Journal of Microbiology (2020) Vol. 58, No. 9, pp. 772–779
DOI 10.1007/s12275-020-0292-0

eISSN 1976-3794
pISSN 1225-8873

Structural insights into the psychrophilic germinal protease *Pa*GPR and its autoinhibitory loop

Chang Woo Lee[1†], Saeyoung Lee[2,3†], Chang-Sook Jeong[1,4], Jisub Hwang[1,4], Jeong Ho Chang[5], In-Geol Choi[3], T. Doohun Kim[6], HaJeung Park[7*], Hye-Yeon Kim[2,8*], and Jun Hyuck Lee[1,4*]

[1]*Unit of Research for Practical Application, Korea Polar Research Institute, Incheon 21990, Republic of Korea*
[2]*Research Center for Bioconvergence Analysis, Korea Basic Science Institute, Chungbuk 34133, Republic of Korea*
[3]*Department of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Republic of Korea*
[4]*Department of Polar Sciences, University of Science and Technology, Incheon 21990, Republic of Korea*
[5]*Department of Biology Education, Kyungpook National University, Daegu 41566, Republic of Korea*
[6]*Department of Chemistry, College of Natural Science, Sookmyung Woman's University, Seoul 04310, Republic of Korea*
[7]*X-Ray Core, The Scripps Research Institute, Scripps Florida, 130 Scripps Way #1A1, Jupiter, FL 33458, USA*
[8]*Center for Convergent Research of Emerging Virus Infection (CEVI), Korea Research Institute of Chemical Technology, Daejeon 34114, Republic of Korea*

(Received Jun 4, 2020 / Revised Jun 30, 2020 / Accepted Jul 15, 2020)

In spore forming microbes, germination protease (GPR) plays a key role in the initiation of the germination process. A critical step during germination is the degradation of small acid-soluble proteins (SASPs), which protect spore DNA from external stresses (UV, heat, low temperature, etc.). Inactive zymogen GPR can be activated by autoprocessing of the N-terminal pro-sequence domain. Activated GPR initiates the degradation of SASPs; however, the detailed mechanisms underlying the activation, catalysis, regulation, and substrate recognition of GPR remain elusive. In this study, we determined the crystal structure of GPR from *Paenisporosarcina* sp. TG-20 (*Pa*GPR) in its inactive form at a resolution of 2.5 Å. Structural analysis showed that the active site of *Pa*GPR is sterically occluded by an inhibitory loop region (residues 202–216). The N-terminal region interacts directly with the self-inhibitory loop region, suggesting that the removal of the N-terminal pro-sequence induces conformational changes, which lead to the release of the self-inhibitory loop region from the active site. In addition, comparative sequence and structural analyses revealed that *Pa*GPR contains two highly conserved Asp residues (D123 and D182) in the active site, similar to the putative aspartic acid protease GPR from *Bacillus megaterium*. The catalytic domain structure of *Pa*GPR also shares similarities with the sequentially non-homologous proteins HycI and HybD. HycI and HybD are metalloproteases that also contain two Asp (or Glu) residues in their active site, playing a role in metal binding. In summary, our results provide useful insights into the activation process of *Pa*GPR and its active conformation.

***Keywords*:** crystal structure, germination protease, X-ray crystallography, zymogen

Introduction

In harsh environments, including heat, cold, drought, UV radiation, and lack of nutrients, some microorganisms, including *Bacillus*, *Clostridium*, and *Sporosarcina* species, transition into a dormant state and undergo morphological changes through spore formation. These spores are composed of several layers of coat proteins that enable the microorganism to survive environmental stresses (Setlow, 1995). During spore formation, small acid-soluble proteins (SASPs) of about 10 kDa are synthesized. These SASPs directly bind to the nuclear double-stranded DNA of spores, protecting the DNA backbone from external stresses. When an external stress condition disappears, dormant spores begin to sprout. At the initial stage of spore germination, germination protease (GPR) degrades SASPs. During sporulation, GPR exists as an inactive zymogen and it is then transformed into its active form by autoprocessing of the N-terminal pro-sequence region (Loshon and Setlow, 1982; Sanchez-Salas and Setlow, 1993; Illades-Aguiar and Setlow, 1994; Pedersen *et al.*, 1997; Nessi *et al.*, 1998).

GPR from *Bacillus megaterium* (*Bm*GPR) has been extensively studied. The zymogen state of *Bm*GPR is called P_{46} (inactive form) and can be autoprocessed to the active form P_{41} by the removal of 15 N-terminal residues; P_{41}, in turn, can undergo further autoprocessing to yield the N-terminal 22 residues truncated form P_{39} (active form) (Carroll and Setlow, 2005). Autoprocessing is triggered by a low pH and an increased level of dipicolinic acid in the spore and by dehydration of the spore core (Illades-Aguiar and Setlow, 1994). Only activated GPRs recognize and cleave SASPs. *Bm*GPR exhibits no sequence similarity to any known class of proteases. To date, only the crystal structure of the zymogen of *Bm*GPR (PDB ID: 1C8B) was elucidated at 3.0 Å resolution (Ponnuraj *et al.*, 2000b). Notably, this structure is missing many portions (i.e., 50 of 370 resides were not modeled)

