In crystallo-screening for discovery of human norovirus 3C-like protease inhibitors

Jingxu Guoa,1, Alice Douangamatb, Weixiao Songa, Alun R. Cokera, A.W. Edith Chan,a
Steve P. Wooda, Jonathan B. Coopera,c,⁎, Efrat Resnickd, Nir Londond, Frank von Defltb,e,f

a Division of Medicine, UCL, Gower Street, London WC1E 6BT, UK
b Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, UK
c Department of Biological Sciences, Birkbeck, University of London, Malet Street, Bloomsbury, London WC1E 7HX, UK
d Department of Organic Chemistry, Weizmann Institute of Science, Rehovot 7610001, Israel
e Structural Genomics Consortium, University of Oxford, Roosevelt Drive, OX3 7DQ, UK
f Department of Biochemistry, University of Johannesburg, Auckland Park 2006, South Africa

ARTICLE INFO

Keywords:
X-ray crystallography
Structural biology
Fragment screening
Drug discovery
3C-like protease
Anti-virals

ABSTRACT

Outbreaks of human epidemic nonbacterial gastroenteritis are mainly caused by noroviruses. Viral replication requires a 3C-like cysteine protease (3CLpro) which processes the 200 kDa viral polyprotein into six functional proteins. The 3CLpro has attracted much interest due to its potential as a target for antiviral drugs. A system for growing high-quality crystals of native Southampton norovirus 3CLpro (SV3CP) has been established, allowing the ligand-free crystal structure to be determined to 1.3 Å in a tetrameric state. This also allowed crystal-based fragment screening to be performed with various compound libraries, ultimately to guide drug discovery for SV3CP. A total of 19 fragments were found to bind to the protease out of the 844 which were screened. Two of the hits were located at the active site of SV3CP and showed good inhibitory activity in kinetic assays. Another 5 were found at the enzyme's putative RNA-binding site and a further 11 were located in the symmetric central cavity of the tetramer.

1. Introduction

Gastroenteritis accounts for the deaths of over 2000 children every day worldwide, making it the second leading cause of death for children under the age of 5, more than the combination of AIDS, malaria and measles (Liu et al., 2012). Whilst there are many other causes of gastroenteritis, including parasites, bacteria and viruses, human caliciviruses are recognised as the leading cause of gastroenteritis worldwide among people of all ages. The Caliciviridae family contains five genera known as norovirus, vesivirus, nebovirus, sapovirus and lagovirus (Clarke et al., 2012) with norovirus being the most common cause of disease in humans (Lambden et al., 1993).

Noroviruses account for more than 50% of gastroenteritis cases and at least 90% of nonbacterial acute gastroenteritis cases worldwide, as reported by the Centers for Disease Control and Prevention in the US (2011). Scallan et al. (2011) estimated that 99% of all viral foodborne illness incidents are caused by noroviruses which corresponds to 5.5 million per year in the US alone. From 2009 to 2013, around 62.5% of norovirus cases needed long-term care facilities in order to control the transmission (Vega et al., 2014). Statistics are generally similar in Europe (Baert et al., 2009; Phillips et al., 2010). Globally, it is estimated that noroviruses lead to a total of $4.2 billion in direct health system costs and $60.3 billion in social cost per year (Bartsch et al., 2016).

Clinical treatment and intervention is hampered by the lack of licensed vaccines or antivirals. Treatment with human immunoglobulin did show some benefit but did not result in clearance of the virus (Florescu et al., 2008). Whilst development of a vaccine has been hindered by the lack of small-animal models and cell culture systems, a number of norovirus vaccines are yielding promising results in clinical trials (e.g. Bernstein et al., 2015; Mateo et al., 2020). In general, norovirus vaccines are based on the use of virus-like particles formed by the main capsid protein VP1 (Lucero et al., 2018). A recent review of the development of norovirus antiviral agents and their targets is given by Netzler et al. (2019).

Noroviruses are genetically classified into 7 genogroups, GI - GVII, based on the amino acid sequence of the VP1 capsid protein and are
further segregated into at least 40 genotypes (Vijé, 2015). Noroviruses from groups GI (like Southampton virus) and GII infect humans, as do members of the GIV.1 subgroup. GII viruses are the most frequently detected (89%) while GI.4 are the major cause of norovirus outbreaks worldwide (Siebenga et al., 2009). Many noroviruses have been reported such as Norwalk virus (Jiang et al., 1993), Hawaii virus (Lew et al., 1994a), Snow Mountain virus (Lorchide and Hardy, 2003), Desert Shield virus (Lew et al., 1994b), Southampton virus (Clarke and Lambden, 1997) and Lordsdale virus (Lambden et al., 1993).

The norovirus genome consists of a single-stranded positive-sense RNA molecule of 7.5–7.7 kb in length and contains three open reading frames (ORFs) (Lambden et al., 1993), except for the murine norovirus which has a fourth alternative ORF (McFadden et al., 2011). ORF1 encodes a 200 kDa non-structural polyprotein which is co- and posttranslationally cleaved into six or seven non-structural proteins by the viral 3C-like protease (NS6). The seven products of this proteolysis are, from N-terminus to C-terminus: p48 (NS1-2), an NTPase (NS3), a 3A-translationally cleaved into six or seven non-structural proteins by the which has a fourth alternative ORF (McFadden et al., 2011). ORF1 RNAmoleculeof7.5–7.7kbinlengthandcontainsthreeopenreading

2. Methods

2.1. Crystallisation

Expression and purification of SV3CP was conducted using the method described by Hussey et al., (2011). Screening for crystallisation conditions for SV3CP was accomplished using the sitting-drop method at 21 °C with the screening kits: Structure Screen 1 & 2, JCSG-plus, PACT premier, MIDAS and Morpheus from Molecular Dimensions (Suffolk, UK). A TTP Labtech Mosquito crystal screening robot (TTP Labtech, Hertfordshire, UK) was used to dispense 400 nl of the protein, at concentrations of 5 mg/ml and 10 mg/ml, with 400 nl of the correponding well solution into each drop. High quality crystals were obtained in 0.2 M ammonium citrate and 12% (v/v) PEG3350 after approximately one week, although crystals kept appearing over the next 2–3 months prior to screening.

