Structure of RizA, an L-amino-acid ligase from *Bacillus subtilis*

Wataru Kagawa,^{a*} Toshinobu Arai,^b Shun Ishikura,^b Kuniki Kino^{b*} and Hitoshi Kurumizaka^c

Received 5 May 2015

Accepted 1 July 2015

Edited by W. N. Hunter, University of Dundee, Scotland

Keywords: L-amino-acid ligase; ATP-grasp fold; rhizocticin; dipeptide synthesis.

PDB reference: RizA, 4wd3

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

^aDepartment of Interdisciplinary Science and Engineering, Program in Chemistry and Life Science, School of Science and Engineering, Meisei University, 2-1-1 Hodokubo, Hino-shi, Tokyo 191-8506, Japan, ^bDepartment of Applied Chemistry, Faculty of Science and Engineering, Waseda University, 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan, and ^cLaboratory of Structural Biology, Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan. *Correspondence e-mail: wataru.kagawa@meisei-u.ac.jp, kkino@waseda.jp

RizA is an L-amino-acid ligase from *Bacillus subtilis* that participates in the biosynthesis of rhizocticin, an oligopeptide antibiotic. The substrate-free form of RizA has been crystallized and the structure was solved at 2.8 Å resolution. The amino-acid-binding site appears to be capable of accommodating multiple amino acids, consistent with previous biochemical studies.

1. Introduction

L-Amino-acid ligase (LAL; EC 6.3.2.28) catalyzes peptide-bond formation between two L-amino acids. LAL is a member of the ATP-dependent carboxylate–amine/thiol ligase superfamily (Galperin & Koonin, 1997) and catalyzes the ligation reaction, which involves an aminoacyl-phosphate intermediate, in an ATP-dependent manner (Fan *et al.*, 1995). Several LALs have been identified and they exhibit different amino-acid substrate specificities. They include BacD (also named YwfE) from *Bacillus subtilis* 168 (Inaoka *et al.*, 2003; Tabata *et al.*, 2005; Tabata & Hashimoto, 2007), RSp1486a from *Ralstonia solanacearum* JCM10489 (Kino, Nakazawa *et al.*, 2008), BL00235 from *B. licheniformis* NBRC12200 (Kino, Noguchi *et al.*, 2008), PSPPH_4299 from *Pseudomonas syringae* pv. *phaseolicola* 1448A (Arai & Kino, 2008; Bender *et al.*, 1999) and TabS from *P. syringae* 14081 (Arai *et al.*, 2013). LALs are potentially useful for the industrial production of antibiotics, phytotoxins and other medical and healthcare-related peptides (Yagasaki & Hashimoto, 2008; Mills *et al.*, 2011; Santos *et al.*, 2012). For example, BacD is involved in the biosynthesis of the antibiotic bacilysin and is utilized in the industrial production of alanyl-glutamine, which is an important component of parenteral nutrition. Another example is TabS, which is involved in the biosynthesis of tabtoxin. TabS reportedly synthesizes various functional peptides (Arai *et al.*, 2013). However, the detailed mechanisms of the peptide syntheses catalyzed by these enzymes are still unclear. Structural information about these enzymes would facilitate the design of ligases with modified activities that are geared towards the industrial production of useful peptides.

LALs contain the ATP-grasp fold, which is composed of three conserved domains referred to as the A-domain (N-terminal domain), the B-domain (central domain) and the C-domain (C-terminal domain). These three domains commonly grasp the ATP molecule, and also provide binding sites for the Mg²⁺ ion and the amino-acid substrate. In general,

the A-, B- and C-domains each have variable loops that differ in their amino-acid contents and lengths, and form part of the active site. The crystal structures of two LALs, BacD and BL00235, revealed that the variable loop regions are responsible for the different amino-acid substrate specificities displayed by these enzymes (Shomura *et al.*, 2012; Suzuki *et al.*, 2012). In the present study, we focused on the L-amino-acid ligase RizA, which is involved in the biosynthesis of the peptide antibiotic rhizocticin. A previous biochemical study revealed that RizA exhibits strict substrate specificity for arginine as the N-terminal amino acid of the synthesized dipeptide (Kino *et al.*, 2009). In contrast, multiple amino acids were accommodated as the C-terminal amino acid of the synthesized dipeptide. In the present study, we performed an X-ray crystallographic study of the RizA protein and determined its structure in the substrate-free form. The putative binding sites for ATP and the amino-acid substrate were examined.

