

Doctorate Dissertation

博士論文

Structural analysis of self-defense mechanism-related proteins

(自己防御機構関連因子の構造解析)

A Dissertation Submitted for Degree of Doctor of Philosophy

December 2018

平成 30 年 12 月 博士 (理学) 申請

Department of Biological Sciences, Graduate School of Science,

The University of Tokyo

東京大学大学院理学系研究科生物科学専攻

Satoshi Fukuhara

福原 慧



## **Abstract**

Biological organisms generally have self-defense mechanisms to protect themselves. In order for organisms to maintain their own existence, mechanisms are required to eliminate endogenous and exogenous abnormalities. Structural studies of proteins related to biological defense mechanisms have brought about various developments such as understanding of structural basis of biological phenomena, elucidation of molecular basis of disease state, and application to drug discovery research. In this dissertation, the author determined the structures of self-defense mechanism-related proteins by X-ray crystallography.

### Crystal structure of Zucchini from *Drosophila melanogaster*

PIWI-interacting RNAs (piRNAs) bind PIWI proteins and silence transposons to maintain the genomic integrity of germ cells. Zucchini (Zuc), a phospholipase D superfamily member, is conserved among animals and is implicated in piRNA biogenesis. However, the underlying mechanism by which Zuc participates in piRNA biogenesis remains elusive. Here, the crystal structure of *Drosophila melanogaster* Zuc

(DmZuc) is presented at 1.75 Å resolution. The structure reveals that the dimer interface of DmZuc has a positively charged, narrow catalytic groove, which could interact with a single-stranded RNA. Functional analysis revealed that DmZuc and the mouse homologue MmZuc showed endoribonuclease activity for single-stranded RNAs in vitro. The RNA cleavage products bear a 5'-monophosphate group, a hallmark of mature piRNAs. Mutational analyses revealed that the conserved active-site residues of DmZuc are critical for the ribonuclease activity in vitro, and for piRNA maturation and transposon silencing in vivo. Structural and functional analyses revealed that Zuc endoribonuclease has a key role in piRNA maturation.

Crystal structure of Type VI Effector-Immunity complex from *Agrobacterium tumefaciens*

The type VI secretion system (T6SS) comprises needle-shaped multi-subunit complexes that play a role in the microbial defense systems of Gram-negative bacteria. Some Gram-negative bacteria harboring T6SS deliver the toxic effector proteins into the cytoplasm or periplasm of their competing bacteria to lyse and kill them. To avoid

self-cell disruption, these bacteria have the cognate immunity proteins that inhibit their toxic effector proteins. T6SS amidase effector protein 4 (Tae4) and T6SS amidase immunity protein 4 (Tai4) are members of the toxic effector-immunity pairs of T6SS. Here, the three-dimensional structures of Tai4 and the Tae4-Tai4 complex from *Agrobacterium tumefaciens* are reported at 1.55 Å and 1.9 Å resolutions, respectively. A structural comparison with other Tae4-Tai4 homologs revealed similarities and differences in the catalytic and inhibitory mechanisms among the Tae4 and Tai4 family proteins.

## Contents

Abstract.....	2
Contents.....	5
Table of Abbreviations.....	7
Chapter1: General Introduction.....	9
Chapter2: Crystal structure of Zucchini from <i>Drosophila melanogaster</i>	
2.1 Introduction.....	13
2.2 Materials and methods	
2.2.1 Structural analysis	
2.2.1.1 Construction.....	16
2.2.1.2 Expression.....	17
2.2.1.3 Purification.....	17
2.2.1.4 Crystallization.....	19
2.2.1.5 Data collection and crystallographic analysis.....	20
2.2.2 Functional analysis	
2.2.2.1 Nuclease assay.....	21
2.2.2.2 PLD assay.....	23
2.2.2.3 Rescue and dimerization assays in OSCs.....	25
2.2.2.4 RNA immunoprecipitation.....	26
2.3 Results	
2.3.1 Construction.....	28
2.3.2 Expression and purification.....	29
2.3.3 Crystallization, X-ray crystallographic analysis and structure determination..	32
2.3.4 Overall structure of DmZuc.....	38
2.3.5 Functional analysis.....	43
2.4 Discussion	
2.4.1 Structure comparison.....	51
2.4.2 Functional analysis.....	58
2.4.3 Comparison of DmZuc with other solved phospholipase D (PLD) family.....	60
2.5 Conclusion.....	63

