PROTEIN STRUCTURE REPORT

C-terminal domain of SARS-CoV main protease can form a 3D domain-swapped dimer

Nan Zhong,[1,2†] Shengnan Zhang,[1,2†] Fei Xue,[3†] Xue Kang,[1,2] Peng Zou,[1,2] Jiaxuan Chen,[1,4] Chao Liang,[1] Zihe Rao,[3] Changwen Jin,[1,2,4] Zhiyong Lou,[3]* and Bin Xia[1,2,4]*

[1]Beijing Nuclear Magnetic Resonance Center, Peking University, Beijing 100871, People' Republic of China
[2]College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, People' Republic of China
[3]Structural Biology Laboratory, Tsinghua University, Beijing 100084, China
[4]College of Life Science, Peking University, Beijing 100871, People' Republic of China

Received 20 November 2008; Revised 8 January 2009; Accepted 15 January 2009
DOI: 10.1002/pro.76
Published online 10 February 2009 proteinscience.org

Abstract: SARS coronavirus main protease (M^{pro}) plays an essential role in the extensive proteolytic processing of the viral polyproteins (pp1a and pp1ab), and it is an important target for anti-SARS drug development. We have reported that both the M^{pro} C-terminal domain alone (M^{pro}-C) and the N-finger deletion mutant of M^{pro} (M^{pro}-Δ7) exist as a stable dimer and a stable monomer (Zhong et al., J Virol 2008; 82:4227-4234). Here, we report structures of both M^{pro}-C monomer and dimer. The structure of the M^{pro}-C monomer is almost identical to that of the C-terminal domain in the crystal structure of M^{pro}. Interestingly, the M^{pro}-C dimer structure is characterized by 3D domain-swapping, in which the first helices of the two protomers are interchanged and each is enwrapped by four other helices from the other protomer. Each folding subunit of the M^{pro}-C domain-swapped dimer still has the same general fold as that of the M^{pro}-C monomer. This special dimerization elucidates the structural basis for the observation that there is no exchange between monomeric and dimeric forms of M^{pro}-C and M^{pro}-Δ7.

Keywords: SARS; main protease; domain-swapping; dimerization; C-terminal domain

Introduction

SARS coronavirus (SARS-CoV) was identified as the etiological agent of the pandemic transmissible disease, severe acute respiratory syndrome.[1–3] The SARS-CoV 5′ two-thirds genome of the virus encodes two overlapping polyproteins, pp1a (486 kDa) and pp1ab (790 kDa), which are proteolytically processed into 16 matured nonstructural proteins (nsp1-16) by two proteases included in these two polyproteins. These nonstructural proteins mediate the viral replication and transcription.[4] Main protease of SARS-CoV (M^{pro})

[†]Nan Zhong, Shengnan Zhang, and Fei Xue authors contributed equally to the work.

Grant sponsor: 973 Program; Grant number: 2003CB514104; Grant sponsor: NSFC; Grant number: 30125009; Grant sponsor: 863 Program; Grant number: 2006AA02A323.

*Correspondence to: Zhiyong Lou, Laboratory of Structural Biology, Tsinghua University, Beijing 100084, People's Republic of China. E-mail: louzy@xtal.tsinghua.edu.cn or Bin Xia, Beijing Nuclear Magnetic Resonance Center, Peking University, Beijing 100871, People s Republic of China. E-mail: binxia@pku.edu.cn

Published by Wiley-Blackwell. © 2009 The Protein Society

plays an important role in the extensive proteolytic processing of the viral polyproteins, which makes it essential for the viral life cycle and represents an attractive target for antiviral agent development.[5–7]

The first structure of M^{pro} was solved in 2003 and revealed a homodimer which is highly similar to other previously reported coronavirus main proteases structures.[5,8] It has been reported that M^{pro} exists in solution as an equilibrium between monomeric and dimeric forms,[9] and only the dimeric form is enzymatically active.[10] Shi et al.[11] have reported that the M^{pro} N-terminal alone is a monomer, whereas the M^{pro} C-terminal domain alone (M^{pro}-C) exists as a stable dimer. However, our previous studies demonstrated that M^{pro}-C exists as both a stable monomer and a stable dimer in solution, and so does the N-finger deletion mutant of M^{pro} (M^{pro}-Δ7) which can also form a stable dimer through dimerization of the C-terminal domain.[12]

Here, we report structures of monomeric and dimeric forms of the C-terminal domain of M^{pro} (M^{pro}-C). M^{pro}-C monomer maintains the same fold as that in the crystal structure of M^{pro}. On the other hand, the M^{pro}-C dimer has a novel structure characterized by 3D domain-swapping, which provides the structural basis for the dimer stability.

