Discrimination based on Gly and Arg/Ser at Position 673 between dipeptidyl-peptidase (DPP) 7 and DPP11, widely distributed DPPs in pathogenic and environmental Gram-negative bacteria

Shakh M. A. Rouf[a], Yuko Ohara-Nemoto[a], Tomonori Hoshino[b], Taku Fujiwara[b], Toshio Ono[a], and Takayuki K. Nemoto[a,] *

[a]*Department of Oral Molecular Biology*, *Course of Medical and Dental of Sciences, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki 852-8588, Japan*
[b]*Department of Pediatric Dentistry, Course of Medical and Dental of Sciences, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki 852-8588, Japan*

Running title: *Discrimination between DPP7 and DPP11 based on residue at 673*

*Corresponding author. Tel.: +81 95 819 7640. Fax: +81 95 819 7642.
E-mail address: tnemoto@nagasaki-u.ac.jp (T.K. Nemoto)
Department of Oral Molecular Biology, Course of Medical and Dental of Sciences, Nagasaki University Graduate School of Biomedical Sciences, 1-7-1 Sakamoto, Nagasaki 852-8588, Japan

ABSTRACT

Porphyromonas gingivalis, an asaccharolytic gram-negative rod-shaped bacterium, expresses the novel Asp/Glu-specific dipeptidyl-peptidase (DPP) 11 (Ohara-Nemoto, Y. *et al*., (2011) *J. Biol. Chem*. 286, 38115-38127), which has been categorized as a member of the S46/DPP7 family that is preferential for hydrophobic residues at the P1 position. From that finding, 129 gene products constituting five clusters from the phylum *Bacteroidetes* have been newly annotated to either DPP7 or DPP11, whereas the remaining 135 members, mainly from the largest phylum *Proteobacteria*, have yet to be assigned. In this study, the substrate specificities of the five clusters and an unassigned group were determined with recombinant DPPs from typical species, i.e., *P. gingivalis*, *Capnocytophaga gingivalis*, *Flavobacterium psychrophilum, Bacteroides fragilis*, *Bacteroides vulgatus*, and *Shewanella putrefaciens*. Consequently, clusters 1, 3, and 5 were found to be DPP7 with rather broad substrate specificity, and clusters 2 and 4 were DPP11. An unassigned *S. putrefaciens* DPP carrying Ser^{673} exhibited Asp/Glu-specificity more preferable to Glu, in contrast to the Asp preference of DPP11 with Arg^{673} from *Bacteroidetes* species. Mutagenesis experiments revealed that Arg^{673}/Ser^{673} were indispensable for the Asp/Glu-specificity of DPP11, and that the broad specificity of DPP7 was mediated by Gly^{673}. Taken together with the distribution of the two genes, all 264 members of the S46 family could be attributed to either DPP7 or DPP11 by an amino acid at position 673. A more compelling phylogenic tree based on the conserved C-terminal region suggested two gene duplication events in the phylum *Bacteroidetes*, one causing the development of DPP7 and DPP11 with altered substrate specificities, and the other producing an additional DPP7 in the genus *Bacteroides.*

Keywords: dipeptidyl peptidase, S46 peptidase, substrate specificity, *Porphyromonas gingivalis*, *Bacteroidetes*, *Proteobacteria*.

Abbreviations: DPP, dipeptidyl-peptidase; PgDPP7, DPP7 from *P. gingivalis*; PgDPP11 and PeDPP11, DPP11 from *P. gingivalis* and *P. endodontalis*, respectively; Arg^{673}- and Ser^{673}-DPP11, DPP11 carrying Arg^{673} and Ser^{673}, respectively; GluV8, glutamyl endopeptidase from *Staphylococcus aureus*; ac, acetyl; Z, benzyloxycarbonyl; MCA, 4-methycoumaryl-7-amide; NJ, neighbor-joining.

1. Introduction

Porphyromonas gingivalis is a major causative organism of aggressive forms of chronic periodontitis [1,2] leading to loss of permanent teeth [3-5], and also detected from extraoral sources especially occluded arteries [6,7] and appendicitis [8]. This bacterium grows exclusively under an anaerobic condition, forms biofilm [9], and does not utilize sugar, whereas it requires proteinaceous substrates as carbon and energy sources [10,11]. Hence, *P. gingivalis* is considered to be a model organism for studying bacterial pathogens possessing a metabolic network that does not require sugar and oxygen.

Cell-associated peptidases in *P. gingivalis*, such as dipeptidyl-peptidases (DPPs), prolyl tripeptidyl-peptidase A, as well as lysyl and arginyl gingipains, are responsible for the initial step of its energy metabolism, because the organisms predominantly incorporate amino acids as di- and tri-peptides from the surrounding environment [12, 13]. To date, three DPPs in *P. gingivalis* have been well characterized. DPPIV specifically cleaves a C-terminal carbon-nitrogen bond of Pro at the penultimate position (P1 position) from the N-terminus of substrates [14] and DPP7 is preferential for hydrophobic residues at the P1 position [15], while we recently reported Asp/Glu-specific DPP11, which is also expressed in *Porphyromonas endodontalis*, an important pathogen in periapical lesions [16]. The discovery of *P. gingivalis* DPP11 (PgDPP11) solved a long-lasting puzzle related to findings that peptidases responsible for the degradation of polypeptides carrying Asp and Glu were apparently absent, while Glx (Glu and Gln) and Asx (Asp and Asn) were the most abundantly utilized amino acids in *P. gingivalis* [12,13,17]. In addition to these DPPs, genome information suggests existence of the DPPIII gene (*PGN1645*) [11], which is considered to be expressed solely in eukaryotic cells [18-20]. When Pro is located at the third position from the N-terminus, prolyl tripeptidyl-peptidase A specifically cleaves an NH_2-X-X-Pro-|-X bond [21, 22].

PgDPP11 encoded by the *PG1283* gene was previously classified as an unassigned member of DPP7 (*PG0491*, PgDPP7) in the S46 family (MEROPS peptidase database, Release 9.5) [23] or considered to be an isoform of DPP7 [15]. As a consequence of the discovery of PgDPP11, 129 homologues in the family from the phylum *Bacteroidetes*, previously known as the *Cytophaga-Flavobacterium-Bacteroides* (CFB) group, are currently annotated to either DPP7 (S46.001) or DPP11 (S46.002). Intriguingly, though many species possess a pair of DPP7 and

DPP11 genes, the genus *Bacteroides* in this phylum exceptionally contains three S46-family genes. Hence, it is of great interest to determine whether the third gene, which is currently annotated to DPP11, truly possesses the substrate specificity of DPP11 or exhibits a novel specificity. Furthermore, unassigned DPP7/11 members are widely distributed to the largest phylum *Proteobacteria*, which comprises the majority of known gram-negative bacteria with medical, industrial, and agricultural significance [24]. These include the genus *Ferrimonas*, which is chemoorganotrophic [25], *Colwellia*, which adapts to cold marine environments [26], and *Xanthomonas*, which causes plant diseases [27]. Therefore, validation of the hydrolytic activities and specificities of these family members is important to elucidate their amino acid metabolism.

The specificity of *P. endodontalis* DPP11 (PeDPP11) is primarily determined by the interaction of Arg^{673} with the carboxyl group at the P1-position Asp/Glu of a substrate peptide [16]. The amino acid at the corresponding position is Gly^{673} in PgDPP7, and exclusively either Arg or Gly in all DPP7 and DPP11 members from the phylum *Bacteroidetes*. Gly^{673} in DPP7 seems to be compatible for its hydrophobic residue preference, however this notion has not been directly tested because of the unavailability of recombinant DPP7 to date. In addition to the P1 position, there is also evidence that the hydrophobicity of the P2-position residue is important for the activity of DPP11 [16]. In contrast, DPP7 and DPPIV seem to have no preference for other residues except that Pro is unacceptable at the third position (P1' position) [14, 23], although few peptides were tested with DPP7 [15, 28].

In the present study, we determined the specificities of five DPP7/11 clusters from the phylum *Bacteroidetes* and an unassigned DPP7/11 group from the phylum *Proteobacteria* in the latest phylogenic tree of the S46 family. Interestingly, the species of the genus *Bacteroides* carried one DPP11 and two DPP7 genes, one of which is currently annotated to DPP11 by sequence homology. In addition, the specificity of an unassigned *Shewanella* DPP that carries Ser^{673} was determined to be DPP11 subtype with the preference to Glu in contrast to Arg^{673}-DPP11. Based on these biochemical findings, we propose that DPP7 and DPP11 of all 264 S46-family members should be unambiguously attributed to either DPP7 or DPP11.

2. Materials and methods

2.1. Materials

The expression and cloning vectors used were pTrcHisTOPO from Invitrogen and pQE60 from Qiagen. Low-molecular-weight markers were obtained from GE Healthcare. Restriction enzymes and DNA-modifying enzymes came from Takara Bio and New England Biolabs, respectively, while KOD Plus DNA polymerase came from Toyobo (Tokyo, Japan). Met-Leu-MCA was obtained from Bachem (Bubendorf, Switzerland), and Leu-Asp-, Leu-Glu-, acetyl (ac)-Leu-Asp-, and benzyloxycarbonyl (Z)-Leu-Leu-Gln-MCA were synthesized by Thermo Fisher Scientific (Ulm, Germany) and TORAY (Tokyo, Japan). Other MCA peptides were from the Peptide Institute Inc. (Osaka, Japan). Oligonucleotide primers came from FASMAC (Atsugi, Japan). Thermolysin from *Bacillus thermoproteolyticus rokko* and trypsin from bovine pancreas were obtained from Sigma-Aldrich. Recombinant Glu-specific endopeptidase from *Staphylococcus aureus* (GluV8) was expressed and the subsequent conversion to a mature form was performed as previously reported [29].

2.2. Bacterial strains

Capnocytophaga gingivalis JCM12953 and *Flavobacterium psychrophilum* JCM8519 were provided by RIKEN BRC through the National Bio-Resource Project of MEXT, Japan. *P. gingivalis* ATCC33277 was obtained from American Type Culture Collection. Each was cultured in ABCM (anaerobic bacteria culture medium) broth (Eiken Chemical, Tokyo, Japan) in the presence of 1 μg/ml of menadione (Sigma-Aldrich) at 35˚C under an anaerobic condition (90% N_2/5% H_2/5% CO_2).

2.3 Construction of expression plasmids

We prepared genomic DNA from *P. gingivalis*, *C. gingivalis*, and *F. psychrophilum* as previously reported [30]. Genomic DNA from *Bacteroides fragilis* strain Onslow JGD06283, *Bacteroides vulgatus* JGD06281, and *Shewanella putrefaciens* JCM20190T were provided by RIKEN BRC. DNA fragments encoding cluster-1 and -2 DPPs of *F. psychrophilum* (locus tag: FP1355 and FP0382, respectively) and *C. gingivalis* (CAPGI0001_0817 and CAPGI0001_1068, respectively), cluster-5 DPPs from *B. fragilis* (BF_3010) and *B. vulgatus* (BVU_2253), DPP

from *S. putrefaciens* (Sputcn32_0757), and *P. gingivalis* DPPIV (PGN_1469) were amplified by PCR with genomic DNA as templates and the primers listed in supplemental Table S1. PCR fragments of cluster-2 DPP from *C. gingivalis* and cluster-1 and -2 DPPs *F. psychrophilum* were directly inserted into pTrcHisTOPO. The PCR fragments of *P. gingivalis* DPPIV and PgDPP11 (PGN_0607), the cluster-1 DPP fragment from *C. gingivalis*, and the fragments from *B. fragilis* and *S. putrefaciens* were digested with *Bam*HI, and cloned into the *Bam*HI site of pQE60. PCR fragments encoding PgDPP7 (PGN_1479) and cluster-5 DPP from *B. vulgatus* were digested with *Bgl*II, and cloned into the *Bam*HI site of pQE60. Among them, since cluster-1 DPP from *F. psychrophilum* was poorly expressed in this *E. coli* expression system, further biochemical analysis was not performed.