†These authors contributed equally to this work.
*For correspondence. (H.J. Park) E-mail: hajpark@scripps.edu; Tel.: +1-561-228-2121 / (H.Y. Kim) E-mail: hyeyeon@kbsi.re.kr; Tel.: +82-43-240-5087 / (J.H. Lee) E-mail: junhyucklee@kopri.re.kr; Tel.: +82-32-760-5555; Fax: +82-32-760-5509

due to weak electron density. The overall structure of *Bm*GPR exhibits a novel fold made of eight β-strands and eleven α-helices. The catalytic core domain has a central β-sheet composed of eight β-strands and is enclosed by nine α-helices. The structure also contains a cap region consisting of two α-helices. *Bm*GPR forms a functional tetramer (dimer of a dimer) (Ponnuraj *et al.*, 1999, 2000a). Site-directed mutagenesis analyses of *Bm*GPR have shown that two Asp residues (D127 and D193) in the core domain are essential for autoprocessing and enzyme activity. From enzyme activity assay results and available structural information, the authors suggested that *Bm*GPR may be an atypical aspartic protease (Carroll and Setlow, 2005). However, the molecular structure of the active form of a GPR remains unknown, and therefore the exact enzymatic mechanism of GPR has yet to be precisely determined.

In this study, we report the 2.5 Å resolution crystal structure of GPR from *Paenisporosarcina* sp. TG-20 (*Pa*GPR) in its inactive form. *Paenisporosarcina* sp. TG-20 is a spore-forming psychrophilic bacterium isolated from Taylor glacier in the Antarctic. Our *Pa*GPR structure was determined at a higher resolution than that of *Bm*GPR. Thus, we were able to precisely compare the structure of *Pa*GPR with those of GPR homologs. Structural comparison of *Pa*GPR and *Bm*GPR showed that the proteins exhibited slightly different self-inhibition interactions. In addition, *Pa*GPR shares structural similarities with metal-dependent hydrogenase maturation peptidases, such as HycI and HybD (Fritsche *et al.*, 1999; Kumarevel *et al.*, 2009; Kwon *et al.*, 2016, 2018). Taken together, structural comparisons suggest possible enzymatic regulatory and reaction mechanisms of *Pa*GPR.

Materials and Methods

Cloning, expression, and purification of PaGPR

The *gpr* gene encoding residues 1–349 of GPR was amplified by polymerase chain reaction (PCR) from the genomic DNA of *Paenisporosarcina* sp. TG-20 using PrimeSTAR HS DNA polymerase (Takara). The PCR was conducted using the following primers: 5′-CGATAACATATGATGAGTAAAA TTAAG-3′ (forward) and 5′-CGATAACTCGAGTCAGG GTGAAATAGAAC-3′ (reverse). The DNA fragments were digested with *Nde*I and *Xho*I and cloned into the pET-28a vector (Novagen). The construct contained the *gpr* gene with a 6 × His-tag, separated by a 20-amino acid spacer sequence. The cloned *Pa*GPR was transformed into *E. coli* BL21(DE3) cells for protein expression. Cells were grown at 37°C in 4 L of LB medium containing 50 μg/ml kanamycin, until the OD_{600} reached a value of 0.6–0.7. Overexpression of GPR was induced by 1.0 mM IPTG, and the culture was incubated for 20 h at 25°C. The cells were harvested by centrifugation and disrupted by ultrasonication in lysis buffer (50 mM sodium phosphate; pH 8.0, 300 mM NaCl, 5 mM imidazole) containing 0.2 mg/ml lysozyme. Cell debris was removed by centrifugation at 16,000 rpm for 1 h at 4°C. The supernatant was loaded onto a Ni^{2+}-NTA affinity column (Qiagen) pre-equilibrated with lysis buffer and then washed with a buffer containing 50 mM sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole. Bound *Pa*GPR was eluted with elution buffer (50 mM sodium phosphate; pH 8.0, 300 mM NaCl, 300 mM imidazole), and the protein was concentrated using Amicon Ultra Centrifugal Filters (Ultracel-3K; Millipore). Concentrated GPR was then applied to a Superdex 200 column (GE Healthcare) pre-equilibrated with a buffer consisting of 50 mM Tris-HCl; pH 8.0 and 150 mM NaCl.

Analytical ultracentrifugation

To investigate the oligomeric state of *Pa*GPR in solution, we performed analytical ultracentrifugation (AUC) using a ProteomeLab XL-A (Beckman Coulter). The AUC buffer contained 20 mM Tris-HCl; pH 8.0 and 150 mM NaCl, and the experiment was performed at 20°C. The sample was centrifuged at 40,000 rpm for 10 min, and the sedimentation profile was monitored at a wavelength of 280 nm. Data were analyzed using the SEDFIT software (Schuck *et al.*, 2002).

Crystallization of PaGPR

Recombinant *Pa*GPR was concentrated to 12 mg/ml for crystallization. Initial crystallization screening in a 96-well sitting drop plate (Emerald Bio) was performed using several commercial crystallization screening kits, such as the MCSG I-IV (Microlytic), SG-1 Screen (Molecular Dimensions), Morpheus (Molecular Dimensions), PEGRx, SaltRx, PEG/Ion, and Index (Hampton Research). Equal volumes of the reservoir (200 nl) and protein solution (200 nl) were mixed and equilibrated against 80 μl reservoir solution at 20°C. Initial crystals appeared after one day from the following conditions: 3.5 M sodium formate (SG-1 #E5) and 0.1 M Bis-Tris pro-

