論文 / 著書情報
Article / Book Information

題目(和文)	ノロウイルスORF1 polyproteinの自己切断における構造生物学的研究
Title(English)	Structural studies in proteolytic processing of norovirus ORF1 polyprotein : 3C-like protease structure
著者(和文)	中村健太郎
Author(English)	Kentaro Nakamura
出典(和文)	学位:博士（理学）, 学位授与機関:東京工業大学, 報告番号:甲第6733号, 授与年月日:2007年3月26日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor of Science, Conferring organization: Tokyo Institute of Technology, Report number:甲第6733号, Conferred date:2007/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Structural studies in proteolytic processing of norovirus ORF1 polyprotein ~3C-like protease structure~

A Doctoral Thesis

by

Kentaro Nakamura

Graduate School of Bioscience and Biotechnology,

Tokyo Institute of Technology

2007

Contents

Abbreviations

3BC	connected 3B-3C region
3CD	connected 3C-3D region
$3C^{pro}$	3C/3C-like protease
$3D^{pol}$	3D RNA-dependent RNA polymerase
ChV	Chiba virus
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
eIF	eukaryotic translation initiation factor
His-tag	hexa-histidine tag
IPTG	β-D-1-thiogalactopyranoside
MPD	2-methyl-2,4-pentanediol
NV	Norovirus
NVP	Norovirus 3C-like protease
ORF	Open Reading Frame
PDB	Protein Data Bank
TCEP	Tris[2-carboxyethl]phosphine
VPg	genome-linked viral protein
XAFS	X-ray Absorption Fine Structure

Chapter 1: General introduction

1.1 Viral acute gastroenteritis in Japan

There are bacterial- and viral- acute gastroenteritis in humans. In bacterial acute gastroenteritis, the major causative agents are *Salmonella* and *Campyrobacter jejuni/coli*, while in viral are noroviruses. In Japan, 2005, total number of the food poisoning event was 1,547; the patients were 27,019 persons, the number of the events from bacterial agents is 1,065 (16,678 patients), from viral agents is 275 (8,728 patients) (the announcement of The Ministry of Health, Labour and Welfare). In details, in the ranking of the events, the first is *Campyrobacter jejuni/coli* (645 events), the second is norovirus (274 events), and the third is *Salmonella* (144 events). But in the ranking of the patients, the first is norovirus (8,728 patients), the second is *Salmonella* (3,700 patients), and the third is *Campyrobacter jejuni/coli* (3,439 patients). Therefore, in the number of patients, norovirus accounts for 32.2% of the whole. The food poisoning occurrence situation according to each annual was summarized at Table 1-I.

1.2 Norovirus

Noroviruses (NVs; formerly "Norwalk-like viruses"), which belong to the family *Caliciviridae*, are the causative agents of viral acute gastroenteritis that has occurred in various epidemiological settings, including nursing homes (Jiang et al., 1996; Vinje and Koopmans, 1996; Vinje et al., 1997), schools (Kobayashi et al., 1991), hospitals (Stevenson et al., 1994; Vinje et al., 1997), restaurants (Fleissner et al., 1989; Parashar et al., 1998), and cruise ships (Gunn et al., 1980; Ho et al., 1989; Herwaldt et al., 1994; Khan et al., 1994). The transmission was carried out by contaminated food, especially oysters (Dowell et al., 1995; Kohn

et al., 1995) and water (Beller et al., 1997; Gray et al., 1997), and by person-to-person contact (Kaplan et al., 1982). The discovery of norovirus goes back to the acute gastroenteritis which broke in an elementary school of Norwalk, Ohio of the United States in 1968. The virus was detected and discovered from patient feces, and called a Norwalk virus from the name of the town. Thereafter, similar viruses were found in all over the world. The viruses were called "small round-structured viruses (SRSVs) or Norwalk-like viruses," a name that describes the electron micrographic morphology of the fixed virion (Figure 1-1). In August, 2002, they were formally named "Norovirus" in the international virology association.

NVs are serologically and genetically diverse—more than a thousand strains have been isolated worldwide (Vinjé, 2000). However, only a limited number of NV genomes have been sequenced; Southampton virus (SAV) (Lambden et al., 1993; L07418), BS5 strain (unpublished, AF093797), Lordsdale virus (LV) (Dingle et al., 1995; X86557), Camberwell virus (CWV) (Seah et al., 1999; AF145896), and Hawaii virus (HWV) (U07611)) Chiba virus (Someya et al., 2000; AB042808), because appropriate cell culture systems and/or animal models supporting virus propagation have not been developed. But a number of partial genomic sequences derived from the RNA-dependent RNA polymerase and capsid protein region of NV have been cloned, sequenced, and deposited in databases. Members of NV share common features in that their genomes are a positive-sense, single-stranded RNA molecule to which a putative genome-encoded protein called VPg is linked at the 5' end of the genome (Estes et al., 1997; Clarke et al., 1998). Extensive genetic analyses of NV demonstrated

that NV is further classified into two subgroups, genogroup I (GI), which includes NV/68, SAV, and BS5, and genogroup II (GII), which includes LV, CWV, and HWV (Wang et al., 1994; Green et al., 1994; Lew et al., 1994; Liu et al., 1995). The phylogenetic tree of published NV sequences based on part of the RNA-dependent RNA polymerase gene demonstrated that the GI and GII subgroups are composed of at least seven and five genetically distinct viruses, respectively (Green et al., 2000; Vinje et al., 2000). Recently, the Jena strain of bovine calicivirus was cloned, and the nucleotide sequence indicated that it is closely related to GI NV (Liu et al., 1999a; AJ011099).

1.3 Norovirus ORFs

The genome of NV contains three open reading frames (ORFs) (Figure 1-2). ORF1 encodes a large polyprotein containing amino acid sequence motifs observed in many RNA viruses, such as VPg, 3C-like protease and RNA-dependent RNA polymerase. Protease activity is required for the maturation of functional proteins encoded by ORF1. ORF2 is predicted to encode the capsid protein. ORF3 encodes a small protein abundant in basic amino acids. Although the precise role of the ORF3 protein in the virus replication is unclear, it is likely that ORF3 protein is a minor structural protein that interacts with the genome RNA during the virion formation (Jiang et al., 1993; Lambden et al., 1993; Dingle et al., 1995; Seah et al., 1999; Glass et al., 2000).

1.4 Norovirus ORF1 polyprotein

As above mentioned, norovirus ORF1 encodes a large polyprotein, which is

probably processed intracellularly into six proteins by the viral 3C-like protease (Figure 1-3). The six NV ORF1 nonstructural proteins are homologous to picornaviral nonstructural proteins and are named accordingly: N-terminal protein, 2C-like nucleoside triphosphatase, 3A-like protein, 3B VPg (genome-linked viral protein), 3C-like protease ($3C^{pro}$), and 3D RNA-dependent RNA polymerase ($3D^{pol}$). NV 3B VPg binds to both of the 5' end of the genome and eukaryotic initiation factors such as eIF3 (Daughenbaugh et al., 2003). The protein may function like the 5'-cap structure of eukaryotic mRNAs involved in translation initiation. NV $3D^{pol}$ is an RNA-dependent RNA polymerase which is unique in that its polymerase activity is not poly(A) tail/primer dependent (Fukushi et al., 2004). 3C-like proteases are the key enzymes for ORF1 polyprotein processing (Liu et al., 1996; Seah et al., 1999; Someya et al., 2000; Belliot et al., 2003) and also cleave the poly(A)-binding protein, causing cellular translation inhibition (Kuyumcu-Martinez et al., 2004).

1.5 Norovirus 3C-like protease

NV 3C-like proteases (NVP) belong to the chymotrypsin-like protease family, in that they appear to have chymotrypsin-like folds. Picornaviral 3C proteases are also members of the chymotrypsin-like family (Dougherty et al., 1999). Probably, most of these chymotrypsin-like viral proteases have a cysteine, rather than a serine, as the active site nucleophile. By individual replacement of the charged residues and the putative active site cysteine of the $3C^{pro}$ from the norovirus Chiba virus (genogroup I strain) with alanines, His30 and Cys139 were identified as a catalytic dyad, which functions without active participation of an active site

carboxyl moiety (Someya et al., 2002). An active site acidic residue does not seem essential for activity of this $3C^{pro}$, although a Glu54-to-Gly mutation abolishes protease activity (Hardy et al., 2002). The results of the mutagenesis study identified five additional residues (Arg8, Lys88, Arg89, Asp138, and His157) as indispensable for protease activity, but their roles were not ascertained in details (Someya et al., 2002). Characterization of the protease tertiary structure should clarify the functional roles of the residues.

1.6 Objectives of this study

Now, three subjects arose: i) the elucidation of the details of ORF1 polyprotein processing mechanism by NVP, ii) the meaning of the processing in the functions of the products from the polyprotein, iii) the accompanying structural difference/identity between precursors and successors. In order to investigate them, I carried out the structural analysis of NVP with and without a substrate, the crystallization of precursors (3BC, 3CD), and the modelling study of 3CD precursor. I expect that drug development for NV-associated gastroenteritis and other diseases caused by the viruses which have the proteases of this group will be facilitated by a global study of the available protease crystal structures.

In this study, Chiba virus (ChV), detected in an oyster-associated outbreak that occurred in December 1987 in Chiba Prefecture Japan, was used as NV.

Table 1-I. The generation status of food poisoning.

		Number of Events		Number of Patients	
2001	Total	1,924		25,732	
	NV	268	(13.9%)	7,335	(28.5%)
	Salmonella	360	(18.7%)	4,912	(19.1%)
	Campyrobacter jejuni/coli	428	(22.2%)	1,880	(7.3%)
2002	Total	1,850		27,629	
	NV	268	(14.5%)	7,961	(28.8%)
	Salmonella	465	(25.1%)	5,833	(21.1%)
	Campyrobacter jejuni/coli	447	(24.2%)	2,152	(7.8%)
2003	Total	1,585		29,355	
	NV	278	(17.5%)	10,603	(36.1%)
	Salmonella	350	(22.1%)	6,517	(22.2%)
	Campyrobacter jejuni/coli	491	(31.0%)	2,642	(9.0%)
2004	Total	1,666		28,175	
	NV	277	(16.6%)	12,537	(44.5%)
	Salmonella	225	(13.5%)	3,788	(13.4%)
	Campyrobacter jejuni/coli	558	(33.5%)	2,485	(8.8%)
2005	Total	1,545		27,019	
	NV	274	(17.7%)	8,727	(32.3%)
	Salmonella	144	(9.3%)	3,700	(13.7%)
	Campyrobacter jejuni/coli	645	(41.7%)	3,439	(12.7%)

* (%) shows the proportion occupied in the total.

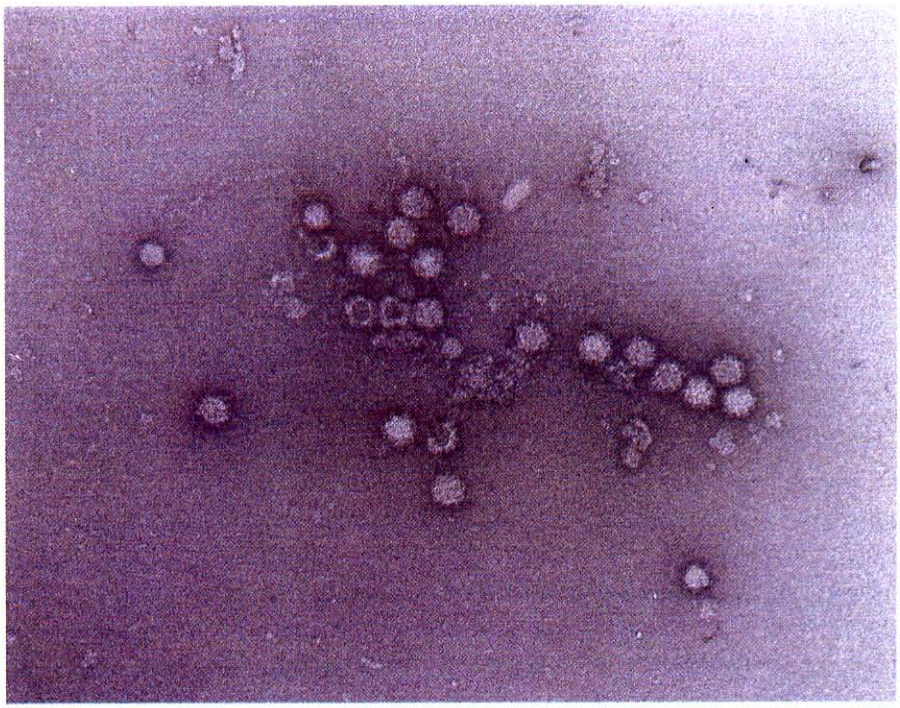


Figure 1-1. The image of NV (ChV) by the electron microscope. Patient's feces were divided by cesium chloride equilibrium density gradient centrifugation, then the viruses were observed by the electron microscope. The diameter of the virion is 38nm, plus single-strand RNA (~7.7 kb) is packed in the inside. The white and black virions are perfect virions and empty which do not have RNA, respectively. (Genetic and antigenic diversities of human caliciviruses. Someya Y, Natori K, Tkeda N, Miyamura T, Jpn. Clinical Biology, vol. 27, No. 4, 1999)

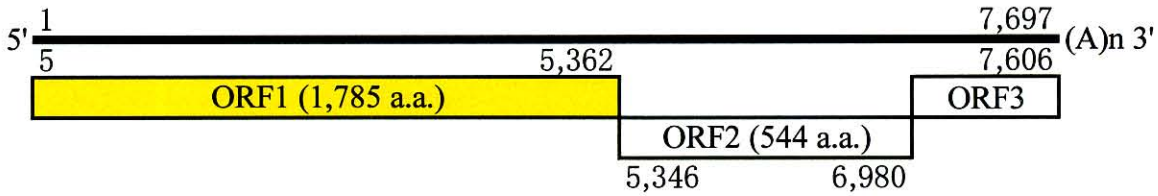


Figure 1-2. ORF1, ORF2, and ORF3 of ChV. ChV has three open reading frames, ORF1 (1,785 a.a.), ORF2 (544 a.a.) and ORF3 (208 a.a.). The base numbers are described in the figure.

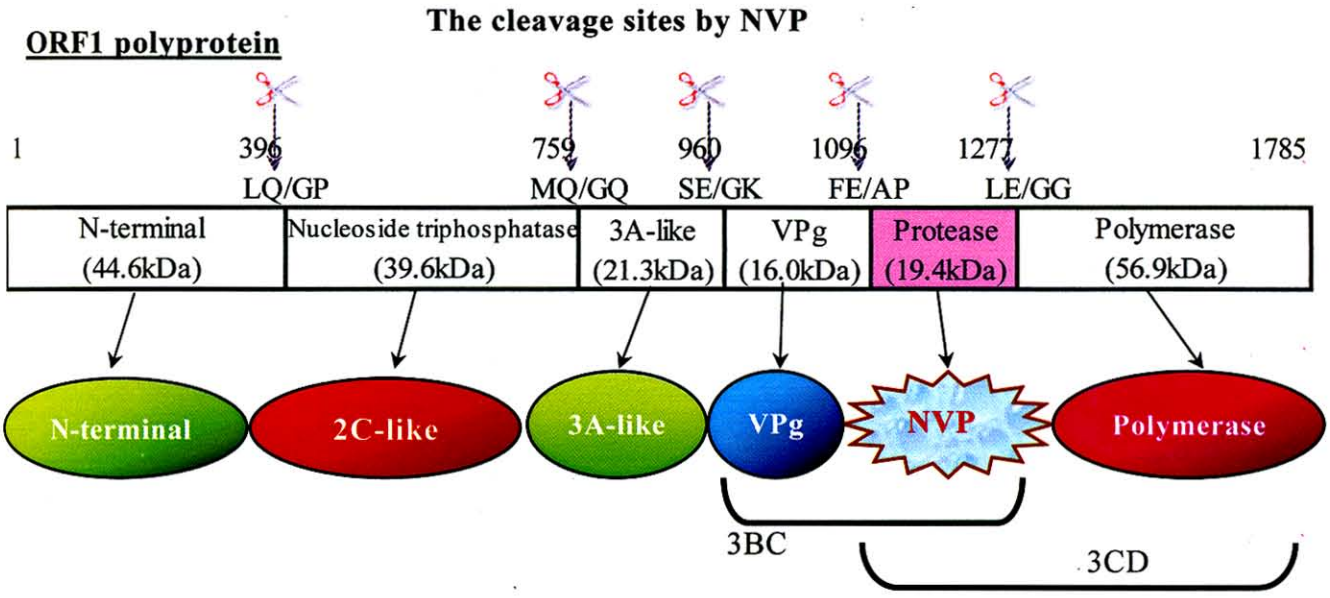


Figure 1-3. The processing of ORF1 polyprotein by NVP. The residue numbers are shown in figure. And the cleavage sites are described by the numbers and the residue types (396: LQ/GP, 759: MQ/GQ, 960: SE/GK, 1096: FE/AP, 1277: LE/GG. The slash shows the scissile site.).

Chapter 2: Structural analysis of NVP in native

2.1 Protein preparation and crystallization of His-tagged NVP labeled by Se-Met

In order to investigate the reaction mechanism and the details of norovirus 3C-like protease (NVP) as the key enzyme in proteolytic processing of norovirus ORF1 polyprotein, we carried out the expression, purification and crystallization of NVP. The Se-Met substitute of His-tagged NVP (His-NVP(Se-Met)) was prepared to utilize the anomalous scattering effect of Se atoms in the structure determination by X-ray crystallography. Inhibition of the biosynthetic pathway of the methionine in the over-expression system using *Escherichia coli*, and the addition of Se-Met into the broth induce the Se-Met substitution of NVP.

2.1.1 Materials & methods

2.1.1.2 Expression and purification of His-NVP(Se-Met)

The expression vector (pT7-HisPro: Someya et al., 2000) was introduced into *E. coli* strain BL21(DE3), and the protein was expressed as a His-tagged protein at the N-terminus. As pre-preculture, BL21(DE3)/pT7-HisPro was cultivated at 37°C in 6 ml LB (Luria-Bertani) broth supplemented with 100 μg/ml ampicillin. The preculture, culture and expression of the protein was carried out at 30°C in M9 minimal medium supplemented with 100 μg/ml ampicillin using the pre-preculture solution as the bacterial cells. When cultures were grown to an OD_{600} of 0.5~0.6 at 30°C, the inhibition of the biosynthetic pathway of the

methionine in *E. coli* and the Se-Met substitution of the proteins were induced by the addition of the amino acids (the inhibitor: 100 mg/L Lys, 100 mg/L Phe, 100 mg/L Thr, 50 mg/L Ile, and 50 mg/L Val) and 50 mg/L Se-Met as final concentration, respectively. Then the expression of His-tagged NVP was induced by the addition of IPTG (isopropyl-D-thiogalactopyranoside) to a final concentration of 0.5 mM at 30°C for overnight. The harvested cell was lysed by sonication in buffer A (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 20 mM 2-ME (2-mercaptoethanol), and 10 mM imidazole). The resulting cell homogenate was centrifuged at 44,200 x g for 30 min. After loading the supernatant on a Ni-NTA Super Flow column, it was washed with buffer A and then eluted with buffer B (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 20 mM 2-ME, 100 mM imidazole). The fractions containing His-NVP(Se-Met) were concentrated and further purified with the gel filtration column Superdex75 (Amercham Biosciences) in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM DTT (1,4-dithiothreitol). The purity of His-NVP(Se-Met) was confirmed by SDS-PAGE analysis (Figure 2-1). The fractions containing His-NVP(Se-Met) were combined and concentrated to 10 mg/ml.

2.1.1.2 Crystallization and data collection

Crystallization trials were performed at 25°C using the hanging-drop or sitting-drop vapor-diffusion method. The crystallization conditions were determined by initial screenings in 96-well plates for the sitting drop vapor diffusion using Crystal Screen I & II (Hampton Research). Each drop was

prepared by mixing 1.5 μl of protein solution, described above, with 1.5 μl precipitant solution, and the reservoir was 100 μl precipitant solution. As the result, His-NVP(Se-Met) was crystallized at 25°C when using the crystallization solution (20 mM Tris-HCl, pH 8.0, 1.2 M Ammonium Sulfate, 0.5 M NaCl) as the precipitant solution. The optimization of the condition gave the best crystallization solution, 0.9 M AS, 0.4 M NaCl, 5% MPD, 0.1 M Tris-HCl, pH 8.0 by the hanging-drop vapor diffusion method. The crystals were grown to the approximate size of 0.3 x 0.1 x 0.1mm in a week (see Figure 2-2).

2.1.1.3 XAFS and diffraction data collection

XAFS data of the crystals were collected around the *K*-absorption edge of Se atoms in the 8 GeV synchrotron radiation facility SPring-8 (BL26B1), so that the existence of Se in the crystal confirmed (Figure 2-3) and the wavelengths, edge, peak, high energy remote, and low energy remote were determined as 0.97954 Å, 0.97934 Å, 0.96927 Å, and 0.98294 Å, respectively. The diffraction data set was collected at three wavelengths (edge, peak, high remote) at -170°C using the flash-frozen method with the cryoprotectant solution (30 % glycerol, 1.0 M AS, 0.4 M NaCl, 5 % MPD, 0.1 M Tris-HCl, pH 8.0).

The diffraction images were processed by D*TREK in the program CrystalClear (RIGAKU/MSC). The three data (edge, peak, high remote) were merged and scaled by the programs in CCP4 (see Table 2-I) (CCP4. 1994). The crystal belong to the hexagonal space group $P6_1$ with the unit cell dimensions a = b = 224.8 and c = 118.7 Å. The search of Se atom positions and the initial

phasing were performed by the multiwavelength anomalous dispersion method (MAD) using program SOLVE (Terwilliger et al., 1994 and 1999).

2.1.2 Results & Discussion

The Se-Met substitute of His-tagged NVP was successfully obtained without using Met auxotroph strain, and the existence of Se atoms of NVP in the crystal was confirmed by XAFS measurement. However, the position of Se atoms was not determined and the phasing by MAD was failed. This might be caused by the fact that the lattice constants was enlarged, and the estimated number of NVP molecule in the asymmetric unit was eighteen, and the number of Se atom in the asymmetric unit exceeded ninety. In addition, the resolution was low, 3.6 Å.

In comparison with the crystal of native NVP, described in Chapter 2.2~2.3, it was shown that the lattice parameter a and b of His-NVP(Se-Met) crystal was respectively about $\sqrt{3}$ times longer than ones of the native crystal. This suggested the possibility that the large lattice, $a = b = 224.8$ Å, would be taken instead of the lattice, $a = b = 129$ Å, by the delicate change/difference in the intermolecular interactions which might be generated by additional Se atoms and/or His-tag, though the molecule ranged in the almost same arrangement in crystals (Figure 2-4 and Figure 2-5B). The arrangement of structures in native crystal supported this suggestion. Actually, in native crystal structure three methionines (Met107, Met120 and Met130) were on the molecular surface and involved in the intermolecular interactions (Figure 2-5).

2.2 Material and methods in native NVP

Expression, purification and crystallization of non-tagged NVP were performed to obtain higher resolution crystals than His-NVP(Se-Met) and determine the crystal structure of NVP. The structure determination was tried by MAD method using soaked/co-crystallized mercury atoms.

