Crystal Structures Reveal an Induced-fit Binding of a Substrate-like Aza-peptide Epoxide to SARS Coronavirus Main Peptidase

Ting-Wai Lee[1], Maia M. Cherney[1], Jie Liu[2], Karen Ellis James[3] James C. Powers[3], Lindsay D. Eltis[2] and Michael N. G. James[1]*

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

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

[3]*School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta GA 30332-0400, USA*

The SARS coronavirus main peptidase (SARS-CoV M^{pro}) plays an essential role in the life-cycle of the virus and is a primary target for the development of anti-SARS agents. Here, we report the crystal structure of M^{pro} at a resolution of 1.82 Å, in space group $P2_1$ at pH 6.0. In contrast to the previously reported structure of M^{pro} in the same space group at the same pH, the active sites and the S1 specificity pockets of both protomers in the structure of M^{pro} reported here are in the catalytically competent conformation, suggesting their conformational flexibility. We report two crystal structures of M^{pro} having an additional Ala at the N terminus of each protomer (M$^{pro}_{+A(-1)}$), both at a resolution of 2.00 Å, in space group $P4_32_12$: one unbound and one bound by a substrate-like aza-peptide epoxide (APE). In the unbound form, the active sites and the S1 specificity pockets of both protomers of M$^{pro}_{+A(-1)}$ are observed in a collapsed (catalytically incompetent) conformation; whereas they are in an open (catalytically competent) conformation in the APE-bound form. The observed conformational flexibility of the active sites and the S1 specificity pockets suggests that these parts of M^{pro} exist in dynamic equilibrium. The structural data further suggest that the binding of APE to M^{pro} follows an induced-fit model. The substrate likely also binds in an induced-fit manner in a process that may help drive the catalytic cycle.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: X-ray crystallography; SARS coronavirus main peptidase; aza-peptide epoxide; structure-based drug design; induced-fit binding

*Corresponding author

Introduction

Severe acute respiratory syndrome (SARS) is a highly transmissible, infectious and often fatal disease (World Health Organization–Severe acute respiratory syndrome†). Since its outbreak in 2002 and rapid spread throughout early 2003, efforts in the development of anti-SARS vaccines and drugs have taken on paramount importance.

SARS is caused by a coronavirus (CoV);[1–3] it is an enveloped, positive-sense single-stranded RNA virus. Anti-SARS therapeutics could target any one of several major steps in the viral life-cycle, such as virus–cell interactions, virus entry, intracellular viral replication, virus assembly and exit.[4] Extensive studies have been carried out on the proteins involved in these steps.[5] The intracellular replication of CoV is mediated by a replicase complex derived from two virally coded polyprotein precursors, pp1a (486 kDa) and pp1ab (790 kDa).[6,7] The formation of this replicase complex requires the extensive processing of the two polyproteins by two cysteine peptidases encoded within them; namely, the main peptidase (M^{pro}), also known as the 3C-like peptidase (3CL^{pro}) because of its similarity to the 3C peptidases of Picornaviridae,[8] and the accessory papain-like peptidase 2 (PL2^{pro}),[7] which cleaves at three sites in the N-proximal regions of the two polyproteins. By contrast, M^{pro} cleaves at 11 sites in the central and C-proximal regions of the two polyproteins, releasing key viral replication pro-

teins, such as an RNA polymerase and a helicase.[7] As an essential protein, M[pro] is an attractive molecular target for the development of anti-SARS drugs.

SARS-CoV M[pro] is catalytically active only in the homodimeric form; each protomer has a molecular mass of 33.8 kDa.[9–14] Considerable efforts have been directed to the X-ray structural studies of M[pro], resulting in the availability of many crystal structures of M[pro] and its variants over a pH range of 5.9–9.0 in a variety of space groups. All of these structures show that the two protomers of M[pro] are oriented almost perpendicular to each other. The N-terminal residues 1–7 of each protomer constitute the N-finger, of which Arg4 was shown to be mandatory for the dimerization and the exhibition of the catalytic activity of M[pro].[14] Beyond the N-finger, each protomer consists of three domains. Domain I (residues 8–101) and domain II (residues 102–184) comprise a two-β-barrel fold similar to that of the chymotrypsin-type serine peptidases. Domain III (residues 201–300) has five α-helices and is connected to domain II by a long loop (residues 185–200). Each protomer has its own substrate-binding region situated in the cleft between domains I and II. Recent mutagenesis studies have confirmed that, similar to the main peptidases from human coronavirus strain 229E,[15,16] and porcine transmissible gastroenteritis coronavirus,[17] SARS-CoV M[pro] is a cysteine peptidase with a Cys-His catalytic dyad at the active site.[16,18] As suggested by the structure-based sequence alignment of the main peptidases (including their flanking residues in the polyproteins) from SARS-CoV and other coronaviruses,[16] and confirmed by *in vitro* studies,[7,9] M[pro] cleaves preferentially at a consensus sequence for the P4 to P1′ residues of substrates (the nomenclature is based on that used by Schechter & Berger[19] with the arrow indicating the cleavage site): (amino acid with a small side-chain)-(any amino acid)-Leu-Gln↓(Ala, Ser, Gly).

A number of M[pro] inhibitors have been proposed using structure-based discovery[20,21] and experimental screening.[22–24] The crystal structures of M[pro] inhibited by some peptidomimetic compounds have been determined.[25–27] Recently, virtual screening followed by experimental evaluations have identi-

fied the old drug cinanserin as a strong inhibitor of the replication of SARS-CoV; cinanserin likely targets M[pro].[28]

We have reported on the kinetic and structural characterization of the inhibition of SARS-CoV M[pro] by an aza-peptide epoxide (APE, Figure 1).[29] APEs were synthesized as a new class of inhibitors apparently specific for clan CD cysteine peptidases,[30,31] including the legumains,[32] and the caspases.[33] Each APE has an aza-peptide component, with an epoxide moiety attached to the carbonyl group of the P1 residue. The side-chain of the P1 residue predominantly determines the target-peptidase specificity of an APE. The substituent on the epoxide C2 atom also allows some tuning of both the inhibitory activity and specificity of APE towards a particular target peptidase. In the APE, the C[α] atom of the P1 residue is replaced by a nitrogen atom, yielding an aza-amino acid residue. This introduces trigonal planar geometry to the α-atom of the P1 residue and reduces the electrophilicity of the carbonyl C atom of the P1 residue, thereby making the carbonyl group of the P1 residue resistant to nucleophilic attack.[34] It has been proposed that APEs inhibit their target peptidases irreversibly by a mechanism in which the catalytic Cys S[γ] atom nucleophilically attacks one of the two epoxide carbon atoms (C2 or C3) of APE.[30,32,33] This results in the opening of the conformationally strained epoxide ring, with the formation of a covalent bond between the Cys S[γ] atom and the attacked APE atom. An APE containing Gln at P1 inhibits M[pro] with a $k_{inact}/K_i = 1900(\pm 400)$ M^{-1} s^{-1}. In this reaction, the catalytic Cys145 S[γ] atom attacks the epoxide C3 atom of the APE.[29]

Results and Discussion

Structure determination

The parameters and statistics derived from data processing and structure refinement are summarized in Table 1. SARS-CoV M[pro] crystallizes in space group $P2_1$ in conditions slightly different from
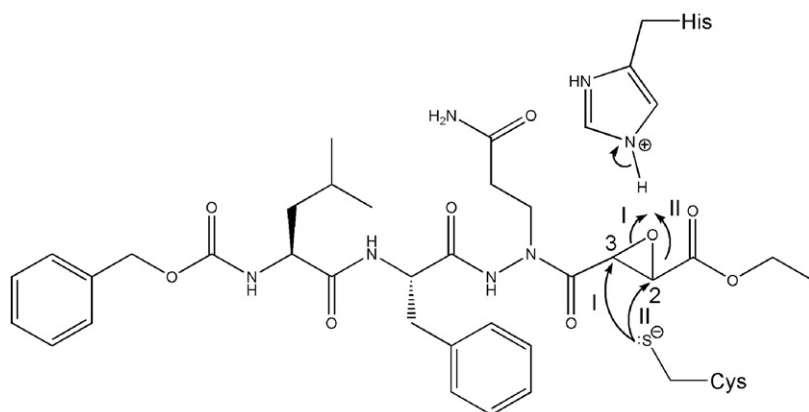
Figure 1. Inhibition of SARS-CoV M[pro] by the aza-peptide epoxides (APEs) synthesized for our study, Cbz-Leu-Phe-AGln-EP-Coo-Et. The epoxide carbon atoms are numbered and their stereochemistry is omitted for simplicity. The proposed mechanism for the irreversible inhibition of clan CD cysteine peptidases by APEs is indicated by arrows. In the inhibition of M[pro], route I was adopted. Cbz, the benzyloxycarbonyl group; AGln, aza-glutamine; EP, epoxide; COOEt, ethyl ester.

Table 1. Parameters and statistics derived from X-ray diffraction data processing and structure refinement

	Unbound M^{pro}	Unbound M$^{pro}_{+A(-1)}$	M$^{pro}_{+A(-1)}$:APE
A. *Data processing*			
Wavelength (Å)	1.116	1.116	1.116
Resolution limit[a] (Å)	41.25–1.82	24.85–2.00	24.78–2.00
	(1.89–1.82)	(2.07–2.00)	(2.07–2.00)
Space group	P2$_1$	P4$_3$2$_1$2	P4$_3$2$_1$2
Unit-cell parameters			
a (Å)	52.39	70.29	70.09
b (Å)	96.19	70.29	70.09
c (Å)	67.91	102.87	103.86
α (deg.)	90.00	90.00	90.00
β (deg.)	102.91	90.00	90.00
γ (deg.)	90.00	90.00	90.00
Mosaicity (deg.)	0.63	0.65	0.58
No. unique reflections	56,644 (5313)	17,964 (1739)	18,098 (1760)
Redundancy	2.0 (1.8)	6.8 (6.2)	7.1 (6.5)
Completeness (%)	96.4 (90.6)	99.3 (97.8)	99.8 (99.4)
R_{sym}[b] (%)	3.6 (20.1)	8.8 (48.8)	7.4 (44.4)
$\langle I/\sigma(I)\rangle$	20.0 (3.4)	19.6 (4.2)	23.7 (4.9)
B. *Structure refinement*			
Resolution range (Å)	41.25–1.82	24.85–2.00	24.78–2.00
R_{work}[c] (%)	17.0	17.5	19.6
R_{free}[c] (%)	21.7	24.1	27.0
Number of non-hydrogen atoms per asymmetric unit (average *B*-factor, Å^2)			
Protein	4753 (30.12)	2332 (23.00)	2332 (23.93)
APE	Not applicable	Not applicable	46 (22.69)
Solvent	591 (37.81)	162 (34.38)	127 (32.49)
rms deviation from ideal geometry			
Bond lengths (Å)	0.019	0.019	0.022
Bond angles (deg.)	1.778	1.765	2.064
Ramachandran plot			
Favored (%)	91.3	90.0	90.0
Allowed (%)	7.3	8.5	8.1
Generously allowed (%)	0.6	0.8	0.8
Disallowed (%)	0.8	0.8	1.2

[a] Numbers in parentheses refer to the highest resolution bins.
[b] $R_{sym} = \sum_{hkl}\sum_i |I_{hkl,i} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_i I_{hkl,i}$, where $I_{hkl,i}$ and $\langle I_{hkl}\rangle$ are the *i*th observed intensity and average intensity of the reflection *hkl*, respectively.
[c] $R_{work} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes of a particular reflection, respectively. The summation is over 95% of the reflections in the specified resolution range. The remaining 5% of the reflections were selected randomly before the structure refinement and are not included in the structure refinement. R_{free} was calculated over these reflections with the equation used for R_{work}.[54]

those reported previously.[25] Each asymmetric unit contains both protomers (A and B) of the physiological dimer. In the electron density maps of the unbound M^{pro}, residues 1A–44A, 50A–305A and 1B–302B could be identified. In the Ramachandran plot of this structure, Asp33A, Asn84A and Asn84B are in the generously allowed regions, and Asp33B, Glu47B, Tyr154A and Tyr154B are in the disallowed regions. In both independent protomers, the Asp33 N atom forms a hydrogen bond with the Thr98 O$^{\gamma1}$ atom, and the Asp33 carbonyl O atom forms

hydrogen bonds with the Trp31 N$^{\varepsilon1}$ and the Asn95 N atoms. Also, the Asn84 N$^{\delta2}$ atom forms a hydrogen bond with the Glu178 carbonyl O atom. The poorly defined electron densities of the side-chains of Glu47B, Tyr154A and Tyr154B indicate dynamic disorder.

