Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

doi:10.1016/j.jmb.2007.06.001 J. Mol. Biol. (2007) 371, 1060–1074

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

A Mechanistic View of Enzyme Inhibition and Peptide Hydrolysis in the Active Site of the SARS-CoV 3C-like Peptidase

Jiang Yin[1], Chunying Niu[1]†, Maia M. Cherney[1]†, Jianmin Zhang[2] Carly Huitema[3], Lindsay D. Eltis[3], John C. Vederas[2] and Michael N. G. James[1,4]*

[1]*Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, AB Canada T6G 2H7*

[2]*Department of Chemistry University of Alberta Edmonton, AB, Canada T6G 2G2*

[3]*Department of Microbiology and Immunology, University of British Columbia, Vancouver BC, Canada V6T 1Z3*

[4]*Alberta Synchrotron Institute, University of Alberta, Edmonton, AB Canada T6G 2E1*

**Corresponding author*

The 3C-like main peptidase $3CL^{pro}$ is a viral polyprotein processing enzyme essential for the viability of the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV). While it is generalized that $3CL^{pro}$ and the structurally related $3C^{pro}$ viral peptidases cleave their substrates *via* a mechanism similar to that underlying the peptide hydrolysis by chymotrypsin-like serine proteinases (CLSPs), some of the hypothesized key intermediates have not been structurally characterized. Here, we present three crystal structures of SARS $3CL^{pro}$ in complex with each of two members of a new class of peptide-based phthalhydrazide inhibitors. Both inhibitors form an unusual thiiranium ring with the nucleophilic sulfur atom of Cys145, trapping the enzyme's catalytic residues in configurations similar to the intermediate states proposed to exist during the hydrolysis of native substrates. Most significantly, our crystallographic data are consistent with a scenario in which a water molecule, possibly *via* indirect coordination from the carbonyl oxygen of Thr26, has initiated nucleophilic attack on the enzyme-bound inhibitor. Our data suggest that this structure resembles that of the proposed tetrahedral intermediate during the deacylation step of normal peptidyl cleavage.

Keywords: SARS; 3C-like proteinase; inhibitor design; tetrahedral intermediate; episulfide

Introduction

The 2002–2003 epidemic of Severe Acute Respiratory Syndrome (SARS) marked the worldwide debut of a newly recognized member of the *Coronaviridae* viral family, SARS-associated coronavirus (SARS-CoV).[1] This highly contagious yet previously uncharacterized virus likely originated from some animal coronavirus that fortuitously crossed the animal–human species barrier. The disease was eventually brought under control by strict enforcement of medical containment and tight screening of travelers, but not before the world had witnessed over 8000 cases including 774 SARS-related deaths. At present, there is no specific and effective treatment against SARS-CoV.

SARS-CoV is an enveloped, positive single-stranded RNA virus that replicates in the cytoplasm of the host cell. Similar to those of picornaviruses, coronaviral RNA genomes encode not only the capsid proteins required for virion assembly but also the non-structural proteins involved in viral RNA replication, including two large viral polyproteins,

† C.N. and M.M.C. contributed equally to this work.

Abbreviations used: Ac, acetyl; BCM7, β-casomorphin 7; CLSP, chymotrypsin-like serine peptidase; CMK, chloromethylketone; FMDV, Foot-and-Mouth Disease Virus; FMK, fluoromethylketone; HAV, Hepatitis A virus; MHV, Mouse Hepatitis Virus; mph, methylphthalhydrazide; Q^c, cycloglutamine; SARS, Severe Acute Respiratory Syndrome; T^b, (O-benzyloxy) threonine; TFLA, trifluoroacetal-leucyl-alanyl-*p*-trifluoromethylphenylanilide; TGEV, transmissible gastroenteritis coronavirus; TI, tetrahedral intermediate.

E-mail address of the corresponding author: michael.james@ualberta.ca

replicases pp1a and pp1ab. In these viruses, the polyproteins are processed by virally encoded peptidases *via* proteolysis. It is thought that the SARS 3C-like main proteinase performs 11 peptide cleavages in the viral polyproteins to generate individual viral proteins that subsequently assemble into functional complexes required for the replication of the viral RNA genome.[2–4]

The crystal structures of four coronaviral 3CL^{pro} enzymes have been reported: those of the transmissible gastroenteritis virus (TGEV), human coronavirus (H-CoV 229E), SARS-CoV, and the mouse hepatitis virus (MHV).[5–8] In all cases, the N-terminal domains I and II are each composed of a chymotrypsin-like β-barrel. The C-terminal domain III is mainly helical and mediates the homodimerization of coronaviral 3CL^{pro} in the crystal structures, an interaction believed to be important for its proteolytic activity *in trans*.[9] The catalytic residues are thought to be the His41-Cys145 dyad, not a triad as in the structurally similar CLSPs and cysteine peptidases including the picornaviral 3C^{pro} enzymes. This hypothesis seems to be supported by various coronaviral 3CL^{pro} crystal structures, which showed that His41 does not interact directly with any acidic residues. However, a water molecule has been consistently observed forming a hydrogen bond (2.6 to 2.9 Å) with $N^{\delta 1}$ of His41. Additionally, this water molecule is stabilized *via* hydrogen bonds formed with two other residues near the catalytic residues, i.e. His164 and Asp187. The role of this water molecule in the proteolytic reactions catalyzed by 3CL^{pro} enzymes has not been thoroughly investigated.

The hydrolysis of peptide substrates by viral 3C^{pro} or 3CL^{pro} peptidases is thought to occur in a manner analogous to the proteolysis by CLSPs. In the first half of the reaction or the acylation step, a histidine general base (His41 in SARS 3CL^{pro} numbering) assists the nucleophilic attack on the carbonyl carbon of the scissile bond by the S^{γ} atom of the cysteine nucleophile (Cys145 in SARS 3CL^{pro} numbering), leading to the formation of the first tetrahedral intermediate (TI1). The ensuing collapse of the TI1 and the departure of the C-terminal product give rise to a covalent thioester enzyme-substrate complex. In the second half of the catalysis, or the deacylation step, a solvent molecule, activated to a nucleophilic OH^{-} ion by His41, attacks the carbonyl carbon of the thioester, forming the second tetrahedral intermediate (TI2), which is followed by the release of the N-terminal product and the regeneration of the catalytic cysteine.

The 3CL^{pro} enzymes have been targeted for drug design against various members of the *Coronavirus* genus due to the extensive structural conservation in their active sites and the apparent absence of human homologues. A few non-covalent, competitive inhibitors[10] and peptidic, covalent inhibitors have been visualized in SARS 3CL^{pro} crystal structures.[7,8,11–13] The covalent inhibitors studied to date carry either a halomethyl ketone, an epoxide or a 1,4 Michael acceptor function as the reactive "warheads". Such functionalities permanently inactivate the viral peptidase *via* the formation of a non-hydrolysable covalent linkage to Cys145 as the result of the nucleophilic attack by its S^{γ} atom. As shown by X-ray crystallography, these inhibitors yield a complex analogous to the thioacyl intermediate formed during proteolysis. Although these crystal structures have provided valuable insights into how the residues in the active site of SARS 3CL^{pro} are organized after the inhibition reaction is completed, little is known about the mechanisms of inhibition at the molecular level or by analogy about the details of the acylation step of normal catalysis. Furthermore, none of these crystal structures (including those of the serine peptidases) offered any substantial structural insights into the deacylation of the acyl enzyme intermediate.

We recently designed a new class of covalent inhibitors against 3C^{pro} (HAV) and 3CL^{pro} (SARS-CoV) based on the phthalhydrazide function.[14,15] Initially designed as non-covalent peptidic inhibitors against the 3C^{pro} and 3CL^{pro} enzymes, we recently demonstrated the formation of an acylated enzyme in HAV 3C^{pro} with concomitant elimination of the phthalhydrazide group.[16] Two species of modified enzymes were observed by high resolution X-ray crystallography: (1) a thioacyl-like species similar to those reported in other covalently inhibited 3C^{pro} and 3CL^{pro} and (2) a species in which the inhibitor is linked to HAV 3C^{pro} by an episulfide cation ring.[16] In the latter complex, the S^{γ} of the catalytic Cys172 is directly attached to the carbonyl carbon, leading to the formation of an oxyanion in the active site. To our knowledge, this is the first structure to show a linkage between the nucleophilic sulfur atom and the sp^3 hybridized carbonyl carbon, in a 3C^{pro} or 3CL^{pro} peptidase. While this structure may be an analogue of the tetrahedral intermediate that occurs during acylation (TI1), the relevance of the episulfide ring complex to normal substrate hydrolysis is limited due to the lack of structures corresponding to other catalytic intermediates.

Here we extend our structural studies on the HAV 3C proteinase to another viral peptidase, SARS 3CL^{pro}. The crystal structures of SARS 3CL^{pro} bound to two phthalhydrazide-charged peptidyl inhibitors, acetyl-valyl-(O-benzyloxy)threonyl-leucyl-ketomethyl(cycloglutamine)-phthalhydrazide (inhibitor 1, Ac-VT^{b}LQ^{c}mph) and acetyl-leucyl-alanyl-alanyl-ketomethyl(cycloglutamine)-phthalhydrazide (inhibitor 2, Ac-LAAQ^{c}mph) are presented. The significance of these structures in relation to the catalytic steps during normal peptide hydrolysis by viral cysteine proteinases is discussed.