2.2.1 Cloning, expression and purification of NVP

Recombinant intein-fusion NVP was expressed in *Escherichia coli* BL21(DE3)-CodonPlus-RIL using the plasmid pTYB11[NVP], which contains a Chiba virus protease gene insert (NCBI/Gen Bank accession number of Chiba virus, AB042808). The gene had been amplified from pUC[HisPro] (Someya et al., 2000) and inserted into pTYB11 of the IMPACT system (New England BioLabs) as the N-terminal segment fused to the intein and chitin-binding domain unit. When cultures were grown to an OD600 of 0.6-0.7 at LB broth supplemented with 100 μg/ml ampicillin, the expression of recombinant intein-fusion NVP was induced by the addition of IPTG to a final concentration of 0.1 mM at 17°C overnight. The harvested *E. coli* was lysed by treatments with sonication in a lysis buffer containing 20 mM HEPES-NaOH, pH 8.2, 0.5 M NaCl, 1 mM EDTA, 1 mM TCEP. The cell homogenate was centrifuged at 20,000 x g for 30 min. The supernatant was applied to a chitin bead column (New England BioLabs) equilibrated with lysis buffer and was washed with a wash

buffer, 20 mM HEPES-NaOH, pH 8.2, 1.0 M NaCl, 1 mM EDTA, 1 mM TCEP. On-column self-cleavage was carried out for 40 hours using a cleavage buffer of 20 mM HEPES-NaOH, pH 8.2, 0.5 M NaCl, 1 mM EDTA and 50 mM DTT. The fraction containing NVP was collected and concentrated. Further purification was performed by gel filtration over a HiLoad 16/60 Superdex (75 pg; Amersham Pharmacia Biotech) column equilibrated with 20 mM Tris-HCl, pH 8.2, 200 mM NaCl, 5 mM DTT. The fractions containing NVP were combined and concentrated to 20 mg/ml. Enzymatic activity was spectrophotometrically assayed using the commercially available substrate H-Glu-Ala-Leu-Phe-Gln-pNA (Bachem). The purity of NVP was confirmed by SDS-PAGE analysis (Figure 2-6).

2.2.2 Crystallization, derivatization, data collection and processing.

The crystallization conditions were determined by a sparse-matrix screen using Crystal Screen I and II (Hampton Research). Although crystals were obtained under various conditions, the best crystals were grown at 20°C using a hanging drop vapor diffusion system consisting of a 3 µl protein solution and 3 µl crystallization buffer drop, equilibrated against crystallization buffer (1 ml). The buffer that produced the best crystals contained 20 mM HEPES, pH 7.5, 0.9 M sodium potassium tartrate ($KNaC_4H_4O_6$). Crystals grew within 2 to 3 weeks and typically had dimensions of approximately 0.3 x 0.3 x 0.4 mm^3 (Figure 2-7).

For multiwavelengh anomalous dispersion (MAD) data, mercury derivatives were prepared by soaking native crystals in 1mM mercury chloride

crystallization buffer for 24 h at 20°C or by cocrystallization of the protein with at mercury derivative in a solution similar to that used to prepare the underivatized crystal.

X-ray diffraction data were collected at SPring-8 (Hyogo, Japan). All data were collected using crystals flash cooled at -170°C, which were prepared by rapidly transferring the crystal into a cryoprotectant containing up to 30 % (vol/vol) glycerol. MAD data were collected at four wavelengths at beamline BL26B2, which was equipped with a Jupiter charge-coupled device detector (Rigaku MSC). MAD data were collected for the f'' maximum, for the f' minimum, and for remote wavelengths below and above the Hg L_{III} edge. All data were processed using the program CrystalClear (Rigaku MSC). The data collection statistics are given in Table 3-I. The crystals belong to the hexagonal space group $P6_1$ with unit cell dimensions $a = b = 128.89$ and $c = 118.08$ Å and the diffract X-ray beyond 2.8 Å resolution. Subsequent calculations were performed using the CCP4 suite of programs (CCP4. 1994).

2.2.3 Structure determination

The Hg MAD data were treated as four derivatives, each with an anomalous scattering contribution. All data sets were scaled to a common level using the CCP4 program SCALEIT. Six of the eight mercury atom positions and the initial phases were identified using SOLVE (Terwilliger et al., 1994 and 1999). Next, the minor sites were identified using the initial phases by anomalous D-fourier synthesis. Then, the phases were calculated using the refined positions of the eight Hg atoms, the program MLPHARE and SHARP (Fortelle et al., 1997).

Additionally, density modification was carried out in several steps using the CCP4 program DM. After the non-crystallographic symmetries operator was determined from the eight Hg positions, solvent flattening, histogram matching, and three twofold averaging were performed by the program DM (operation-1). To avoid the overestimation of the figure of merit, the figure of merit was forced to decrease 0.4-fold manually (operation-2), and then operation-1 was carried out again. These operations, operation-1 and -2, were repeated three times. The electron density map showed almost the entire polypeptide chain for each of the four molecules of the asymmetric unit. The initial model was built using XtalView (McRee 1999). The model was refined at 2.8 Å resolution using Refmac5 and CNS (Murshudov 1997, Brűnger et al., 1998). The crystallographic R factor converged to a value of 0.208, with an associated R_{free} of 0.240. Data quality and refinement statistics are given in Table 2-II and 2-III, respectively. The quality of the structural model and its agreement with the structure factors were checked using the programs PROCHECK, PROMOTIF, and SFCHECK.

2.2.4 Modelling study with substrate

An oligopeptide, corresponding to the third domain of the turkey ovomucoid inhibitor (OMKTY3) residues 14 to 21, which are the P5-P3' residues, was docked onto the NVP substrate binding site. Docking was performed by the CCP4 program lsqkab so that alignment between the oligopeptide and NVP residues of the active site triad, the oxyanion hole, and the β-strand NVP residues 158 to 161 mimicked that of the oligopeptide and corresponding *Streptomyces griseus* protease (SGPB) residues. Then, the OMKTY3 side chains were replaced

with the corresponding side chains of the peptide Glu-Thr-Thr-Leu-Glu-Gly-Gly-Asp, which is the P5-P3' cleavage sequence for 3CD. Finally, the substrate conformation was first manually adjusted and then energy minimized to remove atomic overlaps while retaining the antiparallel β-strand interaction.

Figures were made using PyMOL (DeLano 2002), DINO (http://www.dino3d.org), MSMS, MEAD, and Seqseq (Bashford et al., 1992, Duncan et al., 1999).

Protein structure accession number. The coodinates and structure factors were deposited in the Protein Data Bank (entry 1WQS).

2.3 Results & discussion

2.3.1 Global and domain structures

There are four NVP monomers (designated A, B, C and D) per asymmetric unit. Each monomer consists of an N-terminal domain and a C-terminal domain (Figure 2-8A). The secondary structure composition is primarily and extensively β-strand. A secondary structure topological representation and an assignment are shown in Figure 2-8B.

The N-terminal domain has two α-helices and seven β-strands (aI, bI, cI, dI, eI, fI, and gI). The β-strands form a twisted antiparallel β-sheet resembling an incomplete β-barrel. N-terminal domain antiparallel β-barrels exist in the viral chymotrypsin-like cysteine proteases—HAV $3C^{pro}$ (Allaire et al., 1994;

Bergmann et al., 1997), foot-and-mouth disease virus $3C^{pro}$ (Birtley et al., 2005), tobacco etch virus protease (Phan et al., 2002), HRV $3C^{pro}$ (Cheah et al., 1990; Matthews et al., 1994) and PV $3C^{pro}$ (Ivanoff et al., 1986; Hämmerle et al., 1991; Lawson and Semler 1991; Blair et al., 1996; Mosimann et al., 1997), in the coronavirus 3C-like proteases (also known as main proteases, M^{pro}; Anand et al., 2002; Anand et al., 2003; Yang et al., 2003). However, the NVP N-terminal domain incomplete β-barrel is intermediate in structure between the N-terminal domain four β-strands of HRV $2A^{pro}$ (Petersen et al., 1999) and the corresponding β-barrels of other chymotrypsin-like proteases. The core of the NVP incomplete β-barrel contains the hydrophobic residues, Phe12, Phe25, Phe39, Phe40, Phe58 and Phe60, Trp19, Ile32, Ile44, Ile47, and Ile49 (Figure 2-9). The active site residue His30 is found in the N-terminal domain.

The C-terminal domain is a six-stranded antiparallel β-barrel, formed by strands aII, bII, cII, dII, eII, and fII. The active site Cys139 is found in the C-terminal domain. The catalytic site formed is situated deep within a cleft between the N- and C-terminal domains.

When the four monomers of the asymmetric unit are examined, the following structural features are found noteworthy. The root mean square deviation (r.m.s.d.) for the positions of seventy-one core C_α atoms, with four equivalent atoms per asymmetric unit, is 0.33 (±0.02) Å. The r.m.s.d.s given above and below—with the associated uncertainties reported as standard deviations—are the averages of the r.m.s.d.s for the coordinates of included C_α positions. Whereas, the r.m.s.d. for 162 equivalent C_α atoms is 0.64 (±0.10) Å (cut-off 2.0 Å). Excluding residues sequentially adjacent to the N- and C-termini, the largest

conformational variations are found for: (i) a flexible surface loop (residues 33 to 36; r.m.s.d. 1.68 ± 1.09 Å) following dI strand; (ii) a hairpin loop (residues 107-113; r.m.s.d. 1.33 ± 0.62 Å); (iii) a long loop (residues 124-132; r.m.s.d. 2.10 ± 0.98 Å); (iv) a loop (residues 147-150; r.m.s.d. 1.17 ± 0.55 Å); and (v) a β-hairpin loop (residues 162-164; r.m.s.d. 1.42 ± 0.89 Å) (Figure 2-10). These loops are all solvent accessible and probably flexible. The C-terminal residues, 174-181, could not be traced in the electron density map for any of the four crystallographic independent molecules, perhaps because of disorder.

2.3.2 Active site

For most chymotrypsin-like proteases, a catalytic triad, composed of Ser or Cys, His, and Asp or Glu, is responsible for proteolysis (Dougherty and Semler 1999). However, no obligatory acidic residue, other than Asp138, that affected NVP activity was identified by alanine scanning (Someya et al., 2002). Therefore, the general base catalyst His30 and the nucleophilic Cys139 were proposed as a catalytic dyad. Asp138 could not be a third active-site residue, because, sequentially, it is next to Cys139. Its side chain cannot be adjacent to that of His30. Examination of the NVP structure confirms that His30 and Cys139 are catalytic residues and shows that Glu54 is in the position of the third active site residue (Figure 2-8C, and the stereo view around Cys139 is shown in Figure 2-11). The active site is located at the center of a deep cleft between the N- and C-terminal domains, as are the active sites of other chymotrypsin-like proteases. His30 is part of a short α-helix that follows strand cI and Glu54 is part of a loop connecting strands fI and gI. Cys139 is part of a hairpin turn between an α-helix

(Pro136-Asp138) and strand dII.

His30, Glu54, and Cys139 are conserved in all NV 3C-like proteases (Figure 2-12). Mutagenesis of His30 to other residues always caused the enzyme to lose activity, indicating that His30 is indispensable (Someya et al., 2002). A mutant Ser139 NVP retains activity, whereas introduction of a threonine, tyrosine, or methionine at position 139 abolishes activity (Someya et al., 2002). Cys139 and His30 seem to work the nucleophilic residue and the general acid-base catalyst residue, respectively. On the other hand, the replacement of Glu54 by an alanine does not significantly affect protease activity. Only two residues, Cys139 and His30, are essential for NVP activity, and although is not essential, Glu54 seems to fulfill the additional role in the activity. The situation for trypsin is different. In that case, if the active site aspartic acid is replaced by an asparagine, activity is severely diminished (Sprang et. al., 1987). The pK_a of a cysteine thiol is usually about 8.3, whereas that of a serine hydroxyl is about 13. Therefore, a cysteine thiol will ionize more readily than will a serine hydroxyl. This difference in ionization strengths probably accounts for the difference in the requirement for a catalytic carboxyl moiety. In addition, it suggests the possibility that the His30 imidazolium might exist and the chemical state of the Cys139/His30 interaction might be a thiolate–imidazolium ion pair, which is also found in papain-like proteases (Polgár 1974). For HAV $3C^{pro}$, although the position of Asp84 is spatially equivalent to the third triad member, its side chain points away from the general base His44 (Bergmann et al., 1997). Additionally, for coronavirus M^{pro}, the residue equivalent to Glu54 is Val84. Its aliphatic side chain obviously cannot participate in catalysis (Anand et al., 2002), and a water

molecule occupies the position of the third catalytic residue of the conventional triad. A catalytic dyad must be operational in M^{pro}. Therefore, in general, a carboxyl moiety may not be necessary for catalysis by viral chymotrypsin-like cysteine proteases. For some proteases, an asparagine residue occupies the position corresponding to the active site carboxyl moiety of chymotrypsin-like proteases and can only hydrogen bond with the active site histidine. Sárkány et al. (2000) suggested that a catalytically competent thiolate–imidazolium ion pair, not an imidazole general base catalyst, functions in the HRV $2A^{pro}$ catalytic mechanism. While, in the C139S mutant, His30 is likely to work as a general base, rather than an imidazolium ion, it is still uncertain whether His30 works as a general base or an imidazolium ion in native NVP, which has Cys139 (Someya et al., 2002). Kinetic and mutagenesis studies are needed to confirm that a third member is not necessary for the activity, and the possibility of a NVP thiolate–imidazolium ion pair.

The average distance between the sulfur of Cys139 and the $N^{\varepsilon 2}$ of His30 is 3.4 Å in molecules B and C, which is that expected theoretically (Hamilton and Ibers 1968) and is similar to the distances measured for HRV $3C^{pro}$ (3.5 Å; Matthews et al., 1994) and for PV $3C^{pro}$ (3.4 Å; Mosimann et al., 1997). The Cys sulfur/$HisN^{\varepsilon 2}$ distance in molecule A (3.2 Å) is a little shorter than the distances found for the other molecules of the asymmetric unit. Tartrate was included in the crystallization buffer and electron density corresponding to tartrate is found between Cys139 and His30 of molecule A (Figure 2-13). It is possible that an interaction between molecule A and a tartrate mimics the substrate binding state. For molecule D, the Cys sulfur/$HisN^{\varepsilon 2}$ distance (approximately 4 Å) is longer

than average. The structures of coronavirus M^{pro} and a part of HAV $3C^{pro}$ are inactivated and the distance separation between their catalytic Cys sulfur/$HisN^{\varepsilon 2}$ increased when their active site sulfurs are oxidized to either sulfinic acid or sulfonic acid during crystallization.

As noted above, although Glu54 is not essential for NVP activity, it is plausible that Glu54 decreases the range of motion for His30 if a negatively charged carboxylate and the positively charged imidazolium interact. The average distance between the Glu54 carboxyl oxygen that is closer to the $His30N^{\delta 1}$ of molecule A, B, C, D are 3.4 Å, 2.8 Å, 2.8 Å, 4.1 Å, many of the values are within hydrogen bonding distance. Therefore, as indicated in Someya et. al., (2002), NVP with Glu54 could be more active than without Glu54.

Examination of the NVP crystal structure shows that several hydrogen bonds help maintain the integrity of the active site and/or substrate binding site. These hydrogen bonds include one between the side chain nitrogen of Lys88 and the backbone carbonyl oxygen of Val9 and a second one between the guanidinium nitrogen of Arg8 and the backbone carbonyl oxygen of Thr 69 (Figure 2-8D). These hydrogen bonds involve residues bridging the N- and C-terminal domains. Identification of these hydrogen bonds clarifies why Arg8 and Lys88 are conserved in NV $3C^{pro}$s and why when mutated to other residues in NVP, activity is lost (Someya et al., 2002). An interaction between the Asp90 side chain oxygens and an Arg11 guanidinium nitrogen helps orient the N- and C-terminal domains. Mutants with either the first five residues or the first eight residues deleted retain a low level of activity; whereas as a mutant missing the first eleven residues is inactive (Someya et al., 2002). Therefore, the hydrogen bond made by

the main chain oxygen of Val9 and the side chain amino group of Lys88 appears vital for activity. The integrity of the active site and/or of substrate binding site and indeed the entire molecule may depend heavily on the hydrogen bonds described above.

2.3.3 Oxyanion hole

The amino acid sequence, Gly-Xaa-Ser/Cys-Gly, which includes the active-site nucleophile (underlined), is highly conserved in chymotrypsin-like proteases (Bazan and Fletterick 1998). Aspartic acid occupies the Xaa position, in most eukaryotic proteases, such as chymotrypsin and trypsin, in bacterial serine proteases, and in viral 2A cysteine proteases. However, for viral 3C or 3C-like cysteine proteases, the residue occupying the Xaa position varies, with Asp, Gln, Met, Tyr, or Trp often present. For NVP, the Gly-Asp-Cys-Gly motif forms the pocket of the 'oxyanion hole'. The oxyanion hole helps to bind the substrate tightly and to stabilize the tetrahedral transition state by hydrogen bonding with the negatively charged P1-backbone carbonyl oxygen.

The NVP oxyanion hole is formed by a region preceding Cys139 and two backbone amides (those of Gly137 and Cys139), which point towards the P1-carbonyl oxygen. Additionally, a side chain Asp138 oxygen, clearly forms a good hydrogen bond with the Arg89 side chain nitrogen (Figure 2-8E) as reflected by the well-ordered corresponding regions of the electron density map. These two residues are conserved in all noroviruses, and no other residues tested could replace either Arg89 or Asp138 (Someya et al., 2002). Neither an R89K mutant nor a D138N mutant was active, although the R89K mutant had a positive

charge at position 89 and the D138N should have been able to form a hydrogen bond between residues 89 and 138. Clearly, Arg89 and Asp138 are essential for activity and function by stabilizing the oxyanion hole.

In molecule A, the oxyanion hole is large enough to accommodate the main chain carbonyl oxygen of the P1 residue. However, the oxyanion holes of molecules B, C and D are much narrower than that of molecule A. The volume differences are attributed to the conformations of Pro136 and Gly137 (Figure 2-8F). The Pro136 carbonyl oxygen is displaced to the exterior of the molecule A oxyanion hole; whereas, for the other molecules, the oxygen faces inward. In the electron density maps, the different orientations could be clearly observed. Recall that a tartrate was found only in the active site of molecule A and mimics a substrate binding state. Therefore, in the absence of substrate, the oxyanion hole is stabilized by an electrostatic interaction between the carbonyl oxygen of Pro136 and the numerous peptide amides that line the hole, plus the aforementioned interaction involving Asp138/Arg89. Perhaps, substrate binding reduce the structural flexibility of the main chain of Pro136 and Gly137, so that the carbonyl oxygen of Pro136 rotates to the exterior of the oxyanion hole by almost 180 degrees, while the Gly137 backbone amide turns inward. The effect upon substrate binding in the oxyanion hole structure, as discussed in HRV $3C^{pro}$, has been deduced (Matthews et al., 1994).

2.3.4 Substrate binding site

We built a model of an oligopeptide bound to NVP. This peptide, Glu-Thr-Thr-Leu-Glu*Gly-Gly-Asp, corresponds to P5–P3' of a substrate and

the asterisk marks the scissile bond. The notation, Pn–P1–P1'–Pn', for the residues of substrate or inhibitor is that of Schechter and Berger (1967). (P1–P1' are the scissile bond residues.) We felt that a peptide/substrate model would increase our understanding of NVP substrate specificity. To model the peptide/NVP complex we consulted various protease/inhibitor complex studies performed previously and used, as the starting point for the substrate conformation, the conformation of residues 14 to 21 of OMTKY3 when bound to SGPB (Read et al., 1983) (See Materials and methods for modeling details). For the peptide/NVP model, canonical substrate-binding interactions exist, including interactions between the enzyme and the substrate P4–P2 residues. In order to confirm the validity of this model, we determined the NVP-substrate complex crystal structure (H-Glu-Ala-Leu-Phe-Gln-pNA (Bachem) was used as a substrate). As the result, although the resolution is low, the determined electron density maps indicate the main-chain of the substrate and support this substrate binding model (Figure 2-14). Figure 2-15C portrays the docked substrate, with the NVP oxyanion hole and eII strand visible. The substrate fills the NVP binding pocket of molecule A without any severe van der Waals clashes. However, for molecules B, C, and D, the P1 amino acid is too close to the Pro136 carbonyl oxygen to fit in the substrate binding pocket. The inability of molecules B, C, and D to accommodate the P1 residue is a reflection of the Pro136 carbonyl oxygen position, which, as noted previously, protrudes into the oxyanion hole. The oxyanion hole of molecule A in the substrate binding model is shown in Figure 2-15B.

Examination of the peptide/NVP model suggests that substrate binding is

stabilized by interactions involving hydrogen bonds formed by the backbone of residues P5–P2 and NVP residues 158-162 (Figures 2-15A and 2-15C). The principal hydrogen bonds involve: the backbone oxygen of P5 and the backbone nitrogen of Lys162; the backbone nitrogen of P3 and the backbone oxygen of Ala160; and the backbone oxygen of P3 and the backbone nitrogen of Ala160. The P3 and P4 backbone atoms lie in a tunnel formed by residues 159-162 and 108-110 of NVP (Figure 2-15D). This tunnel has a wall is built of Gln110 and Lys162 side chains. The wall seems to have the potential to open and close because the Gln110 and Lys162 side chains do not interact with other residues, which, in turn, provides freedom of movement. Furthermore, since the pockets for the side chains of P4 and P3 are large, the pockets can accommodate variously sized residues (e.g., Thr, Leu, Ala, Met or Phe in the P4 pocket, and Thr, Ser, Gln or His in the P3 pocket). In the model the allowed volume for the P1' residue is small. This observation is consistent with the fact that glycine or alanine is the P1' residue. The structure of the model explains the variety in P3 and P4 residues and why a glycine or an alanine is required at P1'.

2.3.5 The S1 and S2 sites

The residues at NV ORF1-polyprotein cleavage sites are glutamine or glutamic acid at the P1 site and glycine or alanine at the P1' site. No other residues occupy these sites (Figure 2-12). His157, which is part of the S1 site, is important to substrate binding (Figure 2-15C). Mutagenesis of His157 to any other tested residue severely reduces activity (Someya et al., 2002). For HRV $3C^{pro}$, HAV $3C^{pro}$, PV $3C^{pro}$ and coronavirus M^{pro}, there are histidines at the S1 sites and their

imidazoles interact with substrate P1 carboxamide side chains (Allaire et al., 1994; Matthews et al., 1994; Bergmann et al., 1997; Mosimann et al., 1997; Anand et al., 2002; Anand et al., 2003; Yang et al., 2003).

In the NVP/peptide model, the His157 imidazole is positioned to interact with the P1 glutamine side chain. His157 is centered in a hydrophobic pocket, which is composed of Ile85, Ile87, Leu97, Val99, Leu121, Pro136, Tyr143, Ala160, and Val167. Together, the Leu135 backbone, the Pro136 pyrrolidine ring, and the Ala160 methyl form the entrance to the S1 pocket. To ensure proper placement of the His157 imidazole, it is stabilized by a hydrogen bond with the buried Tyr143 hydroxyl, which has no other hydrogen-bonding partner. The average distance of His157$N^{\delta 1}$–Tyr143O^{ζ} is 2.77 Å ± 0.12 Å (Figure 3-11). NV $3C^{pro}$ S1 pocket residues are highly conserved and Tyr143 is always present. The S1 site histidines are similarly stabilized in other 3C or 3C-like proteases. For PV $3C^{pro}$, the hydroxyl of Tyr138 hydrogen bonds to the S1 His161 imidazole and for coronavirus M^{pro}, the analogous interaction is between the side chains of Tyr160 and His162. For HAV $3C^{pro}$, Glu132 interacts through two water molecules with the S1 His191. Examination of the NVP/peptide model shows that the oxygen of the P1 glutamine side chain can hydrogen bond with the hydroxyl of Thr134, which is a residue that is conserved in all noroviruses (Figure 2-15C). The distance between the two atoms is ~2.7 Å. A P1 residue may be stabilized by both His157 and Thr134 in the NV $3C^{pro}$s. In addition, as mentioned above, NVs have glutamic acid or glutamine at the P1 site, but PV, HAV, HRV and coronavirus have just a glutamine at the P1 site. In comparison of NVP structure with those $3C^{pro}$s, two major differences were found. The one is the space size of the S1 site.

The S1 site of NVP is larger than other viral $3C^{pro}$s, therefore the $\chi 1$ of the P1 residue seems to be easy to rotate. The other point is Asp138 of NVP. Although other residues of the S1 site are similar to other $3C^{pro}$s, NVP-Asp138, which is located in facing of His157, is different. Other viral $3C^{pro}$s have Gln at the position. If $\chi 1$ of the P1 residue rotates, Asp138 of NVP seems to be easy to receive the side chain of the P1 glutamine and form hydrogenbond further than Gln of other viral $3C^{pro}$s. Additional structural and biochemical studies are required to elucidate this speculation.

