Mechanism for Controlling the Dimer-Monomer Switch and Coupling Dimerization to Catalysis of the Severe Acute Respiratory Syndrome Coronavirus 3C-Like Protease[▽]

Jiahai Shi,[1] J. Sivaraman,[1]* and Jianxing Song[1,2]*

Department of Biological Sciences, Faculty of Science, National University of Singapore,[1] and Department of Biochemistry, Yong Loo Lin School of Medicine and National University of Singapore,[2] 10 Kent Ridge Crescent, Singapore 119260, Republic of Singapore

Unlike 3C protease, the severe acute respiratory syndrome coronavirus (SARS-CoV) 3C-like protease (3CLpro) is only enzymatically active as a homodimer and its catalysis is under extensive regulation by the unique extra domain. Despite intense studies, two puzzles still remain: (i) how the dimer-monomer switch is controlled and (ii) why dimerization is absolutely required for catalysis. Here we report the monomeric crystal structure of the SARS-CoV 3CLpro mutant R298A at a resolution of 1.75 Å. Detailed analysis reveals that Arg298 serves as a key component for maintaining dimerization, and consequently, its mutation will trigger a cooperative switch from a dimer to a monomer. The monomeric enzyme is irreversibly inactivated because its catalytic machinery is frozen in the collapsed state, characteristic of the formation of a short 3_{10}-helix from an active-site loop. Remarkably, dimerization appears to be coupled to catalysis in 3CLpro through the use of overlapped residues for two networks, one for dimerization and another for the catalysis.

Severe acute respiratory syndrome (SARS) is the first emerging infectious disease of the 21st century and is caused by a novel coronavirus termed SARS-CoV. It suddenly broke out in 2002 and then rapidly spread to 32 countries, causing ~8,500 infections and over 900 deaths (http://www.who.int/csr/sars /en/). Despite being contained by the summer of 2003, several infections were reported later, a warning that SARS may return. So far, neither a vaccine nor an efficacious therapy has been available. Therefore, it is urgently necessary to design the potential therapeutic agents against SARS.

The SARS-CoV 3C-like protease (3CLpro), also known as the main protease (Mpro), plays a vital role in processing two viral polyproteins, pp1a (486 kDa) and pp1ab (790 kDa), into active subunits required for genome replication and transcription. As a consequence, it has been validated to be a key target for development of antiviral therapies (3). Although the CoV 3CLpro is so named to reflect the similarity of its active site to that of the picornavirus 3C protease, 3CLpro acquires an extra C-terminal domain in addition to the chymotrypsin fold adopted by the 3C protease to host the complete catalytic machinery (1, 2, 31). Intriguingly, unlike 3C protease, only the homodimeric form of the SARS-CoV 3CLpro has been characterized to be catalytically active (10, 25, 26) and the unique extra domain was found to play a key role in controlling dimer-monomer equilibrium (7, 11, 14, 15, 25, 26). However, despite intense studies of the SARS-CoV 3CLpro (2, 6, 7, 10, 11, 13–15, 19, 20, 23, 25–27, 28, 29, 31, 32), two puzzles still remain unsolved: (i) how the dimer-monomer switch is controlled and (ii) why dimerization is absolutely essential for catalysis. To best address these two issues, a high-resolution monomeric structure is highly demanded.

The SARS-CoV 3CLpro assembles into a homodimer with an interface area of over 1,000 Å^2, and dissociation constants have been estimated to range from ~nM to μM by different groups (7, 10, 11, 13–16, 19, 20, 27, 28, 31). Previously we have conducted a systematic mapping of the interfacial residues critical for dimerization, and the results led to a proposal that the residues critical for maintaining dimerization and regulating catalysis are organized as a nano-scale-channel-like network (25). Noticeably, a recent study suggests that the residues for controlling dimerization of the SARS-CoV 3CLpro might not be limited just to the interfacial ones, because Ser147 located away from the interface was also identified to be crucial for dimerization (4). In our previous study, the residue Arg298, located at the end of the extra domain, was also uncovered to be exceptionally important for both dimerization and enzymatic activity. Replacement of Arg298 by Ala suddenly rendered the enzyme into an inactive and monomeric form (25). However, due to the absence of a high-resolution three-dimensional structure of the monomeric enzyme, it is difficult to rationalize how a point mutation is capable of abolishing the SARS-CoV 3CLpro dimer with a relatively large interface and high affinity. Moreover, it is completely unclear how the enzyme rearranges its structure in response to the loss of dimerization as well as what structural changes are accountable for inactivating the catalytic machinery in the monomeric enzyme.

In the present study, we succeeded in determining the crystallographic structure of the monomeric R298A mutant at a resolution of 1.75 Å. Detailed analysis of this structure in conjunction with previously determined dimeric forms offers

* Corresponding author. Mailing address: Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Republic of Singapore. Phone for Jianxing Song: 65 6516 1013. Fax: 65 6779 2486. E-mail: bchsj@nus.edu.sg. Phone for J. Sivaraman: 65 6516 1163. Fax: 65 6779 5671. E-mail: dbsjayar@nus .edu.sg.

key insights into why the extra domain is able to mediate the dimerization as well as how the dimerization is coupled to the catalysis of the SARS-CoV 3CLpro.

MATERIALS AND METHODS

Generation of the recombinant R298A mutant. Recently the extra N-terminal residues left over from the fusion proteins were found to induce the collapsed conformation of the active site; mutagenesis was thus performed on our previous plasmid (25) to convert the thrombin cleavage site (CTG GTT CCG CGT GGA TCC [LVPRGS]) to a factor Xa cleavage site (GAA GGT CGT [IEGR]) by using two primers, 5′-GGC GAC CAT CCT CCA AAA TCG GAT CTG ATC GAA GGT CGT AGT GGT TTT AGG AAA ATG GCA TTC CCG-3′ (forward) and 5′-CGG GAA TGC CAT TTT CCT AAA ACC ACT ACG ACC TTC GAT CAG ATC CGA TTT TGG AGG ATG GTC GCC-3′ (reverse). The GST-fused R298A protease was expressed in *Escherichia coli* strain BL21(DE3) with induction by 0.4 mM isopropyl-1-thio-D-galactopyranoside. The R298A protease was obtained by in-gel cleavage with factor Xa of the GST-R298A protein attached to glutathione Sepharose beads (Amersham Biosciences), followed by a fast protein liquid chromatography purification on a gel filtration column (HiLoad 16/60 Superdex 200). With the factor Xa cleavage, except for the R298A mutation site, the R298A protease studied here has a sequence identical to that of the authentic SARS-CoV 3CLpro whose structure was newly reported (Protein Data Bank code 2H2Z) (29). The molecular weight of the R298A protein was determined by a Voyager STR MALDI-TOF mass spectrometer (Applied Biosystems) using a single protease crystal (data not shown).

