Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

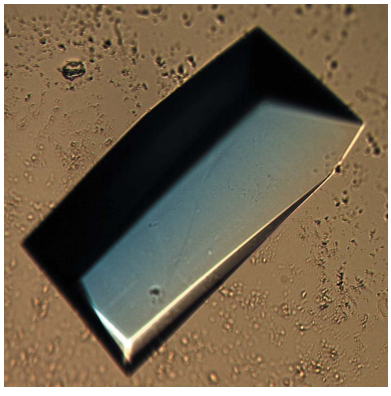
Crystallization and preliminary X-ray diffraction analysis of the protease from Southampton norovirus complexed with a Michael acceptor inhibitor

R. J. Hussey,[a] L. Coates,[b] R. S. Gill,[c] J. N. Wright,[a] M. Sarwar,[a] S. Coker,[c] P. T. Erskine,[c] J. B. Cooper,[c]* S. Wood,[c] I. N. Clarke,[d] P. R. Lambden,[d] R. Broadbridge[e] and P. M. Shoolingin-Jordan[a]

[a]School of Biological Sciences, University of Southampton, Southampton SO16 7PX, England, [b]Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA, [c]Laboratory of Protein Crystallography, Centre for Amyloidosis and Acute Phase Proteins, UCL Department of Medicine (Royal Free Campus), Rowland Hill Street, London NW3 2PF, England, [d]Molecular Microbiology Group, Division of Infection, Inflammation and Immunity, University Medical School, Southampton General Hospital, Southampton SO16 6YD, England, and [e]Peptide Protein Research Ltd, E2 Knowle Village Business Park, Wickham, Hants PO17 5DY, England

Correspondence e-mail: jbcooper@medsch.ucl.ac.uk

Received 17 June 2010
Accepted 29 September 2010



Noroviruses are the predominant cause of human epidemic nonbacterial gastroenteritis. Viral replication requires a cysteine protease that cleaves a 200 kDa viral polyprotein into its constituent functional parts. Here, the crystallization of the recombinant protease from the Southampton norovirus is described. Whilst the native crystals were found to diffract only to medium resolution (2.9 Å), cocrystals of an inhibitor complex diffracted X-rays to 1.7 Å resolution. The polypeptide inhibitor (Ac-EFQLQ-propenyl ethyl ester) possesses an amino-acid sequence designed to match the substrate specificity of the enzyme, but was synthesized with a reactive Michael acceptor group at the C-terminal end.

1. Introduction

Noroviruses are the most common cause of acute viral gastroenteritis in humans, with epidemics commonly occurring in hospitals and on ocean liners (Clarke & Lambden, 2005). The virus, which is transmitted through contaminated food and water, can infect and replicate in enterocytes of the epithelial cell lining of the small and large intestine (Green, 2007). Currently, there is neither vaccine nor antiviral therapy available.

The norovirus genome consists of a molecule of single-stranded positive-sense RNA (7.7 kb) comprising three open reading frames ORF 1, ORF 2 and ORF 3 (Lambden *et al.*, 1993, 1995). ORF 1, which is located at the 5′-terminus of the genome, encodes a large nonstructural 200 kDa polyprotein. ORF 2 encodes the major capsid protein VP1 and ORF 3 codes for a small basic protein VP2 that is thought to assist in the viral assembly process (Bertolotti-Ciarlet *et al.*, 2003). *In vitro* translation and mutagenesis studies indicated that the 200 kDa ORF 1 polyprotein is cleaved by the action of the viral protease to generate initially three separate functional protein products (Liu *et al.*, 1996). Full processing of the precursor polyprotein generates an N-terminal protein (p48), an NTPase (p41), a 3A-like protein (p22), a Vpg protein (p16), a 3C-like protease (p19) and an RNA polymerase (p57) (Liu *et al.*, 1999). The protease also inhibits cellular translation by cleavage of the poly(A)-binding protein, thereby allowing preferential viral protein expression compared with host proteins (Kuyumcu-Martinez *et al.*, 2004). Since processing of the 200 kDa precursor polyprotein is essential to yield functional viral proteins, the viral protease presents itself as an attractive target for antiviral strategies.