Table 1. Data collection statistics and refinement statistics

Data set	*Pa*GPR
X-ray source	PAL 5C beam line
Space group	$P3_221$
Wavelength (Å)	0.97933
Resolution (Å)	50.00–2.50 (2.54–2.50)
Total reflections	610305
Unique reflections	46,339 (2,020)
Average I/σ (I)	29.4 (1.8)
R_{merge}[a]	0.106 (0.392)
Redundancy	13.2 (6.5)
Completeness (%)[b]	98.9 (86.7)
Refinement	
Resolution range (Å)	50.01–2.50 (2.57–2.50)
No. of reflections of working set	44061
No. of reflections of test set	2,257 (146)
No. of amino acid residues	662
No. of water molecules	39
R_{cryst}[b]	0.186 (0.347)
R_{free}[c]	0.227 (0.320)
R.m.s. bond length (Å)	0.0188
R.m.s. bond angle (°)	2.0737
Average B value ($Å^2$) (protein)	74.950
Average B value ($Å^2$) (solvent)	60.504

[a] $R_{merge} = \Sigma \mid <I> - I \mid / \Sigma <I>$.
[b] $R_{cryst} = \Sigma \mid |Fo| - |Fc| \mid / \Sigma |Fo|$.
[c] R_{free} calculated with 5% of all reflections excluded from refinement stages using high-resolution data.
Values in parentheses indicate the highest resolution shells.

pane pH 7.0 with 2.8 M sodium acetate (MCSG #B7). The crystallization conditions were optimized using the hanging-drop vapor diffusion method in 24-well crystallization plates (Molecular Dimensions). The crystallization drops consisted of equal volumes of the reservoir (1 μl) and protein solution (1 μl) and were equilibrated against 500 μl reservoir solution at 20°C. The final optimized hexagonally shaped crystals were obtained from 3.2 M sodium acetate.

X-ray diffraction and data collection

GPR crystals were harvested, briefly soaked into Paratone-N oil (Hampton Research) for cryoprotection, and flash-frozen in a nitrogen-gas stream. X-ray data were collected on beamline 5C at the Pohang Accelerator Laboratory (PAL). A total of 360 images were collected with 1 oscillation and an exposure time of 1 sec per image. The wavelength of the X-ray beam was 0.9796 Å, and the detector-to-crystal distance was

