic increase in the number of unique reflections from 8555 in the “pre-treated” data to 12786.

MNV NS6 4' or NS6 4' 2|3 data sets plots of F/σ against resolution for each of the 3 principal axes revealed only mild anisotropy of the diffraction from both crystals (anisotropic

ΔB values of 13.3 Å² and 15.4 Å² for MNV NS6 4' and NS6 4' 2|3, respectively). In these cases no directional-dependent truncation of the data was applied.

After taking account of the anisotropy of the data, the axes-dependent truncated and corrected structure factors were then used to repeat the molecular replacement phasing of the NS6 1' data and resulted in significantly better solutions. The best solution had an LLG of 3051 and TFZ of 32.3. Moreover, the model showed immediate improvement in the early cycles of refinement ($R_{\text{work}} \sim 31\%$; $R_{\text{free}} \sim 36\%$) and yielded electron density maps that were much more interpretable (Fig. 1e, f). Multiple cycles of refinement and model building lead to a structure characterized by R_{work} of 23.0% and R_{free} of 27.8%.

Anisotropic treatment of the MNV NS6 2' data also improved the molecular replacement outcome, with a unique solution found by Phaser in contrast with the two possible solutions of the uncorrected data. The LLG value increased markedly from 385 to 887 while there was a slight drop in the value of TFZ (from 19.2 to 18.8). Nevertheless, this resulted in immediate improvements in refinement: R_{work} dropped from 32.5 to 30.3% while R_{free} was reduced from 42.0 to 39.0%. These statistics were further improved to R_{work} of 25.4% and R_{free} of 30.3% in the final structure.

The translation Z-scores and LLG's for the Phaser molecular replacement solutions for MNV NS6 4' are 13.8 and 1042, respectively, and for MNV NS6 4' 2|3 27.1 and 1394, respectively. The crystals structures of NS6 4' and NS6 4' 2|3 were refined to R factors ($R_{\text{work}}/R_{\text{free}}$) of 21.3/26.1 and 17.3/26.3 %, respectively.

271 **Structure analysis**

272 Overall, the structure of the protease core domains for NS6 1', NS6 2', NS6 4' and NS6 2|3 4'
 273 is very similar to that reported previously for MNV NS6 (Leen, Baeza & Curry, 2012) – (root
 274 mean squared differences are less than 1 Å) (Fig. 2a). For three of the four new structures
 275 (NS6 2', NS6 4' and NS6 2|3 4') it was possible to identify conserved crystal contacts,
 276 involving a two-fold symmetric packing arrangement largely mediated by the short N-
 277 terminal helix and the loop connecting β -strand aII and bII (Fig. 2b). In each case the C-
 278 terminus of at least one protein molecule in the asymmetric unit was observed to extend away
 279 from the body of the protease and to interact with a neighbouring molecule in the crystal —
 280 as had been observed previously for MNV NS6. Moreover, the extended C-terminus
 281 invariably inserted into the peptide-binding groove formed between loops bII-cII and eII-fII
 282 loops (Fig. 3). Disappointingly however, in no case did the extended peptide make
 283 interactions that were consistent with specific contacts with the protease active site. We will
 284 briefly describe each of the four new structures before turning to the question of why in no
 285 case a protease-substrate complex was obtained.

286 MNV NS6 1' crystallized in space group *C2* with four molecules in the asymmetric unit. This
 287 structure is notable for the fact that, in contrast to the other three structures reported here and
 288 the structures previously reported for Norwalk virus NS6 protease (Zeitler, Estes &
 289 Venkataram Prasad, 2006; Muhaxhiri et al., 2013), the electron density map was of sufficient
 290 quality in all four molecules of the asymmetric unit to permit the incorporation of the loop
 291 between the β -strands cII and dII (Fig. 2a). Superposition of the four copies of NS6 1'
 292 indicates some structural variation in this loop — the C α positions vary by 1-2 Å — consistent
 293 with the notion that it is rather flexible (Fig. 2a).

In three of the molecules in the asymmetric unit the C-terminus is disordered beyond residue 173 (chains A and D) or residue 174 (chain C). The electron density for the last four residues modelled in chain D (Ala 170 to Gly 173) is poor, presumably due to disorder. Nevertheless, we included these residues in the final model because removal increased the *R*-factor.