Enzymes in this family are cysteine proteases that display a trypsin-like or chymotrypsin-like serine protease fold, a property which distinguishes them from other viral proteases (Matthews *et al.*, 1994). The Southampton norovirus protease has a preference for cleavage at LQ–GP and LQ–GK sequences, but it can also cleave at ME–GK, FE–AP and LE–GG (where '–' indicates the scissile bond). In the nomenclature of Schechter & Berger (1967), the substrate residues each side of the scissile bond are labelled P1 and P1′ and the remainder are labelled according to the scheme ...P3, P2, P1, P1′, P2′, P3′.... The corresponding subsites in the enzyme are labelled S3, S2 *etc*. It appears that the Southampton norovirus protease

preferentially accommodates a glutamine or glutamate residue at the P1 position, a small amino acid at P1′ and a hydrophobic residue at P2. Modified peptide inhibitors that include the preferred amino-acid recognition sequence but possess a C-terminal moiety capable of reacting with the active-site cysteine residue have been developed for other viral cysteine proteases and *in vitro* studies have shown that these completely inhibit the catalytic activity and have antiviral properties *in vivo* (Dragovich *et al.*, 1998*a*,*b*, 2003). One such modified peptide inhibitor includes a Michael acceptor group at its C-terminus, which undergoes nucleophilic attack by the active-site thiol, resulting in the inhibitor becoming irreversibly bound to the enzyme (Fig. 1; Dragovich *et al.*, 1998*a*).

A number of noroviral proteases have been analysed by X-ray diffraction, *e.g.* those from the Chiba and Norwalk viruses (Nakamura *et al.*, 2005; Zeitler *et al.*, 2006). In this paper, we describe the crystallization of the Southampton norovirus protease, initially in a form that diffracted to medium resolution. A marked improvement in crystal quality was achieved by cocrystallization of the enzyme with the Michael acceptor peptide inhibitor (MAPI) acetyl-Glu-Phe-Gln-Leu-Gln-*X*, in which a peptide mimicking part of the natural substrate consensus sequence is coupled to a propenyl ethyl ester moiety (*X*) in order to modify the active-site cysteine. The resulting cocrystals belonged to space group $P2_12_12_1$ and diffracted synchrotron radiation to 1.7 Å resolution.

2. Protein expression and purification

The protease from Southampton virus was expressed in *Escherichia coli* BL21 (DE3) pLysS transformed with a plasmid pSV3C derived from pT7-7 (USB Corp.) harbouring DNA for the protease gene flanked by *Nde*I and *Bam*HI restriction sites that were introduced during amplification of the gene using standard PCR methods. Cells transformed with the plasmid were grown in Luria–Bertani medium with 50 µg ml^{-1} ampicillin in shaken flasks at 310 K and induced using isopropyl β-D-1-thiogalactopyranoside (IPTG), which was added to a final concentration of 1 m*M* for the last 3 h of bacterial cell growth. The harvested cells were sonicated and the supernatant was applied onto a column of SP Sepharose cation-exchange matrix (GE Healthcare) in phosphate buffer pH 7.65 containing 5 m*M* β-mercaptoethanol, followed by elution with a gradient to 1 *M* NaCl. After desalting with a Sephadex G25 column, the enzyme was applied onto a Source 15S column (GE Healthcare) in the same buffer, followed by elution with a gradient to 1 *M* NaCl. After a final desalting step with Sephadex G25, the protein was obtained with a yield of approximately 20 mg per litre of culture and was concentrated to 10 mg ml^{-1} and stored in 50% glycerol. The molecular weight of the purified enzyme was determined by electrospray mass-spectrometry as 19 258, which is consistent with the predicted amino-acid sequence of 181 residues and confirms that the protease has self-excised from the flanking sequences encoded by the expression construct.

Table 1
Kinetic data for the hydrolysis of pNA substrates by the Southampton virus protease.

The smallest substrate Ac-QLQ-pNA was not detectably cleaved and therefore no parameters could be determined.