2.2. Data collection, data processing and structure determination

Selected crystals were cryo-protected in 30% glycerol and mounted in loops before flash-cooling. X-ray data were collected at beamline I04-1 at Diamond Light Source (DLS, Didcot, England). Fine-sliced data were collected as guided by the strategy suggested by the program EDNA (Incarnona et al., 2009). Data were processed automatically by the program Xia2 (Winter, 2010) at DLS, which revealed the space group to be C2, as shown in Table 1. Further analysis using Phenix.xt-rage (Zwart et al., 2005) suggested that the data were of good quality. The solvent content of this crystal form was estimated to be 44.9% using Matthews_coef (Kantardjieff and Rupp, 2003).

The structure was determined by use of the program Phaser MR (McCoy et al., 2007) using the protein moity of the published SV3CP-MAPI complex (PDB ID: 2ipb) as a search model. Several rounds of manual rebuilding and correction were performed using Coot (Emsley and Cowtan, 2004) followed by restrained refinement using Refmac5 (Murshudov et al., 2011) and Phenix.refine (Afonine et al., 2012). Since the crystal diffracted to near atomic resolution, the temperature factors were refined anisotropically. Structure validation was performed with MolProbity (Chen et al., 2010). The statistics for data collection, data processing and refinement are shown in Table 1.

2.3. In crystallo fragment screening

2.3.1. Crystal preparation

Crystals were prepared in Swisssi 3-drop crystallisation plates (Hampton Research, CA, USA) in 200 nl droplets containing 100 nl of the protein (4 mg/ml) and 100 nl of well solution [0.2 M ammonium citrate, 12% (v/v) PEG3350]. Since all of the fragments were dissolved in 100% dimethyl sulfoxide (DMSO), crystal stability in this solvent was first tested in the range (v/v) of 0%, 10%, 20%, 30% and 40%, and on soaking time scales of 1 h, 3 h and overnight. In order to make the experiment more efficient, the crystals were also tested with and without additional cryo-protectant for data collection. It was found that these crystals could survive in 40% DMSO for many hours and additional cryo-protection was not required.

2.3.2. Fragment soaking, crystal harvesting and data collection

The plates containing crystals were imaged using a Rock imager system (Formulatrix, USA). All the crystals were then ranked using the
Table 1

X-ray statistics for the native SV33CP structure and fragment complexes. Values in parentheses are for the high resolution shell. For the minority of structures where the overall fragment occupancy was either refined or is less than unity due to proximity with a symmetry axis, the fractional occupancy is shown following the mean fragment B-factor.

| Fragment | Native | J01 | J02 | J03 |
|----------|--------|-----|-----|-----|
| PDB-ID   | 61q    | 6 14| 616w| 62b |
| a (Å)    | 63.14  | 62.72| 62.94| 62.73|
| b (Å)    | 89.37  | 89.35| 89.46| 89.35|
| c (Å)    | 61.60  | 60.69| 61.20| 60.92|
| β (%)    | 96.50  | 96.57| 96.89| 96.95|
| Resolution (Å) | 41.26–1.30 | 60.29–1.56 | 44.73–1.80 | 51.09–1.61 |
| Completeness (%) | 99.9 (99.9) | 99.8 (99.8) | 98.6 (99.4) | 98.9 (99.7) |

Continued on next page.
program TexRank (Ng et al., 2014) and positional coordinates for the injection of the fragments were manually defined in the drop. Each injection from the DSLP library (776 fragments) (Cox et al., 2016) and, due to time constraints, a subset of the Maybridge Ro3 core set (first 68 fragments) (Fisher Scientific UK Ltd, Loughborough, UK) was acoustically dispensed to the corresponding target position in droplets of 2.5 nl volume using a Labcyte Echo 550 liquid handler (Labcyte Inc, CA, USA) which gave an estimated final fragment concentration of 200 mM (Collins et al., 2017). Fragment soaking was conducted in batches to give an average soaking time of approximately 2.5 h prior to crystal mounting. Crystal harvesting was aided by the use of a crystallisation plate shifter (Oxford Lab Technologies, Oxford, UK). All the crystals were mounted in loops of about the same size as the crystals or slightly smaller to allow for automated, unattended data collection in which the X-ray beam was aimed at the centre of each loop. A total of 180° of data were collected for each crystal, taking approximately 60 s per crystal using DLS beamline I04-1.

### 3. Results

#### 3.1. Structure of native SV3CP

The structure of native SV3CP has been determined for the first time at the near-atomic resolution of 1.3 Å resolution (Fig. 1a) revealing a crystallographic tetramer (Fig. 1b). The monomers consist of an N-terminal and a C-terminal domain with the active site cleft located in between. As found in other noroviral 3CLpro structures, the N-terminal domain contains an α-helix and a twisted 7-stranded antiparallel β-sheet forming an incomplete β-barrel (Anand et al., 2002; Birtley et al., 2005; Mosimann et al., 1997). The C-terminal domain is made up of 6 β-strands forming an antiparallel β-barrel and contains the catalytic cysteine residue (Cys139) which makes a catalytic triad with two residues from the N-terminal domain (His30 and Glu54; Fig. 1a). Interestingly, the β-hairpin formed by β9 and β10, which is involved in binding the N-terminal side of the substrate peptide, adopts an appreciably different conformation from that observed in an earlier inhibitor-complexed structure (Fig. 1c; Hussey et al., 2011). It is now clear that the backbone of this β-hairpin moves by 7 Å to open up the active site cleft for substrate binding and movements of some of the side chain atoms exceed of 12 Å. Indeed, in the native enzyme, residues Met107 to Glu110 occupy very approximately the same positions as the P2–P3 residues of the bound substrate analogue and the sidechain of Arg112 occupies the position of the P3 sidechain (Fig. 1d). In addition to the movement of β9 and β10, the β-hairpin formed by strands β11 and β12 also moves to some extent. These effects open up the active site, suggesting that a fairly marked conformational change occurs upon binding of substrate. The Michael acceptor inhibitor also pushes His30 away from the other members of the catalytic triad (Cys139 and Glu54, Fig. 1c).