2.4. Amino acid numbering and in vitro mutagenesis

To specify the positions of amino acid residues involved in the specificity of the DPPs, the numbering was unified to that of PgDPP11 throughout this study. *In vitro* mutagenesis was performed using a PCR technique with appropriate primers (supplemental Table S1) to introduce an amino acid substitution of Gly^{673} in PgDPP7 to Arg, Ala, or Asp. Arg^{673} in PgDPP11 was substituted to Ser. Ser^{673} and Lys^{680} in the *S. putrefaciens* DPP were substituted to Arg and Gly, respectively, or simultaneously. All substitutions were confirmed by nucleotide sequencing.

2.5. Expression and purification of DPPs

Recombinant DPPs were expressed in *Escherichia coli* XL1-blue by induction with 0.2 mM isopropyl-thiogalactopyranoside at 30 ˚C for 4 h. Recombinant proteins were purified by Talon affinity chromatography, as previously reported [29].

2.6. Measurement of hydrolyzing activity

Reaction was started by addition of DPPs to reaction solution (200 μL) contained 50 mM sodium phosphate (pH 7.0), 5 mM EDTA, and 20 μM MCA-substrate. After 1 h incubation at 37 ˚C, fluorescence intensity was measured with excitation at 380 nm and emission at 460 nm. Several dipeptidyl-MCA substrates were prepared from oligopeptidyl-MCA (0.4 mM) as previously reported [16]. In practical terms, boc-Ile-Glu-Gly-Arg- and boc-Glu-Lys-Lys-MCA were cleaved with GluV8 to produce Gly-Arg- and Lys-Lys-MCA, respectively, and

Z-Leu-Arg-Gly-Gly-MCA with trypsin to produce Gly-Gly-MCA, while ac-Lys-Thr-Lys-Gln-Leu-Arg-, Z-Val-Val-Arg-, and boc-Val-Leu-Lys-MCA were pretreated with thermolysin to produce Leu-Arg-, Val-Arg-, and Leu-Lys-MCA, respectively.

2.7. Construction of phylogenic trees of the S46 family

PeDPP11, which was cloned recently in our laboratory [16], and 263 members of the S46 family in the MEROPS database (Release 9.6) were used for construction of phylogenic trees. ClustalX software [31] (http://www.ebi.ac.uk) was used to align the sequences. The full-length amino acid sequences and sequences of the conserved regions were used for construction of the phylogenic trees. Phylogenic analysis with the neighbor-joining (NJ) algorithm [32] was conducted using MEGA version 5 [33].

3. Results

3.1. Peptidase activities of DPPs belonging to clusters 1-5 from the phylum Bacteroidetes

In the S46 family listed in the previous MEROPS database (Release 9.5), 71 members derived from the genus *Bacteroides* constituted only a single group, S46/DPP7, with 3 members left unassigned [16]. In the version (Release 9.6), revised after the finding of DPP11, the number of S46 family members reached 263, which were newly classified into S46.001/DPP7, S46.002/DPP11, and unassigned members (supplemental Fig. S1). Among them, 129 homologues from the phylum *Bacteroidetes* constitute five clusters, in which cluster 3 includes PgDPP7 and cluster 4 includes PgDPP11. Members of clusters 2 and 5 are newly annotated to DPP11, whereas it has not been decided where cluster 1 should reside in either of the two DPPs. Although there are several members with ambiguous locations in the branch, those are currently annotated to DPP11 (supplemental Fig. S1, asterisks 1-4). The remaining 135 members, mainly from the largest phylum *Proteobacteria*, could be classified into three groups, of which only group 3 is annotated to DPP7, while groups 1 and 2 remain unannotated (supplemental Fig. S1).

Thus, since the S46 family members, except for PgDPP7 (PG_0491/PGN_1479), PgDPP11 (PG_12831/PGN_0607), and PeDPP11, have not been biochemically characterized, it is worthwhile to determine directly the enzymatic activities of all five clusters. We chose DPPs

from typical bacterial species as follows; cluster-1 and -2 DPPs from *C. gingivalis* and *F. psychrophilum*, cluster-5 DPPs from *B. fragilis* and *B. vulgatus*, as well as PgDPP7 in cluster 3, and PgDPP11 in cluster 4. Furthermore, *P. gingivalis* DPPIV (PG_1361/PGN_1469), which belongs to the S9 peptidase family, was also examined as a representative of non-S46 family. All DPPs were successfully expressed by use of an *E. coli* expression system and purified as 70-80-kDa species, except for cluster-1 DPP from *F. psychrophilum* (supplemental Fig. S2).

The substrate specificity of these recombinant DPPs was characterized using 11 dipeptidyl MCA substrates (Fig. 1). In accord with a previous report [14], recombinant DPPIV most preferentially hydrolyzed Gly-Pro-MCA, while that of Lys-Ala-MCA was less potent. Hydrolysis of Met-Leu-MCA was nearly negligible as compared to these two substrates. These results indicate that recombinant DPPIV faithfully reproduced the property of DPPIV expressed in *P. gingivalis.* DPP7 from *P. gingivalis* has been reported to preferentially cleave a peptide bond next to a hydrophobic amino acid and Ala at the P1 position [15]. This property was reproduced with recombinant DPP7, which most potently cleaved Met-Leu-MCA and Lys-Ala-MCA to a lesser extent. In addition, DPP7 hydrolyzed dipeptidyl MCA carrying charged residues at the P1 position, such as Leu-Arg-, Leu-Asp-, Leu-Glu-, Val-Arg-, and Leu-Lys-MCA, to some extent. An apparent characteristic of these substrates was the presence of hydrophobic residues at the N-terminus (P2 position). PgDPP11 solely cleaved Leu-Asp- and Leu-Glu-MCA, as previously reported [16].

The substrate specificity of *C. gingivalis* DPP in cluster 1, which is annotated as DPP11 in the latest classification (Release 9.7), was similar to that of PgDPP7, with the most potency toward Met-Leu-MCA, followed by Lys-Ala-, Leu-Arg- > Leu-Asp-, Leu-Glu-, > Leu-Lys-, and >Val-Arg-MCA, while this DPP did not hydrolyze Gly-Arg-, Gly-Gly-, Lys-Lys-, or Gly-Pro-MCA (Fig. 1). Therefore, this *C. gingivalis* DPP could be classified to DPP7. *C. gingivalis* and *F. psychrophilum* DPPs in cluster 2 exhibited the specificity as that of PgDPP11 in cluster 4. Finally, *B. fragilis* and *B. vulgatus* cluster-5 DPPs demonstrated a substrate preference similar to that of cluster-1 and -3 DPP7, although they have been annotated to DPP11. This finding indicates that the discrimination between DPP7 and DPP11 is not simply achieved by the comparison of full-length amino acid sequences, and thus, we suspect that the discrimination might be achieved by the amino acid sequence region critical for the substrate specificity of DPP7 and DPP11.

3.2. Amino acid residue at position 673 critical to define substrate specificity of DPP7 and DPP11

The alignment of the DPP sequences of 129 members of clusters 1-5 shows that three conserved regions, each of which includes one of the essential residues of His^{85}, Asp^{198}, and Ser^{655} of serine proteases (Fig. 2). Forty-three amino acid residues are conserved in more than 95% of the members. These highly conserved residues may be indispensable for functions common to both DPP7 and DPP11, whereas residues specific for DPP7 or DPP11 may be involved in each DPP-specific role. In this respect, the amino acid residue at position 673 was unique, and Arg^{673} was entirely conserved in members of clusters 2 and 4 (DPP11), whereas the position was exclusively Gly in clusters 1, 3, and 5 (DPP7).

We previously demonstrated an ionic interaction between Arg^{673} of DPP11 and the carboxyl group of the P1-position Asp/Glu of a peptide substrate [16]. By analogy, it was postulated that the tiny and hydrophobic residual group (hydrogen) of Gly^{673} is a prerequisite for DPP7 activity to accept hydrophobic and bulky residues of a substrate at the P1 position. Hence, this hypothesis was examined by use of a single amino acid substitution of Gly^{673} of PgDPP7 to Asp, Ala, and Arg (Fig. 3). The hydrolyzing activity of PgDPP7 toward Met-Leu-MCA was abolished by the Gly673Asp substitution (1.7%) and reduced to 22% by the Gly673Arg substitution, indicating the need for a non-charged residue at this position. In the Gly673Ala substitution, 61% of the Met-Leu-MCA hydrolyzing activity was maintained. These results strongly suggest that the tiny hydrogen group (-H) of Gly^{673} is more suitable than a methyl group ($-CH_3$) of Ala, presumably because of acceptance of the bulky hydrophobic residue characteristic of DPP7. Furthermore, it is of interest to note that the Gly673Arg substitution to mimic Arg^{673} of DPP11 could acquire the activity toward Leu-Asp- and Leu-Glu-MCA. Thus, the residue at position 673 is critical for both hydrolyzing activity and the substrate specificity of DPP7 as well as DPP11. Although we also tested whether PgDPP7 Gly673Asp hydrolyzed the peptide carrying basic residue at the P1 position, it failed to hydrolyze Leu-Arg-MCA (data not shown), indicating that the residue at 673 is not the unique determinant of the substrate specificity.

Based on these biochemical findings together with the sequence homology, we considered reasonably proposing that newly discovered DPP7/11 homologues from the phylum *Bacteroidetes* can be readily assigned based on the amino acid residue at position 673.

3.3. DPP from the phylum Proteobacteria and expression of Shewanella putrefaciens DPP carrying Ser^{673}

The S46 family contains 139 unassigned members mainly from the phylum *Proteobacteria*, which constitutes three major groups (supplemental Fig. S1, Table S1). Among them, most members in groups 1 and 3 carry Gly at position 673, thus they should be DPP7. On the other hand, 30 members constituting group 2 were shown to possess neither Gly nor Arg, but rather Ser at the position. Since most members of group 2 were from the genus *Shewanella*, we expressed DPP (MEROPS ID: MER096825) from *S. putrefaciens* to determine the peptidase activity of these members carrying Ser^{673}. *S. putrefaciens* is a typical species of the genus that has been isolated mostly from fish, poultry, and meats, but is also associated with a broad range of human infections [34, 35]. It was found that *S. putrefaciens* DPP specifically cleaved Leu-Asp- and Leu-Glu-MCA, demonstrating its entity as DPP11 (Fig. 4). This conclusion was compatible with the fact that 18 of 19 *Shewanella* species possess a pair of genes encoding DPP7 with Gly^{673} and DPP11 with Ser^{673} (supplemental Table S2).

