Structural insights into the multispecific recognition of dipeptides of deep-sea Gram-negative bacterium *Pseudoalteromonas* sp. SM9913

Chun-Yang Li¹,², Xiu-Lan Chen¹,²,³, Qi-Long Qin¹,², Peng Wang¹,², Wei-Xin Zhang¹,², Bin-Bin Xie¹,², Hai-Nan Su¹,², Xi-Ying Zhang¹,²,†, Bai-Cheng Zhou² and Yu-Zhong Zhang¹,²,³

¹State Key Laboratory of Microbial Technology, ²Marine Biotechnology Research Center, ³Collaborative Innovation Center of Deep Sea Biology, Shandong University, Jinan 250100, China

†Address correspondence to: Xi-Ying Zhang, State Key Laboratory of Microbial Technology, Marine Biotechnology Research Center, Shandong University, Jinan 250100, P. R. China. Tel: +86-531-88364326, Fax: +86-531-88564326, E-mail: zhangxiying@sdu.edu.cn.

Running title: Dipeptides recognition of marine bacteria

Key words: marine bacteria, dipeptide permease (Dpp), dipeptide transporter, dipeptide recognition, bacterial peptide uptake
Peptide uptake is important for nutrition supply for marine bacteria. It is also an important step in marine nitrogen cycling. However, how marine bacteria absorb peptides is still not fully understood. DppA is the periplasmic dipeptide binding protein of dipeptide permease (Dpp, an important peptide transporter in bacteria), and exclusively controls the substrate specificity of Dpp. Here, the substrate binding specificity of deep-sea *Pseudoalteromonas* sp. SM9913 DppA (*Ps*DppA) was analyzed for 25 different dipeptides with varying properties by using isothermal titration calorimetry measurements. *Ps*DppA showed binding affinities for 8 dipeptides. To explain the multispecific substrate recognition mechanism of *Ps*DppA, we solved the crystal structures of unliganded *Ps*DppA and of *Ps*DppA in complex with 4 different types of dipeptides (Ala-Phe, Met-Leu, Gly-Glu and Val-Thr). *Ps*DppA alternates between an “open” and a “closed” form during substrate binding. Structural analyses of the 4 *Ps*DppA-substrate complexes combined with mutational assays indicate that *Ps*DppA binds to different substrates through a precise mechanism: dipeptides are bound mainly by the interactions between their backbones and *Ps*DppA, in particular by anchoring their N- and C-termini through ion-pair interactions; hydrophobic interactions are important in binding hydrophobic dipeptides; Lys457 is necessary for the binding of dipeptides with a C-terminal glutamic acid or glutamine. Additionally, sequence alignment suggests that the substrate recognition mechanism of *Ps*DppA may be common in Gram-negative bacteria. Altogether, our results provide structural insights into the multispecific substrate recognition mechanism of marine Gram-negative bacterial DppA, which provides a better understanding of the mechanisms of marine bacterial peptide uptake.
Peptide uptake plays a significant role in nutrition supply for marine bacteria. It is also an important step in marine nitrogen cycling. However, how marine bacteria recognize and absorb peptides is still unclear. This study analyzed the substrate binding specificity of deep-sea *pseudoalteromonas* sp. SM9913 DppA (*Ps*DppA, the dipeptide-binding protein of dipeptide permease), and solved the crystal structures of unliganded *Ps*DppA and *Ps*DppA in complex with 4 different types of dipeptides. The multispecific recognition mechanism of *Ps*DppA for dipeptides is explained based on structural and mutational analyses. We also find that the substrate-binding mechanism of *Ps*DppA may be common in Gram-negative bacteria. This study sheds light on marine Gram-negative bacterial peptide uptake and marine nitrogen cycling.
INTRODUCTION

The largest components of marine organic matter are referred to as detrital proteins (1, 2). The utilization of oceanic non-living organic nitrogen is vital for marine nitrogen recycling. A portion of this oceanic organic pool is consumed directly by grazers, and bacterial extracellular enzymes degrade the rest into small peptides and amino acids. Bacteria and/or other microorganisms take in these small peptides and amino acids and use them for their growth and reproduction (1). Bacteria absorb peptides through peptide transporters, which are an important means of bacterial nutrition supply (3, 4).

Three types of peptide transporters have been identified in bacteria: oligopeptide permease (Opp), dipeptide permease (Dpp) and di/tripeptide permease (Dtp). Dtp belongs to the secondary active transporters, and Opp and Dpp belong to the ATP-binding cassette (ABC) superfamilies. Opp transports oligopeptides that can contain up to 35 residues (5). Dpp is primarily responsible for transporting dipeptides into cells and can also transport tripeptides to a lesser extent (4, 6). Dpp consists of five subunits: DppA is a periplasmic peptide binding protein that performs the first step in the recognition and binding of substrates; DppB and DppC are two integral membrane proteins that together provide a translocation pathway; DppD and DppF are two cytoplasmic proteins that bind and hydrolyze ATP to provide the energy needed for peptide transport.

OppA, the periplasmic binding protein of Opp, has been widely studied in both Gram-negative bacteria and Gram-positive bacteria (5, 7-13). The OppA of Salmonella typhimurium had a higher affinity for tri- and tetrapeptides than for di- and pentapeptides (7-9). The OppA of Lactococcus lactis was able to bind peptides from 4 residues up to 35 residues, and had an optimal affinity for nonameric peptides (5, 13). In contrast, DppA had a preference to dipeptides. The current understanding of DppA comes primarily from Gram-negative Escherichia coli and Gram-positive L. lactis (4, 6, 14-17). DppA from E. coli or L. lactis exhibits broad substrate specificity for various dipeptides (6, 14). Previous studies indicate that DppA exclusively controls the specificity of Dpp (6).

