

Structural and biochemical analyses of the tetrameric carboxypeptidase S9Cfn from *Fusobacterium nucleatum*

Xin Wang,^{a‡} Meng-Ting Cheng,^{b‡} Zhi-Peng Chen,^b Yong-Liang Jiang,^b Yu-Shu Ge,^b Rong Xia^{a*} and Wen-Tao Hou^{b*}

Received 7 July 2021

Accepted 19 October 2021

Edited by C. S. Bond, University of Western Australia, Crawley, Australia

‡ These authors contributed equally to this work.

Keywords: carboxypeptidases; serine peptidases; *Fusobacterium nucleatum*; crystallography; oral disease.

PDB reference: carboxypeptidase from *Fusobacterium nucleatum*, 7ep9

Supporting information: this article has supporting information at journals.iucr.org/d

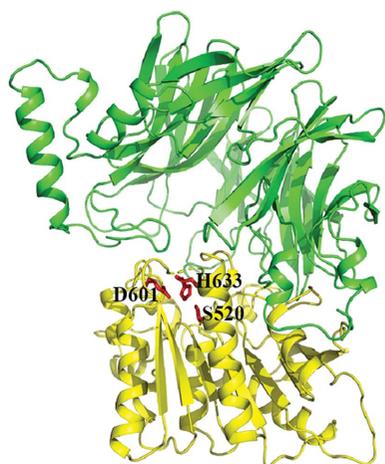
^aDepartment of Stomatology, The Second Affiliated Hospital of Anhui Medical University, Hefei, Anhui 230601, People's Republic of China, and ^bHefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China. *Correspondence e-mail: xiarongqh@aliyun.com, todvince@mail.ustc.edu.cn

As one of the most abundant bacteria in the human oral cavity, *Fusobacterium nucleatum* is closely involved in various oral diseases and is also a risk factor for other diseases. The peptidases of *F. nucleatum* can digest exogenous peptides into amino acids to satisfy its nutrient requirements. Here, a putative *F. nucleatum* peptidase, termed S9Cfn, which belongs to the S9C peptidase family was identified. Enzymatic activity assays combined with mass-spectrometric analysis revealed that S9Cfn is a carboxypeptidase, but not an aminopeptidase as previously annotated. The crystal structure of the S9Cfn tetramer was solved at 2.6 Å resolution and was found to contain a pair of oligomeric pores in the center. Structural analysis, together with site-directed mutagenesis and enzymatic activity assays, revealed a substrate-entrance tunnel that extends from each oligomeric pore to the catalytic triad, adjacent to which three conserved arginine residues are responsible for substrate binding. Moreover, comparison with other S9 peptidase structures indicated drastic conformational changes of the oligomeric pores during the catalytic cycle. Together, these findings increase the knowledge of this unique type of tetrameric carboxypeptidase and provide insight into the homeostatic control of microbiota in the human oral cavity.

1. Introduction

Fusobacterium nucleatum is a Gram-negative anaerobe that is one of the most abundant bacteria in the oral cavity in both diseased and healthy individuals (Griffen *et al.*, 2012). This bacterium is associated with various forms of periodontal disease, including mild reversible gingivitis, advanced irreversible periodontitis (Feng *et al.*, 2014; Kistler *et al.*, 2013; Liu *et al.*, 2014; Yang *et al.*, 2014) and endodontic infections (Rôças *et al.*, 2011; Triches *et al.*, 2014). In addition to oral diseases, *F. nucleatum* has also been reported to be a risk factor for adverse pregnancy outcomes (Haar *et al.*, 2018), gastrointestinal disease (Mira-Pascual *et al.*, 2015; Tahara *et al.*, 2015; Zhong *et al.*, 2014), atherosclerosis (Figuro *et al.*, 2011), rheumatoid arthritis (Témoïn *et al.*, 2012) and Alzheimer's disease (Sparks Stein *et al.*, 2012).

The low levels of free amino acids that are normally present in the oral environment can only meet the basic nutrient requirements of *F. nucleatum* and other oral bacteria (Syrjänen *et al.*, 1990). Thus, *F. nucleatum* has to obtain energy primarily through the fermentation of peptides (Rogers *et al.*, 1992). Notably, most strains prefer peptides rather than free



amino acids, which indicates the important role of peptidases in metabolism (Gharbia *et al.*, 1989). However, due to an absence of systematic studies, the underlying functions of peptidases in *F. nucleatum* remain unknown.

We identified a peptidase encoded by the BKN39_05345 gene in *F. nucleatum*, and a BLAST analysis of its sequence obtained from the NCBI (<https://www.ncbi.nlm.nih.gov/>) showed that this peptidase is putatively a member of the S9 family belonging to the DAP2 superfamily. BKN39_05345 was annotated as an aminopeptidase or acylaminoacyl peptidase catalyzing the removal of amino acids from the N-terminus of peptides. The members of this family are composed of two structural domains: an α/β catalytic hydrolase domain harboring the active site and a cylindrical β -propeller domain (Page & Di Cera, 2008). In the MEROPS database (<http://merops.sanger.ac.uk>), the S9 family can be divided into subfamilies S9A (GGSXGGLL, where X represents Asn or Ala), S9B (GWSYGGY), S9C (GGSYGG) and S9D (SYGAFMT) based on the sequence close to the catalytic serine residue. BKN39_05345 encodes a protein belonging to the S9C subfamily, which was thus termed S9Cfn.

Despite varying in substrate specificity, the peptidases of the S9C subfamily utilize a similar mechanism to regulate substrate access to the covered active site. They allow short peptides to enter the active-site lumen, while excluding longer or structured peptides (Harmat *et al.*, 2011; Yadav *et al.*, 2019). Recently, a member of the S9C subfamily from *Deinococcus radiodurans*, namely S9Cdr, has been shown to have a tetrameric structure and exhibit carboxypeptidase activity (Yadav *et al.*, 2019).

To elucidate the substrate specificity of S9Cfn, which is required in order to combat periodontal disease and deepen our knowledge of its underlying catalytic mechanism, here we report the crystal structure of S9Cfn at 2.6 Å resolution. Structure-guided site-directed mutagenesis combined with activity assays enabled us to identify the first carboxypeptidase in *F. nucleatum* and to shed light on its catalytic mechanism from a structural perspective.

2. Materials and methods

2.1. Cloning and expression

The gene sequence encoding S9Cfn was obtained from the UniProt database (<https://www.uniprot.org/>). The full-length gene encoding S9Cfn (UniProt ID A0A1S1MBE1) was synthesized by General Biosystems after codon optimization for the *Escherichia coli* expression system and was cloned into the pET-29b expression vector with a C-terminal 6×His tag. All mutants were generated via a standard PCR-based strategy (primers are listed in Supplementary Table S1).