A bulky hydrophobic amino acid, such as Leu, Met, and Phe, preferentially occupies the P2 positions of the polyprotein cleavage sites. The S2 site is also a hydrophobic pocket. It consists of Ile109, Val114, and Ala159 side chains and the Arg112 C^{γ} and C^{δ} atoms. The S2 pocket is large enough to accommodate a bulky P2 side chain (Figure 2-15C).

2.3.6 Zn binding

During structural analysis, region of electron density was found that did not correlate either with a water molecule or with protein atoms. Analysis, using XAFS and MAD, indicated that the electron density is associated with a Zn atom. However, the protein had not been deliberately exposed to zinc during either purification or crystallization. To determine the source of the Zn contamination, we prepared buffers, used in the initial purification step, with and without EDTA. Only the crystals of the protein purified in the absence of EDTA contained zinc as detected using XAFS. Therefore, the zinc contaminant was probably introduced when NVP was expressed in E. coli.

The Zn atom is located at the center of the asymmetric unit NVP quartet. It is coordinated to the Ala125 main-chain amides of molecules B and D and to two water molecules which are hydrogen bonded to the Arg100 guanidinium nitrogens of molecules A and C (Figure 2-17). The zinc coordination seems to be a distorted tetrahedron, which is characterized as an intermediate coordination between tetrahedral and square planar.

NVP samples, with and without zinc, were prepared as previous described and their activities assayed spectrophotometrically using H-Glu-Ala-Leu-Phe-Gln-pNA. Both NVP preparations had the same level of activity. Therefore, zinc is not required for proteolysis. Instead, zinc probably stabilizes the NVP structure by tethering the monomers, and, with the result that the quartet is more stable than the monomer. The greater degree of stability would explain why a quartet of protease molecules forms the asymmetric unit.

Table 2-I. Crystallographic data of His-NVP(Se-Met).

Space group	$P6_1$
Unit cell *a*, *b*, *c* (Å)	224.79, 224.79, 118.69
α, β, γ (°)	90.00, 90.00, 120.00
Resolution range (Å)	39.58 – 3.60
Total number of reflections	145,080
Number of unique reflections	73,714
Completeness (%)	97.5 (95.2)
R_{merge}	0.161 (0.410)
$\langle I/\sigma I \rangle$	3.8 (1.8)

The values in parentheses are for the high-resolution shell.

$R_{merge} = \sum\sum[|I(h)-I(h)_i| / I(h)]$, where $I(h)$ is the mean intensity after rejections.

Table 2-II. Data collection [b]

Crystal data				
Space group	$P6_1$			
Unit cell dimensions (Å)	$a = b = 128.89$ and $c = 118.08$			
Data collection				
Crystal	Native	Hg-peak	Hg-edge	Hg-remote
Wavelength (Å)	1.0000	1.0050	1.0090	0.9800
Resolution (Å)	42.41-3.00	43.53-2.80	43.53-2.80	43.53-2.80
	(3.11-3.00)	(2.90-2.80)	(2.90-2.80)	(2.90-2.80)
Reflections	122454	148926	147395	147765
Unique	22748	53872	53774	53342
Completeness (%)	99.7 (99.9)	99.8 (100)	99.7 (100)	99.8 (100)
R_{merge} [a]	0.089 (0.484)	0.088 (0.438)	0.088 (0.429)	0.089 (0.435)
$I/\sigma(I)$	11.6 (3.6)	8.8 (2.5)	8.9 (2.5)	8.8 (2.5)

[a] $R_{merge} = \sum_i\sum_{hkl}|I_i-\langle I\rangle|/\sum_i\sum_{hkl}I_i$: I_i and $\langle I\rangle$ are the observed intensity and the average intensity obtained from multiple measurements, respectively.

[b] Numbers in parentheses indicate the highest-resolution shell.

Table 2-III. Refinement parameter

	Native		Hg-peak	
Resolution range (Å)	8.00-3.20		8.00-2.80	
R_{work}[a], R_{free}[a]	0.194 (0.272), 0.225 (0.321)		0.206 (0.299), 0.240 (0.342)	
Ramachandran plot[c]	Allowed regions	100 %	Allowed regions	100 %
	Disallowed regions	0 %	Disallowed regions	0 %
Number of atoms	Protein atoms	5212	Protein atoms	5212
or molecules	Tartrate atoms	20	Tartrate atoms	20
	Water molecules	37	Water molecules	127
			Hg	8
Average B (Å^2)	67.7		64.4	
Matthews coefficient / solvent %	3.83 / 67.59 %		3.77 / 67.15 %	

[a] $R = \sum_{hkl}||F_{obs}(h,k,l)|-k|F_{calc}(h,k,l)||/ \sum_{hkl}|F_{obs}(h,k,l)|$: R_{work} was calculated using a set of reflections where 10% of the total reflections had been randomly omitted from the refinement and used to calculate R_{free}.

[b] Percentage of residues in the "allowed" and "disallowed" region of the Ramachandran plot (PROCHECK)

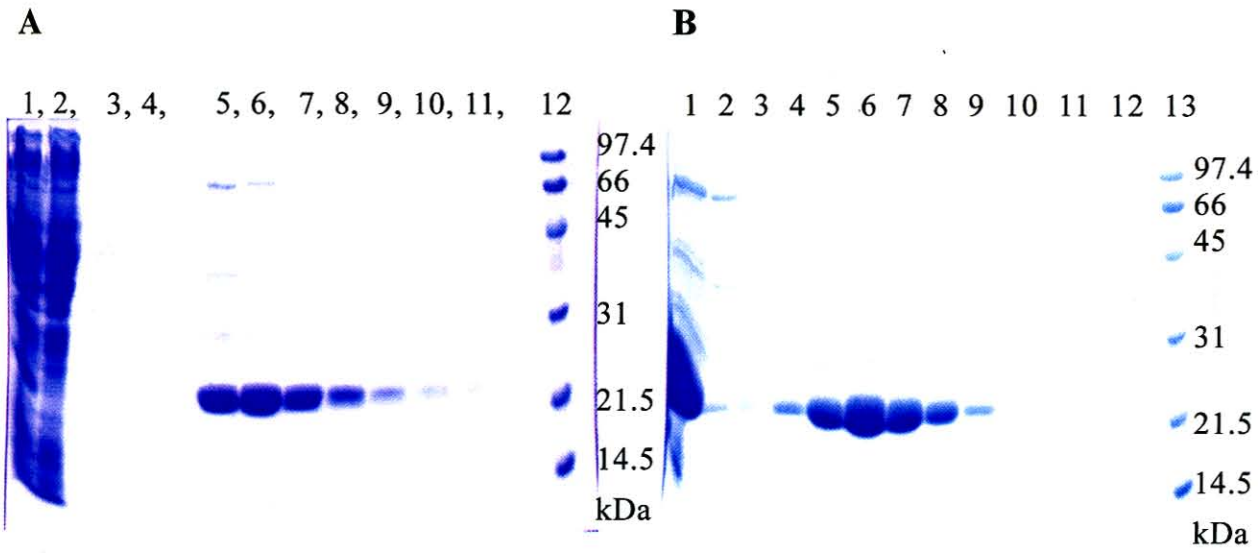


Figure 2-1. Coomassie-stained SDS-PAGE from purification of His-NVP (Se-Met). (A) The purification by a Ni-NTA Super Flow column. Lane 1, supernatant of cell lysis; Lane 2 flow-through; Lane 3~4, wash; Lane 5~11, elusion; Lane 12, molecular-weight markers. (B) The purification by a gel filtration column Superdex75. Lane 1, combined fractions (Lane 5~11 of A); Lane 2~12, fractions by gel filtration; Lane 13, molecular-weight markers. Lane 5~9 were used for the crystallization.

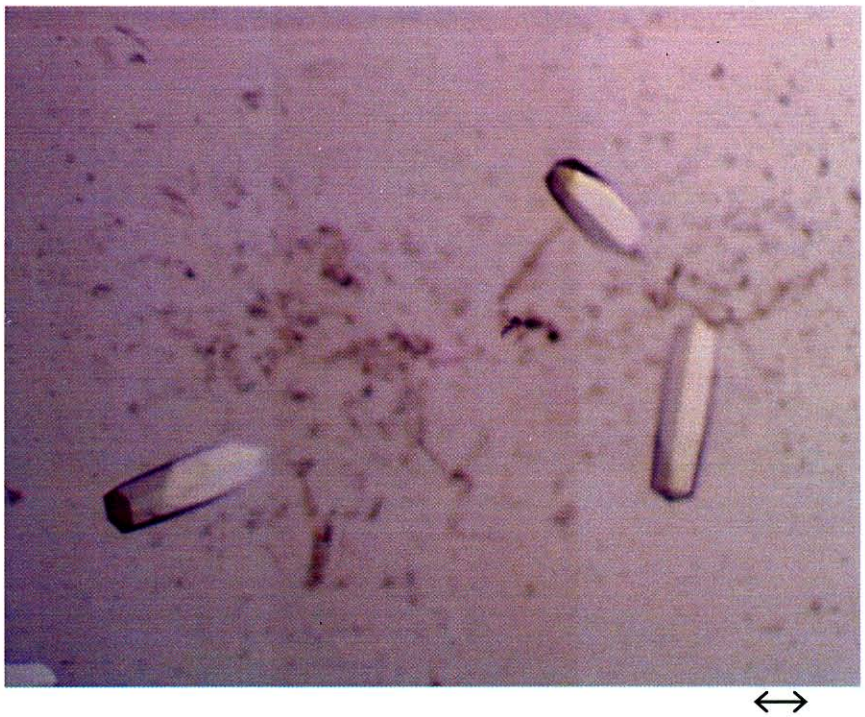


Figure 2-2. His-NVP (Se-Met) crystals. His-NVP (Se-Met) crystals were grown by the hanging-drop vapor diffusion method in a week. The size of the crystals is approximately 0.3 x 0.1 x 0.1 mm^3.

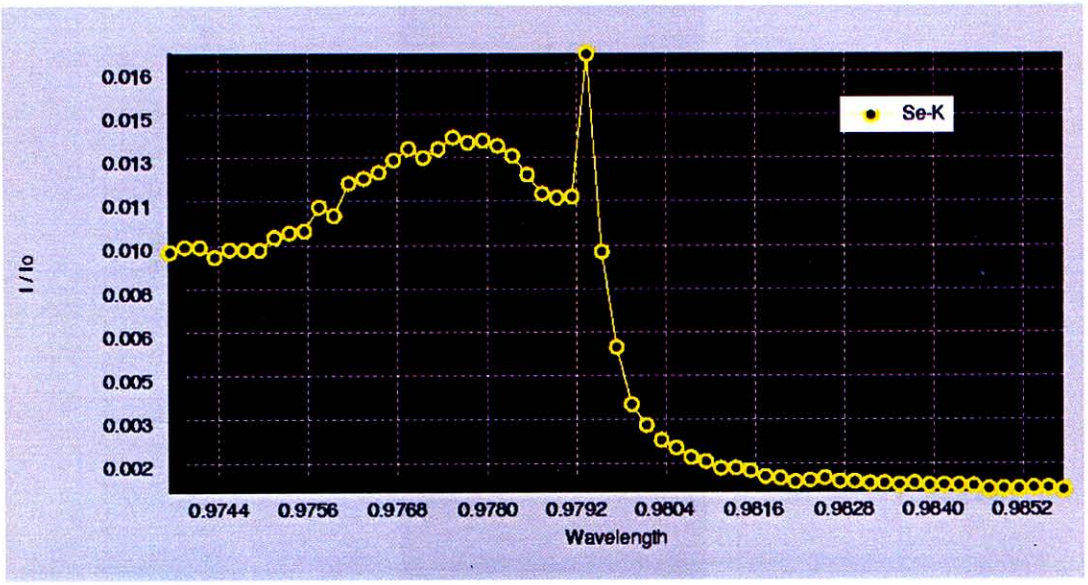


Figure 2-3. XAFS data of His-NVP (Se-Met) crystal around the K absorption edge of selenium. It was confirmed by the measurement of the radiation of the fluorescence X-ray, XAFS, that the absorption of X-ray was observed in His-NVP (Se-Met) crystal around the K absorption edge of selenium, 0.9797 Å.

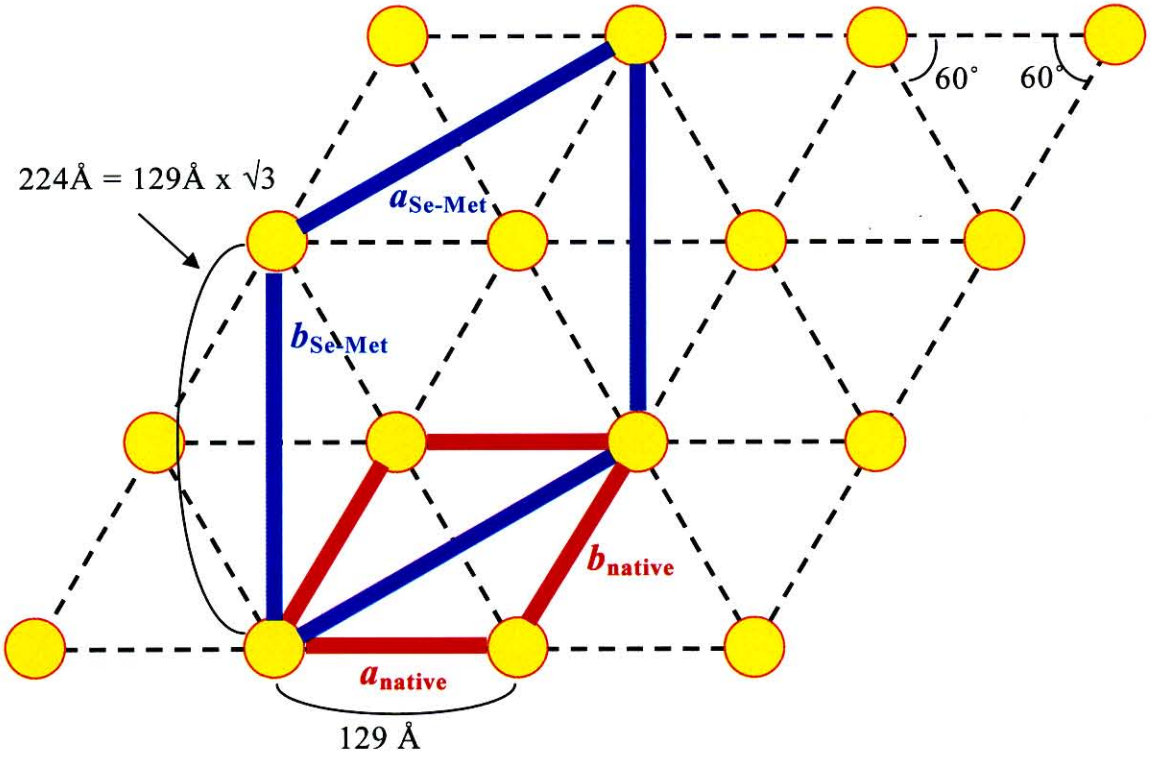
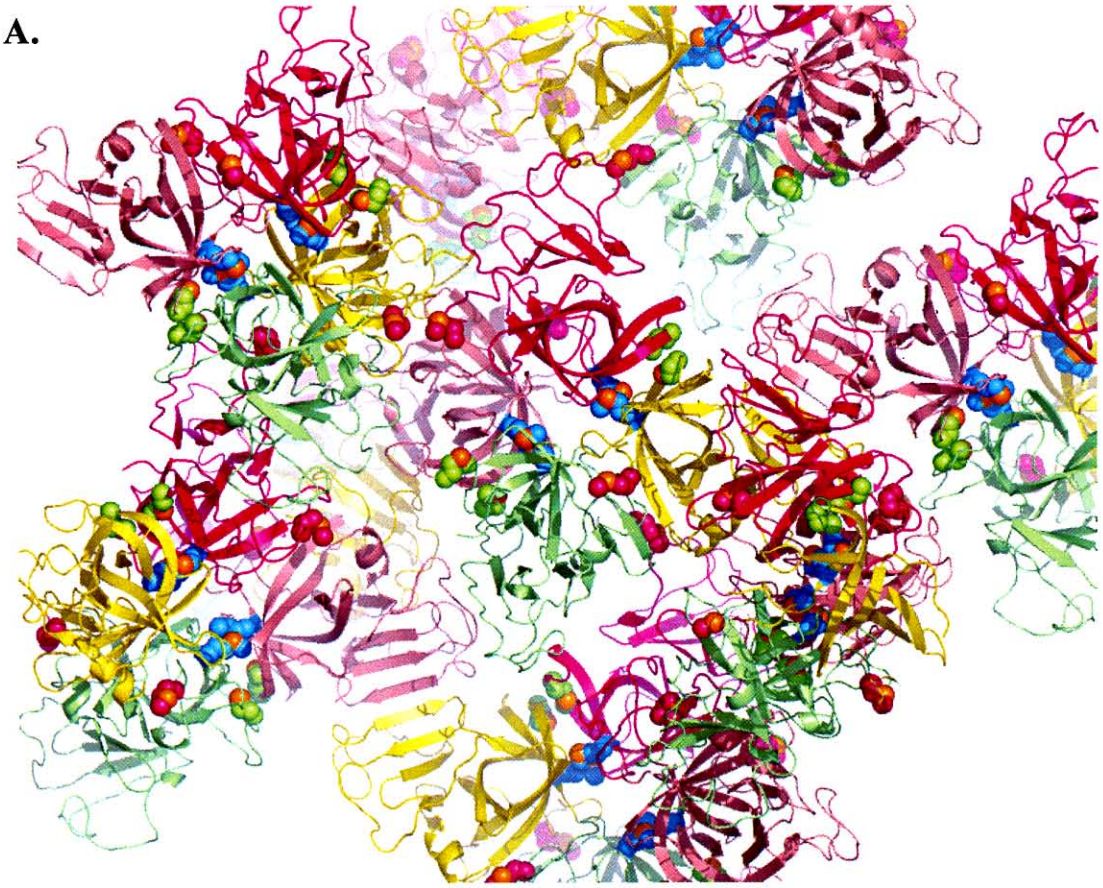


Figure 2-4. The comparison between the lattice (*a* and *b*) of the native crystal and of the Se-Met substitute crystal. The lattice condition of the hexagonal ($P6_1$) crystals was illustrated as the XY-plane formed by *a* and *b* axial. Yellow circles show the estimated asymmetric units on the XY-plane. Red lines (a_{native} and b_{native}) and blue lines ($a_{Se\text{-}Met}$ and $b_{Se\text{-}Met}$) show the lattice parameters of the native crystal and the Se-Met substitute crystal, respectively.

A.



B.

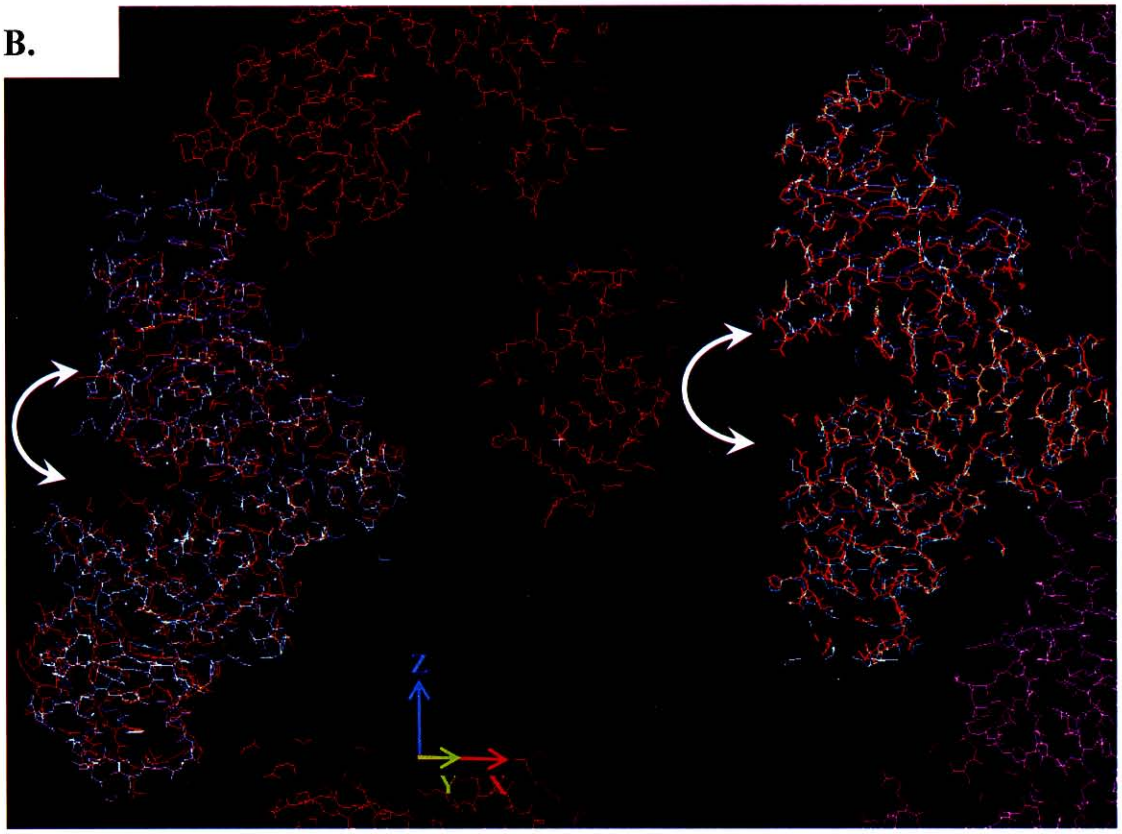


Figure 2-5. The methionine residues in native crystal, and the comparison of His-NVP(Se-Met) structure and native NVP structure.

(A) The methionine residues on the surface of NVP in native crystal. The carbon atoms of Met107, Met120 and Met130 in native NVP structure (the details are shown in Chapter 2.3) are colored green, light blue and magenta, respectively. The sulfur atoms of Met107, Met120 and Met130 are represented as orange spheres. (B) After native NVP structure was determined (Chapter 2.2), the His-NVP(Se-Met) structure was determined by MR method using native NVP structure. The superimposed His-NVP(Se-Met) structure was shown as light-blue sticks. The native NVP structure and the molecules generated by symmetric operation were shown as red- and purple-sticks. In comparison with native structure, the His-NVP(Se-Met) structure spreads to the direction of the white arrow.

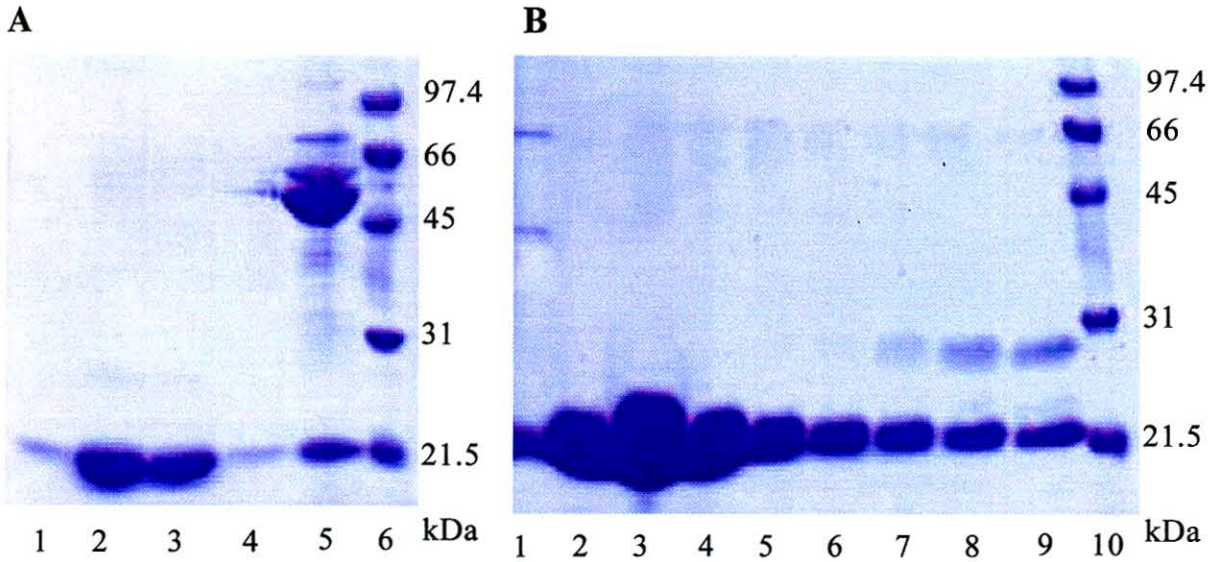


Figure 2-6. Coomassie-stained SDS-PAGE from purification of NVP. (A) The purification by a chitin column. Lane 1~4, elution after 40h cleavage; Lane 5, chitin beads after 40h cleavage; Lane 6, molecular-weight markers. (B) The purification by a gel filtration column Superdex75. Lane 1~9, fractions by gel filtration; Lane 10, molecular-weight markers. Lane 2~6 were used for the crystallization.