Although the full C-termini of chains A, C and D of the NS6 1' could not be modelled, the electron density map revealed additional features that could be built as short stretches of polypeptide. These correspond to portions of the missing C-termini but, because of discontinuities in the electron density, it is not possible to unambiguously ascribe which monomer they belong to. These short segments were present in the asymmetric unit, and were modelled as (i) a stretch of five residues (Chain E) that could be assigned confidently as corresponding to residues Leu 180 to Gly 184 that lies near to the active site of a neighbouring chain A; and (ii) an extended portion of electron density that was not of sufficient quality to identify the amino acid side chains and was built as a seven-residue poly-Ala peptide (chain F), that lies between chains A and D and may be part of a symmetry-related molecule of chain A. However, in neither case were specific interactions with the putative active sites of neighbour monomers observed.

In contrast the electron density for the B chain in the asymmetric unit was sufficiently clear to permit inclusion of all 183 residues of NS6 and the extra first residue of NS7 (here labelled Gly 184) in the refined model (Fig. 3b). The C-terminus of Chain B of NS6 1' beyond Gly 173 extends to make contacts with chain B from an adjacent asymmetric unit to which it is related by a two-fold symmetry axis. This two-fold symmetric packing arrangement is reminiscent of the inter-protein contacts previously observed for the full length NS6 (Leen, Baeza & Curry, 2012) (Fig. 3a). However, although the C-terminus of NS6 1' comes close to the substrate-binding site of neighbouring proteases in the crystal, it does not reach far

enough and makes none of the specific contacts needed to position the scissile bond

(Gln 183-Gly 184) in the active site.

MNV NS6 2' crystallized in space group $P6_122$ with two molecules in the asymmetric unit.

The electron density indicates the positions of residues 3-181 in chain A and 4-181 in chain

B, but is not good enough to allow complete modelling of the N- or C-termini. The two

monomers found in the asymmetric unit are related by a quasi-2-fold symmetry axis and form

an apparent homodimer with an interface area of 1097 \AA^2 (Fig. 3c). In addition to

neighbourly contacts made by the two C-termini, dimerization within the crystal is stabilised

by hydrogen bonds involving Ser51, Ser111 and Val113 of chain A and Ser51, Ser52, Ser111

and Val113 of chain B. Although the C-termini within the asymmetric unit each embrace the

other monomer, reaching into the groove formed between the bII-cII and eII-fII loops, they

are again not located within the substrate binding site.

MNV NS6 4' crystallized in space group $C2$ but in a unit cell that is very different from the

$C2$ crystals obtained with MNV NS6 1' (Table 1) and that has only one molecule in the

asymmetric unit. The electron density was sufficient to build a model that starts at residue 4

and ends at residue 179. Although the extended C-terminus of NS6 once again reaches

between loops bII-cII and eII-fII of its near neighbour, the relation between the two

molecules does not involve two-fold or quasi two-fold symmetry (Fig. 3d). However, as

before no specific contacts are made by the C-terminal peptide with the substrate binding site.

NS6 4' 2/3 crystallised with two molecules in a $P1$ unit cell. The modelled monomers start at

residue 4 and end at residue 182; in contrast to the other structures reported here, the A chain

lack density for the tip of the eII-fII loop so residues 162-163 were omitted. The A and B

chains each extend their C-termini into the cleft formed between the bII-cII and eII-fII loops

of a neighbouring molecule (Fig. 3e, f). The packing arrangements are distinct but rather

similar and, once again, lack specific contacts that are consistent with productive substrate binding.

Implications of the structures

With the extended NS6 protease constructs designed for this study we aimed to exploit our earlier finding that the mature protease crystallised with the C-terminus of one molecule inserted into peptide binding site of a neighbouring protein in a way that revealed specific protease-product interactions (Leen, Baeza & Curry, 2012). We had hoped to obtain the same packing interaction with the C-terminally extended complexes in order to determine the structures or protease-substrate complexes that would reveal the details of the interactions made by the amino acids in positions P1'-P4' of the substrate. A similar approach has since been used successfully to look at different protease-product complexes (Muhaxhiri et al., 2013) but in each of the four cases that we probed, extension of the C-terminus resulted in novel packing arrangements, none of which captured a proper protease-substrate reaction.