Both the wild-type and mutant proteins were overexpressed in *E. coli* strain BL21 (DE3) (Novagen). Cells were grown in Luria–Bertani (LB) culture medium (10 g NaCl, 10 g Bacto Tryptone and 5 g yeast extract per litre) containing 30 $\mu\text{g ml}^{-1}$ kanamycin at 37°C. When the optical density at 600 nm (OD_{600}) reached approximately 0.6–0.8, protein expression

was induced with 0.2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 16°C for a further 20 h. The cells were collected and resuspended in the lysis buffer [20 mM Tris–HCl pH 7.5, 100 mM NaCl, 5 mM β -mercaptoethanol (β -ME)]. The suspension was frozen in liquid nitrogen and stored at -80°C for further use.

Selenomethionine (SeMet)-derived protein was overexpressed in *E. coli* strain B834 (DE3) (Novagen). Transformed cells were cultured in LB medium at 37°C overnight, harvested and washed twice with M9 medium (1 g NH_4Cl , 3 g KH_2PO_4 , 17 g Na_2HPO_4 , 1% glucose, 2 mM MgSO_4 , 0.1 mM CaCl_2 , 1 mg vitamin B₁ and 7.5 mg FeSO_4 per litre). The cells were then cultured in SeMet medium (M9 medium with 25 mg l^{-1} SeMet and 50 mg l^{-1} other essential amino acids) containing 30 $\mu\text{g ml}^{-1}$ kanamycin to an OD_{600} of 0.8–1.0. The subsequent expression and cell-harvesting steps were the same as those described for the native protein.

2.2. Protein purification

For protein crystallization, the thawed suspension was lysed by ultrasonication for 15 min on ice, and the cell lysate was then centrifuged at 12 000g for 30 min at 4°C. The supernatant was loaded onto a Ni–NTA column (GE Healthcare) equilibrated with the binding buffer (20 mM Tris–HCl pH 7.5, 100 mM NaCl, 5 mM β -ME). The target protein was eluted with 6 ml binding buffer (20 mM Tris–HCl pH 7.5, 100 mM NaCl, 5 mM β -ME) plus 400 mM imidazole and further subjected to gel filtration on a Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with the buffer consisting of 20 mM Tris–HCl pH 7.5, 100 mM NaCl, 5 mM β -ME. The purity of the proteins was assessed by 12% (w/v) gel electrophoresis. Peak fractions were collected and concentrated to 10 mg ml^{-1} for further procedures.

For the enzymatic activity assays, the wild-type and mutant enzymes were purified according to a procedure similar to that described above, except that the binding buffer was 20 mM Tris–HCl pH 7.5, 100 mM NaCl. Peak fractions were collected, concentrated to approximately 1.8 mg ml^{-1} and mixed with an equal volume of 100% glycerol. These enzyme mixtures were frozen in liquid nitrogen and stored at -80°C until use.

2.3. Enzymatic activity assays

Alanine standard solutions at various concentrations were prepared in reaction buffer (20 mM MES pH 6.0, 100 mM NaCl, 5% glycerol) and 240 μl of 20% (w/v) ninhydrin reagent was mixed with 60 μl standard solution followed by heating in a 95°C water bath for 10 min for color development. The color intensity was measured as the OD_{504} after cooling. Linear fitting was carried out with the standard solution concentration as the abscissa and OD_{504} as the ordinate.

Enzymatic activity was determined via the ninhydrin assay, which detects the free α -amino groups of amino acids liberated from peptides (Doi *et al.*, 1981). A total of 900 ng purified enzyme was incubated with various custom-synthesized peptides (GL Biochem) at 4 mM in 60 μl reaction buffer (20 mM MES pH 6.0, 100 mM NaCl, 5% glycerol) at 37°C for

30 min. At the end of each reaction, 240 μl 20% (w/v) ninhydrin reagent was added to the reaction mixture followed by heating in a 95°C water bath for 10 min to develop the color. The intensity of the color produced was measured as the OD_{504} after cooling. The amount of product was quantified based on a standard curve.

For the ion-inhibition assays, a total of 900 ng purified enzyme was pre-incubated with various metal ions (1 mM) in 60 μl reaction buffer (20 mM MES pH 6.0, 100 mM NaCl, 5% glycerol) at 37°C for 30 min. The reaction was then incubated with 4 mM $\text{ML}(\text{A})_3$ at 37°C for 5 min. No metal ions were added to the control group. The reaction conditions and detection method were the same as those described above.

To confirm the important role of the oligomeric pores, the activities of the wild-type and G148C variant were detected in the presence or absence of 5 mM β -ME. The reaction conditions and detection method were the same as those described above.

At least three independent measurements were performed for all assays.

2.4. Mass-spectrometric analysis

The cleavage pattern of substrates was monitored by mass spectrometry (MS) using a Q Exactive Plus (Thermo Fisher, USA). A total of 900 ng purified enzyme was incubated with 10 mM $\text{ML}(\text{A})_3$ in 60 μl reaction buffer (20 mM MES pH 6.0, 100 mM NaCl, 5% glycerol) for 5, 10 or 30 min at 37°C and then immediately heated in a water bath at 95°C for 10 min to stop the reaction. No enzyme was added to the control group. The reaction products were then diluted tenfold with the reaction buffer and concentrated by ultrafiltration (Millipore Amicon) to a final volume of 200 μl . Before the products were detected by MS, they were again diluted tenfold with 50% (v/v) acetonitrile containing 1‰ formic acid. Finally, the samples were loaded using an injection pump. The MS parameters were as follows: flow rate 5 $\mu\text{l min}^{-1}$, scan range 100–700 m/z , resolution 70 000, spray voltage (|kV|) 3.5, capillary temperature 320°C, sheath gas flow rate 3 and acquisition time 5 min.

2.5. Crystallization, data collection and processing

The SeMet-derived S9Cfn protein was concentrated to 7 mg ml^{-1} for crystallization. Initial crystal screening was performed by the sitting-drop vapor-diffusion method using commercial crystallization screening kits (Hampton Research). A Mosquito robot (TTP Labtech) was used for the initial screening. Crystals were grown at 16°C using the hanging-drop vapor-diffusion method, with a 1 μl drop of protein solution mixed with an equal volume of reservoir solution. After exhaustive optimization, the reservoir solution used to obtain crystals of SeMet-derived S9Cfn with good diffraction quality was 21% (w/v) polyethylene glycol (PEG) 2000, 0.1 M Tris–HCl pH 7.9. A single crystal was transferred to the cryoprotectant (reservoir solution supplemented with 30% glycerol) and flash-cooled in liquid nitrogen. The X-ray diffraction data were collected at -173°C in a liquid-nitrogen

stream using an ADSC Q315r CCD detector on beamline BL17U1 at the Shanghai Synchrotron Radiation Facility (SSRF). All diffraction data were integrated and scaled with *HKL-2000* (Otwinowski & Minor, 1997).