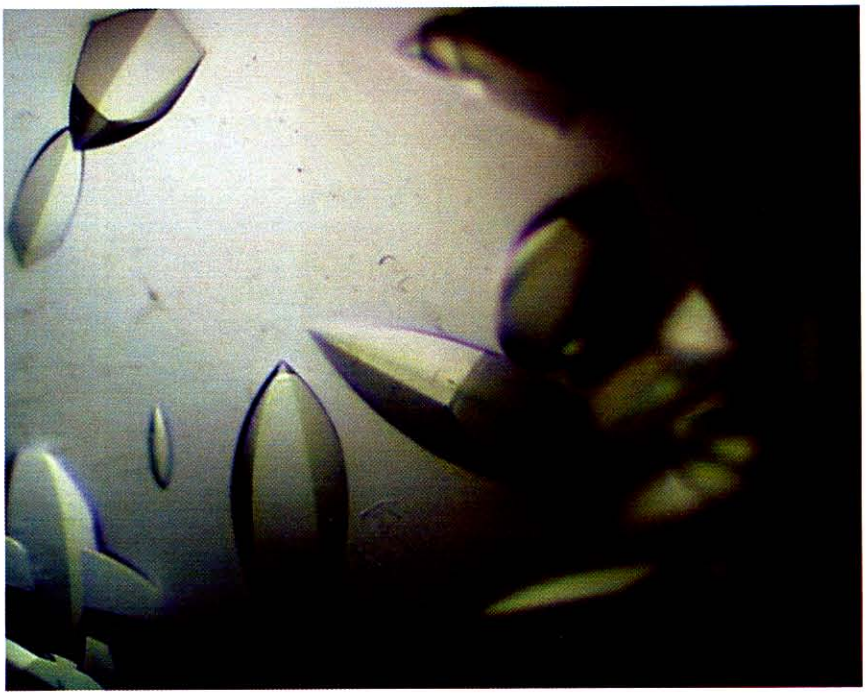


Figure 2-7. Native NVP crystals. NVP crystals were grown by the hanging-drop vapor diffusion method in two week. The size of the crystals is approximately 0.4 x 0.3 x 0.3 mm^3.

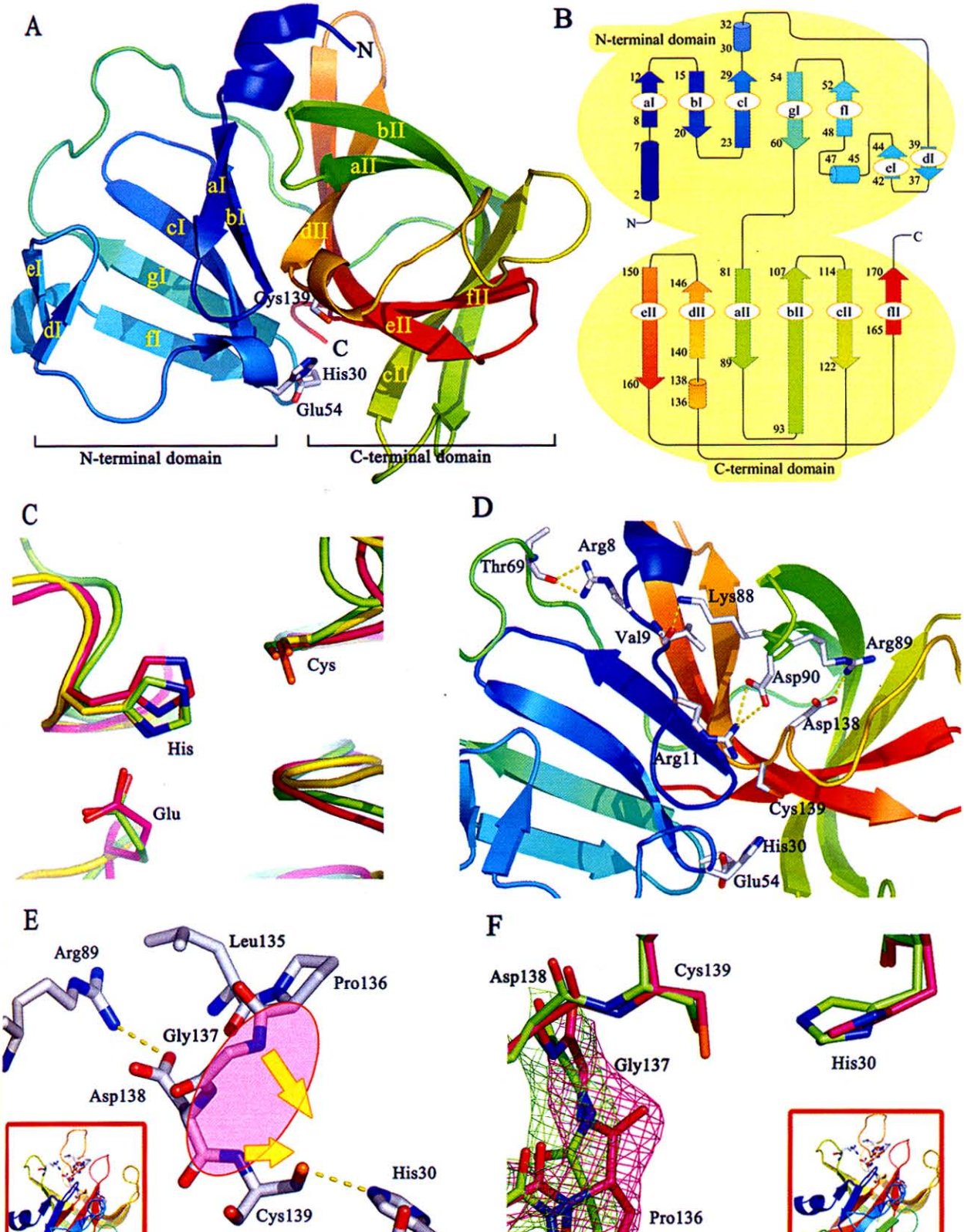

A
N
bII
aII
aI
bI
cI
dII
eI
gI
Cys139
fII
dI
fI
eII
C
His30
cII
Glu54
N-terminal domain
C-terminal domain
B
N-terminal domain
C-terminal domain
C
His
Glu
Cys
D
Thr69
Arg8
Lys88
Val9
Arg89
Asp90
Asp138
Arg11
Cys139
His30
Glu54
E
Arg89
Leu135
Pro136
Gly137
Asp138
His30
Cys139
F
Asp138
Cys139
Gly137
His30
Pro136
Leu135

Figure 2-8. The NVP crystal structure. (A) The NVP global structure. The N- and C-terminal domains are identified in the figure. Helices and β-strands are represented as ribbons and arrows, respectively. The three active site residues are labeled and represented as stick figures. (B) The NVP secondary structure topology. Residue numbers delineate the beginning and end of each the secondary structure element. (C) Superimposition of the NVP active site residues (green) onto the spatially equivalent HAV $3C^{pro}$ residues (yellow) and onto the HRV $3C^{pro}$ residues (purple). (D) The hydrogen bond network that stabilizes the active site or substrate binding site. Dashed yellow lines indicate hydrogen bonds; residues, contributing to the network, are labeled. (E) The oxyanion hole. The catalytic residues, Cys139 and His30, are shown. Dashed yellow lines indicate hydrogen bonds. The oxyanion hole is highlighted in pink. (F) Superimposition of molecules A and C oxyanion holes. The carbons of molecule A and C are colored green and purple, respectively. The $2|F_o|-|F_c|$ electron density maps of the Gly137 and Pro136 backbones are represented by colored mesh.

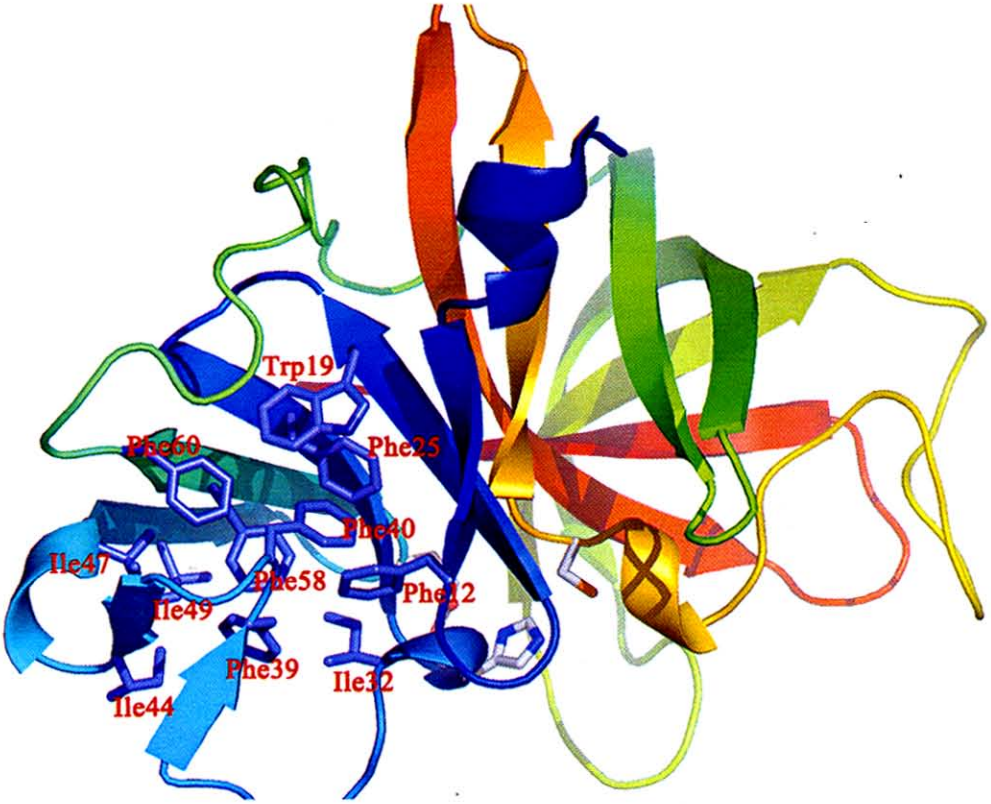


Figure 2-9. The NVP N-terminal domain hydrophobic core. The core hydrophobic residues are represented as stick figures and are the color of slate.

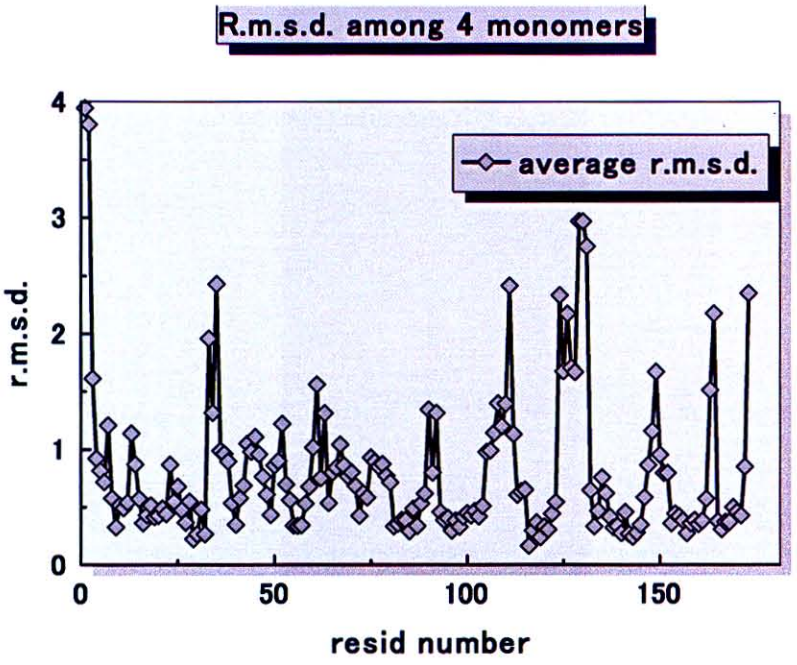


Figure. 2-10. Root mean square differences among the C_α atomic coordinates after superimposition of monomers A, B, C, and D.

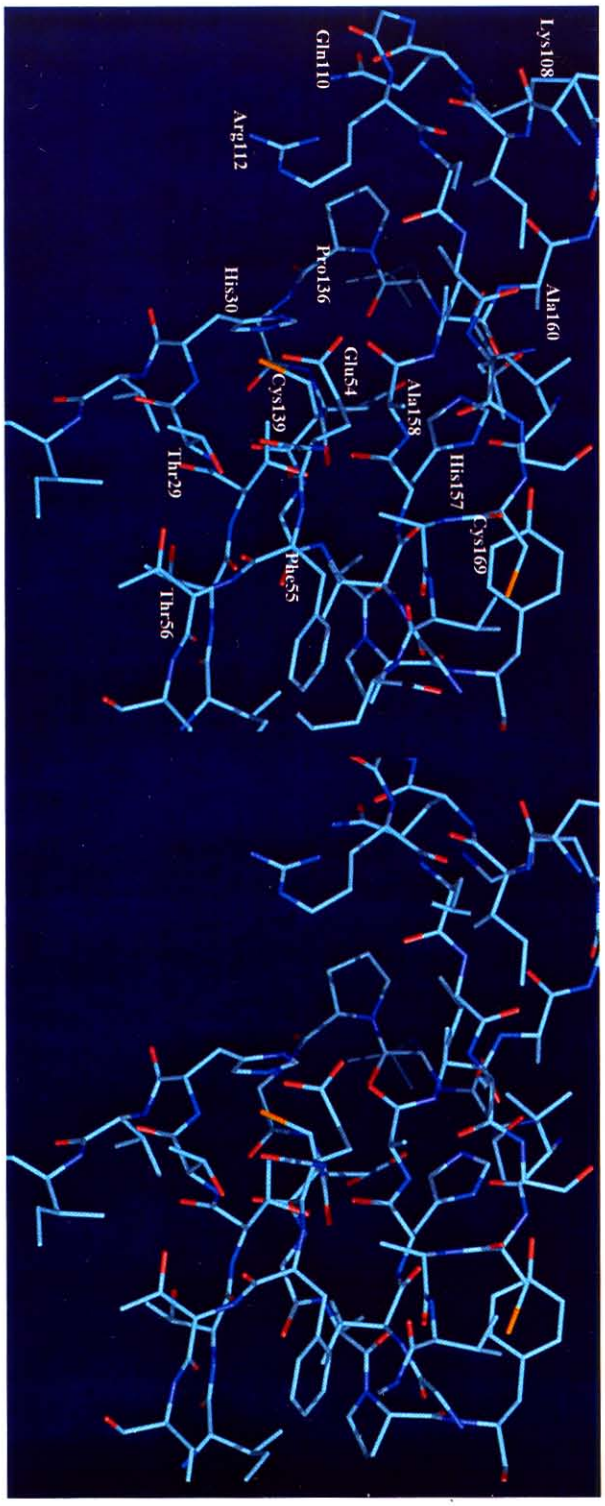


Figure 2-11. A stereo view of the structure of NVP around active site Cys139.

The main chain of the structure is shown in blue.

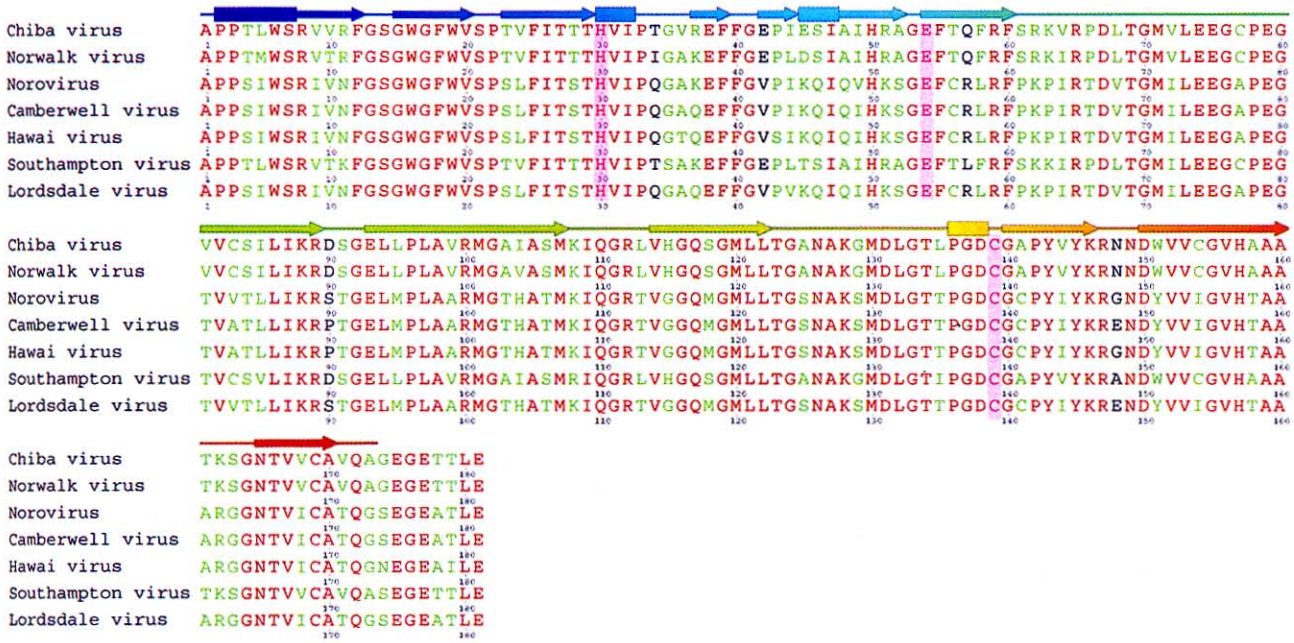


Figure. 2-12. Sequence alignment of Norovirus 3C proteases. The initial alignment, followed by adjustments to improve agreement with the NVP structure, were performed using CLUSTAL X, version 1.81 (Thompson *et al.*, 1997) and Secseq (D. E. Brodersen, unpublished software, http://xray.imsb.au.dk/~deb/secseq). NVP β-strands and α-helices were introduced into the figure using PROMOTIF (Hutchinson and Thornton, 1996) and Secseq. The amino acid sequences were translated from the genomic polyprotein sequences that are available at the NCBI/GenBank (accession Nos: Norwalk, NP_786949; Southampton; A37491, BS5; AAC64602, Camberwell; AAD33960, Lordsdale; P54634, Hawaii; AAB97767).

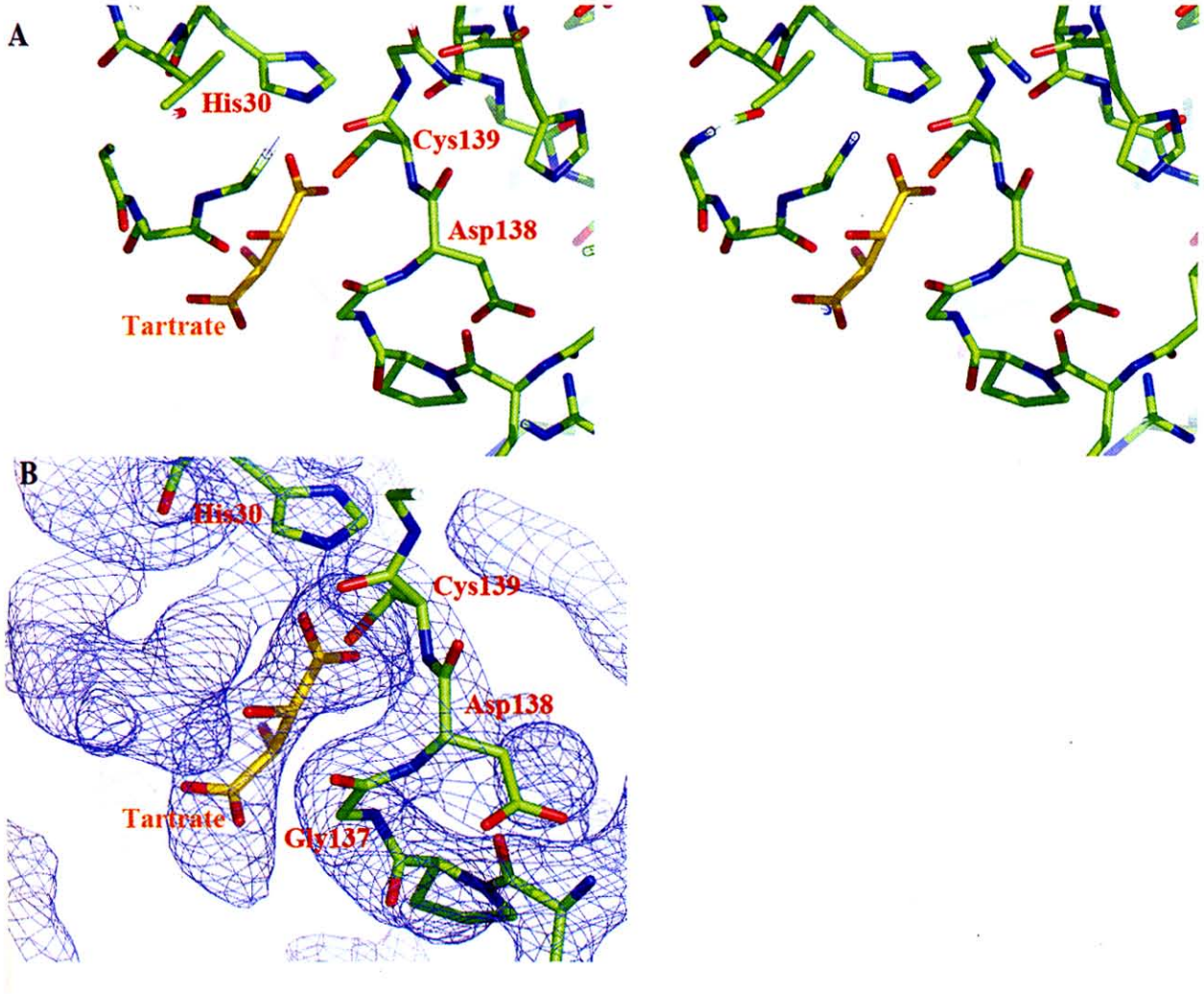


Figure 2-13. (A) A stereo view of tartrate and NVP. The carbon atoms of tartrate and NVP are colored green and yellow, respectively. (B) The $2|F_o|-|F_c|$ electron density maps of tartrate and NVP.