Chapter3: Crystal structure of Type VI Effector-Immunity complex from *Agrobacterium tumefaciens*

3.1 Introduction.....	64
3.2 Materials and methods	
3.2.1 Macromolecule production.....	70
3.2.2 Crystallization.....	72
3.2.3 Data collection and processing.....	73
3.2.4 Structure determination.....	74
3.3 Results	
3.3.1 Structure determination.....	75
3.3.2 Overall structure of AtTai4 and the AtTae4-AtTai4 complex.....	83
3.3.3 Structure comparison.....	85
3.3.4 Catalytic site.....	91
3.4 Discussion	
3.4.1 Protein expression.....	92
3.4.2 Overall structure.....	92
3.4.3 Structure comparison.....	93
3.4.4 Catalytic site.....	94
3.5 Conclusion.....	97
Chapter4: General Discussion.....	98
References.....	102
Original papers.....	110
Acknowledgements.....	111

### Table of abbreviations

Ago3	Argonaute 3
<i>At</i>	<i>Agrobacterium tumefaciens</i>
Aub	Aubergine
CIP	calf intestinal phosphatase
CL	cardiolipin
<i>Dm</i>	<i>Drosophila melanogaster</i>
EGFP	enhanced green fluorescent protein
<i>Ec</i>	<i>Enterobacter cloacae</i>
IP	Immunoprecipitation
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
LB	Luria-Bertani
LC	liquid chromatography
MAD	Multi-wavelength anomalous diffraction
MLS	mitochondrial localization sequence
MRM	multiple reaction monitoring
MS/MS	tandem mass spectrometry
OSCs	ovarian somatic cells
PA	phosphatidic acid
PC	phosphatidylcholine
PDB	Protein Data Bank
PLD	phospholipase D
RNAi	RNA-mediated interference
RT-PCR	quantitative PCR with reverse transcription
SeMet	Selenomethionine
<i>Sm</i>	<i>Serratia marcescens</i>
<i>St</i>	<i>Salmonella typhimurium</i>
T6SS	type VI secretion system
TEV	tobacco etch virus
Tae	T6SS amidase effector protein

Tae4	T6SS amidase effector protein 4
Tai	T6SS amidase immunity protein
Tai4	T6SS amidase immunity protein 4
Tde	T6SS-related DNase effectors
Tdi	T6SS-related DNase immunity
Tge	T6SS-related glycoside hydrolase effectors
Tle	T6SS-related lipase effectors
Ter. Exo.	Terminator Exonuclease
Zuc	Zucchini
n.i.	non-immune IgG
nt	nucleotide
piRNAs	PIWI-interacting RNAs
r.m.s.d.	root-mean-square deviation
siRNA	short interfering RNA

## **Chapter1: General introduction**

Every organism has a biological defense mechanism to protect itself from endogenous or exogenous abnormalities (Beck *et al.*, 1996). When considering this biological defense mechanism from the viewpoint of the cell, which is the basic unit of life, it can roughly be divided into two categories: mechanism against abnormalities occurring inside of the cell and mechanism that responds to extra-cellular abnormalities.

One of the abnormalities arising from the inside of the cell is activation of transposon (McCLINTOCK, B., 1950). A transposon is a moving gene that exists in the animal genome. When the transposon is transferred, the genome is damaged, resulting in various diseases and infertility (Kazazian *et al.*, 1988; Miki *et al.*, 1992; Kazazian *et al.*, 2002; Belancio *et al.*, 2008; Sun *et al.*, 2018). Therefore, cells have various mechanisms to suppress transposon expression and transfer activity (Chung *et al.*, 2008). In germ cells, it is necessary to inherit genetic information accurately to the next generation. It is known that small RNA called piRNA plays a role in protecting the genome from damage by transposons (Malone & Hannon, 2009; Senti & Brennecke, 2010; Siomi *et al.*, 2011; Pillai & Chuma, 2012). However, in 2012, there were many

unclear points in the pathway of piRNA production.