2.6. Structure determination and refinement

The crystal structure of SeMet-derived S9Cfn was determined by the single-wavelength anomalous dispersion (SAD) phasing method. The *AutoSol* program from *Phenix* (Liebschner *et al.*, 2019) was used to locate the Se atoms and calculate the initial phases. Automatic model building was carried out using *AutoBuild* in *Phenix*. Refinement was carried out using the *CCP4* suite (Winn *et al.*, 2011), and the model was rebuilt using *Coot* (Emsley *et al.*, 2010). The final model was evaluated with *MolProbity* (Chen *et al.*, 2010) and *PROCHECK* (Laskowski *et al.*, 1993). All figures showing structures were prepared with *PyMOL* (<http://www.pymol.org>).

3. Results

3.1. Enzymatic activity assays reveal that S9Cfn is a carboxypeptidase

Sequence analysis predicted that S9Cfn might be an aminopeptidase. We synthesized and cloned the gene encoding S9Cfn into the plasmid pET-19b with a C-terminal His tag and expressed the recombinant protein in *E. coli* (Fig. 1*a*). The purified proteins degraded peptide substrates, which were detected by MS. We used the $\text{ML}(\text{A})_3$ peptide as a substrate to record the cleavage profile. The substrate was treated with the enzyme for 5, 10 or 30 min and the elution profiles of the reaction products were analyzed by MS (Fig. 1*b*). The results showed that the amino acids at the C-terminus of the peptide were released sequentially during the reaction, indicating that S9Cfn is a carboxypeptidase. To confirm this carboxypeptidase activity, we further tested the enzymatic activity towards C-terminally amidated $\text{ML}(\text{A})_3$ and $\text{ML}(\text{A})_3\text{-NH}_2$. The enzyme could not hydrolyze the C-terminally blocked peptides (Fig. 1*c*). Furthermore, the addition of Cu^{2+} or Fe^{3+} , which have been reported to inhibit the activity of carboxypeptidases (Larsen & Auld, 1991), led to an approximately 90% loss of activity (Fig. 1*d*). Moreover, the reported carboxypeptidase activity enhancer Mg^{2+} caused a 13% increase in activity. These results confirmed that S9Cfn is a carboxypeptidase. Similarly, S9Cdr from *D. radiodurans*, which has been annotated as an acylaminoacyl peptidase, was also found to exhibit carboxypeptidase activity (Yadav *et al.*, 2019).

We then examined the substrate preference of S9Cfn using peptides of different lengths, such as ML, MLA, $\text{ML}(\text{A})_2$, $\text{ML}(\text{A})_3$, $\text{ML}(\text{A})_5$ and $\text{ML}(\text{A})_7$. Almost no enzymatic activity was found towards the dipeptide ML. As the peptide length increased, the cleavage efficiency of S9Cfn gradually increased, and the highest efficiency was observed for $\text{ML}(\text{A})_5$ (Fig. 1*e*). The cleavage efficiency of S9Cfn was reduced once the peptide reached nine residues, or longer than $\text{ML}(\text{A})_7$.