Interestingly, *S. putrefaciens* Ser^{673}-DPP11 preferentially cleaved Leu-Glu-MCA as compared to Leu-Asp-MCA, in contrast to the Leu-Asp-MCA preference of cluster-2 and -4 DPP11 (Figs. 1 and 5). In order to investigate the relationship between the amino acid at position 673 and Asp- and Glu-preferences, an amino acid swapping experiment was performed. In the Arg673Ser mutant of PgDPP11, the Asp/Glu preference was inverted, i.e., peptidase activity toward Leu-Asp-MCA nearly disappeared, whereas that to Leu-Glu-MCA partially remained and *vice versa* for *S. putrefaciens* DPP11 (Fig. 5B).

These results of specificity described above strongly suggest that the amino acid residue at position 673 is critical not only for discrimination between DPP7 and DPP11, but also for the preference of either Asp or Glu in DPP11. Ser^{673} was specifically located in DPP11 of group 2 distributed in the genera *Shewanella*, *Ferrimonas*, *Pseudoalteromonas*, *Stenotrophomonas*, and *Pseudoxanthomonas*, while a few members carrying Ser^{673} were exceptionally observed in group 1 from the genera *Rhodothermus*, *Salinibacter*, and *Brevundimonas* (supplemental Fig. S1, asterisks 5 and 7; supplemental Table S2).

As for the interaction with the Asp/Glu residue at the P1 position of substrates, a specific basic residue was surmised in Ser^{673}-DPP11. Based on the alignment, we found that Lys^{680} is

conserved in 21 Ser^{673}-DPP11 members of all *Shewanella* species, and two other Ser^{673}-DPP11 from *Ferrimonas balearica* and *Pseudoalteromonas tunicate* (supplemental Table S2), and that most Arg^{673}-DPP11 carry Gly^{680} (Fig. 2). In fact, substitution experiment findings demonstrated that Ser^{673}-DPP11-Lys680Gly abrogated that activity (Fig. 5). These results suggested that Lys^{680}, but not Ser^{673}, could manage interactions with carboxyl group of substrates. Furthermore, Lys680Gly substitution disrupted the acidic residue specificity, as it more efficiently degraded Ala-Asn- and Leu-Gln-MCA (insert in Fig. 5). Finally, the double mutations (Ser673Arg and Lys680Gly) that mimicked the amino acids of PgDPP11 induced a partial restoration of activity for Leu-Asp-MCA. Therefore, these results indicate the importance of Lys^{680} in the enzymatic activity of Ser^{673}-DPP11.

3.4. A new phylogenic tree of S46 family

We constructed phylogenic trees based on each or total of the three conserved regions (see Fig. 2). With region 1, cluster-4 members of the phylum *Bacteroidetes* were split into three distinct locations: For instance, PgDPP11 was separated from PeDPP11 (data not shown). With region 2, each cluster was distributed same as that classified with the full-length sequences (data not shown). In contrast, the tree constructed using region 3 (Pro^{571}/Ser^{571}-Leu^{700}) that carries an essential Ser^{655} and specificity determining Gly/Arg/Ser^{673} (Fig. 6) may unify *Bacteroidetes* DPP7 members into a single group comprising clusters 1, 3, and 5, and further accumulates cluster-2 and -4 DPP11 at adjacent locations, as compared to the tree based on full-length sequences (supplemental Fig. S1). In addition, the phylogenic tree obtained from the combination of three regions (supplemental Fig. S3) was basically identical to that of region 3.

Accordingly, the phylogenic tree based on region 3 as well as that of the combination of three regions enabled us to propose a simple and systematic evolutionary route of DPP7 and DPP11 in the phylum *Bacteroidetes*. That is, the single ancestor gene of the S46 family was initially duplicated into DPP7 and DPP11 genes in the phylum *Bacteroidetes.* Thereafter, a second gene duplication may occur only in the DPP7 gene, which produced two DPP7 molecules belonging to clusters 3 and 5 (Fig. 6).

4. Discussion

In order to determine the substrate specificities of cluster-1 to -5 DPPs from the phylum *Bacteroidetes* of the S46 family, we expressed at least two members from each of cluster 1, 2, and 5, while cluster-1 DPP from *Flavobacterium psychrophilum* was not successfully expressed. Moreover, we confirmed that the substrate specificities of *Bacteroides fragilis* DPP7 in cluster 3 (T. K. Nemoto and H. Nishimata, unpublished observation) and *P. endodontalis* DPP11 in cluster 4 [16] were identical to those from *P. gingivalis*. Taken together with the sequence similarity within each cluster, it was considered reasonable to conclude that members of one cluster share an unique substrate specificity with either DPP7 or DPP11. In addition, the finding that most species possess a set of DPP7 and DPP11 genes, except for the genus *Bacteroides*, confirmed the validity of the classification.

The present study demonstrated that the amino acid at position 673 is involved in enzyme activity together with the catalytic triad His^{85}, Asp^{198}, and Ser^{655}, and is tightly associated with the substrate specificities of DPP7 and DPP11 (Figs. 3 and 5). The substrate preference of DPP7 has been reported to be aliphatic or aromatic amino acid residues at the P1 position [15]. However, recombinant PgDPP7 cleaved Leu-Arg-MCA at an efficiency of 55% as compared to that of Met-Leu-MCA (Fig. 1). DPP7 from *Capnocytophaga gingivalis* (cluster 1), *Bacteroides fragilis* (cluster 5*)*, and *Bacteroides vulgatus* (cluster 5) also cleaved Leu-Arg-MCA, thus DPP7 probably possesses a broader specificity, even for negatively and positively charged substrates at the P1 position, than previously reported. An extensive analysis with various dipeptidyl MCA substrates indicated the hydrophobic preference of DPP7 at the P2 position as well as the P1 position, which partly overcame the inaccessibility to the P1 position residue (S.M.A. Rouf, Y. Ohara-Nemoto, and T. K. Nemoto, unpublished observation). These results may be compatible with a previous observation [15], in which native DPP7 produced dipeptides from oligopeptides with hydrophobic residues at the P2 position.

Our study also demonstrated that Ser^{673}-DPP11 is capable of Asp/Glu-specific hydrolysis and that Ser^{673}-DPP11 exhibits a Glu preference in contrast to the Asp preference of authentic Arg^{673}-DPP11 (Fig. 5). We previously reported that the size of Arg^{673} defined the preference to Asp, because the Asp preference was gradually shifted to Glu by the substitution of Arg^{673} to the smaller amino acid residues His, Gln, and Gly, in that order [16]. By analogy, the Glu preference

of Ser^{673}-DPP11 would be explained by the conversion from Arg^{673} to Ser. In addition, based on the amino acid sequence alignment, the ε-amino group of Lys^{680} in Ser^{673}-DPP11 is suspected to locate at the position of the guanidinium group of Arg^{673} of PgDPP11 in a steric manner. As expected, we found that substitution of Lys^{680} to Gly, which mimicked the amino acid conversion from Ser^{673}-DPP11 to Arg^{673}-PgDPP11, abrogated the activity, suggesting that Lys^{680} is involved in substrate specificity and enzymatic activity (Fig. 5).

Among three conserved regions of the S46 family members, C-terminal conserved region 3 composed of 230 residues seems to be most important for activity and specificity, since the alkoxide of catalytic Ser^{655} is directly associated with the carbonyl carbon of Asp/Glu in a substrate and initiates the cleavage of peptide bond [40]. Moreover, the complete sequence homology around Ser^{655}, i.e., Thr^{650}-Thr-Gly-Gly-Asn-Ser^{655}-Gly-Ser-Pro-Val^{659}, between PgDPP11 and GluV8 indicates the importance of this sequence in enzymatic activity [36]. In particular, using the analogy of Gly^{193} and Ser^{195} of trypsin [37, 38], the backbone amide hydrogen atoms of Gly^{653} and Ser^{655} may serve to stabilize the developing negative charge on the carbonyl oxygen atom of the cleaved amides. Furthermore, Gly^{673} and Arg^{673}/Ser^{673}, critical for specificity, were localized in the same region (Fig. 2). The residue at position 673 may equivalent to Asp^{185} of trypsin located in the catalytic pocket responsible for attracting and stabilizing the charged P1-position residue of a substrate. This assumption may be also true for Lys^{680} of Ser^{673}-DPP11.

Hoshino *et al.* [39] reported the evolution of the cariogenic characteristics of *Streptococcus mutans* using glucosyltransferase. In that study, a middle catalytic domain (residues 400-900) instead of the 1500-amino acid full-length sequence of glucosyltransferase was used for the alignment, because the rest of the sequence varied enormously among species. In accord with that study, the present phylogenic tree based on the conserved region 3 as well as the combination of three conserved regions indicates an evolutionary route of DPP7 and DPP11, which could more reasonably explain the relationships among the five clusters of the S46 family compared to the tree based on the full-length sequence. Proteins generally accumulate amino acid substitutions at variable rates based on the requirements of the sequences. Hence, for multi-domain proteins, specificity-determining restricted regions may be preferable for study of their classification. By use of the C-terminal conserved region 3 for comparison, we could figure a simple evolutionary route of DPP7 and DPP11 genes in phylum *Bacteroidetes* (Fig. 6). We

further examined whether classification based on the conserved region 3 could be successfully expanded to DPPs from phyla other than *Bacteroidetes*. They were tentatively separated into three groups, of which groups 2 and 3 were exclusively composed of Ser^{673}-DPP11 and Gly^{673}-DPP7, respectively (Fig. 6).

Acknowledgements

We greatly acknowledge Mariko Naito (Nagasaki University) for valuable discussion and Takeshi Kobayakawa for technical assistance. This work was supported in grants-in-aid for scientific research from the Ministry of Education, Science, Sports, and Culture of Japan (to Y. O.-N., S. K. and T. K. N.) and from the Institute for Fermentation, Osaka (to T. K. N.). S. M. A. Rouf is a graduate student supported by Japanese Government (MEXT) Scholarship.

References

[1] P. Kastelein, T.J. van Steenbergen, J.M. Bras, P, Kastelein, J. de Graaff, An experimentally induced phlegmonous abscess by a strain of *Bacteroides gingivalis* in guinea pigs and mice. Antonie Van Leeuwenhoek 47 (1981) 1-9.

[2] T.J.M. van Steenbergen, A.J. van Winkelhoff, D. Mayrand, D. Grenier, J. de Graaff, *Bacteroides endodontalis* sp. nov., an asaccharolytic black-pigmented *Bacteroides* species from infected dental root. Canals Int. J. Syst. Bacteriol. 34 (1984) 118-120.

[3] D. White, D. Mayrand, Association of oral *Bacteroides* with gingivitis and adult periodontitis. J. Periodont. Res. 16 (1981) 259-265.

[4] W.E.C. Moore, L.V. Holdeman, R.M. Smibert, D.E. Hash, J.A. Burmeister, R.R. Ranney, Bacteriology of severe periodontitis in young adult humans. Infect. Immun. 38 (1982) 1137-1148.

[5] W.J. Loesche, S.A. Syed, E. Schmidt, E.C. Morrison, Bacterial profiles of subgingival plaques in periodontitis. J. Periodontol. 56 (1985) 447-456.

[6] T. Y. Inoue, M. Umeda, Y. Huang, N. Kurihara, M. Koike, I. Ishikawa, Oral bacteria in the occluded arteries of patients with Buerger disease. J. Vasc. Surg. 42 (2005) 107-115.