SARS-CoV M$^{pro}_{+A(-1)}$ crystallized in space group $P2_12_12_1$[29] as well as $P4_32_12$. Each asymmetric unit of the latter contains only one protomer of the dimer; the two protomers of each dimer are related by the crystallographic 2-fold symmetry axis parallel with the C-face diagonal of the unit cell. In the electron density maps of the unbound M$^{pro}_{+A(-1)}$, residues 3–300 of the protomer were clearly identified. In the Ramachandran plot of this structure, Asp33 and Ser139 are in the generously allowed regions, and Asn84 and Tyr154 are in the disallowed regions. The side-chain of Tyr154 makes contact with those of Ile78 and probably Arg76 from a neighboring asymmetric unit. The electron density of Ser139 is not well defined. In the electron density maps of the M$^{pro}_{+A(-1)}$:APE complex, residues 2–300 of the protomer were identified. In the Ramachandran plot of this structure, Asp33 and Asn277 are in the generously allowed regions, and Asn84, Tyr154 and Ile286 are in the disallowed regions. The side-chains of Thr285 and Ile286 from opposite protomers of the dimer contact each other. The electron density of Asn277 is not well defined.

Superpositions of the crystal structures reported here with all of those previously reported show no significant difference in the protomer orientation and the overall fold.

Active sites and substrate-binding regions of the unbound SARS-CoV M^{pro}

The previously reported crystal structure of SARS-CoV M^{pro} in space group $P2_1$ at pH 6.0 showed the collapse of the active site and S1 specificity pocket of one of the protomers, whereas the structures in the same space group at pH 7.6 and at pH 8.0 showed all the active sites to be in the catalytically competent conformation (Table 2). On the basis of this trend, a pH-triggered switch for the catalytic activity of M^{pro} was proposed.[25] We have now grown crystals of M^{pro} in the same space group ($P2_1$ with the same unit-cell constants) at pH 6.0 under slightly different conditions; the active sites and the S1 specificity pockets of both protomers are in the catalytically competent conformation (Figures 2(a) and (b), 3(a) and (b), 4(a) and (b)). More specifically, superposition of protomers A and B of the resulting M^{pro} structure (rmsd 0.50 Å for 1070 out of 1172 main-chain atoms) shows good agreement in most atomic positions. In both independent protomers, the catalytic dyad has a distance of 3.6 Å between the His41 N$^{\varepsilon2}$ atom and the Cys145 S$^{\gamma}$ atom, and the Cys145 S$^{\gamma}$ atom is coplanar with the atoms of the His41 imidazole ring. Residues Gly143 to Cys145 are in the proper conformation to form the oxyanion hole that

Table 2. Crystal structures of SARS-CoV M^{pro} (wild-type and variants) determined in different conditions

Ref.	Inhibitor	Space group	pH	In the catalytically competent conformation?	
				Protomer A[a]	Protomer B[a]
Wild-type SARS-CoV M^{pro}					
[b]	–	*P*2$_1$	**6.0**	**Yes**	**Yes**
36	–	*P*4$_3$2$_1$2	5.9	No	No
25	–	*P*2$_1$	6.0	Yes	No
25	CMK	*P*2$_1$	6.0	Yes	No
29	–	*C*2	6.5	Yes	Yes
29	APE	*C*2	6.5	Yes	Yes
37	–	*P*2$_1$2$_1$2	6.5	Yes	Yes
36	–	*P*2$_1$2$_1$2	6.6	No	No
42	–	*P*2$_1$2$_1$2	7.0	Yes	Yes
25	–	*P*2$_1$	7.6	Yes	Yes
25	–	*P*2$_1$	8.0	Yes	Yes
SARS-CoV M$^{pro}_{+A(-1)}$					
[b]	–	*P*4$_3$2$_1$2	**6.5**	**No**	**No**
[b]	**APE**	***P*4$_3$2$_1$2**	**6.5**	**Yes**	**Yes**
29	APE	*P*2$_1$2$_1$2$_1$	6.5	Yes	Yes
SARS-CoV M^{pro} with additional Ser-Leu at the N termini of both protomers					
[c]	–	*P*2$_1$	6.5	Yes	Yes
SARS-CoV M^{pro} C145A variant					
42	product	*C*2	9.0	Yes	Yes
SARS-CoV M^{pro} with additional Gly-Pro-Leu-Gly-Ser at the N termini of both protomers					
27	I2	*P*2$_1$	6.0	Yes	Yes
27	N1	*P*2$_1$	6.0	Yes	Yes
27	N3	*P*2$_1$	6.0	Yes	Yes
27	N9	*P*2$_1$	6.0	Yes	Yes

[a] For any structure whose asymmetric unit contains only one protomer, that protomer represents both protomer A and protomer B.
[b] These are the structures reported here.
[c] PDB accession code 1Q2W; J. B. Bonanno *et al.*, unpublished results.

accommodates the carbonyl O atom of the scissile peptide bond of the substrate; the position to be occupied by the carbonyl O atom is occupied by a water molecule. The O atom of the water molecule forms hydrogen bonds at distances of 3.0 Å with the Gly143 N atom and with the Cys145 N atom (Figure 3(a) and (b)).

Previous studies suggest strongly that the predominant S1 specificity of SARS-CoV M^{pro} for Gln is determined primarily by the conserved residue His163.[15,16,35] In both independent protomers, the orientation of the imidazole ring of His163 is determined by the hydrogen bonding between its N$^{\delta 1}$ atom and the OH group of Tyr161 (3.0 Å) and by its π-stacking with the phenyl ring of Phe140 (the distance between the geometric centers of the aromatic rings is 3.9 Å). The position to be occupied by the side-chain carbonyl O atom of P1-Gln of the substrate is occupied by a water molecule, whose O atom forms a hydrogen bond with the His163 N$^{\varepsilon 2}$ atom (2.9 Å); this water molecule is coplanar with the atoms of the imidazole ring of His163. In both protomers, Phe140 and Glu166 interact with Ser1 of the opposite protomer to form the "floor" of the S1

specificity pocket (Figure 3(a) and (b)). In protomer A, the Glu166A O$^{\varepsilon 1}$ atom forms a hydrogen bond (2.8 Å) with the His172A N$^{\varepsilon 2}$ atom, thereby constituting a "side-wall" of the S1 specificity pocket similar to that in protomer A of previously reported structures and the molecular-dynamic simulation model of M^{pro} at pH 6.0 (Figure 3(a)).[25,36] However, in protomer B of the structure of M^{pro} reported here, the side-chain of Glu166B rotates away and the His172B N$^{\varepsilon 2}$ atom forms a hydrogen bond (3.0 Å) with the Ser1A O$^{\gamma}$ atom of protomer A instead. The Ser1A O$^{\gamma}$ atom also forms a hydrogen bond (2.8 Å) with the Gly170B carbonyl O atom (Figure 3(b)). The difference in the side-chain conformation of Glu166 in protomers A and B of the structure of M^{pro} reported here indicates the weakness of the interaction between the side-chains of Glu166 and His172, even though this interaction may acquire some ionic character by the protonation of the His172 N$^{\delta 1}$ atom. Contrary to those in protomer B of both the previously reported structure and the molecular-dynamic simulation model of M^{pro} at pH 6.0,[25,36] the side-chains of His163B and Glu166B do not interact in protomer B of the structure of M^{pro} reported here, probably because His163B is protonated only at its N$^{\varepsilon 2}$ atom and carries no charge.

As in the previously reported structure of M^{pro} at pH 6.0,[25] protomers A and B in the current structure show good agreement in the rest of the substrate-binding region, including the S2, S4 and S1′ specificity pockets. Superpositions with the crystal structures of M^{pro} in the other conditions show that the rest of the substrate-binding regions in the structure of M^{pro} reported here are in the catalytically competent conformation.

Active sites and substrate-binding regions of the unbound SARS-CoV M$^{pro}_{+A(-1)}$

We have reported the crystal structure of the SARS-CoV M$^{pro}_{+A(-1)}$:APE complex in space group *P*2$_1$2$_1$2$_1$, whose asymmetric unit contains both protomers of the M$^{pro}_{+A(-1)}$ dimer.[29] In that structure, the active sites and the substrate-binding regions of both protomers are in the catalytically competent conformation (Table 2). In both independent protomers, the additional Ala at the N terminus blocks Ser1 and disrupts its interactions with Phe140 and Glu166 of the opposite protomer; however, the floor of the S1 specificity pocket is only partly disrupted. More importantly, the presence of a ten-residue affinity tag at the N terminus of both independent protomers reduces the specific activity of M^{pro} by less than an order of magnitude.[29] Similar observations are given by the crystal structure of M^{pro} with additional residues Ser-Leu at the N terminus of both independent protomers (PDB accession code 1Q2W; J. B. Bonanno *et al.*, unpublished results).