The substrate specificity of E. coli DppA (EcDppA, MW = 57,400 Da) was determined by measuring the relative abilities of peptides to compete with a standard radioactively labeled peptide (Gly[125I]Tyr) for binding to DppA, and these indicated that EcDppA had stereochemical preference to various dipeptides containing L-residues, and that the binding affinity of EcDppA was enhanced by alkylation of the N-terminal α-amino group (6). The substrate specificity of the DppA from L. lactis (MW ≈ 59,900 Da) was analyzed by native cationic gel electrophoresis, which showed that the DppA from L. lactis presented a preference for dipeptides containing methionine or arginine (14). Therefore, these results indicate that DppA has preference to some dipeptides although it shows broad substrate
specificity towards many dipeptides. To date, two crystal structures of EcDppA (unliganded/liganded with Gly-Leu) have been reported, which indicated that the binding site for dipeptide in EcDppA is designed to recognize the ligand’s backbone, and that a hydrophobic pocket accepts the leucine side chain of the dipeptide Gly-Leu (16). However, due to the lack of the structures of DppA complex with different dipeptides, the mechanism of DppA for multispecific recognition of dipeptides is still largely unknown.

Deep-sea sediment ecosystems are mainly bacteria and archaea, which contribute to many fundamental processes in nitrogen cycling including organic nitrogen assimilation and remineralization (1, 18-20). It is estimated that the total input of particulate organic nitrogen (PON) to deep-sea sediment is 24 ~ 80 µmol m⁻² d⁻¹ in different areas and at different times (21). Most of this PON is degraded by sedimentary bacterial extracellular enzymes, which is the first step in sedimentary nitrogen cycling. The resulting peptides and amino acids are absorbed by sedimentary microorganisms and then ammonified, nitrified and denitrified by various bacteria and their enzymes. Although PON degradation, ammonification, nitrification and denitrification in deep-sea sediments have been studied (22-25), the mechanism by which deep-sea bacteria absorb peptides and amino acids is seldom studied.

*Pseudoalteromonas* sp. SM9913 is a psychrophilic bacterium that has been isolated from a deep-sea sediment (26). Genomic analyses indicate that *P. sp. SM9913 possesses two peptide transport systems: a Dpp and a Dtp. DppA from *P. sp. SM9913 (PsDppA) is cold-adapted compared to its homolog from *E. coli* (27). In this study, the specificities of *PsDppA to different types of dipeptides were analyzed by using isothermal titration calorimetry (ITC). In light of ITC analysis, we crystallized the unliganded *PsDppA* and *PsDppA* in complex with 4 selected dipeptides (Ala-Phe, Met-Leu, Gly-Glu, Val-Thr) possessing different properties. After structural and mutational analyses, the mechanism of the multispecific recognition of dipeptides by *PsDppA* was illustrated.

**MATERIALS AND METHODS**

Gene Cloning, Site-directed Mutagenesis, Protein Expression and Purification. Gene cloning and heterologous expression of *PsDppA* in *E. coli* strain Rosetta-gami (DE3) were performed as previously described (27). Site-directed mutations were introduced in *PsDppA* using PCR-based methods and were verified using DNA sequencing. All of the mutants were expressed in *E. coli* Rosetta-gami (DE3) using the same methods as for *PsDppA* (27). Recombinant proteins were purified with Ni²⁺-NTA resin (Qiagen, Germany) and then fractionated with gel filtration on a Superdex G200 column (GE Healthcare, the US) with an elution buffer containing 10 mM Tris-HCl (pH 8.0) and 100 mM NaCl.
ITC Measurements. ITC measurements were performed at 25°C using a MicroCal iTC200 System (GE Healthcare, the US). The sample cell was loaded with 250 μl of protein sample (100 μM) and the reference cell contained distilled water. The syringe was filled with 40.07 μl of dipeptides (1 mM). The 25 dipeptides (Ala-Ala, Ala-Gly, Ala-Gln, Ala-Phe, Gly-Gly, Gly-Ser, Gly-Glu, Gly-Leu, Gly-Phe, Val-Thr, Glu-Glu, Phe-Ala, Phe-Val, Tyr-Ala, Asp-Glu, Val-Ser, Met-Leu, Ala-Glu, Ala-Asp, Ala-Lys, Lys-Lys, Glu-Lys, Asp-Lys, Leu-Leu and Gln-Gln) were either purchased from Sigma-Aldrich (the US) or synthesized domestically. All proteins (PsDppA and its mutants) and the 25 dipeptides were kept in the same buffer containing 10 mM Tris-HCl (pH 8.0) and 100 mM NaCl. Titrations were carried out by adding 0.8 μl of dipeptides in the first injection and 2 μl for the following 19 injections with stirring at 1000 rpm.

Crystallization and Data Collection. Purified protein was concentrated to 12 mg ml⁻¹ in a buffer containing 10 mM Tris-HCl (pH 8.0) and 100 mM NaCl. Initial crystallization trials were conducted using the sitting drop vapor diffusion method. Crystals of unliganded PsDppA were obtained at 4°C after one week in hanging drops with a reservoir containing 0.2 M potassium phosphate dibasic (pH 9.2) and 20% polyethylene glycol 3,350. To obtain crystals of the PsDppA-substrate complexes, we mixed the purified PsDppA protein (12 mg ml⁻¹) with Ala-Phe (at a molar ratio of 1:1), Met-Leu (at a molar ratio of 1:1), Gly-Glu (at a molar ratio of 1:2) and Val-Thr (at a molar ratio of 1:2) for 15 min at 20°C. Diffraction-quality crystals of these 4 complexes were obtained at 20°C after one week in hanging drops with a reservoir containing 0.5 M sodium phosphate monobasic monohydrate and 0.9 M potassium phosphate dibasic (pH 7.0).