[7] T. Toyofuku, Y. Inoue, N. Kurihara, T. Kudo, M. Jibiki, N. Sugano, M. Umeda, Y. Izumi, Differential detection rate of periodontopathic bacteria in atherosclerosis. Surg Today. 41 (2011) 1395-1400.

[8] J. Mättö, S. Asikainen, M.L. Vaisanen, M. Rautio, M. Saarela, P. Summanen, S. Finegold, H. Jousimies-Somer, Role of *Porphyromonas gingivalis*, *Prevotella intermedia*, and *Prevotella nigrescens* in extraoral and some odontogenic infections. Clin. Infect. Dis. 25(Suppl. 2) (1997) S194-S198.

[9] H.K. Kuramitsu, W. Chen, A. Ikegami, Biofilm formation by the periodontopathic bacteria *Treponema denticola* and *Porphyromonas gingivalis*. J. Periodontol. 76 (2005) 2047-2051.

[10] K. Nelson, R. Fleischmann, R. DeBoy, I. Paulsen, D. Foots, J. Eiken, S. Daugherty, R. Dodson, A. Durkin, M. Gwinn, D. Haft, J. Kolonay, W. Nelson, O. White, T. Mason, L. Tallon, J. Gray, D. Granger, H. Tettelin, H. Dong, J. Galvin, M. Duncan, F. Dewhirst, C. Fraser, C. Complete genome sequence of the oral pathogenic bacterium *Porphyromonas gingivalis* strain W83. J. Bacteriol. 185 (2003) 5591-5601.

[11] M. Naito, H. Hirakawa, A. Yamashita, N. Ohara, M. Shoji, H. Yukitake, K. Nakayama, H. Toh, F. Yoshimura, S. Kuhara, M. Hattori, T. Hayashi, K. Nakayama, Determination of the genome sequence of *Porphyromonas gingivalis* strain ATCC 33277 and genomic comparison with strain W83 revealed extensive genome rearrangements in *P. gingivalis.* DNA Res. 15 (2008) 215-225.

[12] N. Takahashi, T. Sato, T. Yamada, Metabolic pathways for cytotoxic end product formation from glutamate- and aspartate-containing peptides by *Porphyromonas gingivalis.* J. Bacteriol. 182 (2000) 4704-4710.

[13] N. Takahashi, T. Sato, Preferential utilization of dipeptides by *Porphyromonas gingivalis.* J. Dent. Res. 80 (2001) 1425-1429.

[14] A. Banbula, M. Bugno, J. Goldstein, J. Yen, D. Nelson, J. Travis, J. Potempa, Emerging family of proline-specific peptidases of *Porphyromonas gingivalis*: purification and characterization of serine dipeptidyl peptidase, a structural and functional homologue of mammalian prolyl dipeptidyl peptidase IV. Infect. Immun. 68 (2000) 1176-1182.

[15] A. Banbula, J. Yen, A. Oleksy, P. Mak, M. Bugno, J. Travis, J. Potempa, *Porphyromonas gingivalis* DPP-7 represents a novel type of dipeptidylpeptidase. J. Biol. Chem. 276 (2001) 6299-6305.

[16] Y. Ohara-Nemoto, Y. Shimoyama, S. Kimura, A. Kon, H. Haraga, T. Ono, T.K. Nemoto, Asp- and Glu-specific novel dipeptidyl peptidase 11 of *Porphyromonas gingivalis* ensures utilization of proteinaceous energy sources. J. Biol. Chem. 286 (2011) 38115-38127.

[17] V. Mazumdar, E.S. Snitkin, S. Amar, S. Segre, Metabolic network model of a human oral pathogen. J. Bacteriol. 191 (2009) 74-90.

[18] S. Ellis, J.M. Nuenke, Dipeptidyl arylamidase III of the pituitary. Purification and characterization. J. Biol. Chem. 242, (1967) 4623-4629.

[19] C.M. Lee, S.H. Snyder, Selective affinity for enkephalin and angiotensin. J. Biol. Chem. 257 (1982) 12043-12050.

[20] Y. Shimamori, Y. Watanabe, Y. Fujimoto, Purification and characterization of dipeptidyl aminopeptidase III from human placenta. Chem. Pharm. Bull. (Tokyo) 34, (1986) 3333-3340.[21] A. Banbula, P. Mak, M. Bugno, J. Silberring, A. Dubin, D. Nelson, J.

[21] Travis, J. Potempa, Prolyl tripeptidyl peptidase from *Porphyromonas gingivalis*. A novel enzyme with possible pathological implications for the development of periodontitis. J. Biol. Chem. 274 (1999) 9246-9252.

[22] K. Ito, Y. Nakajima, Y. Xu, N. Yamada, Y. Onohara, T. Ito, F. Matsubara, T. Kabashima, K. Nakayama, T. Yoshimoto, Crystal structure and mechanism of tripeptidyl activity of prolyl tripeptidyl aminopeptidase from *Porphyromonas gingivalis*. J. Mol. Biol. 362, (2006) 228-240.

[23] N.D. Rawlings, A.J. Barrett, A. Bateman, MEROPS: the peptidase database. Nucleic Acids Res. 38 (2010) D227-D233.

[24] P.A.R. Vandamme, Taxonomy and classification of bacteria, in: P.R. Murray, E.J. Baron, J.H. Jorgensen, M.L. Landry, M.A. Pfaller (Eds.), Manual of Clinical Microbiology, 9th ed., Vol. 1, American Society for Microbiology Press, Washington, D.C., 2007, pp. 275-290.

[25] M. Nolan, J. Sikorski, K. Davenport, S. Lucas, T.G. Del Rio, H. Tice, J.F. Cheng, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y.J. Chang, C.D. Jeffries, R. Tapia, T. Brettin, J.C. Detter, C. Han, M. Yasawong, M. Rohde, B.J. Tindall, M. Göker, T. Woyke, J. Bristow, J.A. Eisen, V. Markowitz, P. Hugenholtz, N.C. Kyrpides, H.P. Klenk, A. Lapidus, Complete genome sequence of *Ferrimonas balearica* type strain (PAT). Stand. Genomic. Sci. 3 (2010) 174-182.

[26] B.A. Methé, K.E. Nelson, J.W. Deming, B. Momen, E. Melamud, X. Zhang, J. Moult, R. Madupu, W.C. Nelson, R.J. Dodson, L.M. Brinkac, S.C. Daugherty, A.S. Durkin, R.T. DeBoy, J.F. Kolonay, S.A. Sullivan, L. Zhou, T.M. Davidsen, M. Wu, A.L. Huston, M. Lewis, B. Weaver, J.F. Weidman, H. Khouri, T.R. Utterback, T.V. Feldblyum, C.M. Fraser, The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 10913-10918.

[27] J. Boch, U. Bonas, *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. Ann. Rev. Phytopathol. 48, (2010) 419-436.

[28] Y. Kumagai, K. Konishi, T. Gomi, H. Yagishita, A. Yajima, M. Yoshikawa, Enzymatic properties of dipeptidyl aminopeptidase IV produced by the periodontal pathogen

Porphyromonas gingivalis and its participation in virulence. Infect Immun. 68, (2000) 716-724.

[29] T.K. Nemoto, Y. Ohara-Nemoto, T. Ono, T. Kobayakawa, Y. Shimoyama, S. Kimura, T. Takagi, Characterization of the glutamyl endopeptidase from *Staphylococcus aureus* expressed in *Escherichia coli.* FEBS J. 275, (2008) 573-587.

[30] Y. Ikeda, Y. Ohara-Nemoto, S. Kimura, K. Ishibashi, K. Kikuchi, PCR-based identification of *Staphylococcus epidermidis* targeting *gseA* encoding the glutamic-acid-specific protease. Can. J. Microbiol. 50 (2004) 493-498.

[31] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, (1997) 4876–4882.

[32] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4 (1997) 406–425.

[33] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, (2011) MEGA5: molecular evolutionary genetics analysis using maximum likehood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28 (2011) 2731-2739.

[34] J.K. Fredrickson, J.M. Zachara, D.W. Kennedy, H. Dong, T.C. Onstott, N.W. Hinman, S.-M. Li, Biogenic iron mineralization accompanying the dissimilatory reduction of hydrous ferric oxide by a groundwater bacterium - Sorption of Mn^{2+} on $FeCO_3$. Geochim. Cosmochim. Acta 62 (1998) 3239-3257.

[35] P.C. Schreckenberger, M.I. Daneshvar, D.G. Hollis, *Acinetobacter*, *Achromobacter*, *Chryseobacterium*, *Moraxella*, and other nonfermentative gram-negative rods. in: P.R. Murray, E.J. Baron, J.H. Jorgensen, M.L. Landry, and M.A. Pfaller (Eds.), Manual of Clinical Microbiology, 9th ed., Vol. 1, American Society for Microbiology Press, Washington, D.C., 2007, pp. 770-802.

[36] C. Carmona, G.L. Gray, Nucleotide sequence of the serine protease gene of *Staphylococcus aureus*, strain V8. Nucleic Acids Res. 15 (1987) 6757.

[37] H. Fehlhammer, W. Bode, R. Huber, Crystal structure of bovine trypsinogen at 1.8Å resolution. J. Mol. Biol. 111 (1977) 415-438.

[38] W. Bode, P. Schwager, R. Huber, The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-

pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9Å resolution. J. Mol. Biol. 118 (1978) 99-112.

[39] T. Hoshino, T. Fujiwara, S. Kawabata, Evolution of cariogenic character in *Streptococcus mutans*: horizontal transmission of glycosyl hydrolase family 70 genes. Sci. Rep. (2012) 2, 518; DOI: 10.1038/strep00518 (2011) 1-7.

[40] L. Polgar, The catalytic triad of a serine peptidase. Cell. Mol. Life. Sci. 62 (2005) 2161-2172.

FIGURE 1

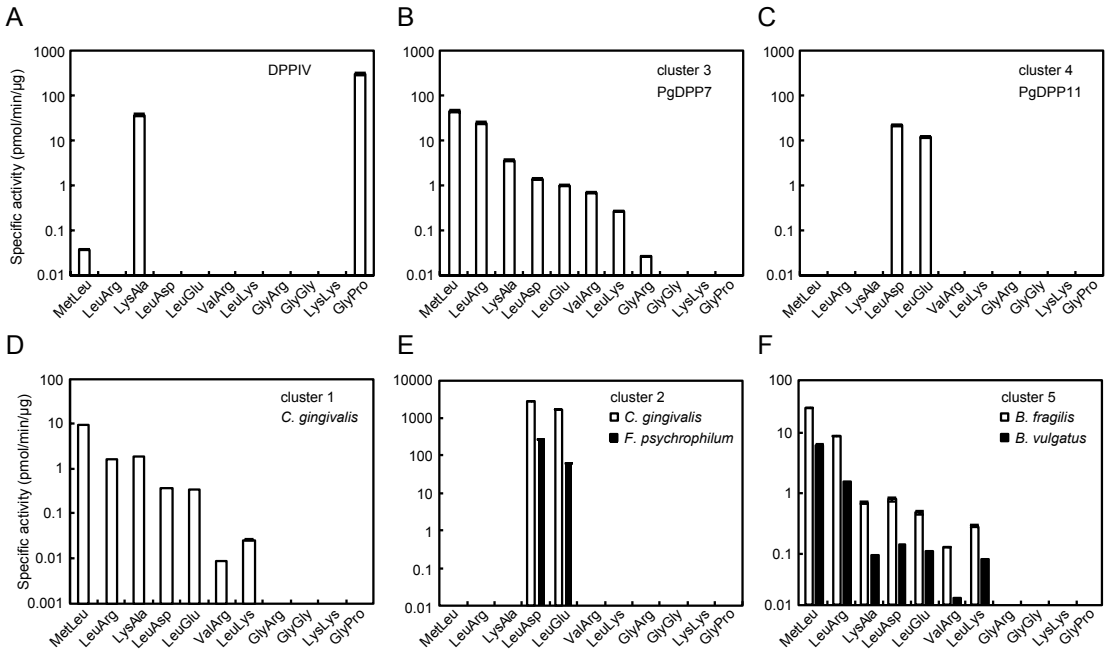


Fig. 1. Substrate specificity of DPPIV and DPP7/11 belonging to the five clusters of the S46 family. The hydrolyzing activities of *P. gingivalis* (A) DPPIV, (B) DPP7 in cluster 3, and (C) DPP11 in cluster 4 toward dipeptidyl-MCA substrates were determined. (D) Cluster-1 DPP from *C. gingivalis*, (E) cluster-2 DPPs from *C. gingivalis* and *F. psychrophilum*, and (F) cluster-5 DPPs from *B. fragilis* and *B. vulgatus*. Values are shown as the mean ± S.D. (n=3).