2. Materials and methods

2.1. Protein expression and purification

The *rizA* gene [DDBJ (DNA Data Bank of Japan) accession No. AB437349] was amplified from the genomic DNA of *B. subtilis* (strain NBRC3134) by PCR using forward and reverse primers with sequences 5'-GTTTATTTCCCATGGT-GCGTATTTTACTCATTAATTC-3' and 5'-CATAGGATC-CTTATTGGTTCATCATATGATCCTCTG-3', respectively. The second codon of the *rizA* gene, CTT (encoding leucine), was changed to GTG (valine) to avoid protein degradation by the bacterial N-end rule (Tobias *et al.*, 1991). The PCR fragment was inserted between the NcoI and BamHI sites in the pET-21d vector (Merck). The RizA protein expressed from the resulting plasmid lacked affinity-tag sequences. The selenomethionine-substituted RizA was overexpressed in *Escherichia coli* strain B834(DE3) cultured in LeMaster medium containing 50 µg ml⁻¹ ampicillin (final concentration) (LeMaster & Richards, 1985; Hendrickson *et al.*, 1990; Leahy *et al.*, 1992). *rizA* expression was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (final concentration) when the culture reached an optical density (A_{660}) of between 0.5 and 0.7. The cultured cells were harvested by centrifugation (4160g, 10 min, 277 K), resuspended in 0.1 M Tris-HCl buffer pH 8.0 and lysed by sonication at 277 K. The cell lysate was cleared of insoluble material by centrifugation at 17 000g for 30 min at 277 K. All purification procedures were performed at 277 K or on ice. The cleared cell lysate was loaded onto a Toyopearl DEAE 650S (Tosoh) column equilibrated with buffer A (50 mM Tris-HCl buffer pH 8.0). The RizA-bound column was washed with buffer A. RizA was eluted using a linear gradient of 0–0.6 M KCl in buffer A. The fractions were assayed for ligase activity using a colorimetric assay that involves a reaction between an amino acid and hydroxylamine (NH₂OH) to yield an aminoacyl hydroxamate, which forms a red complex with iron(III) (Kino *et al.*, 2009). Briefly, a reaction mixture consisting of 30 mM arginine,

200 mM NH₂OH, 30 mM ATP, 30 mM MgSO₄·7H₂O and a small portion of the DEAE column fraction was incubated overnight at 303 K. To this reaction mixture, 8% trichloroacetic acid (75 µl) and 3.4% FeCl₃ dissolved in 2 N HCl (75 µl) were added and mixed vigorously, and the mixture was then centrifuged. The absorbance at 490 nm of the resulting supernatant was measured. The DEAE fractions containing ligase activity were pooled and ammonium sulfate was added to a final concentration of 1 M. This protein solution was loaded onto a HiLoad 16/10 Phenyl Sepharose HP column (GE Healthcare) equilibrated with buffer A containing 1.5 M ammonium sulfate. After washing the column with the same buffer, RizA was eluted with a linear gradient of 1.5–0 M ammonium sulfate in buffer A. The fractions with ligase activity were combined and concentrated by ultrafiltration. RizA was then fractionated through a HiLoad 16/60 Superdex 200 pg column (GE Healthcare) equilibrated with buffer A containing 150 mM NaCl. The fractions with ligase activity were combined and applied onto a Mono Q 5/50 GL column (GE Healthcare) equilibrated with buffer A. After washing the RizA-bound Mono Q column with buffer A, the protein was eluted with a linear gradient of 0–0.6 M KCl in buffer A. The fractions containing ligase activity were pooled and concentrated by ultrafiltration. The RizA concentration was determined by the Bradford method with bovine serum albumin as the standard.

2.2. Crystallization and diffraction data collection

The purified selenomethionine-substituted RizA (1 mg ml⁻¹ final concentration) was mixed with ADP (1 mM final concentration) and tris(2-carboxyethyl)phosphine-HCl (2.5 mM final concentration) prior to crystallization. Tris(2-carboxyethyl)phosphine-HCl, which is a reducing agent, was added to prevent RizA from aggregating prior to and during crystallization. Crystals of selenomethionine-substituted RizA were grown at 293 K using the hanging-drop vapour-diffusion method by mixing 1 µl protein solution and 1 µl reservoir solution consisting of 0.1 M sodium HEPES pH 6.8, 0.2 M MgCl₂·6H₂O, 26% (w/v) polyethylene glycol 3350, 0.1 M sodium iodide. The crystals were cryoprotected in reservoir solution supplemented with 20% (v/v) glycerol and flash-cooled in a stream of nitrogen gas (~100 K). X-ray diffraction data were collected using an ADSC Quantum 315 CCD detector on beamline 17A of the Photon Factory, Tsukuba, Japan. The data sets were processed and scaled using the HKL-2000 program suite (Otwinowski & Minor, 1997).