We added 15% glycerol to the crystallization buffer as a cryoprotectant. X-ray diffraction data were collected on the BL17U1 beam line at the Shanghai Synchrotron Radiation Facility using detector ADSC Quantum 315r. HKL2000 was used to process the data sets (28).

Structure Determination and Refinement. The crystal of PsDppA is in space group P1, but all 4 crystals of the PsDppA-substrate complexes belong to space group P3₂. The structure of PsDppA-AF was solved by molecular replacement using the CCP4 suite with a molecule of EcDppA (PDB code: 1dpp) used as the search model (29). Structures of PsDppA-ML, PsDppA-GE and PsDppA-VT were determined using PsDppA-AF as a search model. Attempts to solve the structure of unliganded PsDppA using PsDppA-AF as a search model were unsuccessful. We divided the model into two parts (domains I and II as one part, and domain III as the second part) and each was used independently as a search model to solve the structure of unliganded PsDppA. The refinement of all of the structures were performed using
Circular Dichroism Spectroscopy Assays. Wild type \( P.s \) DppA and all of the mutants were tested by circular dichroism (CD) spectroscopy assays at 25°C on a J-810 spectropolarimeter (Jasco, Japan). The CD spectra of the \( P.s \) DppA proteins were collected at a final protein concentration of approximately 10 \( \mu M \) from 250 to 200 nm at a scan speed of 500 nm min\(^{-1}\) with a path length of 0.1 cm. All \( P.s \) DppA proteins were in buffer containing 10 mM Tris–HCl (pH 8.0) and 100 mM NaCl.

Protein Structure Accession Numbers. The structures of \( P.s \) DppA, \( P.s \) DppA-AF, \( P.s \) DppA-GE, \( P.s \) DppA-ML and \( P.s \) DppA-VT were deposited in the Protein Data Bank under the accession codes 4QFK, 4QFL, 4QFN, 4QFO and 4QFP.

RESULTS

Substrate Specificity of \( P.s \) DppA. To explore the substrate specificity of \( P.s \) DppA, we tested the affinities of \( P.s \) DppA for 25 dipeptides by using ITC measurements. These dipeptides possess various properties because of their different side chains. \( P.s \) DppA presented affinities for 8 dipeptides under the experimental conditions, which were Ala-Phe, Gly-Phe, Met-Leu, Gly-Leu, Val-Thr, Gly-Glu, Ala-Glu and Ala-Gln (Fig. 1). Besides the high \( K_d \) (138.7 \( \mu M \)) with Ala-Gln, the \( K_d \) values of \( P.s \) DppA with the other 7 dipeptides were in the range of 6-76 \( \mu M \) (Table 1). It can be concluded from the results that \( P.s \) DppA binds a variety of peptides, because the N-terminal residues of these peptides possess short side chains (Gly or Ala), moderate hydrophobic side chains (Val), or long hydrophobic side chains (Met) and that the C-terminal residues are either hydrophilic (Thr, Glu, or Gln) or hydrophobic (Phe or Leu). Therefore, \( P.s \) DppA showed a broad substrate binding specificity. However, we also observed that \( P.s \) DppA exhibited large differences in its affinity for other dipeptides possessing similar properties, e.g., it had an affinity for Ala-Glu but not for Ala-Asp; it had an affinity for Val-Thr but not for Val-Ser. The affinity of \( P.s \) DppA did appear to relate to the amino acid sequence of dipeptides. For example, \( P.s \) DppA showed high affinity to Ala-Phe but little to Phe-Ala.

The thermodynamics parameters for the peptide binding of \( P.s \) DppA were also determined by ITC measurements, including changes in enthalpy (\( \Delta H \)) and entropy (\( \Delta S \)) (Table 1). The result showed that the peptide binding of \( P.s \) DppA was mainly contributed by enthalpic changes, indicating that the main interactions between \( P.s \) DppA and the substrates should be hydrogen bonds and/or electrostatic interactions. However, entropic changes...
contributed significantly to the binding of Met-Leu and Gly-Leu, with the $\Delta S$ of 10.7 and 12.3 cal mol$^{-1}$ deg$^{-1}$, respectively (Table 1). Under the experimental conditions ($T = 298$ K), entropic changes and enthalpic changes contributed approximately equally to the binding of Met-Leu and Gly-Leu, indicating that hydrophobic interactions also played an important role in the binding.

**Overall Structure of PsDppA.** To identify the dipeptide binding mechanism of PsDppA, we obtained the crystal structures of unliganded PsDppA and the complexes of PsDppA with 4 selected dipeptides (Ala-Phe, Met-Leu, Gly-Glu, Val-Thr) possessing different properties. The crystal structures of unliganded PsDppA, a PsDppA-AF complex, a PsDppA-ML complex, a PsDppA-GE complex and a PsDppA-VT complex were determined to 2.30 Å, 1.75 Å, 2.30 Å and 1.90 Å, respectively (Table 2). The 4 structures of the PsDppA-dipeptide complexes are quite similar except for the different substrates, with a root mean square deviation (rmsd) of 0.176-0.360 Å between any two complexes. The structural analyses of the PsDppA-substrate complexes described below are based on the structure of PsDppA-AF complex because the structure of the PsDppA-AF complex had the highest resolution. The final models show that there are eight molecules in the asymmetric unit of PsDppA and two molecules in the asymmetric unit of the PsDppA-AF complex. Each molecule contains three distinct domains that all contain $\alpha/\beta$ folding motifs (Fig. 2A and B). Gel filtration analysis indicated that PsDppA was monomeric in solution (Fig. 2C), indicating that the octamer/dimer PsDppA observed in the structure was a result of crystal packing; therefore, we will limit our discussion to the monomer. The monomers of PsDppA and the PsDppA-AF complex have a similar makeup. Domain I includes residues Gly30-Leu60 (domain I-I), Gly211-Thr288 (domain I-II), and Met507-Lys534 (domain I-III); domain II includes residues Tyr61-Thr210; and domain III includes residues Asn289-Gly506. In domains I and II, $\beta$-sheets are situated at the solvent-accessible surface area, whereas in domain III three anti-parallel $\beta$-sheets ($\beta12$, $\beta15$ and $\beta16$) are surrounded by an outer $\alpha/\beta$ folds, constituting the hydrophobic core.