Why did we not get the same crystal form as for the full-length protein? At present it is difficult to give a definitive answer to this question. Although we screened extensively for crystallisation conditions, we cannot claim to have searched exhaustively and it may be that further efforts might yet succeed. Moreover, it is well known that even very modest changes to protein constructs may alter substantially the way that they crystallise, though it is worth noting that a similar strategy to explore the specific interactions of different protease-product complexes was applied successfully to Norwalk virus NS6 (Muhaxhiri et al., 2013). In preliminary investigations using nuclear magnetic resonance (NMR) to investigate solution structures of extended MNV NS6 protease structures, we have obtained evidence to suggest that their C-termini may have a propensity to fold into the active site of the molecule that it

belongs to, making a *cis* interaction (Fernandes, Pastore & Curry, unpublished data). This may explain why *trans* interactions were not observed in our crystals but in turn raises a further question: why was this conformational state not captured in the crystal form? One possibility is that the concentration of proteins inevitably involved in crystallisation somehow destabilizes the *cis* interaction but does not necessarily capture a catalytically competent *trans* interaction between the cleavage junction within the extended C-terminus and a protease neighbour. Why this would be the case, particularly when a protease-substrate complex would be expected to be a relatively stable state, remains a mystery.

Acknowledgements

We thank staff at the Diamond Light Source (Didcot, UK) for assistance with data collection.

References

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, and Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66:213-221.
- Allaire M, Chernaia MM, Malcolm BA, and James MN. 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* 369:72-76.
- Belliot G, Sosnovtsev SV, Mitra T, Hammer C, Garfield M, and Green KY. 2003. In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. *J. Virol.* 77:10957-10974.
- Collaborative Computational Project N. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50:760-763.
- Emsley P, and Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60:2126-2132.
- Glass PJ, White LJ, Ball JM, Leparc-Goffart I, Hardy ME, and Estes MK. 2000. Norwalk virus open reading frame 3 encodes a minor structural protein. *J. Virol.* 74:6581-6591.
- Hussey RJ, Coates L, Gill RS, Erskine PT, Coker SF, Mitchell E, Cooper JB, Wood S, Broadbridge R, Clarke IN, Lambden PR, and Shoolingin-Jordan PM. 2011. A structural study of norovirus 3C protease specificity: binding of a designed active site-directed peptide inhibitor. *Biochemistry* 50:240-249.
- Jiang X, Wang M, Wang K, and Estes MK. 1993. Sequence and genomic organization of Norwalk virus. *Virology* 195:51-61.
- Karst SM. 2010. Pathogenesis of noroviruses, emerging RNA viruses. *Viruses* 2:748-781.
- Kim Y, Lovell S, Tiew KC, Mandadapu SR, Alliston KR, Battaile KP, Groutas WC, and Chang KO. 2012. Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J. Virol.* 86:11754-11762.