Substrate	K_M (*M*)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
Ac-QLQ-pNA	—	—	—
Ac-FQLQ-pNA	1.5×10^{-3}	0.08	55
Ac-EFQLQ-pNA	3×10^{-4}	0.14	463
Ac-DEFQLQ-pNA	8×10^{-4}	0.33	416

3. Chromogenic substrate synthesis, kinetic assay and inhibitor synthesis

For kinetic studies of the protease specificity, a series of peptides were synthesized that provided a convenient spectrophotometric assay of its proteolytic activity. The chromogenic peptides Ac-QLQ-pNA, Ac-FQLQ-pNA, Ac-EFQLQ-pNA and Ac-DEFQLQ-pNA were synthesized using a combination of standard Fmoc solid-phase chemistry and synthetic techniques (Merrifield, 2007). Each peptide mimics residues of the protease-recognition sequence within the 200 kDa ORF 1 polyprotein which experiences the greatest rate of cleavage (DEFQLQ–GKMYDF; Liu *et al.*, 1999). All peptides were synthesized with an acetylated N-terminus and a C-terminus linked to a *para*-nitroaniline group (Whitmore *et al.*, 1995; Kaspari *et al.*, 1996). MALDI-Q-TOF-MS was used to confirm correct synthesis following reverse-phase purification of each product in DMSO. In the assay, cleavage of the C-terminal *para*-nitroanilide (pNA) group yields free *para*-nitroaniline, which can be followed spectrophotometrically at 405 nm. The C-terminal chromogenic glutamine-pNA is a complicating factor during peptide-chain extension since the C-terminal residue of the peptide would normally be attached to the support resin *via* its main-chain carboxyl. In this instance, pNA has to be attached at the glutamine carboxyl and therefore this residue must be linked to the support resin by its side chain. Hence, the pNA derivative of glutamate, rather than glutamine, was synthesized first using Fmoc-L-Glu(γOtBu)OH, in which the side-chain carboxyl is protected with a tertiary butyl ester group. The glutamate side chain

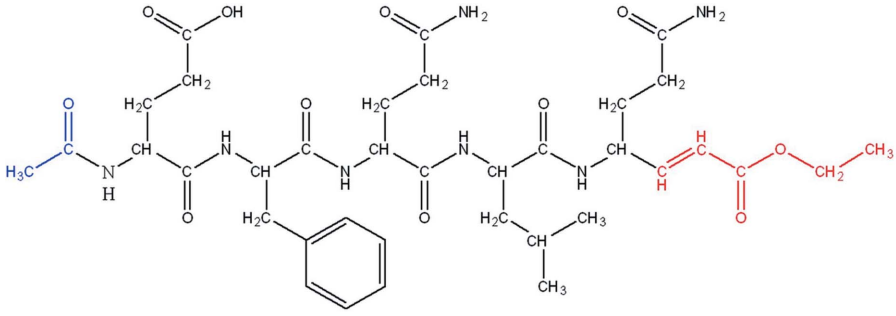


Figure 1
Structure of the Michael acceptor peptide inhibitor (MAPI) designed for the Southampton virus protease.

of the Fmoc-L-Glu(γOtBu)-pNA product was then selectively deprotected (with 95% trifluoroacetic acid) and linked to a Rink amide MBHA resin (Rink, 1987) using a standard procedure for the synthesis of peptide amides. Following completion of the remaining synthesis and deprotection cycles, cleavage of the desired peptide from the resin (with 95% trifluoroacetic acid) results in amination of the carboxyl linking group, which in this case yielded a glutamine residue at the C-terminus of the peptide-pNA.

The rate of cleavage of each pNA substrate by the protease was monitored using a Nanodrop ND1000 spectrophotometer. The assay involved dissolving the substrates in DMSO and diluting them into a solution containing 100 m*M* Tris pH 8.5 and 5 m*M* β-mercaptoethanol to give final substrate concentrations in the 0.1–3.0 m*M* range and a final enzyme concentration of 0.1 mg ml^{-1}. The absorbance at 405 nm of 2 µl samples taken from the reaction mixture was measured at 1 min intervals over a 10 min period. The initial rates of cleavage of the chromogenic pNA peptides established that Ac-EFQLQ-pNA was the best substrate in terms of specificity constant (k_{cat}/K_M ratio; see Table 1).