Results and Discussion

In vitro inhibition of SARS 3CL^{pro} by inhibitor 1

The inhibition of SARS 3CL^{pro} by inhibitor 1 was characterized using a FRET-based fluorogenic

method reported previously.[12] When added to a mixture of protease and substrate, inhibitor 1 inhibited 3CL^{pro} in a competitive fashion with a K_{ic} of 250(±50) nM. No kinetic evidence of covalent inhibition was observed when 10 uM 3CL^{pro} was pre-incubated with 100 uM inhibitor 1 (20 mM Bis–Tris (pH 7.0), 2 mM DTT, 37 °C) for 15 to 60 min, nor under solution conditions that mimicked the crystallization buffer.

Inhibitors 1 and 2 are covalently attached to SARS 3CL^{pro} in crystals

We examined the chemical nature of the reaction between the inhibitors and SARS 3CL^{pro} by subjecting SARS 3CL^{pro} crystals soaked in inhibitors 1 or 2 to electron spray ionization-mass spectrometry (ESI-MS). For SARS 3CL^{pro}-inhibitor 1 complex, a mass difference of 616 Da was observed before and after soaking inhibitor 1 into 3CL^{pro} crystals. This corresponds to a covalent adduct between 3CL^{pro} and inhibitor 1 without the phthalhydrazide moiety (616 Da). Similarly, the covalent linkage between inhibitor 2 and 3CL^{pro} was established by a mass difference of 466 Da (unreacted *versus* inactivated 3CL^{pro}), which is within the experimental error (the calculated mass of inhibitor 2 after the removal of the phthalhydrazide function is 467 Da). These results are in line with our crystallographic observations (see below).

The discrepancy between the findings of solution studies and those of crystallography may be explained by several hypotheses: (1) while fast, competitive inhibition predominates in an aqueous environment, slow, selective crystallization[17] of covalently inactivated enzyme molecules takes place in the course of a 24–72 h period; (2) the "local concentration effect" in crystal soaking experiments, plus the relatively high concentration of inhibitor used in crystallographic study may foster the formation of the covalent linkage between the inhibitors and the enzyme; and (3) the local conformation of the residues in the active site of crystallized SARS 3CL^{pro} may be restricted in some way that facilitates the covalent attachment of the inhibitors to the enzyme.

Structural overview

The three SARS 3CL^{pro}-inhibitor complexes are very similar in overall protein fold to each other and to the unliganded structure (PDB code 2A5A) as evidenced by their r.m.s.d. values (less than 0.3 Å calculated over all 306 C^{α} atoms Table 1)). Two regions show significant structural divergences: those comprising residues 45-TAEDM-49 and residues 277-NG-278. Residues N277 and G278 are located in a flexible surface loop in the C-terminal helical domain, where residues 276-MNGR-279 form a type II turn. The electron density for N277 was previously noted to be poorly defined and this seems to be attributable to the relatively higher inherent flexibility in this residue.[12] Although the loop containing N277 is situated near the dimerization interface of SARS 3CL^{pro} in the crystal, N277 itself is not directly involved in crystal packing. Residues 45-TAEDM-49 form the outer lid of the S2 pocket and their atoms have been associated with temperature factors significantly higher than those of the neighboring residues in various SARS 3CL^{pro} crystal structures. The significance of the flexibility in these residues in the recognition of the P2 residues of peptidyl substrates is discussed below.

Table 1. Alignment statistics of various complexes discussed in this study

	2A5I	XXXX	YYYY	ZZZZ
2A5A	0.28[a]	0.24	0.28	0.23
XXXX	0.27/1.16[b]		0.18/1.92	0.18/2.13
YYYY	0.23/0.93			0.16/0.30
ZZZZ	0.23/1.12			

[a] r.m.s.d. values (C^{α} positions) over all protein residues in Å.
[b] r.m.s.d. values for inhibitors alone.

The interactions between the peptidyl portions of the inhibitors and the specificity pockets of SARS 3CL^{pro}

The three refined structures show that the inhibitors are located in the substrate-binding cleft between the two chymotrypsin-like β-barrels. Excluding the interactions inside the S1 pocket, four and three hydrogen bonds were observed between the enzyme and the peptide backbone of inhibitors 1 and 2, respectively (Table 2). This structural organization has the general resemblance to that of the picornaviral 3C-peptidyl inhibitor complexes. In the latter, the peptidyl portions of the inhibitors form a parallel β-sheet with a strand (residues 141-ATYVHK-146 in HAV 3C) of the β-hairpin substructure, and an antiparallel β-sheet with a strand (residues 192-VAGGN-196 in HAV 3C) from the C-terminal β-barrel. In contrast, in the SARS 3CL^{pro} structures, the substrate/inhibitor peptide is antiparallel to both of the two interacting strands, residues 189-QTA-190 and residues 164-HMELP-168, an arrangement similar to that observed in the subtilisin family of serine proteinases.[18] A recent theoretical model also suggested that peptide substrates likely also form an antiparallel β-sheet in the active site cleft of caliciviral 3C-like proteinases.[19]

Upon substrate binding, small shifts were observed in the residues forming one of the two β-strands in α-lytic protease interacting with the peptidyl substrates in the active site.[20] The movement of this β-strand (residues 214-SGGNV-218 in α-lytic protease) causes the specificity pockets of the enzyme to collapse onto the substrate, a mechanism proposed to confer substrate specificity. A similar motion seems at play in the active sites of SARS 3CL^{pro} as well; residues 164-HMELP-168 of a structurally equivalent β-strand to that in α-lytic protease exhibit shifts of 0.3–0.4 Å from their positions in the unliganded 3CL^{pro} structure (PDB

Table 2. Interactions between the tetrapeptidyl inhibitors and SARS 3CL proteinase

	Thioacyl-like	Episulfide (inhibitor 1)	Episulfide (inhibitor 2)	Deacylating
P4[a]	21/2[b]	20/1	30/3	30/3
P3	16/4	12/3	8/4	8/4
P2	7/2	10/4	12	12
P1	45/4	42/1	44	45
H bonds	P4O:Gln189NE2 (2.7)[c]	P4O:Gln189NE2 (2.9)		
	P3N:Glu166O (2.8)	P3N:Glu166O (2.9)	P3N:Glu166O (3.0)	P3N:Glu166O (3.0)
	P3O:Glu166N (2.8)	P3O:Glu166N (2.9)	P3O:Glu166N (2.9)	P3O:Glu166N (2.9)
	P1N:Glu164O (3.5)	P1N:Glu164O (3.1)	P1N:Glu164O (3.0)	P1N:Glu164O (3.0)
	P1O:Gly143N (3.0)	P1O:Gly143N (3.1)	P1O:Gly143N (2.6)	P1O:Gly143N (2.9)
	P1O:Cys145N (3.1)	P1O:Cys145N (2.9)	P1O:Cys145N (3.2)	P1O:Cys145N (2.9)
	P1OE1:His163NE2 (2.6)	P1OE1:His163NE2 (2.6)	P1OE1:His163NE2 (2.8)	P1OE1:His163NE2 (2.7)
	P1NE2:Glu166OE2 (3.0)	P1NE2:Glu166OE2 (3.0)	P1NE2:Glu166OE2 (2.9)	P1NE2:Glu166OE2 (3.0)
	P1NE2:Phe140O (2.9)	P1NE2:Phe140O (3.5)	P1NE2:Phe140O (3.3)	P1NE2:Phe140O (3.4)

[a] Residue positions with respect to scissile bond as defined by Schechter and Berger.[34]
[b] Total number of van der Waals interactions/van der Waals interactions with solvent (less than or equal to 4 Å).
[c] Parentheses indicate the distances in Å.

code 2A5A) towards the inhibitor. Although inhibitors 1 and 2 consist of different peptidyl sequences, the size of these shifts is comparable among all three complexes described here. This suggests that the driving force of the motion in residues 164-HMELP-168 is the hydrogen bonding formed between these residues and the main chain atoms of the peptide substrates.

The S1 specificity pocket is primarily responsible for discriminating among peptide substrates. Virtually all native coronaviral $3CL^{pro}$ cleavage sites have a P1-Gln residue. With the exception of the side-chain atoms of Asn142, all atoms forming the S1 pocket show below average conformational changes upon inhibitor binding, suggesting that the screening of prospective substrates at the P1 position is largely based on structural complementarity. Indeed, we found that the side-chain atoms of the P1-cycloglutamine (P1-Gln^c) residues of both inhibitors 1 and 2 occupy almost identical positions in the S1 pocket. Two hydrogen bonds contribute to the specific recognition and stabilization of the P1-Gln^c in the S1 pocket: the side-chain carbonyl oxygen accepts a strong hydrogen bond (2.5 to 2.7 Å) from $N^{\varepsilon 2}$ of His163 and the amide nitrogen of the pyrrole ring donates a hydrogen bond (~3 Å) to a carboxylate oxygen of Glu166. The S1 pocket and the oxyanion hole in SARS $3CL^{pro}$ are essentially unchanged for the unliganded enzyme (PDB code 2A5A) and the three complex structures presented here.

In contrast to S1, the S2 pocket, although isostructural between the unliganded enzyme and the $3CL^{pro}$-inhibitor 2 complex, showed significant structural change upon the binding of inhibitor 1. Leucine occurs most frequently at the P2 position in the cognate cleavage sites for coronaviral $3CL^{pro}$ enzymes. Such preference can be best explained by the induced orderliness in the S2 pocket. The side-chain of Met165, the residue that forms the bottom of the S2 pocket, takes on two alternate conformations (with $\chi 1$ angles ~0° and 80°, respectively) in both the native enzyme and $3CL^{pro}$-inhibitor 2 complex. The presence of the P2-Leu of inhibitor 1 in the S2 pocket completely eliminates the first conformation ($\chi 1$ angle ~0°) of Met165. This is consistent with the structural changes in the S2 pocket observed in the structures of the SARS $3CL^{pro}$-azapeptide epoxide complexes[12]: in addition to the restriction in the side-chain placement of Met165, the S2 "lid", a surface loop consisting of residues 45-TAEDM-49, moves up towards the solvent to accommodate the bulkier P2-Phe of the epoxide inhibitor. These structural adjustments in the S2 pocket with regard to the incumbent P2 residue indicate that the recognition/binding of the P2 residue in the S2 pocket occurs in an "induced-fit" fashion.

