

Ting Xu,<sup>a</sup> Amy Ooi,<sup>a</sup> Hooi Chen  
Lee,<sup>a</sup> Rupert Wilmoth,<sup>a</sup>  
Ding Xiang Liu<sup>b</sup> and Julien  
Lescar<sup>a\*</sup>

<sup>a</sup>School of Biological Sciences, Nanyang  
Technological University, 60 Nanyang Drive,  
Singapore 637551, Singapore, and <sup>b</sup>Institute of  
Molecular and Cell Biology, Singapore

Correspondence e-mail: julien@ntu.edu.sg

Received 26 August 2005

Accepted 17 October 2005

Online 20 October 2005

**PDB Reference:** SARS coronavirus main  
proteinase, 2c3s, 2c3ssf.

## Structure of the SARS coronavirus main proteinase as an active C<sub>2</sub> crystallographic dimer

The 34 kDa main proteinase (M<sup>Pro</sup>) from the severe acute respiratory syndrome coronavirus (SARS-CoV) plays an important role in the virus life cycle through the specific processing of viral polyproteins. As such, SARS-CoV M<sup>Pro</sup> is a key target for the identification of specific inhibitors directed against the SARS virus. With a view to facilitating the development of such compounds, crystals were obtained of the enzyme at pH 6.5 in the orthorhombic space group *P*2<sub>1</sub>2<sub>1</sub>2 that diffract to a resolution of 1.9 Å. These crystals contain one monomer per asymmetric unit and the biologically active dimer is generated *via* the crystallographic twofold axis. The conformation of the catalytic site indicates that the enzyme is active in the crystalline form and thus suitable for structure-based inhibition studies.

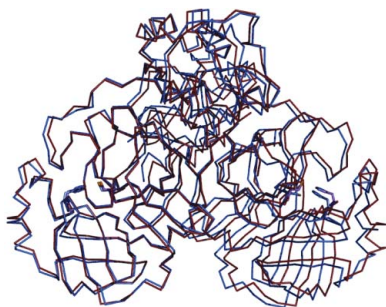
### 1. Introduction

Severe acute respiratory syndrome (SARS) is a severe form of pneumonia. Its transmission pattern, high mortality rate and possible re-emergence in the future make SARS a serious threat for which neither efficient therapy nor vaccine is currently available. The disease is caused by a member of the coronavirus family: the SARS coronavirus (SARS-CoV; Fouchier *et al.*, 2003). Following viral entry into cells, two polyproteins named pp1a and pp1ab, with molecular weights of 486 and 790 kDa, respectively, are synthesized (Rota *et al.*, 2003). During the viral life cycle, pp1a and pp1ab are processed into 15 putative non-structural proteins by two viral proteases: the papain-like protease and the main proteinase M<sup>Pro</sup> (also named the 3C-like protease; 3CL<sup>Pro</sup>; reviewed in Ziebuhr *et al.*, 2000). In SARS-CoV, M<sup>Pro</sup> is responsible for the cleavage of 11 sites in the replicase polyproteins (Snijder *et al.*, 2003), releasing viral enzymes needed for replication, such as the RNA-dependent RNA polymerase and the helicase, as well as other accessory proteins and non-structural proteins the functions of which are not fully understood. Thus, given its pivotal role in the viral life cycle, M<sup>Pro</sup> is an attractive target for the development of drugs directed against the SARS virus. Three-dimensional structures of M<sup>Pro</sup> enzymes have been reported for several coronaviruses including human CoV (HCoV229E; Anand *et al.*, 2003), porcine transmissible gastroenteritis virus (TGEV; Anand *et al.*, 2002) and SARS-CoV (Yang *et al.*, 2003). In this study, using an *Escherichia coli* overexpression system, we purified the SARS-CoV M<sup>Pro</sup> and obtained a novel crystal form at pH 6.5 that diffracts to high resolution and contains one monomer per asymmetric unit. The active-site residues and the oxyanion hole adopt a functional conformation, indicating that this crystal form might be useful for structure-based drug design.