In contrast, there are various types of extracellular abnormalities against cells.

For example, multicellular organisms such as mammals have various immune mechanisms to eliminate non-self from outside, for example adaptive immune system and innate immune system (Litman *et al.*, 2005; Pancer *et al.*, 2006). In unicellular organisms, the presence of extraneous substances and heterogenous unicellular organisms presents a serious threat to their survival (Hood *et al.*, 2010; Ma *et al.*, 2014).

When bacteria are exposed to survival competition with other bacteria that share a growing environment, the bacteria attack enemy bacteria, to secure an advantageous environment for them. In detail, bacteria cause virulence by secreting or injecting various proteins that disturb physiological functions against host bacteria (Hood *et al.*, 2010; Russell *et al.*, 2011; Russell *et al.*, 2012; Ma *et al.*, 2014). These proteins are synthesized in bacteria and transported outside the bacteria via secretion systems (Pukatzki *et al.*, 2007; Hood *et al.*, 2010; Russell *et al.*, 2011; Russell *et al.*, 2012).

*Pseudomonas aeruginosa*, which causes opportunistic infections, injects effectors using a type VI secretion systems (T6SS) and attacks enemy bacteria (Hood *et al.*, 2010). As

the effector has cell-wall digesting and lipid lysis activity, the bacteria attacked from *P. aeruginosa* are lysed and killed (Russell *et al.*, 2011). On the other hand, since *P. aeruginosa* carries immune proteins that interact with effectors, *P. aeruginosa* can prevent attacks by its own endogenous effectors (Russell *et al.*, 2011). Although many structures of effector-immunity complexes have been reported, there are still many unclear points in the interaction of effector-immunity complexes (Zhang, Gao, Wang *et al.*, 2013; Zhang, Gao, Wei *et al.*, 2013; Benz *et al.*, 2013; Srikannathasan *et al.*, 2013).

In this dissertation, the author focused on the protein Zucchini (Zuc), which is related to piRNA production and transposon suppression, and an effector-immunity complex, which is used for competition for survival among different bacteria. In chapter 2, the author determined the crystal structure of Zuc from *Drosophila melanogaster*. In addition, functional analysis based on the crystal structure showed that the Zuc has enzymatic activity to cleave single-stranded RNA, and gave the insight into the role in piRNA production pathway in germ cells. In chapter 3, the author determined the crystal structure of T6SS amidase effector protein 4 (Tae4) - T6SS amidase immunity protein 4 (Tai4) complex from *Agrobacterium tumefaciens*. The structural comparison with other

effector-immunity homologs revealed similarities and differences in the catalytic and inhibitory mechanisms among the Tae4 and Tai4 family proteins.

## **Chapter2: Crystal structure of Zucchini from *Drosophila melanogaster***

### **2.1 Introduction**

PIWI-interacting RNAs (piRNAs), which are germline-specific small noncoding RNAs, bind to PIWI Argonaute proteins and silence transposons in animal germ cells to maintain genomic integrity (Malone & Hannon, 2009; Senti & Brennecke, 2010; Siomi *et al.*, 2011; Pillai & Chuma, 2012). piRNAs are produced through two distinct biogenetic pathways: the primary processing pathway and the secondary ping-pong cycle (Vagin *et al.*, 2006; Gunawardane *et al.*, 2007; Brennecke *et al.*, 2007; Malone *et al.*, 2009; Li *et al.*, 2009; Saito *et al.*, 2009). *Drosophila melanogaster* has three PIWI proteins: Piwi, Aubergine (Aub) and Argonaute 3 (Ago3). Piwi is expressed in ovarian somatic cells and participates in the primary pathway, whereas Aub and Ago3 are expressed in ovarian germ cells and function in the ping-pong cycle. In the primary processing pathway, piRNAs are transcribed as long single-stranded precursors from intergenic repetitive elements known as piRNA clusters (Brennecke *et al.*, 2007; Malone *et al.*, 2009; Li *et al.*, 2009). Genetic and biochemical studies revealed that the putative nuclease Zucchini (Zuc), the RNA helicase Armitage and the Tudor

domain-containing RNA helicase Yb, are involved in the primary processing pathway (Olivieri *et al.*, 2010; Saito *et al.*, 2010; Haase *et al.*, 2010; Qi *et al.*, 2011).