Attempts to crystallize M$^{pro}_{+A(-1)}$ in space group *P*2$_1$ have not been successful. Interestingly, M$^{pro}_{+A(-1)}$ crystallizes in space group *P*4$_3$2$_1$2 as well as in space group *P*2$_1$2$_1$2$_1$ under the same conditions. In space group *P*4$_3$2$_1$2, each asymmetric unit has one

protomer of $M_{+A(-1)}^{pro}$. Superpositions of the resulting structure of the unbound $M_{+A(-1)}^{pro}$ with the crystal structures of M^{pro} in the other conditions show no difference in the overall fold or the protomer orientation. Crystal contacts of the unbound $M_{+A(-1)}^{pro}$ in space group $P4_32_12$ do not involve any residues forming the active sites and the S1 specificity pockets of the peptidase. The structure of the unbound $M_{+A(-1)}^{pro}$ in space group $P4_32_12$ shows good agreement in most atomic positions with the structure of the APE-bound $M_{+A(-1)}^{pro}$ in space group $P2_12_12_1$. In the structure of the unbound $M_{+A(-1)}^{pro}$ reported here, the catalytic dyad has a distance of 3.9 Å between the His41 $N^{\epsilon2}$ atom and the Cys145 S^{γ} atom, and the Cys145 S^{γ} atom is coplanar with the atoms of the His41 imidazole ring. However, the oxyanion hole and the S1 specificity pocket are distorted (Figures 2(c) and 3(c)). The φ and ψ angles of residues Lys137 to Ser144 show dramatic differences compared with those of the previously reported structure of the APE-bound $M_{+A(-1)}^{pro}$[29] (Figure 4(c)). The above-average *B*-factors of these residues in the unbound $M_{+A(-1)}^{pro}$ indicate their high mobility relative to the rest of the peptidase (Figure 5(c)). The oxyanion hole is not distorted as much; the N atoms of Gly143 and Cys145 are still oriented to donate hydrogen bonds that would stabilize the negatively charged carbonyl O atom of the scissile peptide bond of the substrate, although no water molecule is found at the position to be occupied by the carbonyl O atom (Figure 3(c)).

The hydrogen bond between the $N^{\delta1}$ atom of His163 and the OH group of Tyr161 is preserved (3.1 Å); however, the imidazole ring of His163 is no longer π-stacked with the phenyl ring of Phe140; it makes contacts with the side-chain of Leu141 instead (Figure 3(c)). Here the phenyl ring of Phe140 makes contacts with the side-chains of Val114, Tyr126, Ile136 and His172, and the main-chain atoms of Lys137 and Gly138 from the parent protomer, and with the side-chain of Arg4 from the opposite protomer. The Phe140 carbonyl O atom also forms a long hydrogen bond with the OH group of Tyr118 (3.4 Å). The electron density maps show two possible conformers of the side-chain of Glu166; the occupancies of both conformers were fixed at 0.5
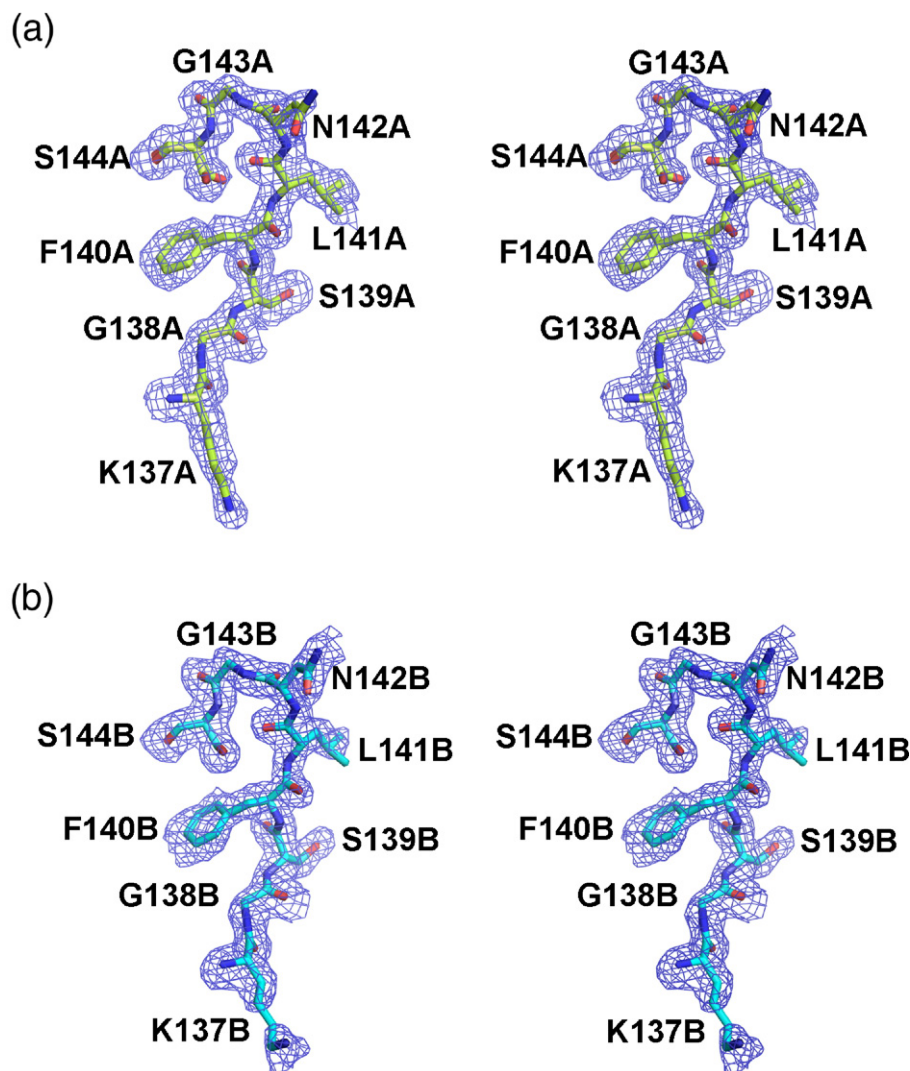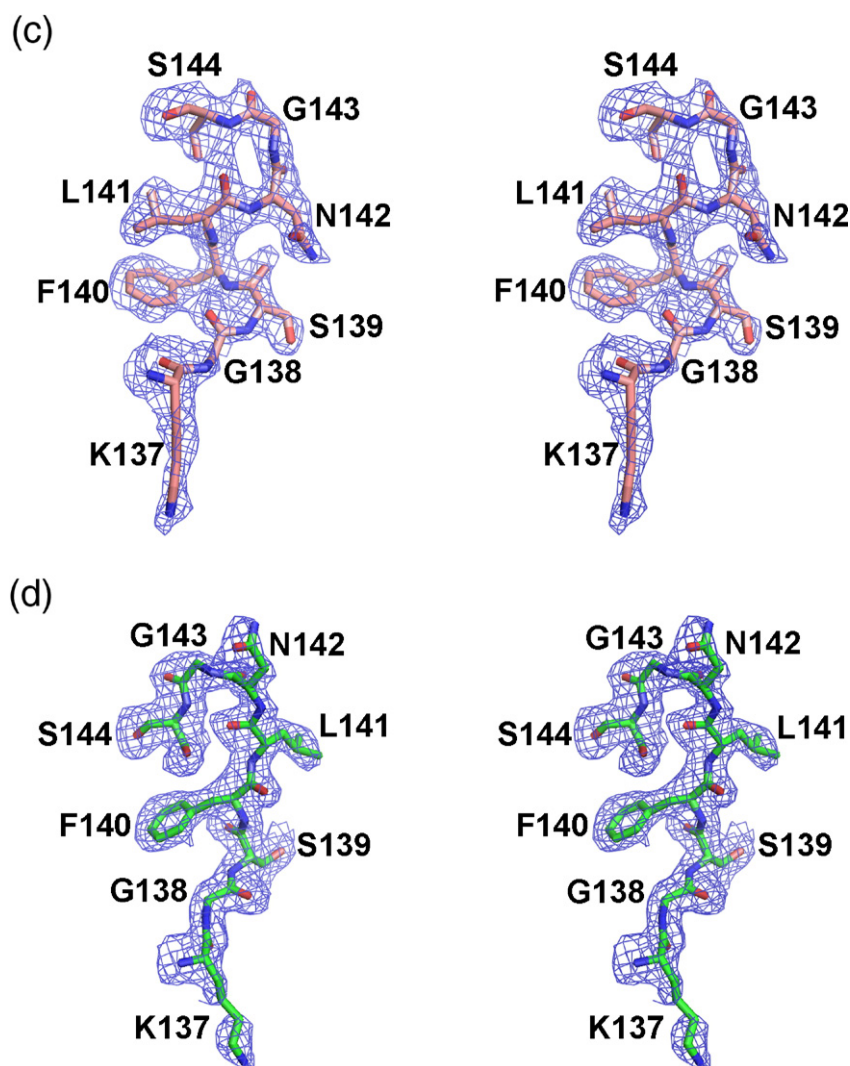

Figure 2 (*legend on next page*)

(c)

(d)

Figure 2. Electron densities in the F_o–F_c omit maps for residues Lys137 to Ser144 of SARS-CoV M^{pro} and $M^{pro}_{+A(-1)}$. (a) Protomer A of the unbound M^{pro}. (b) Protomer B of the unbound M^{pro}. (c) Unbound $M^{pro}_{+A(-1)}$. (d) APE-bound $M^{pro}_{+A(-1)}$. The residues in (c) are in the catalytically incompetent conformation.

without refinement. In conformer 1, the Glu166 $O^{\varepsilon 1}$ atom forms a hydrogen bond with the His172 $N^{\delta 1}$ atom (2.3 Å), and the Glu166 $O^{\varepsilon 2}$ atom forms a hydrogen bond with the His163 $N^{\varepsilon 2}$ atom (2.8 Å); whereas in conformer 2, the side-chain of Glu166 protrudes into the solvent (Figure 3(c)). Probably because of their high mobility, the additional Ala at the N terminus, Ser1 and Gly2 of the protomer could not be identified in the electron density maps of the unbound $M^{pro}_{+A(-1)}$. Interactions similar to those forming the floors of the S1 specificity pockets of M^{pro} are not observed in the unbound $M^{pro}_{+A(-1)}$.

The rest of the substrate-binding region in the structure of the unbound $M^{pro}_{+A(-1)}$ reported here is in the catalytically competent conformation, in good agreement with those in the previously reported structures of the APE-bound $M^{pro}_{+A(-1)}$ and of M^{pro}, except for some side-chains whose conformational rearrangements are necessary in order to accommodate APE or the other inhibitors.

Binding of APE to SARS-CoV $M^{pro}_{+A(-1)}$

Both the $P2_1$ crystals of SARS-CoV M^{pro} and the $P4_32_12$ crystals of SARS-CoV $M^{pro}_{+A(-1)}$ were soaked

in solutions of the APE synthesized for this study in three stereochemical versions: the two diastereomers Cbz-Leu-Phe-AGln-(2S,3S)EP-COOEt and Cbz-Leu-Phe-AGln-(2R,3R)EP-COOEt, and the racemic mixture of these two diastereomers. Outstanding electron density for APE was observed only in the electron density maps of $M^{pro}_{+A(-1)}$; it could be fit only by the 2S,3S diastereomer (Figure 6(a)). The latter is consistent with the results of the studies on the inhibition of M^{pro} by the APE in these versions. These results can be explained with the models of all four of the possible diastereomers of the APE binding to M^{pro}.[29] Superpositions of the structures of the unbound and the APE-bound $M^{pro}_{+A(-1)}$ reported here show that the binding of the APE does not grossly affect the overall fold or the protomer orientation of the $M^{pro}_{+A(-1)}$ dimer. The same observations are found in the superpositions of the previously reported crystal structures of the unbound and the APE-bound M^{pro} in space group C2.[29] The structure of the APE-bound $M^{pro}_{+A(-1)}$ in space group $P4_32_12$ also agrees well in most atomic positions with that of the APE-bound $M^{pro}_{+A(-1)}$ in space group $P2_12_12_1$.