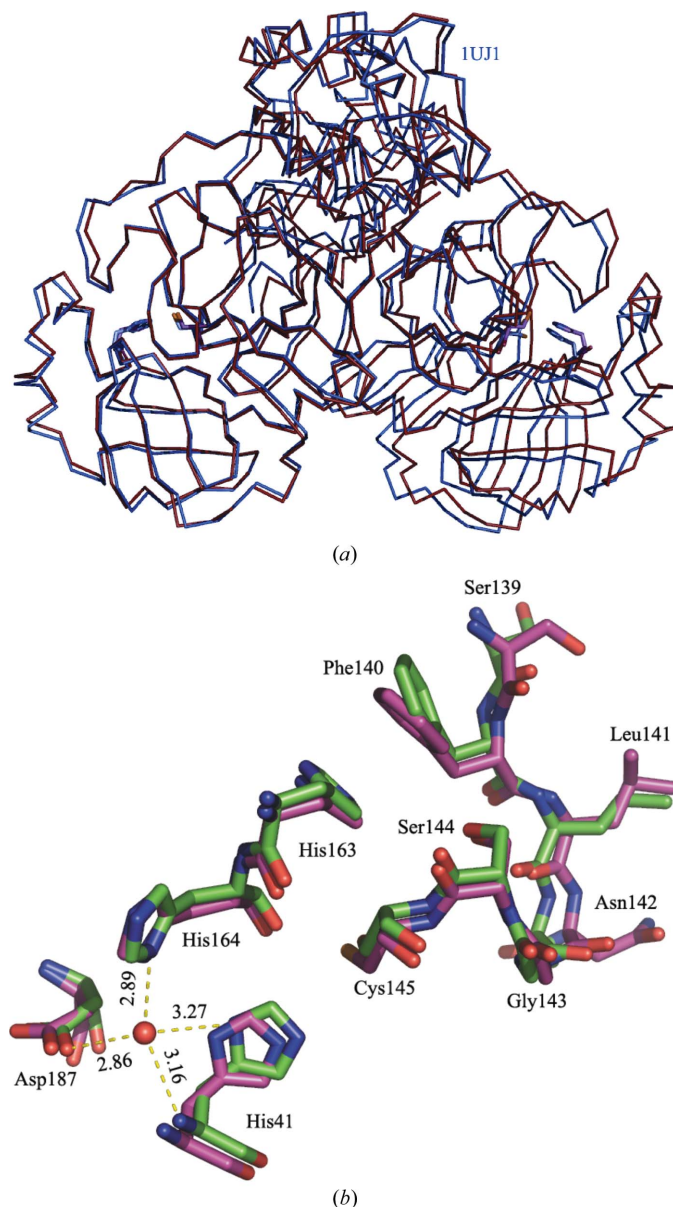
### 2. Experimental

#### 2.1. Protein expression and purification

The DNA fragment encoding the SARS-CoV M<sup>Pro</sup> strain SIN 2774 (Ruan *et al.*, 2003) was amplified by PCR using *Pfu* polymerase (Stratagene) and cloned into pMAL-c2x (New England Biolabs) incorporating the maltose-binding protein (MBP) at the N-terminus of SARS-CoV M<sup>Pro</sup>. The forward primer (5'-TACTAATTGAAGGA GTTCGGGTTTTAGGAAAATGG-3') contains an *Xmn*I site



(bold). The reverse primer (5'-AGCCGGATCCTTATTGGAAGGT AACACCAG-3') contains a *Bam*HI site (bold) downstream of the stop codon TAA. Four additional amino acids (IEGR) were introduced to facilitate the removal of MBP by factor Xa. Transformed BL21(DE3) *E. coli* cells were grown at 310 K in LB media supplemented with 0.2% glucose until an OD<sub>600 nm</sub> of 0.6–0.8 was attained. IPTG was added to a final concentration of 1 mM and the temperature was lowered to 303 K. After 2 h, cells were harvested by centrifugation at 8000g for 10 min, resuspended in buffer A (20 mM Tris–HCl pH 7.4, 50 mM NaCl, 1 mM EDTA) and lysed by sonication for 20 min, followed by centrifugation at 20 000g for 20 min at 277 K. The supernatant was loaded onto an Econo-column (Bio-Rad)



**Figure 1**  
(a) Overall superposition of the C $\alpha$  traces from the SARS-CoV M<sup>Pro</sup> monomer present in our asymmetric unit (coloured red, PDB code 2c3s) with the active monomer A of Yang *et al.* (2003) (shown in blue, PDB code 1uj1, chain A). A residual rotation of 4.5° is needed to then bring the two equivalent monomers B into coincidence. The active-site residues of each monomer are represented as sticks. (b) Detailed view of the active site represented as green sticks (2c3s, this work) superimposed onto the active monomer A of SARS-CoV M<sup>Pro</sup> (1uj1, chain A). The putative hydrogen bonds (dashed lines) formed by the spatially conserved water molecule (red sphere) are shown.

**Table 1**

Data-collection and refinement statistics.