The recombinant PsDppA does not contain the 19 N-terminal amino acid residues because they are predicted to be a signal peptide that is not recognized by *E. coli* (27). The following 10 residues (Lys20-Gln29) were not observed in the electron density, suggesting that these residues may be disordered. The crystal structure begins at strand $\beta1$ (Gly30-Ala35) of domain I-I. The first $\alpha$-helix ($\alpha1$, Gly50-Leu60) following $\beta1$ connects domain I and domain II; domain I and domain III are connected by two segments that presumably function as a hinge (Thr288-Leu290 and Gly506-Arg508); there is no direct connection between domain II and domain III. At the C-terminus of PsDppA, residues Ala514-Lys534 form a long loop. The substrate binding site is located in the cleft between domain III and domains I and
II.

The conformational differences between PsDppA and the PsDppA-substrate complex were analyzed by superposing the molecules of PsDppA and the PsDppA-AF complex. We observed that most of domains I and II are completely superposed and that domain III rotates ~30° as a rigid-body (Fig. 2D). The hinge region connecting domains I and III of PsDppA enables the domain rotation and the conformational change between “open” and “closed” forms. In the “open” form, the substrate binding site is solvent-accessible, making it possible for the substrate to enter. After binding the substrate, PsDppA changes to the “closed” state, combines with the transmembrane proteins (DppB and DppC) and releases the substrate. It then returns to the “open” form and prepares for binding to a new ligand.

Structural Analysis of the Interactions between PsDppA and Different Dipeptides Based on PsDppA-Substrate Complexes. Structural analysis of the 4 PsDppA-substrate complexes indicated that the main chains of the 4 types of dipeptides bind in the same position on PsDppA with the same orientation, even though they have different side chains (Fig. 3). Six residues in PsDppA (Thr48, Gly50, Arg383, Tyr385, Ala434 and Asp436) can form specific interactions with these substrates. Of these residues, Thr48, Gly50 and Ala434 can form hydrogen bonds with dipeptides through their backbone carbonyl or amino groups. The other three residues can form specific interactions with dipeptides through their side chains; the hydroxyl O of Tyr385 can form a hydrogen bond with dipeptides, and the side chains of Arg383 and Asp436 can participate in the formation of salt bridges to the oppositely charged termini of the ligands (Fig. 3). Hydrogen bonds formed between the main chains of the dipeptides and the residues of PsDppA, and salt bridges formed between the N- and C-termini of the dipeptides and specific acidic and basic residues of PsDppA should help explain the position and orientation of dipeptides in PsDppA.

The interactions between the PsDppA residues and the side chains of the ligands were analyzed (Fig. 4). In the PsDppA-AF complex, several hydrophobic residues are around the C-terminal phenylalanine of the ligand, including Trp414, Ile418, Leu431, Trp433 and Pro459. Together, they form a hydrophobic pocket that can accommodate and stabilize the hydrophobic side chain of phenylalanine by the hydrophobic effects (Fig. 4A). In the PsDppA-ML complex, the C-terminal leucine of the ligand is nestled in the same hydrophobic pocket, and Ile53, Phe182 and Ala526 hydrophobically interact with the N-terminal methionine of the ligand (Fig. 4B). In the PsDppA-GE complex, the negatively charged side chain carboxyl group of Glu in the Gly-Glu ligand interacts with Lys457 and Thr48 by forming a salt bridge and a hydrogen bond, respectively (Fig. 4C). In the PsDppA-VT complex, the N-terminal valine interacts with Ile53 and Phe182 hydrophobically and the C-terminal threonine forms a hydrogen bond with the carbonyl O of Thr48 (Fig. 4D).
When the 4 PsDppA-substrate complexes are aligned, we see that residues involved in forming interactions with substrates (hydrophobic interactions, hydrogen bonds or salt bridges) are almost completely superposed, indicating that these residues do not occupy different positions when PsDppA binds to different ligands (Fig. 5). Still, slight changes to protein side chains can happen. The side chain of Lys457 in the PsDppA-GE complex adopts a different conformation from those in the other three complexes. In the PsDppA-GE complex, the Cβ and the Cα of the C-terminal glutamic acid residue are in almost the same position as the C-terminal amino acid residues of the other dipeptides; however, the carboxyl group in the side chain of glutamic acid moves towards the adjacent residues Thr48 and Lys457 and adopts a different conformation than in other dipeptides (Fig. 5). Additionally, the amino group in the side chain of Lys457 moves 1.7 Å toward the C-terminal glutamic acid of the ligand, shortening the distance 2.85 Å between the amino group of Lys457 and the carboxyl group of glutamic acid, making it possible for the formation of a salt bridge between them (Fig. 5).

Lys457 may play a key role in recognizing the C-terminal glutamic acid of dipeptides.