The structural consequences of the binding of the APE to $M^{pro}_{+A(-1)}$ are very similar to those observed in the crystal structures of the unbound and the APE-bound M^{pro} in space group $C2$, and the APE-bound $M^{pro}_{+A(-1)}$ in space group $P2_12_12_1$.[29] A covalent bond forms between the Cys145 S$^\gamma$ atom of $M^{pro}_{+A(-1)}$ and the epoxide C3 atom of the APE (2.15 Å; Figure 6(b) and (c)). In the refinement of all of the structures of the APE-bound M^{pro} and $M^{pro}_{+A(-1)}$, a restraint was applied to this C-S bond on the basis that the length of a C-S single bond is normally about 1.8 Å. Interestingly, the opened epoxide moiety of the APE always tended to be tilted away from Cys145 of M^{pro} or $M^{pro}_{+A(-1)}$, thereby lengthening this C-S bond by 0.2–0.3 Å, even though the opened epoxide moiety of the APE was manually moved back towards Cys145 of M^{pro} or $M^{pro}_{+A(-1)}$ regularly during the refinement process. Although such lengthening is not considered significant as the overall positional uncertainties (based on maximum likelihood) of all these structures are in the range of 0.2–0.3 Å, it does suggest the possibility of this C-S bond being under strain and vulnerable to rupture caused by a second nucleophilic attack at the epoxide C3 atom of the APE (say, by an activated water molecule), leading

to the speculation that the APE could act as a reversible inhibitor of M^{pro}.

The conformation of the opened epoxide moiety and the main-chain conformation of P1-AGln of the APE in the structure of the $M^{pro}_{+A(-1)}$:APE complex reported here are essentially the same as those in the previously reported structures of the M^{pro}:APE and the $M^{pro}_{+A(-1)}$:APE complexes (Figure 7(a)–(d)). Unlike the P1 C$^\alpha$ atom of other inhibitors, the P1-AGln N$^\alpha$ atom of the APE is sp^2-hybridized and has a trigonal planar geometry; in order to be accommodated by the S1 specificity pocket of the peptidases, the side-chain of P1-AGln has to adopt a different conformation, in particular the equivalent to χ (N-N$^\alpha$-C$^\beta$-C$^\gamma$). This angle is −123.6° in the structure of the $M^{pro}_{+A(-1)}$: APE complex reported here; by contrast, in the crystal structures of the complexes of an M^{pro} variant with a series of peptidomimetic inhibitors, this angle is in the range of −65° to −80°.[27]

Interestingly, the binding of the APE induces the recovery of the catalytically competent conformation of the oxyanion holes and the S1 specificity pockets of $M^{pro}_{+A(-1)}$ (Figures 2(d) and 3(d)). The P1-AGln main-chain carbonyl O atom of the APE is accommodated in the oxyanion hole of $M^{pro}_{+A(-1)}$,

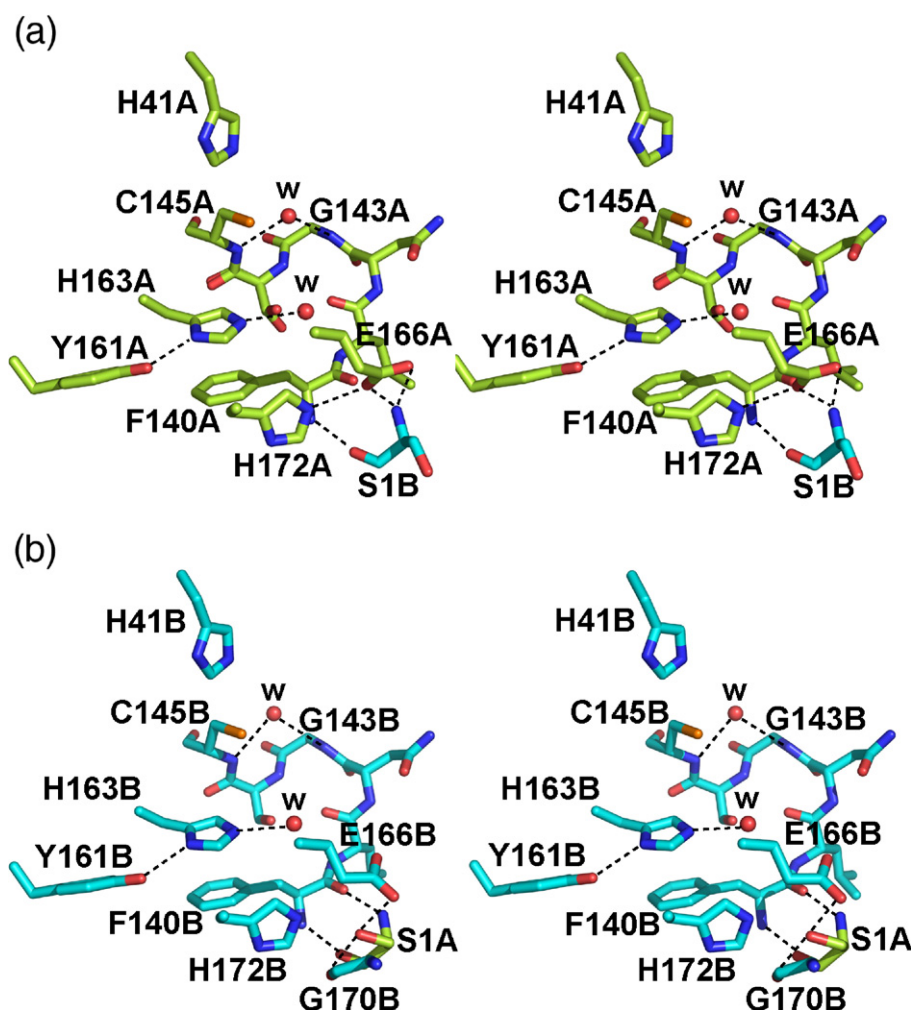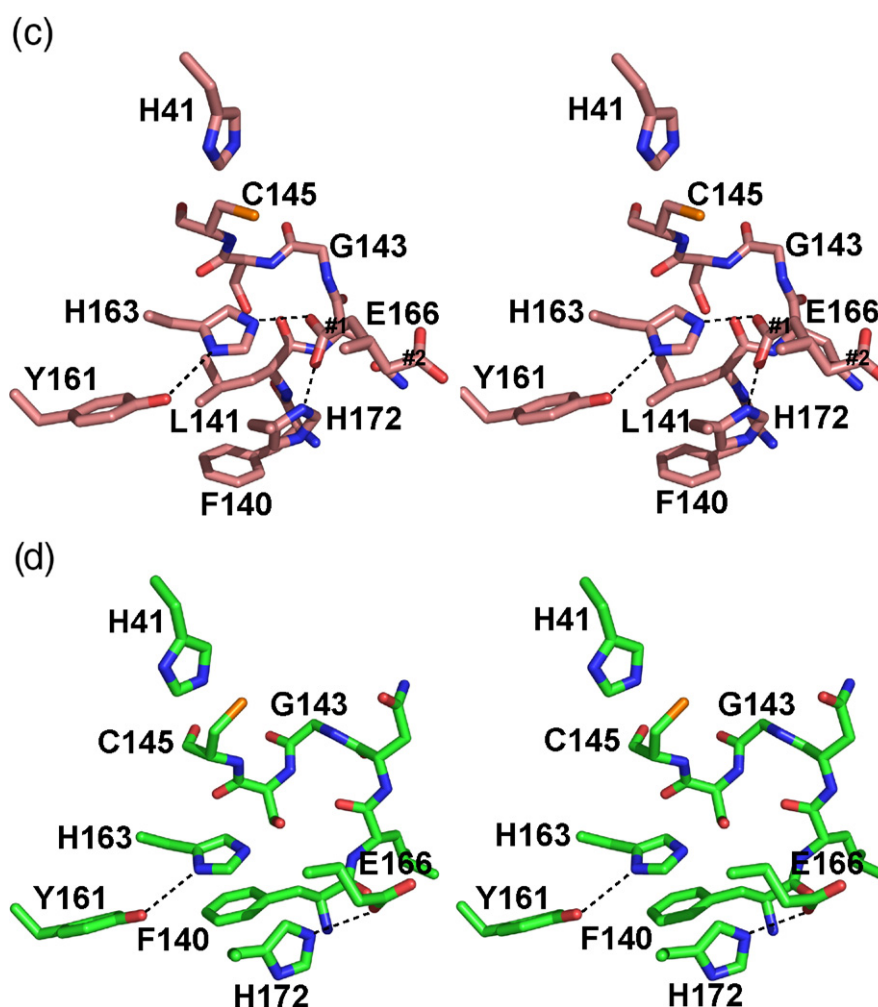Figure 3 (*legend on next page*)

Figure 3. Active sites and S1 specificity pockets of SARS-CoV M^{pro} and $M^{pro}_{+A(-1)}$, viewed from the S2 specificity pockets. (a) Protomer A of the unbound M^{pro}. (b) Protomer B of the unbound M^{pro}. (c) Unbound $M^{pro}_{+A(-1)}$. (d) APE-bound $M^{pro}_{+A(-1)}$ (For clarity, the APE is not shown.). Hydrogen bonds are indicated by broken lines. Water molecules are labeled w.

forming hydrogen bonds with the N atoms of Gly143 (2.8 Å) and Cys145 (3.2 Å) (Figure 6(b) and (c)). The ϕ and ψ angles of residues Lys137 to Ser144 in the structure of the APE-bound $M^{pro}_{+A(-1)}$ reported here are essentially the same as those in the previously reported structure of the APE-bound $M^{pro}_{+A(-1)}$ (Figure 4(d)). In contrast to those in the structure of the unbound $M^{pro}_{+A(-1)}$, the *B*-factors of these residues in the structures of the APE-bound $M^{pro}_{+A(-1)}$ are close to the averages, indicating that the mobility of these residues is reduced upon the binding of the APE (Figure 5(d)).

The hydrogen bond between the $N^{\delta 1}$ atom of His163 and the OH group of Tyr161 is preserved (3.3 Å), and the π-stacking of the imidazole ring of His163 with the phenyl ring of Phe140 is recovered (distances between the geometric centers of the aromatic rings: 3.7 Å) (Figure 3(d)). The His163 $N^{\epsilon 2}$ atom no longer interacts with the Glu166 $O^{\epsilon 2}$ atom, but forms a hydrogen bond (2.7 Å) with the side-chain carbonyl O atom of P1-AGln of the APE instead (Figure 6(b) and (c)). The side-chains of

Glu166 and His172 interact with each other (3.0 Å; Figure 3(d)). The additional Ala at the N terminus and Ser1 of the protomer could not be identified in the electron density maps of the $M^{pro}_{+A(-1)}$:APE complex. Interestingly, the side-chain amide N atom of P1-AGln of the APE forms hydrogen bonds, though not in ideal geometry, with the Phe140 carbonyl O atom (3.4 Å) and the Glu166 $O^{\epsilon 1}$ and $O^{\epsilon 2}$ atom (3.5 Å and 2.9 Å, respectively) (Figure 6(b)). Therefore, without the participation of Ser1 of the opposite protomer, Phe140 and Glu166 are tied together to form parts of the floors of the S1 specificity pockets of $M^{pro}_{+A(-1)}$.