Values in parentheses refer to the highest resolution shell.

| Data-collection statistics                               |  |
|--|--|
| Space group  | <i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2           |
| Unit-cell parameters (Å)                                 | <i>a</i> = 107.7, <i>b</i> = 44.9, <i>c</i> = 54.2 |
| Resolution range (Å)                                     | 28–1.90 (1.95–1.90)                                |
| Unique reflections                                       | 19895  |
| Redundancy   | 8.0 (5.9)  |
| Completeness (%)   | 97.9 (88.2)  |
| <i>I</i> σ( <i>I</i> )                                   | 4.6 (2.2)  |
| <i>R</i> <sub>merge</sub> <sup>†</sup> (%)               | 7.8 (31.4)   |
| <i>V</i> <sub>M</sub> (Å <sup>3</sup> Da <sup>-1</sup> ) | 1.95   |
| Refinement statistics                                    |  |
| <i>R</i> <sub>free</sub> (%)                             | 22.5 (27.5)  |
| <i>R</i> <sub>free</sub> value (%)                       | 26.4 (31.2)  |
| No. of protein atoms                                     | 2302 [301 residues]                                |
| No. of solvent molecules                                 | 211  |
| No. of reflections in working set                        | 19880  |
| No. of reflections in test set                           | 1077   |
| Mean temperature factor (Å <sup>2</sup> )                | 35.21  |
| R.m.s.d. bond lengths (Å)                                | 0.006  |
| R.m.s.d. bond angles (°)                                 | 1.31   |
| R.m.s.d. dihedral angles (°)                             | 24.7   |
| Ramachandran plot  |  |
| Most favoured region (%)                                 | 87.7   |
| Additionally allowed regions (%)                         | 11.1   |
| Generously allowed regions (%)                           | 0.8  |
| Disallowed regions (%)                                   | 0.4  |

$$^{\dagger} R_{\text{merge}} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I, \quad ^{\ddagger} R = \sum |F_{\text{calc}} - F_{\text{obs}}| / \sum |F_{\text{obs}}|.$$

packed with amylose resin (New England Biolab) equilibrated with buffer A and incubated overnight at 277 K. The fusion protein was eluted at 277 K using 20 mM Tris–HCl pH 7.4, 50 mM NaCl, 1 mM EDTA, 10 mM maltose and loaded onto a HiPrep 16/10 Q Sepharose FF column (Amersham) equilibrated with buffer B (20 mM Tris–HCl pH 8.0, 50 mM NaCl, 1 mM EDTA). Proteins were eluted using a linear NaCl concentration gradient in buffer C (20 mM Tris–HCl pH 8.0, 1 M NaCl, 1 mM EDTA). Fractions containing MBP-SARS-CoV M<sup>Pro</sup> were pooled, concentrated by ultrafiltration at 3000g (Centricon, Vivascience) and desalted in 20 mM Tris–HCl pH 7.0, 50 mM NaCl, 1 mM CaCl<sub>2</sub> using PD-10 columns (Amersham). One unit of factor Xa was added per 142 μg of fusion protein for 6 h at 297 K. After cleavage, factor Xa was removed using a resin (Qiagen). Cleaved products were loaded onto an XK 16/20 phenyl Sepharose resin column (Amersham) equilibrated in buffer D (12.5 mM Tris–HCl pH 7.0, 300 mM NaCl, 1 mM DTT, 0.1 mM EDTA). The recombinant SARS-CoV M<sup>Pro</sup> was eluted using buffer E (12.5 mM Tris–HCl pH 7.0, 1 mM DTT, 0.1 mM EDTA). Fractions containing SARS-CoV M<sup>Pro</sup> were pooled and the buffer changed to 10 mM Tris–HCl pH 7.4, 1 mM EDTA, 1 mM DTT for concentration to 5 mg ml<sup>-1</sup> as determined using the Bradford method (Bio-Rad) with BSA as a standard and stored at 193 K.

## 2.2. Crystallization and data collection

Crystals of SARS-CoV M<sup>Pro</sup> were grown using the hanging-drop vapour-diffusion method. Equal volumes (1 μl) of protein and mother liquor were mixed over wells containing 0.1 M MES pH 6.5 and 0.6 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> at 291 K. Macro seeding produced thin elongated plate-like crystals over a period of one week. For data collection, crystals were soaked in a cryoprotecting solution containing 30% glycerol, 0.1 M MES, 0.6 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> pH 6.5, before being mounted and cooled to 100 K in a nitrogen-gas stream (Oxford Cryosystems). Diffraction intensities were recorded at beamline ID14-4 at the European Synchrotron Radiation Facility, Grenoble, France on an ADSC CCD detector using an attenuated beam of 0.125 × 0.050 mm. Integration, scaling and merging of the intensities

were carried out using programs from the CCP4 suite (Collaborative Computational Project, Number 4, 1994). Data-collection and refinement statistics are presented in Table 1.

### 2.3. Structure determination and refinement

The structure of SARS-CoV M<sup>Pro</sup> was readily solved by molecular replacement using the program *AMoRe* from the CCP4 suite with the SARS-CoV M<sup>Pro</sup> structure deposited as PDB code 1uj1 as a search model. The program *REFMAC5* was used for refinement cycles, which were alternated with rebuilding sessions using the program *O* (Jones *et al.*, 1991). 5% of the reflections were set aside to monitor the progress of refinement using the  $R_{\text{free}}$  factor. Water molecules, added automatically using *ARP/wARP* (Perrakis *et al.*, 1999), were checked by visual inspection. The quality of the model was assessed using *PROCHECK* (Laskowski *et al.*, 1993). Structure superposition was performed with *LSQKAB* (Collaborative Computational Project, Number 4, 1994).