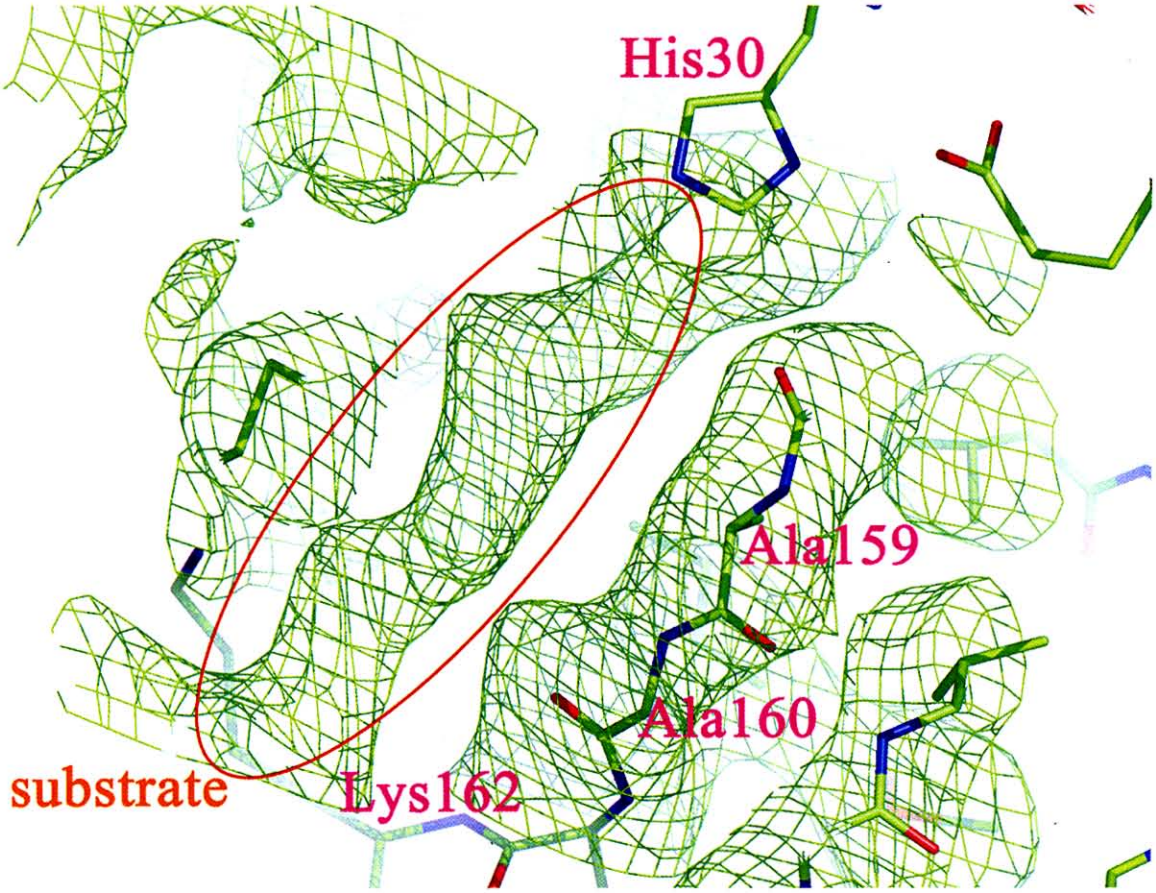


Figure 2-14. NVP/substrate complex structure. The $2|F_o|-|F_c|$ electron density maps are shown by green mesh. The carbon atoms of NVP are colored green.

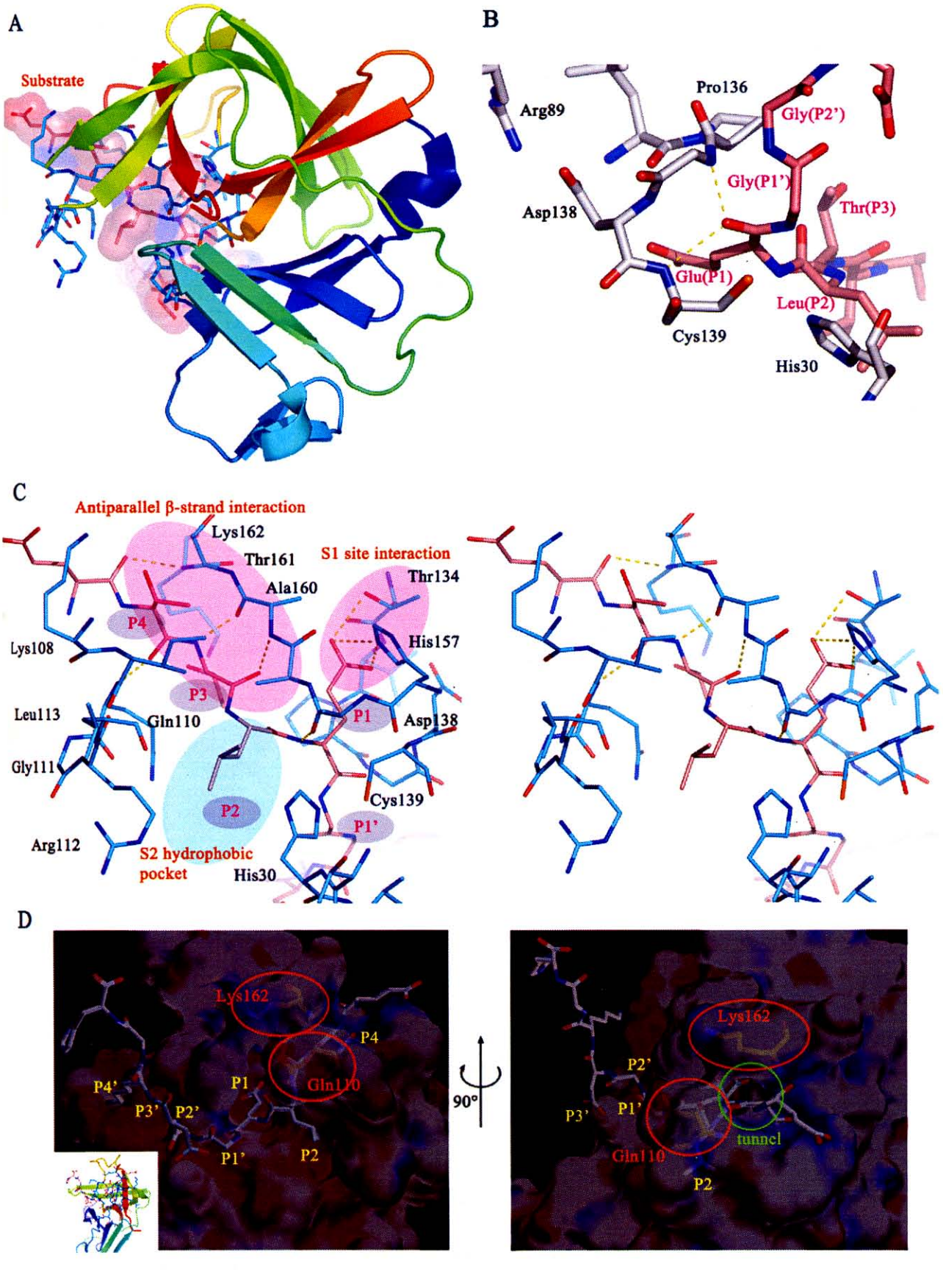

A
Substrate
B
Arg89
Pro136
Gly(P2')
Gly(P1')
Asp138
Thr(P3)
Glu(P1)
Leu(P2)
Cys139
His30
C
Antiparallel β-strand interaction
Lys162
Thr161
Ala160
S1 site interaction
Thr134
His157
P4
Lys108
P3
Leu113
Gln110
P1
Asp138
Gly111
Cys139
P2
P1'
Arg112
S2 hydrophobic pocket
His30
D
Lys162
P4
P4'
P1
Gln110
P3'
P2'
P1'
P2
90°
Lys162
P2'
P3'
P1'
funnel
Gln110
P2

Figure. 2-15. (A) The NVP/substrate model. The substrate is represented by stick figures and is surrounded by a molecular surface representation. Secondary structures are colored as in Figure 1. (B) The NVP oxyanion hole complexed with substrate. The substrate is predominantly colored pink and NVP is predominantly colored white. (C) A stereo view of the model showing substrate bound to NVP. The substrate is predominantly colored pink and NVP is predominantly colored blue. The antiparallel β-strand interactions and S1 site interactions are highlighted pink. The S2 hydrophobic pocket is highlighted sky blue. (D) The electrostatic potential of the NVP/substrate model. Positive and negative potentials are colored blue and red, respectively. 'P' denotes substrate residues. An orange circle surrounds Lys162 and Gln110 of NVP. The carbon atoms of Lys162 and Gln110 are colored yellow. The tunnel of the substrate binding site is shown an green circle.

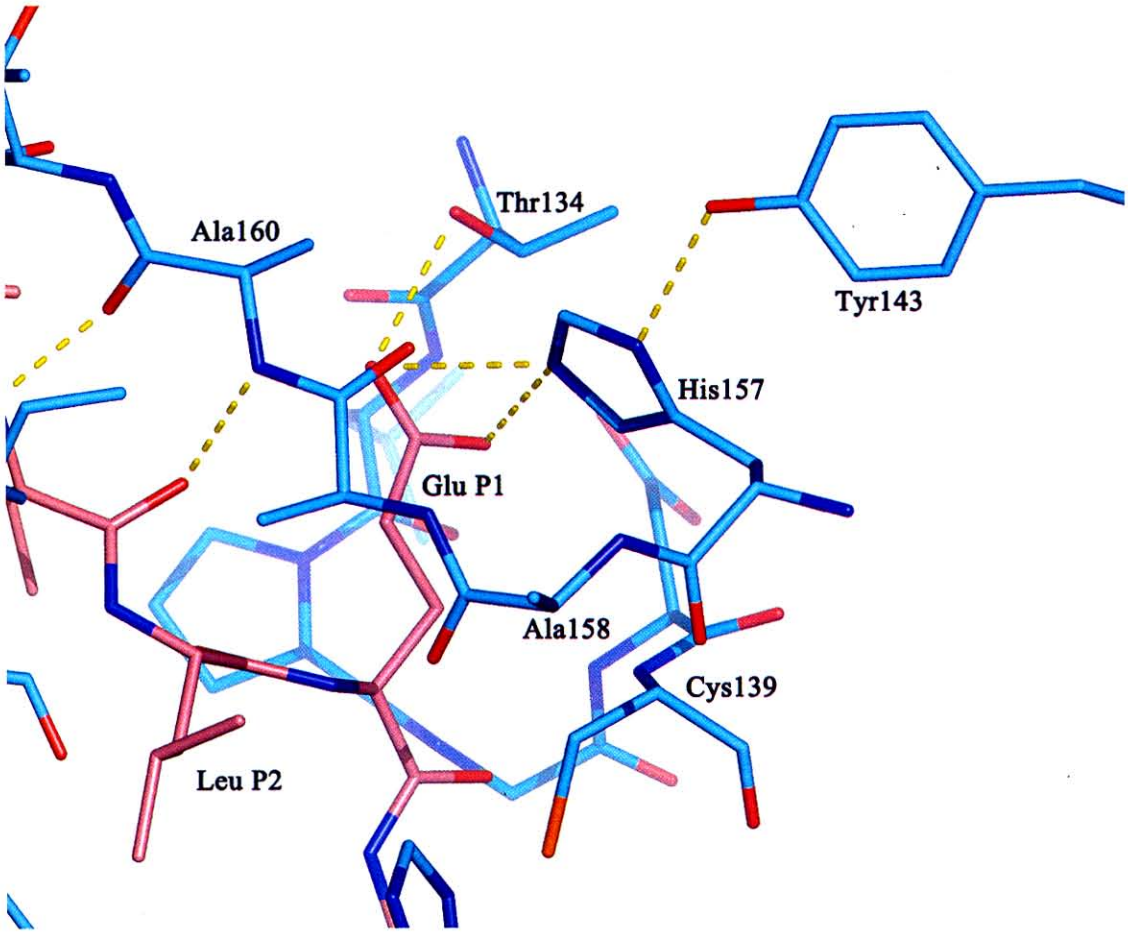


Figure 2-16. The NVP/substrate model. The carbon atoms of substrate and NVP are colored red and blue, respectively.

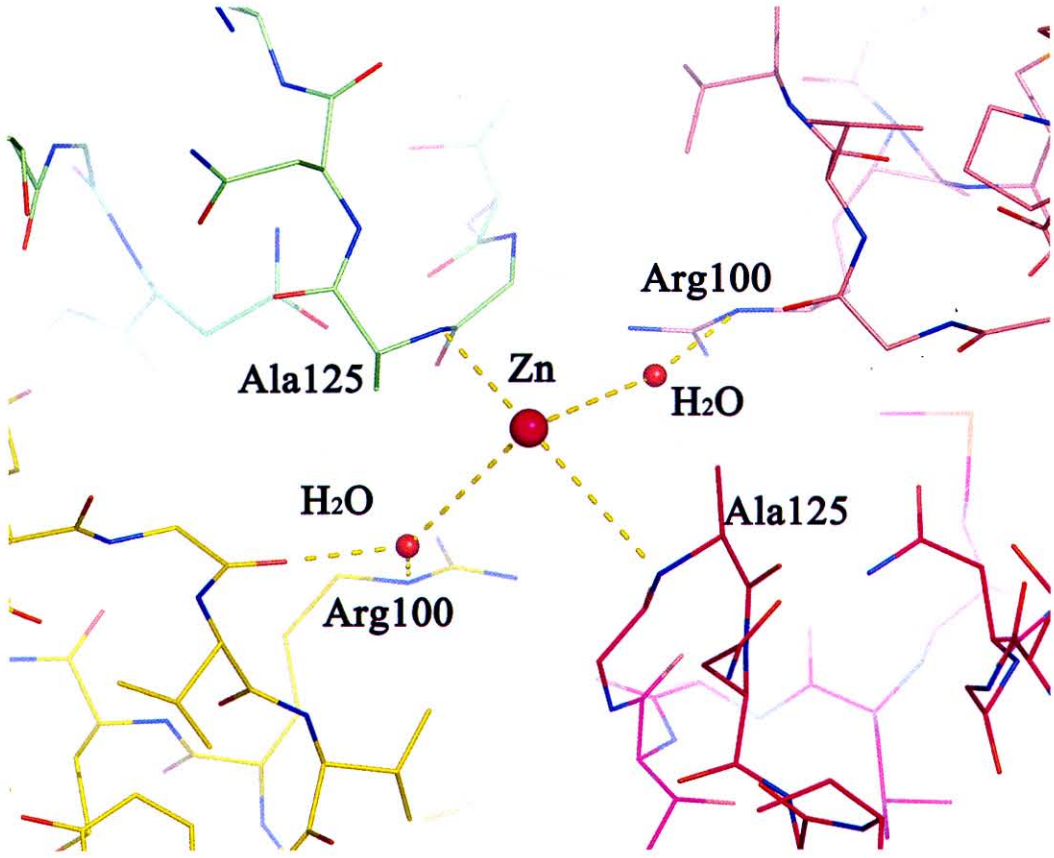


Figure 2-17. Zn coordinated to the four molecules of the asymmetric unit. The carbon atoms of molecules A, B, C, and D are colored yellow, purple, red, and green, respectively. Water molecules (red) and the Zn (purple) are represented as spheres.

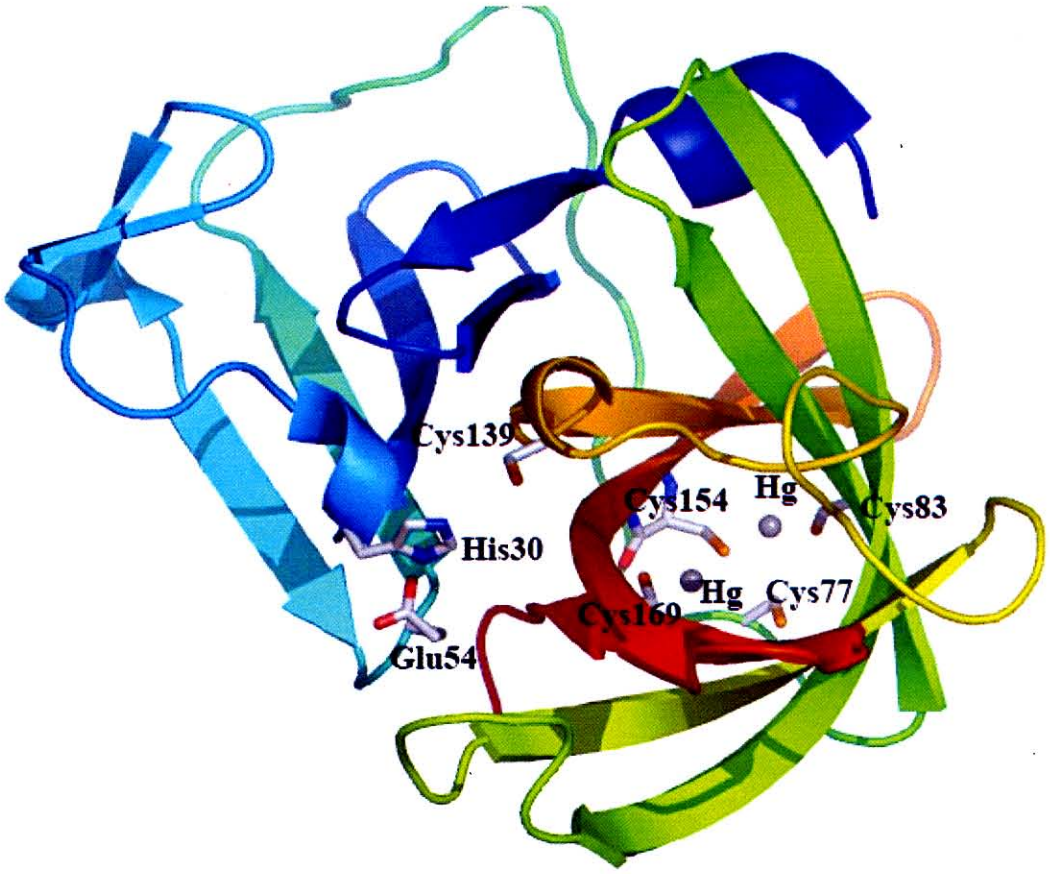


Figure 2-18. Hg coordinated away from the active site of NVP. Hg atoms are represented as grey spheres.

Chapter 3: Structural analysis of 3BC precursor and crystallization of 3B(VPg)

As described in Chapter 1, the details of the cleavage by NVP and the function of 3B(VPg) are almost still unknown. This chapter focused on there. In addition, I considered the structure analysis of NVP with a substrate to reinforce the research of substrate-binding in Chapter 2.

In order to investigate the details of *trans/cis* cleavage by NVP at 3B-3C cleavage site, the structural features of 3B(VPg), and the complex structure of NVP with a substrate, I tried to determine the crystal structure of 3BC precursor, 3B(VPg), and NVP with a substrate.

3.1 Material and methods

3.1.1 Cloning, expression and purification of 3BC precursor

ChV ORF1-C139A which had Ala139 mutation at the Cys139 position of 3C protease was prepared using the vector pUC18[ORF1-C139A] from National Institute of Infectious Diseases. 3BC-C139A gene encoding 3B region and 3C-C139A of ORF1 had been amplified from ORF1-C139A, and inserted into pTYB11 of the IMPACT system (New England BioLabs) as the N-terminal segment fused to the intein and chitin-binding domain unit. The recombinant vector pTYB11[3BC-C139A] was introduced in *E. coli* strain ER2566 (New England BioLabs) and cultures were grown to an OD600 of 0.6-0.7 at LB broth supplemented with 100μg/ml ampicillin. The expression of recombinant 3BC-C139A, called as 3BC precursor, was induced by the addition of IPTG to a final concentration of 0.1 mM at 17°C overnight. The harvested *E. coli* was lysed by treatments with sonication in a lysis buffer containing 20 mM HEPES-NaOH,

pH 8.2, 0.5 M NaCl, 1 mM EDTA. The cell homogenate was centrifuged at 20,000 x g for 30 min. The supernatant was applied to a chitin bead column (New England BioLabs) equilibrated with lysis buffer and was washed by a wash buffer, 20 mM HEPES-NaOH, pH 8.2, 1.0 M NaCl, 1 mM EDTA. On-column self-cleavage was carried out for 40 hours using a cleavage buffer of 20 mM HEPES-NaOH, pH 8.2, 0.5 M NaCl, 1 mM EDTA and 50 mM DTT. The fraction containing 3BC-C139A was collected and concentrated. Further purification was performed by gel filtration over a HiLoad 16/60 Superdex (75 pg; Amersham Pharmacia Biotech) column equilibrated with 20 mM HEPES-NaOH, pH 8.2, 200 mM NaCl. The fractions containing 3BC-C139A were combined and concentrated to 10 mg/ml. The purity of 3BC-C139A was confirmed by SDS-PAGE analysis (Figure 3-1).

3.1.2 Cloning, expression and purification of 3B(VPg)

3B(VPg) gene encoding 3B region of ORF1 had been excised from ORF1-C139A, and inserted into pTYB11 of the IMPACT system (New England BioLabs). The recombinant vector pTYB11[3B(VPg)] was introduced in the *E. coli* strain ER2566 (New England BioLabs) and cultures were grown to an OD600 of 0.6-0.7 at LB broth supplemented with 100 μg/ml ampicillin. The expression of recombinant 3B(VPg) was induced by the addition of IPTG to a final concentration of 0.1 mM at 17°C overnight. The harvested *E. coli* was lysed by treatments with sonication in a lysis buffer containing 20 mM HEPES-NaOH, pH 8.2, 0.5 M NaCl, 1 mM EDTA. The cell homogenate was centrifuged at 20,000 x g for 30 min. The supernatant was applied to a chitin bead column (New

England BioLabs) equilibrated with lysis buffer and was washed by a wash buffer, 20 mM HEPES-NaOH, pH 8.2, 1.0 M NaCl, 1 mM EDTA. On-column self-cleavage was carried out for 40 hours using a cleavage buffer of 20 mM HEPES-NaOH, pH 8.2, 0.5 M NaCl, 1 mM EDTA and 50 mM DTT. The fraction containing 3B(VPg) was collected and concentrated. Further purification was performed by gel filtration over a HiLoad 16/60 Superdex (75 pg; Amersham Pharmacia Biotech) column equilibrated with 20 mM HEPES-NaOH, pH 8.2, 200 mM NaCl, 5 mM DTT. The fractions containing 3B(VPg) were combined and concentrated to 10 mg/ml. The purity of 3B was confirmed by SDS-PAGE analysis (Figure 3-2).

3.1.3 Crystallization and data collection of 3BC-C139A

The crystallization conditions were determined by sparse-matrix screening using Crystal Screen I and II (Hampton Research) and sitting drop vapor diffusion plates. As a result of the optimization of these conditions, the best crystals were grown at 20°C using a hanging drop vapor diffusion system consisting of a 3 µl protein solution and 3 µl crystallization buffer drop, equilibrated against crystallization buffer (1 ml). The buffer that produced the best crystals contained 0.1 M Na citrate, pH 5.6, 0.7 M lithium sulfate (Li_2SO_4), 0.3 M AS (ammonium sulfate). Crystals grew within 1 to 2 weeks and typically had dimensions of approximately 0.1 x 0.1 x 0.3 mm^3 (Figure 3-3).

3.1.4 Crystallization of 3B(VPg)

The screening of the crystallization conditions were performed by sitting drop vapor diffusion method using Crystal Screen I & II, Grid Screen AS, Grid Screen MPD, PEG/Ion Screen (Hampton Research), Wizard I & II (Emerald Biostructures), and Structure Screen I & II (Molecular Dimensions Ltd). The concentration of 3B(VPg) was screened from 5 to 20 mg/ml every 5 mg/ml. However a crystal was not observed in any conditions.

3.1.5 Diffraction data collection, processing, and structure determination of 3BC precursor

X-ray diffraction data were collected at in-house X-ray system. All data were collected using crystals flash cooled at -170°C, which were prepared by rapidly transferring the crystal into a cryoprotectant containing up to 30% (vol/vol) glycerol. Diffraction data was collected at 1.5418 Å wavelength (CuKα), using a rotating anode X-ray generator and an R-AXIS IV device detector (Rigaku). The data was processed using the program CrystalClear. The data collection statistics are given in Table 3-I. The crystals belong to the hexagonal space group $P6_5$ with unit cell dimensions $a = b = 131.15$ and $c = 82.27$ Å and the diffract X-ray beyond 3.3 Å resolution. Subsequent calculations were performed using the CCP4 suite of programs (CCP4. 1994).

The F data processed by dtreck2mtz was used for structure determination by molecular replacement method (MR). MR was performed by MOLREP using 3C protease structure (NVP, see Chapter 2) as the partial structure of the search model. The initial R factor, R_{free} and correlation was 0.43, 0.49 and 0.65,

respectively. The model was refined at 3.3 Å resolution using Refmac5 and CNS (Murshudov 1997, Brűnger et al., 1998). The crystallographic *R* factor converged to a value of 0.237, with an associated R_{free} of 0.294. Data quality and refinement statistics are given in Table 3-I. The quality of the structural model and its agreement with the structure factors were checked using the programs PROCHECK, PROMOTIF, and SFCHECK.