Similar to the APE in protomer A of the previously reported structure of the APE-bound $M^{pro}_{+A(-1)}$, the benzyloxycarbonyl (Cbz) group of the APE in the structure of the APE-bound $M^{pro}_{+A(-1)}$ reported here squeezes into and slightly widens the S4 specificity pockets of $M^{pro}_{+A(-1)}$ and, as a result, making contacts with Leu167, Pro168, Gln192 and Ala193 of $M^{pro}_{+A(-1)}$. Otherwise, the interactions of the APE with the rest of the substrate-binding region observed in the
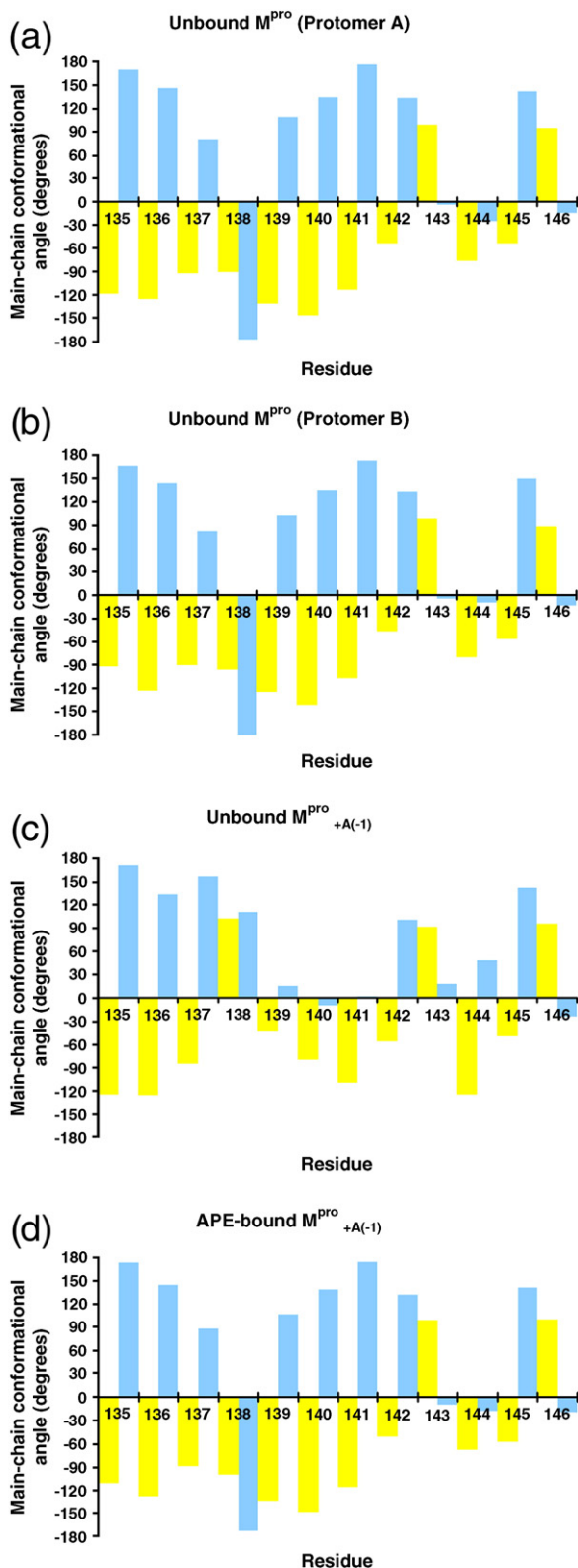
Figure 4. Main-chain conformational angles of residues Thr135 to Gly146 in the crystal structures of SARS-CoV M^{pro} and M$^{pro}_{+A(-1)}$ reported here. The ϕ and the ψ angles are represented by the yellow and the blue bars, respectively. (a) Protomer A of the unbound M^{pro} (overall positional uncertainty estimated based on maximum likelihood, σ_r: 0.15 Å). (b) Protomer B of the unbound M^{pro} (σ_r: 0.15 Å). (c) Unbound M$^{pro}_{+A(-1)}$ (σ_r: 0.23 Å). (d) APE-bound M$^{pro}_{+A(-1)}$ (σ_r: 0.28 Å).

structure of the APE-bound M$^{pro}_{+A(-1)}$ reported here are essentially the same as those observed in the previously reported structures of the APE-bound M$^{pro}_{+A(-1)}$ and M^{pro} (Figure 6(b) and (c)).

Dynamic equilibrium for the conformation of the active sites and the S1 specificity pockets of SARS-CoV M^{pro}

In contrast to the previously reported crystal structure of SARS-CoV M^{pro} in space group $P2_1$ at pH 6.0,[25] the crystal structure of M^{pro} reported here (in the same space group and at the same pH) shows that the active sites and the S1 specificity pockets of both protomers are in the catalytically competent conformation (Table 2). In the determinations of both structures, the complete wild-type sequence (residues 1–306) of M^{pro} was over-expressed, purified and crystallized. Although different strategies might have been used in the X-ray diffraction data collection and processing, and in the structure solution and refinement, this could not result in the dramatic structural differences observed. The structural differences probably arise from the differences in the conditions of the preparation and the crystallization of M^{pro}. Similarly, the crystal structure of M^{pro} in space group $P2_12_12$ at pH 6.5 shows that both protomers are in the catalytically competent conformation,[37] whereas that in the same space group $P2_12_12$ and at a slightly higher pH (6.6) shows that both protomers of the latter are in the catalytically incompetent conformation.[36] The conformations of the collapsed active sites and S1 specificity pockets observed in the catalytically incompetent protomers show some variability. This variability indicates that the active site and the S1 specificity pocket of each protomer of M^{pro} do not adopt a single conformation in solution at pH 5.9–6.6, but instead there is an ensemble of conformations. A particular conformation might be favored by a particular set of crystallization conditions.

However, the various conformations of the collapsed active sites and S1 specificity pockets share a common feature: the imidazole ring of His163 is not π-stacked with the phenyl ring of Phe140. π-stacking is an example of aromatic interactions.[38] In the structures of the unbound M^{pro}, an aromatic interaction is observed between these two rings in an offset-stacked (i.e. π-stacking) or an edge-to-face fashion wherever the protomer is in the catalytically competent conformation (Figure 8). In both orientations, one or two hydrogen atoms (with partial positive charge) on the phenyl ring of Phe140 are positioned near the central region (with partial negative charge) of the imidazole ring of His163. According to the results of previous studies,[39,40] a single aromatic interaction as such is weak (the interaction energy may be in the range of 1–2 kcal/mol only), in contrast to the clusters of aromatic interactions commonly involved in the stabilization of protein structures.[41] Therefore, this interaction is susceptible to disruptions that could be caused by changes in a number of factors; the formation of this

interaction can be viewed as a reversible process in dynamic equilibrium:

Aromatic interaction ↔ No aromatic interaction

(1)
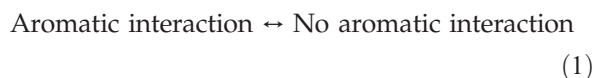
The position of the equilibrium could be determined, in part, by pH. At low pH (near or below

(a) Unbound M^pro (Protomer A)

(b) Unbound M^pro (Protomer B)

(c) Unbound M^pro_{+A(-1)}

(d) APE-bound M^pro_{+A(-1)}

the pK_{a1} of His163; Figure 9), the His163 N$^{\delta 1}$ atom in a significant number of the M^pro protomers is protonated, thereby introducing a positive charge on the imidazole ring of His163. This would disfavor its aromatic interaction with the phenyl ring of Phe140, and the position of the equilibrium would shift to the right. This is consistent with the trend exhibited by the structures of the unbound M^pro at various pH values (Table 2). In the pH range of 7.0–9.0, most of the M^pro protomers have the aromatic interaction and are in the catalytically competent conformation (the left-hand side of the equilibrium predominates); whereas in the pH range of 5.9–6.6, some of the M^pro protomers lose the aromatic interaction and are in the catalytically incompetent conformation (the position of the equilibrium shifts to the right).