- 404 Lambden PR, Caul EO, Ashley CR, and Clarke IN. 1993. Sequence and genome organization of a
405 human small round-structured (Norwalk-like) virus. *Science* 259:516-519.
- 406 Leen EN, Baeza G, and Curry S. 2012. Structure of a murine norovirus NS6 protease-product
407 complex revealed by adventitious crystallisation. *PloS ONE* 7:e38723.
- 408 Liu B, Clarke IN, and Lambden PR. 1996. Polyprotein processing in Southampton virus:
409 identification of 3C-like protease cleavage sites by in vitro mutagenesis. *J. Virol.* 70:2605-2610.
- 410 Matthews BW. 1968. Solvent content of protein crystals. *J. Mol. Biol.* 33:491-497.
- 411 Matthews DA, Smith WW, Ferre RA, Condon B, Budahazi G, Sisson W, Villafranca JE, Janson CA,
412 McElroy HE, Gribskov CL, and et al. 1994. Structure of human rhinovirus 3C protease reveals a
413 trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein.
414 *Cell* 77:761-771.
- 415 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ. 2007. Phaser
416 crystallographic software. *J. Appl. Crystallogr.* 40:658-674.
- 417 McFadden N, Bailey D, Carrara G, Benson A, Chaudhry Y, Shortland A, Heeney J, Yarovinsky F,
418 Simmonds P, Macdonald A, and Goodfellow I. 2011. Norovirus regulation of the innate immune
419 response and apoptosis occurs via the product of the alternative open reading frame 4. *PLoS*
420 *Pathogens* 7:e1002413.
- 421 Muhaxhiri Z, Deng L, Shanker S, Sankaran B, Estes MK, Palzkill T, Song Y, and Prasad BV. 2013.
422 Structural basis of substrate specificity and protease inhibition in Norwalk virus. *J. Virol.*
423 87:4281-4292.
- 424 Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F,
425 and Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta*
426 *Crystallogr. D Biol. Crystallogr.* 67:355-367.
- 427 Nakamura K, Someya Y, Kumasaka T, Ueno G, Yamamoto M, Sato T, Takeda N, Miyamura T, and
428 Tanaka N. 2005. A norovirus protease structure provides insights into active and substrate
429 binding site integrity. *J. Virol.* 79:13685-13693.
- 430 Sawaya MR. 2014. Methods to refine macromolecular structures in cases of severe diffraction
431 anisotropy. *Methods Mol Biol* 1091:205-214.

- 432 Scheffler U, Rudolph W, Gebhardt J, and Rohayem J. 2007. Differential cleavage of the norovirus
433 polyprotein precursor by two active forms of the viral protease. *J. Gen. Virol.* 88:2013-2018.
- 434 Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 1.3r1.
- 435 Sosnovtsev SV, Belliot G, Chang KO, Prikhodko VG, Thackray LB, Wobus CE, Karst SM, Virgin
436 HW, and Green KY. 2006. Cleavage map and proteolytic processing of the murine norovirus
437 nonstructural polyprotein in infected cells. *J. Virol.* 80:7816-7831.
- 438 Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, and Eisenberg D. 2006. Toward the structural
439 genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium*
440 *tuberculosis*. *Proc. Natl. Acad. Sci. USA* 103:8060-8065.
- 441 Takahashi D, Hiromasa Y, Kim Y, Anbanandam A, Yao X, Chang KO, and Prakash O. 2013.
442 Structural and dynamics characterization of norovirus protease. *Protein Sci.* 22:347-357.
- 443 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB,
444 Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR,
445 Read RJ, Vagin A, and Wilson KS. 2011. Overview of the CCP4 suite and current developments.
446 *Acta Crystallogr. D Biol. Crystallogr.* 67:235-242.
- 447 Zeitler CE, Estes MK, and Venkataram Prasad BV. 2006. X-ray crystallographic structure of the
448 Norwalk virus protease at 1.5-Å resolution. *J. Virol.* 80:5050-5058.