3.1.6 Structure determination of the 3BC-C139A complex with the substrate peptide for 3C

In order to determine the complex structure of 3BC-C139A with the substrate peptide for 3C protease, I carried out cocrystallization of the 3BC-C139A with the substrate peptide in a solution similar to that used to prepare the 3BC-C139A crystal. The substrate peptide, T-A-T-S-E-G-K in the 3A-3B cleavage site, was prepared by peptide chemical synthesis. The peptide (final concentration 0.5 mg/ml) was mixed with 10 mg/ml 3BC-C139A so that the mole ratio was 1:1. The crystal was gained in the same condition as the 3BC-C139A crystals. X-ray diffraction data were collected at SPring-8 (Hyogo, Japan). All data were collected using crystals flash cooled at -170°C, which were prepared by rapidly transferring the crystal into a cryoprotectant containing up to 30 % (vol/vol) glycerol. Diffraction data was collected at 1.000 Å wavelength at beamline BL41XU, which was equipped with Jupiter charge-coupled device detector (Rigaku MSC). The data was processed using the program HKL2000 (HKL Research, Inc.). The data collection statistics are given in Table 3-II. The crystals belong to the hexagonal space group $P6_522$ with unit cell dimensions $a = b =$

131.25 and c = 82.85 Å and the diffract X-ray beyond 2.75 Å resolution. Subsequent calculations were performed using the CCP4 suite of programs (CCP4. 1994).

The phase determination was performed by molecular replacement method (MR). MR was performed by MOLREP using 3C protease structure (NVP, see Chapter 2) as the partial structure of the search model. The initial R factor, R_{free} and correlation was 0.45, 0.48 and 0.68, respectively. The model was refined at 2.75 Å resolution using Refmac5 and CNS (Murshudov 1997, Brűnger et al., 1998). The crystallographic R factor converged to a value of 0.216, with an associated R_{free} of 0.255. The density of the substrate peptide was observed in $|Fo|$-$|Fc|$ map, the model was build manually using XtalView (McRee 1999). Data quality and refinement statistics are given in Table 3-II. The quality of the structural model and its agreement with the structure factors were checked using the programs PROCHECK, PROMOTIF, and SFCHECK.

3.2 Results and discussion

3.2.1 Into *cis*- or *trans*-cleavage at the 3B-3C cleavage site, insights were provided by the 3BC-C139A structure and the structural feature of 3B(VPg).

3BC-C139A crystal structure was determined by molecular replacement. However, although the 3C region was well ordered, the 3B region was almost

disordered in the electron density map (Figure 3-4). The structure of the 3C region was not different from the structure of NVP only (see NVP structure in Chapter 2). The N-terminal region of $3C^{pro}$ in 3BC-C139A structure had an α-helix as well as native NVP in Chapter 2. In explanation about the 3B region, there were three possibilities; 1) the region was disordered: although the region existed, was disordered by the high flexibility or did not have a rigid structure, 2) the region was processed and removed in the crystallization and/or purification: 3) the exegesis/solution of the structure analysis by MR method, the region could not be observed, was wrong.

By SDS-PAGE analysis of the crystal, it was confirmed that 3BC-C139A in the crystal had full length. And the lattice parameters of 3BC-C139A crystal, $a = b =$ 131.15 and c = 82.27 Å, were different from the crystal of NVP only, $a = b =$ 128.89 and c = 118.08 Å. In addition, in spite of not building the 3B region in the model, the R factor of the structure model fell down to ~0.24. The 3B region, 136 residues, occupied ~43 % in the 3BC full length, 317 residues. If the 3B region had a rigid structure and affected the diffraction data, R factor could not fall down to the value. And there is enough space for 3B region in the crystal (Figure 3-4). Therefore it was suggested that the region was disordered in crystal structure and that the 3B region usually would be very flexible and not have a rigid structure.

The results of the folding prediction for the 3B region using FoldIndex (Vladimir et al., 2000) indicated that the region might belong to natively unfolded proteins (Figure 3-5B). Furthermore I carried out the expression, purification and crystallization of 3B only. As the result, it was found that the

crystallization of 3B was very difficult and 3B had high solubility. It is known that natively unfolded protein has high solubility depend on the high structural flexibility. The results strictly supported the prediction that 3B belonged to natively unfolded proteins and the explanation that the 3B region was disordered in the 3BC-C139A crystal. Therefore it was found that, in 3BC precursor, the 3C region had the same structure as NVP only and the 3B region had high structural flexibility.

These results give a new insight in whether the 3B-3C self-processing is performed by *trans*- (intermolecular) or *cis*-cleavage (intramolecular). In a *trans*-cleavage, 3BC precursor does not need any conformational changes. On the other hand, in a *cis*-cleavage, 3BC precursor needs conformational changes in the N-terminal region of 3C. The N-terminal regions of viral 3C proteases, such as PV $3C^{pro}$ and α-lytic protease, undergo dramatic conformational changes upon processing (Khan et al., 1999). The N-terminal region of the aforementioned $3C^{pro}$ are extended before self-processing with the 3B-3C cleavage site positioned at the catalytic site of the protease. When self-processing (*cis*-cleavage) occurs, the N-terminal region folds into an α-helix (Khan et al., 1999). The crystal structure of the NVP N-terminal region is also helical. Therefore, it was discussed that NVP may also be able to self-process at the 3B-3C cleavage site. However, 3BC-C139A structure denied the *cis*-cleavage at the 3C-3C cleavage site and suggested *trans*-cleavage, because the 3C N-terminal region in 3BC-C139A was α-helix, the same as the N-terminal region of NVP only. Actually, when 3BC-C139A was processed by NVP, the cleavage at the 3B-3C site occurred, so that the separated 3B and 3C detected by SDS-PAGE analysis.

Therefore, the 3B-3C self cleavage would be performed by not *cis*- but *trans*-cleavage.

On the 3B(VPg) region, it was observed that both crystallization as 3B only and 3BC were very difficult. It will be required in order to determine the structure that the co-crystallization is tried with the proteins/ligands which interact with 3B(VPg). eIF3-subunit7 may fulfill the role as a candidate. Although the function of 3B(VPg) is almost still unknown, it is discussed that 3B(VPg) may function in human translation complex. It was reported that NV 3B VPg binds to the 5'-end of the viral genome and to eukaryotic initiation factors such as eIF3 (Daughenbaugh et al., 2003) (Figure 3-6). The identify of the binding region of eIF3-subunit7 to 3B(VPg) and co-crystallization with eIF3-subunit7 will be required.

3.2.2 The complex structure of 3BC-C139A and the substrate for NVP, 3AB cleavage site polypeptide.

It was identified that the 3C region of 3BC-C139A had the same structure as native NVP structure. In addition, 3BC-C139A had the Ala mutation at the C139 active site so that the 3C region did not have protease activity. Therefore I carried out the structure determination of the complex with the substrate peptide (T-A-T-S-E-G-K in the 3A-3B cleavage site) for NVP using 3BC-C139A crystal.

There was one 3BC-C139A molecule in asymmetric unit, and the electron density of the substrate peptide was observed in $2|Fo|-|Fc|$ map (Figure 3-7). The substrate binding in S2'-S1'-S1-S2 agreed very well with the substrate binding model in Chapter 2, so that the modelling study in the site was proven. However,

in S3-S4 site, the complex structure was different from the modelling study and native NVP. The loop M (residues Ser106 to Val114) of the complex structure formed opener structure than native NVP (Figure 3-8). As the result, hydrogen bonds were formed between $O^{\varepsilon}/N^{\varepsilon}$ of Gln110 and the main chain oxygen of P4 Ala. This was just a snapshot that the tunnel-wall formed by Gln110 and Lys162, opened and trapped the substrate in. This structure strictly supported the discussion in Chapter 2 that the tunnel-wall would have the potential to open and close. Therefore it is found that the loop Ser106-Val114 plays the role as a wall of the tunnel and a lid, does structural conformation change dependent on substrate binding in the open and close form. Furthermore, Gln110, the key residue in the substrate trap, gripes and fixes the main chain of the substrate as a magic-hand. Thus the mutation at Gln110 might reduce the binding ability with substrates and protease activity. Although the site-directed mutagenesis study in NVP was carried out in Someya et al., 2002 etc., Gln110 mutation was not reported. The mutagenesis study at Gln110 and the activity assay will be required to investigate the effect of Gln110 to substrate binding ability.

Table 3-I. Data collection [c]	
Crystal data	
Space group	$P6_5$
Unit cell dimensions (Å)	$a = b = 131.15$ and $c = 82.27$
Data collection	
Crystal	3BC-C139A
Wavelength (Å)	1.54056
Resolution (Å)	111.8-3.30 (3.39-3.30)
Reflections	126742
Unique	12186
Completeness (%)	100 (100)
R_{merge} [a]	0.128 (0.481)
$I/\sigma(I)$	7.7 (4.7)
Refinement parameter	
R_{work} [b], R_{free} [b]	0.237 (0.283), 0.294 (0.341)

[a] $R_{merge} = \sum_i \sum_{hkl} |I_i - \langle I \rangle| / \sum_i \sum_{hkl} I_i$: I_i and $\langle I \rangle$ are the observed intensity and the average intensity obtained from multiple measurements, respectively.

[b] $R = \sum_{hkl} ||F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|| / \sum_{hkl} |F_{obs}(h,k,l)|$: R_{work} was calculated using a set of reflections where 10% of the total reflections had been randomly omitted from the refinement and used to calculate R_{free}.

[c] Numbers in parentheses indicate the highest-resolution shell.

Table 3-II. Data collection [c]

Crystal data	
Space group	$P6_522$
Unit cell dimensions (Å)	$a = b = 131.25$ and $c = 82.85$
Data collection	
Crystal	3BC-C139A with the substrate peptide "3A-3B" (cocrystallization)
Wavelength (Å)	1.0000
Resolution (Å)	50-2.75 (2.82-2.75)
Reflections	150474
Unique	10830
Completeness (%)	99.8 (99.9)
R_{merge} [a]	0.088 (0.453)
$I/\sigma(I)$	8.4 (4.2)
Refinement parameter	
R_{work} [b], R_{free} [b]	0.216 (0.331), 0.255 (0.371)

[a] $R_{merge} = \sum_i\sum_{hkl}|I_i-\langle I\rangle|/\sum_i\sum_{hkl}I_i$: I_i and $\langle I\rangle$ are the observed intensity and the average intensity obtained from multiple measurements, respectively.

[b] $R = \sum_{hkl}||F_{obs}(h,k,l)|-k|F_{calc}(h,k,l)||/\sum_{hkl}|F_{obs}(h,k,l)|$: R_{work} was calculated using a set of reflections where 10% of the total reflections had been randomly omitted from the refinement and used to calculate R_{free}.

[c] Numbers in parentheses indicate the highest-resolution shell.

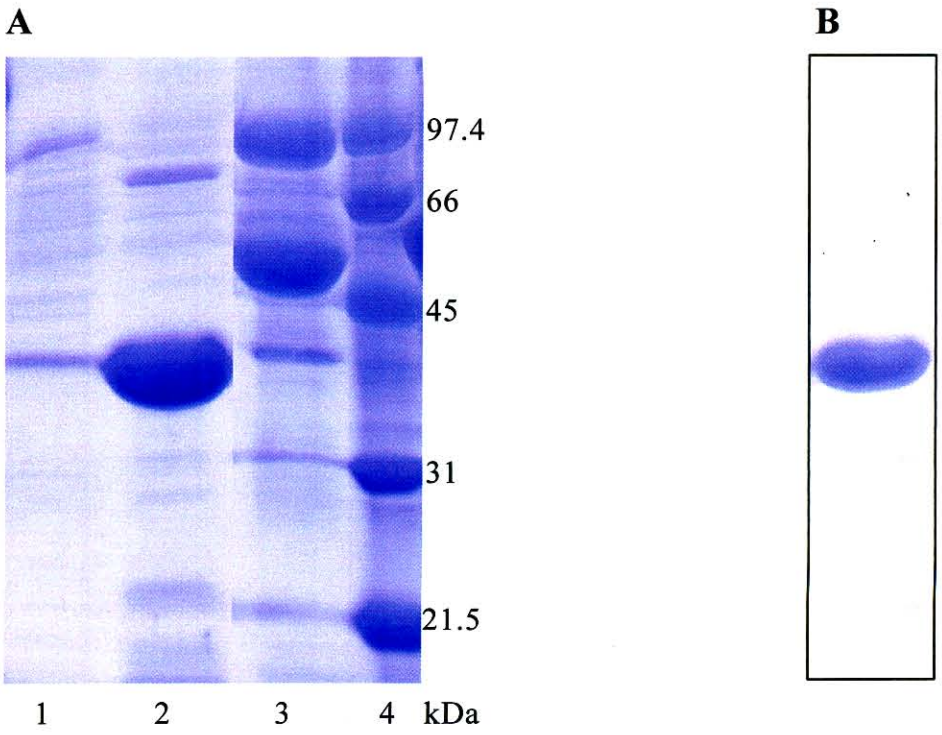


Figure 3-1. Coomassie-stained SDS-PAGE from purification of 3BC-C139A. (A) The purification by a chitin column. Lane 1~2, elution after 40h cleavage; Lane 3, chitin beads after 40h cleavage; Lane 4, molecular-weight markers. (B) The purification by a gel filtration column Superdex75. This is combined fractions by gel filtration. It was used for the crystallization.

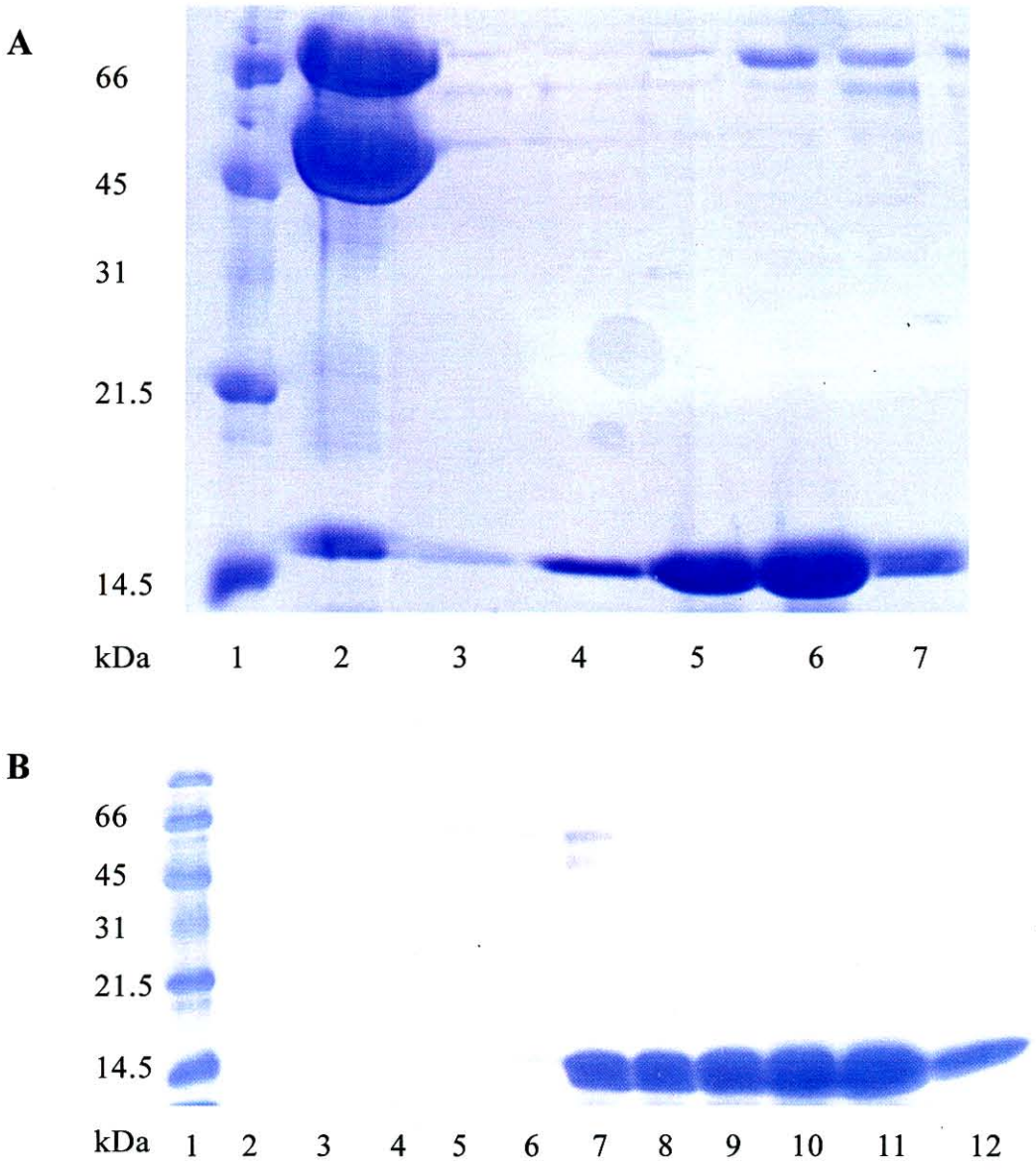


Figure 3-2. Coomassie-stained SDS-PAGE from purification of 3B(VPg). (A) The purification by a chitin column. Lane 1, molecular-weight markers ; Lane 2 chitin beads after 40h cleavage; Lane 3~7, elution after 40h cleavage. (B) The purification by a gel filtration column Superdex75. Lane 1, molecular-weight markers; Lane 2~12, fractions by gel filtration;. Lane 8~12 were used for the crystallization.

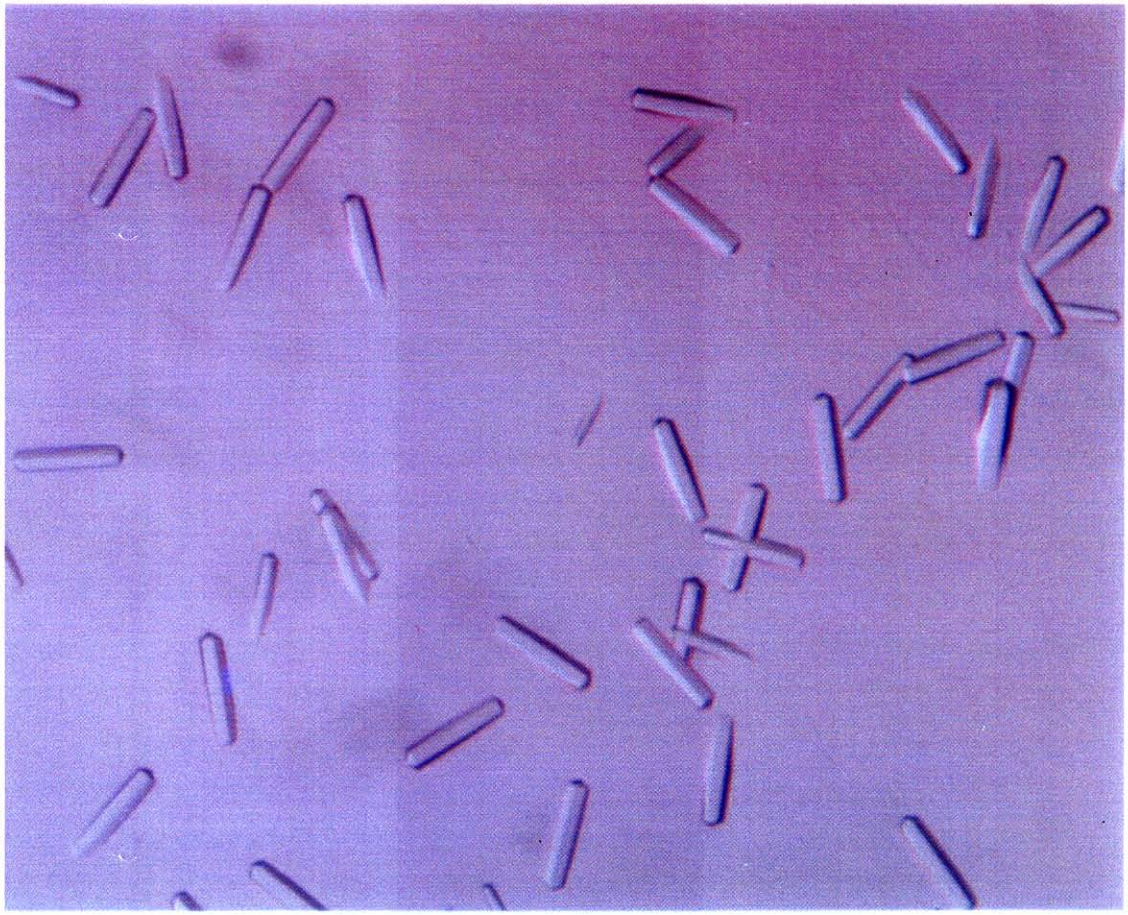
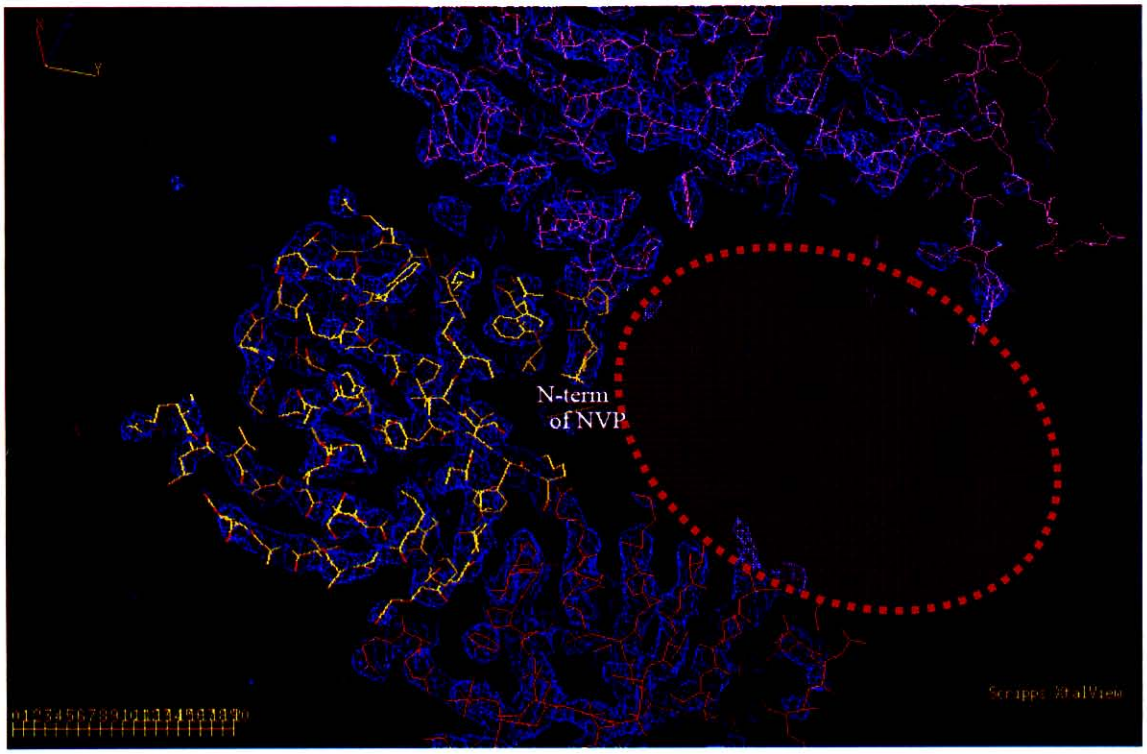


Figure 3-3. 3BC-C139A crystals. 3BC-C139A crystals were grown by the hanging-drop vapor diffusion method in a week. The size of the crystals is approximately 0.3 x 0.1 x 0.1 mm^3.