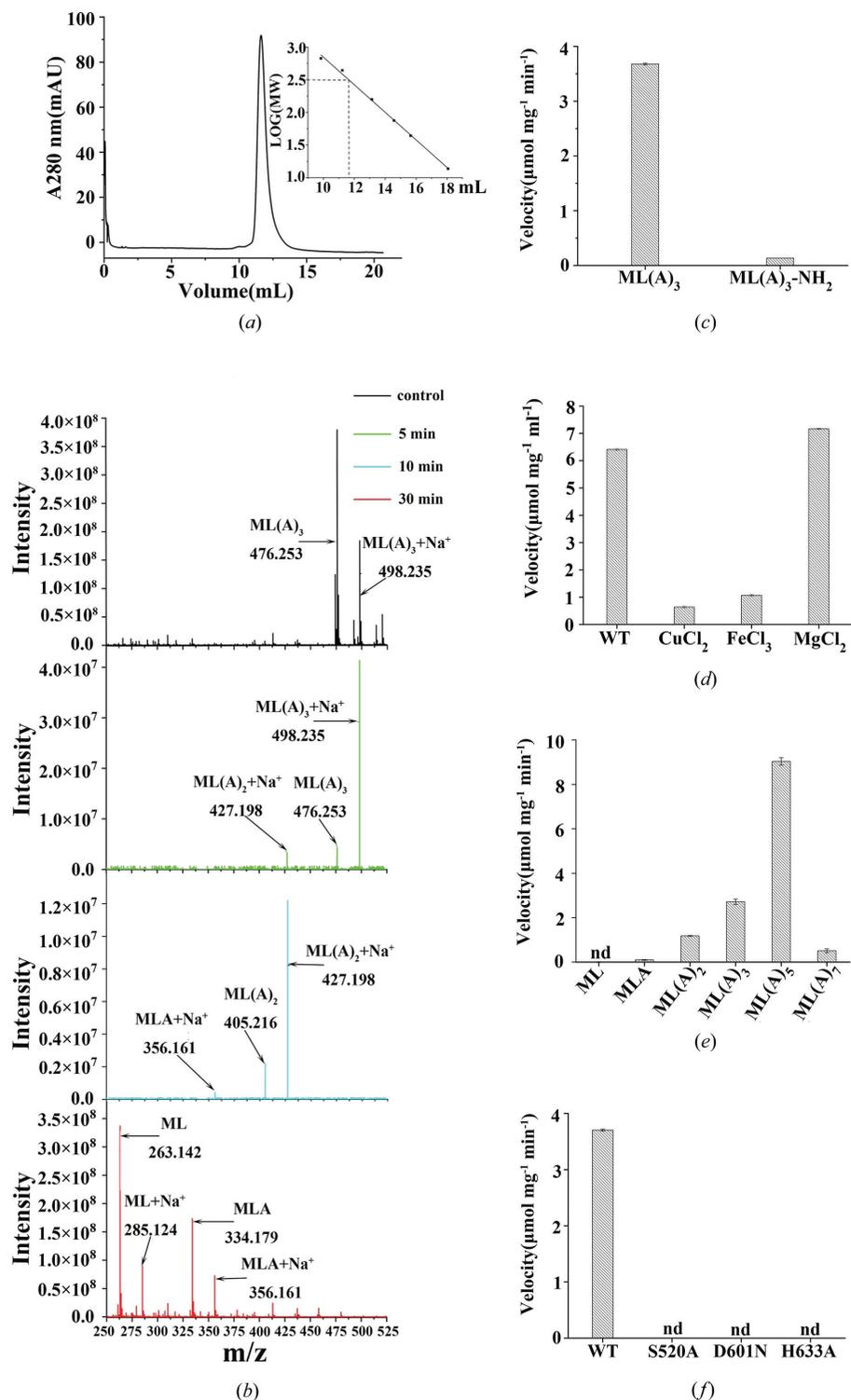


Figure 1 Biochemical characterization of S9Cfn. (a) Molecular-mass determination of S9Cfn by analytical gel-filtration chromatography. The size-exclusion chromatography profile of S9Cfn is shown using a Superdex 200 Increase 10/300 column (GE Healthcare). (b) Mass-spectrometric analysis showing the carboxypeptidase activity of S9Cfn. S9Cfn was incubated with ML(A)₃ and the products were analyzed by mass spectrometry at different times as shown. The representative mass profiles from the different times are shown with the corresponding sequence and *m/z* ratio labeled at the top. (c) The activity of S9Cfn towards C-terminally blocked ML(A)₃. (d) The activity of S9Cfn with the addition of various metal ions. (e) The activity of S9Cfn towards peptides with different chain lengths. (f) The enzyme activities of wild-type S9Cfn and its mutants toward ML(A)₃. The error bars depict the standard deviation between replicates (*n* = 3). Nondetectable activity is depicted 'nd'.

These results indicated that the optimal substrate of S9Cfn is a heptapeptide, similar to S9Cdr (Yadav *et al.*, 2019).

Sequence analysis revealed the conserved catalytic triad consists of Ser514, Asp601 and His633 (Supplementary Fig. S1). To test the roles of these residues, we performed site-directed mutagenesis and determined the enzymatic activities. Purification of the three site-directed mutagenesis mutants of the catalytic triad proceeded similarly to that of the wild type (Supplementary Fig. S2). However, as expected, the results showed that mutation of any of these residues to alanine totally abolished the enzymatic activity (Fig. 1f).

3.2. Overall tetrameric structure of S9Cfn

To decipher the underlying catalytic mechanism of S9Cfn, we employed purified protein samples in crystallization trials. We ultimately obtained high-quality crystals using 0.1 M Tris-HCl pH 7.9, 21% (*w/v*) PEG 2000. The structure was determined at 2.6 Å resolution (PDB entry 7ep9; Table 1) using the single-wavelength anomalous dispersion method with SeMet-substituted crystals. Each asymmetric unit contained four S9Cfn subunits, which are assembled into a tetramer. Consistently, S9Cfn is indeed a tetramer in solution, as confirmed by size-exclusion chromatography (~316 kDa; Fig. 1a).

The tetrameric structure can be recognized as a dimer of dimers. Each subunit is composed of an N-terminal β-propeller domain and a C-terminal α/β catalytic domain (Fig. 2a). The β-propeller domain is composed of seven β-sheets (residues 5–399), each of which consists of four β-strands. The C-terminal α/β catalytic domain is composed of eight β-sheets and six α-helices (Figs. 2b and 2c). The catalytic triad (Ser520, Asp601 and His633; Figs. 2b and 2c) is located in the α/β catalytic domain, which is a common feature among the members of the S9 family. Two subunits form a dimer via extensive interactions between the catalytic domains of each subunit, with an interface area of ~2500 Å². The two

Table 1

Data-collection and refinement statistics for S9Cfn.

Values in parentheses are for the highest resolution shell.

Data collection	
Space group	$P2_12_12$
a, b, c (Å)	192.004, 124.797, 128.898
α, β, γ (°)	90, 90, 90
Resolution range (Å)	50.0–2.6 (2.7–2.6)
No. of reflections	1213183
No. of unique reflections	182885 (95180)
$\langle I/\sigma(I) \rangle$	10.2 (1.7)
R_{merge}^\dagger	0.24 (0.76)
Completeness (%)	100.0 (100.0)
$CC_{1/2}$	0.99 (0.84)
Average multiplicity	12.7 (12.7)
Wilson B factor (Å ²)	37.3
Structure refinement	
Resolution range (Å)	44.8–2.6
R factor ‡ / R_{free}^\S	0.23/0.27
No. of protein atoms	21636
No. of water atoms	490
R.m.s.d. ¶ , bond lengths (Å)	0.002
R.m.s.d., bond angles (°)	0.532
Average B factor (Å ²)	38.5
Ramachandran plot ††	
Favored regions (%)	93.39
Allowed regions (%)	6.19
Outliers (%)	0.42
PDB code	7ep9

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i th observation of reflection hkl and $\langle I(hkl) \rangle$ is the mean value for reflection hkl . Summations are over all reflections. $^\ddagger R$ factor = $\sum_{hkl} (|F_{\text{obs}}| - |F_{\text{calc}}|) / \sum_{hkl} |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. $^\S R_{\text{free}}$ was calculated using 5% of the data, which were excluded from refinement. ¶ R.m.s.d. from ideal values. †† Categories were defined by *MolProbity*.

dimers further form a tetramer via interactions between the propeller domains, with an interface area of $\sim 620 \text{ \AA}^2$ (Fig. 2a).

The catalytic triad is covered by the β -propeller domain as a large dome hindering the approach of a peptide substrate to the domain interface (Fig. 2b). A tunnel was calculated by *MOLEonline* (Pravda *et al.*, 2018) from the catalytic triad in the cavity between the two domains (Fig. 2d), with an expansion of 54 Å (Supplementary Fig. S3). This tunnel, which extends for almost 40 Å (with a diameter greater than 4 Å) to the exit, provides sufficient space to accommodate the peptides. The outlet of the tunnel is highly hydrophilic, in contrast to the neutral core region. The hydrophilic outlet and neutral core region of the tunnel indicate that S9C is active towards a variety of peptides. The exit of the tunnel is restricted by $\alpha 6$ from the catalytic domain and $\alpha 2$ and the re-entrant $\alpha 3$ ($\alpha 3a$ and $\alpha 3b$) in the third propeller blade from the β -propeller domain (Fig. 2e), in particular the gating loop (residues 163–181) between $\alpha 3b$ and $\beta 3$ in the third propeller blade.