449

450

Figures

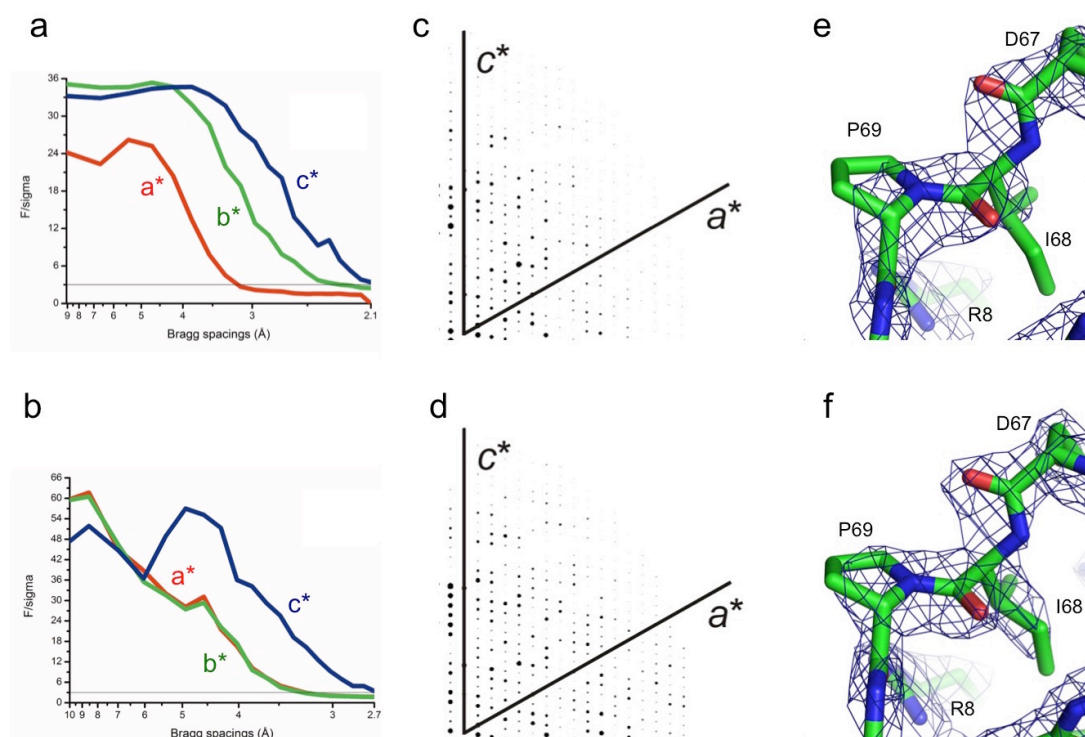


Figure 1. Analysis and correction of the anisotropic diffraction observed for crystals of NS6 1' and NS6 2'. F/σ vs. Bragg spacings for each of the cell directions for (a) NS6 1' and (b) NS6 2' respectively. Pseudo-precession images of the anisotropy in the a^*c^* ($h0l$) plane for NS6 1' (c) before and (d) after correction. $2F_o - F_c$ electron density maps contoured at 2σ after one round of refinement of the molecular replacement solutions obtained with Phaser (McCoy et al., 2007) for NS6 1' (e) before and (f) after anisotropic correction.

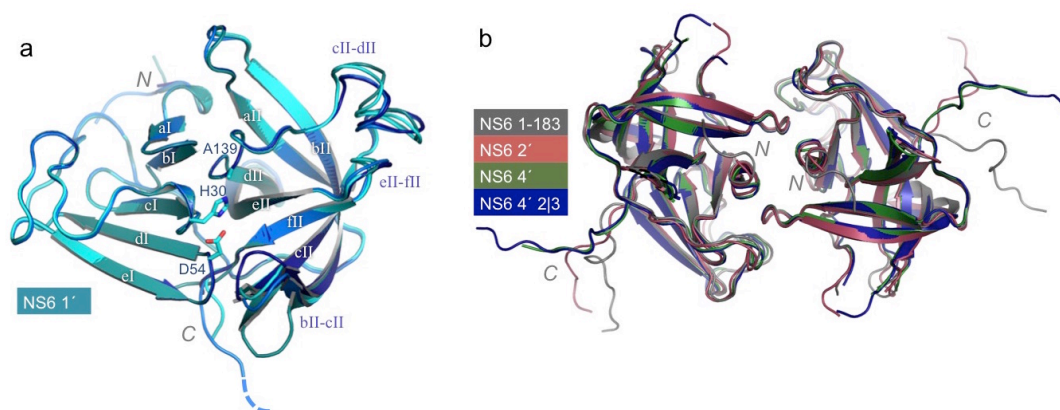


Figure 2. Structure and conserved packing interfaces of C-terminally extended NS6 proteases. (a) Superposition of the four molecules in the asymmetric unit of crystals of NS6 1 (coloured various shades of blue). The β -strands are labelled, as are the N- and C-termini and the conserved side-chains of catalytic triad. Note that the active site C139 has been replaced by A139 in all structures reported here. (b) A conserved packing arrangement observed for the mature NS6 protease (NS6 1-183) (Leen, Baeza & Curry, 2012) and three of the structures solved in the present work (NS6 2', NS6 4' and NS6 4' 2|3). All structural figures were made with PyMOL (Schrodinger, 2010).