The SV3CP enzyme has approximately 90% sequence identity with other GI noroviral 3Cproteases and an identity of the order of 68% with the enzyme from the GII genotype. SV3CP has approximately 58% identity with the mouse norovirus enzyme. The monomer structures of these enzymes superpose with SV3CP with a Cα RMSD of typically 1.0 – 1.2 Å for virtually all of the amino acids in the chains. The structures differ most noticeably in the hairpin linking strands β9 and β10 which is close to the active site.

In line with other noroviral 3Cproteases which have been analysed by gel-filtration, it is highly likely that SV3CP forms dimers in solution or, at least, exists in a monomer – dimer equilibrium (Chang et al., 2012; Leen et al., 2012; Zeitler et al., 2006). Accordingly, a dimer is observed in the crystallographic asymmetric unit of SV3CP (Fig. 2, chains A and B). However, analysis with the PDBePISA website (Krissinel and Henrick, 2007) suggested a tetrameric form (Fig. 1b) which might also be stable in solution. The interface area between the chains of the crystallographically observed dimers (formed by chains A and B) is 883.0 Å². However, a neighbouring dimer in the crystal structure forms an interface of comparable buried surface area (692.3 Å²) between chains labelled A and D chains and likewise for chains labelled the B and C. This result indicates that higher order oligomers may possibly be formed by SV3CP dimers, such as the putative tetramer shown in Fig. 1b. Intriguingly, a number of other human GI and GII noroviral protease structures (Nakamura et al., 2005, RCSB ID: 1wqs; Muzzarelli et al., 2019, RCSB ID: 6661; Viskovska et al., 2019, RCSB ID: 6nir) form essentially the same tetrameric assembly in the crystals, as shown in Supplementary Fig. 1. The majority of the amino acids forming the tetramer contacts (Glu79, Met101, Ala103, Ala105, Ser106, Met107, Arg108, Met120, Leu121, Leu122, Thr161, Ser163, Asn165, Thr166) are either invariant or conservatively substituted. The same tetramer is also observed in mouse norovirus protease (Fernandes et al., 2015,...

### Table 1 (continued)

| Fragment | J16 | J17 | J18 | J19 |
|----------|-----|-----|-----|-----|
| N observed | 175,458 | 106,011 | 163,299 | 89,757 |
| N unique | 56,264 | 31,951 | 52,139 | 27,348 |
| R-factor | 13.75 | 14.52 | 14.35 | 13.70 |
| Test set size | 2826 | 1559 | 2622 | 1339 |
| R.m.s.d. bond lengths (Å) | 0.015 | 0.011 | 0.013 | 0.010 |
| Mean protein β-factor (Å²) | 26.59 | 28.41 | 20.74 | 34.66 |
| Mean fragment β-factor (Å²) | 26.53, 0.50 | 40.47 | 23.78 | 56.50 |

### 2.4. Activity assay

The protease (0.5 mg/ml final concentration) in a buffer containing 100 mM Tris, pH 8.5, and 5 mM β-mercaptoethanol was mixed with the fragment (dissolved in DMSO at concentrations of 0.027, 0.135, 0.27, 0.405 and 0.54 mM) for 20 min at RT. The solution was then mixed with the chromogenic substrate (Ac-EFQLQ-para-nitroaniline; Peptide Protein Research Ltd, Southampton, UK), which was dissolved in DMSO to give final concentrations of 0.4, 0.9, 1.4, 1.9, 2.5 and 3.0 mM, in a 1:1 ratio and the absorbance at 405 nm was measured at 20 s intervals over a 3 min period, using a Nanodrop ND1000 spectrophotometer. The Kᵢ values were determined using GraphPad Prism (www.graphpad.com).
Fig. 1. (a) The overall structure of SV3CP with the β-strands labelled. The protein is composed of an N-terminal domain (red) containing a twisted 7-stranded antiparallel β-sheet and a C-terminal domain (green) consisting of a 6-stranded β-barrel. The catalytic triad of Cys139, His30 and Glu54 is shown in stick representation. (b) The putative tetrameric form of SV3CP. The asymmetric unit is composed of a dimer formed by chains denoted A and B. Crystallographic symmetry generates another identical dimer formed by chains denoted C and D. Both the AB and CD dimers have significant buried surface area suggesting that the dimer is physiologically significant. However another stable interface is formed between chains A and D and between chains B and C, resulting a tetrameric ensemble which may also have physiological significance. The domains of chain A are coloured and oriented the same as in (a) whereas chains B, C and D are coloured yellow, cyan and ochre. (c) Superimposition of the protein moieties of ligand-free SV3CP (green) and the polypeptide inhibitor-bound structure (Hussey et al., 2011; RCSB ID: 2iph) (pale brown). The ligand-free SV3CP structure has a shorter C-terminal end labelled C1 whereas this part of the complex with the substrate-analogue inhibitor is labelled C2. The red asterisk indicates the β-hairpin (between strands β9 and β10) which undergoes substantial movement upon binding the inhibitor. (d) A close-up of the superposition showing the polypeptide Michael acceptor inhibitor (MAPI) coloured cyan and a number of residues in the β-hairpin, including Arg112 which moves drastically upon binding the inhibitor. (e) A surface representation of ligand-free SV3CP demonstrating the relatively closed state of the active site cleft (arrowed A). (f) A surface representation of the complex with MAPI in which the active site has opened appreciably to accommodate the polypeptide moiety of the inhibitor.
RCSB ID: 4x2v) which has lower sequence identity with SV3CP (~58%) than do the other human GI or GII proteases (~91% and 68%, respectively). These findings, along with the ability of the tetrameric cavity to bind small molecule fragments (see later), suggest that this tetrameric form may have functional significance for 3CLpro. Indeed, in the structure of the Minerva virus enzyme (RCSB ID: 6b6i) the tetrameric assembly allows the C-terminal end of two monomers (equivalent to B and D) to extend into the active sites of adjacent monomers (C and A, respectively) across the dimer-dimer interface (Muzzarelli et al., 2019). Similar tail-interdigitating effects are observed in the structures of the protease from Houston virus (RCSB ID: 6nir; Viskovska et al., 2019) and mouse norovirus (RCSB ID: 4x2v; Fernandes et al., 2015). Given that localised replication centres are known to form within norovirus-infected cells (e.g. Thorne and Goodfellow, 2014), a high local concentration of 3CLpro may allow the enzyme to tetramerise.