3.3. Three arginine residues are responsible for substrate binding

The tripeptide substrate Met-Leu-Ala was docked into our structure using *HADDOCK* (Van Zundert *et al.*, 2016). According to this docking model (Fig. 3a), three arginine residues (Arg543, Arg603 and Arg637) located near the active site could bind to the substrate peptide. The side chain of

Arg603 forms a hydrogen bond to the carbonyl O atom of Met at the P2 position, while the free α -carboxylate of Ala at the P1' position interacts with the side chains of Arg543 and Arg637. Notably, the carbonyl C atom of the P1 leucine is placed at the position of nucleophilic attack by O $^\gamma$ of Ser520. Sequence alignment revealed that these three arginine residues are conserved in the S9C carboxypeptidase family (Supplementary Fig. S1). To identify the roles of these three arginine residues, site-directed mutagenesis was performed combined with enzymatic activity assays (Fig. 3b). All mutants showed significantly reduced activities (Fig. 3c). These results revealed that Arg543, Arg603 and Arg637 play an important role in binding peptides and preparing them for the attack by Ser520. Based on this binding pattern, in which the three arginine residues only coordinate to the main chain of the peptide, we surmised that S9Cfn could bind a vast array of peptides. In addition, the cavity loop (residues 301–318) protruding from the β -propeller domain was parallel to the tripeptide in this model (Fig. 3a). This loop might help to reorient the substrates after their entrance into the substrate channel and binding to the three arginine residues. The function of this cavity loop has also been suggested in a previous report (Yadav *et al.*, 2019).

3.4. The oligomeric pores of the tetramer are the site of substrate entry

Notably, S9Cfn displays an active state without the addition of any substrates during crystallization, *i.e.* the catalytic triad is assembled in a conformation that is ready for catalysis. In the inactive state, which was reported for S9Cdr, the catalytic triad was disassembled and nonfunctional (Yadav *et al.*, 2019). Notably, the conformation of the catalytic triad of S9Cfn is equivalent to that in the active structure of S9Cdr (PDB entry 5yzn), which could only be obtained by crystal soaking with the substrate. Superposition of the two monomers revealed that the catalytic triads could perfectly overlap, including the gating loop (Fig. 4a). Despite sharing 29% sequence identity, S9Cfn and S9Cdr have a quite similar tetrameric structure, with a root-mean-square deviation (r.m.s.d.) of 1.96 Å over 1731 C $^\alpha$ atoms (Supplementary Fig. S4). The major differences between the two tetramers are the oligomeric pores, which are located at the centers of the tetrameric assemblies (Fig. 4b). The oligomeric pore in S9Cfn is restrained by a pair of re-entrant $\alpha 3$ helices in the third propeller blade of the β -propeller domain, with a narrowest distance of 8.6 Å (Fig. 4b). In contrast, in the previously reported S9Cdr structure $\alpha 3$ is much shorter, and the oligomeric pore is restrained by the loop between $\beta 2$ and $\beta 3$ (residues 42–53) in the first propeller blade from the β -propeller domain, leading to a wider distance of 18.6 Å (Fig. 4b). The apparent distinction between the two oligomeric pores indicated that (i) although the catalytic triads are assembled in both structures, they might still represent distinct states, (ii) the oligomeric pore might undergo conformational changes during the catalytic process and (iii) the tunnels of each monomer would extend to the oligomeric pore (Fig. 4b, Supplementary Fig.

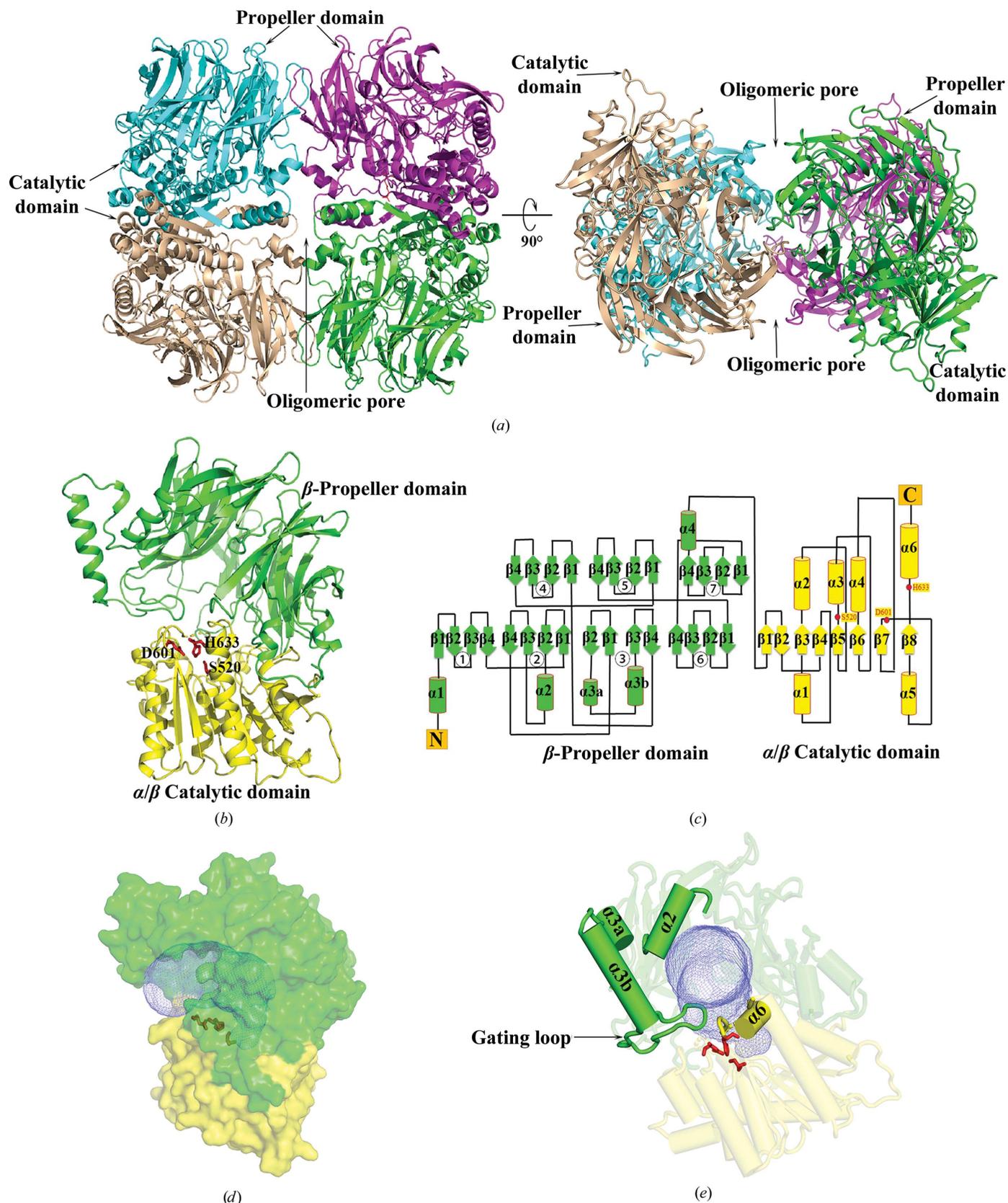


Figure 2
 Overall structure of S9Cfn. (a) Top and side views of the quaternary structure of S9Cfn. The four subunits are shown in different colors. The propeller domain, catalytic domain and oligomeric pore of the tetramer are marked. (b) Cartoon representation of one S9Cfn subunit. The propeller domain and the catalytic domain are shown in green and yellow, respectively, and the catalytic triad is shown in red sticks. The same color scheme is used throughout this manuscript. (c) Topology diagram of S9Cfn. (d) Substrate-channel simulation of S9Cfn. The channel was calculated by *MOLEonline* (Pravda *et al.*, 2018). The surface of the S9Cfn monomer is shown and the channel is displayed as a blue mesh. (e) The exit of the substrate tunnel of one subunit.