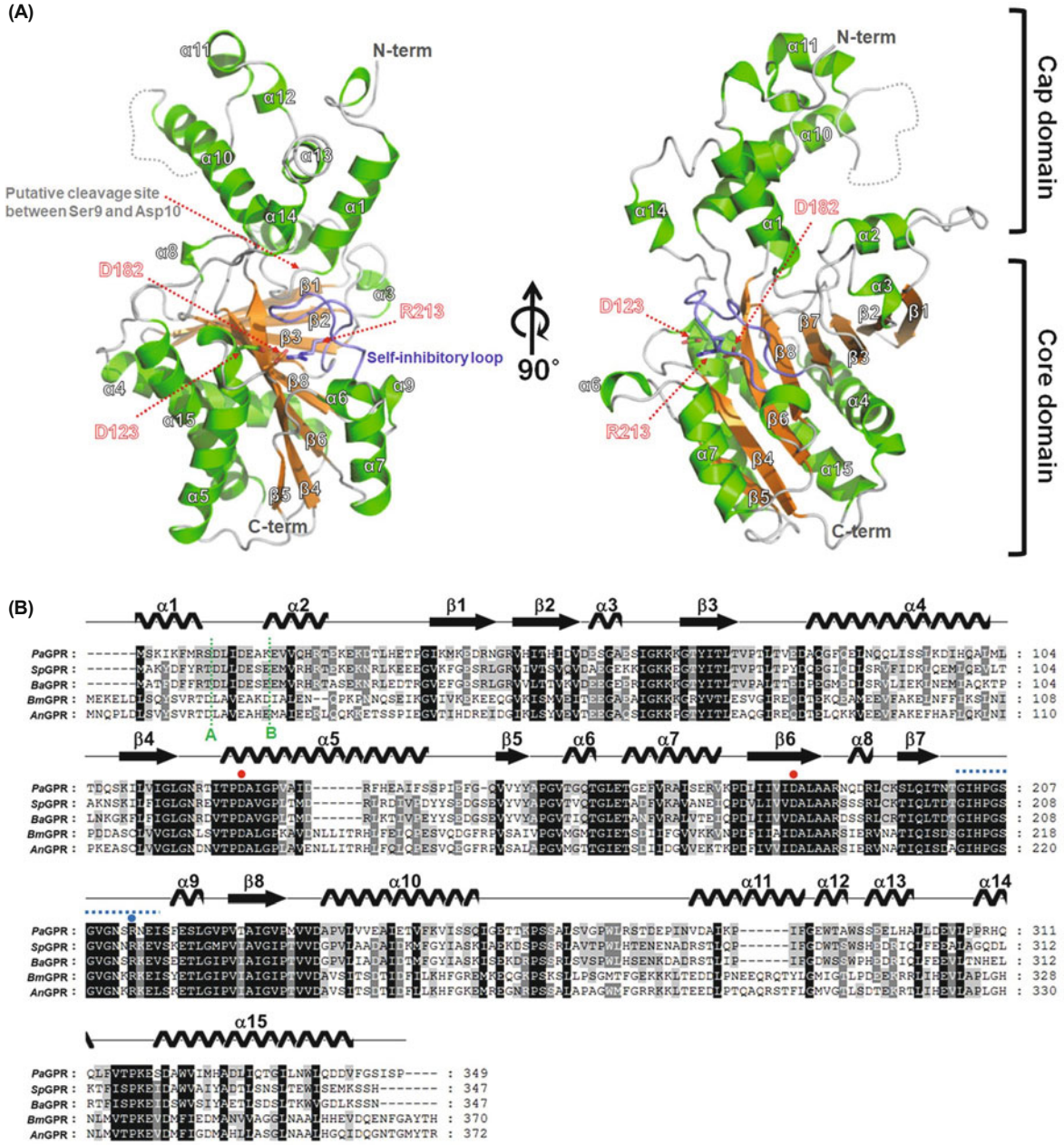

Fig. 1. Crystal structure of *Pa*GPR and sequence alignments. (A) Overall structure of *Pa*GPR is shown as a cartoon, with α-helices colored green and β-stands colored orange. The conserved aspartic acids (Asp123 and Asp182) and Arg213 (located on the self-inhibitory loop region) are shown as sticks. The self-inhibitory loop region (residues 202–216) is shown in blue. (B) Multiple sequence alignments of *Pa*GPR and other GPRs, including *Sp*GPR endopeptidase from *Sporosarcina* sp. HYO08 (NCBI accession code: WP_067403391.1), *Ba*GPR endopeptidase from *Bacillus* sp. OxB-1 (NCBI accession code WP_041075278.1), *Bm*GPR (PDB ID: 1C8B, UniProtKB code: P22321), and *An*GPR endopeptidase from *Anoxybacillus* sp. P3H1B (NCBI accession code: WP_066145953.1). The secondary structures of *Pa*GPR are shown above the aligned sequences. Asp123 and Asp182 are indicated by red dots and Arg213 by a blue dot. The self-inhibitory loop region is indicated by a dotted blue line. The N-terminal pro-peptide cleavage sites for P_{41} *Bm*GPR and P_{39} *Bm*GPR are indicated by a green (A) and (B).

350 mm. Diffraction data were indexed, integrated, and scaled using the *HKL-2000* software (Otwinowski and Minor, 1997). Data collection statistics are provided in Table 1. The best diffraction dataset was collected at a resolution of 2.5 Å. The crystal belonged to space group $P3_221$ exhibiting the following unit-cell parameters: a = 157.059 Å, b = 157.059 Å, c = 95.933 Å. The Matthews coefficient and solvent content were calculated to be 2.20 $Å^3$/Da and 44%, respectively, assuming four protein molecules per asymmetric unit.

Structure determination and refinement

Initial phase information was obtained by the molecular replacement method using the *MOLREP* (Vagin and Teplyakov, 2010) from the *CCP4* suite (Winn *et al.*, 2011). The core domain of *Bm*GPR (PDB ID: 1C8B; 36% sequence identity) was used as a search model. The initial partial model of *Pa*GPR was then iteratively manually rebuilt and refined using the programs *REFMAC5* (Murshudov *et al.*, 2011), *phenix.reifne* (Afonine *et al.*, 2012), and *COOT* (Emsley and Cowtan, 2004). The final model of *Pa*GPR had a R_{work} and R_{free} of 0.186 and 0.227, respectively. The quality of the refined model was validated by *MolProbity* (Chen *et al.*, 2010). The detailed statistics are presented in Table 1. Structural representations were generated using *PyMOL* (DeLano, 2002). The coordinates and structure factors of *Pa*GPR were deposited in the RCSB Protein Data Bank under accession code 7C4X.

Results and Discussion

Overall structure of *Pa*GPR

The 2.5-Å resolution crystal structure of *Pa*GPR was determined by molecular replacement. The truncated core domain of *Bm*GPR was used as a search model (PDB ID: 1C8B) (Ponnuraj *et al.*, 2000b). The amino acid sequences of *Pa*GPR and *Bm*GPR share approximately 31% sequence identity and 52% sequence similarity. Each monomer in the *Pa*GPR structure contained 325 of the 349 amino acid residues. The *Pa*GPR monomer structure contains 15 α-helices and eight β-strands, and the overall structure is divided into two domains: the catalytic core and cap domain (Fig. 1A). The core domain consists of eight central β-strands (the strands β1–β4 form an antiparallel β-sheet and the strands β5–β8 form a parallel β-sheet.) surrounded by eight α-helices (α3–α9 and α15). The cap domain comprises seven α-helices (α1, α2, α10–α14). The two domains are linked by three hinge regions (residues 9–15, 232–235, and 313–319) (Fig. 1B). However, 20 residues (256–275) of the α10–α11 loop region were not built in the final model due to weak electron density.

The space group of the *Pa*GPR structure is $P3_221$, and the asymmetric unit contains two protomers. The results of the AUC experiment demonstrated that the purified *Pa*GPR protein was a stable tetramer in solution (Fig. 2). Additionally, by applying a crystallographic symmetry operator, a tight tetrameric arrangement (dimer of a dimer) can be generated, as shown in Fig. 2. PISA analysis showed that the in-