In the native SV3CP structure, no electron density is visible for the last 8 residues (ASEGETTL) at the C-terminal end of the protein. Since these residues are well-defined in the complex with a substrate analogue (Hussey et al., 2011), their absence in the native structure might be due to autolysis during storage or crystallisation of the uninhibited protease. In this region of the structure, there is a minor consensus sequence for SV3CP cleavage with the following amino acids VQ-AS corresponding to the P2-P1-P1′-P2′ positions (Hussey et al., 2011; Kankanamalage et al., 2015) suggesting that slow autolysis prior to crystal growth is possible. Mass spectrometric analysis of the purified protein yielded a molecular mass of 19,290 Da (Supplementary Fig. 2) confirming that the protease was indeed fully intact at the time of crystallisation. Therefore another possibility is that this region of the molecule is simply disordered in the new crystal form. However, it is not clear why this should be since this region of both monomers is not involved in crystal contacts in either crystal form.

3.2. Crystal-based fragment screening

Most crystals used in the non-covalent fragment screening experiment diffracted to resolutions ranging from 1.5 to 1.8 Å with good

Fig. 2. Binding sites of the fragment screening hits. All the ligands bind in five sites labelled as A–E which are shown on a surface representation of the dimer. Site A coincides with the active site of the protease and site B is a putative RNA binding site. Site C coincides with the centre of the tetramer.

Fig. 3. Molecular structures of the fragment screening hits.
crystallographic statistics (Table 1). Fragment J12 is the worst in terms of resolution, diffracting to approximately 2.1 Å, although the electron density is still of good quality. Screening with the DSPL library and part of the Maybridge Ro3 library identified 19 ligands in total which bind in five different sites, as illustrated in Fig. 2. The majority of fragments have mean $B$-factors which are comparable with those of the protein moieties (Table 1). In only one case (J02) was the occupancy of the fragment refined, although for several others it was set to 0.5 due to the fragments residing on a 2-fold axis.

Site A, the protease active site, is a long groove containing the catalytic Cys139 residue. Two fragments (J01 and J02) were found to bind here, each on different sides of the catalytic cysteine (Fig. 2). Five hits (J03-J07) were found to bind in the putative RNA binding site (site B) including one (J07) which also binds in another site, site C. Site C lies in a pocket between chains A and B and the symmetry related chains A' and B', with 11 hits being identified (J07-J17) here. Two other fragments were found at additional sites: D (J18) and E (J19). Molecular structures of the ligands J01-J19 are given in Fig. 3.

3.2.1. Active site-binding fragments (site A)

Two non-covalently bound fragments were identified in the active site of the protease named as J01 and J02, as indicated by their omit maps (Fig. 4a and c). J01 binds in the S1 subsite where its carboxyl group is oriented towards S2 and S3. J01 forms several direct hydrogen bonds with the side chains of Gln110 and Arg112 and makes some additional hydrogen bonds mediated by a water molecule (Fig. 4a and b). These residues are at the tip of the functionally important $\beta$-hairpin (connecting strands $\beta 9$ and $\beta 10$) that is involved in substrate recognition and moves substantially upon binding of polypeptide substrate analogues (Fig. 1). However, in the presence of J01, the $\beta$-hairpin adopts the same conformation as the ligand-free SV3CP, suggesting that binding of this fragment does not alter its conformation. Since the carboxyl group of J01 appears to hold the $\beta$-hairpin loop (residues 109 to 112) in the closed conformation, this must help to prevent the enzyme from adopting the ‘open’ conformation that can accommodate the substrate. The ligand –NH group (N1) is also within hydrogen bonding distance of the main chain carbonyl group of Thr134. The benzoic acid moiety of J01 makes many hydrophobic interactions with the active site residues including Pro136, Cys139 and Ala160. In contrast, the 5-methyl-2-thienyl group forms fewer contacts with the enzyme than the aromatic group since it points away from the active site towards a large solvent channel.

J02 resides on the other side of the long active site, where it occupies the S2 subsite without forming any hydrogen bonds (Fig. 4c and d). Instead, the phenyl ring is sandwiched between the side chains of His30 of the catalytic triad and Arg112 from the $\beta$-hairpin loop by $\pi$-$\pi$ interactions.
stacking and cation-π interactions. Interestingly, the guanidinium group of Arg112 has moved from its position in the other fragment complex to accommodate J02. Several hydrophobic interactions are formed between this fragment and Glu54 from the catalytic triad and Val114, and a number of contacts are made with a symmetry-related molecule.