Analytical ultracentrifuge. Sedimentation velocity experiments were done using a Beckman Coulter XL-I analytical ultracentrifuge. Briefly, the R298A protease was dialyzed against 10 mM Tris-HCl (pH 7.5) containing 14.4 mM β-mercaptoethanol and subsequently concentrated to 3 mg/ml for preparing samples at various protease concentrations. For a sedimentation velocity experiment, sample (400 μl) and reference (440 μl) solutions were loaded into standard double-sector centerpieces and mounted in a Beckman An-50 Ti rotor. Experiments were conducted at 20°C with a rotor speed of 42,000 rpm. The UV absorbance of the samples was monitored in a continuous mode without delay, and a step size of 0.003 cm was used without averaging. To avoid imprecise UV absorbance readings at 280 nm for samples at high concentrations, a wavelength of 290 nm was used to monitor the velocity experiments at protein concentrations of 2 and 3 mg/ml, while 280 nm was used for those at lower concentrations (0.2, 0.5, 1.0, and 1.5 mg/ml). Multiple scans at different time points were fitted into a continuous size distribution using SEDFIT 9.4c (24). A partial specific volume of 0.7311 cm^3/g was calculated from the amino acid sequence of the R298A protease; the solvent density and viscosity were also calculated by use of SEDNTERP (http://www.jphilo.mailway.com/).

Crystallization, structure determination, and analysis of the R298A protease. The R298A protease with a concentration of 10 mg/ml was crystallized in a 2-μl hanging drop using a condition slightly modified from that previously reported (29), namely 6% polyethylene glycol 6000, 2% (±)-2-methyl-2,4-pentanediol, 5% isopropanol, and 0.1 M morpholineethanesulfonic acid at pH 6.0. A single plate-like crystal (0.4 mm by 0.1 mm by 0.005 mm) was picked up from the crystal clusters for diffraction after a 3-day growth period, and then the coverslip was moved to a new well with buffer (0.1 M morpholineethanesulfonic acid, 10% polyethylene glycol 6000, pH 6.0) for 10 more days. The X-ray diffraction data were collected at beam line X29C, National Synchrotron Light Source, Brookhaven National Laboratory, using a Q315 charge-coupled-device detector (Area Detector Systems Corp., Poway, CA) at a wavelength of 1.0810 Å.

The obtained data set was processed using the program HKL2000 (22). Initially, an attempt to solve the structure using a single protomer from the native 3CLpro structure (Protein Data Bank code 1UJ1) as a search model by molecular replacement was unsuccessful. As such, we have subsequently treated the catalytic fold (residues 12 to 180) and the extra domain (residues 200 to 300) as two independent search models in the program Phaser (21) by assuming that these two domains are in a different disposition compared to the native 3CLpro structure, which readily gave the structure solution. Further, model building was done using the program COOT (9) and was followed by refinement using the program CNS (5). Finally, 457 well-defined water molecules were added, and refinement was continued until the *R* value converged to 0.168 (R_{free} = 0.208) for all reflections up to a 1.75-Å resolution. The final model contained all 306 residues of the mutant R298A protease and had good stereochemistry, with all residues falling within the allowed regions of the Ramachandran plot except for Tyr154, which was located at a tight turn as analyzed by PROCHECK (18). Details of the data collection and refinement statistics are presented in Table 1.

TABLE 1. Crystallographic data and refinement statistics for the mutant R298A

Parameter	Value(s)[b]
Data collection	
Wavelength (Å)	1.0810
Resolution limit (Å)	50–1.75 (1.81–1.75)
Space group	P2$_1$
Cell parameters	
a (Å)	33.39
b (Å)	66.23
c (Å)	62.21
β (°)	100.77
No. of observed reflections	186,691
No. of unique reflections	26,899
Completeness (%)	99.4 (95)
Redundancy	6.9 (4.5)
Linear R-factor	0.067 (0.245)
Overall *I*/(*I*)	12.3 (2.2)
Refinement	
Resolution range (Å)	20.0–1.75 (1.77–1.75)
R_{work}[c]	0.168 (0.179)
No. of reflections	23,600
R_{free}[d]	0.209 (0.187)
No. of reflections	1,452
RMSD[e] bond lengths (Å)	0.008
RMSD[e] bond angles (°)	1.43
Ramachandran plot	
Favored (%)	89.4
Allowed (%)	9.4
Generously allowed (%)	0.8
Disallowed (%)	0.4[a]

[a] Only one residue, Tyr154, is in the disallowed region. This residue is located at the tight turn region, which leads to the abnormal dihedral angle.
[b] Values in parentheses are for the highest-resolution shell.
[c] R_{work} = Σ|F$_{obs}$ − F$_{calc}$|/ΣF$_{obs}$, where F$_{calc}$ and F$_{obs}$ are the calculated and observed structure factor amplitudes, respectively.
[d] R_{free} was calculated in the same manner as R_{work}, but for 6% of the total reflections chosen at random and omitted from refinement.
[e] RMSD, root mean square deviation.

The structure overlay was done by LSQKAB and CCP4mg from the CCP4 suite (8). The extra domain motion analysis was determined by DynDom 1.5 (12). The volume of the active pocket was calculated using SURFNET (17), and all the figures were prepared using the PyMOL molecular graphics system (W. L. DeLano, DeLano Scientific LLC, San Carlos, CA).

Protein Data Bank accession number. The atomic coordinates have been deposited in the Protein Data Bank with the code 2QCY.