Although the structural differences in the S4 pocket are minor between the complexes with inhibitors 1 and 2, the only hydrogen bond formed between $3CL^{pro}$ and the carbonyl oxygen of the P4 residues of the inhibitors is lost in the inhibitor 2 complex. Asn189, the residue that forms the mobile lid of the S4 pocket, normally makes a hydrogen bond (2.8–2.9 Å) with the P4 carbonyl oxygen *via* its $N^{\varepsilon 2}$ atom, which is indeed observed in the inhibitor 1 complexes. The distance between these two atoms is 3.9 Å in the inhibitor 2 complex because of a shift in the backbone atoms of the P4 and P3 residues of inhibitor 2. This is possibly due to the fact that inhibitor 2 contains a larger hydrophobic residue at the P4 position (Leu) than the P4-Val of inhibitor 1. Leucine does not fit well inside the normally shallow S4 pocket. Consequently, the acetyl modification of P4-Leu, which resembles the peptide bond between the P5 and P4 residues of a substrate, does not lie in the same volume as that of the (acetyl)-P4-Val of inhibitor 1. It is of interest to mention here that the S4 pocket of SARS $3CL^{pro}$ does occasionally exhibit inducible conformational change to accommodate larger residues such as the N-terminal benzene moiety in the epoxide inhibitor bound to SARS $3CL^{pro}$.[12] In that structure, only one of the two protomers in the asymmetric unit showed an enlarged S4 pocket, indicating that either the S4 pocket of SARS $3CL^{pro}$ is less susceptible to induced structural adjustment or the binding energy for a

fully buried leucine side-chain in inhibitor 2 is not sufficient to drive the corresponding conformational changes in the S4 pocket.

Covalent linkages between the inhibitors and SARS 3CL^{pro} catalytic cysteine Structures of SARS 3CL^{pro} in complex with inhibitor 1

The SARS 3CL^{pro}-inhibitor 1 complex obtained *via* the co-crystallization method shows predominantly a thioacyl-like connectivity between S^{γ} of Cys145 and the C^{i} atom of the inhibitor (Figure 1 for nomenclature, Table 3 for bond geometry and Figures 2, 3 and 4 for visualization). This structure mirrors those previously published of SARS 3CL^{pro} that has reacted with peptidyl chloromethyl ketone inhibitors,[7] with an epoxide inhibitor[12] and with Michael acceptor inhibitors.[8] In all of these structures, the C^{β}–S^{γ} bond of Cys145 shows a dramatic rotation (~90° in the χ1 angle) from its position in the unliganded enzyme. Similar conformational changes in the catalytic residues of the CLSPs are rarely reported in crystal structures. For example, chloromethyl ketone inhibitors usually form two covalent attachments each to one of the His57:Ser195 catalytic pair in their native conformation, possibly through a double displacement mechanism.[21] Interestingly, in the crystal structure of elastase complexed with the non-covalent inhibitor, trifluoroacetal-leucyl-alanyl-*p*-trifluoromethylphenylanilide (TFLA), the C^{β}–O^{γ} bond of Ser195 does undergo a similar rotation to avoid steric clashes with the trifluoro function of the inhibitor (PDB code 7EST).[22] This rotation in the side-chain of the catalytic serine/cysteine residue not only increases slightly the distance between $N^{\varepsilon 2}$ of the assisting general base (His57 of elastase/His41 of 3CL^{pro}) and the nucleophilic atom, but also places the nucleophilic O^{γ}/S^{γ} atom in a position less coplanar with the imidazole ring of the histidine residue. Consequently, the hydrogen bond between the catalytic His:Ser/Cys pair is weakened. When the usual analogy is drawn between these inhibitor-enzyme complexes and intermediate states of peptide hydrolysis, an attractive hypothesis arises stating that the demonstrated innate structural plasticity in the catalytic serine/cysteine residue effectively decreased the backward reaction from the acyl enzyme stage. However, that these conformational changes occur only in inactivated enzymes argues that the non-native conformation is unique to enzyme-inhibitor complexes and may not bear much relevance to the hydrolysis of peptide substrates. Indeed, the crystal structures of serine proteinase complexes showed that the formation of neither the acyl complex nor the tetrahedral intermediate requires the side-chain movement in the catalytic serine to a magnitude similar to that observed in the thioacyl-like complexes with viral 3C^{pro} or 3CL^{pro} enzymes. In SARS 3CL^{pro} and picornaviral 3C^{pro}, because of the geometric constraints imposed by the relatively greater rigidity of the oxyanion hole and the catalytic dyad, the additional methylene carbon C^{i} of inhibitor 1 makes it impossible for S^{γ} of Cys145 to maintain its native conformation in the thioacyl-like complex. It

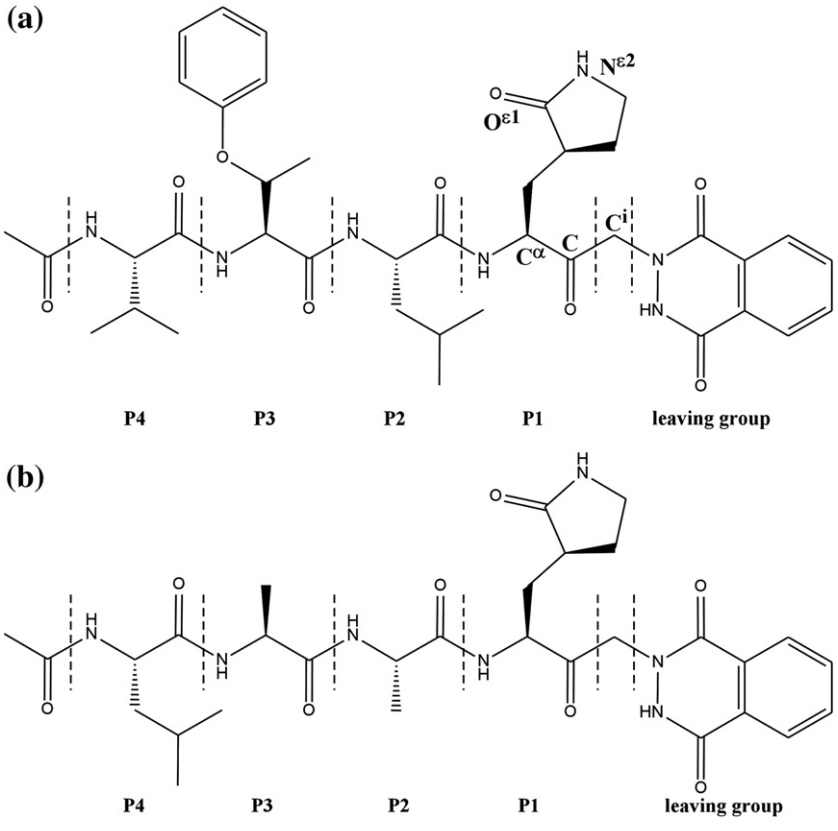

Figure 1. The chemical formulae for the inhibitors used in this study. Chiral centers are indicated where necessary.

Table 3. Bond geometry of the three covalent linkages

Bond lengths	Thioacyl-like	Episulfide	Deacylating
S^{γ}-C^{i}	1.81	1.83	1.81
S^{γ}-C		1.80	1.83
C=O	1.23		
C-O^{-}		1.43	1.45
C^{i}-C	1.51	1.53	1.53
C-C^{α}	1.50	1.52	1.55
C^{i}-OH			1.45
Cys145N···O	3.11	3.14	2.92
Gly143N···O	3.01	3.07	2.91
His41$N^{\varepsilon 2}$···S^{γ}	3.07		
His41$N^{\varepsilon 2}$···C^{i}		2.97	
Bond angles			
S^{γ}-C^{i}-C	108.9	64.0	65.4
C^{β}-S^{γ}-C^{i}	117.3	104.3	105.7
O-C-C^{i}	123.1	113.3	108.1
O-C-C^{α}	118.4	111.3	110.9
C^{i}-C-C^{α}	118.5	111.4	110.9
S^{γ}-C-O		123.4	113.0
S^{γ}-C-C^{i}		66.3	107.4
S^{γ}-C-C^{α}		125.0	121.0
C^{β}-S^{γ}-C		111.5	107.5
S^{γ}-C-C^{α}		121.0	106.4
O^{h}-C^{i}-C			109.7
Dihedrals[a]	**Thioacyl-like**	**Episulfide**	**Deacylating**
His41…S^{γ}	161.5	173.5	173.3
His41…C^{i}	164.6	155.0	151.0
His41…C	163.5	159.4	160.1
His41…O^{h}			144.8

C^{i}, C, C^{α}, O are atoms of the inhibitors, C^{β} and S^{γ} are atoms of Cys145, O^{h} is the oxygen atom of the hydrolytic water, all bond lengths are in Å, angles are in °. Covalent bonds are shown as dashes, whereas hydrogen bonds are shown in dotted lines.

[a] Dihedrals are defined as the torsion angles determined by $N^{\delta 1}$, $C^{\varepsilon 1}$, $N^{\varepsilon 2}$ of His41 and the fourth atom.

is noteworthy that two recently reported atomic resolution X-ray crystal structures of α-lytic protease showed a modified version of the aforementioned mechanism by which proteinases may promote the formation of the acyl enzyme over back-protonation of the catalytic Ser195.[20] The authors observed a translational shift in the position of O^{γ} of the catalytic Ser195 residue that weakens the His57:Ser195 hydrogen bond through a significant decrease in the calculated His57$N^{\varepsilon 2}$-H···O^{γ}Ser195 angle from the ideal linear value of 180° to 127°.