Zuc is conserved among animals and is composed of an N-terminal mitochondrial localization sequence (MLS) and a catalytic domain belonging to the phospholipase D (PLD) superfamily (Pane *et al.*, 2007). *D. melanogaster* Zuc (DmZuc) and the mouse homologue MmZuc (also known as Pld6 and MitoPLD) are localized on the mitochondrial surface and have been implicated in piRNA biogenesis (Pane *et al.*, 2007; Malone *et al.*, 2009; Saito *et al.*, 2009, 2010; Olivieri *et al.*, 2010; Haase *et al.*, 2010; Huang *et al.*, 2011; Watanabe *et al.*, 2011). The PLD superfamily is characterized by a conserved HKD (His-Lys-Asp) motif and includes various proteins with diverse functions, such as PLD, cardiolipin synthases, phosphatidylserine synthases, nucleases, toxins and viral envelope proteins (Ponting & Kerr, 1996). Most of the superfamily members have two domains with a similar fold containing the HKD motif, whereas the bacterial nuclease Nuc has only one copy of the HKD motif and dimerizes to create the active site at the dimer interface (Stuckey & Dixon, 1999). Like Nuc, Zuc has one copy of the HKD motif, suggesting that Zuc functions as a dimer. MmZuc reportedly

hydrolyzes the mitochondrial lipid cardiolipin to generate the signalling lipid phosphatidic acid (Choi *et al.*, 2006), whereas Zuc has been considered to be a candidate for the piRNA-processing nuclease, since it shares the highest sequence similarity with Nuc among the superfamily members (Pane *et al.*, 2007). However, the mechanism by which Zuc participates in piRNA biogenesis remains elusive, since its enzymatic activity has not been fully explored. Here, the author reported the expression, purification, crystallization, X-ray crystallographic analysis, and functional analysis of DmZuc. The structural analysis, together with a functional analysis, will help in elucidating the molecular mechanism by which Zuc participates in piRNA biogenesis.

## 2.2 Materials and methods

### 2.2.1 Structural analysis

#### 2.2.1.1 Construction

The author constructed expression plasmids encoding residues 41–253, 89–250 and 89–253 of DmZuc (DmZuc<sub>41–253</sub>, DmZuc<sub>89–250</sub> and DmZuc<sub>89–253</sub>, respectively) and residues 35–221 and 74–221 of MmZuc (MmZuc<sub>35–221</sub> and MmZuc<sub>74–221</sub>, respectively). The DNA fragments encoding these regions were amplified by PCR from the cDNAs of DmZuc (UniProt accession code Q9VKD7) and MmZuc (UniProt accession code Q5SWZ9) using PrimeSTAR MAX DNA Polymerase (Takara Bio Inc.). The PCR products were cloned between the SacI and XhoI sites of the pCold-GST vector (Hayashi & Kojima, 2008) and thus all of the constructs contained the sequence **LEVLFQGP**GHMEL (the Turbo3C protease-recognition sequence is indicated in bold) between the His<sub>6</sub>-GST tag and the Zuc protein. The DmZuc<sub>41–253</sub> K171A mutant was generated by a PCR-based method using pCold-DmZuc<sub>41–253</sub> as the template. The insert sequences were verified by DNA sequencing.

### 2.2.1.2 Expression

DmZuc<sub>41–253</sub>, DmZuc<sub>89–253</sub> and MmZuc<sub>35–221</sub> were overexpressed in *Escherichia coli* Rosetta2 (DE3) cells (Novagen). The *E. coli* cells were cultured at 310 K in LB medium containing 100 g L<sup>-1</sup> ampicillin. When the culture reached an OD<sub>600</sub> of 0.5–0.8, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM. The cells were further cultured at 293 K for 24 h and were harvested by centrifugation. Selenomethionine (SeMet) labelled DmZuc<sub>89–250</sub> was overexpressed in *E. coli* B834 (DE3) cells in SeMet core medium (Wako) supplemented with 10 g L<sup>-1</sup> D-glucose, 1× MEM vitamin solution (Sigma), 0.25 g L<sup>-1</sup> MgSO<sub>4</sub>, 4 mg L<sup>-1</sup> FeSO<sub>4</sub>, 100 g L<sup>-1</sup> ampicillin and 25 mg L<sup>-1</sup> L-SeMet (Wako).