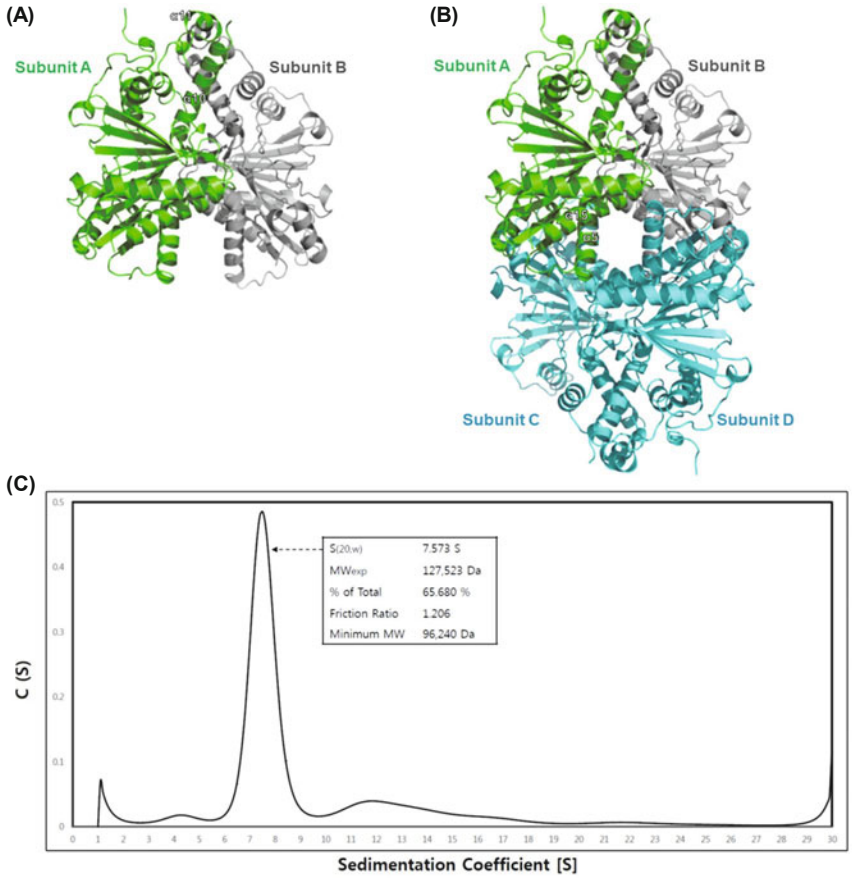

Fig. 2. Oligomeric state of *Pa*GPR. (A) The asymmetric unit of *Pa*GPRs contains two monomers. (B) Tetrameric structure of *Pa*GPR. (C) The result of AUC shows that *Pa*GPR exhibited a tetrameric state in solution. AUC with 0.5 mg/ml *Pa*GPR gave a mass of 127.5 kDa (sedimentation coefficient: 7.573 S; frictional ratio: 1.206), indicating that *Pa*GPR is a stable tetramer in solution. The predicted molecular weight of *Pa*GPR based on its amino acid composition is 38.3 kDa.

Table 2. **DALI search results**

Protein	PDB ID	DALI Z-score	UniProtKB code	Sequence similarity (%) with *Pa*GPR (aligned residue number)	Reference
Germination protease from *Bacillus megaterium*	1C8B	29.0	P22321	26 (282/320)	Ponnuraj *et al.* (2000b)
[NiFe] hydrogenase maturation protease HycI from *Thermococcus kodakarensis*	5ZBY	12.4	Q5JIH7	19 (134/156)	Kwon *et al.* (2016)
Hydrogenase 2 maturation peptidase from *Thaumarchaeota archaeon*	5TTX	11.2	M7TUZ8	21 (111/154)	Not published
Uncharacterized protein from *Wolinella succinogenes*	3PU6	10.1	Q7MR08	12 (122/138)	Not published

terface area generated by the crystallographic symmetry was 26.1% of the total surface (51424.3 Å^2), indicating a biologically relevant interface. The dimer interface in *Pa*GPR is formed by α10–α12, α12–α13 loop, α13, α13–α14 loop, α14, and α14–α15 loop regions and dominated by hydrophobic interactions (Val241, Ile244, Val247, Phe248, Phe287, Trp290, Trp293, Leu298, Leu301, Leu302, and Val305). In addition, several polar interactions are involved in dimerization. Remarkably, the residues Arg309 and Lys318 protrude toward the neighboring subunit and exhibit hydrogen bonds and salt bridges. The dimer-dimer interface is formed by two different patches. In the first patch, the β5 strand and the α-helices α6 and α7 interact with the equivalent regions of the adjacent dimer. Similarly, the second patch is formed by interactions between the helices α5 and α15 of both dimers.

Next, a structure homology search was performed with the *Pa*GPR monomer structure using the DALI server. The search identified *Bm*GPR (Z-score = 29.0) as the closest structural homolog of *Pa*GPR. Additionally, the top four hits of the DALI search included HycI protease (PDB ID: 5ZBY, Z-score = 12.4), hydrogenase 2 maturation peptidase (PDB ID: 5TTX, Z-score = 11.2), and an uncharacterized protein (PDB ID: 3PU6, Z-score = 10.1) (Table 2).

Structural comparison of *Pa*GPR and *Bm*GPR

The *Pa*GPR structure was superimposed with that of *Bm*GPR, and the r.m.s. deviation over 307 Cα atoms was 0.96 Å. Structural comparisons of *Pa*GPR and *Bm*GPR showed that *Pa*GPR contains two conserved Asp residues (D123 and D182) in the active site. As both crystal structures represent the inactive forms of GPR, their active sites are occupied by an internal self-inhibitory loop region (residues 217–231 in *Bm*GPR and residues 202–216 in *Pa*GPR). Thus, proper substrate binding may not be possible without further activation and conformational changes of the inhibitory loop. The self-inhibition of *Pa*GPR occurs by direct interactions between the two catalytic Asp residues and Arg213. *Bm*GPR contains Lys223 at the corresponding position. Multiple sequence alignments showed that Arg213 in *Pa*GPR is highly conserved among GPR homologs (Fig. 3).