The structure of the $3CL^{pro}$-inhibitor 1 complex obtained *via* the soaking method indicates that there is an alternate bonding mode. This is clear in the $||F_o|-|F_c||,\alpha_c$ difference map when refinement was performed using the coordinates of the thioacyl-like complex. As shown in Figure 4(a), even when the occupancies of both inhibitor and Cys145 were set to 1.0, a significant (>4σ) positive peak was observed near the position of S^{γ} of the unliganded enzyme. Correspondingly, there is a small negative density peak near the S^{γ} atom at its inhibited, "thioacyl"-like position. A similar result was reported by us previously for the complexes of HAV $3C^{pro}$ with phthalhydrazide inhibitors.[16] In the SARS $3CL^{pro}$ complex, we have determined that an episulfide cation (thiiranium ring) best explains the electron density near S^{γ} of Cys145. Subsequent structural refinement has shown that both the negative and positive peaks in the $||F_o|-|F_c||,\alpha_c$ difference map disappear when we set the episulfide bonding mode to be the major species (>70%). The majority of the episulfide species is also confirmed in another crystal structure of the same complex acquired using a lower concentration of inhibitor 1 during soaking, in which the ratio between thiomethyl acyl and episulfide bonding modes remains roughly the same after refinement. This suggests that either linkage can be the "terminal" product of the inhibition reaction in this form of SARS $3CL^{pro}$ crystals. The distinct difference in the distribution of the two bonding modes between the structures of co-crystals and soaked crystals suggests the intermediacy of the episulfide linkage in solution. Additionally, the collapse of the episulfide cation into the thiomethyl ketone linkage may require conformational changes in the $3CL^{pro}$ enzyme, which are more attainable in aqueous solution than in crystals (Table 3).

In the episulfide linkage, the C^{i} atom of inhibitor 1 is 3.0 Å away from $N^{\varepsilon 2}$ of His41, a distance shorter than that observed in the thioacyl-like bonding mode (>4 Å). The shorter distance indicates a possible CH···N type hydrogen bond that would help to stabilize the episulfide cation structure. Two lines of argument support this proposal: (1) C^{i} is an electron-deficient atom due to the positive charge associated with the episulfide ring and the relatively larger electronegativity of S^{γ} and (2) $N^{\delta 1}$ of His41 donates a strong hydrogen bond to a nearby water (~2.7 Å), which is, in turn, hydrogen bonded to D187 (~2.8 Å to $O^{\delta 2}$), H164 (~2.9 Å to $N^{\delta 1}$) and the main chain nitrogen of His41. $N^{\varepsilon 2}$ of His41 is closer to S^{γ} of Cys145 than to C^{i} of inhibitor 1 in the thioacyl-like complex, whereas the reverse is true for the episulfide complex. This is in agreement with the proposed catalytic roles of His41: it acts as a general base to activate S^{γ} of Cys145 for its nucleophilic attack on the peptide bond, and then as a general acid to protonate the scissile peptide amide nitrogen for the release of the leaving group, the C-terminal product. The imidazole ring of His41 is roughly coplanar with both S^{γ} of Cys145 and C^{i}. Moreover, both the $N^{\varepsilon 2}$–S^{γ} and the $N^{\varepsilon 2}$–C^{i} distances in the two complexes are between 3.0 and 3.6 Å, suggesting that nucleophilic attack and leaving group protonation are closely coordinated.

Structure of SARS $3CL^{pro}$ in complex with inhibitor 2

The $|F_o|-|F_c|,\alpha_c$ difference map generated using an episulfide linkage as the refinement model clearly shows that the thioacyl-like bonding mode probably does not exist at a level significant enough to be taken into account. Surprisingly, a strong positive electron density peak appeared near the C^{i} atom of inhibitor 2 (Figure 4(b)). This peak extends into the electron density surrounding the imidazole ring of His41 on the corresponding $|2|F_o|-|F_c||,\alpha_c$

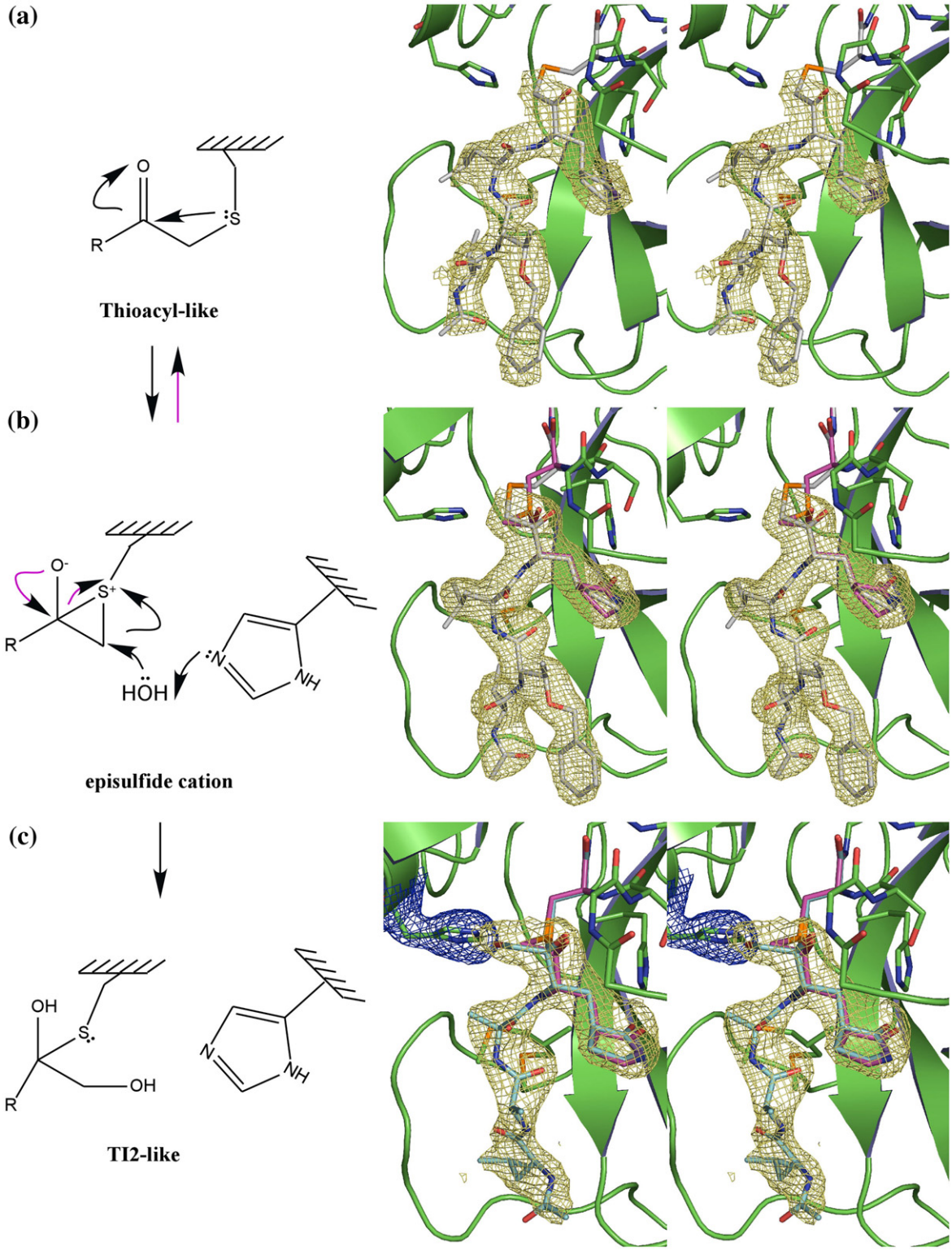

Figure 2. The electron densities of the inhibitors on the $|2|F_o|-|F_c||,\alpha_{calc}$ map. (a) The thioacyl form of SARS $3CL^{pro}$-inhibitor 1 complex; (b) the episulfide (major) and thioacyl mixed forms of SARS $3CL^{pro}$-inhibitor 1 complex; (c) the episulfide and the deacylating (major) mixed forms of SARS $3CL^{pro}$-inhibitor 2 complex. The protein is shown in cartoon with the carbon atoms colored green. The inhibitors are shown in sticks. Colors grey, purple and cyan are used to distinguish the carbon atoms in the thioacyl, episulfide and the deacylating complexes, respectively. In addition, the carbon atoms in portions of the inhibitors corresponding to the P4-P2 residues of the native substrates are colored grey. Cys145 is colored similarly as the bonding mode of the inhibitor to which it is linked. Oxygen and nitrogen atoms are colored red and blue, respectively.