## 3. Results and discussion

### 3.1. Overall structure of SARS-CoV M<sup>Pro</sup>

The model comprises one monomer per asymmetric unit (residues 1–301). Five residues from the C-terminus are not visible in the electron-density map and have been omitted. Residues 1–101 (domain I) and 102–184 (domain II) form the chymotrypsin-like double- $\beta$ -barrel structure which is observed in several viral proteases including picornaviruses, togaviruses and flaviviruses (Babe & Craik, 1997). The C-terminal  $\alpha$ -helical domain (residues 201–301) of SARS-CoV M<sup>Pro</sup> is required for activity, since a truncated fragment comprising only its catalytic domain displays a significant decrease in enzymatic activity (Bacha *et al.*, 2004). Structural and functional studies of coronavirus M<sup>Pro</sup> have shown that dimerization is required for maximal protease activity. In this respect, a prominent role is played by the seven amino-terminal amino acids, which adopt an extended conformation making extensive contacts with domain II of the other monomer and ensuring the formation of a catalytically competent active site (Yang *et al.*, 2003). In our crystal form, the active dimer is generated through the crystallographic twofold. No contact is established by Ser1, which is mobile as shown by a higher than average temperature factor. The path of the main chain, however, closely follows that observed in previously reported active monomers, with an r.m.s deviation of 0.80 Å for 300 equivalent main-chain atoms (PDB code 1uj1 chain A; Yang *et al.*, 2003) (Fig. 1a). This latter crystal form belongs to space group  $P2_1$  and contains one dimer in the asymmetric unit with quasi-twofold symmetry. This indicates that the N-terminal residue is not absolutely required for the active site to adopt an active conformation.

A figure showing the distribution of the thermal factors of the SARS-CoV main proteinase is available as supplementary material.<sup>1</sup>

<sup>1</sup>Supplementary material is available from Crystallography Journals Online (Reference: SW5004).

### 3.2. Structure of the active site

The substrate-binding site is located in a cleft between the two  $\beta$ -barrels. The catalytic Cys145–His41 dyad (with the cysteine thiol acting as the nucleophile) is used instead of the classical Ser–His–Asp triad of serine proteases (Fig. 1b). Although the crystals were obtained at pH 6.5, a value which is presumably near the  $pK_a$  value of His residues in the substrate-binding site and where the enzyme shows a slightly reduced activity, the conformation of the active site indicates an active enzyme (Fig. 1b). The immediate vicinity of the active site is involved in extensive intermolecular contacts with neighbouring molecules. Thus, this crystal form is likely to be more suitable for studies involving soaking or co-crystallization of small compounds rather than long peptides. Interestingly, during the course of preparation and submission of this manuscript, related crystal forms of SARS-CoV M<sup>Pro</sup> have been reported by Hsu *et al.* (2005) and by Tan *et al.* (2005).

We are grateful to Dr Ed Liu from the Genome Institute of Singapore for providing us with the clones from the SARS virus. Financial support from the Singapore Biomedical (03/1/21/20/291) and National Medical Research Councils (NMRC/SRG/001/2003 to JL laboratory) is acknowledged as well as access to data-collection facilities at ESRF.

## References

- Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). *EMBO J.* **21**, 3213–3224.
- Anand, K., Ziebuhr, J., Wadhvani, P., Mesters, J. R. & Hilgenfeld, R. (2003). *Science*, **300**, 1763–1767.
- Babe, L. M. & Craik, C. S. (1997). *Cell*, **91**, 427–430.
- Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. (2004). *Biochemistry*, **43**, 4906–4912.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Fouchier, R. A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J., van den Hoogen, B. G., Peiris, M., Lim, W., Stohr, K. & Osterhaus, A. D. (2003). *Nature (London)*, **423**, 240.
- Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T.-P., Shr, H. L., Chang, G. G., Wang, A. H.-J. & Liang, P. H. (2005). *J. Biol. Chem.* **280**, 31257–31266.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- Rota, P. A. *et al.* (2003). *Science*, **300**, 1394–1399.
- Ruan, Y. J. *et al.* (2003). *Lancet*, **361**, 1779–1785.
- Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. & Gorbalenya, A. E. (2003). *J. Mol. Biol.* **331**, 991–1004.
- Tan, J., Verschuere, K. H. G., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B., Mesters, J., Chen, K., Shen, X., Jiang, H. & Hilgenfeld, R. (2005). In the press.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13190–13195.
- Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). *J. Gen. Virol.* **81**, 853–879.