2.3. Structural determination and refinement

The structure of selenomethionine-derivatized RizA was determined using SAD data collected from a single crystal. Ten of 13 selenium positions were determined by *SHELXD* (Sheldrick, 2010). The phase data were calculated using *SOLVE* (Terwilliger & Berendzen, 1999), followed by density modification using *RESOLVE* (Terwilliger, 2000). After density modification, the figure of merit improved from 0.27 to 0.68. Automatic model building and refinement were

performed using the *AutoBuild* option of *PHENIX* (Adams *et al.*, 2010). The partial model was subjected to iterative rounds of manual model building using *Coot* (Emsley *et al.*, 2010) and refinement using *PHENIX* using individual isotropic atomic displacement parameters. All structure figures were created using *PyMOL* (<http://www.pymol.org>). Structural superpositions of RizA, BL00235 and BacD were performed using the align command in *PyMOL*. Data-collection and refinement statistics are given in Table 1.

3. Results and discussion

3.1. Structure of substrate-free RizA

The overall structure of RizA is similar to those of the previously determined LALs BL00235 (PDB entry 3vot; Suzuki *et al.*, 2012) and BacD (PDB entry 3vmm; Shomura *et al.*, 2012) (Fig. 1). Like BL00235 and BacD, RizA has an ATP-grasp fold composed of three $\alpha+\beta$ domains referred to as the A-, B- and C-domains. A visual inspection, along with a comparison of the RizA structure with those of BL00235 and BacD, revealed that in RizA the A-, B- and C-domains are composed of amino-acid residues 1–121, 122–195 and 196–413, respectively, with folds that are essentially identical to those of BL00235 and BacD. However, a superposition of the RizA structure with the structures of BL00235 and BacD revealed that large regions of the α -helices and the β -strands in the A-, B- and C-domains of RizA superimposed poorly on those of BL00235 and BacD (Figs. 2*a*, 2*b* and 2*c*). The root-mean-square difference (r.m.s.d.) between BL00235 and BacD was 2.195 Å (over 1368 atoms), whereas the r.m.s.d.s between RizA and BL00235 and between RizA and BacD were 6.124 Å (over 1505 atoms) and 5.274 Å (over 1482 atoms), respectively. The larger r.m.s.d. values observed between RizA and the other two LALs are probably owing to the absence of ADP binding in the RizA structure. In the BL00235 and BacD structures the ADP molecule binds to the interface of the A-, B- and C-domains and probably draws the three domains close together (Figs. 2*d*, 2*e* and 2*f*). By contrast, the three domains in the ADP-free RizA are more open compared with those of BL00235 and BacD. Although ADP and Mg^{2+} ion were included during the crystallization of RizA, both were absent

Table 1

Statistics for crystallographic structural determination.

Values in parentheses are for the highest resolution shell.

Data collection	
Beamline	BL17A, Photon Factory
Wavelength (Å)	0.9788
Space group	$P2_1$
Unit-cell parameters (Å, °)	$a = 58.4$, $b = 87.3$, $c = 85.3$, $\beta = 97.5$
Resolution range (Å)	48.24–2.80 (2.95–2.80)
Unique reflections	20941
Multiplicity	7.3 (7.4)
Completeness (%)	99.5 (99.7)
$R_{\text{merge}}^{\dagger}$ (%)	9.1 (29.4)
$\langle I/\sigma(I) \rangle$	15.5 (7.0)
Wilson B factor (Å ²)	54.3
Refinement	
Resolution (Å)	48.24–2.80
R factor/free R factor (%)	20.1/25.5
B factor (Å ²)	37.0
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.678
Ramachandran plot, residues (%)	
Favoured	97.9
Allowed	2.1
Outliers	0

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th observation of reflection hkl and $\langle I(hkl) \rangle$ is the weighted average intensity for all i observations of reflection hkl .

in the crystal structure for unknown reasons. However, we were not able to crystallize RizA in the absence of ADP, suggesting that ADP may have weakly bound to RizA. In the RizA structure, amino-acid residues 222–230, 368–375 and 406–413 in the C-domain were not visible in the final electron-density map. Amino-acid residues 222–230 and 368–375 are part of the loop regions that are proximal to the putative ATP and the amino-acid substrate-binding pockets. Notably, the corresponding regions in the BL00235 and BacD structures bound to ADP are clearly visible. This difference suggested that the conformations of these regions become fixed when ADP binds to its binding pocket.