Subsequently, a polypeptide inhibitor (MAPI) with the same sequence as the optimal substrate and a Michael acceptor group at the C-terminus was synthesized, essentially by the methods described in Dragovich *et al.* (1998*a*), and purified by reverse-phase chromatography. The inhibitor (shown in Fig. 1) has the sequence Ac-EFQLQ-*X*, where *X* is the propenyl ethyl ester extension (the Michael acceptor) which undergoes nucleophilic attack by the active-site thiol. This generates a stable covalent complex between the enzyme and the inhibitor linked by a thioether bond (Govardhan & Abeles, 1996). For synthetic convenience, the peptide region of the inhibitor mimics only the nonprime residues of the bound substrate, *i.e.* P5–P1. The success of each step in the synthesis and the purity of the final compound was confirmed by MALDI-Q-TOF-MS. Since this inhibitor was found to inactivate the enzyme irreversibly and rapidly (<1 min), no further compounds were synthesized for the crystallographic work.

4. Crystallization and preliminary X-ray analysis

For all crystallization studies, samples of protease that were stored in glycerol were exchanged using a Sephadex G25 mini-column into 10 m*M* phosphate buffer pH 7.45 containing 5 m*M* β-mercaptoethanol and then concentrated to 3 mg ml^{-1}. Crystals of the native enzyme were obtained by the vapour-diffusion method in several conditions at room temperature using the Jena Biosciences JBScreen Classic screens; following further screening the optimum conditions were found to be 7% PEG 8000, 0.1 *M* HEPES pH 7.5 with 8%(*v*/*v*) ethylene glycol (Fig. 2). Crystals were frozen by transferring them

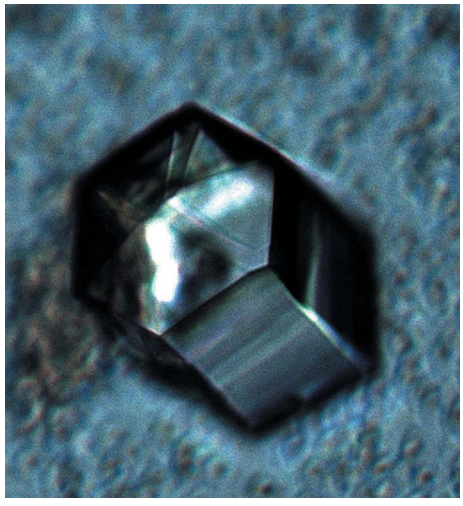


Figure 2
A crystal of native Southampton virus protease.