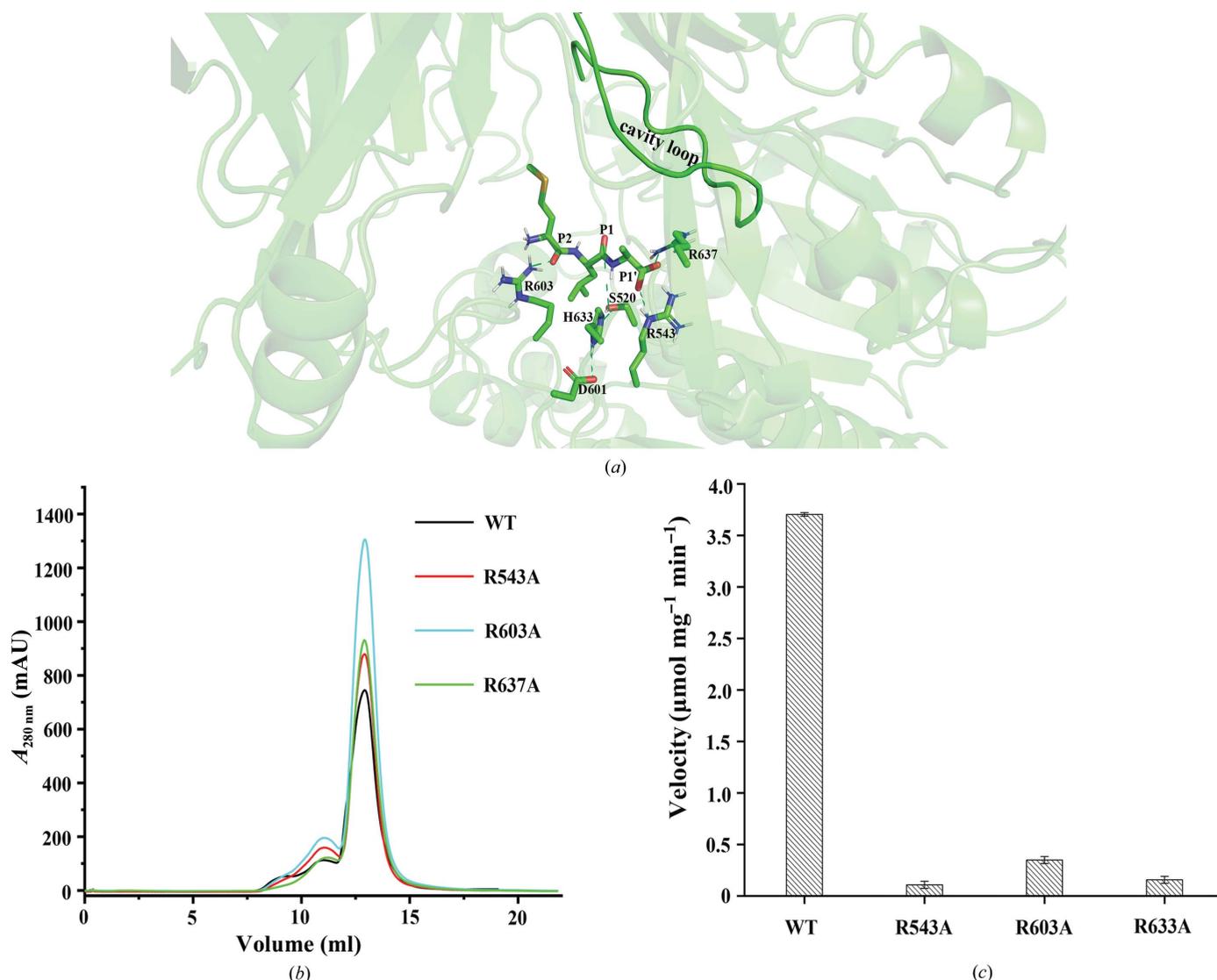


Figure 3 Three arginine residues are responsible for substrate binding. (a) The substrate (Met-Leu-Ala; blue sticks) was modeled in our active S9Cfn structure using *HADDOCK* (Van Zundert *et al.*, 2016). Substrates binding to the arginine residues and the catalytic triad residues are shown as green sticks. (b) Size-exclusion chromatography profiles of wild-type S9Cfn and its mutants using a Superdex 200 Increase 10/300 column (GE Healthcare). (c) The enzymatic activities of wild-type S9Cfn and its mutants towards ML(A)₃.

S5). Interestingly, the previously reported hexameric active structure of serine protease from *Pyrococcus horikoshii* (PhAAP; PDB entry 4hxg; Menyhárd *et al.*, 2013) possesses an oligomeric pore with an opening of 13.4 Å in diameter, which is between those of active S9Cfn and S9Cdr (Supplementary Fig. S6). When S9Cfn was colored by *B* factor, the gating loop was relatively mobile compared with the core region (Supplementary Fig. S7a). We thus proposed that the substrates should enter through the oligomeric pores before approaching the catalytic triad of each subunit. To confirm this hypothesis, we mutated Gly148 at the narrowest point (Figs. 4a and 4b) of the oligomeric pores to cysteine (Supplementary Fig. S7b), which might form an intermolecular disulfide bond under oxidative conditions. Site-directed mutagenesis of this glycine residue gave a similar size-exclusion chromatography profile to the wild type (Supplementary Fig. S8). SDS-PAGE

revealed that a proportion of the mutant protein formed a covalent dimer under nonreducing denaturing conditions, while the wild-type protein migrated as a monomer. The dimer band could be disassembled upon the addition of β-ME (Fig. 4c) by SDS-PAGE. The results indicated that ~30% of the mutant formed disulfide bonds, leading to blockage of the oligomeric pores. The subsequent enzymatic activity assays showed that the G148C variant displayed an ~30% decrease in activity compared with the wild type (Fig. 4d). Moreover, the activity of G148C was comparable to that of the wild type in the presence of 5 mM β-ME, which could break the disulfide bond. These results indicated that the oligomeric pore should play a pivotal role in controlling the entry of peptide substrates into the catalytic pocket. Notably, the substrate tunnel of each subunit is constrained by the other two adjacent subunits (Fig. 4b, Supplementary Fig. S5).