**Mutational Analysis of the Interactions between PsDppA and Different Dipeptides.** As described above, structural analysis indicated that some residues of PsDppA might play important roles in substrate recognition. To verify the roles of these residues, we constructed 14 mutants that were classified into three groups. Group I contains mutated residues that interact with the main chains of the dipeptides, including Thr48Ala, Arg383Ala, Tyr385Phe, Tyr385Ala and Asp436Ala. Group II contains mutated residues that form hydrophobic interactions with the N-termini of the dipeptides, including Phe182Tyr, Phe182Ala and Ile53Ala. Group III contains mutated residues that interact with the C-termini of the dipeptides, including Trp414Ala, Ile418Ala, Leu431Ala, Pro459Ala, Trp433Ala and Lys457Ala. CD spectroscopy analysis showed that the secondary structures of most mutants showed little change from wild type PsDppA, indicating that the changes in the ligand binding ability of the mutants resulted from residue substitutions rather than overall structural changes (Fig. 6). We also noticed that mutants Trp414Ala and Asp436Ala presented some differences from the wild type PsDppA in their CD spectra (Fig. 6). As tryptophane possesses strong hydrophobic side chain and aspartic acid possesses strong hydrophilic side chain, it is possible that these two mutations cause some changes of the secondary structure of PsDppA.

The substrate binding specificity of each mutant for the 4 selected dipeptides (Ala-Phe, Met-Leu, Gly-Glu and Val-Thr) was analyzed by ITC measurements. The dipeptide binding affinity of most of the mutants was abolished, indicating the importance of these mutated residues for dipeptide recognition (Table 3). On the other hand, some mutants still retained binding affinities for the dipeptides (Table 3). In Group I, mutant Thr48Ala had weak binding affinities for Ala-Phe, and mutant Arg383Ala had weak binding affinities for dipeptides.
Ala-Phe and Met-Leu (Fig. 7). In Group II, mutants Ile53Ala and Phe182Ala had weak
binding affinities for Ala-Phe. Moreover, mutant Phe182Tyr had high binding affinities for
dipeptides Ala-Phe, Gly-Glu and Val-Thr but little for Met-Leu (Fig. 7). Because
phenylalanine and tyrosine both contain a benzene ring, this result suggests that the aromatic
benzene ring of Phe182 is essential for stabilizing the micro-environment of the binding
pocket. The long hydrophobic chain of methionine in the dipeptide Met-Leu may conflict
with the hydrophilic hydroxyl of tyrosine, resulting in a loss in affinity of mutant Phe182Tyr
for the Met-Leu dipeptide. Thermodynamic analysis also indicated that hydrophobic
interactions are very important in the binding of Met-Leu (Table 1). In Group III, the
Pro459Ala mutant had almost the same dipeptide binding affinity as wild type PsDppA,
which indicates that Pro459 does not significantly contribute to binding dipeptides (Fig. 7).
Mutant Lys457Ala had similar affinities to wild type PsDppA for dipeptides Ala-Phe,
Met-Leu and Val-Thr (Fig. 7), but no detectable affinities for Gly-Glu or Ala-Gln, indicating
an important role of Lys457 in binding dipeptides containing a C-terminal glutamic
acid/glutamine. Therefore, mutational analysis provided additional information about the
multiple substrate recognition of PsDppA.

The Multispecific Recognition Mechanism of PsDppA to Dipeptides. Based on
biochemical results, structural analysis and mutational assays, we propose a substrate
recognition mechanism for PsDppA (Fig. 8). Even though PsDppA possesses a large binding
pocket, different types of dipeptides are located in the same position and orientation in the
pocket and are anchored by hydrogen bonds or salt bridges with six adjacent residues (Group
I). Mutations to any of these six residues lead to a large decrease in the binding affinity,
indicating that dipeptide backbone recognition is essential for substrate recognition of
PsDppA. A large hydrophobic pocket formed by Trp414, Ile418, Leu431 and Trp433 in
PsDppA around the side chain of the C-terminal residue of dipeptides plays a key role in
substrate recognition; mutation of any of the 4 residues resulted in a complete loss of binding
affinity. Some mutants in groups I and II retained weak affinities for dipeptides, indicating
that this intact hydrophobic pocket may mitigate the loss of binding affinity caused by other
mutations to some extent (Table 3). Lys457 is necessary for PsDppA to bind dipeptides with a
C-terminal glutamic acid or glutamine. In addition, a weaker hydrophobic pocket around the
N-terminal residue of dipeptides also contributes to substrate binding. In conclusion, PsDppA
possesses a multispecific substrate recognition mechanism. Generally, PsDppA prefers
hydrophobic dipeptides and dipeptides containing a C-terminal glutamic acid or glutamine.

DISCUSSION
Interactions between proteins and peptide ligands can be amino acid sequence dependent or independent (13). Sequence-dependent peptide binding relies on specific interactions between the side chains of ligands and those of proteins, while sequence-independent peptide binding is achieved through interactions between the backbones of ligands and the proteins (13). In the case of DppA, despite the broad substrate specificity, its binding affinities to different substrates differ based on peptide sequence and length (6, 14). EcDppA presented high affinity for Ala-Ala, Ala-Lys, Lys-Ala and Asp-Phe but little affinity for Gly-Gly and Ala-Asp (6). The DppA from L. lactis presented a preference for dipeptides containing methionine or arginine but not for Glu-Ala, Glu-Glu and Gly-Gly (14). In this study, among the 25 dipeptides, PsDppA only exhibited affinity for 8 dipeptides under the experimental conditions.

The 8 dipeptides are various in amino acid composition and sequence. Therefore, bacterial DppAs have multiple specificity for dipeptides.