RESULTS AND DISCUSSION

Crystallization and structure determination. In the present study, we have attempted to crystallize all monomeric mutants that we previously documented (25), and subsequently this has led to the acquisition of high-quality crystals for the R298A mutant. This success thus allowed us to determine the monomeric crystal structure of the SARS-CoV 3CLpro in the space group P2$_1$ at a resolution of 1.75 Å (Table 1) with one molecule per asymmetry unit. The new construct of the R298A mutant with the authentic sequence was also found to be completely inactive even at high protein concentrations. On the other hand, the present observation that the R298A mutant exists as a monomer in crystal is consistent with the results that we previously obtained in solution by gel filtration chromatography and dynamic light scattering (25), as well as by analytical
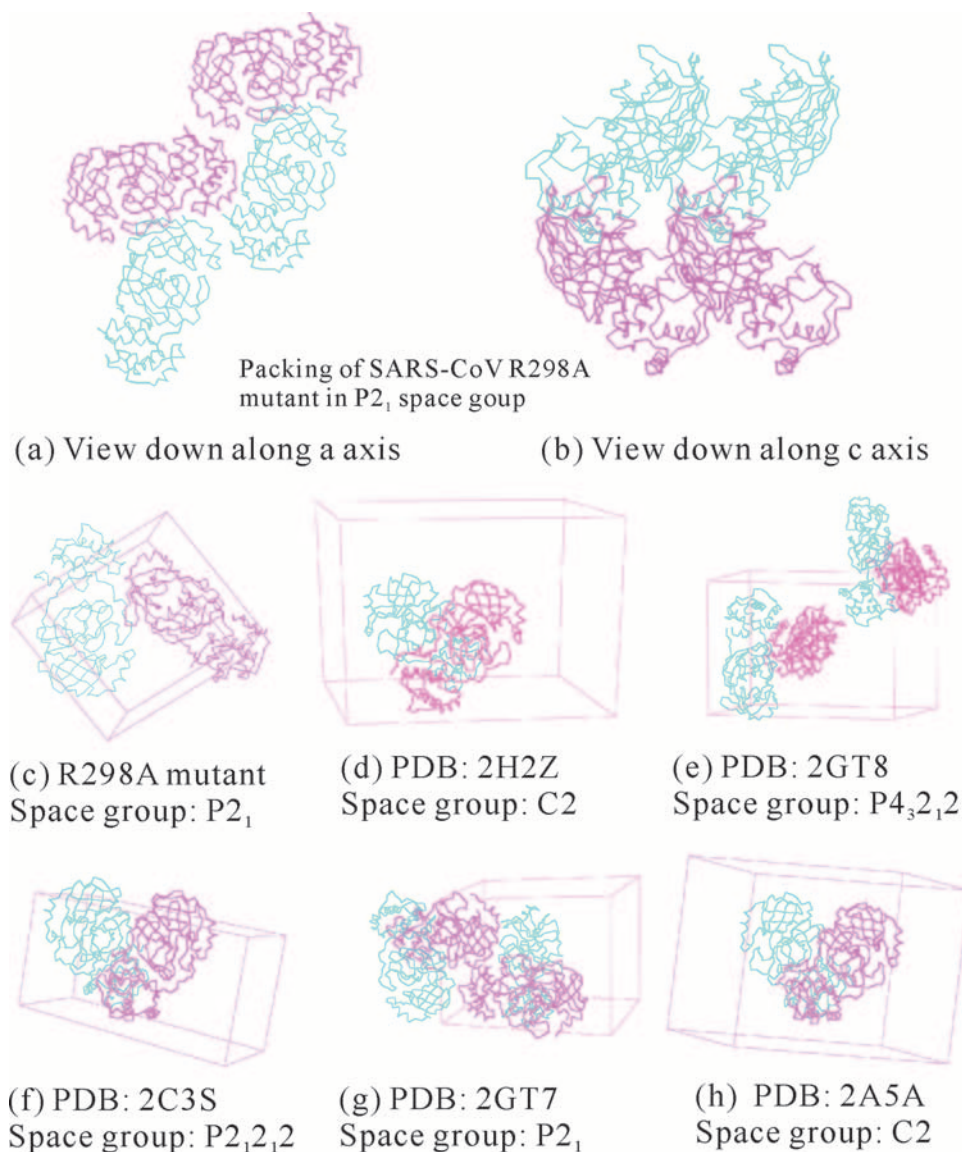
FIG. 1. Crystal packing of the R298A mutant and comparison with other native SARS-CoV 3CLpro structures. Crystal packing at plane bc (a) and at plane ab (b). In a unit cell, the arrangement of the molecules of the R298A mutant in space group $P2_1$ (c) is different from the previously reported native SARS-CoV 3CLpro structures in space group C2 (d and h), $P4_32_12$ (e), $P2_12_12$ (f), and $P2_1$ (g). For the native protease structures, two molecules form a biological dimer in the unit cell. One molecule is magenta, while the other molecule in crystallography twofold symmetry is cyan.

ultracentrifuge (see below). The R298A crystal structure showed that no biological dimer unit like that of the wild-type 3CLpro could be formed in the unit cell even by considering the crystallography twofold symmetry between the adjacent asymmetric units (Fig. 1). Moreover, the present crystal packing has never been found in any previously reported structures of the SARS-CoV 3CLpro crystallized in various space groups, including $P2_1$.

Analytical ultracentrifuge characterization. To further characterize the actual status of the R298A mutant in solution, analytical ultracentrifuge experiments were carried out to measure sedimentation coefficients at six different protein concentrations. Analysis of the sedimentation velocity data collected at all these concentrations gave rise to peaks at ~2.8S by use of