In kinetic assays both J01 and J02 showed inhibitory activity against SV3CP with $K_i$ values of 0.37 mM and 0.34 mM, respectively. These values are typical of initial hits in crystallographic fragment-screening studies targeting catalytic- or allosteric-sites of enzymes (Bauman et al., 2013; Delbert et al., 2018; Zhang et al., 2019) suggesting that the binding modes we observe in 3CLpro are highly relevant. Since J01 and J02 bind in the active site cleft and maintain the closed conformation of the hairpin, they are good candidates for developing further inhibitors and linking them into a new compound could also improve the bioactivity. A superposition of their binding modes on that of the covalently bound Michael acceptor inhibitor (Fig. 5) demonstrates how these two fragments occupy the S1 and S2 subsites, respectively. J02 does not overlap with the P2 residue of the polypeptide inhibitor as well as J01 and the P1 residue do, since it appears to lie somewhere between the spatially adjacent S2 and S1′ subsites.

### 3.2.2. Fragments binding at the putative RNA binding site (site B)

In addition to the protease activity, studies on viral 3C proteases suggested that they or their larger precursors can bind specifically to the 5′-terminal nucleotides of the viral RNA (Leong et al., 1993; Nayak et al., 2006). The interaction occurs only on the plus strand which forms a ribonucleoprotein (RNP) complex that is necessary for the initiation of the plus strand synthesis (Andino et al., 1990). It has been shown that

![Fig. 5](image-url) A superposition of the two active-site binding fragment structures on a covalently bound substrate analogue inhibitor. The complexes with J01 and J02 are coloured magenta and green, respectively, while the polypeptide Michael acceptor inhibitor structure (RCSB ID: 2iph; Hussey et al., 2011) is shown in cyan. The C-terminal end of ligand-free SV3CP is labelled C1 and the corresponding part of the polypeptide inhibitor complex is labelled C2. The β-hairpin loop connecting strands β9 and β10 moves significantly from its position in the native structure (which is very close to its position in the J01 and J02 complexes) upon binding the polypeptide inhibitor.

![Fig. 6](image-url) Interactions between the putative RNA binding site (site B) of the protease and fragments J03-J07. These are shown in 3D with the omit electron density contoured at 1.0 RMSD as (a, c, e, g, i) and in 2D with interacting residues shown in (b, d, f, h, j), respectively. Hydrogen bonds are indicated by dashed lines in cyan and hydrophobic interactions are indicated by red eyebrow-like icons. Protein chain identifiers are indicated by the letters A and B in brackets and those with a prime are from symmetry-related chains.
human noroviral RNA non-competitively inhibits the protease activity with an IC50 of in the µM range (Viswanathan et al., 2013). The RNA binding site has been studied by mutagenesis in other homologous 3C proteases, in which a key arginine residue was identified in the conserved sequence, KF/VRDI (F/V represents F or V) (Bergmann et al., 1997; Leong et al., 1993; Nayak et al., 2006). Structural comparison of SV3CP with HRV 3CLpro (PDB ID: 5fx5; Kawatkar et al., 2016) and FMDV 3CLpro (PDB ID: 2j92; Nayak et al., 2006) identified Arg65 as the equivalent residue in SV3CP, which is within a KIRPDL sequence that has similarity with the consensus. The Rand D residues in this sequence interact by a salt-bridge that forms one side of the putative RNA binding site of SV3CP (site B) which is shown in Fig. 2 and, as for the FMDV and HRV proteases, it is a shallow groove. In addition, these sites are on the surface of the SV3CP tetramer and form deep channels with the neighbouring symmetry-related molecules in HRV, FMDV and Southampton virus 3CLpro. Inhibitors binding in the RNA binding site have the potential to inhibit noroviral replication and are therefore of interest as a separate class of drug.

Fragments J03-J06 were found to reside at this site and their contact residues are shown in Fig. 6. All the fragments form hydrophobic contacts with Arg65 and other residues in the KIRPDL sequence. While J03 (Fig. 6a and b) and J06 (Fig. 6g and h) are mainly involved in hydrophobic interactions, J04 (Fig. 6c and d) and J05 (Fig. 6e and f) also form many hydrogen bonds with the neighbouring residues, potentially making them stronger binders. The carbonyl group (O1) of J04 is involved in three hydrogen bonds formed, directly or mediated by a water molecule, with Thr10, Lys11 and Ser91 (although the latter residue is from a symmetry related molecule). The N1 atom forms two hydrogen bonds with Ser7 and Pro3 (also from the symmetry mate) with the participation of a water molecule. A hydrogen bond is also seen between the fluorine substituent in the indole ring of J04 and the NE1 atom on the side chain of Trp19. This residue is one of a number of quite solvent-exposed aromatic residues including phenylalanines 12, 25, 39 and 40 which form the putative RNA-binding site. J05 also forms water-mediated hydrogen bonds with Ser91 from the symmetry related molecule. Unlike the active site fragments which bind in different subsites of the substrate-binding channel, these four fragments bind in approximately the same position with their aromatic ‘heads’ overlapping to a large degree but their aliphatic ‘tails’ pointing away in different directions. Since binding of viral RNA inhibits the protease activity (Viswanathan et al., 2013), ligands binding at this site have the potential both to interfere both with RNA binding and with the protease activity. However, since this site is of the order of 20 Å from the catalytic centre the mechanism of protease inhibition is currently difficult to explain. Fragment J07 was found to bind in both the putative RNA binding site (B, Fig. 6i and j) and site C (Fig. 7a and b) in the centre of the putative tetramer.