3.2. Putative binding sites for ATP and the amino-acid substrate

To examine the putative binding sites for ATP and the amino-acid substrate in the RizA structure in more detail, the

Figure 1

Overall structure of RizA. The RizA structure determined at 2.8 Å resolution is depicted as a ribbon representation. RizA has an ATP-grasp fold consisting of the A-, B- and C-domains (coloured blue, purple and yellow, respectively).

amino-acid sequences of RizA, BL00235 and BacD were aligned based on the superimposed structures of these proteins. We then inspected the amino-acid residues of the BacD protein that directly interact with ADP and the amino-acid substrate and identified the corresponding residues of RizA from the aligned sequence. In the BacD structure, Lys138, Lys178, Glu226, Glu311 and Glu324 interact with the ADP molecule mainly by hydrogen bonding (Fig. 3*b*). These amino-acid residues are conserved in BL00235 as well as in RizA, as revealed from the sequence alignments and the

superimposed structures (Figs. 3*a* and 3*c*). Thus, the ATP-binding pocket appears to be conserved among the various LALs. By contrast, the substrate-binding pockets of these proteins have significantly different sizes and amino-acid compositions. In the BacD structure, the amino-acid substrate is recognized by Leu12, Trp76, Leu110, Glu273, Glu311, Glu324 and Arg328. Glu311, Glu324 and Arg328 are bound to the peptide bond of the dipeptide substrate and are conserved in BL00235 and RizA. However, Glu273, Leu12 and Leu110, which recognize the side chains of the dipeptide substrate, are

Figure 2
Comparison of the domain orientations in the RizA structure with those of other LALs. (*a–c*) Structural superpositions of RizA (*a*) with BL00235 (*b*) and BacD (*c*). RizA is coloured according to the A-, B- and C-domains (blue, purple and yellow, respectively). BL00235 and BacD are coloured light cyan and light green, respectively. (*d–f*) Surface representations of RizA (*d*), BL00235 (*e*) and BacD (*f*). For the three proteins, the A-, B- and C-domains are coloured blue, purple and yellow, respectively. ADP (coloured red), divalent ions (Mg^{2+} or Ca^{2+} , coloured green) and the amino-acid substrate (BacD only, coloured dark blue) bound to BL00235 (*g*) and BacD (*h*) are viewed from the same angle as in (*e*) and (*f*), respectively.

Figure 3

The putative ATP-binding site of RizA. The conserved amino-acid residues involved in ADP binding, revealed from the crystal structures of BacD and BL00235, are shown in stick representations. All three domains appear to contribute to ATP binding. (a) Lys111 (coloured blue) is located in the A-domain, Lys151 and Glu193 (coloured purple) are located in the B-domain, and Glu259 and Glu272 (coloured yellow) are located in the C-domain. The surrounding protein regions are shown in ribbon representations and are coloured according to the three domains. The green spheres in (b) and (c) depict Mg^{2+} and Ca^{2+} ions, respectively. Dashed lines (coloured light blue) indicate hydrogen-bonding or ionic interactions.

Figure 4

The putative dipeptide-binding site of RizA. The predicted dipeptide-binding site of RizA is depicted as surface (a) and ribbon (b) representations. The phosphorylated phosphinate L-alanyl-L-phenylalanine and ADP from the superpositioned BacD structure are shown. The A-domain (white dashed circle) is the predicted binding site for the arginine side chain. Several hydrophobic residues are located in this region that may interact with the alkyl chain of the arginine residue. In addition, the disordered amino-acid residues 222–230 (VPLLKGMK) and 368–375 (DEKQSSSF) are near this region and contain several hydrophobic amino-acid residues that may interact with arginine. The carboxyl group of the Asp376 residue is coloured red. The B-domain (white dashed circle) is the predicted binding site for the C-terminal amino-acid residue of the dipeptide.

not conserved in BL00235 and RizA. Previous biochemical studies demonstrated that these enzymes exhibit different substrate specificities. Thus, it is reasonable to conclude that the amino-acid residues constituting the substrate-binding pockets of these enzymes are different.