the continuous sedimentation coefficient distribution $[c(s)]$ model (Fig. 2). Moreover, based on the continuous molar mass distribution $[c(M)]$ model, the molecular masses of the R298A mutant at different concentrations were determined to range from 31.1 to 33.7 kDa. Data fitting both the monomer and monomer-dimer equilibrium models were tested, but only the monomer model was found to be valid. Most important, even at a protease concentration of up to 3 mg/ml (89.6 μM), no peak could be detected at ~4S, which was expected for the wild-type 3CLpro dimer. The results clearly indicate that the R298A mutant exists as a complete monomer in solution even at a concentration of 3 mg/ml, completely in agreement with the crystallographic result of the R298A mutant remaining as a monomer in the saturation crystallization condition with a
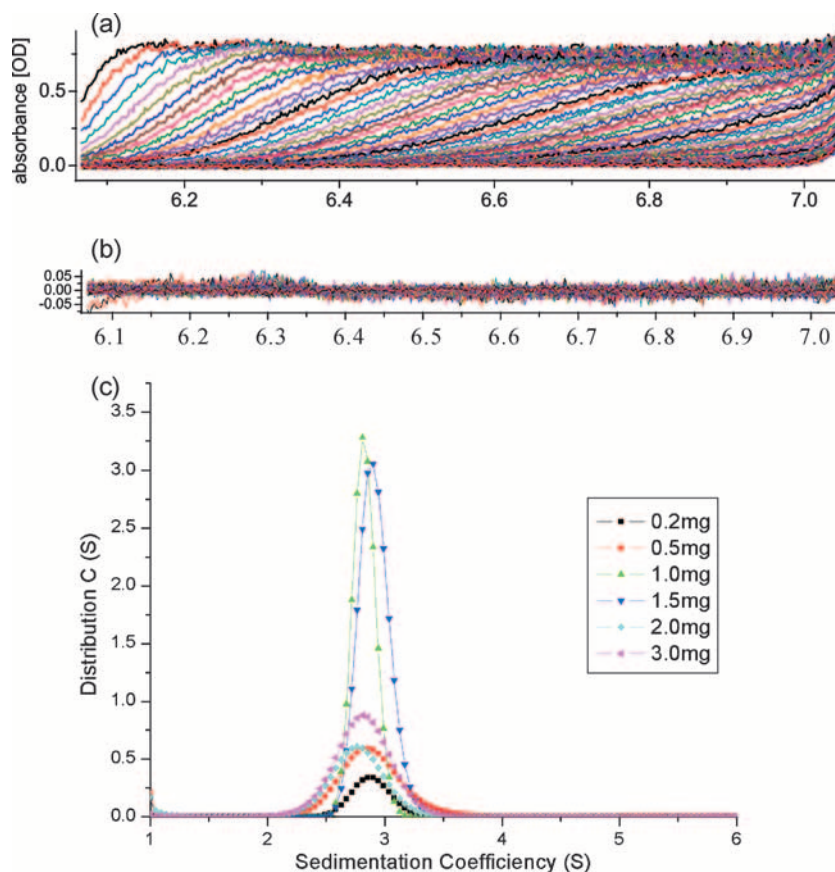
FIG. 2. Sedimentation velocity ultracentrifugation of the R298A mutant. The sedimentation experiments were carried out with a Beckman Coulter XL-I analytical ultracentrifuge at 20°C and 42,000 rpm at protease concentrations from 0.2 to 3 mg/ml. A wavelength of 290 nm was used to monitor the velocity experiments at protein concentrations of 2 and 3 mg/ml, while 280 nm was used for those at lower concentrations (0.2, 0.5, 1.0, and 1.5 mg/ml). (a) Sedimentation velocity absorbance trace of the R298A mutant (1 mg/ml) at 280 nm. OD, optical density. (b) Residuals of the experimental fit of the R298A mutant at 1 mg/ml (29.8 μM). (c) Continuous sedimentation coefficiency (S) distributions [$c(s)$] at six different protease concentrations (from 0.2 to 3 mg/ml).

protein concentration at least higher than that of the initial solution used for crystallization (10 mg/ml).

Structural comparison. Remarkably, the R298A mutant still adopts the characteristic architecture common to all coronavirus 3CLpros, with all residues (1 to 306) well defined in the electron density maps, as exemplified in Fig. 3A. However, as seen in Fig. 3B, due to the large conformational changes and a 40-degree reorientation between the catalytic fold (residues 12 to 180), holding the entire catalytic dyad and a major portion of the substrate binding pocket, and the extra domain (residues 200 to 306) in the R298A structure, it is impossible to superimpose its full structures with other dimeric 3CLpro structures. However, a domain-based superposition with a fully active 3CLpro structure that has the authentic sequence and a pH of 6.0 (Protein Data Bank code of 2H2Z) (29) reveals that three regions undergo radical conformational changes in the monomeric R298A structure: namely the N and C termini, as well as the loop (181 to 199) connecting the catalytic fold and extra domain (Fig. 3). On the other hand, a comparison of the tertiary contact maps of the mutant R298A and the active enzyme (Protein Data Bank code 2H2Z) structures shows that almost no alternation of the packing pattern occurs within the catalytic fold and extra domain (data not shown). Surprisingly,

although the mutation R298A is located on the extra domain (200 to 300), this domain has a backbone root mean square deviation of 0.41 Å, much lower than 1.08 Å for the catalytic fold (12 to 190). It is particularly worthwhile to note that even the side-chain orientations of Arg298 in 2H2Z and Ala298 in R298A are highly similar (Fig. 3C). Examination of the residue-specific root mean square deviations indicates that while no large change is detected within the extra domain, relatively large variations can be recognized within the catalytic fold, in particular over several loops constituting the catalytic machinery, including residues 117 to 125, 133 to 144, and 166 to 169 (data not shown). Very interestingly, these segments are close to the dimerization interface in the dimeric structure (Fig. 4A).

How does the R298A mutation trigger a dimer-to-monomer switch? These findings open up an interesting question as to how the R298A mutation can dramatically eliminate dimerization without deconstructing the structural architecture and tertiary packing. Previous characterization suggests that the dimerization interface consists of residues from both the catalytic fold and the extra domain of the SARS-CoV 3CLpro. Both the N and C termini have been found to play a key role in maintaining the dimeric structure (7, 14, 15, 25, 26, 28). In particular, as shown in Fig. 4B, the N terminus (also termed
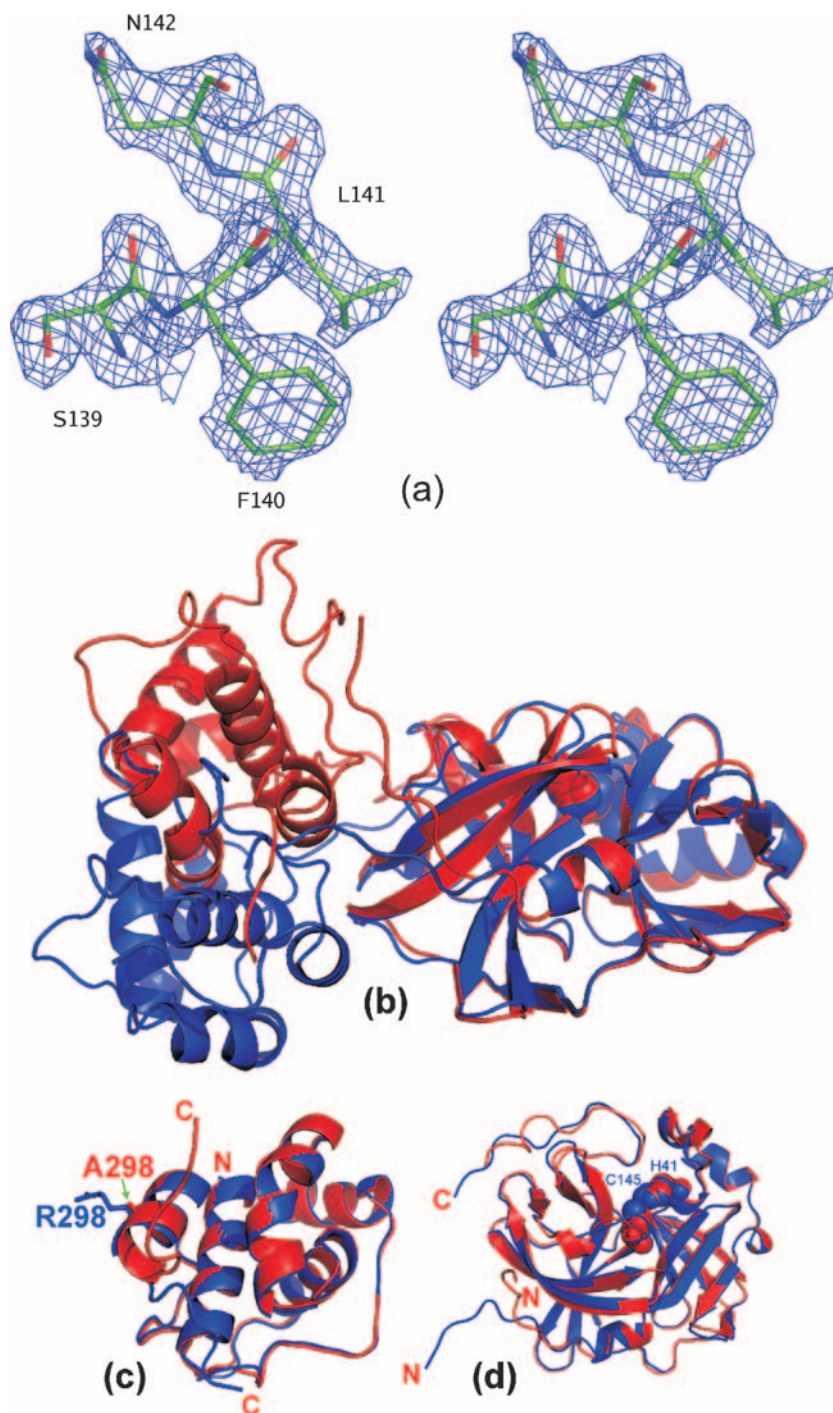
FIG. 3. Crystal structure of the monomeric R298A mutant. (a) Simulated-annealing Fo-Fc map (contoured at 3.5 sigma) for part of the oxyanion loop (138 to 145) in the R298A structure. All residues within 3.5 Å of the Calpha atom of Phe140 were omitted prior to refinement. (b) Complete structures of the mutant R298A (red) and a native SARS 3CLpro (blue) (Protein Data Bank code of 2H2Z) superimposed over the catalytic fold (12 to 180). (c) Comparison of the extra domain (200 to 300) of the two structures. The side chains of the mutation site are displayed as sticks. C, C terminus; N, N terminus. (d). Comparison of the catalytic fold (200 to 300) with the N terminus and the connecting loop (181 to 199) displayed.