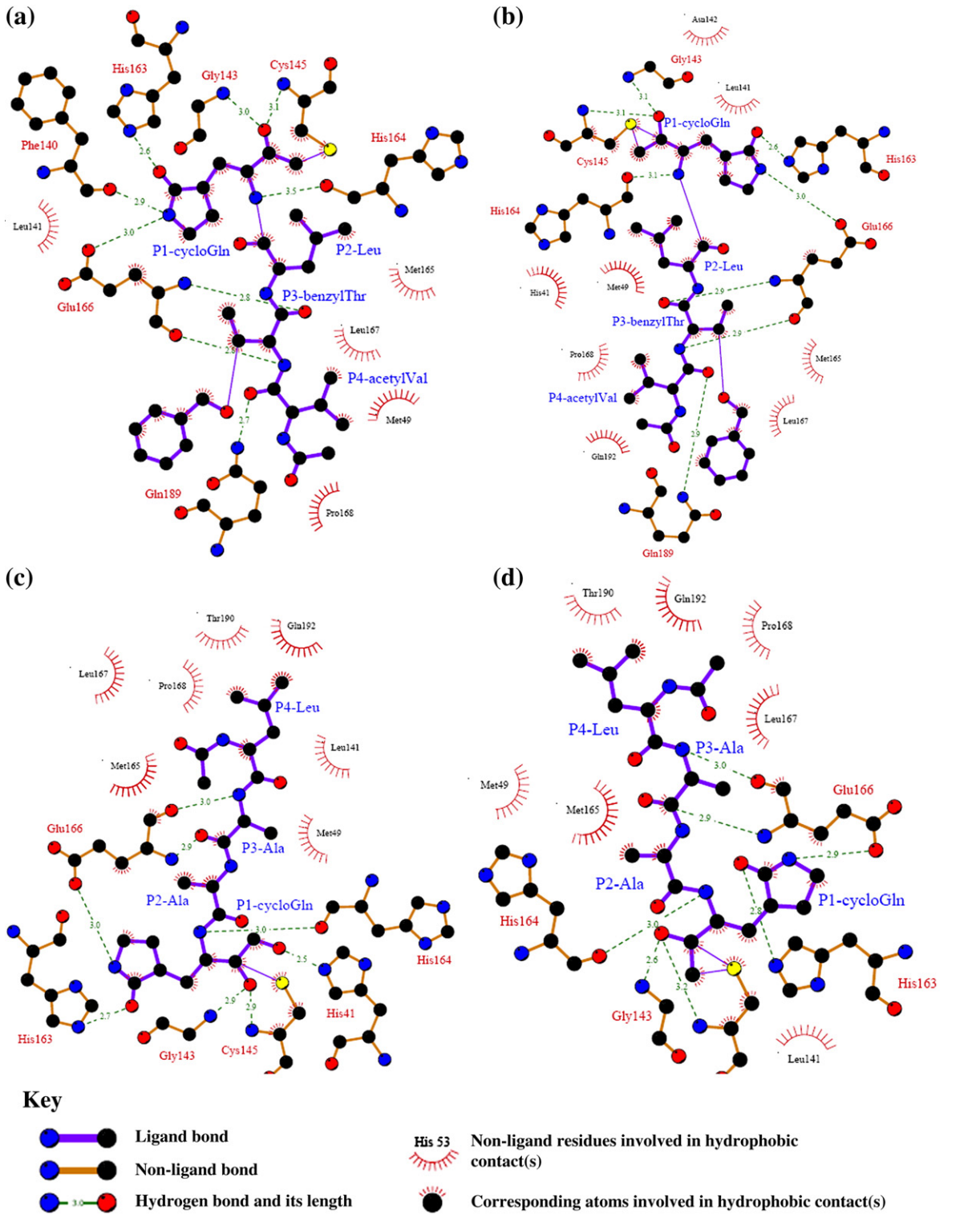

Figure 3. The interactions between the inhibitors and SARS 3CL^{pro}. (a) The thioacyl form of SARS 3CL^{pro}-inhibitor 1 complex; (b) the episulfide form of SARS 3CL^{pro}-inhibitor 1 complex; (c) the deacylating form of SARS 3CL^{pro}-inhibitor 2 complex; (d) the episulfide form of SARS 3CL^{pro}-inhibitor 2 complex.

difference map. The continuous electron density led to our hypothesis that this peak represents a possible position of a water molecule, which, with the assistance of His41, has attacked and opened the episulfide ring formed between S^{γ} of Cys145 and the methyl ketone function of inhibitor 2. Subsequent refinement of the modified structure in the active site of SARS 3CL^{pro} proved successful; the positive peak disappeared and no negative peaks were observed nearby, indicating that our model provides a reasonable explanation for the additional electron densities near C^{i} on the initial $|F_o|-|F_c|,\alpha_c$ map.

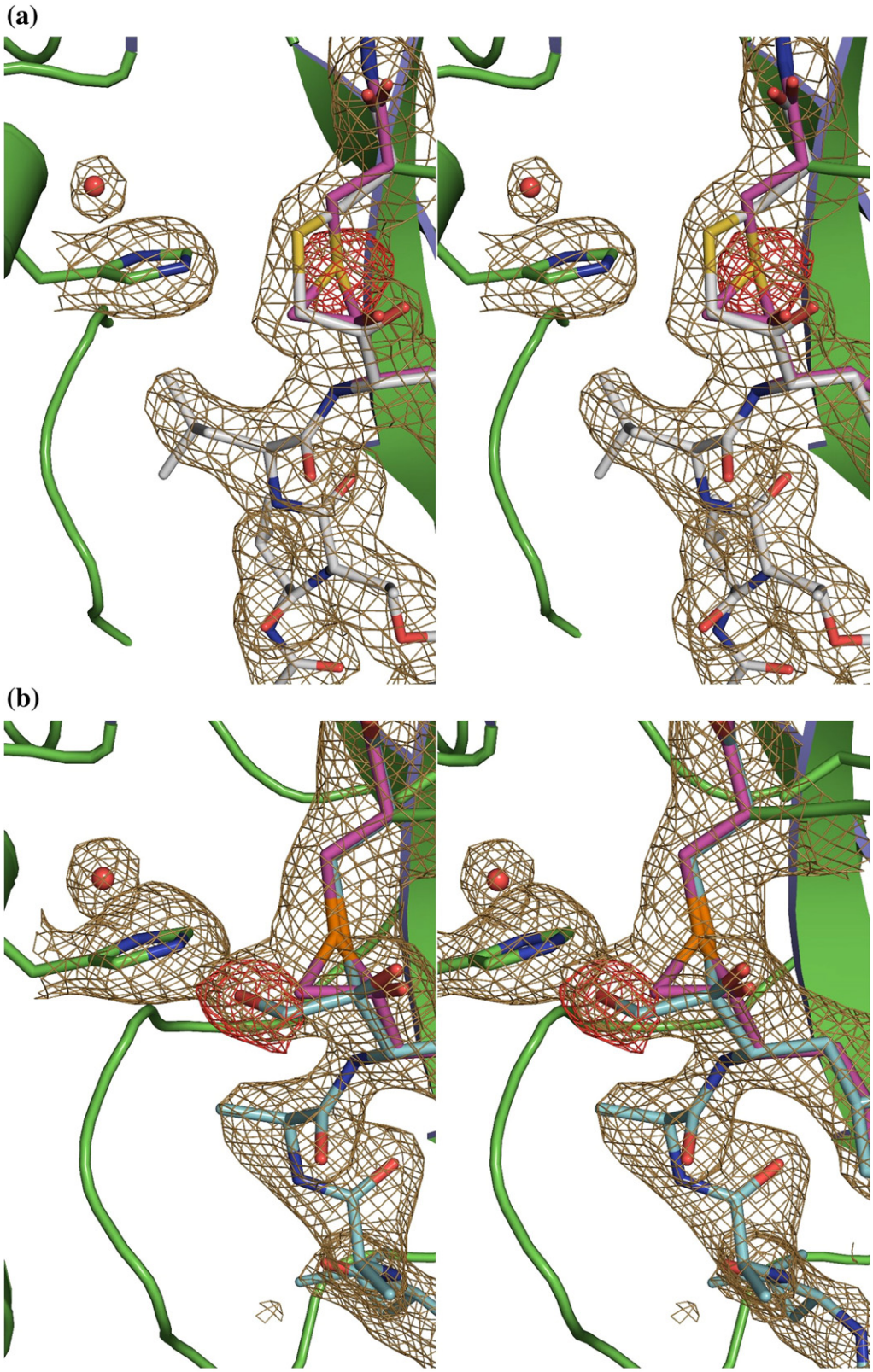

Figure 4. Positive electron density peaks on the $|F_o|-|F_c|,\alpha_{calc}$ difference maps for (a) the episulfide linkage of SARS 3CL^{pro}-inhibitor 1 complex and (b) the deacylating form of SARS 3CL^{pro}-inhibitor 2 complex overlaid with their corresponding refined structures. The protein and inhibitor structures are similarly colored as in Figure 2. Positive densities on the $2||F_o|-|F_c||,\alpha_{calc}$ difference maps (contoured at 1 σ) are shown in sandy color, whereas those on the $|F_o|-|F_c|,\alpha_{calc}$ difference maps (contoured at 3.5 σ) are colored red. The water molecule hydrogen bonded to H41 is shown as a red sphere.

The final structure shows a distance of ~2.5 Å between the hydroxyl oxygen on atom C^i and $N^{\varepsilon 2}$ of His41. This indicates that it is a strong hydrogen bond, considering an overall coordinate error of ~0.15 Å based on maximum likelihood.[23] Although solvent-directed nucleophilic attack during the deacylation stage of peptide hydrolysis by proteinases has been an unchallenged textbook concept, this is the first crystal structure that provides clear visual demonstration of such critical reaction step. Furthermore, the reaction pathway proposed here parallels the posited mechanism underlying the inhibition of chymotrypsin by chloromethyl ketone inhibitors.[24] In the latter hypothesis, an epoxy ether intermediate, resulting from the internal displacement of the chloride by the oxyanion, is subsequently attacked (thus opened) by $N^{\varepsilon 2}$ of the histidine residue of the catalytic triad to form a second covalent linkage between the inhibitor and the His: Ser catalytic pair of chymotrypsin.