FIGURE 2-1

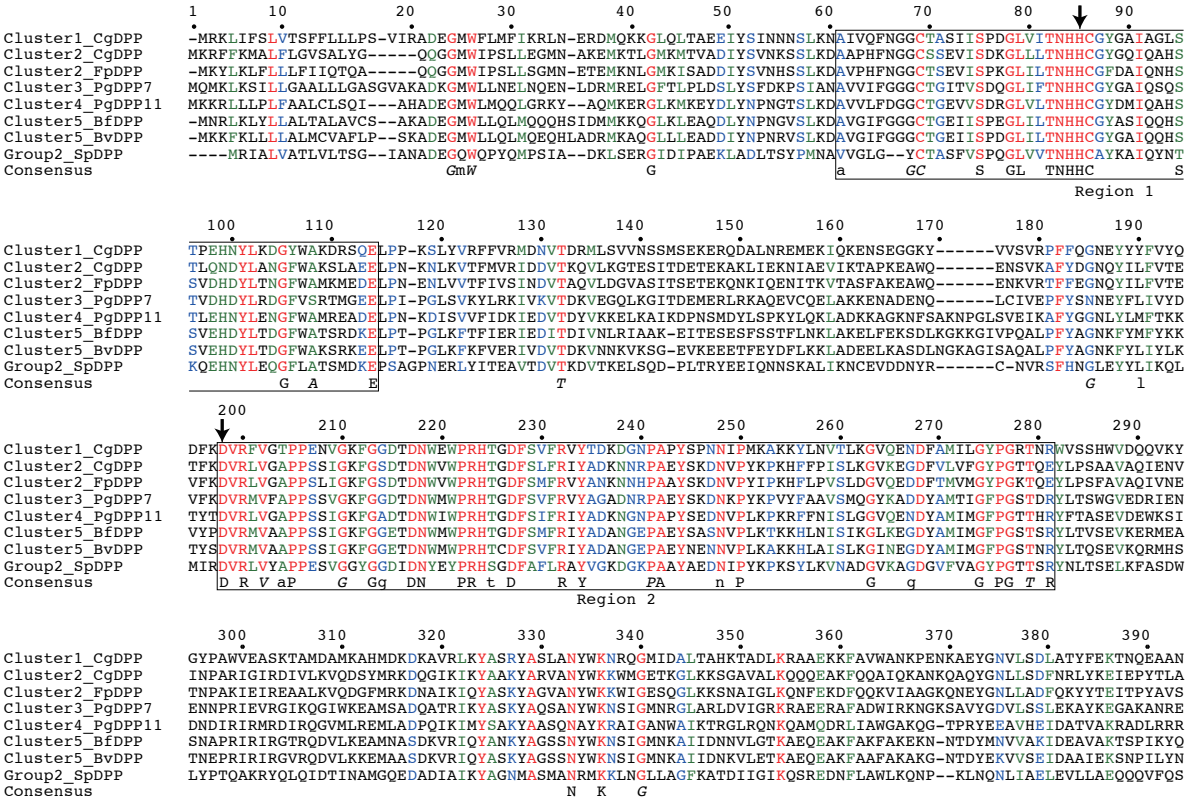


FIGURE 2-2

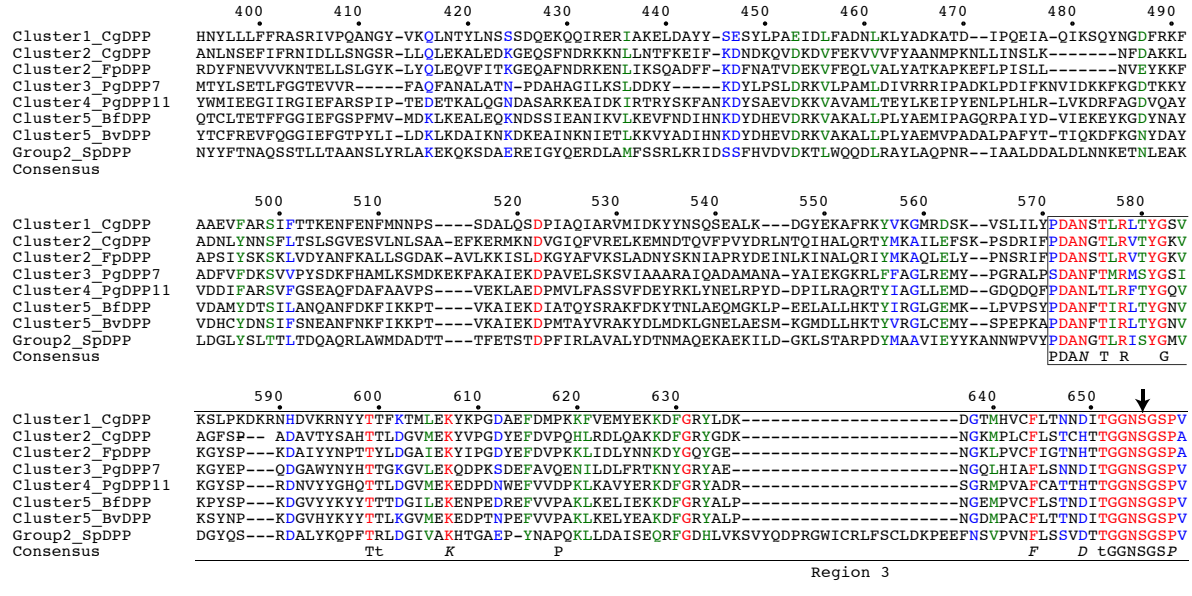


Fig. 2. Alignments of amino acid sequences of expressed DPPs and conserved amino acid residues in S46 family members. Amino acid sequences of recombinant DPPs expressed in this study were shown. Amino acid number is represented as that of PgDPP11. Conserved residues, highly similar, and weakly similar residues among 8 DPPs are shown in red, green, and blue, respectively. As consensus residues, identical amino acids in 264 members of the S46 family in more than 95%, 80%, and 70% are indicated by capital, capital-italic, and small letters, respectively. Conserved regions 1 (Ala^{61}-Glu^{114}), 2 (Asp^{198}-Arg^{281}), and 3 (Pro^{571}/Asp^{571}-Leu^{700}) are boxed. Arrowheads indicate essential amino acids forming the catalytic triad of the serine protease family. Entirely conserved Gly^{673} in clusters-1, -3, and -5 DPP7 members, or Arg^{673} in cluster-2 and -4 DPP11 members, but was converted to Ser^{673} in Group-2 DPP are boxed. An asterisk indicates Gly/Ser^{680}, which is substituted by Lys^{680} in *S. putrefaciens* DPP (see text for details). Cluster-1 CgDPP (*C. gingivalis*, MEROPS ID: MER217135), Cluster-2 CgDPP (*C. gingivalis*, MER217541), Cluster-2 FpDPP (*F. psychrophilum*, MER109496), Cluster-3 PgDPP7 (*P. gingivalis*, MER14366), Cluster-4 PgDPP11 (*P. gingivalis*, MER34628), Cluster-5 BfDPP (*B. fragilis*, MER039993), Cluster-5 BvDPP (*B. vulgatus*, MER61211), Group-2 SpDPP (*S. putrefaciens*, MER096825).

FIGURE 3

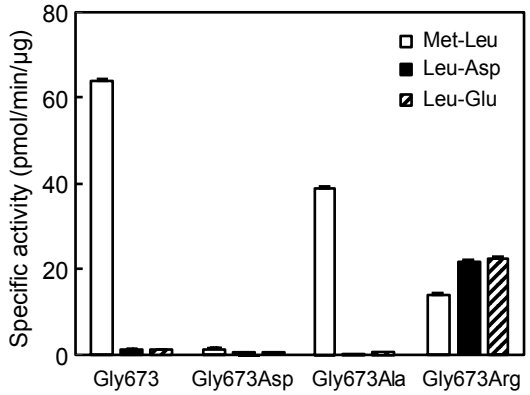


Fig. 3. Role of Gly^{673} in enzyme activity of DPP7. Hydrolyzing activities of PgDPP7 wild-type Gly^{673}, and single amino acid substitution mutants Gly673Asp, Gly673Ala, and Gly673Arg were determined with Met-Leu-, Leu-Asp , and Leu-Glu-MCA. Values are shown as the mean ± S.D. (n=3).

FIGURE 4

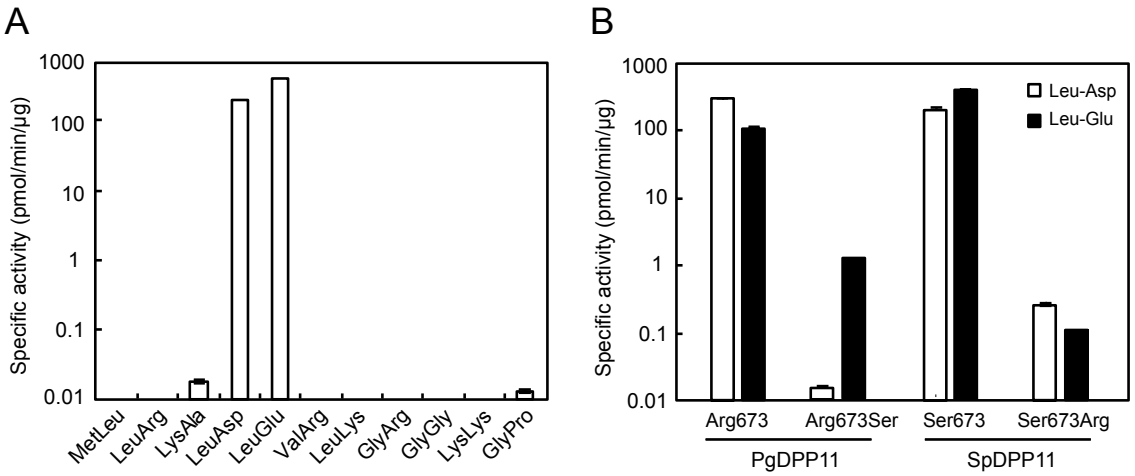


Fig. 4. Alterations of Asp/Glu preference by substitution of Arg673/Ser673 of PgDPP11 and SpDPP11 to Ser^{673}/Arg^{673} respectively. (A) The substrate specificity of *S. putrefaciens* DPP (MER096825) was determined with dipeptidyl MCA substrates. (B) The hydrolyzing activities of PgDPP11 wild-type Arg673, substitution mutant Arg673Ser, *S. putrefaciens* DPP11 wild-type Ser^{673}, and substitution mutant Ser673Arg were determined. Values are shown as the mean ± S.D. (n=3).