Table I. *Restraints and Structural Statistics for M^{pro}-C Monomer*

Distance restraints		
Intra-residue	1492	
Sequential ($\|i-j\|=1$)	656	
Medium ($1<\|i-j\|<5$)	449	
Long-range ($\|i-j\|>5$)	620	
Ambiguous	823	
Total	4040	
Hydrogen bond restraints		
Hydrogen bond	32	
Dihedral angle restraints		
Φ	58	
Ψ	58	
Total	116	
Violation statistics		
NOE violation (>0.2 Å)	1	
Torsion angle violation (>2°)	0	
Ramachandran plot statistics		
Residues in most favored regions (%)	85.3	
Residues in allowed regions (%)	13.7	
Residues in generously allowed regions (%)	0.7	
Residues in disallowed regions (%)	0.3	
RMSD from mean structure (Å)		
	Backbone	Heavy atoms
Global	3.7 ± 0.8	4.0 ± 0.7
Structured region	0.6 ± 0.1	0.9 ± 0.1
Regular secondary structure	0.20 ± 0.05	0.62 ± 0.08

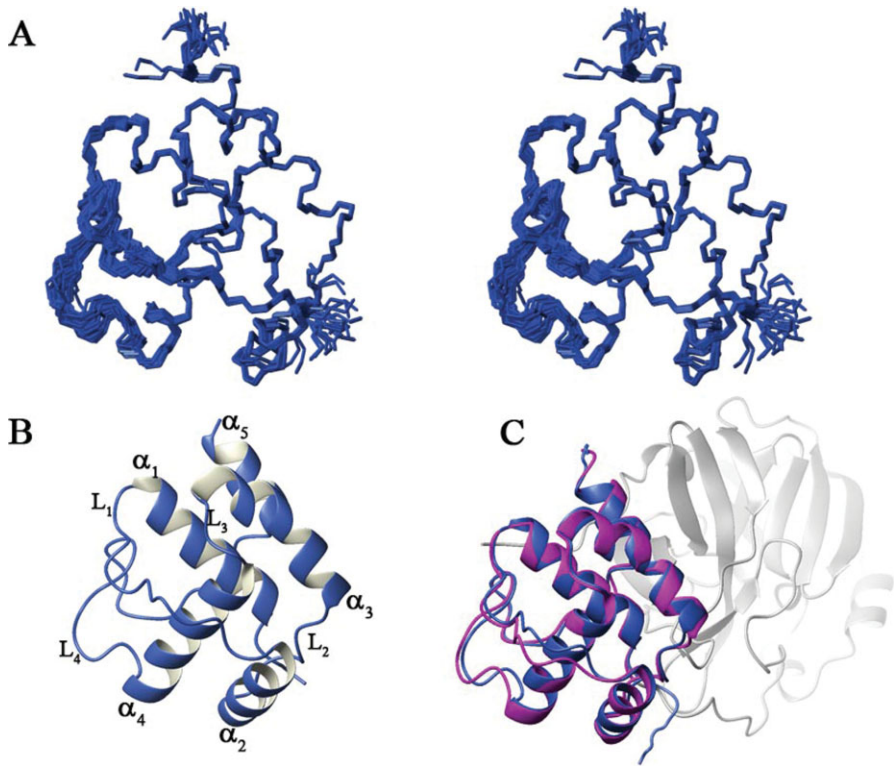

Figure 1. Solution structure of M^{pro}-C monomer. (A) Superimposition of the backbone trace of 20 representative structures; (B) Ribbon diagram of the mean structure with secondary structural elements labeled; (C) Comparison of the structure of M^{pro}-C monomer (blue) with its corresponding part in the crystal structure of M^{pro} (1UK2, magenta). domains I and II of M^{pro} are shown in light gray.