### 2.2.1.3 Purification

*E. coli* cells expressing DmZuc<sub>89–253</sub> were resuspended in buffer A (50 mM Tris–HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 3 mM β-mercaptoethanol), lysed by sonication and centrifuged at 40000×g for 30 min. The supernatant was incubated with 4 ml Ni–NTA Superflow resin (Qiagen) and the mixture was loaded into an Econo-Column (Bio-Rad). The resin was washed with buffer A and the protein was

eluted with buffer B (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole, 3 mM  $\beta$ -mercaptoethanol). The eluted protein was dialyzed against buffer A and treated with Turbo3C protease (Nacalai Tesque) at 277 K overnight in order to cleave the N-terminal His<sub>6</sub>-GST tag. The dialyzed protein was passed through a Ni-NTA column and was concentrated using an Amicon Ultra 3K filter (Millipore). The protein was further purified by chromatography on a HiLoad Superdex 75 16/60 column (GE Healthcare) equilibrated with buffer C (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT). The purified protein was concentrated using an Amicon Ultra 3K filter and was stored at 193 K until use. SeMet-labelled DmZuc<sub>89-250</sub> was purified using a protocol similar to that used for DmZuc<sub>89-253</sub>.

*E. coli* cells expressing wild-type (WT) DmZuc<sub>41-253</sub> were resuspended in buffer D (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 3 mM  $\beta$ -mercaptoethanol), lysed by sonication and centrifuged at 20000g for 30 min. The supernatant was incubated with 2 ml Glutathione Sepharose Fast Flow resin (GE Healthcare) and the mixture was loaded into an Econo-Column. The resin was washed with buffer D and the protein was eluted with buffer D supplemented with 10 mM glutathione. The eluted protein was treated

with Turbo3C protease at 277 K overnight to cleave the N-terminal His<sub>6</sub>-GST tag and was then passed through an Ni-NTA column. The protein was diluted with a fivefold volume of buffer E (20 mM Tris-HCl pH 8.0, 1 mM DTT) to reduce the NaCl concentration to 50 mM and was loaded onto a Resource S column (GE Healthcare) equilibrated with buffer F (20 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM DTT). The column was washed with buffer F and the protein was eluted with a linear gradient of 20–600 mM NaCl in buffer F. The protein was concentrated using an Amicon Ultra 10K filter (Millipore) and was further purified by chromatography on a Superdex 200 10/300 column (GE Healthcare). The purified protein was concentrated using an Amicon Ultra 10K filter and was stored at 193 K until use. MmZuc<sub>35–221</sub> and the DmZuc<sub>41–253</sub> K171A mutant were purified using a protocol similar to that used for WT DmZuc<sub>41–253</sub>. All of the purified proteins were stored in buffer C at 193 K until use. The Superdex 200 10/300 column was calibrated using a Gel Filtration Calibration Kit LMW (GE Healthcare).

#### **2.2.1.4 Crystallization**

Initial crystallization screening was performed at 293 K by the sitting-drop

vapour-diffusion method in a 96-well MRC Crystallization Plate (Hampton Research) using screening kits including Crystal Screen, Crystal Screen 2, PEG/Ion, Natrix, Index (Hampton Research), PACT, JCSG+ (Qiagen), MemGold (Molecular Dimensions), Wizard (Emerald BioSystems), JBScreen Classic and JBScreen Membrane (Jena Bioscience). Crystallization drops were prepared by mixing 100 nl protein solution and 100 nl reservoir solution using a Mosquito crystallization robot (TTP LabTech). Initial crystallization conditions were optimized at 293 K by varying the volumes of the drops, the protein concentration, the pH and the type and concentration of PEG using the sitting-drop or hanging-drop-vapor-diffusion methods. Crystallization conditions were also optimized with the Additive Screen kit (Hampton Research).