Comparison of the three complex structures reported here suggests that the activated water molecule carrying out the attack on the acyl enzyme approaches from the C-terminal side of the scissile bond, i.e. the S′ side. The extrapolated angle of attack (O···C=O), using the coordinates of the hydroxyl oxygen attached to the C^i atom and those of the carbonyl function in the episulfide bonding mode (the *bona fide* pre-attack scenario), is ~115°. This angle is satisfyingly close to the calculated optimal angle (~109°) for nucleophilic addition onto carbonyl groups.[25] The slightly larger angle in our structure may reflect either minor coordinate error or the fact that the attacking hydroxyl oxygen is still ~2.3 Å from the P1 carbonyl carbon and separated by the C^i atom or both. During normal deacylation, the actual line of attack on the P1 carbonyl group by the hydrolytic water from its position observed in our "deacylating" complex structure will be inevitably affected by the bulky S^γ of Cys145. Therefore, the attacking angle of the hydrolytic water on P1 carbonyl of peptide substrates may change slightly from that estimated in the "deacylating" complex. More importantly, the establishment of C-side entrance of the hydrolytic water and its position relative to His41 and Cys145 as observed in the deacylating complex are in accordance with the hypothesized "minimal energy pathway" mechanism, in which the solvent activation by $N^{\varepsilon 2}$ (His41) and re-protonation of S^γ (Cys145) are tightly coordinated. The distance between $N^{\varepsilon 2}$ (His41) and S^γ (Cys145) in the deacylating complex is 3.7 Å, indicating that it is a better mimic of tetrahedral intermediate 2 (TI2) instead of TI2′.[26] This distance is longer than that between $N^{\varepsilon 2}$ (His57) and O^γ (Ser195) in CLSPs, reflecting the larger size of S^γ *versus* O^γ and the necessity to reduce possible steric hindrance between the hydrolytic water and the catalytic pair during its entrance into the active site from the S′ sites.

In the active sites of CLSPs, the carboxylate of Asp102 (PPE numbering) forms hydrogen bonds with $N^{\delta 1}$ of His57 and O^γ of Ser214 (the S1 residue), whose carbonyl oxygen interacts with the amide nitrogen of the P1 residue *via* a hydrogen bond (Figure 5(a)). Adding the hydrogen bond between the carboxyl group of Asp102 and main chain amide of His57, the hydrogen bonding network converges on the catalytic histidine residue and was thought to couple acyl enzyme hydrolysis and product release. An extensive hydrogen bonding network involving the catalytic histidine residue is also observed in the crystal structures of coronaviral $3CL^{pro}$. However, coronaviral 3CL proteinases do not have an acidic residue as the third member of the catalytic triad; Asp187 (SARS $3CL^{pro}$ numbering), the equivalent residue to Asp102 in CLSPs, does not interact directly with the catalytic His41. Instead, a solvent molecule is consistently observed in the vicinity of $N^{\delta 1}$ of His41 in various coronaviral $3CL^{pro}$ crystal structures.[5–8] In the deacylating complex, this water (WAT22) forms hydrogen bonds with $N^{\delta 1}$ of His41 (2.5 Å), $N^{\delta 1}$ of His164 (2.9 Å), $O^{\delta 2}$ of Asp187 (2.7 Å), and main-chain NH of His41 (3.3 Å) (Figure 5(b)). In addition, the carbonyl oxygen of His164 forms a hydrogen bond with the main-chain NH of the P1-Gln^c (3.0 Å). Therefore, the solvent-directed hydrogen bonding network in SARS $3CL^{pro}$ closely mimics the Ser214···Asp102··· His57 interactions in the CLSPs. This suggests that SARS $3CL^{pro}$ may employ a similar mechanism such as that proposed for the CLSPs underlying the concerted progression from acyl enzyme hydrolysis to product release. Interestingly, the residues in the picornaviral $3C^{pro}$ corresponding to Ser214 in the CLSPs are usually aliphatic (e.g. Val192 in HAV $3C^{pro}$). This precludes the possibility of linking peptide substrate binding/dissociation to the catalytic events concerning the scissile bond *via* the hydrogen-bonding network described above.

The origin of the hydrolytic water

It has been proposed that the solvent molecules hydrogen bonding to the carbonyl oxygen of residue 41 in the CLSPs (Thr41 in PPE) are involved in acyl enzyme hydrolysis (references in Perona *et al.*[27]). The distances between O of residue 41 and $N^{\varepsilon 2}$ of His57 are ~7.0 Å in various CLSP crystal structures, implying that there are likely two water molecules bridging these two residues. This is indeed observed in the structure of the PPE-β-casomorphin (BCM7) acyl enzyme; the main-chain carbonyl oxygen of Thr41 coordinates the hypothetical hydrolytic water (WAT317) through another solvent molecule (Figure 6(a)). This intermediary water (WAT318) forms two strong hydrogen bonds to the main-chain oxygen of Thr41 (2.7 Å) and to WAT317 (2.7 Å), which is at a distance of 2.9 Å from $N^{\varepsilon 2}$ of His57. We found a similar disposition of two water molecules in the active site of the thioacyl-like complex of the SARS $3CL^{pro}$. Thr26, the structurally equivalent residue of Thr41 in PPE, tightly binds a water molecule, WAT58 (~2.6 Å) that interacts with another solvent molecule, WAT81. The distance between WAT81

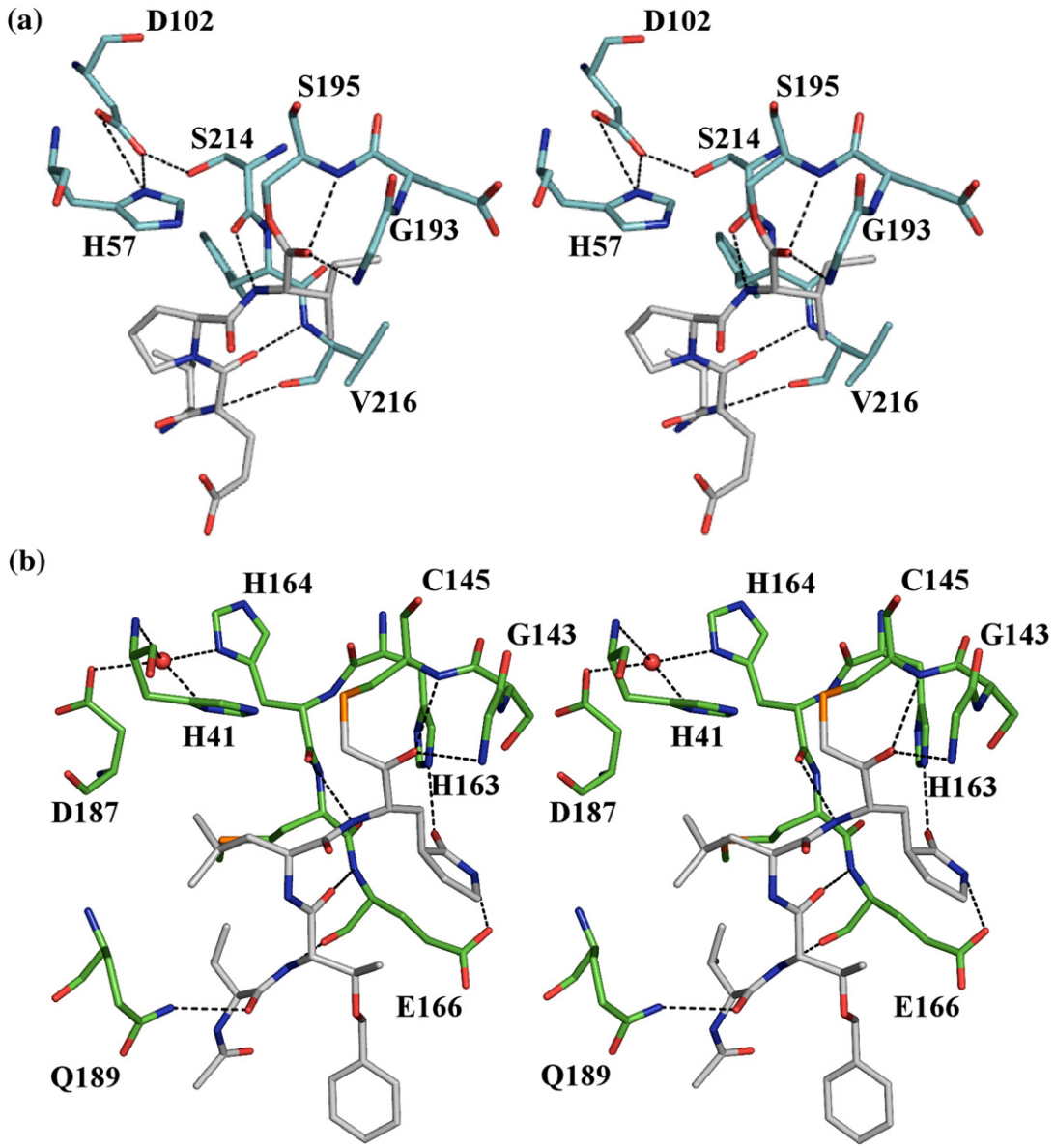

Figure 5. Similar hydrogen bonding network converged on the catalytic histidine residue in PPE and SARS 3CL^{pro}. (a) The PPE-BCM7 acyl complex and (b) the SARS 3CL^{pro}-inhibitor 1 thioacyl-like complex. The PPE and SARS 3CL^{pro} residues are distinguished by the color of the carbon atoms as cyan and green, respectively. The carbon atoms in the inhibitors are colored grey. Hydrogen bonds are shown in black broken lines. WAT22 is shown in red sphere.

and $N^{\varepsilon 2}$ of H41 in 3CL^{pro} is 4.8 Å, ruling out the possibility that this solvent molecule serves as the hydrolytic water. However, it is important to take into account that the observed position of WAT81 is a consequence of both the conformational change in the side-chain atoms of Cys145 and the intercalation of C^{i} between His41 and the carbonyl group of the P1 residue in the thioacyl-like complex (Figure 5(b)). Strikingly, although the amino acid identity of the structurally equivalent residues of Thr41 in PPE varies from virus to virus (Thr26 in SARS-CoV, Lys24 in HRV2, and Val28 in HAV), the relative position of their main-chain carbonyl oxygen atoms with regard to the catalytic residues is highly conserved. This reiterated structural resemblance among viral cysteine proteinases and the CLSPs seems to strengthen the notion that their catalytic mechanisms are likely to be quite similar as well.