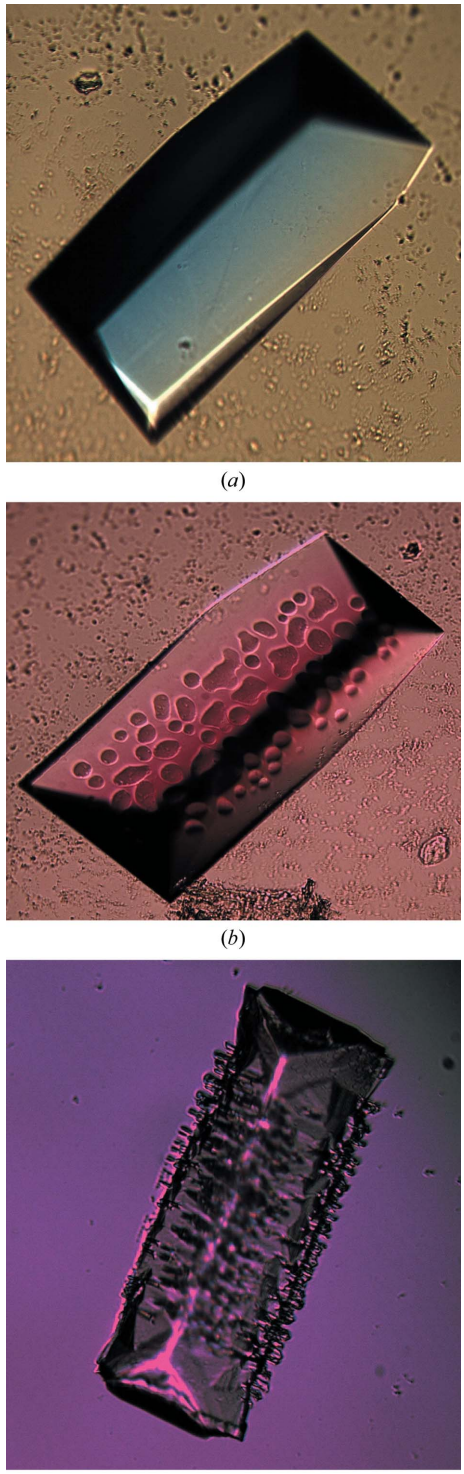


Figure 3
A cocrystal of Southampton virus protease and the inhibitor MAPI. A freshly grown crystal is shown in (*a*) and its progressive deterioration over several days is shown in (*b*) and (*c*).

into 10 µl well solution using a loop and stirring in four 1 µl droplets of glycerol stepwise to give a final glycerol concentration of approximately 30%(*v*/*v*). They were then frozen by plunging them into liquid ethane in a liquid-nitrogen bath. Preliminary data collection using station ID14-1 at the ESRF (Grenoble, France) revealed that the crystals belonged to a hexagonal point group, with unit-cell parameters $a = b = 129.5$, $c = 119.7$ Å. The best crystal only produced diffraction data to medium resolution of rather marginal quality ($d_{\rm min} = 2.9$ Å; $R_{\rm merge} = 14.5\%$ assuming space group *P*6) and most of the crystals tested only diffracted to between 3 and 4 Å resolution. All attempts at structure analysis by molecular replacement using several search models with this data set were unsuccessful. Ultimately, it was not possible to determine the exact space group of this crystal form or to analyse its structure.

Fortunately, crystals of vastly improved diffraction quality were obtained by forming a complex of the protease with the inhibitor MAPI. To obtain these cocrystals, NaCl was added to the protein sample to a final concentration of 300 m*M*, which allowed the protease to be concentrated to 17 mg ml^{-1} using a 10 kDa cutoff Centricon concentration vessel. To provide suitable conditions to complex the protease with the essentially insoluble inhibitor, it was necessary to include 10% DMSO in the buffer. An amount of inhibitor giving a threefold molar excess over the protein was dissolved in a volume of DMSO that would, once added to the protein sample, result in the final buffer containing 10% DMSO. The inhibitor in DMSO was added to the protein sample in ten equal volumes at 10 min intervals. The sample was then passed through a Sephadex G25 Minispin column to rid the complex of any excess unbound inhibitor and DMSO. Incubation of a small volume of protease complex with the chromogenic substrate AcEFQLQ-pNA demonstrated that 100% inhibition had been achieved. Further confirmation and accurate assessment of MAPI binding was accomplished by mass spectrometry, which revealed a single major peak of molecular weight 20 045 corresponding to one molecule of protease covalently linked to one molecule of the inhibitor.

Further screening for crystallization conditions with the enzyme–inhibitor complex at a concentration of 15 mg ml^{-1} yielded large rhomboid crystals in 25%(*w*/*v*) PEG 5000 MME, 100 m*M* Tris–HCl pH 8.5, 200 m*M* lithium sulfate (Jena Bioscience JBScreen Classic 4 condition A1). The cocrystals were extremely clean in appearance and measured between 200 and 400 µm in all dimensions (Fig. 3). The cocrystals were cryoprotected in the same manner as the native enzyme. It was found that freezing the crystals within a week of their appearance was crucial to achieving high-quality data. Hence, this was usually performed between 18 and 21 d of setting up the crystallization experiments. If this time limit was exceeded, the crystals were seen to degrade in an unusual manner, apparently involving a phase change prior to collapse of the crystal (Fig. 3). The proneness of the enzyme to oxidation may be responsible for this deterioration over time.