Results

Solution structure of M^{pro}-C monomer

We have obtained nearly complete backbone resonance assignments for the M^{pro}-C monomer except those for residues F219 and E288, and more than 95% of side-chain resonances were assigned. The solution structures of the M^{pro}-C monomer were calculated using interproton NOE-derived distance restraints together with the dihedral angle and hydrogen bond restraints (Table I). The 20 structures with the lowest energies are shown in Figure 1(A), together with the ribbon diagrams of the mean structure [Fig. 1(B)]. The Ramachandran plot indicates that a majority of residues (99.7%) have their Phi and Psi angles in allowed regions, and only 0.3% of them are in disallowed region. The root mean square deviation (RMSD) for backbone heavy atoms in secondary structure elements is 0.20 ± 0.05 Å and that for all heavy atoms in secondary structure elements is 0.62 ± 0.08 Å (Table I). All these indicate that the solution structure is determined with good quality.

Just as the C-terminal domain in the full-length M^{pro}, the M^{pro}-C monomer adopts a globular all-alpha fold consisting of five α-helices (α_1 (T201-N214), α_2 (L227-Y237), α_3 (Q244-T257), α_4 (V261-N274), and α_5 (P293-Q299)), two well-defined loops (L_2 (N238-T243) and L_3 (G258-A260)), and two flexible loops (L_1 (D216-T226) and L_4 (G275-T292)). Helix α_1 is enwrapped by helices α_2,α_3,α_4, and α_5, along with L_4, which forms the hydrophobic core [Fig. 1(B)]. The mean structure of the M^{pro}-C monomer overlaps well with the corresponding part in the structure of M^{pro} (1UK2) except for loops L_1 and L_4, with a 0.5 Å RMSD for backbone heavy atoms in secondary structure elements [Fig. 1(C)]. As loops L_1 and L_4 are relatively flexible in solution structure, the structural difference for these two loops is not unexpected. Therefore, the structure of M^{pro}-C monomer should remain the same as that of the C-terminal domain of M^{pro}.

Crystal structure of M^{pro}-C dimer

As the NMR data quality for the M^{pro}-C dimer is very poor, we have determined its crystal structure with a resolution of 2.4 Å. The crystal structure of the M^{pro}-C dimer was determined by molecular replacement using the structure of the C-terminal 120 residues from M^{pro} crystal structure as the search model (Table II).

The N-terminal 11 residues and C-terminal 8 or 9 residues in two protomers are invisible on the electron density map. To our surprise, the structure of the M^{pro}-C dimer is characterized by 3D domain-swapping with the two helices α_1 of the two molecules interchange their positions (Fig. 2). Each helix α_1 is now surrounded by helices α_2–α_5 of the other molecule and these five helices from two molecules adopts a compact globular all-alpha fold the same as that of M^{pro}-C monomer, which is called a "folding subunit." Thus, the M^{pro}-C dimer is consisted of two identical folding subunits and each is reconstituted by chain fragments from both protomers. In other words, helix α_1 and helices α_2–α_5 from one protomer are now located in different folding subunits for the M^{pro}-C dimer, and the two protomers are symmetrically related by a crystallographic twofold axis. The two folding subunits of the M^{pro}-C dimer are linked by two hinge loops consisting of residues D216-T226 which form loop L_1 in the structure of the M^{pro}-C monomer. As a result, a new short 3_{10} helix is formed by residues Trp218, Phe219, and Leu220 on both hinge loops. Consequently, helix α_4, which is covered by loop L_1 in the M^{pro}-C monomer, is now exposed in both folding subunits of the M^{pro}-C dimer [Fig. 2(D)].

According to the nomenclature of Liu et al.,[13] the interface between domains of a domain-swapped oligomer with the structural characteristics presented in the monomer is termed the closed interface and that found only in the oligomer is termed the open interface. The M^{pro}-C dimer has an extensive close

Table II. *X-Ray Diffraction Data and Model Refinement Statistics for M^{pro}-C dimer*

Parameters	
Data collection statistics	
Cell parameters	a = 51.4 Å, b = 51.3 Å, c = 51.4 Å, α = 112.2°, β = 112.0°, γ = 104.4°
Space group	*P1*
Wavelength used (Å)	1.5418
Resolution (Å)	50.0 (2.5)[a] – 2.4
No. of all reflections	44,600
No. of unique reflections	15,252
Completeness (%)	96.0 (88.1)
Average $I/\sigma(I)$	27.3 (5.2)
R_{merge}[b] (%)	5.1 (19.0)
Refinement statistics	
No. of reflections used ($\sigma(F) > 0$)	14,727
R_{work}[c] (%)	20.8
R_{free}[c] (%)	24.3
No. of protein residues	402
No. of protein atoms	3162
No. of solvent molecules	60
RMSD bond distance (Å)	0.007
RMSD bond angle (°)	1.322
Average B-value (Å^2)	40.2
Ramachandran plot statistics	
Residues in most favored regions	91.9%
Residues in additionally allowed regions	7.8%
Residues in generously allowed regions	0.3%