Materials and Methods

Chemical syntheses of SARS 3CL^{pro} inhibitors

The syntheses of the phthalhydrazide inhibitors used in this study were done as described.[14]

Expression, purification, and crystallization of SARS 3CL^{pro}

SARS 3CL^{pro} protein was expressed in *Escherichia coli* in its cognate coding sequence with no additional protein tag

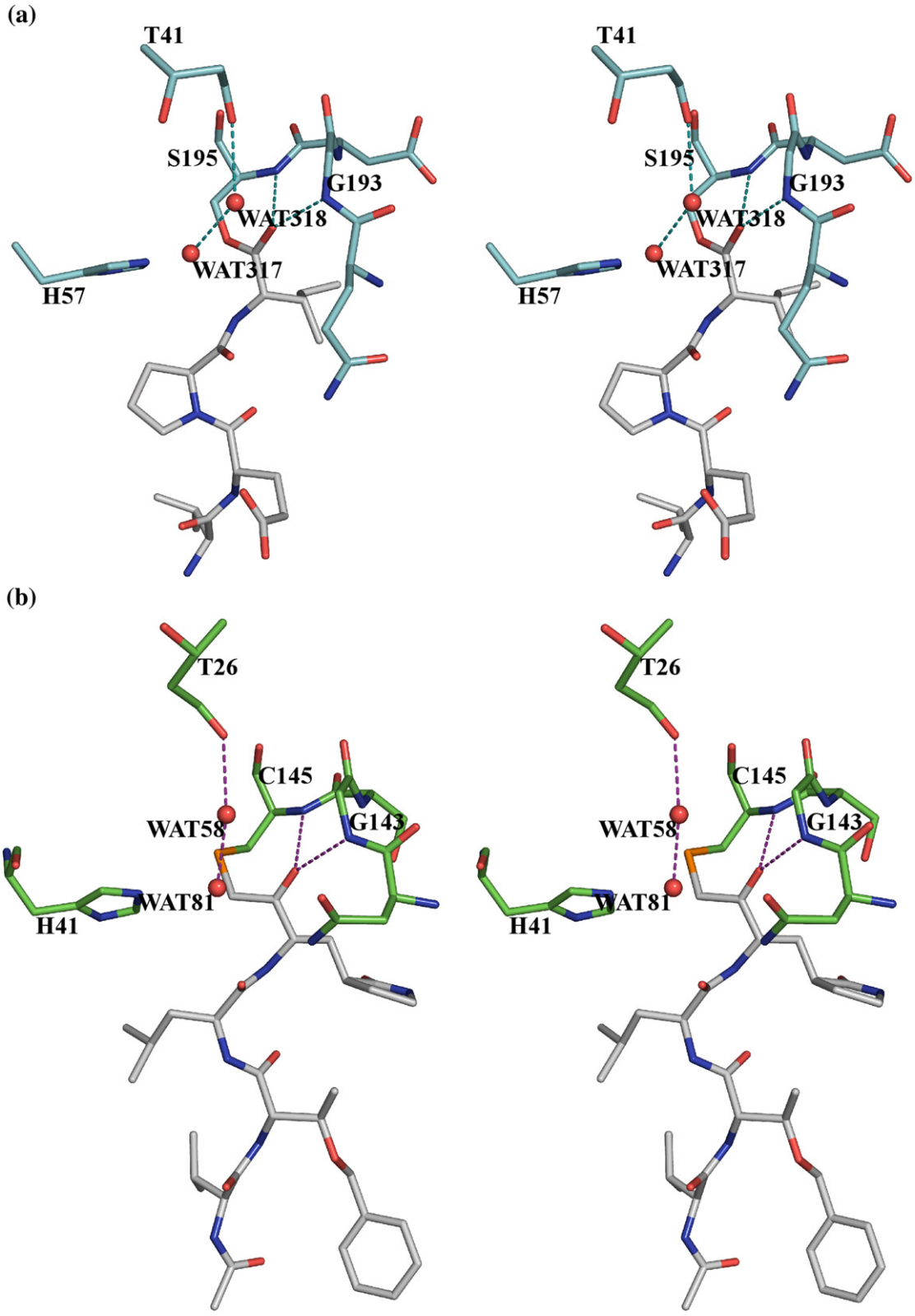

Figure 6. Similar solvent molecules near the catalytic residues in the active sites of PPE and SARS 3CL^{pro} complexes. (a) The PPE-BCM7 acyl complex and (b) the SARS 3CL^{pro}-inhibitor 1 thioacyl-like complex. Carbon atoms in PPE, BCM7, SARS 3CL^{pro}, and inhibitor 1 are distinguished by the colors yellow, cyan, green and grey, respectively. Hydrogen bonds in the PPE-BCM7 acyl complex and those in the SARS 3CL^{pro}-inhibitor 1 thioacyl-like complex are colored purple and dark teal, respectively. The interacting solvent molecules are shown as spheres and distinguished by the color of their hydrogen bonds.

attached. The purification procedure of SARS 3CL^{pro} essentially followed the one published.[12] The complexes were obtained either through co-crystallizing each of the inhibitors with SARS 3CL^{pro} or *via* soaking the inhibitors individually into the pre-grown SARS 3CL^{pro} crystals. The crystallization conditions were derived from a previously published condition that is conducive for the growth of crystals in the *C*2 space group (one molecule/asymmetric unit).[12] Co-crystals of SARS 3CL^{pro} with inhibitor 1 were obtained two to three days after mixing the protein with mother liquor containing the compound (2 to 5 mM). For soaking, crystals were incubated in drops containing either inhibitor 1 or 2 (2–5 mM) overnight before being flash cooled for data collection at the synchrotron X-ray source.

Mass spectrometry of SARS 3CL^{pro} inactivated by phthalhydrazide inhibitors

Pre-grown SARS 3CL^{pro} crystals were soaked in solutions containing either inhibitor 1 or 2 for ~18 h before being collected in a PCR tube. A large volume (v/v 1:200) of wash buffer (1 mM Tris pH 7.5) was used to wash the crystals four to five times. Samples were loaded onto a C4 ziptip (Millipore, MA, USA) and eluted with 0.1% formic acid. Electron spray ionization-mass spectrometric (ESI-MS) analyses were performed on eluted protein samples using a Waters (Micromass) Q-TOF Premier with an infusion rate of 0.5–1 ml/min.

X-ray data collection, processing and structure refinement

X-ray diffraction data were collected at beamline 12.3.1 of the Advanced Light Source (ALS). The data were indexed, integrated and scaled with the programs MOSFLM and SCALA.[28] The native SARS 3CL^{pro} structure in the *C*2 space group (PDB code 2A5A) was chosen as the search model for molecular replacement using the program MOLREP.[28] Refinement was done using cycles of REFMAC 5[28] and manual adjustment of the model using Xtalview.[29] The crystallographic statistics of data collection and model refinement are listed in Table 4.

Structural analysis and Figure generation

The quality of the final models was assessed using the program PROCHECK.[30] The interactions between the inhibitors and the enzyme were calculated using the program CONTACT.[28] The program LIGPLOT[31] was used

Table 4. Crystallographic statistics of data collection and structure refinement

PDB code	2Z3E	2Z3C	2Z3D
Inhibitor solution	1, co-crystallized	1, soaking	2, soaking
Space group	*C*2	*C*2	*C*2
a (Å)	108.26	108.56	108.27
b (Å)	82.24	81.46	81.84
c (Å)	53.49	53.36	53.42
α (°), β (°), γ (°)	90, 104.45, 90	90, 104.45, 90	90, 104.17, 90
No. molecules/asymmetric unit	1	1	1
V_m (Matthews' coefficient)/ %solvent content	3.41/63.9	3.38/63.6	3.39/63.7
Data collection		ALS Beamline 12.3.1	
Resolution range (Å)	15.47–2.32 (2.45–2.32)[a]	18.37–1.79 (1.89–1.79)	18.43–2.10 (2.21–2.10)
Wavelength (Å)	0.97946	0.97946	0.97946
Observations	45,213 (6602)	305,138 (33,941)	135,574 (19,673)
Unique reflections	18,207 (2662)	41,831 (6038)	23,907 (3584)
$I/\sum(I)$	8.9 (5.4)	11.6 (2.7)	9.7 (5.1)
Data completeness (%)	92.5 (92.8)	98.8 (97.6)	91.2 (93.8)
R_{merge}[b]	0.061 (0.186)	0.078 (0.643)	0.070 (0.328)
Refinement			
No. reflections used	16,658 (1090)	39,711 (2888)	22,259 (1716)
Resolution range (Å)	15.47–2.32 (2.37–2.32)	18.37–1.79 (1.84–1.79)	18.43–2.10 (2.16–2.10)
Free set size (%)	5.0	5.0	5.0
No. atoms (protein+ligand)	2652	2758	2666
No. waters	227	295	212
$R_{working}$[c] (%)	18.7 (23.4)	19.0 (29.8)	19.8 (25.4)
R_{free} (%)	23.1 (28.2)	23.9 (30.4)	24.8 (30.7)
Mean *B* value (Å^2)[d]	32.66/46.32/36.28	31.49/40.81/50.72	31.55/44.94/41.88
r.m.s.d. from ideal geometry			
Bond length (Å)	0.020	0.015	0.012
Bond angle (°)	1.956	1.700	1.481
Chirality	0.127	0.116	0.101
Ramachandran plot (%most favored/%disallowed)	90.3/0.0	90.3/0.0	91.0/0.0

[a] Parentheses indicate values for the highest resolution shell.
[b] $R_{merge} = \sum_h \sum_j | I_{hj} - \langle I_h \rangle | / \sum_h \sum_j I_{hj}$, where $\langle I_h \rangle$ is the weighted mean intensity of the symmetry-related reflections I_{hj}.
[c] $R_{working} = \sum_h | | F_o - | F_c | | / \sum_h F_o$, where $| F_o |$ and $| F_c |$ represent the observed and calculated structure factor amplitudes, respectively. R_{free} is $R_{working}$ calculated with the reference set.
[d] Average *B* factors of the complex/tetrapeptidyl inhibitor/solvent molecules.

to illustrate these interactions. Structures were aligned using the program ALIGN.[32] Figures containing structural models and electron density maps were generated by the program Pymol‡.