A previous study of RizA revealed that the enzyme exhibited high specificity towards arginine as the N-terminal side of the synthesized dipeptide (Kino *et al.*, 2009). The putative substrate-binding site is a hydrophobic cavity that is apparently large enough to accommodate the arginine residue (Fig. 4). The cavity contains hydrophobic residues, which may interact with the hydrocarbon portion of the arginine side chain. Within this hydrophobic environment, Asp376 is located deep within the cavity and may be involved in electrostatic interactions with the positively charged portion of the arginine side chain. The conformation of the putative binding site for the C-terminal amino-acid residue of the dipeptide appears to be open. Therefore, the precise conformational changes that occur near this binding site will be identified by co-crystallizing RizA with various dipeptides and/or an ATP analogue.

Acknowledgements

The synchrotron-radiation experiments were performed on the BL41XU beamline of SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI; Proposal Nos. 2010B1375, 2011A1528 and 2011B1133) and on the AR-NW12 and BL-17A beamlines of the Photon Factory (Proposal No. 2010G615). We are grateful to the beamline scientists for their assistance in data collection. This work was supported in part by JSPS KAKENHI Grant Nos. 24570138 (to WK) and 26289318 (to KK). This work was also supported in part by MEXT KAKENHI Grant Nos. 26116521 (to WK) and 25116002 (to HK). KK, TA and HK were also supported by the Waseda Research Institute for Science and Engineering.

References

Adams, P. D. *et al.* (2010). *Acta Cryst. D* **66**, 213–221.

- Arai, T., Arimura, Y., Ishikura, S. & Kino, K. (2013). *Appl. Environ. Microbiol.* **79**, 5023–5029.
- Arai, T. & Kino, K. (2008). *Biosci. Biotechnol. Biochem.* **72**, 3048–3050.
- Bender, C. L., Alarcón-Chaidez, F. & Gross, D. C. (1999). *Microbiol. Mol. Biol. Rev.* **63**, 266–292.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Fan, C., Moews, P. C., Shi, Y., Walsh, C. T. & Knox, J. R. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 1172–1176.
- Galperin, M. Y. & Koonin, E. V. (1997). *Protein Sci.* **6**, 2639–2643.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Inaoka, T., Takahashi, K., Ohnishi-Kameyama, M., Yoshida, M. & Ochi, K. (2003). *J. Biol. Chem.* **278**, 2169–2176.
- Kino, K., Kotanaka, Y., Arai, T. & Yagasaki, M. (2009). *Biosci. Biotechnol. Biochem.* **73**, 901–907.
- Kino, K., Nakazawa, Y. & Yagasaki, M. (2008). *Biochem. Biophys. Res. Commun.* **371**, 536–540.
- Kino, K., Noguchi, A., Nakazawa, Y. & Yagasaki, M. (2008). *J. Biosci. Bioeng.* **106**, 313–315.
- Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992). *Science*, **258**, 987–991.
- LeMaster, D. M. & Richards, F. M. (1985). *Biochemistry*, **24**, 7263–7268.
- Mills, S., Stanton, C., Hill, C. & Ross, R. P. (2011). *Annu. Rev. Food Sci. Technol.* **2**, 299–329.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Santos, S., Torcato, I. & Castanho, M. A. (2012). *Biopolymers*, **98**, 288–293.
- Sheldrick, G. M. (2010). *Acta Cryst. D* **66**, 479–485.
- Shomura, Y., Hinokuchi, E., Ikeda, H., Senoo, A., Takahashi, Y., Saito, J., Komori, H., Shibata, N., Yonetani, Y. & Higuchi, Y. (2012). *Protein Sci.* **21**, 707–716.
- Suzuki, M., Takahashi, Y., Noguchi, A., Arai, T., Yagasaki, M., Kino, K. & Saito, J. (2012). *Acta Cryst. D* **68**, 1535–1540.
- Tabata, K. & Hashimoto, S. (2007). *Appl. Environ. Microbiol.* **73**, 6378–6385.
- Tabata, K., Ikeda, H. & Hashimoto, S. (2005). *J. Bacteriol.* **187**, 5195–5202.
- Terwilliger, T. C. (2000). *Acta Cryst. D* **56**, 965–972.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst. D* **55**, 849–861.
- Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. (1991). *Science*, **254**, 1374–1377.
- Yagasaki, M. & Hashimoto, S. (2008). *Appl. Microbiol. Biotechnol.* **81**, 13–22.