[a] Numbers in parentheses are corresponding values for the highest resolution shell.
[b] $R_{merge} = \Sigma_h\Sigma_1 \mid I_{ih} - \langle I_h \rangle \mid / \Sigma_h\Sigma_I \langle I_h \rangle$, where $\langle I_h \rangle$ is the mean of the observations I_{ih} of reflection h.
[c] $R_{work} = \Sigma(||F_p(obs)| - |F_p(calc)||)/ \Sigma|F_p(obs)|$; $R_{free} = R$ factor for a selected subset (5%) of the reflections that was not included in prior refinement calculations.

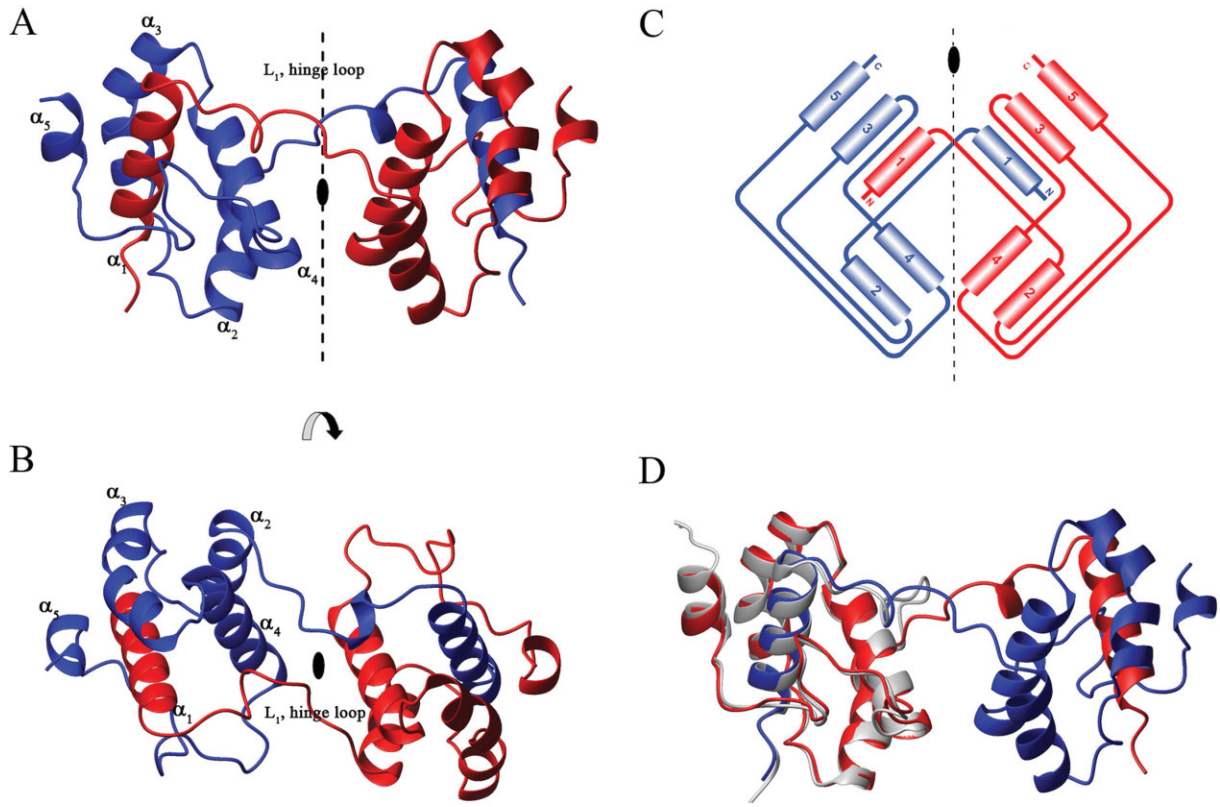

Figure 2. Crystal structure of M^{pro}-C dimer. (A and B). Ribbon diagrams of the 3D domain-swapped dimer are shown with 90° rotation. The secondary structural elements are labeled. The two protomers are shown in red and blue, respectively. (C). Topology diagrams of M^{pro}-C dimer with the same color scheme as in A and B. (D). Superimposition of M^{pro}-C dimer structure with the mean structure of M^{pro}-C monomer. The two protomers of M^{pro}-C dimer are shown in red and blue, respectively. M^{pro}-C monomer is in light grey.