Autoprocessing of *Bm*GPR is induced by low pH, increased level of dipicolinic acid, and under dehydration conditions. Upon autoprocessing, the N-terminal domain of *Bm*GPR (P_{41}) is truncated by 15 residues. To produce the active conformation of *Pa*GPR protein, we tried to recombinantly express N-terminally truncated constructs; however, the con-

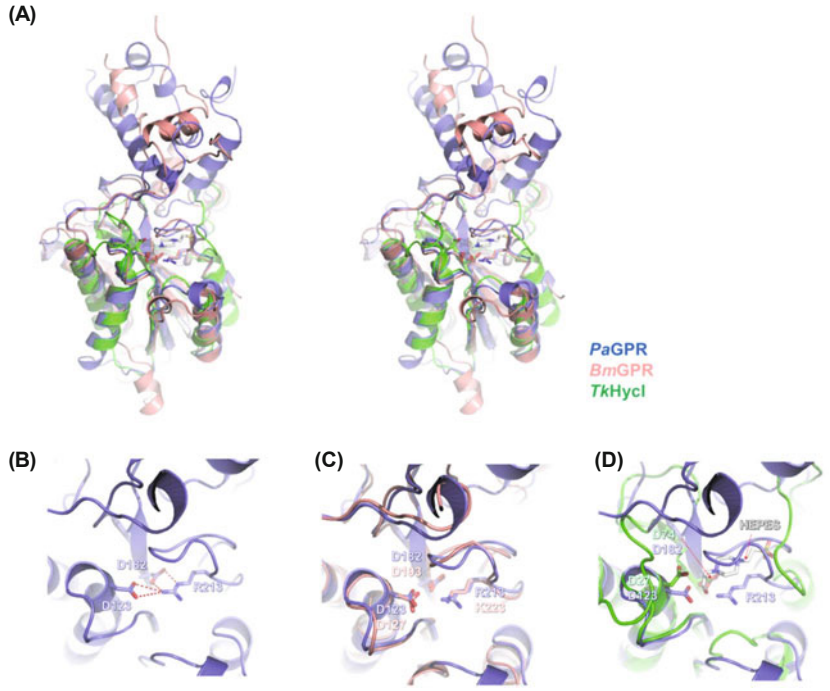

Fig. 3. **Self-inhibitory loop region of *Pa*GPR.** (A) Superposition of *Pa*GPR (green), *Bm*GPR (salmon), and hydrogenase maturation peptidase *Tk*HycI (green). (B) Close-up view of the self- inhibitory loop region of *Pa*GPR. The residues of interest Asp123, Asp182, and Arg213 are shown as sticks. (C) Superimposition of *Pa*GPR and *Bm*GPR. (D) Superimposition of *Pa*GPR and *Tk*HycI. The HEPES molecule in the *Tk*HycI structure is shown as white sticks.

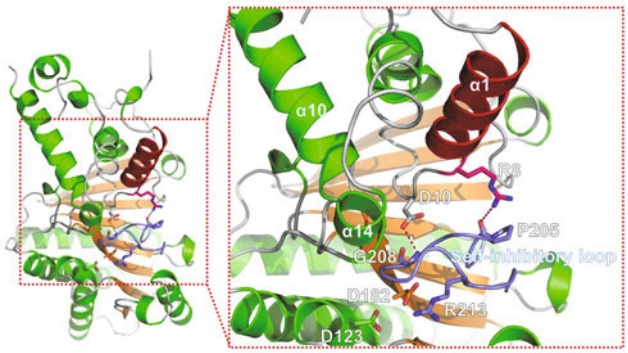

Fig. 4. Interaction between the N-terminal pro-peptide region and the self-inhibitory loop in *Pa*GPR. The pro-peptide region is colored red, and the self-inhibitory loop region is colored slate blue. Interacting residues are shown as sticks.

structs were expressed in insoluble inclusion bodies. In many zymogen proteases, the pro-peptide region is essential for proper folding, which explains the expression of truncated *Pa*GPR in inclusion bodies (Smith and Gottesman, 1989; Tao *et al.*, 1994; Ogino *et al.*, 1999). We tested the activation of the inactive zymogen *Pa*GPR by low pH and dipicolinic acid treatment *in vitro*, but with no success. It is thought that the activation process of *Pa*GPR differs from that of *Bm*GPR. In the present *Pa*GPR structure, the pro-peptide region (residues 1–9) forms the α1-helix, which strongly interacts with the α13-helix and the self-inhibitory loop region. In detail, Arg8 establishes polar interactions with the backbone carbonyl atom of Pro205. In addition, Asp10 exhibits polar interactions with the backbone nitrogen atom of Gly208 (Fig. 4). These observations suggest that the cleavage of the pro-peptide affects the conformation of the self-inhibitory loop region. Autoprocessing of *Pa*GPR may induce the displacement of the self-inhibitory loop region from the active site.