A



B

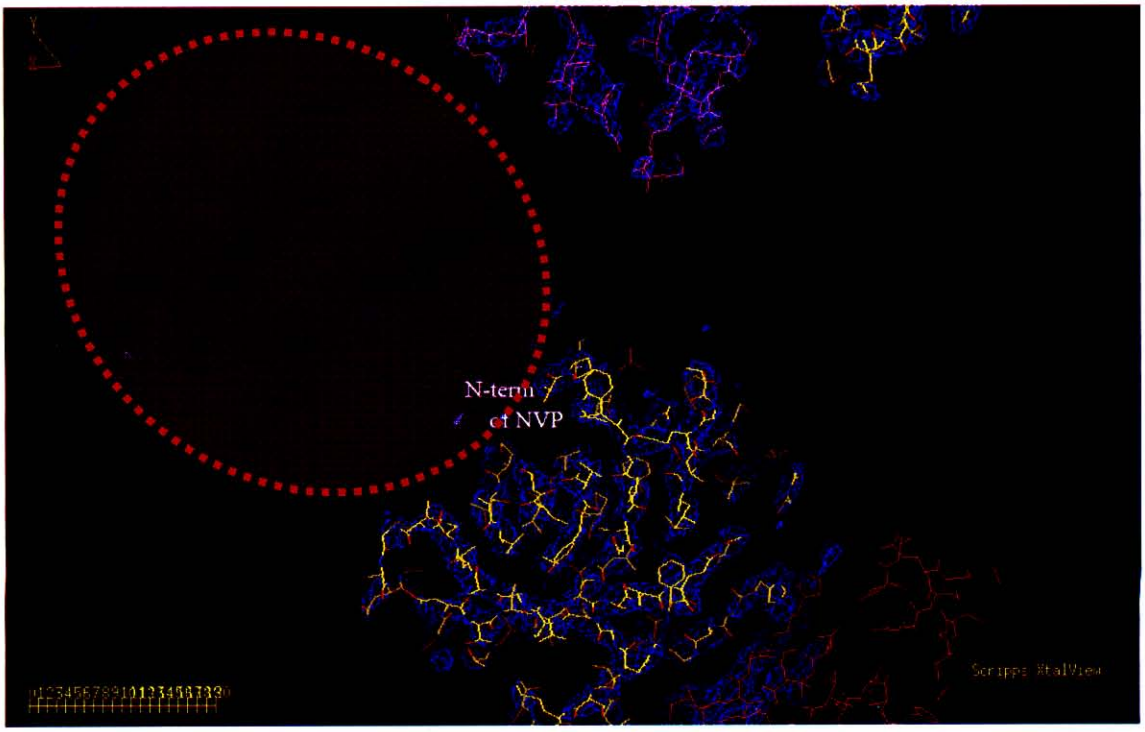


Figure 3-4. The electron density map for 3BC-C139A. The model of 3C region in 3BC-C139A crystal structure are shown with the electron density map (blue map). N-terminus of 3C region and the disordered space for 3B region are described as "N-terminal of NVP" and red-circle, respectively. (A) and (B) displayed them from the different direction.

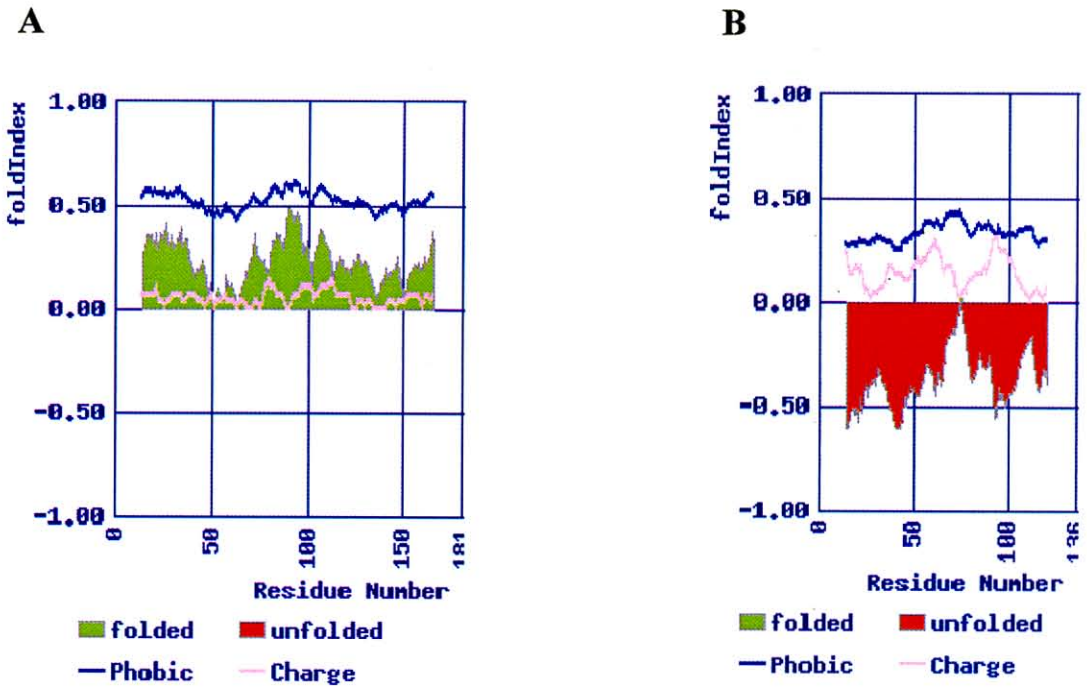


Figure 3-5. Folding indexes of NVP and 3B(VPg). They were calculated from the amino acid sequence (Vladimir et al., 2000). (A) Folding index of NVP. The detail values are: 181 residues, fold-ability 0.271, Charge 0.006, Phobic 0.513. The prediction showed that NVP had high folding ability. (B) Folding index of 3B(VPg). The detail values are: 136 residues, fold-ability -0.290, Charge 0.022, Phobic 0.317. The prediction showed that 3B(VPg) would be natively unfolded protein.

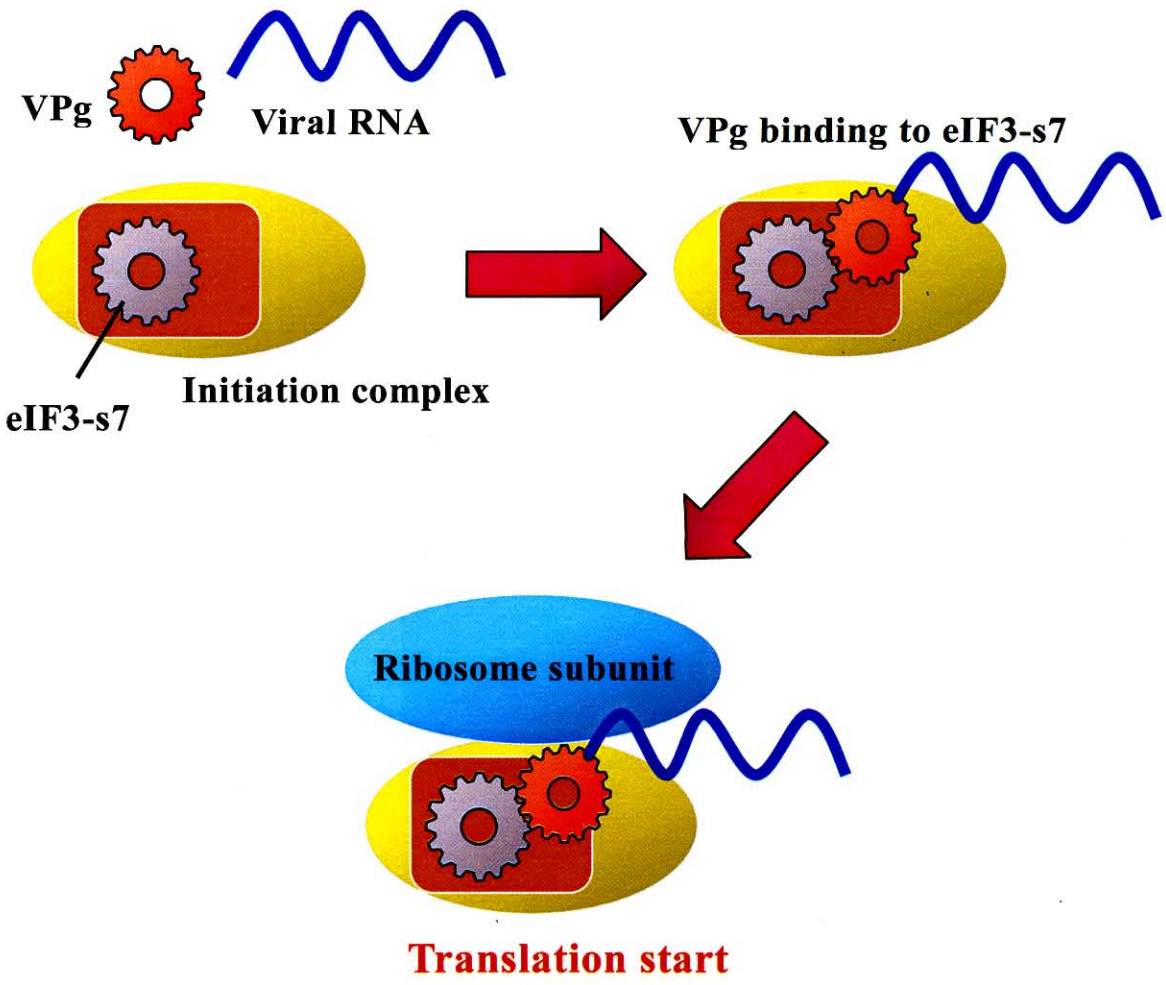


Figure 3-6. The now believed hypothesis of the action of VPg in human translation complex. It was known that VPg had sequence similarity to eIF1A and bound to the eIF3-subunit7. Therefore, the role described here may be the one of the functions of VPg.

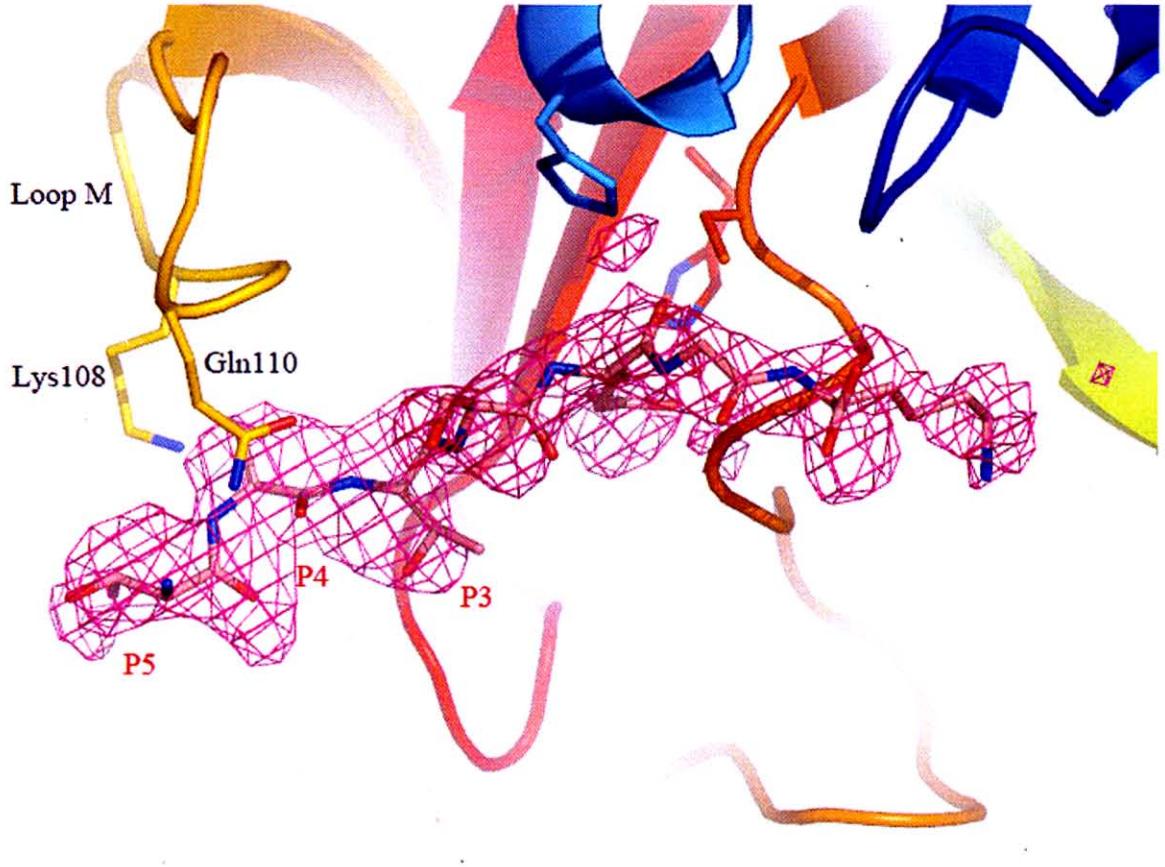


Figure 3-7. The NVP/substrate complex structure with the electron density map for the substrate peptide. The carbons of the substrate peptide and Lys108/Gln110 of NVP are colored pink and yellow, respectively. The $2|F_o|-|F_c|$ electron density map of the substrate is represented by colored mesh.

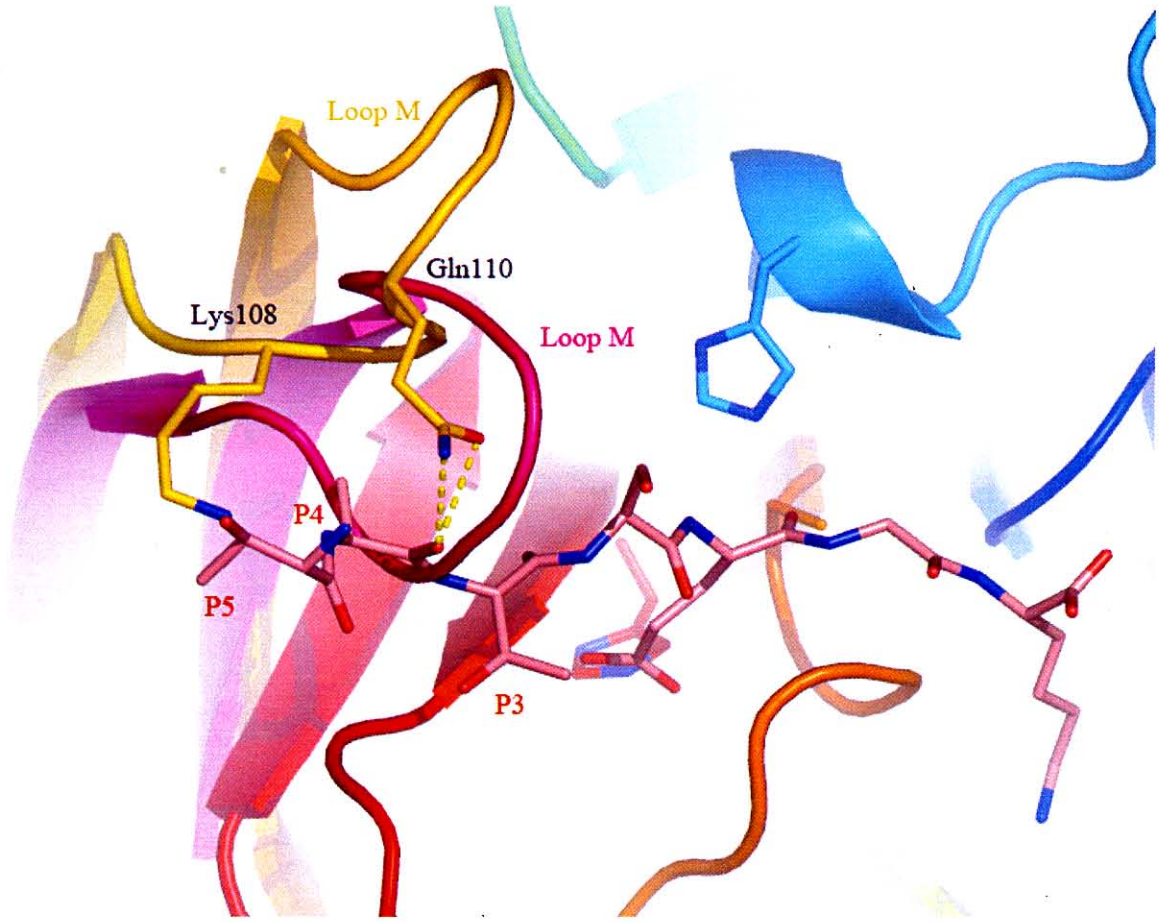


Figure 3-8. The NVP/substrate complex structure with superimposed native NVP. The carbons of the substrate peptide and Lys108/Gln110 of the NVP/substrate complex structure are colored pink and yellow, respectively. The region around loop M (residues Ser106 to Val114) of the complex structure, and superimposed native NVP without substrate are represented by yellow and purple, respectively.

Chapter 4: Protein preparation and crystallization of 3CD precursor, and modelling study of before (precursor) and after (successor) processing structures of 3CD

4.1 Materials and methods

4.1.1 Cloning, expression and purification of 3CD precursor

ChV ORF1-C139A which had Ala139 mutation at the Cys139 position of 3C protease was prepared as described in Chapter 3.1.1. 3CD-C139A gene encoding 3C-C139A and 3D polymerase region of ORF1 had been excised from ORF1-C139A. As the result of trial and error in the expression system of strain and vector including pQE30Xa/BL21(DE3), pQE30Xa/Rosetta-gami, and pUC18/BL21(DE3), 3CD-C139A gene was inserted into pTYB11 of the IMPACT system (New England BioLabs) as the N-terminal segment fused to the intein and chitin-binding domain unit (Figure 4-1). The recombinant vector pTYB11[3CD-C139A] was introduced in the *E. coli* strain ER2566 (New England BioLabs) and cultures were grown to an OD600 of 0.6-0.7 at LB broth supplemented with 100μg/ml ampicillin. The expression of recombinant 3CD-C139A, called as 3CD precursor, was induced by the addition of IPTG to a final concentration of 0.1 mM at 17°C overnight. The harvested *E. coli* was lysed by treatments with sonication in a lysis buffer containing 20 mM HEPES-NaOH, pH 8.2, 0.5 M NaCl, 1 mM EDTA. The cell homogenate was centrifuged at 20,000 x g for 30 min. The supernatant was applied to a chitin bead column (New England BioLabs) equilibrated with lysis buffer and was washed by a wash buffer, 20 mM HEPES-NaOH, pH 8.2, 1.0 M NaCl, 1 mM EDTA. On-column self-cleavage was carried out for 40 hours using a cleavage buffer of 20 mM HEPES-NaOH, pH 8.2, 0.5 M NaCl, 1 mM EDTA and 50 mM DTT. The fraction containing 3CD-C139A was collected and concentrated. Further purification was

performed by gel filtration over a HiLoad 16/60 Superdex (200 pg; Amersham Pharmacia Biotech) column equilibrated with 20 mM HEPES-NaOH, pH 8.0, 200 mM NaCl. The fractions containing 3CD-C139A were combined and concentrated to 10mg/ml. The purity of 3CD-C139A was confirmed by SDS-PAGE analysis (Figure 4-2).

4.1.2 Crystallization of 3CD precursor

The screening of the crystallization conditions were performed at 4°C and 20°C by sitting drop vapor diffusion method using Crystal Screen I & II, Grid Screen Ammonium Sulfate, Grid Screen MPD, MembFac, PEG/Ion Screen (Hampton Research), Wizard I & II (Emerald Biostructures), and Structure Screen I & II (Molecular Dimensions Ltd). The concentration of 3CD-C139A was screened from 5 to 20 mg/ml every 5 mg/ml. However a crystal was not observed in any conditions.

4.2 Results and discussion

It was known that expression of the region with 3D polymerase was difficult in *E. coli* expression system. As the result of trying various expression strains and vectors, it was successfully achieved using IMPACT system, pTYB11[3CD-C139A]/ER2566. Although over 1,000 crystallization conditions were tested, any crystal was not obtained. Then, I carried out the modelling study of 3CD precursor and successor.

Modelling study of before and after processing structures of 3CD

To understand the structural aspects of the 3C-3D polyprotein and the cleavage process, I constructed before and after processing models of NV 3C-3D. For the modeling study, I considered the NVP structure reported herein; a crystal structure of Norwalk virus RNA polymerase as the NV 3D polymerase ($3D^{pol}$) after processing model (Ng et al., 2004); the PV $3C^{pro}$ structure (Mosimann et al., 1997); and two PV RNA polymerase structures (PV $3D^{pol}$) for the before and after processing models (Hansen et al., 1997; Thompson and Peersen 2004). I also considered a study about PV 3C-3D interactions (Yang et al., 2004). The PV $3D^{pol}$ structure (PV $3D^{pol}$-I; Hansen et al. 1997) had been used to model the PV 3C-3D complex (Yang et al. 2004). However, a second PV $3D^{pol}$ structure has been characterised recently (PV $3D^{pol}$-II; Thompson and Peersen 2004). The conformations of the N-terminal regions of these two structures differ dramatically. Based on the 3C-3D interaction study (Yang et al., 2004), the PV $3D^{pol}$-I N-terminal and the PV $3C^{pro}$ C-terminal are physically close; whereas, because the PV $3D^{pol}$-II N-terminal moves to the opposite side of the fingers domain, the two termini cannot interact. Therefore, the PV $3D^{pol}$-I and PV $3D^{pol}$-II structures may be appropriate models for the before and after processing states, respectively. Superimposition of the NV $3D^{pol}$ and PV $3D^{pol}$-II crystal structures is a good fit; therefore, the NV $3D^{pol}$ structure (Ng et al., 2004) may be the processed form.

To create the before processing NV $3D^{pol}$ model, the NV $3D^{pol}$ structure (Ng et al., 2004) was first superimposed on the PV $3D^{pol}$-I crystal structure and then the conformation of NV $3D^{pol}$ was altered manually and residues inserted, as needed,

in the NV $3D^{pol}$ N-terminal region, to improve the superimposition of the two proteins. The structure of NV $3D^{pol}$ was then energy minimised to eliminate potential van der Waals overlaps. Then I adjusted the eight C-terminal NVP residues, which are disordered in the crystal structure, to mimic the conformation of the corresponding PV $3C^{pro}$ region. Finally, NVP and $3D^{pol}$ were docked (Figure 5-3) using the PV 3C-3D model as the reference (Yang et al., 2004). The primary interactions between the two protein components are: hydrophobic interactions among 3C-Pro78, 3D-Pro421 and 3D-Met430 and between 3C-Ala103 and 3D-Pro432; and hydrogen bonds made by Asn147, Asn148, Asn149, Asn150 of $3C^{pro}$ and Arg126, Asp129 of $3D^{pol}$ (Figure 4-4). The 3CD connecting peptide, which is the C-terminal region of $3C^{pro}$ and the N-terminal region of $3D^{pol}$, lies in the centre of a large tunnel found in $3D^{pol}$. The active site of $3D^{pol}$ is part of the tunnel and is near the finger, palm and thumb domains through which RNA passes Thus, given that in the before processing model the connecting 3CD peptide obstructs the tunnel, separation of $3C^{pro}$ and $3D^{pol}$ is required before viral RNA replication is possible. Belliot et al., (2003) found that the 3CD covalent complex is stable and suggested that this stability imparts a function other than proteolysis or RNA polyermization. The modelled 3C/3D cleavage site is distant from the NVP active site; therefore, *trans*-cleavage is possible. It was confirmed that 3CD-C139A was processed by NVP at the 3C-3D site (*trans*-cleavage). Additional structural and biochemical studies are required to elucidate the proteolytic processing mechanism for the NV polyprotein.