the N finger) not only extends beyond the catalytic fold to pack with residues on the extra domain of the same protomer but also forms extensive interactions with residues of the opposite protomer. Of these interprotomer interactions, two have been

documented to significantly stabilize the dimeric structure, namely a salt bridge between the side chains of Arg4 and Glu290 of the opposite protomer (7) and a hydrophobic-aromatic interaction between the side chains of Met6 and Tyr126
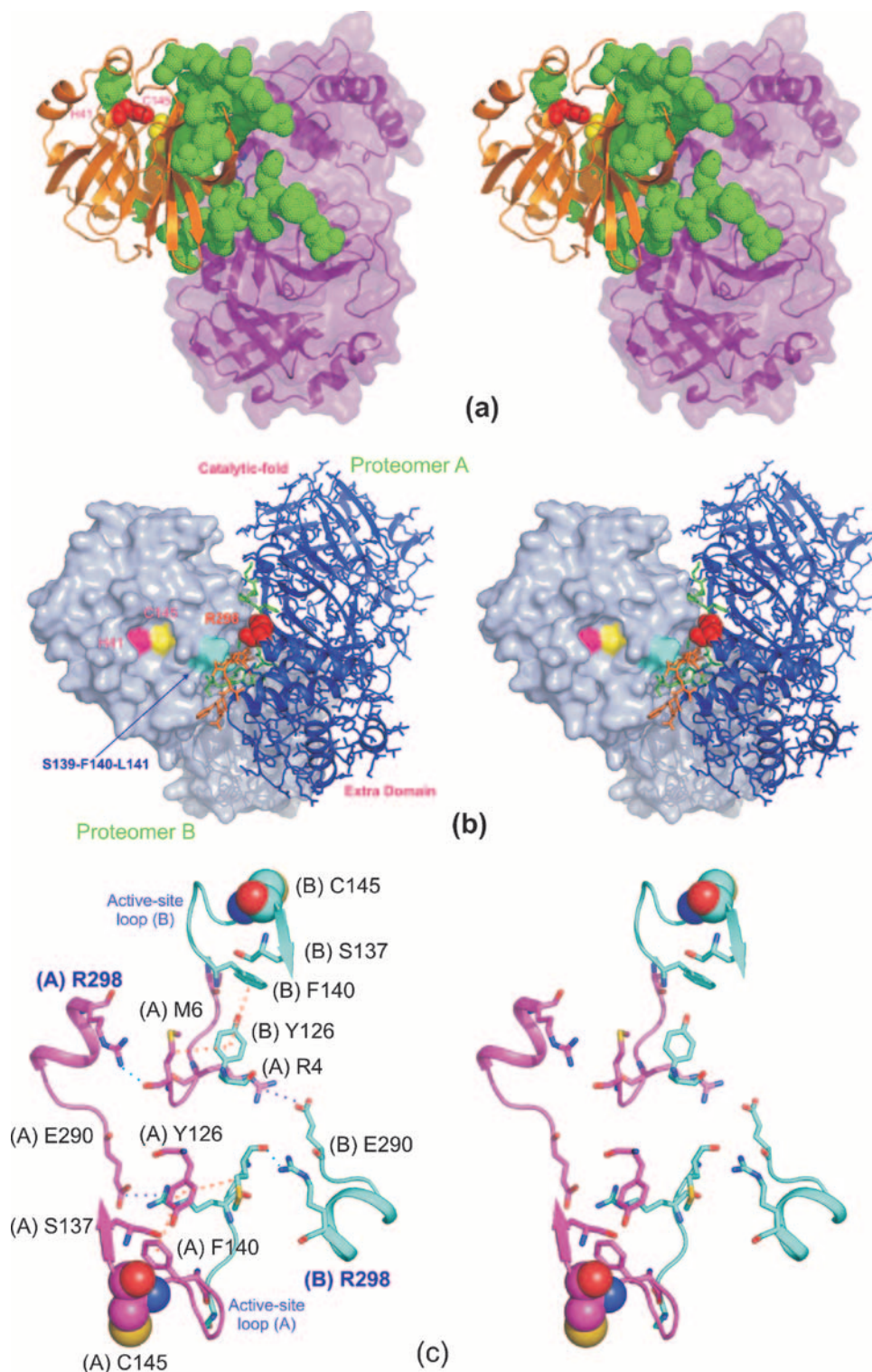
FIG. 4. Structural features associated with dimerization. (a) A stereoview of significantly perturbed residues within the region (1 to 200) of the mutant R298A structure. These residues are mapped to one protomer (brown ribbon) of the dimeric structure (Protein Data Bank code 2H2Z) and displayed as green dots. Another protomer is shown as the violet ribbon and surface. The active-site residues His41 and Cys145 are shown as red and yellow spheres, respectively. (b). A stereoview of the dimeric structure (Protein Data Bank code 2H2Z) with one protomer in ribbon and another in surface modes. The mutation residue R298 is highlighted as red spheres. The N terminus of protomer A is shown as green sticks, while the C terminus is shown as brown sticks. The part of the oxyanion loop Ser139-Phe140-Leu141 which is converted into a short 3_{10}-helix in the mutant R298A structure is labeled and shown in cyan. The active-site residue His41 is pink, and Cys145 is yellow. (c) A stereoview of an interaction network responsible for maintaining the dimeric structure (Protein Data Bank code 2H2Z), with one protomer in purple and another in cyan. Hydrogen bonds are indicated by cyan dashed lines, hydrophobic interactions by red dashed lines, and salt bridges by blue dashed lines. The active-site Cys145 residues of both protomers are displayed as spheres. Interestingly, it appears that Tyr126 is utilized for maintaining dimerization via an aromatic-hydrophobic interaction with Met6 of the opposite protomer, as well as for stabilizing the catalytic machinery via an aromatic stack interaction with Phe140 of the same protomer.
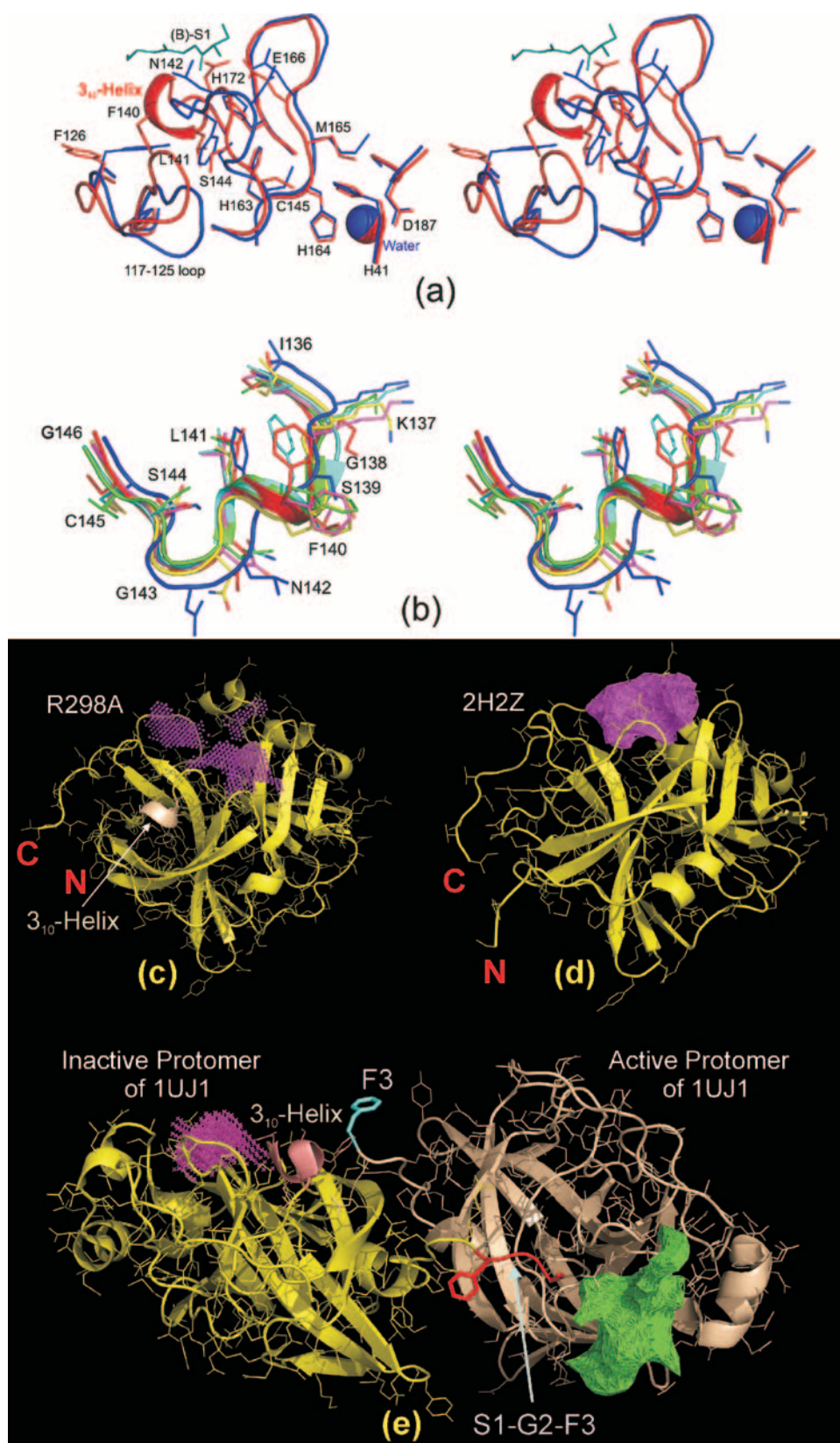