4. Discussion

A variety of enzymes are grouped into the S9 family, including well studied endopeptidases, dipeptidyl peptidases, tripeptidyl peptidases and acylaminoacyl peptidases (Li *et al.*, 2010; Nemoto & Nemoto, 2021; Ekino *et al.*, 2018; Harmat *et al.*, 2011). However, carboxypeptidases have rarely been assigned to this group, with the exception of the recently reported S9Cdr from *D. radiodurans*. Interestingly, similar to S9Cdr, our enzyme S9Cfn was also annotated as an aminopeptidase in the NCBI and UniProt databases, despite a sequence identity

of <30%. Our biochemical results clearly showed that S9Cfn is indeed a carboxypeptidase, rather than the previously annotated aminopeptidase. Together with the data for the previously reported S9Cdr, we concluded that some members of the S9C subfamily are most likely to be carboxypeptidases.

Peptidases of the S9 family are proposed to adopt sophisticated mechanisms to regulate substrate access to the buried active site and permit peptides with a proper length while preventing inappropriate peptides from nontargeted digestion (Harmat *et al.*, 2011; Yadav *et al.*, 2019). However, the detailed

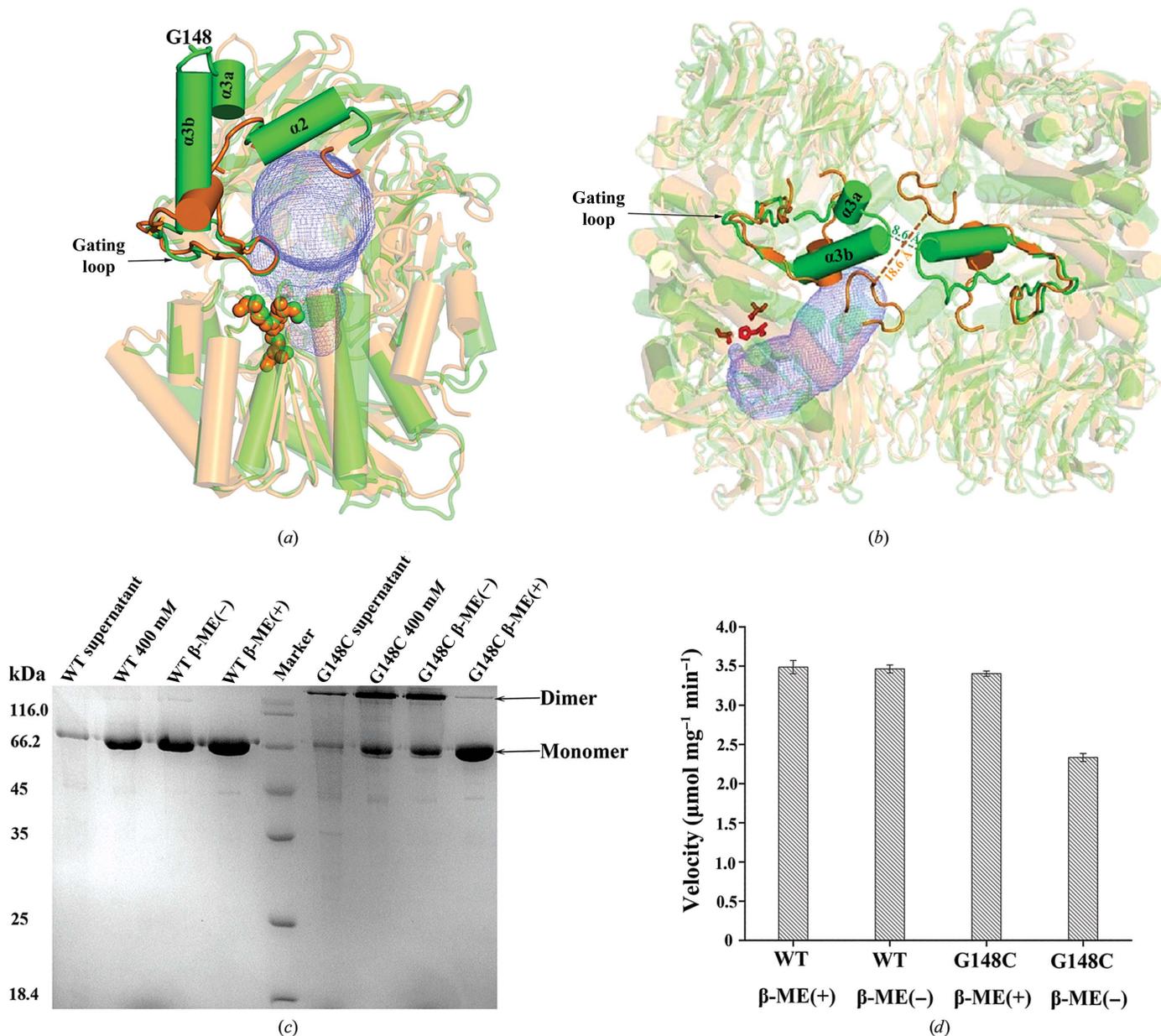


Figure 4

The oligomeric pores of the tetramer are essential for substrate entry. (a) Superposition of the S9Cfn monomer with S9Cdr (PDB entry 5yzn). The S9Cfn monomer is shown as a green cartoon, while the S9Cdr monomer is shown in orange. The catalytic triad residues are shown as spheres in the same color as their respective monomers. Gly148 is shown as sticks. (b) Superposition of the S9Cfn tetramer with the S9Cdr tetramer (PDB entry 5yzn). The cartoons are displayed with the same color scheme as described in (a). $\alpha 3a$, $\alpha 3b$ and the gating loop responsible for oligomeric pore formation from the contralateral monomer are displayed. The narrowest distances in the pores are shown. (c) Disulfide-bond formation in the S9Cfn G148C mutant. Wild-type S9Cfn and the G148C mutant were purified in the absence of β -mercaptoethanol (β -ME). The purified wild-type S9Cfn and G148C mutant proteins were analyzed by SDS-PAGE with or without β -ME. The bands corresponding to the monomer and the covalent dimer are labeled. (d) The enzymatic activity of wild-type S9Cfn and the G148C mutant in the presence or absence of β -ME toward ML(A)₃.

mechanism of tetrameric enzymes in the S9 family is still unclear due to a lack of sufficient structural information. It has been proposed that peptides approach the catalytic triad via the oligomeric pore of the tetramer and through the lateral opening of each monomer (Menyhárd *et al.*, 2013; Yadav *et al.*, 2019). For the first time, using site-directed mutagenesis combined with biochemical assays, we demonstrated that the oligomeric pore is indeed the substrate entrance, which is controlled by a pair of re-entrant α 3 helices from the adjacent pair of β -propeller domains.