Structural comparison of *Pa*GPR and hydrogenase maturation peptidases

DALI search showed that *Pa*GPR shares remarkable similarity with hydrogenase maturation peptidase structures (HycI and HybD) (Fritsche *et al.*, 1999; Yang *et al.*, 2007; Kumarevel *et al.*, 2009; Kwon *et al.*, 2016, 2018). Although hydrogenase maturation peptidases lack a cap domain, the overall topology of their catalytic domains seems to be very similar to that of *Pa*GPR. In addition, hydrogenase maturation peptidases also have two Asp (or Glu) residues in their active sites. Hydrogenase maturation peptidases are metal-dependent endopeptidases, and the conserved Asp (or Glu) residues are involved in metal binding. Structural superimposition of *Pa*GPR and HycI from *Thermococcus kodakarensis* (*Tk*HycI, PDB ID: 5ZBY) (Kwon *et al.*, 2018) showed that the bound HEPES molecule in the putative substrate-binding site of *Tk*HycI is located at the same position as the inhibitory loop region in *Pa*GPR. This supports the assumption that the self-inhibitory loop region inhibits substrate binding in *Pa*GPR (Fig. 3D). Structural comparison between the metal-bound (Fritsche *et al.*, 1999) and metal-free HybD from *E. coli* (*Ec*HybD) (Yang *et al.*, 2007) is shown in Fig. 5. In the metal-free structure, Gln94 occupies the metal-binding site. However, during metal binding, Gln94 moves away from the metal-binding site, and the adjacent His93 residue moves into the site for metal coordination. Based on the structural similarities between the active sites of *Pa*GPR and hydrogenase maturation peptidase, we suggest that *Pa*GPR may have a similar activation process, involving the displacement of the self-inhibitory loop region from the active site and the entrance of catalytically essential residues into the active site via conformation changes triggered by pro-peptide cleavage.

Two possible enzymatic mechanisms of *Pa*GPR

Hitherto, the PDB contained only one GPR structure, namely, that of *Bm*GPR determined at 3.0 Å resolution (PDB

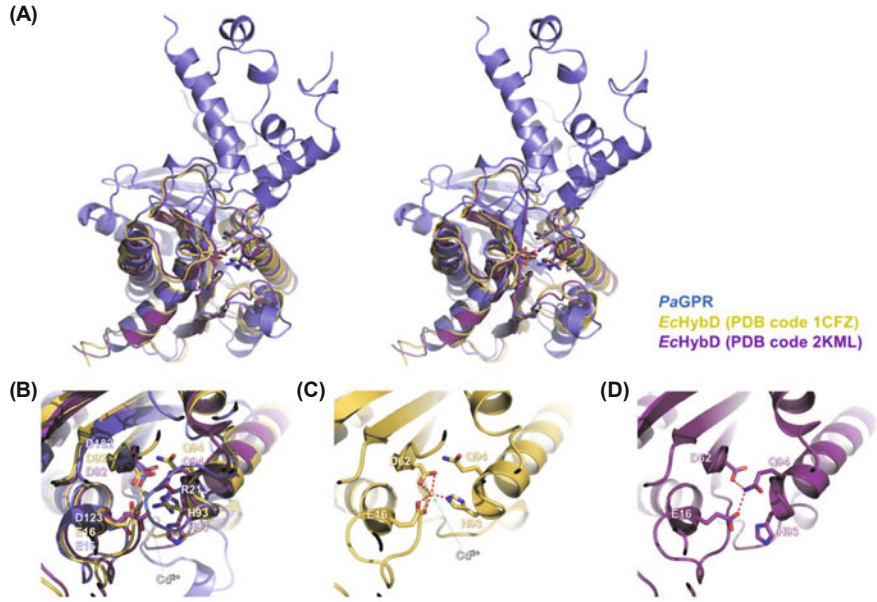

Fig. 5. Structural comparisons of *Pa*GPR and *Ec*HybD. (A) Superimposition of *Pa*GPR (slate), metal-bound *Ec*HybD (yellow), and metal-free *Ec*HybD (purple). (B) Close-up view of the superimposed self-inhibitory loop region. (C) Close-up view of the metal-bound *Ec*HybD. The bound cadmium metal ion is shown as a sphere. (D) Close-up view of the metal-free *Ec*HybD.

ID: 1C8B). Our higher resolution structure of *Pa*GPR enabled intensive structural comparisons. Based on structural analyses, it seems that the self-inhibited structure of *Pa*GPR differs from that of *Bm*GPR. Additionally, the *Pa*GPR structure provided details of the interactions between the N-terminal pro-peptide region and the self-inhibitory loop. We suggested two possible catalytic mechanisms for *Pa*GPR: (1) *Pa*GPR is an aspartic acid protease, similar to *Bm*GPR, or (2) the two conserved Asp residues in the active site of *Pa*GPR may bind metal ions such as *Ec*HybD (Fritsche *et al.*, 1999). Thus, *Pa*GPR may act as a metal-dependent protease. However, metal-binding by *Pa*GPR warrants further investigation.

Acknowledgments

We would like to thank the staff at the X-ray core facility of Korea Basic Science Institute (KBSI) (Ochang, Korea) and that at BL-5C of the Pohang Accelerator Laboratory (Pohang, Korea) for their kind help and support during data collection. This work was supported by the Korea Polar Research Institute (KOPRI; grant numbers PM20030 and PE20040), the Korea Basic Science Institute (grant numbers C030130, PN2019122), and the National Research Council of Science & Technology (CRC-16-01-KRICT).

Author Contribution

Chang Woo Lee: Investigation, visualization, writing - original draft. **Saeyoung Lee:** Data curation, writing - original draft. **Chang-Sook Jeong:** Data curation, investigation. **Jisub Hwang:** Resources, investigation. **Jeong Ho Chang:** Methodology, Investigation. **In-Geol Choi:** Investigation. **T. Doohun Kim:** Validation, investigation. **HaJeung Park:** Writing - review & editing. **Hye-Yeon Kim:** Supervision, writing - review & editing, acquisition of funding. **Jun Hyuck Lee:** Supervision, project administration, acquisition of funding, writing - review & editing

Conflict of Interest

The authors declare no conflict of interest.