A second factor in determining the position of the equilibrium could be the integrity of the interactions among Phe140 and Glu166 of the parent protomer, and Ser1 of the opposite protomer. In the structure of M^pro_{+A(-1)} reported here, the additional Ala at the N terminus of the opposite protomer blocks Ser1 and disrupts its normal interactions with Phe140 and Glu166 of the parent protomer. This probably weakens the conformational anchor of Phe140, as indicated by its above-average *B*-factors (Figure 5(c)), making the aromatic interaction of its phenyl ring with the imidazole ring of His163 vulnerable to disruption. In all of the crystal structures of the unbound M^pro (both the wild-type and the variants) determined so far, wherever Phe140 and Glu166 of the parent protomer, and Ser1 of the opposite protomer do not interact normally, the *B*-factors of residues Lys137 to Ser144 of the parent protomer are above the average, even if the parent protomer is in the catalytically competent conformation (e.g. the M^pro structure of M^pro at pH 6.5 in the space group $P2_12_12$);[37] in contrast, wherever the three residues interact normally, the *B*-factors of residues Lys137 to Ser144 of the parent protomer are close to the average (e.g. protomer A of the M^pro structure reported here; Figure 5(a)). These observations suggest that the interactions among the three residues can immobilize residues 137 to 144. Apparently, both the ionic interaction between the N terminus of Ser1 and the side-chain of Glu166 of opposite protomers, and the amide hydrogen-carbonyl oxygen hydrogen bonds between Ser1 and Phe140 of opposite protomers contribute to the immobilization of residues Lys137 to Ser144. The
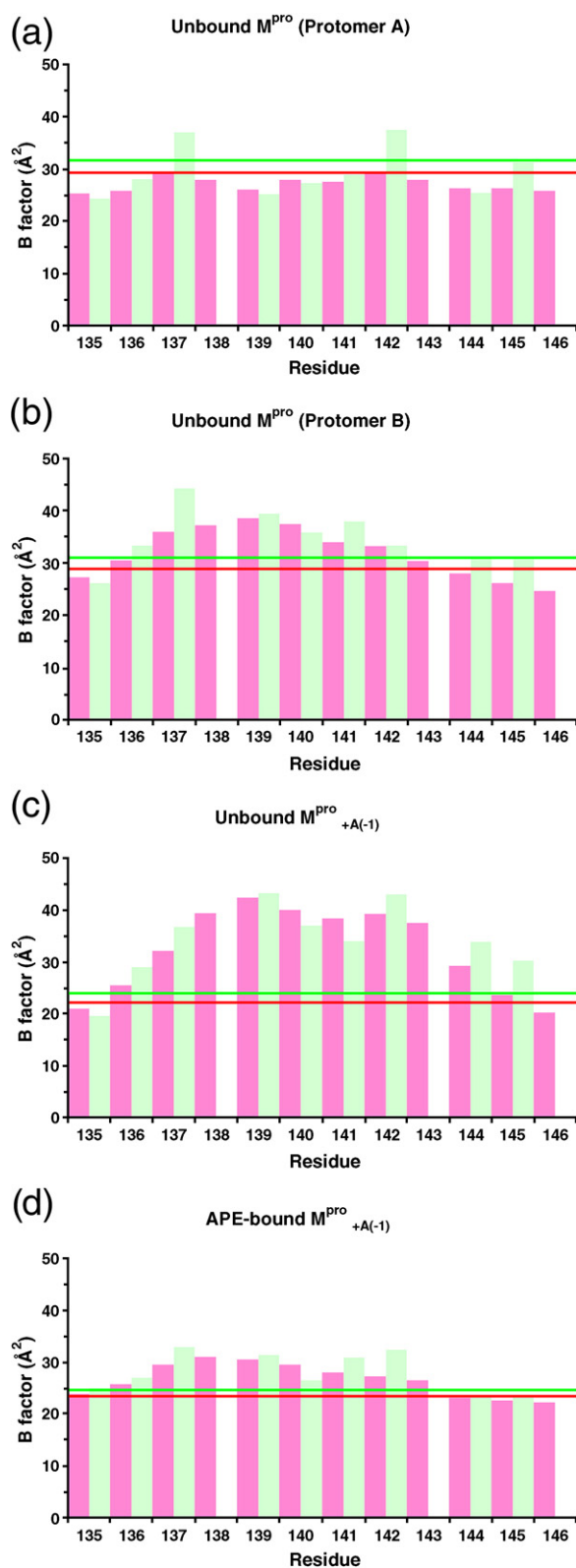
Figure 5. The *B*-factors of residues Thr135–Gly146 in the crystal structures of SARS-CoV M^pro and M^pro_{+A(-1)} reported here. The main chain and the side-chain *B*-factors are represented by the red and the green bars, respectively. The average values of the main chain and the side-chain *B*-factors are indicated by the red and the green lines, respectively. There is no green bar for Gly138, Gly143 or Gly146, because these residues do not have side-chains. (a) Protomer A of the unbound M^pro. (b) Protomer B of the unbound M^pro. (c) Unbound M^pro_{+A(-1)}. (d) APE-bound M^pro_{+A(-1)}.
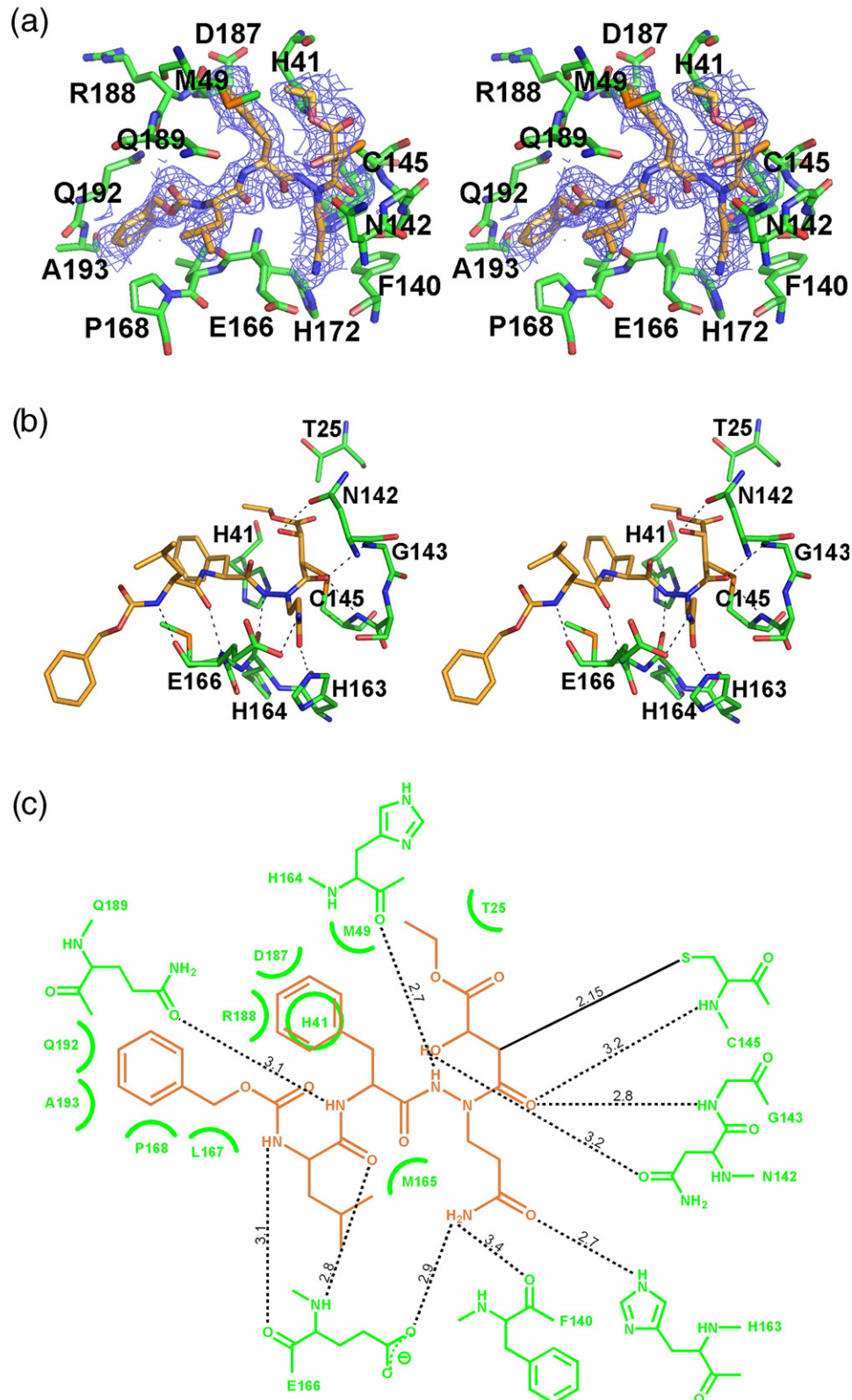
Figure 6. Interactions of the APE (orange) with SARS-CoV $M^{pro}_{+A(-1)}$ (green). (a) Outstanding density in the $F_o - F_c$ map of the $M^{pro}_{+A(-1)}$:APE complex. (b) View from the 'floors' of the S1 specificity pockets of $M^{pro}_{+A(-1)}$. (c) A diagram of the interactions. Hydrogen bonds are shown as broken lines, with their distances (in Å) given alongside. The residues of $M^{pro}_{+A(-1)}$ in contact with the APE are shown as arcs.
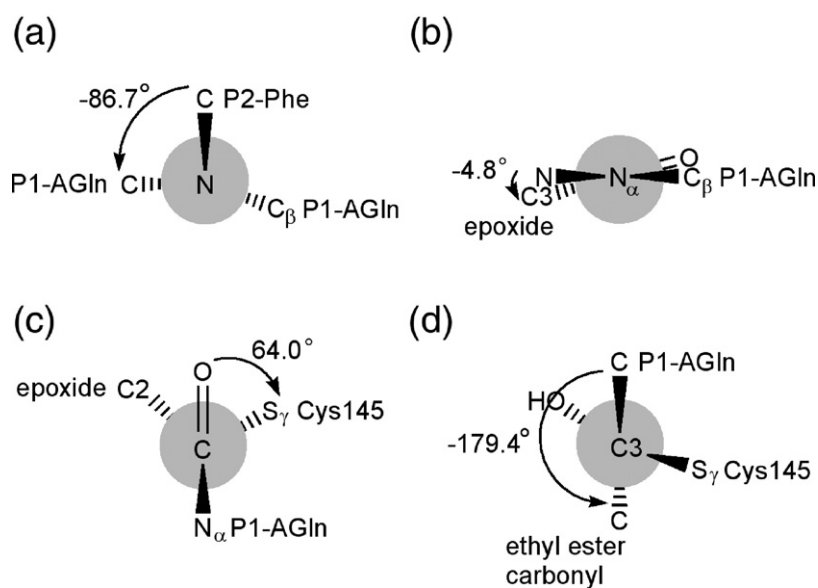
Figure 7. Newman projections of the APE. (a) The equivalent to ϕ of P1-AGln, C(P2-Phe)-N(P1-AGln)-N$^\alpha$(P1-AGln)-C(P1-AGln); (b) the equivalent to ψ of P1-AGln, N(P1-AGln)-N$^\alpha$(P1-AGln)-C(P1-AGln)-C3(epoxide); (c) the torsion angle O=C(P1-AGln)-C3(epoxide)-S$^\gamma$(Cys145); and (d) the torsion angle C(P1-AGln)-C3(epoxide)-C2(epoxide)-C(ethyl ester carbonyl).

M^{pro} structure at pH 7.0 in space group $P2_12_12$[42] and protomer B of the M^{pro} structure reported here (Figure 5(b)) suggest that the absence of either interaction slightly compromises the immobilizing effect.

The increased vulnerability of the aromatic interaction to disruption would shift the position of the equilibrium to the right. This probably explains why the active sites and the S1 specificity pockets of both protomers are collapsed in the crystal structure of M$^{pro}_{+A(-1)}$ reported here. M^{pro} with its three N-terminal residues truncated, M$^{pro}_{\Delta(1-3)}$, exists mainly as a dimer in solution but its activity is 24% lower than that of M^{pro}.[14] Without the interactions among Phe140 and Glu166 of the parent protomer, and Ser1 of the opposite protomer, the position of the equilibrium would shift to the right, resulting in the decreased availability of the active sites and the S1 specificity pockets in the catalytically competent conformation. The binding of a substrate to M$^{pro}_{\Delta(1-3)}$ may induce the recovery of the catalytically competent conformation of the active sites and the S1 specificity pockets (to be discussed below), at an energetic cost accounted for by the reduced activity of M$^{pro}_{\Delta(1-3)}$.

Induced-fit binding of the APE to SARS-CoV M^{pro}

The crystal structures of the unbound and the APE-bound SARS-CoV M$^{pro}_{+A(-1)}$ reported here show that the binding of the APE to M$^{pro}_{+A(-1)}$ follows an induced-fit model, not the lock-and-key model. More particularly, the binding of the APE induces the recovery of the catalytically competent conformation of the oxyanion holes and the S1 specificity pockets of both protomers of M$^{pro}_{+A(-1)}$. This induced fit is likely relevant to the binding of APE to M^{pro}, because of the conformational flexibility of the active sites and the S1 specificity pockets of M^{pro} discussed above. In all of the crystal structures of

the inhibitor-bound M^{pro} (both the wild-type and the variants) determined so far,[25–27,29,42] except for protomer B in the complex of M^{pro} with a chloromethyl ketone (CMK),[25] the active sites and the S1 specificity pockets are in the catalytically competent conformation (Table 2). The induced-fit binding of the APE to M$^{pro}_{+A(-1)}$ probably is driven by the formation of the covalent bond between the Cys145 S$^\gamma$ atom of M$^{pro}_{+A(-1)}$ and the epoxide C3 atom of the APE. It would be energetically costly to break this C-S bond in order to repel the APE and preserve the conformational flexibility of the active sites and the S1 specificity pockets of M$^{pro}_{+A(-1)}$. Therefore, the active sites and the S1 specificity pockets of M$^{pro}_{+A(-1)}$ have to adopt the catalytically competent conformation in order to accommodate a substrate mimic like APE. This explanation probably applies to the actual catalytic cycle of M^{pro} as well: the acylation step (i.e. formation of
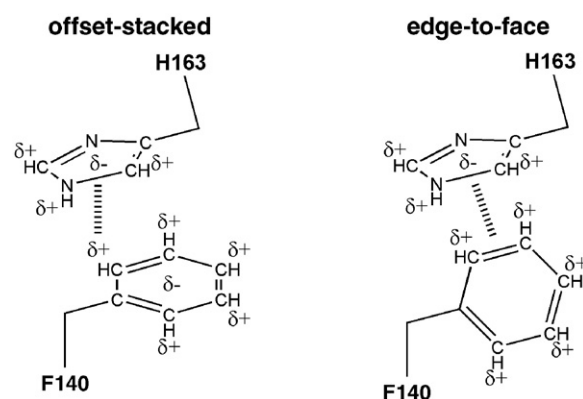