FIG. 5. Structural characteristics of the R298A active site. (a) A stereoview of the substrate-binding pockets of the structures of the mutant R298A (red) and the native enzyme (Protein Data Bank code 2H2Z) (blue). Cyan sticks represent the two N-terminal residues Ser1-Gly2 of the opposite protomer in the dimeric structure (Protein Data Bank code 2H2Z), which is completely lacking in the monomeric R298A structure. The characteristic 3_{10}-helix found in the mutant R298A structure is shown in ribbon mode and labeled. Spheres are used to indicate a buried water

of the opposite protomer (28) (Fig. 4C). Interestingly, mutation of either Met6, Glu290, or Arg298 has been demonstrated to trigger the dimer-to-monomer switch of the enzyme (7, 25, 28). On the other hand, within the same protomer, a hydrogen bond is formed between the side chain NH_2 of Arg298 and the backbone oxygen of Met6. Furthermore, as seen in Fig. 4B, Arg298 is located at the end of the extra domain and folds back to interact with the catalytic fold. As such, it is likely that the residue Arg298 serves as a key component in maintaining dimerization by integrating the N and extra domain of the same protomer together and in this way tremendously stabilizes the precise positioning and orientation of the N finger and extra domain, which are absolutely essential for the dimer formation. Also very strikingly, it appears that the intraprotomer interaction between R298A and Met6 can be further connected to the catalytic machinery of the opposite protomer through a two-step relay: an interprotomer hydrophobic-aromatic interaction between the side chains of Met6 and Tyr126 followed by an intraprotomer aromatic side chain interaction of Tyr126 to Phe140, an essential component of substrate-binding pocket (Fig. 4C). The phenyl ring of Phe140 interacts with the imidazole ring of His163 at the bottom of the S1 specificity pocket, guaranteeing that this histidine remains uncharged over a wide range of pH values. Indeed, likely due to the loss of dimerization, the aromatic ring of Tyr126 undergoes a significant reorientation in the monomeric R298A structure. Nevertheless, the interactions shown in Fig. 4C may represent only part of a large network mediating dimerization, and other residues such as Ser147 are further involved in modulating dimerization via a long-range effect (4).