It is noteworthy that although the catalytic triad of the S9Cfn structure is in the active state, as in the structure of S9Cdr, the oligomeric pore becomes much more constrained, from ~ 18 Å in diameter in S9Cdr to ~ 8 Å in our structure. It would be difficult for peptides to enter such a narrow opening, suggesting that induced fit would occur upon approach of the substrate to the guard loop and helices of the oligomeric pore. Comparing our structure with other known structures (Supplementary Fig. S6), we found that the oligomeric pore might undergo significant conformational changes during the catalytic cycles. Accordingly, the oligomeric pore may sense substrates and facilitate their entry into the tunnel with an open conformation. The dynamic gating mechanism might also contribute to determining the optimal peptide length. These results deepen our knowledge of carboxypeptidases from the S9C subfamily.

5. Related literature

The following reference is cited in the supporting information for this article: Robert & Gouet (2014).

Acknowledgements

We thank the staff of beamline BL-17U1 at the Shanghai Synchrotron Radiation Facility. Author contributions were as follows. W-TH and RX conceived, designed and supervised the project. XW and M-TC designed and performed the experiments. Z-PC and Y-LJ collected the X-ray data and solved the structure. XW, M-TC, W-TH and Y-SG analyzed the data. XW, M-TC and W-TH prepared the manuscript. All authors discussed the data and read the manuscript. The authors declare no competing interests.

Funding information

This work was supported by the Chinese National Key R&D Program (2018YFB0407204).

References

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
 Doi, E., Shibata, D. & Matoba, T. (1981). *Anal. Biochem.* **118**, 173–184.
 Ekino, K., Yonei, S., Oyama, H., Oka, T., Nomura, Y. & Shin, T. (2018). *Appl. Biochem. Biotechnol.* **184**, 239–252.
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.

Feng, X., Zhang, L., Xu, L., Meng, H., Lu, R., Chen, Z., Shi, D. & Wang, X. (2014). *J. Periodontol.* **85**, 150–159.
 Figuero, E., Sánchez-Beltrán, M., Cuesta-Frecho, S., Tejerina, J. M., del Castro, J. A., Gutiérrez, J. M., Herrera, D. & Sanz, M. (2011). *J. Periodontol.* **82**, 1469–1477.
 Gharbia, S. E., Shah, H. N. & Welch, S. G. (1989). *Curr. Microbiol.* **19**, 231–235.
 Griffen, A. L., Beall, C. J., Campbell, J. H., Firestone, N. D., Kumar, P. S., Yang, Z. K., Podar, M. & Leys, E. J. (2012). *ISME J.* **6**, 1176–1185.
 Harmat, V., Domokos, K., Menyhárd, D. K., Palló, A., Szeltner, Z., Szamosi, I., Beke-Somfai, T., Náray-Szabó, G. & Polgár, L. (2011). *J. Biol. Chem.* **286**, 1987–1998.
 Kistler, J. O., Booth, V., Bradshaw, D. J. & Wade, W. G. (2013). *PLoS One*, **8**, e71227.
 Larsen, K. S. & Auld, D. S. (1991). *Biochemistry*, **30**, 2613–2618.
 Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
 Li, M., Chen, C., Davies, D. R. & Chiu, T. K. (2010). *J. Biol. Chem.* **285**, 21487–21495.
 Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
 Liu, P., Liu, Y., Wang, J., Guo, Y., Zhang, Y. & Xiao, S. (2014). *PLoS One*, **9**, e85280.
 Menyhárd, D. K., Kiss-Szemán, A., Tichy-Rács, É., Hornung, B., Rádi, K., Szeltner, Z., Domokos, K., Szamosi, I., Náray-Szabó, G., Polgár, L. & Harmat, V. (2013). *J. Biol. Chem.* **288**, 17884–17894.
 Mira-Pascual, L., Cabrera-Rubio, R., Ocon, S., Costales, P., Parra, A., Suarez, A., Moris, F., Rodrigo, L., Mira, A. & Collado, M. (2015). *J. Gastroenterol.* **50**, 167–179.
 Nemoto, T. K. & Ohara Nemoto, Y. (2021). *Mol. Oral Microbiol.* **36**, 145–156.
 Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
 Page, M. J. & Di Cera, E. (2008). *Cell. Mol. Life Sci.* **65**, 1220–1236.
 Pravda, L., Sehnal, D., Toušek, D., Navrátilová, V., Bazgier, V., Berka, K., Svobodová Vařeková, R., Koča, J. & Otyepka, M. (2018). *Nucleic Acids Res.* **46**, W368–W373.
 Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
 Rôças, I. N., Siqueira, J. F. Jr & Debelian, G. J. (2011). *J. Endod.* **37**, 1206–1212.
 Rogers, A., Gully, N., Pfennig, A. & Zilm, P. (1992). *Oral Microbiol. Immunol.* **7**, 299–303.
 Sparks Stein, P., Steffen, M. J., Smith, C., Jicha, G., Ebersole, J. L., Abner, E. & Dawson, D. III (2012). *Alzheimers Dement.* **8**, 196–203.
 Syrjänen, S., Alakujala, L., Alakujala, P., Markkanen, S. & Markkanen, H. (1990). *Arch. Oral Biol.* **35**, 189–193.
 Tahara, T., Shibata, T., Kawamura, T., Okubo, M., Ichikawa, Y., Sumi, K., Miyata, M., Ishizuka, T., Nakamura, M., Nagasaka, M., Nakagawa, Y., Ohmiya, N., Arisawa, T. & Hirata, I. (2015). *Dig. Dis. Sci.* **60**, 205–210.
 Témoins, S., Chakaki, A., Askari, A., El-Halaby, A., Fitzgerald, S., Marcus, R. E., Han, Y. W. & Bissada, N. F. (2012). *J. Clin. Rheumatol.* **18**, 117–121.
 Triches, T. C., de Figueiredo, L. C., Feres, M., de Freitas, S. F. T., Zimmermann, G. S. & Cordeiro, M. M. R. (2014). *J. Dent. Child.* **81**, 14–19.
 Vander Haar, E. L., So, J., Gyamfi-Bannerman, C. & Han, Y. W. (2018). *Anaerobe*, **50**, 55–59.
 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235–242.

- Yadav, P., Goyal, V. D., Gaur, N. K., Kumar, A., Gokhale, S. M., Jamdar, S. N. & Makde, R. D. (2019). *J. Biol. Chem.* **294**, 89–100.
- Yang, N.-Y., Zhang, Q., Li, J.-L., Yang, S.-H. & Shi, Q. (2014). *Int. J. Paediatr. Dent.* **24**, 226–233.
- Zhong, D., Brower-Sinning, R., Firek, B. & Morowitz, M. J. (2014). *J. Pediatr. Surg.* **49**, 441–446.
- Zundert, G. C. P. van, Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastiris, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J. & Bonvin, A. M. J. J. (2016). *J. Mol. Biol.* **428**, 720–725.