FIGURE 5

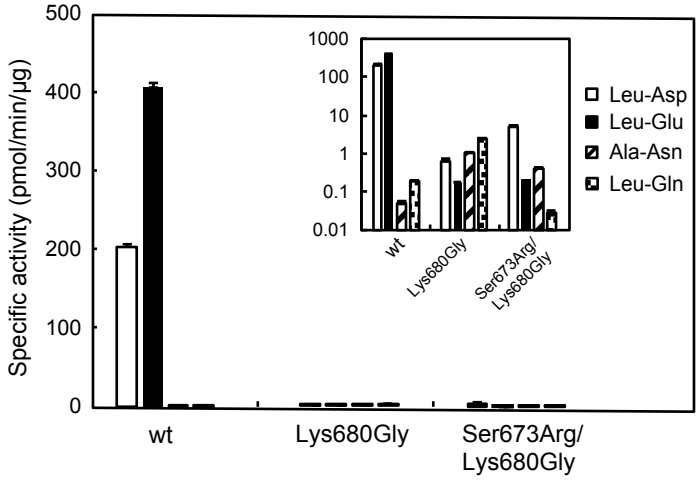


Fig. 5. Effect of Lys680Gly substitution on hydrolyzing activities of *S. putrefaciens* Ser^{673}-DPP11. The hydrolyzing activity of *S. putrefaciens* DPP11 wild-type (wt), substitution mutants Lys680Gly and Ser673Arg/Lys680Gly were determined with four dipeptidyl MCA substrates. Inserts are represented as logarithmic scales. Values are shown as the mean ± S.D. (n=3).

FIGURE 6

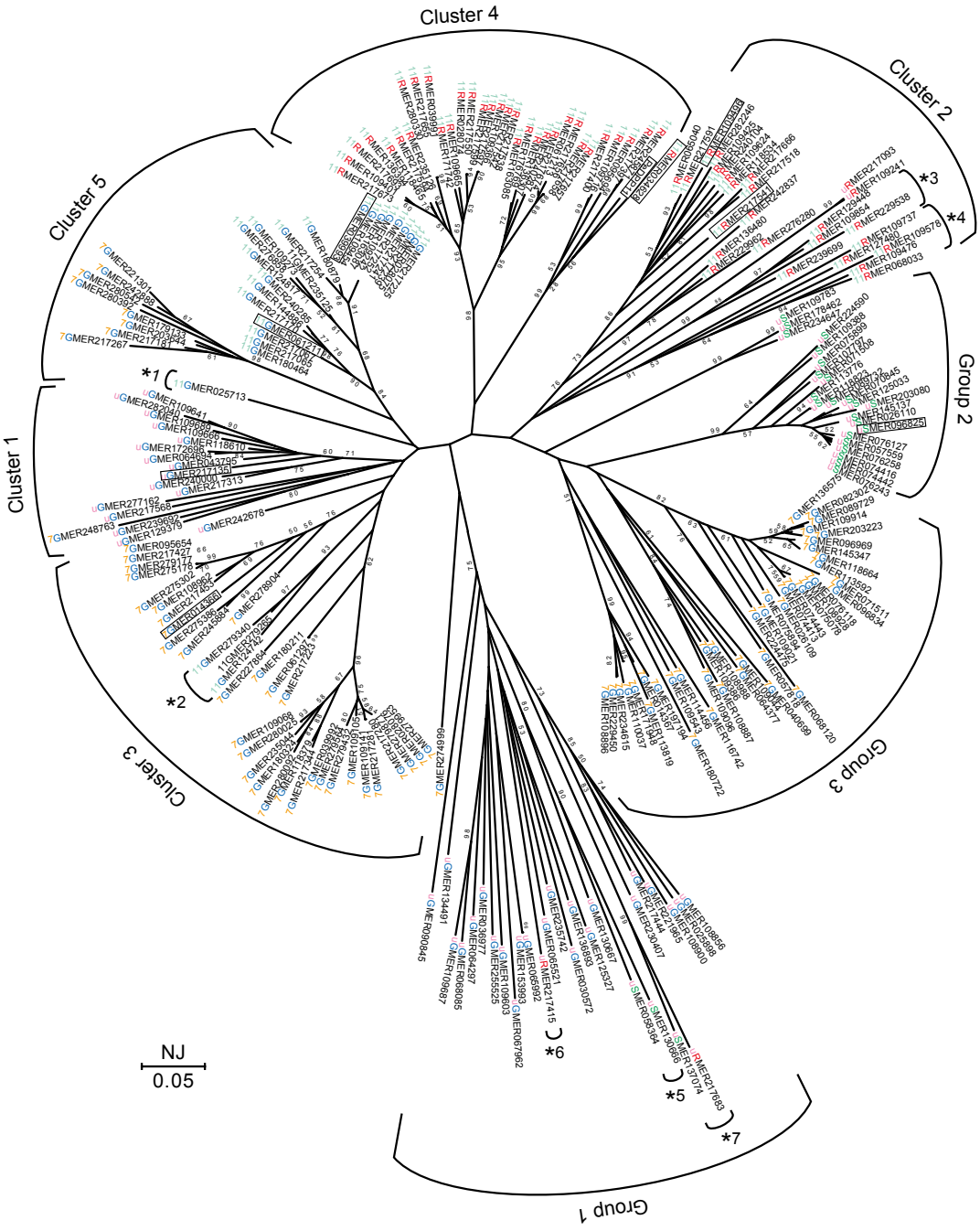


Fig. 6. Phylogenic tree of S46 family based on sequences of conserved region 3. A phylogenic tree was constructed using an NJ method [32] with ClustalX software [31] based on conserved region 3 (Pro/Ser^{571}-Leu^{700}). Basic information for DPPs is included in the names, as follows: first 7 (DPP7), 11 (DPP11), or u (unassigned) in the MEROPS database (Release 9.6), followed by G (Gly^{673}), R (Arg^{673}), or S (Ser^{673}), then finally the MEROPS codes. For example, PgDPP7 and PgDPP11 were written as 7GMER014366 and 11RMER034628, respectively. As an exception, PeDPP11 is simply expressed the same as that cloned in the laboratory [16] and is not yet registered in the database. DPPs expressed in this study are boxed. The phylum *Bacteroidetes* consisted of five clusters and others mainly comprising the phylum *Proteobacteria* consisted of three groups. See text and supplemental Fig. S1 for members marked by asterisks. Basic information and the names of the DPPs are described in supplemental Table S2.

SUPPLEMENTAL DATA

Supplemental Table S1

Primers used for expressions of DPPIV, DPP7, and DPP11 from *P. gingivalis*, and DPPs from clusters 1, 2, and 5 of the phylum *Bacteroidetes*. Restriction sites are indicated by italics.

Primer	DPP	Sequence (5'-3')
5PgDPPIVD23Bam	PgDPPIV	GGCTCAG*GGATCC*GACAAACCCGTAGATTT
3PgDPPIVL723Bam	PgDPPIV	AGAGATACT*GGATCC*AAGATTGTCGAACAA
5PgDPP7Q2Bgl	PgDPP7	GACTGCT*AGATCT*CAAATGAAATTAAAAAGTATTC
3PgDPP7I712Bgl	PgDPP7	GGCGAAT*AGATCT*GATCAACTTCAGCTC
5PgDPP7Q2	PgDPP7 deletion	CAAATGAAATTAAAAAGTATTCTTCTCG
pQE60-M1Rev	PgDPP7 deletion	CATGGTTAATTTCTCCTCTTTAATGA
5PgDPP11K2Bam	PgDPP11	GTA*GGATCC*AAAAAAGACTATTGCTCCCCCTC
3PgDPP11P720Bam	PgDPP11	ATA*GGATTC*GGGAACGATATTCATTTCATCCAAC
5CgDPPG1R2Bam	*C. gingivalis* cluster-1	TAATTG*GGATCC*AGAAAACTTATTTTTTC
3CgDPPG1Q715Bam	*C. gingivalis* cluster-1	AATAG*GGATCC*CTGGATAATAGTCATTTC
5FpDPPC1D22	*F. psychrophilum* cluster-1	GATGAAGGAATGTGGTTTTTAATGT
3FpDPPC1K716	*F. psychrophilum* cluster-1	CTATTATTTAACTAAAGTCATTTCA
5CgDPPC2L16	*C. gingivalis* cluster-2	CTCTATGGACAGCAAGGCGGGAT
3CgDPPC2K712	*C. gingivalis* cluster-2	CTATTACTTGATGAGTTTCATTT
5FpDPPC2Q18	*F. psychrophilum* cluster-2	CAACAAGGGGGCATGTGGATTCC
3FpDPPC2K713	*F. psychrophilum* cluster-2	CTATTATTTTTTCTTAGGGTGAA
5BfDPPC5N2Bam	*B. fragilis* cluster-5	ATATA*GGATCC*AACAGACTCAAACTTTAC
3BfDPPC5E721Bam	*B. fragilis* cluster-5	TCATA*GGATCC*TTCAACAATCGTCATTTC
5Bv DPPC5K2Bgl	*B. vulgatus* cluster-5	TTATG*AGATCT*AAAAAATTCAAACTTTTATTG

3BvDPPC5E721Bgl	*B. vulgatus* cluster-5	TACAT*AGATCT*TTCAACAATAGTCATTTC
5SpDPP SR2Bam	*S. putrefaciens*	TAATCA*GGATCC*CGTATTGCACTGGTTGCGAC
3SpDPP SN732Bam	*S. putrefaciens*	ACTATA*GGATCC*ATTTCTCACTAAATCCAGTTC

Supplemental Table S2

Classification of DPP7 and DPP11 based on amino acid residue at position 673. DPPs in the S46 family are classified as DPP7 carrying Gly^{673}, and DPP11 carrying either Asp^{673} or Ser^{673}. DPPs are represented by the MEROPS codes and numbers 1-263 are identical to those of the MEROPS database (Release 9.6). Annotation in the database is indicated by italics (unassigned), bold (DPP11), and underlining (DPP7). MEROPS codes in grey backgrounds are DPPs with substrate specificity directly identified in the present and previous [refs. 14-16] studies. [a]Nos. 61 and 62 were 93.44% identical, presumably due to the different strains of a single organism. [b]Nos. 98 and 231 were 25.81% identical. [c]Composed of 1200 residues, of which residues 1-626 were homologous to the S46 family members.