### 3.2.3. Fragments binding in the tetramer cavity (site C)

The finding that the native crystals of the enzyme are formed by a tetrameric assembly of monomers is suggestive of a physiological role for the tetramer. We were also intrigued to find that the majority of the fragments binding to the protease (J07 - J17, Fig. 7) were located in a cavity at the centre of the putative tetramer, site C. The site is characterised by the convergence of two-fold symmetry axes, both crystallographic and non-crystallographic, since the NCS two-fold relating the monomers in each dimer and the crystallographic two-fold relating both dimers in the tetramer meet at this point. The binding site is formed by four copies of the hydrophobic amino acids Leu122 and Val82 as well as Arg100 which are provided by all chains of the
formed at the convergence of 2-fold axes, two copies of each ligand are present at this site and sometimes the two symmetry-related copies of the fragment interact extensively with each other. Since the same tetrameric assembly is observed in other GI and GII norovirus proteases, this binding site may be a conserved feature of these enzymes. Given its ability to bind so many heteroaromatic fragments and the diverse functions which noroviral proteins and their precursors are known to have (e.g. Emmott et al., 2019), it is tempting to speculate that the tetramer cleft has a physiological role, perhaps even as a secondary substrate- or RNA-binding site.

3.2.4. Other fragment binding sites (D and E)

Two of the fragments (J18 and J19, Fig. 8) were found to bind at unrelated sites involving crystal contacts which are probably not of physiological significance. Site D lies close to Lys11, Lys88 and Glu93 whereas site E lies between Arg59 and the C-terminal end of the enzyme. The amide bond within J18 has apparently been cleaved and the resulting fragments, trifluoroacetic acid and 2-ethyl-1,3,4-thiadiazole, bind at sites C and D, respectively. Interestingly, it appears that the amide bond in J11 has also been cleaved and the resulting 2-ethyl-1,3,4-thiadiazole binds instead at site C. A check on the stock solution of this compound was made mass spectrometry and this yielded a main mass of 130 Da, which is within a dalton of the predicted molecular mass of the observed fragment. It is possible that the electron-withdrawing groups on the amino terminal side of the amide bonds of these two compounds may render them unstable in water.

4. Discussion

The X-ray structure of the Southampton virus 3CL\textsuperscript{pro} has been determined at 1.3 Å resolution in a crystal form that has allowed fragment-screening for novel inhibitors to be undertaken at similar resolutions. Two fragments were found to bind in the active site cleft of the protease. J01 and J02 bind in different subsites of the long active site (see Fig. 5) but both of them interact with the functionally important β-hairpin linking strands β9 and β10. J01 occupies S\textsubscript{1} and forms hydrophobic interactions with catalytic Cys139 while J02 occupies S\textsubscript{2} and forms hydrophobic and π-π interactions with Glu54 and His30, which are also from the catalytic triad. Both J01 and J02 could potentially be developed into more potent norovirus protease inhibitors, however, a better ligand might ultimately be obtained by coupling them together, given that the distance between the closest two atoms is slightly less than 3.8 Å.

Some of the remaining fragments were found to interact with the protease at its putative RNA-binding site. Whilst these compounds are likely to have less effect on the protease activity than J01 and J02, which bind in the active site, RNA binding to the enzyme has been shown to cause non-competitive inhibition of the protease (Viswanathan et al., 2013). Other fragments were found to bind at an additional site which is buried deeply in the centre of the crystallographic tetramer. The fact that a C193A mutant of the Minerva virus protease forms the same tetramer in the crystal with the C-terminus of one subunit occupying the active site cleft of another monomer (Muzzarelli et al., 2019), suggests that this assembly may also be involved in proteolytic maturation of noroviruses. Hence, compounds that have the potential to interfere with formation of the tetramer or affect its stability may impact on noroviral replication and therefore deserve to be screened for in vivo activity, e.g. against mouse norvirus, which can be cultured, or in a suitable replicon assay. If such studies were to be successful, the highly symmetric nature of the binding site is something that could, in principle, be exploited in drug design.

Given the recent COVID-19 pandemic, it is potentially useful to compare our results on SV3CP with the 3CL\textsuperscript{pro} of coronavirus (e.g. Yang et al., 2013). The two enzymes have quite low sequence identity of approximately 12% within the common protease moieties and superimpose with an RMSD of 2.4 Å for 126 structurally aligned residues.
The coronavirus protease is considerably larger (303 residues) than SV3CP due to the presence of a C-terminal domain which is involved in dimerisation. Although topologically similar, the protease moieties of both structures differ very substantially in the loop regions connecting the core β-strands. In spite of these differences, coronavirus protease also has specificity for Gln at the P1 position of substrate. In very recent fragment screening of the SARS-CoV-2 protease, 23 active site hits were obtained which span the S3 to S1’ subsites of the enzyme, thus providing

Fig. 8. Interactions made by fragments J18 and J19. The amide bond within J18 has apparently been cleaved and the resulting fragments, 2-ethyl-1,3,4-thiadiazole and trifluoroacetic acid, bind at sites D (a, b) and C (c, d), respectively. J19 binds at site E (e, f). These are shown in 3D with the omit electron density contoured at 1.0 RMSD. Hydrogen bonds are indicated by dashed lines in cyan and hydrophobic interactions are indicated by red eyebrow-like icons. Protein chain identifiers are indicated by the letters A and B in brackets and those with a prime are from symmetry-related chains.
somewhat better coverage of the active site cleft than we have achieved with SV3CP (Douangamath et al., 2020). Other SARS-CoV-2 protease inhibitor structures have also been reported in recent months (Dai et al., 2020; Jin et al., 2020a,b; Zhang et al., 2020). This resurgence of interest in rational 3CL**pro drug design is likely to have combined benefits for what are currently intractable and severe viral infections. These studies provide a rational basis on which compounds with improved potency can be designed by medicinal chemists.