Since the R298A mutation will release the constrained N finger and C terminus and then switch from a dimer to a monomer, it is not surprising to see large rearrangements over the N and C termini, as well as a 40-degree rotation between the catalytic fold and extra domain in the monomeric R298A structure. Based on the present results, we propose here that although the dimerization interface may be relatively large and its affinity is high, the network mediating dimerization is organized in a way which contains many key components. Disruption of any of them may result in a switch from a dimer to a monomer. This device may offer a significant advantage to facilitating the enzyme to act by following an association-activation-catalysis-dissociation catalytic cycle as previously proposed (6, 13) and/or under extensive regulation by unknown in vivo binding partners (25, 26). On the other hand, this also implies that it would be relatively easier to obtain dimerization inhibitors for designing highly specific bifunctional inhibitors

for the SARS-CoV 3CLpro by linking with the active-site inhibitors in the future (25).

Why is dimerization absolutely essential for catalysis? The CoV 3CLpro shares a similar catalytic machinery with the picornavirus 3C protease, accomplishing catalytic tasks through a catalysis mechanism similar to chymotrypsin (1, 2, 31). However, instead of having the catalytic triad found in chymotrypsin, they possess only a catalytic dyad, composed of the residues His41 and Cys145, with a buried water molecule at the position normally occupied by the side chain of Asp, the third member of the catalytic triad (19, 20, 23, 27, 31, 32). This water molecule forming three hydrogen bonds with His41, His164, and Asp187 is proposed to be a key component of the catalytic machinery. On the other hand, in the 3C protease, the extra domain is completely absent, and also there is no evidence suggesting that dimerization is required to activate its catalysis. Therefore, it is of fundamental interest to understand how in the SARS-CoV 3CLpro, dimerization suddenly becomes coupled to catalysis.

Surprisingly, although the mutation R298A is located at the end of the extra domain, in the monomeric R298A structure, extensive structural variations are observed in the catalytic fold, in particular, the loops constituting the catalytic machinery. This observation is consistent with the previous proposal that the active-site loops of the SARS-CoV 3CLpro are very dynamic and thus sensitive to the variation of environmental conditions (19, 27, 29–31, 32). As seen in Fig. 5A, many residues made up of the catalytic machinery undergo significant changes in the monomeric R298A structure. More specifically, the loop residues over positions 138 to 144 have much more profound changes than other regions, such as the buried water molecule and its hydrogen-bonded His41, His164, and Asp187. This is reasonable because the loop residues, Gly138-Ser139-Phe140-Leu141, themselves directly contact the surface of the opposite protomer. Similarly, the antiparallel β sheet over residues 111 to 129 surrounding the catalytic machinery also has large rearrangements because many residues, such as Tyr126 on this β sheet, also have direct contacts with the other protomer. Nevertheless, as seen in Fig. 5A, the most distinguishable characteristic in the mutant R298A catalytic machinery is the formation of a short 3_{10}-helix by residues Ser139-Phe140-Leu141, which instead adopt a loop conformation in the active enzyme structure (Protein Data Bank code 2H2Z). Most importantly, this chameleon conversion abolishes the key stack interaction between the rings of Phe140 and His163, as well as significantly twists the conformation of the residues Gly143-Ser144-Cys145, thus leading to the collapsed substrate-

binding pocket and oxyanion hole. This amazing observation inspired us to examine all available crystal structures of the SARS-CoV 3CLpro. To our great surprise, the results indicate that in the dimeric structures, this characteristic is absent in the active protomers but can be found in the collapsed protomers. For example, in the dimeric structure (Protein Data Bank code 1UJ1), which possesses a dimeric structure with two different protomers, one catalytically active and another collapsed (31), the conformation assumed by residues 136 to 146 in the active protomer is highly similar to that of the fully active enzyme (Protein Data Bank code 2H2Z), while the conformation in the collapsed protomer highly resembles that of the R298A structure (Fig. 5B). On the other hand, it was also recently shown that the collapsed protomer in the 1UJ1 structure could be successfully converted to the catalytically active one by removing the extra N-terminal residues left over from cleavage of the affinity tag (29). This strongly implies that although it is in the dimeric enzyme, the dynamic catalytic machinery can be trapped in the collapsed conformations by a variety of unfavorable factors. However, the collapsed enzyme still has the potential to be reactivated once the unfavorable factors are removed. By contrast, it appears that if dimerization is lacking, the catalytic machinery will be permanently frozen in the collapsed state, thus leading to an irreversible inactivation of the enzyme as observed on the R298A mutant, which has no catalytic activity even at a very high concentration for both previous constructs (25) as well as the authentic sequence studied here. During a revision of the manuscript, a very unique crystal structure was reported for the avian infectious bronchitis virus 3CLpro (30). In each asymmetric unit, three protease molecules, named A, B, and C, exist. While molecules A and B form a typical catalytically active and symmetrical homodimer like the dimeric SARS-CoV protease, molecule C is not involved in such a dimer. Molecule C was thus proposed to be a monomer frozen in crystal by the binding and fixation of its C terminus into the active site of the dimer formed by molecules A and B. Very interestingly, similar to our present observation for the R298A mutant, this trapped IBV 3CLpro monomer (molecule C) also shows significant differences from molecules A and B over termini and the loop region connecting domains II and III, as well as the active-site pocket. In particular, the short 3_{10}-helix that we found to be characteristic of the collapsed enzymes is also presente in molecule C but not in molecules A and B.