	DPP7	DPP11	
Species and strains	Gly^{673}	Arg^{673}	Ser^{673}
Shewanella sediminis	*1 (MER109914)*		*36 (MER070845)*
Shewanella benthica	*2 (MER136575)*		*37 (MER125033)*
Shewanella woodyi	*3 (MER082302)*		*35 (MER076258)*
Shewanella violacea	*4 (MER203223)*		*38 (MER203080)*
Shewanella loihica	*5 (MER089729)*		*34 (MER089732)*
Shewanella putrefaciens	*6 (MER096834)*		*44 (MER096825)*
Shewanella sp. W3-18-1	*7 (MER076118)*		*45 (MER076127)*
Shewanella baltica	*8 (MER108928)*		*46 (MER057559)*
Shewanella sp. ANA-3	*9 (MER075078)*		*41 (MER076243)*
Shewanella sp. MR-7	*10 (MER074443)*		*40 (MER074442)*
Shewanella sp. MR-4	*11 (MER074413)*		*39 (MER074416)*
Shewanella oneidensis	*12 (MER026109)*		*42 (MER026110)*
Shewanella halifaxensis	*13 (MER118664)*		*32 (MER118823)*
Shewanella pealeana	*14 (MER113592)*		*31 (MER113776)*
Shewanella piezotolerans	*15 (MER145347)*		*33 (MER145137)*
Shewanella denitrificans	*16 (MER071511)*		*47 (MER071508)*
Shewanella frigidimarina	*17 (MER096969)*		*49 (MER107797)*
Shewanella amazonensis	*18 (MER075894)*		*48 (MER075899)*
Shewanella sp. HN-41			*43 (MER276432)*
Ferrimonas balearica	*19 (MER224475)*		*50 (MER224590)*
Pseudoalteromonas tunicata	*20 (MER109021)*		*52 (MER109388)*
Kangiella koreensis	*21 (MER068120)*		
Rheinheimera sp. A13L	*22 (MER276692)*		*51 (MER276752)*
Colwellia psychrerythraea	*23 (MER057818)*		
Glaciecola nitratireducens	*24 (MER281964)*		
gamma proteobacterium			*25 (MER276352)*

IMCC3088			
Erythrobacter sp. NAP1	*26 (MER109386)*		
Erythrobacter sp. SD-21	*27 (MER108887)*		
Parvularcula bermudensis	*28 (MER108988)*		
Alteromonas macleodii	*29 (MER109096)*		
Idiomarina loihiensis	*30 (MER040699)*		
Burkholderia sp. JV3	*53 (MER272800)*		
Stenotrophomonas sp. SKA14	*54 (MER171948)*		*64 (MER178462)*
Stenotrophomonas maltophilia	*55 (MER110037)*		*65 (MER109783)*
Pseudoxanthomonas suwonensis	*56 (MER234615)*		*66 (MER234647)*
Xanthomonas albilineans	*57 (MER197194)*		
Xanthomonas oryzae	*58 (MER108896)*		
Xanthomonas alfalfae	*59 (MER274065)*		
Xanthomonas fuscans	*60 (MER229450)*		
Xylella fastidiosa	*61 (MER014367)*[a] *62 (MER113819)*[a]		
Burkholderia oklahomensis	*63 (MER180722)*		
Haliangium ochraceum	*67 (MER116742)*	*72 (MER127480)*	
Sorangium cellulosum	*68 (MER114456)*		
Plesiocystis pacifica	*69 (MER109543)*	*73 (MER109737)*	
Solibacter usitatus	*70 (MER109044)*	*76 (MER109578)*	
Anaeromyxobacter sp. Fw109-5	*71 (MER064377)*		
Myxococcus xanthus	*249 (MER068085)*	*74 (MER068033)*	
Stigmatella aurantiaca	*250 (MER109687)*	*75 (MER109476)*	
Candidatus Chloracidobacterium thermophilum	*77 (MER274999)*		
Pedobacter sp. BAL39		*78 (MER109854)*	
Pedobacter heparinus	*213 (MER129379)*	*79 (MER129448)*	
Sphingobacterium spiritivorum		*80 (MER229538)*	
Pedobacter saltans	*212* (*MER239692*)	*81 (MER239699)*	
Aureococcus anophagefferens		*82 (MER243787)*	
Alistipes shahii	**83 (MER279340)**	**162 (MER229962)**	
Alistipes sp. HGB5	**84 (MER279265)**		
Alistipes putredinis	**85 (MER124742)**	**163 (MER136480)**	
Cytophaga hutchinsonii	**86 (MER025713)**		
Bacteroides coprophilus	**135 (MER217170)**	**87 (MER176769)**	
Bacteroides plebeius	**134 (MER136226)**	**88 (MER163085)**	
Bacteroides salanitronis	**132 (MER240285)**	**89 (MER240287)**	

Bacteroides coprocola	**133 (MER144886)**	**90 (MER136307)**	
Bacteroides sp. 9_1_42	**137 (MER217067)** 184 (MER217223)	**91 (MER217658)**	
Bacteroides dorei	**139 (MER180464)** 185 (MER180211)	**92 (MER217657)**	
Bacteroides sp. 3_1_33		**93 (MER217656)**	
Bacteroides vulgatus	**136 (MER061211)** 186 (MER061297)	**94 (MER061213)**	
Bacteroides sp. 4_3_47	**138 (MER217065)**	**95 (MER217674)**	
Bacteroides sp. D2	**126 (MER217252)**	**96 (MER217599)**	
Bacteroides sp. 2_2_4	165 (MER217280)	**97 (MER217597)**	
Bacteroides ovatus	**125 (MER109242)** 166 (MER109141)	**98 (MER109366)**[b] *231 (MER109241)*[b]	
Bacteroides sp. D1	**128 (MER217239)** 167 (MER217281)	**99 (MER217538)**	
Bacteroides sp. 2_1_22	**127 (MER217241)**	**100 (MER217542)**	
Bacteroides finegoldii	**124 (MER163087)**	**101 (MER171742)**	
Bacteroides caccae	**123 (MER109454)** 168 (MER109105)	**102 (MER109665)**	
Bacteroides sp. 1_1_6	**129 (MER217225)** 171 (MER279953)	**103 (MER217550)**	
Bacteroides thetaiotaomicron	**130 (MER028075)** 170 (MER028076)	**104 (MER028074)**	
Bacteroides sp. 2_1_56FAA	175 (MER279492)	**105 (MER280467)**	
Bacteroides sp. 3_2_5	174 (MER279564)	**106 (MER217655)**	
Bacteroides fragilis	**131 (MER039993)** 173 (MER039992)	**107 (MER039991)**	
Bacteroides sp. 2_1_16	172 (MER279432)	**108 (MER280330)**	
Bacteroides stercoris	**120 (MER124817)** 180 (MER118379)	**109 (MER118465)**	
Bacteroides eggerthii	**119 (MER276858)** 181 (MER217344)	**110 (MER217594)**	
Bacteroides uniformis	**117 (MER109273)** 177 (MER109068)	**111 (MER109401)**	
Bacteroides sp. D20	178 (MER280025)	**112 (MER217693)**	
Bacteroides cellulosilyticus	**122 (MER217254)** 183 (MER280092)	**113 (MER217673)**	
Bacteroides intestinalis	**121 (MER180879)**	**114 (MER136394)**	

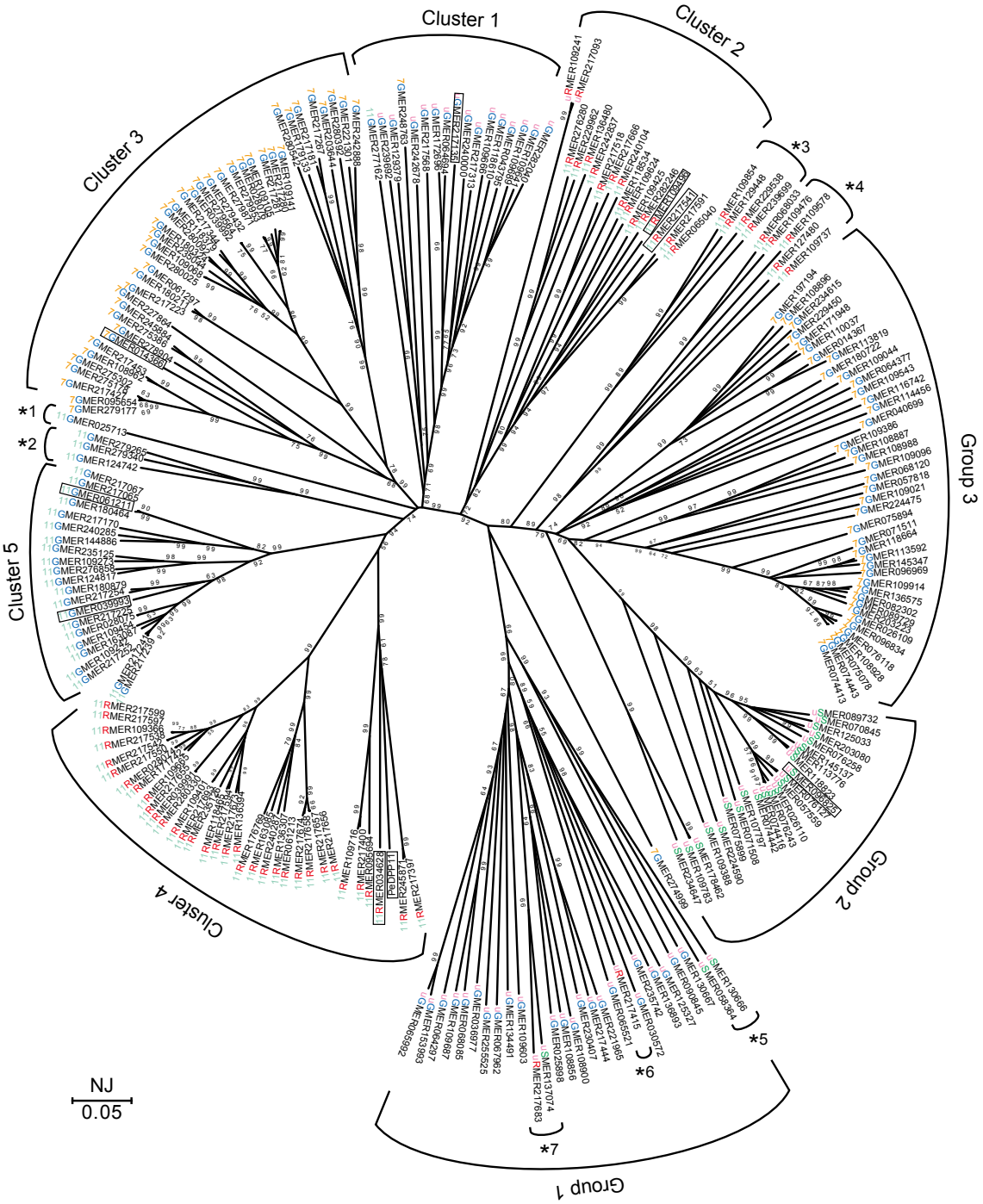
	182 (MER180324)		
Bacteroides helcogenes	**116 (MER235125)** 179 (MER235044)	**115 (MER235126)**	
Bacteroides clarus	**118 (MER277010)**		
Bacteroides sp. 1_1_14	169 (MER279871)		
Bacteroides sp. 4_1_36	176 (MER279717)		
Bacteroides sp. 20_3	194 (MER275178)		
Bacteroides sp. 3_1_19	195 (MER275302)		
Bacteroides sp. 2_1_33B	198 (MER217427)		
Parabacteroides sp. D13	196 (MER279177)		
Bacteroides sp. D2		*232 (MER217093)*	
Parabacteroides merdae	192 (MER108962)	**140 (MER109716)**	
Parabacteroides johnsonii	193 (MER217453)	**141 (MER217400)**	
Parabacteroides distasonis	197 (MER095654)	**142 (MER095694)**	
Porphyromonas gingivalis	189 (MER014366)	**143 (MER034628)**	
Porphyromonas asaccharolytica	191 (MER245884)	**144 (MER245871)**	
Porphyromonas uenonis	190 (MER275386)	**145 (MER217397)**	
Odoribacter splanchnicus	**146 (MER275825)**		
Bacteroidetes oral taxon 274	**147 (MER277162)**		
Flavobacteria bacterium BAL38	*218 (MER109689)*	**148 (MER109425)**	
Flavobacterium columnare	*219 (MER282040)*	**149 (MER282246)**	
Flavobacterium psychrophilum	*217 (MER109641)*	**150 (MER109496)**	
Capnocytophaga gingivalis	*229 (MER217135)*	**151 (MER217541)**	
Capnocytophaga canimorsus	*228 (MER255367)*	**152 (MER255206)**	
Capnocytophaga sputigena	*226 (MER172696)*	**153 (MER217591)**	
Capnocytophaga ochracea	*227 (MER064694)*	**154 (MER065040)**	
Flavobacteriales bacterium ALC-1		**155 (MER118534)**	
Lacinutrix sp. 5H-3-7-4		**156 (MER251262)**	
unidentified eubacterium SCB49		**157 (MER109624)**	
Weeksella virosa		**158 (MER240104)**	
Flavobacteriaceae bacterium 3519-10	*230 (MER217568)*	**159 (MER217518)**	
Chryseobacterium gleum		**160 (MER217666)**	
Alistipes sp. HGB5		**161 (MER276280)**	
Fluviicola taffensis		**164 (MER242837)**	
Paraprevotella xylaniphila	187 (MER277088)		
Porphyromonas endodontalis	*188 (MER278904)*	**264 PeDPP11**	