To identify the multispecific recognition mechanism of DppA, 4 structures of PsDppA-substrate complexes were obtained and compared. Structural analysis and mutational assays identify the specific intermolecular interactions formed in the substrate binding process of PsDppA, which can provide a molecular scale explanation of the multispecific recognition exhibited by PsDppA for different dipeptides. First, dipeptides are bound in the same position and orientation in PsDppA. The structure of EcDppA complexed with Gly-Leu showed that the peptide Gly-Leu in EcDppA also occupied the same position and orientation as those in PsDppA (16). This may explain why PsDppA exhibits a high affinity for Ala-Phe but not for Phe-Ala. Second, the peptides are bound in an extended, identical conformation due to the interactions between their backbones and PsDppA, in particular by anchoring their N- and C-termini through ion-pair interactions. This result is consistent with the peptide binding mode of EcDppA and the OppA of E. coli and S. typhimurium (7-10, 16). However, for the OppA of L. lactis, the peptide termini are not fixed with salt bridges, explaining the wide tolerance of peptides with varied lengths and sequences (11, 32). Moreover, the binding pocket of OppA is significantly larger than that of DppA (11), and a number of water molecules located in the binding pocket will adjust according to the ligand size (7, 33). In the case of DppA, few water-mediated interactions in the peptide binding pocket are observed in either PsDppA or EcDppA (16). Third, the N- and C-terminal pockets for binding the peptide side chains are mostly hydrophobic. The C-terminal pocket is more spacious and capped by a lysine residue, accepting more bulky hydrophobic side chains or long acidic/polar side chains (Phe, Leu, Glu or Gln). This may be the reason why PsDppA has a higher affinity for dipeptides possessing a C-terminal hydrophobic side chain (e. g. Ala-Phe) than those possessing a short C-terminal side chain (e. g. Gly-Gly and Ala-Ala). A similar hydrophobic pocket is also identified to accept the side chains of leucine of Gly-Leu in EcDppA (16).
the OppA of <i>L. lactis</i>, a hydrophobic pocket was also found to play an important role in peptide binding (32). Lys457 is indispensable for the recognition of dipeptides with a C-terminal glutamic acid or glutamine. Lys457 can only make a hydrogen bond with glutamine but not a salt bridge, which probably explains the 20-fold lower affinity of <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b> for Ala-Gln compared to Ala-Glu (Table 1). However, Lys457 does not interact with a C-terminal aspartic acid because the distance between them is too large for a salt bridge to form. Thus, <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b> has a high affinity for Ala-Glu but not for Ala-Asp, despite the similarities between glutamic acid and aspartic acid. We notice that <i>E</i><i><i><i><i><i>DppA</i></i></i></i></i> only presented a moderate affinity to Ala-Glu (16), which might result from the replacement of the basic Lys457 in <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b> by the neutral Ser429 in <i>E</i><i><i><i><i><i>DppA</i></i></i></i></i>. Fourth, the N-terminal pocket of <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b> is relatively narrow, thereby only accepting either short or long, unbranched amino acid side chains (Gly, Ala, Val or Met). Previous structural analysis showed that the substrate binding pocket of <i>E</i><i><i><i><i><i>DppA</i></i></i></i></i> appears to be spacious enough to accept any variety of the 20 natural amino acids (16). The results in this study demonstrate that <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b> binds to different substrates through a precise mechanism in addition to spatial constraints and that the side chains of the dipeptides greatly contribute to the process of substrate recognition of <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b>.

A sequence alignment of <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i> with its homologs from other species shows that, among the residues involved in forming hydrogen bonds or salt bridges with the main-chain of the dipeptides, Arg383, Tyr385, Asp436, Thr48 and Gly50 are highly conserved in Gram-negative bacteria and Ala434 is less conserved (Fig. 9). Because Ala434 forms hydrogen bonds with dipeptides through main-chain-to-main-chain interactions, variations at this position should not affect substrate recognition of DppA. In addition, residues interacting with the side chains of the dipeptides are less conserved in Gram-negative bacteria, indicating that different species may possess different substrate specificities. Thus, the substrate recognition mechanism of DppA may be similar in Gram-negative bacteria. The binding of DppA to dipeptides primarily depends on recognition of the main-chain through hydrogen bonds and salt linkages and residues around the termini of dipeptides are responsible for specific substrate recognition by DppA.

Dpp plays an important role in peptide uptake for bacterial nutrition. Metatranscriptomic and metaproteomic studies indicate that peptide transporters are extensively expressed in the marine environment (34, 35), suggesting that peptide transporters are functionally active in marine bacteria. Because the substrate recognition mechanism of <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b> may be common to Gram-negative bacteria, our results on the substrate recognition mechanism of <i>P</i><i><i><i><i><i>DppA</i></i></i></i></i><b><i><i><i><i><i>Ps</i></i></i></i></i></b> provides a better understanding of how marine Gram-negative bacteria absorb peptides, which is helpful for the study of marine nitrogen cycling. Moreover, because DppA also acts as a primary chemoreceptor toward various peptide analogs that have antimicrobial activity (3, 4, 6), our results may provide relevant information for the design of peptide antibiotics against...
Gram-negative pathogens.

AUTHOR CONTRIBUTIONS

Xi-Ying Zhang and Yu-Zhong Zhang designed research; Chun-Yang Li, Xiu-Lan Chen, Wei-Xin Zhang, and Peng Wang performed research; Chun-Yang Li, Xiu-Lan Chen, Bin-Bin Xie, Hai-Nan Su, Qi-Long Qin, and Bai-Cheng Zhou analyzed data; and Chun-Yang Li and Xiu-Lan Chen wrote the paper.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation of China (31290230, 31290231 and 31170055), China Ocean Mineral Resources Research and Development Association Program (DY125-15-T-05 and DY125-15-R-03), Hi-Tech Research and Development program of China (2012AA092103 and 2014AA093509), and the Fundamental Research Funds of Shandong University (2014QY006-21 and 2014QY006-22).