CRediT authorship contribution statement

Jingyu Guo: Conceptualization, Investigation, Methodology, Writing - original draft. Alice Douangamath: Methodology. Weixiao Song: Methodology, Investigation. Alun R. Coker: Supervision, Investigation. A.W. Edith Chan: Supervision, Investigation. Steve P. Wood: Supervision, Investigation. Jonathan B. Cooper: Supervision, Investigation, Writing - review & editing. Efrat Resnick: Investigation, Methodology. Nir London: Conceptualization, Methodology, Supervision. Frank von Delft: Conceptualization, Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Profs P. M. Shoolingin-Jordan and I. N. Clarke (University of Southampton) for providing the expression construct and for numerous helpful discussions. We also thank Anthony Aimon (DLS Ltd) for fragment mass-spectrometric analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2020.100031.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Mritunjay, N., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular crystallography. Acta Crystallogr. Sect. D. Biol. Crystallogr. 66, 12–21.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Kapral, G.J., Mritunjay, N., Murshudov, G.N., Winn, M.D.,和技术性的扩展导致了第一个报告的抑制剂的PHIP(2)，一个非典型的大肠杆菌。J. Med. Chem. 127, 41–61.

Barta, F., Brans, M., Grütz, H., Nopen, S., Peigné, S., Boland, S., Chaltin, P., Brandao-Neto, J., von Delft, F., Touw, W.G., Joosten, R.P., Liekens, S., Tytgat, J., Ulsen, C., 2018. An allosteric binding site of the a2 nicotinic acetylcholine receptor revealed in a humanized acetylcholine-binding protein. J. Biol. Chem. 293, 2534–2545.

Division of Viral Diseases, N.C.F.I., Respiratory Diseases, C.D.F.C., Prevention, 2011. Updated norovirus outbreak management and disease prevention guidelines. MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports 60, 1–18.

Douangamath, A., Feareon, D., Geelhoft, P., Krojer, T., Lukacik, P., Owen, C.D., Resnick, E., Strain-Damerell, C., Aimon, A., Abrahámy-Balogh, P., Brandão-Neto, J., Carbery, A., Davidson, G., Diza, A., Downes, T.D., Dunnett, L., Fairhead, M., Firth, J.D., Jones, S.P., Keeley, A., Keserü, G.M., Klein, H.F., Martin, M.P., Noble, M.E.M., O’Brien, P., Powell, A., Resti, R., Skynier, R., Snee, M., Waring, M.J., Wild, C., London, N., von Delft, F., Walsh, M.A., 2020. Crystallographic and electrostatic fragment screening of the SARS-CoV-2 main protease. BioRxiv. https://doi.org/10.1101/2020.05.27.118117.

Emmott, E., de Rougemont, A., Hosmillo, M., Lu, J., Fitzmaurice, T., Haas, J., Goodfellow, I., 2019. Polypeptide processing and intermolecular interactions within the viral replication complex spatially and temporally control norovirus protease activity. J. Biol. Chem. 294, 4259–4271.

Emwasy, P., Coktan, K., 2004. Cot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D. Biol. Crystallogr. 60, 2126–2132.

Fernandes, H., Leen, E.N., Cromwell Jr, H., Pfeil, M.P., Carry, S., 2015. Structure determination of Murine Norovirus N6 protease with C-terminal extensions designed to probe protease-substrate interactions. PeerJ 3, e798. https://doi.org/10.7717/peerj.798.

Florescu, D.F., Hill, L.A., McCartan, M.A., Grant, W., 2008. Two cases of Norwalk virus outbreak following small bowel transplantation treated with oral human serum in monomugulin. Pediatr. Transplant. 12, 372–375.

Hu, J., Liao, Z., Jin, Zhenming, Du, Xiaoyu, Xu, Yechun, Deng, Yongqiang, Liu, Meiin, Zhao, Yao, Zhang, Bing, Xu, Xiaofeng, Zhang, Lihua, Peng, Chao, Duan, Yinkai, Yu, Jing, Wang, Lin, Yang, KaiLin, Liu, Jiang, Renyi, Wang, Zhe, Xiao, Xue, Bao, Fang, Liu, Hong, Xiang, Li, Xiao, Xuan, Xinna, Bai, Fang, Liu, Hong, Xiang, Gao, Wei, Xue, Yi, Xue, Lian, Hao, Li, Xue, Xia, Xiao, G., Jiang, H., Rao, Z., Zhang, L.-K., Xu, Y., Yang, H., Liu, H., 2020. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 368 (6497), 1331–1335. https://doi.org/10.1126/science.abb4489.

Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., Li, C., Ge, X., Sun, F., Wang, H., Jiang, Y., Wang, J., Liu, X., Xiao, G., Jiang, H., Rao, Z., Zhang, L.-K., Xu, Y., Yang, H., Liu, H., 2020. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 368 (6497), 1331–1335. https://doi.org/10.1126/science.abb4489.

Douangamath, A.C.G., Lushington, G.H., Neto, J., Battale, K.P., Lovell, S., Chang, K.-O., Goutas, W.C., 2017. Design, synthesis, and evaluation of a novel series of macrocyclic inhibitors of norovirus 3CL protease. Eur. J. Med. Chem. 127, 41–61.

Andino, R., Rieber, T.S., Collisson, T., Stechschulte, P., Carnaghi, P., Harrison, D.C., 2015. The proteolytic activity of coronavirus transmembrane proteins by a 3C-like cysteine proteinase. Virology 308, 216–224.

Birley, J.R., Knox, S.R., Jaulent, A.M., Brick, P., Leatherbarrow, R.J., Curry, S., 2005. Construction of the 3C Gene Product From Hepatitis A Virus: Specific Proteinase Activity and RNA Recognition. J. Virol. 71, 2436–2448.

Bergmann, E.M., Mosimann, S.C., Chernaia, M.M., Malemdin, R.A., James, M.N., 1997. The refined crystal structure of the SC Gene Product From Hepatitis A Virus: Specific Proteinase Activity and RNA Recognition. J. Virol. 71, 2436–2448.