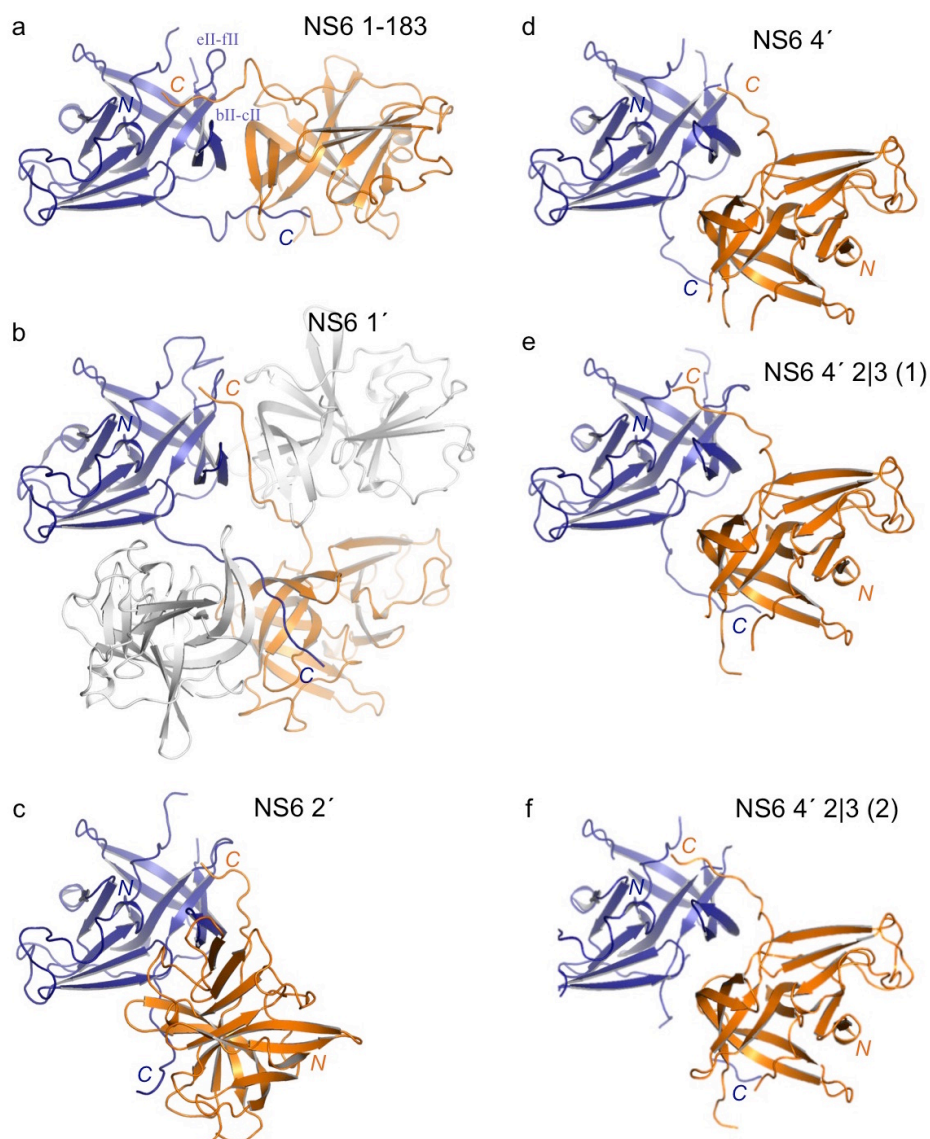


Figure 3. Comparison of the packing arrangements of C-terminally extended NS6

proteases. (a) The mature NS6 protease (NS6 1-183) (Leen, Baeza & Curry, 2012). Chains which interact via C-termini are coloured blue and orange. This colour-scheme is maintained throughout the figure; note also that the orientation of the blue chain is the same in each panel. (b) NS6 1' — the interaction is between a pair of symmetry-related B chains. An additional pair of symmetry-related C chains — which also contact the extended C-terminus of the B chains — is shown in white (c) NS6 2'. In this case the interaction is between the

483 two chains in the asymmetric unit. (d) $NS6\ 4'$ — here there is only one chain in the
 484 asymmetric unit and the interaction is not symmetric. (e) $NS6\ 4'\ 2|3$ — here again the
 485 interaction between chain B (blue) of one asymmetric unit and chain A (orange) of another is
 486 not symmetric. (f) $NS6\ 4'\ 2|3$ — a second but very similar mode of interaction in these
 487 crystals between chain A (blue) of one asymmetric unit and chain B (orange) of another.