REFERENCES


structures of the periplasmic oligopeptide binding protein in the unliganded form and in complex with lysyllysine. Biochemistry 36:9747-9758.


30. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW,


FIGURE LEGENDS

Figure 1. ITC data for titrations of various dipeptides into wild type PsDppA. ITC traces (upper panels) and integrated binding isotherms (lower panels) are shown for each dipeptide. Substrates (S) are shown in each ITC traces.

Figure 2. The overall structural analysis of PsDppA. A. Ribbon representation of a PsDppA monomer. B. Ribbon representation of a PsDppA-AF complex monomer. All α/β motifs have been labeled. Different domains are represented with different colors: cyan for domain I-I (Gly30-Leu60), blue for domain I-II (Gly211-Thr288), yellow for domain I-III (Met507-Lys534), green for domain II (Tyr61-Thr210) and red for domain III (Asn289-Gly506). C. Analysis of the form of PsDppA in solution by gel filtration. Conalbumin (MW = 75,000 Da, GE Healthcare) and ovalbumin (MW = 43,000 Da, GE Healthcare) were used as markers. The predicted molecular mass of PsDppA is ~57,000 Da. D. Superimposed structures of PsDppA (colored in orange) and the PsDppA-AF complex (colored in magenta). Partial α/β folding motifs were labeled.
Figure 3. Interactions between PsDppA residues and the main chains of ligands. A. PsDppA-AF complex. B. PsDppA-ML complex. C. PsDppA-GE complex. D. PsDppA-VT complex. Bound dipeptides are colored in cyan and residues involved in binding substrate are colored in green. The 2Fo-Fc densities for dipeptides and PsDppA residues are contoured in blue at 1.5σ. Hydrogen bonds and salt bridges are represented by black dashed lines.

Figure 4. Interactions between PsDppA residues and the side chains of ligands. Bound dipeptides are colored in cyan, and PsDppA residues interacting with the substrates (hydrophobic interactions, hydrogen bonds and salt bridges) are colored in yellow. The 2Fo-Fc densities for dipeptides are contoured in blue at 1.5σ. Hydrogen bonds and salt bridges are represented by black dashed lines. A. PsDppA-AF complex. B. PsDppA-ML complex. C. PsDppA-GE complex. D. PsDppA-VT complex.

Figure 5. Conformational alignment of dipeptides and PsDppA residue interactions from different PsDppA-substrate complexes. The PsDppA-AF complex is colored in green, the PsDppA-ML complex is colored in magenta, the PsDppA-GE complex is colored in cyan, and the PsDppA-VT complex is colored in yellow.

Figure 6. CD spectra of wild type PsDppA and its mutants. CD spectra of the proteins at a final concentration of approximately 10 μM were collected from 250 to 200 nm.

Figure 7. ITC data for titrations of various dipeptides into PsDppA mutants. ITC traces (upper panels) and integrated binding isotherms (lower panels) are shown. M, mutant; S, substrate.

Figure 8. The multispecific substrate recognition mechanism of PsDppA. The dipeptide molecule is represented by a ball-and-stick model. Residues interacting with the main chains of the dipeptides are colored in red, and the bonds are shown as red arrows. Residues involved in hydrophobic interactions and salt bridges with side chains of dipeptides are colored in green and blue, respectively, and their probable interactions are shown in the same color.