Bertley, J.R., Knox, S.R., Jaulent, A.M., Brick, P., Leatherbarrow, R.J., Curry, S., 2005. Construction of the 3C Gene Product From Hepatitis A Virus: Specific Proteinase Activity and RNA Recognition. J. Virol. 71, 2436–2448.
Nakamura, K., Someya, Y., Kumasaka, T., Ueno, G., Yamamoto, M., Sato, T., Takeda, N., Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., Liu, L., Johnson, H.L., Cousens, S., Perin, J., Scott, S., Lawn, J., Rudan, I., Campbell, H., Lew, J.F., Kapikian, A.Z., Jiang, X., Estes, M.K., Green, K.Y., 1994b. Molecular characterization of human norovirus GII.4 variants, 2001–2007. J. Infect. Dis. 200, 802–812.

Nayak, A., Goodfield, I.G., Woolaway, K.E., Birley, J., Curry, S., Belsham, G.J., 2006. Role of RNA Structure and RNA Binding Activity of Foot-and-Mouth Disease Virus 3C Protein in Vpg Uridylation and Virus Replication. J. Virol. 80, 9865–9875.

Netzer, N.E., Enos Tuipulotu, D., White, P.A., 2019. Norovirus antivirals: Where are we now? Med. Res. Rev. 39, 860–886.

Ng, J.T., Dekker, C., Kroemer, M., Osborne, M., von Delft, F., 2014. Using textons to rank crystalization droplets by the likely presence of crystals. Acta Crystallogr. D Biol. Crystallogr. 70, 2702–2718.

Pearse, N., Bradley, A.R., Collins, P., Trowek, T., Nowak, R., Talon, R., Marden, B.D., Kelm, S., Shi, J., Deane, C., 2016. A Multi-Crystal Method for Extracting Obscured Signal from bioRxiv. Crystallogr. Electr. Density., 073411.

Phillips, G., Tam, C.C., Conti, S., Rodrigues, I.C.L., Brown, D., Irtriz-Gomara, M., Gray, J., Logan, B., 2010. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. Am. J. Epidemiol. kwq021.

Resnick, E., Bradley, A., Gan, J., Douangamath, A., Kroser, T., Sethi, R., Geurink, P.P., Almon, A., Amis, G., Bellini, D., Bennett, J., Fairhead, M., Fedorov, O., Galbizon, B., Gan, J., Guo, J., Plotnikov, A., Reznik, N., Ruda, G.F., Díaz-Sáez, L., Straub, V.M., Szommer, T., Velupillai, S., Zaidman, D., Zhang, Y., Coker, A.R., Dowson, C.G., Barr, J.H., Wang, C., Huber, K.V.M., Brennan, E.O., Oza, H., von Delft, F., Lendon, N., 2019. Rapid Covalent-Peptide Discovery by Electrophile-Fragment Screening. J. Am. Chem. Soc. 141, 8951–8968.

Scalani, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., Jones, J.L., Griffin, P.M., 2011. Foodborne illness acquired in the United States-major pathogens. Emerg. Infect. Dis. 17, 7–15.

Siebenga, J.J., Vennema, H., Zheng, D.P., Vinje, J., Lee, B.E., Pang, X.L., Ho, E.C., Lim, W., Choudaker, A., Broer, S., 2009. Norovirus illness is a global problem: emergence and spread of norovirus GI 4 variants, 2001–2007. J. Infect. Dis. 200, 802–812.

Someya, Y., Takeda, N., Miyamura, T., 2005. An orovirus protease structure provides insights into Antiviral Studies of the Human Norovirus GII.4 Protease. Biochem. 58, 900–907.

Srivastava, P., Zhang, R., McIntyre, P.J., Collins, P.M., Foley, D.J., Arter, C., von Delft, F., Bayliss, R., Lister, T., Olsson, L., Patel, J., 2016. Design and Structure-Activity Relationships of Novel Inhibitors of Human Rhinovirus 3C Protease. Bioorg. Med. Chem. Lett. 26, 3248–3252.

Thorne, L.G., Goodfield, I.G., 2014. Norovirus gene expression and replication. J. Gen. Virol. 95, 278–291.

Trowek, S., Nee, S., Wang, S., Neubauer, S., Kula, M., Stoeck, C., Lopman, B., 2010. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. Am. J. Epidemiol. kwq021.

Vega, E., Barclay, L., Gregorius, N., Shirley, S.H., Lee, D., Vinje, J., 2014. Genotypic and epidemiological trends of norovirus outbreaks in the United States, 2009 to 2013. J. Clin. Microbiol. 52, 147–159.

Vinjé, J., 2015. Advances in laboratory methods for detection and typing of norovirus. J. Clin. Microbiol. 53, 373–381.

Viskovska, M.A., Zhao, B., Shanker, S., Choi, J.-M., Deng, L., Song, Y., Palzkill, T., Hu, L., Estes, M.K., Venkataram, P.B., 2019. Norovirus Protease Shows pH-Sensitive Proteolysis with a Unique Arg-His Pairing in the Catalytic Site. J. Virol. 93, e01479–e1518.

Wang, Y., Yang, Kailin, Liu, Xiang, G., Guddat, L.W., Xiao, Gengfu, Zhang, Leike, Yang, H., Roca, Z., 2020. Structural and molecular crystal structures. Acta Crystallogr. Sect. A 69, s299.

Yelamos, J., John, B., 2015. Advances in laboratory methods for detection and typing of norovirus. J. Clin. Microbiol. 53, 373–381.

Zeitler, C.E., Estes, M.K., Prasad, B., 2006. X-ray crystallographic data from a Desert Shield troop with gastroenteritis. Virology 200, 319–325.

Zhou, S., Liu, P., Kovari, I.A., Yang, Z., Schinazi, R.F., Kovari, L.C., 2019. Structural and molecular crystal structures. Acta Crystallogr. Sect. A 69, s299.