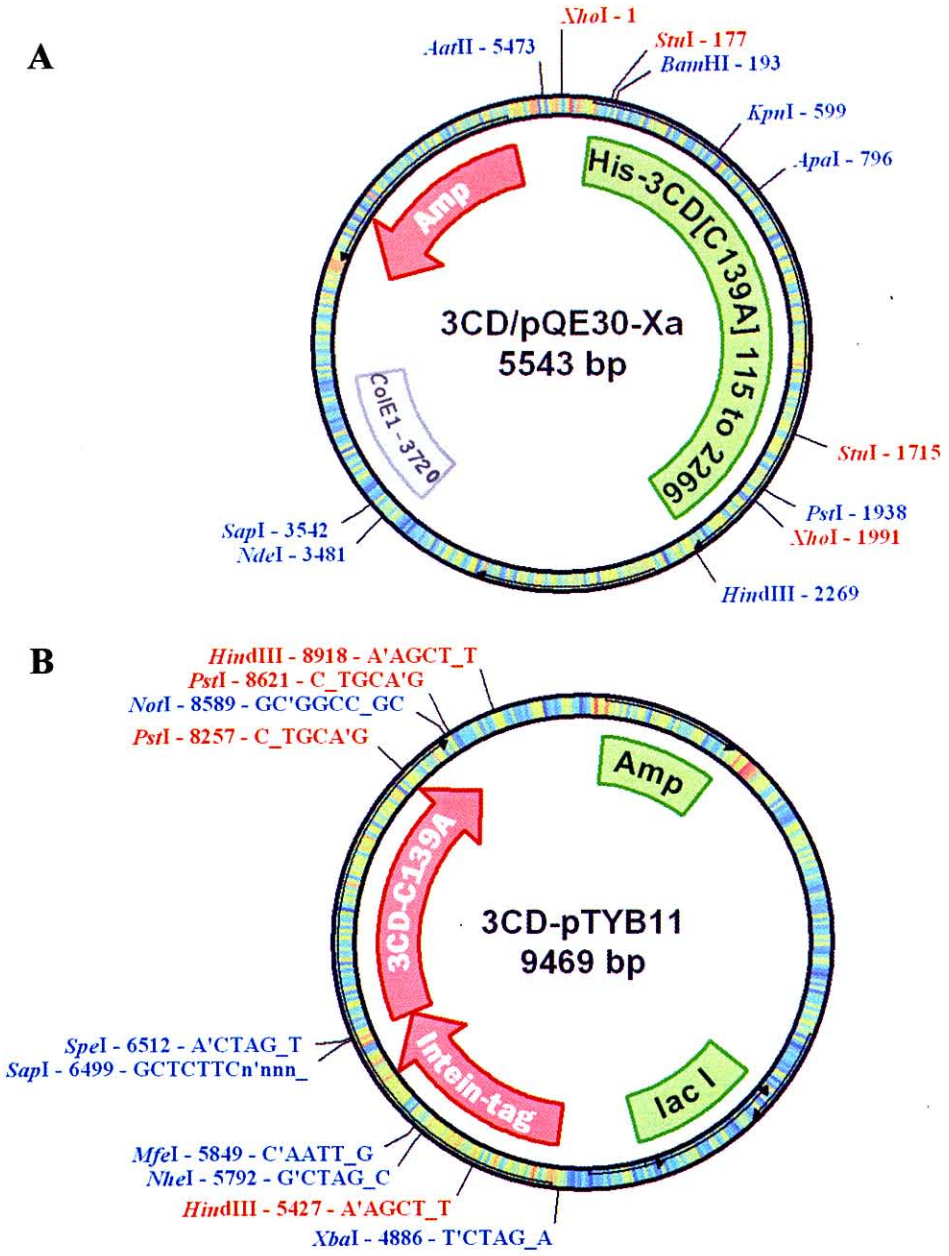


Figure 4-1. The constructs of 3CD-C139A with expression vector. (A) 3CD-C139A inserted the vector pQE30Xa has His-tag at the N-terminus of 3CD-C139A. (B) 3CD-C139A inserted pTYB11 is expressed as an intein-fusion protein in *E. coli*.

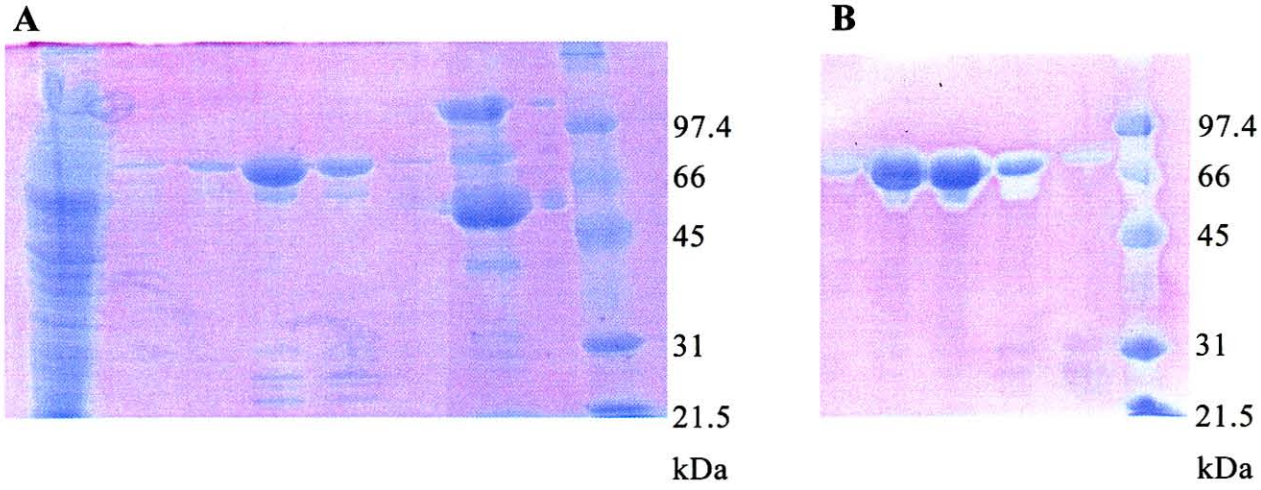


Figure 4-2. Coomassie-stained SDS-PAGE from purification of 3CD-C139A. (A) The purification by a chitin column. Lane 1, supernatant of cell lysis; Lane 2 wash; Lane 3~6, elution after 40h cleavage; Lane 7, chitin beads after 40h cleavage; Lane 8, molecular-weight markers. (B) The purification by a gel filtration column Superdex200. Lane 1~5, fractions by gel filtration; Lane 6, molecular-weight markers. Lane 2~4 were used for the crystallization.

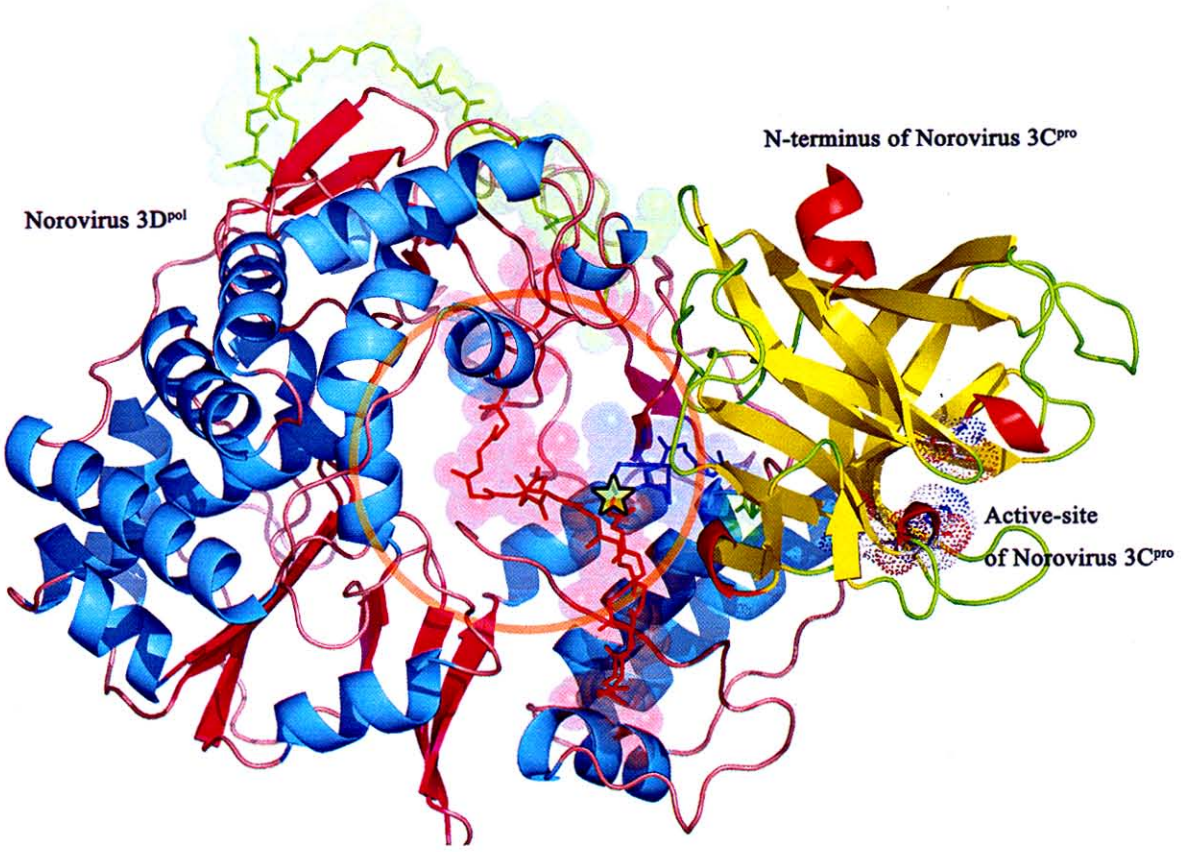


Figure 4-3. The NV 3CD before (precursor) and after (successor) processing models. NVP helices, strands, and loops are colored red, yellow, and light green, respectively. Those of $3D^{pol}$ are colored sky blue, purple, and pink, respectively. The atoms of Cys139, His30, and Glu54, are represented as stick figures and molecular surfaces. For the before processing model, C-terminal NVP residues are colored blue and N-terminal $3D^{pol}$ residues are colored red. These residues are part of the 3CD connecting peptide and are diagramed using stick figures and molecular surfaces. The $3D^{pol}$ N-terminal region of the after processing model, represented as stick figures and molecular surfaces, is colored light green. An orange circle surrounds the $3D^{pol}$ RNA cavity. The 3CD scissile bond is marked with a yellow star.

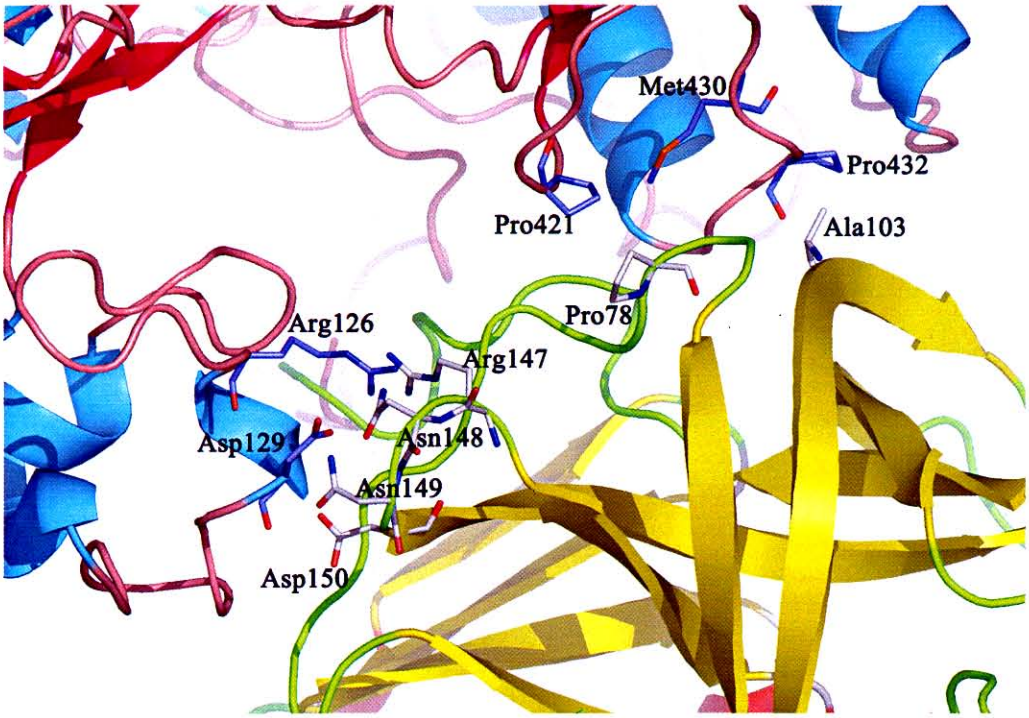


Figure 4-4. The interface between $3C^{pro}$ (NVP) and $3D^{pol}$ in the NV 3CD model. $3C^{pro}$ (NVP) helices, strands, and loops are colored red, yellow, and light green, respectively. $3D^{pol}$ helices, strands, and loops are colored sky blue, purple, and pink, respectively. Residues that interact across the interface are shown as stick figures and are slate-colored for $3D^{pol}$ residues and white for $3C^{pro}$ (NVP) residues.

Chapter 5: Conclusion

NVP is the key enzyme in norovirus ORF1 polyprotein processing. Because many functional proteins are encoded in ORF1, the inhibitions of their maturation crucially affect the survival of the viruses.

In this thesis, to begin with, I determined the first crystal structure of NVP and the complex structure with the substrate peptide, 3A-3B cleavage site. As the result, the catalytic dyad of NVP as a cysteine protease with chymotrypsin-like fold, the conformational change upon substrate binding at oxyanion hole, hydrogen bond network which contributes to the structural stabilization, and the details of the substrate binding/recognition mechanism were described. It was found newly that NVP has not a catalytic triad which is known in serine proteases with chymotrypsin/-like fold but a catalytic dyad at the active site, structural changes are caused by the substrate binding in the oxyanion hole and loop 106-114, Gln110 plays the role as a magic-hand in the substrate binding. I expect that drug development for NV-associated gastroenteritis and other diseases caused by the viruses which have the protease of this group will be facilitated by a global study of the available protease crystal structures, this structure will facilitate further molecular modelling of other members of the huge family of viral 3C/3C-like enzymes for which structural information is still not available.

Secondly, I carried out the construction of the expression system, purification, crystallization and structure determination/modelling of 3BC-C139A and 3CD-C139A, so that indicated that the cleavages at the 3B-3B site and 3C-3D would be performed as an intermolecular cleavage. The 3B-3C cleavage site in the precursor was far from the active site of NVP differently from PV $3C^{pro}$ and α-lytic protease, and the active site exposed to the solvent was

substrate-accessible. Therefore, inhibitors for NVP would be able to inhibit the protease activity in ORF1 polyprotein too. Modelling study of 3CD precursor suggested that the 3C-3D cleavage by NVP would be indispensable for maturation of 3D RNA polymerase, and that the maturation might be trigger on the 3D polymerase function. This prediction also, shows the importance of NVP and the maturation by NVP in the survival of NV,

I expect that further biochemical and structural biology studies will deepen the understanding to NVs, drug development and the effective utilization of the viruses including them as the vector for the gene therapy will advance.

References

Allaire M, Chernaia MM, Malcolm BA, James MN. (1994) Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**: 72-76

Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R. (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21**: 3213-3224

Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**: 1763-1767

Bashford D, Gerwert K. (1992) Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. *J Mol Biol* **224**: 473-486

Bazan JF, Fletterick RJ. (1988) Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: Structural and functional implications. *Proc. Natl. Acad. Sci. U S A* **85**: 7872-7876

Belliot G, Sosnovtsev SV, Mitra T, Hammer C, Garfield M, 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

Bergmann EM, Mosimann SC, Chernaia MM, Malcolm BA, James MN. (1997) The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J Virol* **71**: 2436-2448

Blair WS, Nguyen JH, Parsley TB, Semler BL. (1996) Mutations in the poliovirus 3CD proteinase S1-specificity pocket affect substrate recognition and RNA binding. *Virology* **218**: 1-13

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905-921

Cheah KC, Leong LE, Porter AG. (1990) Site-directed mutagenesis suggests close functional relationship between a human rhinovirus 3C cysteine protease and cellular trypsin-like serine proteases. *J Biol Chem* **265**: 7180-7187

Clarke IN, Lambden PR, Caul EO. (1998) Human enteric RNA viruses: caliciviruses and astroviruses. In *Topley & Wilson's microbiology and microbial infections*, Mahy BW, Collier L (eds) pp 511-535. London, United Kingdom: Arnold

Collaborative Computational Project N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760-763

Cowtan KD. (1996) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr D Biol Crystallogr* **52**: 43-48

Daughenbaugh KF, Fraser CS, Hershey JW, Hardy ME. (2003) The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. *EMBO J* **22**: 2852-2859

DeLano WL. (2002) The PyMOL Molecular Graphics System. . San Carlos, CA, USA: DeLano Scientific

Dougherty WG, Semler BL. (1999) Expression of Virus-Encoded Proteinases: Functional and Structural Similarities with Cellular Enzymes. *Microbiol. Rev.* **57**: 781-822

Duncan BS, Macke TJ, Olson AJ. (1995) Biomolecular visualization using AVS. *J Mol Graph* **13**: 271-82 299

Fortelle E. de La, Bricogne G, (1997) "Maximum-Likelihood Heavy-Atom Parameter Refinement for the Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods" *Methods in Enzymology* **276**

472-494. New York: Academic Press.

Fukushi S, Kojima S, Takai R, Hoshino FB, Oka T, Takeda N, Katayama K, Kageyama T. (2004) Poly(A)- and primer-independent RNA polymerase of Norovirus. *J Virol* **78**: 3889-3896

Glass PJ, White LJ, Ball JM, Leparc-Goffart I, Hardy ME, Estes MK. (2000) Norwalk virus open reading frame 3 encodes a minor structural protein. *J Virol* **74**: 6581-6591

Gosert R, Dollenmaier G, Weitz M. (1997) Identification of active-site residues in protease 3C of hepatitis A virus by site-directed mutagenesis. *J Virol* **71**: 3062-3068

Green KY, Ando T, Balayan MS, Berke T, Clarke IN, Estes MK, Matson DO, Nakata S, Neill JD, Studdert MJ, Thiel HJ. (2000) Taxonomy of the caliciviruses. *J Infect Dis* **181 Suppl 2**: S322-S330

Hamilton WC, Ibers JA. (1968) In *Hydrogen Bonding in Solids*, pp 167. New York and Amsterdam: Benjamin, Inc.

Hämmerle T, Hellen CU, Wimmer E. (1991) Site-directed mutagenesis of the putative catalytic triad of poliovirus 3C proteinase. *J Biol Chem* **266**: 5412-5416

Hansen JL, Long AM, Schultz SC. (1997) Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure* **5**: 1109-1122

Hardy ME, Crone TJ, Brower JE, Ettayebi K. (2002) Substrate specificity of the Norwalk virus 3C-like proteinase. *Virus Res* **89**: 29-39

Hutchinson EG, Thornton JM. (1996) PROMOTIF—a program to identify and analyze structural motif in proteins. *Protein Sci.* **5**: 212-220

Ivanoff LA, Towatari T, Ray J, Korant BD, Petteway SR. (1986) Expression and site-specific mutagenesis of the poliovirus 3C protease in Escherichia coli. *Proc Natl Acad Sci U S A* **83**: 5392-5396

Jiang X, Wang M, Graham DY, Estes MK. (1992) Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J Virol* **66**: 6527-6532

Kamphuis IG, Kalk KH, Swarte MB, Drenth J. (1984) Structure of papain refined at 1.65 A resolution. *J Mol Biol* **179**: 233-256

Kapikian AZ. (1994) Norwalk and Norwalk-like viruses. In *Viral infections of the gastrointestinal tract*, Kapikian AZ (ed) pp 471-518. New York: Marcel Dekker, Inc.

Katayama K, Shirato-Horikoshi H, Kojima S, Kageyama T, Oka T, Hoshino F,

Fukushi S, Shinohara M, Uchida K, Suzuki Y, Gojobori T, Takeda N. (2002) Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. *Virology* **299**: 225-239

Khan AR, Khazanovich-Bernstein N, Bergmann EM, James MN. (1999) Structural aspects of activation pathways of aspartic protease zymogens and viral 3C protease precursors. *Proc Natl Acad Sci U S A* **96**: 10968-10975

Kuyumcu-Martinez M, Belliot G, Sosnovtsev SV, Chang KO, Green KY, Lloyd RE. (2004) Calicivirus 3C-like proteinase inhibits cellular translation by cleavage of poly(A)-binding protein. *J Virol* **78**: 8172-8182

Lambden PR, Caul EO, Ashley CR, Clarke IN. (1993) Sequence and genome organization of a human small round-structured (Norwalk-like) virus. *Science* **259**: 516-519

Lawson MA, Semler BL. (1991) Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. *Proc Natl Acad Sci U S A* **88**: 9919-9923

Liu B, Clarke IN, Lambden PR. (1996) Polyprotein processing in Southampton virus: identification of 3C-like protease cleavage sites by in vitro mutagenesis. *J Virol* **70**: 2605-2610

Matthews DA, Smith WW, Ferre RA, Condon B, Budahazi G, Sisson W, Villafranca JE, Janson CA, McElroy HE, Gribskov CL. (1994) Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **77**: 761-771

McRee DE. (1999) XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125**: 156-165

Mosimann SC, Cherney MM, Sia S, Plotch S, James MN. (1997) Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J Mol Biol* **273**: 1032-1047

Murshudov GN. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255

Ng KK, Pendas-Franco N, Rojo J, Boga JA, Machin A, Alonso JM, Parra F. (2004) Crystal structure of Norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. *J Biol Chem* **279**: 16638-16645.

Petersen JF, Cherney MM, Liebig HD, Skern T, Kuechler E, James MN. (1999) The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J* **18**: 5463-5475

Polgár L. (1974) Mercaptide-imidazolium ion-pair: the reactive nucleophile in papain catalysis. *FEBS Lett* **47**: 15-18

Prasad BV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK. (1999) X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**: 287-290

Read RJ, Fujinaga M, Sielecki AR, James MN. (1983) Structure of the complex of Streptomyces griseus protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-A resolution. *Biochemistry* **22**: 4420-4433

Sárkány Z, Skern T, Polgár L. (2000) Characterization of the active site thiol group of rhinovirus 2A proteinase. *FEBS Lett* **481**: 289-292

Schechter I, Berger A. (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**: 157-162

Seah EL, Marshall JA, Wright PJ. (1999) Open reading frame 1 of the Norwalk-like virus Camberwell: completion of sequence and expression in mammalian cells. *J Virol* **73**: 10531-10535

Someya Y, Takeda N, Miyamura T. (2000) Complete nucleotide sequence of the Chiba virus genome and functional expression of the 3C-like protease in Escherichia coli. *Virology* **278**: 490-500

Someya Y, Takeda N, Miyamura T. (2002) Identification of active-site amino acid residues in the Chiba virus 3C-like protease. *J Virol* **76**: 5949-5958

Sprang S, Standing T, Fletterick RJ, Stroud RM, Finer-Moore J, Xuong N-H, Hamlin R, Rutter WJ, Craik CS. (1987) The Three-Dimensional Structure of Asn102 Mutant of Trypsin: Role of Asp102 in Serine Protease Catalysis. *Science* **237**: 905-909

Terwilliger TC. (1994) MAD phasing: treatment of dispersive differences as isomorphous replacement information. *Acta Crystallogr D Biol Crystallogr* **50**: 17-23

Terwilliger TC, Berendzen J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**: 849-861

Thompson AA, Peersen OB. (2004) Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J* **23**: 3462-3471

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876-4882

Utagawa ET, Takeda N, Inouye S, Kasuga K, Yamazaki S. (1994) 3'-terminal sequence of a small round structured virus (SRSV) in Japan. *Arch Virol* **135**: 185-192

Vinjé J, Green J, Lewis DC, Gallimore CI, Brown DW, Koopmans MP. (2000) Genetic polymorphism across regions of the three open reading frames of "Norwalk-like viruses". *Arch Virol* **145**: 223-241

Vladimir NU, Joel RG, Anthony LF (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**: 415-427

Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, Sun L, Mo L, Ye S, Pang H, Gao GF, Anand K, Bartlam M, Hilgenfeld R, Rao Z. (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* **100**: 13190-13195

Yang Y, Rijnbrand R, Watowich S, Lemon SM. (2004) Genetic evidence for an interaction between a picornaviral cis-acting RNA replication element and 3CD protein. *J Biol Chem* **279**: 12659-12667

Acknowledgements

The present work has been carried out under the direction of Dr. Takashi Kumasaka. I would like to express my sincere gratitude to my supervisor for the continuous and useful research guidance and encouragement.

I also would like to express my sincere gratitude to Professor Nobuo Tanaka for the support and advice on my research during the six years.

I would like to extend sincere thanks to Dr. Yuichi Someya (National Institute of Infectious Diseases) for the support on my research.

I wish to extend my thanks to all the people who are related to this research. And this work was supported in part by The National Project on Protein Structural and Functional Analyses (Priority Research Program, Protein 3000 Project) to N.T. and T.K.

I present the sincere condolences to all the people who departed the life in the direct/indirect cause by the norovirus.

Finally, I wish to thank my parents for their warm continuous supports.

Kentaro Nakamura