interface of about 3000 Å^2, which involves interchanging of α_1 helices of both protomers and produces tight entanglement of the two protomers. This elucidates the structure basis for the exceeding stability of the M^{pro}-C dimer and the lack of exchange between the monomer and dimer.[12] However, there is no apparent interaction to define an open interface between the two hinge loops. In addition, the electron densities of residues on the hinge loops are relatively weak and their main chain B-factors are relatively high (>50 Å^2), which suggest that the hinge loops linking the two folding subunits are rather flexible. Thus, we expect that the relative orientation of two folding subunits in the M^{pro}-C dimer should not be fixed in solution.

Discussion

We have reported that M^{pro}-C exists in both monomeric and a dimeric forms in solution, and here, we have solved the structures of both forms. We found that the M^{pro}-C monomer structure adopts the same conformation as that of the C-terminal domain in the crystal structure of full-length M^{pro}. However, the M^{pro}-C dimer has an unusual structure which is a 3D domain-swapped dimer. We have carried out denature-refolding experiments and found that both stable monomeric and dimeric forms can be regenerated from refolding of either M^{pro}-C monomer or dimer (data not shown). Also, taking into consideration that both M^{pro}-C monomer and dimer are produced in *E. coli* and there is no exchange between the two, it seems that forming domain-swapped dimer is not dependent on other components of the virus or human cells, but should be an intrinsic ability of M^{pro}-C.

Furthermore, our previous NMR studies demonstrated that the two C-terminal domains of the M^{pro}-Δ7 dimer have identical dimerization pattern to that of the M^{pro}-C dimer.[12] Thus, M^{pro}-Δ7 can also form a domain-swapped dimer through domain-swapping of their C-terminal domains, which also explains why the M^{pro}-Δ7 dimer is stable and does not exchange with the M^{pro}-Δ7 monomer. However, what puzzles us is that we failed to detect any domain-swapped dimer for the full-length M^{pro} expressed in *E. coli*. Why could the full-length M^{pro} not form domain-swapped dimer? Could it be that the N-finger of M^{pro} is able to interfere and prevent the 3D domain-swapping dimerization of the C-terminal domain of M^{pro}? Is there a biological relevance for this 3D domain-swapping? It is clear that the two N-termini of both protomers in the crystal structure of the M^{pro}-C dimer stretch out from the proteins in opposite directions, and the relative orientation of the two folding subunits in the domain-

swapped M^{pro}-C dimer is different from that of the two C-terminal domains in the crystal structure of the full-length M^{pro}. Even if the full-length M^{pro} could form a domain-swapped dimer, the two N-terminal domains of the dimer would be far away, and the domain-swapped dimer could not adopt the active conformation. In other words, a domain-swapped dimer of the full-length M^{pro} would not be active. But, why would the virus retain the C-terminal domain within M^{pro} that has the potential to deactivate the enzyme? These are the questions we will pursue in the future.

Materials and Methods

For the M^{pro}-C, the DNA fragment encoding residues 187–306 was cloned into pET21a vector. Samples of M^{pro}-C monomer and dimer were prepared according to previously published method.[12] All NMR samples were at a concentration of about 1 m*M* and were prepared in buffer containing 50 m*M* potassium phosphate (pH 7.0), 1 m*M* EDTA, 0.03% NaN_3, in 90% H_2O/10% D_2O, plus Complete, an EDTA-free Protease Inhibitor Cocktail (Roche, Germany). All NMR experiments were performed at 298 K on a Bruker Avance 500 MHz (with cryoprobe), 600 MHz NMR, and 800 MHz spectrometers. Resonance assignments were obtained using standard methods.[14] The interproton nuclear Overhauser effect (NOE) was employed to generate the distance restraints. Dihedral angles were determined from backbone chemical shifts using TALOS.[15] Hydrogen bond restraints were generated from the H-D exchange experiments in combined with the CSI secondary structural prediction.[16,17] Structures were calculated and refined using the program CYANA and AMBER.[18–20]