Figure 8. Aromatic interactions observed between Phe140 and His163 in the crystal structures of SARS-CoV M^{pro} (wild type and variants). The broken lines indicate the alignment of the partial positive charges ($\delta+$) of the hydrogen atoms on the phenyl ring of Phe140 with the partial negative charges ($\delta-$) in the central part of the imidazole ring of His163.
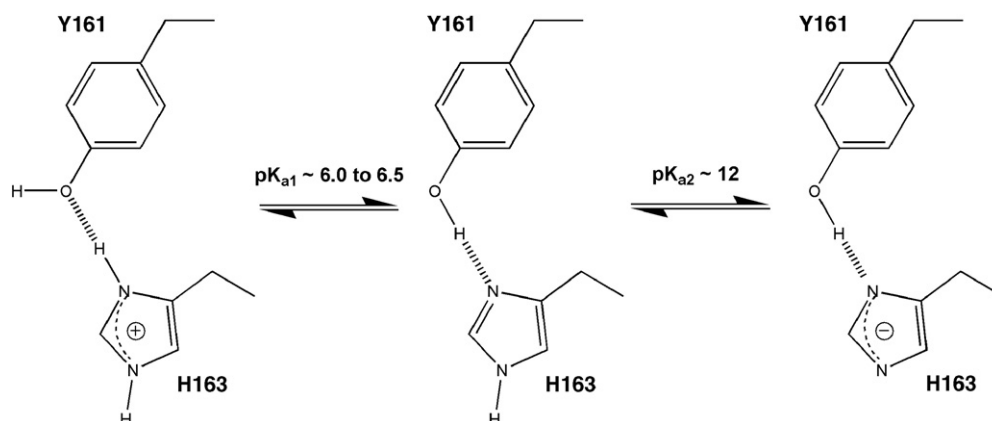
Figure 9. Protonation/deprotonation states of His163 hydrogen bonded to Tyr161 in SARS-CoV M^{pro} (wild-type and variants). K_{a1} is the dissociation constant for the protonation of the His163 $N^{\delta1}$ atom; K_{a2} is the dissociation constant for the deprotonation of the His163 $N^{\varepsilon2}$ atom.

the covalent acyl-enzyme intermediate) may drive the induced-fit binding of a substrate to M^{pro}; after the deacylation step (i.e. breakdown of the covalent acyl-enzyme intermediate), the removal of the product may be favored by the recovery of the conformational flexibility of the active sites and the S1 specificity pockets of M^{pro}.

As the side-chain of P1-AGln of the APE occupies the S1 specificity pockets of $M^{pro}_{+A(-1)}$, the side-chain amide N atom of P1-AGln of the APE donates hydrogen bonds to the Phe140 carbonyl O atom and to the Glu166 $O^{\varepsilon1}$ and $O^{\varepsilon2}$ atoms of $M^{pro}_{+A(-1)}$, in addition to the hydrogen bond between the side-chain carbonyl O atom of P1-AGln of the APE and the His163 $N^{\varepsilon2}$ atom of $M^{pro}_{+A(-1)}$. Similarly, additional hydrogen bonds are observed in the crystal structures of the $M^{pro}_{+A(-1)}$:APE complex in space group $P2_12_12_1$, the M^{pro}:APE complex,[29] protomer A of the M^{pro}:CMK complex,[25] and the complex of M^{pro} with a peptidomimetic inhibitor.[26] In the crystal structures of the complexes of an M^{pro} variant with a series of peptidomimetic inhibitors, these hydrogen bonds seem to be water-mediated.[27] Consequently, even without the participation of Ser1 of the opposite protomer of $M^{pro}_{+A(-1)}$, Phe140 and Glu166 are tied together by these hydrogen bonds to form parts of the floors of the S1 specificity pockets. These hydrogen bonds also conformationally anchor Phe140, as indicated by its close-to-average B-factors (Figure 5(d)), and make the π-stacking of its phenyl ring with the imidazole ring of His163 resistant to disruption. As a result, the catalytically competent conformation of the active sites and the S1 specificity pockets of $M^{pro}_{+A(-1)}$ is rigidified; the position of the equilibrium described by equation (1) would shift to the left. This probably serves as an additional explanation for the predominant S1 specificity of M^{pro} or $M^{pro}_{+A(-1)}$ for Gln, but not for Glu. Comparison of the structure of the unbound M^{pro} with that of the APE-bound M^{pro}, both in space group $C2$ at pH 6.5,[29] does not show an obvious reduction in the B-factors of residues Lys137 to Ser144 upon the binding of the APE to M^{pro}, probably because

Phe140 and Glu166 of the parent protomer, and Ser1 of the opposite protomer in the unbound M^{pro} interact normally and immobilize these residues.

In protomer B of the M^{pro}:CMK complex, the binding of the CMK does not induce the recovery of the catalytically competent conformation of the active site and the S1 specificity pocket of M^{pro} (Table 2). The side-chain of P1-Gln of the CMK cannot be accommodated by the collapsed S1 specificity pocket, so it protrudes into the solvent.[25] This could be attributed to the covalent bond between the Cys145 S^{γ} atom of M^{pro} and the methylene C atom of the CMK, whose rotation can orient the side-chain of P1-Gln of the CMK towards the solvent, thereby avoiding the steric hindrance due to the collapsed S1 specificity pocket. In protomer A of the M^{pro}:CMK complex, although the active site and the S1 specificity pocket of M^{pro} are in the catalytically competent conformation (Table 2), allowing the side-chain of P1-Gln of the CMK to occupy the S1 specificity pocket of M^{pro}, the unoccupied space in the rest of the substrate-binding region of M^{pro} is large enough for the conformational rearrangements in the rest of the CMK, resulting in an unexpected binding mode of the CMK to protomer A of M^{pro} as well.[25] These unexpected binding modes are not observed in the crystal structures of the complexes of M^{pro} (both the wild-type and the variants) with the APE,[29] nor with the peptidomimetic inhibitors,[26,27] probably because their P' ester groups sterically hinder the rotation of the C-S covalent bond formed between the peptidases and the inhibitors. In the case of the APE, the π-conjugation of the N^{α} atom and the carbonyl group of P1-AGln restricts the orientation of the side-chain of P1-AGln to point towards the S1 specificity pockets of M^{pro} or $M^{pro}_{+A(-1)}$. In the case of the peptidomimetic inhibitors, the C atom attacked by the Cys145 S^{γ} atom corresponds to the carbonyl C atom of the scissile peptide bond of a substrate. Therefore, the resulting C-S bond pulls the inhibitor towards the back wall of the substrate-binding region. This reinforces the steric hindrance due to

the P' ester group of the inhibitor against the rotation of the C-S bond.

Thus, a refined explanation could be provided for the pH-dependence of the activity of M^pro[12,18,25,36]: At low pH, a significant number of the M^pro protomers are in the catalytically incompetent conformation. However, these protomers are conformationally flexible; the recovery of their catalytically competent conformation can be induced by the binding of a substrate. The activity of M^pro is still compromised, because of the energetic cost associated with this induced-fit binding. It would be interesting to determine this energetic cost quantitatively, and compare it with the experimentally determined decrease in the activity of M^pro. Before carrying out such studies, the possibility of a second mechanism, such as the change in the protonation/deprotonation state of the catalytic dyad with pH, being involved in the pH-dependence of the activity of M^pro should not be ruled out.

Domain III of SARS-CoV M^pro

The compositions of the dimer interfaces observed in the structures reported here are essentially identical with those observed in all of the previously reported structures. Extensive studies have been carried out on the role of domain III in the dimerization of M^pro. With domain III truncated, the protomer of M^pro does not dimerize over a wide range of concentrations,[10,14] whereas domain III alone forms a stable dimer even at very low concentrations.[10] Truncation of the final C-terminal helix (residues 293–306) strongly disfavors the dimerization of M^pro;[14,43] analysis of the dimer interfaces observed in the structures of M^pro found that most of the residues of domain III involved in dimer interactions are parts of the final C-terminal helix. Superpositions of the previously reported structures of M^pro in space group $P2_1$ at pH 6.0, 7.6 and 8.0,[25] show that domain III of both independent protomers rotates in a rigid-body manner as the pH increases from 6.0 to 8.0, whereas the remainder of the M^pro dimer remains essentially in the same atomic positions (Figure 10). Domain III of protomer B rotates through a slightly greater angle than does domain III of protomer A. This rotation was regarded as a consequence of the recovery of the catalytically competent conformation of the active sites and the S1 specificity pockets of M^pro as the pH increased from 6.0 to 8.0. However, the structure of M^pro in space group $P2_1$ at pH 6.0 reported here, having the active sites and the S1 specificity pockets of both protomers in the catalytically competent conformation, shows good agreement in the atomic positions of domain III of both independent protomers with the previously reported structure of M^pro in space group $P2_1$ at pH 6.0, but not with those at pH 7.6 or pH 8.0. The interactions between the domain-III residues of opposite protomers, and the interactions of the residues of domain III with those of domains I and II of the same protomer have been analyzed, but the mechanics of this rotation is still unclear. The structures of M^pro in the other space groups have not been studied over such a wide range of pH. It would be interesting to explore whether this rotation is an artifact associated with crystal packing in space group $P2_1$, or whether similar rotations occur in the structures of M^pro in the other space groups as well, indicating some biochemical significance of this rotation. For example, the rotation may represent a mechanism underlying the pH-dependence of the dimerization behaviors of M^pro.[12]

Materials and Methods

Preparation of SARS-CoV M^pro, M^pro_{+A(-1)} and APE

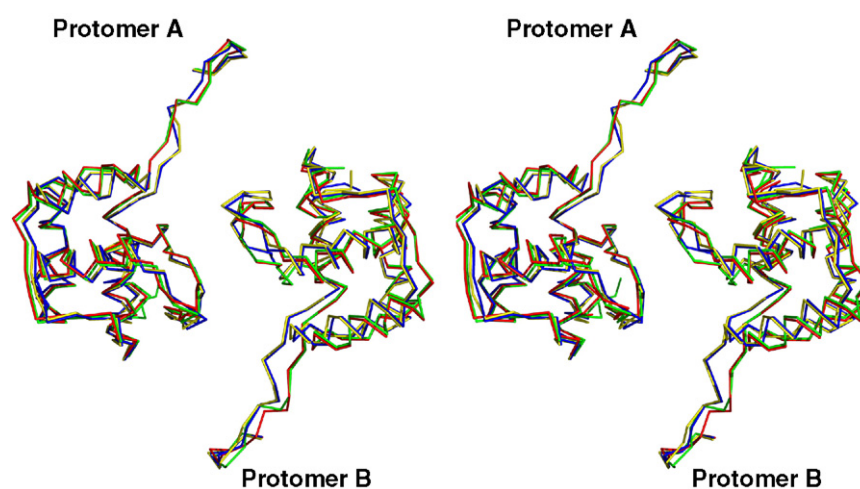SARS-CoV M^pro_{+A(-1)} was cloned, over-expressed and purified as described.[24] A clone expressing M^pro was

Figure 10. C^α traces showing the orientations of domain III of the previously reported structures of SARS-CoV M^pro in space group $P2_1$ at pH 6.0 (red), 7.6 (yellow) and 8.0 (blue)[25], and the structure of M^pro in space group $P2_1$ at pH 6.0 reported here (green). Domain III of both independent protomers rotates as pH increases from 6.0 to 8.0.

generated using oligonucleotide-directed evolution to delete the codon corresponding to the N-terminal Ala of $M^{pro}_{+A(-1)}$. Using this clone, M^{pro} was over-expressed and purified essentially as described for $M^{pro}_{+A(-1)}$. Cbz-Leu-Phe-AGln-(S,S)EP-COOEt was synthesized using the methods established to synthesize other APEs[30,32,33] with minor modifications.