Data collection at station ID14-1 (ESRF, Grenoble) established that the crystals diffracted to high resolution (1.7 Å) and gave sharp non-overlapping reflections (Fig. 4), which was a marked improvement on the native crystal data. Using 1° oscillations, 190° of data were collected from a single crystal using an ADSC Q210 CCD detector with an exposure time of 6 s per image and a crystal-to-detector distance of 160.1 mm. A low-resolution pass was performed using 1 s exposures, 3° oscillations and a crystal-to-detector distance of 391.4 mm to re-collect the spots that were overloaded in the initial high-resolution pass. Data processing in *MOSFLM* (Leslie, 2006), *SCALA* (Evans, 2006) and other programs in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) revealed that the crystals belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 49.5$, $b = 84.1$, $c = 121.5$ Å; the data set had an overall $R_{\rm merge}$ of 5.3% to 1.7 Å resolution (Table 2). Using the method of Matthews (1968) it was estimated that there were two monomers per asymmetric unit, which corresponds to a solvent content of 61%. Accordingly, a self-rotation function calculated at 2.0 Å resolution with a radius of integration of 25 Å using *MOLREP* (Vagin &

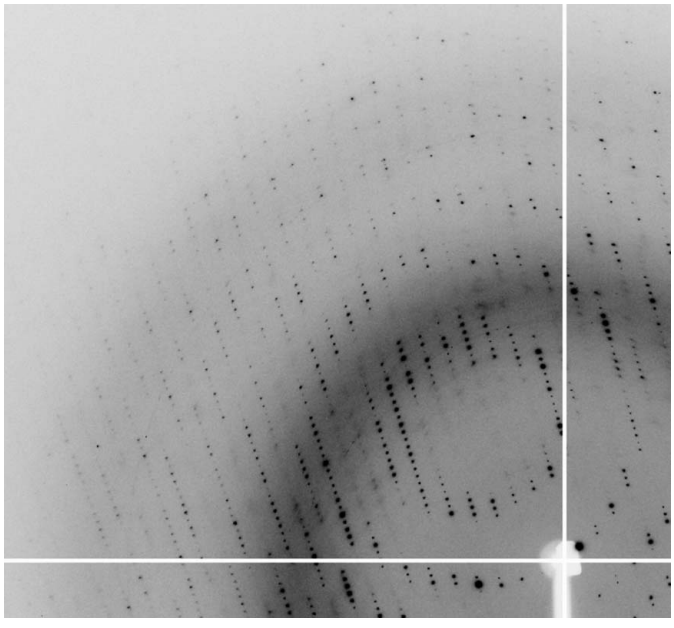


Figure 4
Part of a diffraction image obtained from a cocrystal on beamline ID14-1 at ESRF (Grenoble, France). Diffraction to a resolution of approximately 1.7 Å is visible.

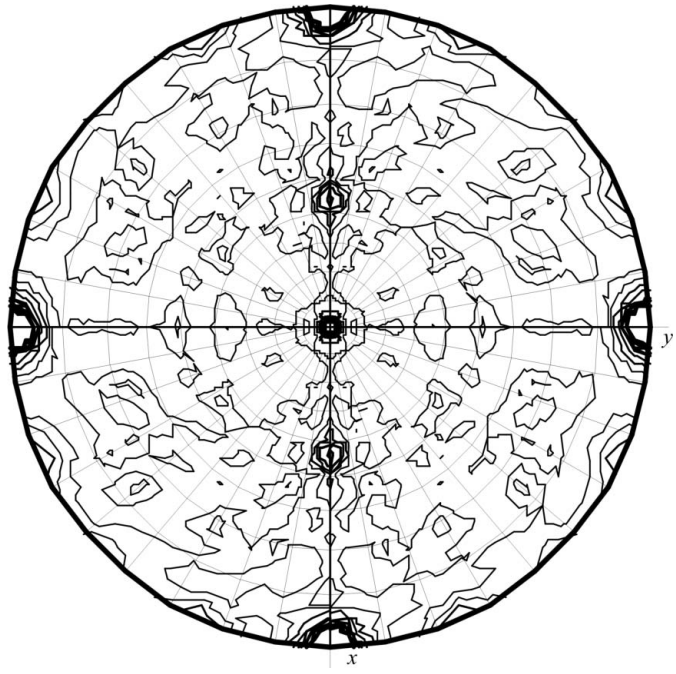


Figure 5
A self-rotation function for the protease–inhibitor complex calculated at 2 Å resolution. The $\chi = 180°$ section is shown.