488

489

Table 1. Data-collection and model refinement statistics.

Data Collection				
	MNV NS6 1'	MNV NS6 2'	MNV NS6 4'	NS6 4' 2/3
Radiation source	Diamond I03	Diamond I03	Rigaku MicroMax-007 HF-M	Rigaku MicroMax-007 HF-M
Wavelength (Å)	1.000	1.000	1.54	1.54
Detector	Pilatus 6M-F	Pilatus 6M-F	Saturn 944+ CCD	Saturn 944+ CCD
Resolution limits ¹ (Å)	70.93-2.3 (2.42 -2.3)	68.02-3.1 (3.31-3.1)	38.11-2.472 (2.561-2.472)	19.27-2.417 (2.503-2.417)
Space group	<i>C</i> 2	<i>P</i> 6 ₁ 22	<i>C</i> 2	<i>P</i> 1
Unit-cell parameters (Å, °)	<i>a</i> = 99.64 <i>b</i> = 111.86 <i>c</i> = 81.29 β = 119.24	<i>a</i> = 136.04 <i>c</i> = 82.39	<i>a</i> = 88.19 <i>b</i> = 35.36 <i>c</i> = 52.81 β = 105.98	<i>a</i> = 35.52, <i>b</i> = 47.32, <i>c</i> = 53.07 α = 104.45 β = 91.53 γ = 110.61
Mosaicity (°)	0.42	0.70	0.67	1.11
No. of unique reflection	34593	8555	5711	10817
Multiplicity	3.5 (3.6)	5.7 (6.0)	3.0 (2.3)	1.8 (1.8)
$\langle I/\sigma(I) \rangle^2$	6.7 (1.2)	10.9 (2.1)	19.63 (6.94)	6.7 (2.31)
Completeness (%)	99.7 (99.8)	99.8 (99.6)	99.03 (92.44)	91.14 (66.47)
R_{merge} (%) ³	8.4 (8.6)	8.5 (91.5)	3.4 (10.6)	9.3 (27.1)
Overall <i>B</i> factor (Å ²)	54.4	94.3	27.1	24.9
Model Refinement				
No. of Non-hydrogen atoms/waters	5356/88	2563/0	1218/22	2559/107
R_{work} (%) ⁴	23.0	25.4	21.3 (27.3)	17.3 (22.4)
R_{free} (%) ⁵	27.8	30.3	26.1 (36.3)	26.3 (31.3)
RMSD bonds (Å) ⁶	0.012	0.012	0.010	0.008
RMSD bond angles (°)	1.605	1.610	1.16	1.19
Ramachandran plot (% favoured/allowed)	89.3/10.7	88.8/11.2	95/5	96/4
PDB Identifier	4x2v	4x2w	4x2x	4x2y

¹ Values in parentheses refer to the highest resolution shell of data.

² $\langle I/\sigma(I) \rangle$ is the mean signal-to-noise ratio, where I is the integrated intensity of a measured reflection and $\sigma(I)$ is the estimated error in the measurement.

³ $R_{merge} = 100 \times \sum_{hkl} |I_j(hkl) - \langle I_j(hkl) \rangle| / \sum_{hkl} \sum_j I_j(hkl)$, where $I_j(hkl)$ and $\langle I_j(hkl) \rangle$ are the intensity of measurement j and the mean intensity for the reflection with indices hkl , respectively.

⁴ $R_{work} = 100 \times \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$.

⁵ R_{free} is the R_{work} calculated using a randomly selected 5% sample of reflection data that were omitted from the refinement.

⁶ $RMSD$, root-mean-squared deviations (from ideality).