References

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. 2012. Towards automated crystallographic structure refinement with *phenix. refine*. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367.

Carroll, T.M. and Setlow, P. 2005. Site-directed mutagenesis and structural studies suggest that the germination protease, GPR, in spores of *Bacillus* species is an atypical aspartic acid protease. *J. Bacteriol.* **187**, 7119–7125.

Chen, V.B., Arendall 3rd, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.

DeLano, W.L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, California, USA.

Emsley, P. and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.

Fritsche, E., Paschos, A., Beisel, H.G., Böck, A., and Huber, R. 1999. Crystal structure of the hydrogenase maturating endopeptidase HYBD from *Escherichia coli*. *J. Mol. Biol.* **288**, 989–998.

Illades-Aguiar, B. and Setlow, P. 1994. Autoprocessing of the protease that degrades small, acid-soluble proteins of spores of *Bacillus* species is triggered by low pH, dehydration, and dipicolinic acid. *J. Bacteriol.* **176**, 7032–7037.

Kumarevel, T., Tanaka, T., Bessho, Y., Shinkai, A., and Yokoyama, S. 2009. Crystal structure of hydrogenase maturating endopeptidase HycI from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **389**, 310–314.

Kwon, S., Nishitani, Y., Hirao, Y., Kanai, T., Atomi, H., and Miki, K. 2018. Structure of a [NiFe] hydrogenase maturation protease HycI provides insights into its substrate selectivity. *Biochem. Biophys. Res. Comm.* **498**, 782–788.

Kwon, S., Nishitani, Y., Watanabe, S., Hirao, Y., Imanaka, T., Kanai, T., Atomi, H., and Miki, K. 2016. Crystal structure of a [NiFe] hydrogenase maturation protease HybD from *Thermococcus kodakarensis* KOD1. *Proteins* **84**, 1321–1327.

Loshon, C.A. and Setlow, P. 1982. *Bacillus megaterium* spore protease: purification, radioimmunoassay, and analysis of antigen level and localization during growth, sporulation, and spore germination. *J. Bacteriol.* **150**, 303–311.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367.

Nessi, C., Jedrzejas, M.J., and Setlow, P. 1998. Structure and mechanism of action of the protease that degrades small, acid-soluble spore proteins during germination of spores of *Bacillus* species. *J. Bacteriol.* **180**, 5077–5084.

Ogino, T., Kaji, T., Kawabata, M., Satoh, K., Tomoo, K., Ishida, T., Yamazaki, H., Ishidoh, K., and Kominami, E. 1999. Function of the propeptide region in recombinant expression of active procathepsin L in *Escherichia coli*. *J. Biochem.* **126**, 78–83.

Otwinowski, Z. and Minor, W.I. 1997. Processing of X-ray diffraction data collected in oscillation mode. *In* Carter, C.Jr. and Sweet, R. (eds.), Macromolecular Crystallography Part A, vol. 276, pp. 307–326. Academic Press, New York, USA.

Pedersen, L.B., Nessi, C., and Setlow, P. 1997. Most of the pro-peptide is dispensable for stability and autoprocessing of the zymogen of the germination protease of spores of *Bacillus* species. *J. Bacteriol.* **179**, 1824–1827.

Ponnuraj, K., Kelly, S., Nessi, C., Setlow, P., and Jedrzejas, M.J. 2000a. Crystallization and preliminary diffraction studies of a truncated form of a novel protease from spores of *Bacillus megaterium*. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 70–72.

Ponnuraj, K., Nessi, C., Setlow, P., and Jedrzejas, M.J. 1999. Structural studies of a novel germination protease from spores of *Bacillus megaterium*. *J. Struct. Biol.* **125**, 19–24.

Ponnuraj, K., Rowland, S., Nessi, C., Setlow, P., and Jedrzejas, M.J. 2000b. Crystal structure of a novel germination protease from spores of *Bacillus megaterium*: structural arrangement and zymogen activation. *J. Mol. Biol.* **300**, 1–10.

Sanchez-Salas, J.L. and Setlow, P. 1993. Proteolytic processing of the protease which initiates degradation of small, acid-soluble proteins during germination of *Bacillus subtilis* spores. *J. Bacteriol.* **175**, 2568–2577.

Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. 2002. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems.

Biophys. J. **82**, 1096–1111.
Setlow, P. 1995. Mechanisms for the prevention of damage to DNA in spores of *Bacillus* species. *Annu. Rev. Microbiol.* **49**, 29–54.
Smith, S.M. and Gottesman, M.M. 1989. Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *J. Biol. Chem.* **264**, 20487–20495.
Tao, K., Stearns, N.A., Dong, J., Wu, Q.L., and Sahagian, G.G. 1994. The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Arch. Biochem. Biophys.* **311**, 19–27.
Vagin, A. and Teplyakov, A. 2010. Molecular replacement with *MOLREP*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25.
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., *et al.* 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242.
Yang, F., Hu, W., Xu, H., Li, C., Xia, B., and Jin, C. 2007. Solution structure and backbone dynamics of an endopeptidase HycI from *Escherichia coli*: implications for mechanism of the [NiFe] hydrogenase maturation. *J. Biol. Chem.* **282**, 3856–3863.