Compared with the collapsed protomers in the dimeric structures, the monomeric R298A structure also has profound structural changes over other regions critical for catalysis. For instance, very large changes are also found over residues 180 to 200 forming the S2 subsite. Therefore, to comprehensively assess the features of the catalytic machinery, we quantitatively calculated the volumes of the active-site cavity of the R298A and other SARS-CoV 3CLpro structures by using the program SURFNET. Interestingly, for the dimeric structure (Protein Data Bank code 1UJ1) composed of two asymmetric protomers, the active protomer without the characteristic 3_{10}-helx has a large and deep cavity over the active site similar to that of the fully active enzyme (Protein Data Bank code 2H2Z). By contrast, the collapsed protomer with the characteristic 3_{10}-helix has an active-site cavity with a significantly reduced volume. More dramatically, the active-site cavity of the R298A

structure is completely fragmentized (Fig. 5C). As a consequence, the catalytic machinery of the monomeric R298A mutant no longer suits the binding substrates and leaves no space to accommodate glutamine at the substrate analog P1 site and a tetrahedral intermediate.

Conclusion. The present study reveals that despite its location at the extra domain, Arg298 is a key component of an integrated network modulating dimerization. By use of overlapped residues such as Tyr126, dimerization may be elegantly coupled to catalysis of the SARS-CoV 3CLpro. As a consequence, the mutation of Arg298 not only eliminates dimerization but also irreversibly inactivates the enzyme by permanently freezing the catalytic machinery in the collapsed state, characteristic of a formation of a short 3_{10}-helix from a chameleon active-site loop. The current results not only lead to the proposal of a possible mechanism of rationalizing the observation that the catalytic machinery of the SARS-CoV 3CLpro is under extensive regulation by the unique extra domain but also bear implications for understanding general principles governing the evolution of the regulatory device for the catalytic machinery by introducing oligomerization.

REFERENCES

1. **Allaire, M., M. M. Chernaia, B. A. Malcolm, and M. N. James.** 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature **369:**72–76.
2. **Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld.** 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J. **21:**3213–3224.
3. **Anand, K., J. Ziebuhr, P. Wadhwani, J. R. Mesters, and R. Hilgenfeld.** 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science **300:**1763–1767.
4. **Barrila, J., U. Bacha, and E. Freire.** 2006. Long-range cooperative interactions modulate dimerization in SARS 3CLpro. Biochemistry **45:**14908–14916.
5. **Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D **54:**905–921.
6. **Chang, H. P., C. Y. Chou, and G. G. Chang.** 2007. Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. Biophys. J. **92:**1374–1383.
7. **Chou, C. Y., H. C. Chang, W. C. Hsu, T. Z. Lin, C. H. Lin, and G. G. Chang.** 2004. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. Biochemistry **43:**14958–14970.
8. **Collaborative Computational Project, Number 4.** 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D **50:**760–763.
9. **Emsley, P., and K. Cowtan.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D **60:**2126–2132.
10. **Fan, K., P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen, and L. Lai.** 2004. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. J. Biol. Chem. **279:**1637–1642.
11. **Graziano, V., W. J. McGrath, L. Yang, and W. F. Mangel.** 2006. SARS CoV main proteinase: the monomer-dimer equilibrium dissociation constant. Biochemistry **45:**14632–14641.
12. **Hayward, S., and R. A. Lee.** 2002. Improvements in the analysis of domain

motions in proteins from conformational change: DynDom version 1.50. J. Mol. Graph. Model. **21**:181–183.

13. **Hsu, M. F., C. J. Kuo, K. T. Chang, H. C. Chang, C. C. Chou, T. P. Ko, H. L. Shr, G. G. Chang, A. H. Wang, and P. H. Liang.** 2005. Mechanism of the maturation process of SARS-CoV 3CL protease. J. Biol. Chem. **280**:31257–31266.

14. **Hsu, W. C., H. C. Chang, C. Y. Chou, P. J. Tsai, P. I. Lin, and G. G. Chang.** 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. **Kuang, W. F., L. P. Chow, M. H. Wu, and L. H. Hwang.** 2005. Mutational and inhibitive analysis of SARS coronavirus 3C-like protease by fluorescence resonance energy transfer-based assays. Biochem. Biophys. Res. Commun. **331**:1554–1559.

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

17. **Laskowski, R. A.** 1995. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J. Mol. Graph. **13**:323–328.

18. **Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. **26**:283–291.

19. **Lee, T. W., M. M. Cherney, C. Huitema, J. Liu, K. E. James, J. C. Powers, L. D. Eltis, and M. N. James.** 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.

20. **Lee, T. W., M. M. Cherney, J. Liu, K. E. James, J. C. Powers, L. D. Eltis, and M. N. James.** 2007. Crystal structures reveal an induced-fit binding of a substrate-like Aza-peptide epoxide to SARS coronavirus main peptidase. J. Mol. Biol. **366**:916–932.

21. **McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni, and R. J. Read.** 2005. Likelihood-enhanced fast translation functions. Acta Crystallogr. D **61**:458–464.

22. **Otwinowski, Z., and W. Minor.** 1997. Processing of x-ray diffraction data collected in oscillation mode, p. 307–326. *In* C. W. Carter, Jr., and R. M. Sweet (ed.), Methods in enzymology, vol. 276. Academic Press, New York, NY.

23. **Pang, Y. P.** 2004. Three-dimensional model of a substrate-bound SARS chymotrypsin-like cysteine proteinase predicted by multiple molecular dynamics simulations: catalytic efficiency regulated by substrate binding. Proteins **57**:747–757.

24. **Schuck, P.** 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. **78**:1606–1619.

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

26. **Shi, J., Z. Wei, and J. Song.** 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.

27. **Tan, J., K. H. Verschueren, K. Anand, J. Shen, M. Yang, Y. Xu, Z. Rao, J. Bigalke, B. Heisen, J. R. Mesters, K. Chen, X. Shen, H. Jiang, and R. Hilgenfeld.** 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.

28. **Wei, P., K. Fan, H. Chen, L. Ma, C. Huang, L. Tan, D. Xi, C. Li, Y. Liu, A. Cao, and L. Lai.** 2006. The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. Biochem. Biophys. Res. Commun. **339**:865–872.

29. **Xue, X., H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, and Z. Rao.** 2007. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J. Mol. Biol. **366**:965–975.

30. **Xue, X., H. Yu, H. Yang, F. Xue, Z. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, X. C. Zhang, M. Liao, M. Bartlam, and Z. Rao.** 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J. Virol. **82**:2515–2527.

31. **Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao.** 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.

32. **Zheng, K., G. Ma, J. Zhou, M. Zen, W. Zhao, Y. Jiang, Q. Yu, and J. Feng.** 2007. Insight into the activity of SARS main protease: molecular dynamics study of dimeric and monomeric form of enzyme. Proteins **66**:467–479.