M^{pro}-C dimer sample was concentrated to 30 mg/mL in 20 m*M* Tris pH 7.0. Crystallization was performed by the sitting-drop vapor-diffusion method at 289 K in 48-well plates. The crystals selected for diffraction studies grew in 0.2*M* Sodium chloride, 0.1*M* BIS-TRIS (pH 5.5), 25% w/v PEG3350. A 2.4 Å resolution diffraction data set was collected at 100 K from a single M^{pro}-C dimer crystal using an in-house Rigaku MM-007 generator and an R-Axis VI++ detector. The beam was focused by osmic mirror. A total of 270 frames of data were collected. Processing of diffraction images and scaling of the integrated intensities were performed using the HKL2000 software package.[21] Initial phases were obtained by molecular replacement with PHASER[22] using the crystal structure of SARS-CoV main protease (PDB code: 2H2Z, excluding the N terminal 186 residues) as the searching model. The final manual rebuilding and refinement were performed in COOT,[23] Refmac,[24] and CNS regarding to the *2Fo-Fc* and *1Fo-Fc* density map.

The solution and crystal structures were analyzed using the program packages PROCHECK[25] and Figures were created by using MOLMOL.[26]

Coordinates

Coordinates and structure factors have been deposited at the PDB under accession code 3EBN for the crystal structure of the M^{pro}-C dimer and 2K7X for the solution structure of the M^{pro}-C monomer.

Acknowledgment

All NMR experiments were carried out at the Beijing NMR Center.

References

1. Chan HL, Tsui SK, Sung JJ (2003) Coronavirus in severe acute respiratory syndrome (SARS). Trends Mol Med 9: 323–325.
2. Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, Laman JD, de Jong T, van Doornum G, Lim W, Ling AE, Chan PK, Tam JS, Zambon MC, Gopal R, Drosten C, van der Werf S, Escriou N, Manuguerra JC, Stohr K, Peiris JS, Osterhaus AD (2003) Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet 362: 263–270.
3. Leng Q, Bentwich Z (2003) A novel coronavirus and SARS. N Engl J Med 349:709.
4. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol 331: 991–1004.
5. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science 300: 1763–1767.
6. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, Sun L, Mo L, Ye S, Pang H, Gao GF, Anand K, Bartlam M, Hilgenfeld R, Rao Z (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc Natl Acad Sci USA 100: 13190–13195.
7. Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, Zhao Q, Zhou Z, Pei D, Ziebuhr J, Hilgenfeld R, Yuen KY, Wong L, Gao G, Chen S, Chen Z, Ma D, Bartlam M, Rao Z (2005) Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS Biol 3:e324.
8. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J 21:3213–3224.
9. Graziano V, McGrath WJ, Yang L, Mangel WF (2006) SARS CoV main proteinase: the monomer-dimer equilibrium dissociation constant. Biochemistry 45:14632–14641.
10. Fan K, Wei P, Feng Q, Chen S, Huang C, Ma L, Lai B, Pei J, Liu Y, Chen J, Lai L (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. J Biol Chem 279:1637–1642.
11. 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.
12. Zhong N, Zhang S, Zou P, Chen J, Kang X, Li Z, Liang C, Jin C, Xia B (2008) Without its N-finger, the main protease of severe acute respiratory syndrome coronavirus can

form a novel dimer through its C-terminal domain. J Virol 82:4227–4234.
13. Liu Y, Gotte G, Libonati M, Eisenberg D (2001) A domain-swapped RNase A dimer with implications for amyloid formation. Nat Struct Biol 8:211–214.
14. Bax A (1994) Multidimensional nuclear-magnetic-resonance methods for protein studies. Curr Opin Struct Biol 4:738–744.
15. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13: 289–302.
16. Wishart DS, Sykes BD, Richards FM (1992) The chemical-shift index—a fast and simple method for the assignment of protein secondary structure through NMR-spectroscopy. Biochemistry 31:1647–1651.
17. Wishart DS, Sykes BD (1994) The C-13 chemical-shift index—a simple method for the identification of protein secondary structure using C-13 chemical-shift data. J Biomol NMR 4:171–180.
18. Guntert P, Mumenthaler C, Wuthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298.
19. Herrmann T, Guntert P, Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319: 209–227.
20. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J Comput Chem 26:1668–1688.
21. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter CW, Jr, Sweet RM, Eds. (1997) Macromolecular crystallography, part A. Academic Press, New York, pp 307–326.
22. McCoy A, Grosse-Kunstleve R, Adams P, Winn M, Storoni L, Read R (2007) Phaser crystallographic software. J Appl Crystallogr 40:658–674.
23. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D 60: 2126–2132.
24. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D 53:240–255.
25. Laskowski R, MacArthur M, Moss D, Thornton J (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283–291.
26. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 14:51–55, 29–32.