Crystallization, crystal soaking and cryo-protection

Before crystallization, both SARS-CoV M^{pro} and $M^{pro}_{+A(-1)}$ were dialyzed against 20 mM Tris–HCl (pH 7.5), 20 mM NaCl and concentrated to 10 mg/ml. All crystals were grown at ambient temperature by the hanging-drop, vapor-diffusion method. For the $P2_1$ crystals, the reservoir solution contained 0.1 M Mes (pH 6.0), 2% (w/v) polyethylene glycol 20,000, 3% (v/v) dimethyl sulfoxide, and 1 mM dithiothreitol. The final crystallization drop contained equal amounts of the M^{pro} solution and the reservoir solution. Crystals of a thick-plate habit grew in three to five days to a size of about 0.3 mm×0.3 mm× 0.1 mm. For the $P4_32_12$ crystals, the growth conditions were the same as those for the $P2_12_12_1$ crystals;[29] the reservoir solution contained 0.1 M Mes (pH 6.5), 50 mM ammonium acetate, 6% (w/v) polyethylene glycol 8000, 3% (v/v) ethylene glycol, and 1 mM dithiothreitol. The drop contained equal amounts of the $M^{pro}_{+A(-1)}$ solution and the reservoir solution. Bipyrimidal crystals grew in three to five days to a size of about 0.1 mm×0.1 mm× 0.1 mm. Macroseeding was carried out to improve the quality of the crystals. Crystals of good quality were selected and soaked overnight in drops having the same compositions as their reservoir solutions plus the APE chosen for this study present at 3 mM. Cryo-protectants had essentially the same compositions as reservoir solutions, except for the inclusion of 25% (v/v) ethylene glycol and the exclusion of dimethyl sulfoxide and dithiothreitol. Crystals were soaked for about 10 s and then immediately shock-cooled in liquid nitrogen for storage and shipment to the synchrotron beamline.

Data collection and processing, structure solution, refinement and analysis

The X-ray diffraction data from all crystals were collected at the synchrotron Beamline 8.3.1 (equipped with an ADSC-Q210 CCD detector) at the Advanced Light Source in the Lawrence Berkeley National Laboratory. All data sets were indexed, scaled and merged using DENZO and SCALEPACK.[44] Structure solution and refinement were carried out in CCP4.[45,46] All structures were solved by the molecular replacement method using MOLREP,[47] with the structure of unbound SARS-CoV M^{pro} in space group $C2$ (PDB accession code 2A5A) being used as the search model.[29] In the structure of the $M^{pro}_{+A(-1)}$:APE complex, APE was located as outstanding electron densities in the substrate-binding region of $M^{pro}_{+A(-1)}$ in both the $|F_o|-|F_c|,\alpha_c$ (contoured at 3σ and 4σ) and the $2|F_o|-|F_c|,\alpha_c$ (contoured at 1σ) maps. A restraint was applied to the covalent bond between the Cys145 S^γ atom of $M^{pro}_{+A(-1)}$ and the epoxide C3 atom of the APE, on the basis that the length of a C-S single bond is normally about 1.8 Å. All structures were then iteratively refined using REFMAC,[48] and manually adjusted when needed using XtalView/Xfit.[49] The stereochemical qualities of the final structures were assessed using PROCHECK.[50] Graphical representations of the structures were prepared using PyMOL‡. Superimpositions of the various structures were carried out using ALIGN,[51,52] based on the main-chain atoms (amide N, C^α, and carbonyl C and O). The surface areas of the structures were calculated using NACCESS§ Dimer interactions and $M^{pro}_{+A(-1)}$:APE interactions were analyzed using DIMPLOT,[53] and LIGPLOT,[53] respectively.

Protein Data Bank accession codes

The atomic coordinates and the structure factors of all the structures have been deposited in the RCSB Protein Data Bank. The accession code is 2GT7 for the structure of the unbound SARS-CoV M^{pro}, 2GT8 for the structure of the unbound $M^{pro}_{+A(-1)}$, and 2GTB for the structure of the $M^{pro}_{+A(-1)}$:APE complex.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/ j.jmb.2006.11.078

References

1. Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S. *et al.* (2003). A novel coro-

‡ http://www.pymol.org/
§ http://wolf.bms.umist.ac.uk/naccess/

navirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953–1966.

2. Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S. *et al.* (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976.

3. Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W. *et al.* (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*, **361**, 1319–1325.

4. Groneberg, D. A., Hilgenfeld, R. & Zabel, P. (2005). Molecular mechanisms of severe acute respiratory syndrome (SARS). *Respir. Res.* **6**, 8.

5. Bartlam, M., Yang, H. & Rao, Z. (2005). Structural insights into SARS coronavirus proteins. *Curr. Opin. Struct. Biol.* **15**, 664–672.

6. Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001). Viral replicase gene products suffice for coronavirus discontinuous transcription. *J. Virol.* **75**, 6676–6681.

7. Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S. *et al.* (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305–2315.

8. Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853–879.

9. 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.

10. Shi, J., Wei, Z. & Song, J. (2004). Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* **279**, 24765–24773.

11. Kuo, C. J., Chi, Y. H., Hsu, J. T. & Liang, P. H. (2004). Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. Biophys. Res. Commun.* **318**, 862–867.

12. Chou, C. Y., Chang, H. C., Hsu, W. C., Lin, T. Z., Lin, C. H. & Chang, G. G. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry*, **43**, 14958–14970.

13. Chen, S., Chen, L., Tan, J., Chen, J., Du, L., Sun, T. *et al.* (2005). Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization: biochemical and thermodynamic investigation in conjunction with molecular dynamic simulations. *J. Biol. Chem.* **280**, 164–173.

14. Hsu, W. C., Chang, H. C., Chou, C. Y., Tsai, P. J., Lin, P. I. & Chang, G. G. (2005). Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *J. Biol. Chem.* **280**, 22741–22748.

15. Ziebuhr, J., Heusipp, G. & Siddell, S. G. (1997). Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J. Virol.* **71**, 3992–3997.

16. 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.

17. 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.

18. Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. (2004). 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry*, **43**, 4568–4574.

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

20. Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. (2004). Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry*, **43**, 4906–4912.

21. 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.

22. Wu, C. Y., Jan, J. T., Ma, S. H., Kuo, C. J., Juan, H. F., Cheng, Y. S. *et al.* (2004). Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl Acad. Sci. USA*, **101**, 10012–10017.

23. Kao, R. Y., Tsui, W. H., Lee, T. S., Tanner, J. A., Watt, R. M., Huang, J. D. *et al.* (2004). Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics. *Chem. Biol.* **11**, 1293–1299.

24. Blanchard, J. E., Elowe, N. H., Huitema, C., Fortin, P. D., Cechetto, J. D., Eltis, L. D. & Brown, E. D. (2004). High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chem. Biol.* **11**, 1445–1453.

25. 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.

26. 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.

27. 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.

28. Chen, L., Gui, C., Luo, X., Yang, Q., Gunther, S., Scandella, E. *et al.* (2005). Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. *J. Virol.* **79**, 7095–7103.

29. 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.

30. Asgian, J. L., James, K. E., Li, Z. Z., Carter, W., Barrett, A. J., Mikolajczyk, J. *et al.* (2002). Aza-peptide epoxides: a new class of inhibitors selective for clan CD cysteine proteases. *J. Med. Chem.* **45**, 4958–4960.

31. Barrett, A. J. & Rawlings, N. D. (2001). Evolutionary lines of cysteine peptidases. *Biol. Chem.* **382**, 727–733.

32. James, K. E., Gotz, M. G., Caffrey, C. R., Hansell, E., Carter, W., Barrett, A. J. *et al.* (2003). Aza-peptide epoxides: potent and selective inhibitors of Schistosoma mansoni and pig kidney legumains (asparaginyl endopeptidases). *Biol. Chem.* **384**, 1613–1618.

33. James, K. E., Asgian, J. L., Li, Z. Z., Ekici, O. D., Rubin, J. R., Mikolajczyk, J. *et al.* (2004). Design, synthesis, and evaluation of aza-peptide epoxides as selective

and potent inhibitors of caspases-1, -3, -6, and -8. *J. Med. Chem.* **47**, 1553–1574.

34. Powers, J. C., Asgian, J. L., Ekici, O. D. & James, K. E. (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* **102**, 4639–4750.

35. Hegyi, A., Friebe, A., Gorbalenya, A. E. & Ziebuhr, J. (2002). Mutational analysis of the active centre of coronavirus 3C-like proteases. *J. Gen. Virol.* **83**, 581–593.

36. Tan, J., Verschueren, K. H., Anand, K., Shen, J., Yang, M., Xu, Y. *et al.* (2005). pH-dependent conformational flexibility of the SARS-CoV main proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *J. Mol. Biol.* **354**, 25–40.

37. Xu, T., Ooi, A., Lee, H. C., Wilmouth, R., Liu, D. X. & Lescar, J. (2005). Structure of the SARS coronavirus main proteinase as an active C2 crystallographic dimer. *Acta Crystallog. sect. F*, **61**, 964–966.

38. Hunter, C. A., Lawson, K. R., Perkins, J. & Urch, C. J. (2001). Aromatic interactions. *J. Chem. Soc., Perkin Trans.* **2**, 651–669.

39. Burley, S. K. & Petsko, G. A. (1986). Dimerization energetics of benzene and aromatic amino acid sidechains. *J. Am. Chem. Soc.* **108**, 7995–8001.

40. Serrano, L., Bycroft, M. & Fersht, A. R. (1991). Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *J. Mol. Biol.* **218**, 465–475.

41. Burley, S. K. & Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science*, **229**, 23–28.

42. 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.

43. Shi, J. & Song, J. (2006). The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *FEBS J.* **273**, 1035–1045.

44. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

45. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.

46. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallog. sect. D*, **59**, 1131–1137.

47. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022–1025.

48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.

49. McRee, D. E. (1999). XtalView/Xfit–a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.

50. 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.

51. Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 A. *J. Mol. Biol.* **190**, 593–604.

52. Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst.* **30**, 1160–1161.

53. 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.

54. Brünger, A. T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Edited by M. Guss