Table 2
Data-collection and processing statistics for the protease–MAPI cocrystal.

Values in parentheses are for the outer resolution shell.

Beamline	ID14-1, ESRF
Wavelength (Å)	0.934
Space group	$P2_12_12_1$
Unit-cell parameters	
a (Å)	49.5
b (Å)	84.1
c (Å)	121.5
Mosaic spread (°)	0.7
Resolution (Å)	40.5–1.7 (1.8–1.7)
$R_{\rm merge}$† (%)	5.3 (62.3)
Completeness (%)	98.8 (92.8)
Average $I/\sigma(I)$	22.4 (2.4)
Multiplicity	6.7 (5.2)
No. of observed reflections	376358 (38859)
No. of unique reflections	55934 (7504)
Wilson plot *B* factor (Å^2)	24.7
Solvent content (%)	61.0
No. of molecules per asymmetric unit	2

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of the scaled observations $I_i(hkl)$.

Teplyakov, 2010) showed significant non-axial peaks in the $\chi = 180^\circ$ section (Fig. 5), suggesting the presence of noncrystallographic twofold symmetry. Structure determination of the selenomethionyl enzyme complexed with the inhibitor in the same crystal form is in progress.

We gratefully acknowledge the School of Biological Sciences, University of Southampton for a studentship award to RJH, Hope (Southampton General Hospital) for a grant to PMSJ and the ESRF (Grenoble, France) for beam time and travel support. The work was also part-supported by a Wellcome Trust grant (reference 086112) to INC and PRL.

References

Bertolotti-Ciarlet, A., Crawford, S. E., Hutson, A. M. & Estes, M. K. (2003). *J. Virol.* **77**, 11603–11615.
Clarke, I. N. & Lambden, P. R. (2005). *Virology*, edited by B. W. J. Mahy & V. ter Meulen, pp. 911–931. London: Hodder.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dragovich, P. S. *et al.* (1998*a*). *J. Med. Chem.* **41**, 2806–2818.
Dragovich, P. S. *et al.* (1998*b*). *J. Med. Chem.* **41**, 2819–2834.
Dragovich, P. S. *et al.* (2003). *J. Med. Chem.* **46**, 4572–4585.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Govardhan, C. P. & Abeles, R. H. (1996). *Arch. Biochem. Biophys.* **330**, 110–114.
Green, K. Y. (2007). *Fields Virology*, edited by D. M. Knipe & P. M. Howley, pp. 949–980. Philadelphia: Kluwer.
Kaspari, A., Schierhorn, A. & Schutkowski, M. (1996). *Int. J. Pept. Protein Res.* **48**, 486–494.
Kuyumcu-Martinez, M., Belliot, G., Sosnovtsev, S. V., Chang, K. O., Green, K. Y. & Lloyd, R. E. (2004). *J. Virol.* **78**, 8172–8182.
Lambden, P. R., Caul, E. O., Ashley, C. R. & Clarke, I. N. (1993). *Science*, **259**, 516–519.
Lambden, P. R., Liu, B. L. & Clarke, I. N. (1995). *Virus Genes*, **10**, 149–152.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Liu, B. L., Clarke, I. N. & Lambden, P. R. (1996). *J. Virol.* **70**, 2605–2610.
Liu, B. L., Viljoen, G. J., Clarke, I. N. & Lambden, P. R. (1999). *J. Gen. Virol.* **80**, 291–296.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. & Worland, S. (1994). *Cell*, **77**, 761–771.
Merrifield, R. B. (2007). *J. Am. Chem. Soc.* **85**, 2149–2154.
Nakamura, K., Someya, Y., Kumasaka, T., Ueno, G., Yamamoto, M., Sato, T., Takeda, N., Miyamura, T. & Tanaka, N. (2005). *J. Virol.* **79**, 13685–13693.
Rink, H. (1987). *Tetrahedron Lett.* **28**, 3787–3790.
Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Whitmore, A. J., Daniel, R. M. & Petach, H. H. (1995). *Tetrahedron Lett.* **36**, 475–476.
Zeitler, C. E., Estes, M. K. & Prasad, B. V. V. (2006). *J. Virol.* **80**, 5050–5058.