Figure 9. Sequence alignment of PsDppA with those in other species. Residues interacting with the main chains of dipeptides are marked in a red box. Residues forming hydrophobic interactions and salt bridges with the side chains of the dipeptides are marked in green and blue boxes, respectively. The numbers above sequences refer to the amino acid numbers in the PsDppA sequence. Sequences 1 to 11 are DppA polypeptides from the following Gram-negative bacteria: 1, Pseudoalteromonas sp. SM9913; 2, Pseudoalteromonas tunicata D2; 3, Alteromonas sp. SN2; 4, Shewanella sediminis HAW-EB3; 5, Escherichia coli MS 78-1; 6, Rheinheimera sp. A13L; 7, Alshewanella jeotgali KCTC 22429; 8, Glaciecola sp. 4H-3-7+YE-5; 9, Colwellia psychrerythraea 34H; 10, Idiomarina baltica OS145; 11, Photobacterium damselae subsp. damselae CIP 102761. Sequence 12 is DppA polypeptide from the Gram-positive bacterium Lactococcus lactis subsp. cremoris NZ9000.
Table 1. Thermodynamic parameters determined by ITC measurements.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>$K_d$ (μM)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$\Delta S$ (cal mol$^{-1}$ deg$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala-Phe</td>
<td>7.7 ± 1.5</td>
<td>-9.5 ± 0.1</td>
<td>-9.3 ± 1.2</td>
</tr>
<tr>
<td>Gly-Phe</td>
<td>39.1 ± 2.7</td>
<td>-8.5 ± 0.7</td>
<td>-8.2 ± 2.4</td>
</tr>
<tr>
<td>Met-Leu</td>
<td>53.4 ± 12.9</td>
<td>-2.7 ± 0.3</td>
<td>10.7 ± 0.5</td>
</tr>
<tr>
<td>Gly-Leu</td>
<td>29.5 ± 5.4</td>
<td>-2.5 ± 0.6</td>
<td>12.3 ± 2.4</td>
</tr>
<tr>
<td>Val-Thr</td>
<td>58.6 ± 14.6</td>
<td>-9.3 ± 0.5</td>
<td>-11.7 ± 2.1</td>
</tr>
<tr>
<td>Gly-Glu</td>
<td>76.0 ± 14.4</td>
<td>-7.4 ± 0.7</td>
<td>-5.6 ± 1.7</td>
</tr>
<tr>
<td>Ala-Glu</td>
<td>6.6 ± 1.4</td>
<td>-13.2 ± 0.1</td>
<td>-20.4 ± 0.4</td>
</tr>
<tr>
<td>Ala-Gln</td>
<td>138.7 ±14.9</td>
<td>-4.9 ± 0.4</td>
<td>1.0 ± 1.0</td>
</tr>
</tbody>
</table>
Table 2. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diffraction data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td>P&lt;sub&gt;3&lt;/sub&gt;</td>
<td>P&lt;sub&gt;3&lt;/sub&gt;</td>
<td>P&lt;sub&gt;3&lt;/sub&gt;</td>
<td>P&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
<tr>
<td>Unit cell</td>
<td>94.0, 90.0, 140.5</td>
<td>106.1, 106.1, 101.0</td>
<td>100.4, 100.4, 106.9</td>
<td>106.0, 106.0, 100.4</td>
<td>105.9, 105.9, 101.3</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>90.0, 90.0, 120.0</td>
<td>90.0, 90.0, 120.0</td>
<td>90.0, 90.0, 120.0</td>
<td>90.0, 90.0, 120.0</td>
<td>90.0, 90.0, 120.0</td>
</tr>
<tr>
<td>Resolution range (Å)*</td>
<td>50.0-2.3 (2.38-2.30)</td>
<td>50.0-1.75 (1.81-1.75)</td>
<td>50.0-2.3 (2.38-2.30)</td>
<td>50.0-2.3 (2.38-2.30)</td>
<td>50.0-1.9 (1.97-1.90)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>549888</td>
<td>414449</td>
<td>286888</td>
<td>155384</td>
<td>279078</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>185390</td>
<td>126624</td>
<td>55190</td>
<td>55803</td>
<td>58026</td>
</tr>
<tr>
<td>Redundancy</td>
<td>1.9 (1.9)</td>
<td>3.3 (3.3)</td>
<td>5.6 (3.3)</td>
<td>2.8 (2.8)</td>
<td>2.8 (2.8)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95.1 (95.4)</td>
<td>94.4 (100.0)</td>
<td>100.0 (100.0)</td>
<td>99.3 (99.6)</td>
<td>99.2 (99.8)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;**</td>
<td>0.09 (0.36)</td>
<td>0.10 (0.57)</td>
<td>0.14 (0.65)</td>
<td>0.07 (0.24)</td>
<td>0.09 (0.58)</td>
</tr>
<tr>
<td>I/σI</td>
<td>9.5 (3.0)</td>
<td>25.4 (2.5)</td>
<td>21.7 (4.7)</td>
<td>24.7 (6.1)</td>
<td>18.9 (2.7)</td>
</tr>
<tr>
<td><strong>Refinement statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-factor</td>
<td>0.19</td>
<td>0.16</td>
<td>0.16</td>
<td>0.15</td>
<td>0.16</td>
</tr>
<tr>
<td>Free R-factor</td>
<td>0.26</td>
<td>0.18</td>
<td>0.21</td>
<td>0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>RMSD from ideal geometry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.008</td>
<td>0.007</td>
<td>0.008</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.18</td>
<td>1.10</td>
<td>1.12</td>
<td>1.07</td>
<td>0.97</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favourable</td>
<td>93.7</td>
<td>93.6</td>
<td>95.5</td>
<td>95.9</td>
<td>95.7</td>
</tr>
<tr>
<td>Allowed</td>
<td>5.5</td>
<td>4.5</td>
<td>4.2</td>
<td>4.1</td>
<td>4.1</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.8</td>
<td>0.10</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Overall B-factors (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>29.3</td>
<td>24.5</td>
<td>28.8</td>
<td>26.7</td>
<td>23.6</td>
</tr>
</tbody>
</table>

*Numbers in parentheses refer to data in the highest resolution shell  ** R<sub>merge</sub>=\sum_{i,j}||I(hkl)-\langle I(hkl)\rangle||/\sum_{i,j}||I(hkl)||
Table 3. Substrate binding specificity of the *Ps*DppA mutants analyzed by ITC measurement.

<table>
<thead>
<tr>
<th></th>
<th>Ala-Phe</th>
<th>Met-Leu</th>
<th>Gly-Glu</th>
<th>Val-Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Group I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr48Ala</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arg383Ala</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyr385Phe</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyr385Ala</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp436Ala</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Group II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe182Tyr</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Phe182Ala</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ile53Ala</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Group III</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp414Ala</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ile418Ala</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leu431Ala</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Trp433Ala</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lys457Ala</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td>Pro459Ala</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>

“++”, no substantial difference in the dipeptide binding affinity was observed between wild type *Ps*DppA and its mutants under the experimental conditions.

“+”, an obviously weaker dipeptide binding affinity was detected on the mutants compared to wild type *Ps*DppA under the experimental conditions.

“-”, little dipeptide binding affinity was detected under the experimental conditions.
- **PsDppA-AF complex**
- **PsDppA-ML complex**
- **PsDppA-GE complex**
- **PsDppA-VT complex**

Diagram showing interacting amino acids and dipeptides with labels:
- F182
- T48
- D436
- W433
- I53
- A526
- R383
- K457
- P459
- L431
- W414
- Y385
- I418

Distance label: 1.7 Å