Dysgonomonas mossii	199 (MER275543)		
Dysgonomonas gadei	200 (MER275470)		
Paludibacter propionicigenes	201 (MER227864)		
Prevotella timonensis	202 (MER280542)		
Prevotella oralis	203 (MER280616)		
Prevotella salivae	204 (MER280687)		
Prevotella copri	205 (MER179133)		
Prevotella sp. oral taxon 472	206 (MER217181)		
Prevotella sp. oral taxon 317	207 (MER217267)		
Prevotella ruminicola	208 (MER203644)		
Prevotella melaninogenica	209 (MER221301)		
Prevotella denticola	210 (MER242888)		
Prevotella bivia	211 (MER280392)		
Runella slithyformis	*214 (MER252347)*		
Fluviicola taffensis	*215 (MER242678)*		
Haliscomenobacter hydrossis	*216 (MER248763)*		
Flavobacteria bacterium BAL38	*218 (MER109689)*		
Flavobacterium columnare	*219 (MER282040)*		
Flavobacterium branchiophilum	*220 (MER274656)*		
Flavobacterium johnsoniae	*221 (MER043795)*		
unidentified eubacterium SCB49	*222 (MER109666)*		
Kordia algicida	*223 (MER118610)*		
Weeksella virosa	*224 (MER240000)*		
Sphingobacterium spiritivorum	*225 (MER217313)*		
Capnocytophaga ochracea	*227 (MER064694)*		
Rhodothermus marinus	*235 (MER130667)*		*233 (MER130666)*
Salinibacter ruber			*234 (MER058364)*
Spirosoma linguale	*236 (MER090845)*		
Opitutus terrae	*237 (MER125327)*		
bacterium Ellin514	*238 (MER136893)*		
Isosphaera pallida	*239 (MER235742)*		
Rhodopirellula baltica	*240 (MER030572)*		
Gemmata obscuriglobus		*241 (MER217415)*	
Acidobacteria bacterium	*242 (MER065521)*		
Anaeromyxobacter dehalogenans	*243 (MER153993)*		
Anaeromyxobacter sp. K	*244 (MER125498)*[c]		
Anaeromyxobacter dehalogenans	*245 (MER065992)*		

Anaeromyxobacter sp. Fw109-5	*246 (MER064297)*		
Gloeobacter violaceus	247 (MER036977)		
Myxococcus fulvus	*248 (MER251918)*		
Collimonas fungivorans	*251 (MER255525)*		
Myxococcus fulvus	*252 (MER251542)*		
Myxococcus xanthus	*253 (MER067962)*		
Rhodospirillum centenum	*254 (MER134491)*		
uncultured bacterium	*255 (MER109603)*		
Brevundimonas subvibrioides	*256 (MER221965)*		*263 (MER217683)*
Brevundimonas sp. BAL3	*257 (MER217444)*		*262 (MER137074)*
Asticcacaulis excentricus	*258 (MER230407)*		
Sphingomonas sp. SKA58	*259 (MER108856)*		
Novosphingobium aromaticivorans	*260 (MER025898)*		
Sphingopyxis alaskensis	*261 (MER108900)*		

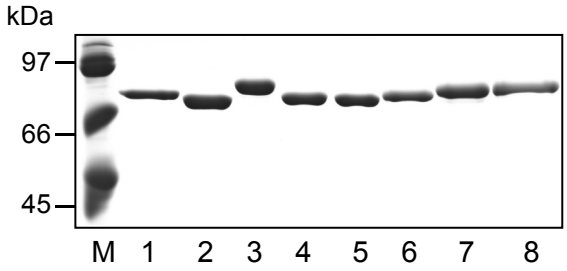
SUPPLEMENTAL FIGURE LEGENDS

Supplemental FIGURE S1



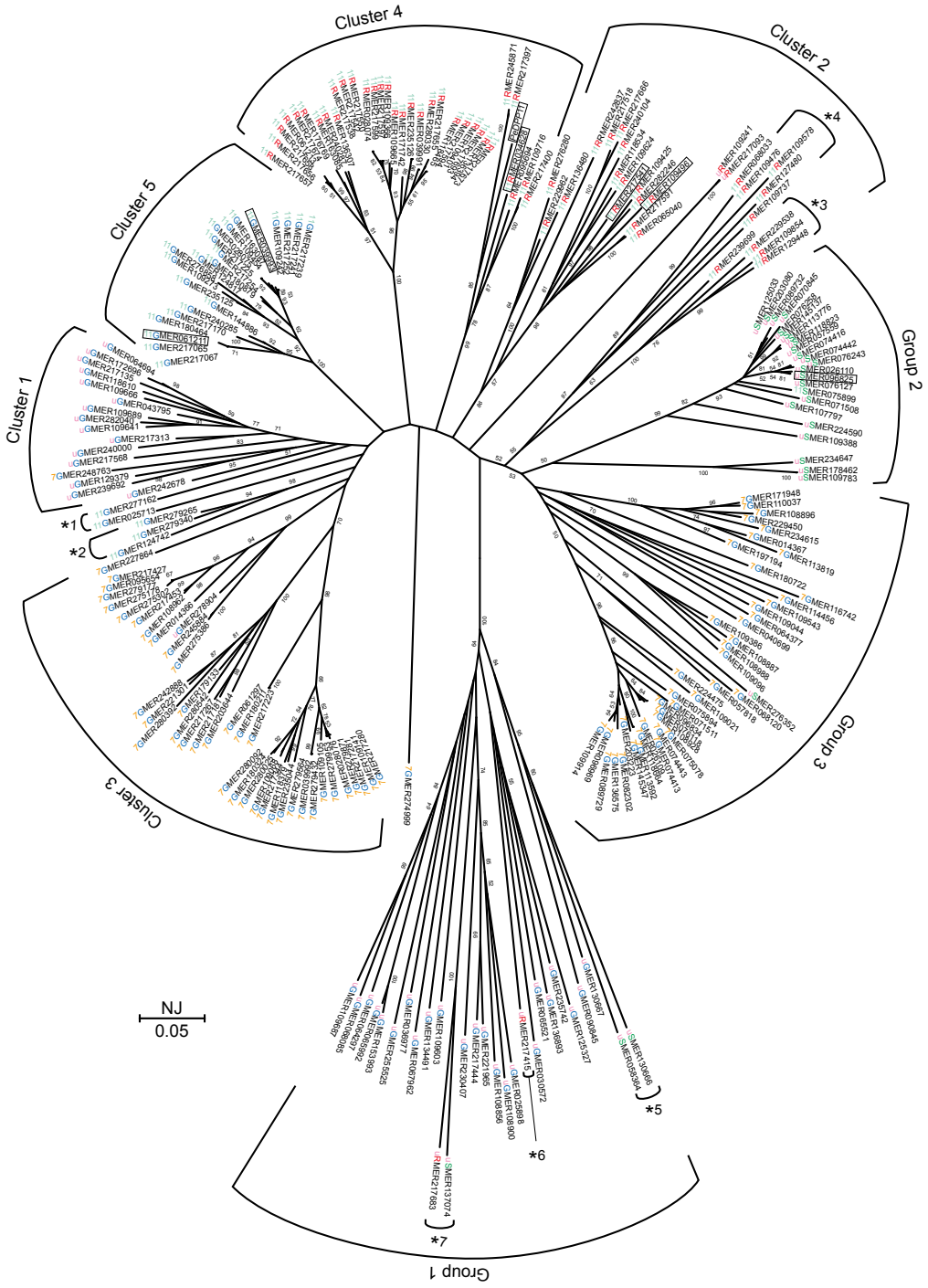
Supplemental Fig. S1. Phylogenic tree of S46 family based on full-length amino acid sequences. The phylogenic tree of the S46 family was constructed using an NJ method by comparisons of the full-length sequences. Basic information for the DPPs is included in the names, as follows: first, 7 (DPP7), 11 (DPP11), or u (unassigned) in the MEROPS database (Release 9.6), followed by G (Gly^{673}), R (Arg^{673}), or S (Ser^{673}), and finally the MEROPS codes. For example, PgDPP7 and PgDPP11 are expressed as 7GMER014366 (DPP7-Gly^{673}-MER014366) and 11RMER034628 (DPP11-Arg^{673}-MER034628), respectively. DPPs expressed in this study are boxed. Members from phylum *Bacteroidetes* were separated into five clusters and others were classified into three groups. Members with asterisks 1-4 were not categorized into any clusters or groups. Members with asterisks 5-7 were categorized into group 1 and possessed Arg^{673} or Ser^{673}.

Supplemental FIGURE S2



Supplemental Fig. S2. SDS-PAGE of recombinant DPPs. *E. coli*-expressed purified samples (0.5 μg) of PgDPPIV (lane 1), PgDPP7 (lane 2), PgDPP11 (lane 3), cluster-1 DPP from *C. gingivalis* (lane 4), cluster-2 DPPs from *C. gingivalis* (lane 5) and *F. psychrophilum* (lane 6), and cluster-5 DPPs from *B. fragilis* (lane 7) and *B. vulgatus* (lane 8) were separated using SDS-PAGE on 10% polyacrylamide gels. Separated proteins were stained with Coomassie-brilliant blue. Lane M, low-molecular-weight marker.

Supplemental FIGURE S3



Supplemental Fig. S3. Phylogenic tree of S46 family based on the three conserved regions. Three conserved regions (Ala^{61}-Glu^{114}, Asp^{198}-Arg^{281}, and Pro^{571}/Asp^{571}-Leu^{700}) were used for calculation. Other notations are same as those in supplemental Fig. S1.