Kinetic assays of SARS 3CL^{pro} activity

The steady-state proteolytic activity of SARS 3CL^{pro} was determined using a fluorogenic peptide substrate Abz-SVTLQSGY(NO_2)R, where Abz is aminobenzoate and Y(NO_2) is nitrotyrosine. The standard assay was performed using 50 nM 3CL^{pro}, 20 mM Bis-Tris (pH 7.0), 2 mM DTT in 100 μl at 37(±0.1) °C in a small volume quartz cuvette. For competitive inhibition experiments, the inhibitor concentration was varied from 0.2 to 3 μM. Fluorescence data were empirically corrected for the inner filter effect.[33] Fluorescence was measured using a Cary Eclipse Fluorescence spectrophotometer (Varian Canada, Mississauga, Ontario, Canada) equipped with a circulating water-bath. The reaction was monitored using an excitation wavelength of 320 nm (2.5 nm bandpass) and an emission wavelength of 420 nm (5 nm bandpass). Initial velocities were determined from a least-squares analysis of the linear portion of the progress curves (at least 1 min) using Excel 2003 (Microsoft, Redmond, WA).

Protein Data Bank accession codes

The coordinates for the structures of SARS 3CL^{pro} in complex with inhibitor 1 and 2 have been deposited in RCSB Protein Data Bank§ and are available under accession codes, 2Z3C, 2Z3D, 2Z3E.

Acknowledgements

We thank Drs E. Bergmann, B. Biswal and R. Sankaranaryanar for insightful discussions, S. Khan for technical assistance and Dr J. Parrish and V. Jbanova at the Alberta Synchrotron Institute for assistance with data collection. Experimental data were collected at the SIBYLS beamline, 12.3.1 at the Advanced Light Source (ALS) at Lawrence Berkeley Laboratory, which is funded by NCI grant CA92584 and DOE contract DE-AC03-76SF00098. Data collection was supported by The Alberta Synchrotron Institute (ASI) synchrotron access program, which is supported by grants from the Alberta Science and Research Authority (ASRA) and the Alberta Heritage Foundation for Medical Research (AHFMR). This research project is supported by Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), AHFMR (including a fellowship to J.Y.) and Killam Trust (Izaak Walton Killam Memorial Postdoctoral Fellowship to J.Y.).

J.C.V. is Canada Research Chair in Bioorganic and Medicinal Chemistry. M.N.G.J. is Canada Research Chair in Protein Structure and Function.

‡ http://www.pymol.sourceforge.net/
§ www.pdb.org

References

1. Fouchier, R. A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J., van den Hoogen, B. G. *et al.* (2003). Aetiology: Koch's postulates fulfilled for SARS virus. *Nature*, **423**, 240.
2. Denison, M. R., Zoltick, P. W., Leibowitz, J. L., Pachuk, C. J. & Weiss, S. R. (1991). Identification of polypeptides encoded in open reading frame 1b of the putative polymerase gene of the murine coronavirus mouse hepatitis virus A59. *J. Virol.* **65**, 3076–3082.
3. Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L. *et al.* (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279**, 1637–1642.
4. Whiting, A. K. & Peticolas, W. L. (1994). Details of the acyl-enzyme intermediate and the oxyanion hole in serine protease catalysis. *Biochemistry*, **33**, 552–561.
5. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., 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.
6. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, **300**, 1763–1767.
7. Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W. *et al.* (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**, e324.
8. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z. *et al.* (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl Acad. Sci. USA*, **100**, 13190–13195.
9. Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P. *et al.* (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **280**, 31257–31266.
10. Lu, I. L., Mahindroo, N., Liang, P. H., Peng, Y. H., Kuo, C. J., Tsai, K. C. *et al.* (2006). Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.* **49**, 5154–5161.
11. Ghosh, A. K., Xi, K., Ratia, K., Santarsiero, B. D., Fu, W., Harcourt, B. H. *et al.* (2005). Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. *J. Med. Chem.* **48**, 6767–6771.
12. Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C. *et al.* (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.* **353**, 1137–1151.
13. Yang, S., Chen, S. J., Hsu, M. F., Wu, J. D., Tseng, C. T., Liu, Y. F. *et al.* (2006). Synthesis, crystal structure, structure-activity relationships, and antiviral activity of a potent SARS coronavirus 3CL protease inhibitor. *J. Med. Chem.* **49**, 4971–4980.
14. Jain, R. P., Pettersson, H. I., Zhang, J., Aull, K. D., Fortin, P. D., Huitema, C. *et al.* (2004). Synthesis and evaluation of keto-glutamine analogues as potent inhibitors of severe acute respiratory syndrome 3CLpro. *J. Med. Chem.* **47**, 6113–6116.
15. Ramtohul, Y. K., James, M. N. & Vederas, J. C. (2002). Synthesis and evaluation of keto-glutamine analogues as inhibitors of hepatitis A virus 3C proteinase. *J. Org. Chem.* **67**, 3169–3178.

16. Yin, J., Cherney, M. M., Bergmann, E. M., Zhang, J., Huitema, C., Pettersson, H. *et al.* (2006). An episulfide cation (thiiranium ring) trapped in the active site of HAV 3C proteinase inactivated by peptide-based ketone inhibitors. *J. Mol. Biol.* **361**, 673–686.
17. Yin, J., Bergmann, E. M., Cherney, M. M., Lall, M. S., Jain, R. P., Vederas, J. C. & James, M. N. (2005). Dual modes of modification of Hepatitis A Virus 3C protease by a serine-derived beta-lactone: selective crystallization and formation of a functional catalytic triad in the active site. *J. Mol. Biol.* **354**, 854–871.
18. McPhalen, C. A. & James, M. N. (1988). Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry*, **27**, 6582–6598.
19. Nakamura, K., Someya, Y., Kumasaka, T., Ueno, G., Yamamoto, M., Sato, T. *et al.* (2005). A norovirus protease structure provides insights into active and substrate binding site integrity. *J. Virol.* **79**, 13685–13693.
20. Fuhrmann, C. N., Daugherty, M. D. & Agard, D. A. (2006). Subangstrom crystallography reveals that short ionic hydrogen bonds, and not a His-Asp low-barrier hydrogen bond, stabilize the transition state in serine protease catalysis. *J. Am. Chem. Soc.* **128**, 9086–9102.
21. Kreutter, K., Steinmetz, A. C., Liang, T. C., Prorok, M., Abeles, R. H. & Ringe, D. (1994). Three-dimensional structure of chymotrypsin inactivated with (2S)-*N*-acetyl-L-alanyl-L-phenylalanyl alpha-chloroethane: implications for the mechanism of inactivation of serine proteases by chloroketones. *Biochemistry*, **33**, 13792–13800.
22. de la Sierra, I. L., Papamichael, E., Sakarellos, C., Dimicoli, J. L. & Prange, T. (1990). Interaction of the peptide CF3-Leu-Ala-NH-C6H4-CF3 (TFLA) with porcine pancreatic elastase. X-ray studies at 1.8 Å. *J. Mol. Recognit.* **3**, 36–44.
23. Pannu, N. S. & Read, R. J. (1996). Improved structure refinement through maximum likelihood. *Acta Crystallog. sect. A*, **52**, 659–668.
24. Prorok, M., Albeck, A., Foxman, B. M. & Abeles, R. H. (1994). Chloroketone hydrolysis by chymotrypsin and *N*-methylhistidyl-57-chymotrypsin: implications for the mechanism of chymotrypsin inactivation by chloroketones. *Biochemistry*, **33**, 9784–9790.
25. Bürgi, H. B., Dunitz, J. D. & Shefter, E. (1973). Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Am. Chem. Soc.* **95**, 5065–5067.
26. Topf, M. & Richards, W. G. (2004). Theoretical studies on the deacylation step of serine protease catalysis in the gas phase, in solution, and in elastase. *J. Am. Chem. Soc.* **126**, 14631–14641.
27. Perona, J. J., Craik, C. S. & Fletterick, R. J. (1993). Locating the catalytic water molecule in serine proteases. *Science*, **261**, 620–622.
28. Collaborative Computational Project, No. 4. (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
29. McRee, D. E. (1999). XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
31. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134.
32. Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallog.* **30**, 1160–1161.
33. Liu, Y., Kati, W., Chen, C. M., Tripathi, R., Molla, A. & Kohlbrenner, W. (1999). Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal. Biochem.* **267**, 331–335.
34. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.

Edited by R. Huber

(Received 28 March 2007; accepted 30 May 2007)
Available online 8 June 2007