#### **2.2.1.5 Data collection and crystallographic analysis**

Crystals were cryoprotected in reservoir solution supplemented with 30% ethylene glycol and flash-cooled in a nitrogen-gas stream at 100 K. X-ray diffraction experiments were performed on BL32XU at SPring-8, Hyogo, Japan using an MX225HE detector. X-ray diffraction data were collected using a helical data-collection strategy involving a microbeam ( $1 \times 15 \mu\text{m}$ ) and were processed using HKL-2000

(HKL Research Inc.). The structure of DmZuc<sub>89-250</sub> was determined by the MAD method, using the 2.2-Å resolution data from the SeMet-labelled crystal. Four Se atoms were located using SHELXD, and the initial phases were calculated using SHARP, followed by automated model building using RESOLVE. The model was further built manually using COOT and refined using PHENIX. The structures of DmZuc<sub>41-253</sub> and DmZuc<sub>41-253</sub> K171A were solved by molecular replacement with MOLREP, using the structure of DmZuc<sub>89-250</sub> as a search model. Ramachandran plot analysis was performed using RAMPAGE (Lovell *et al.*, 2003).

### **2.2.2 Functional analysis**

In the functional analyses (2.2.2.1 to 2.2.2.4), the author expressed and purified recombinant DmZuc or MmZuc proteins. Dr. Hirotsugu Ishizu, Dr. Kuniaki Saito, and Ms. Miharu K. Kamatani conducted biochemical and biological assays. Mr. Keita Nakanaga and Dr. Junken Aoki conducted PLD assay. All protocols of functional analyses were quoted from Nishimasu *et al.*, 2012.

#### **2.2.2.1 Nuclease assay**

The RNA substrates used for nuclease assays were: 5'-AUUUAUCAAGCU

UAUCGAUACCGUCGACCUCGAGGGGGGGC-3' (42-nt), 5'-GGUCUGAUUUCG  
AUCUGGUUCCCUGGAACAAAAGUGGCAG-3' (40-nt sense strand), 5'-CUGCCA  
CUUUUGUUCCAGGGAACCAGAUUCGAAAUCAGACC-3' (40-nt antisense strand),  
and poly(U) (40-nt). ssRNAs were synthesized (Sigma) and labelled at their 5' ends  
using [ $\gamma$ - $^{32}$ P] ATP and T4 polynucleotide kinase (New England Biolabs). For 3' end  
labelling, RNAs were labelled at their 3' ends using 5'-[ $^{32}$ P]pCp and T4 RNA ligase 1  
(New England Biolabs). To make circular ssRNA, 5' end-labelled 42-nt RNAs were  
circularized by intramolecular ligation using T4 RNA ligase 1. The labelled RNAs were  
purified by electrophoresis on 12% denaturing polyacrylamide gels. To make dsRNA, 5'  
end-labelled 40-nt sense strand RNA and non-labelled 40-nt antisense strand RNA were  
mixed at a molar ratio of 1:2 (sense: antisense) in annealing buffer (10 mM Tris-HCl  
pH 7.5, 50 mM NaCl, 1 mM EDTA), heated at 95 °C for 15 min, and then cooled slowly  
to 30 °C.

$^{32}$ P-end-labelled RNA substrate ( $10^4$  c.p.m.) was incubated with purified  
recombinant DmZuc (1.62  $\mu$ M) in buffer A (25 mM HEPES-KOH pH 7.4, 2.5 mM  
EDTA, 5 mM DTT) for 1 h at 26 °C (37 °C for MmZuc). The reaction was terminated

by treatment with proteinase K, followed by extraction with phenol–chloroform and precipitation with ethanol. The products were resolved on 20% denaturing polyacrylamide gels. The effects of NaCl on the nuclease activity were tested using the 42-nt ssRNA substrate in buffer B (25 mM HEPES-KOH pH 7.4, 5 mM magnesium acetate, 5 mM DTT, 0–100 mM NaCl). Terminator 5'-phosphate-dependent exonuclease was used in accordance with the manufacturer's instructions (Epicentre). The DNA substrate used for ssDNase assays was 5'-AATTGGTACGTAGCGCTTGATCTAGAGGGTTTGCCAATAGCAATCCGCACGTTTCCTCGTGCTCGACAAT-3'.

#### **2.2.2.2 PLD assay**

PLD from *Actinomadura* sp. no. 362 was purchased from Meito Sangyo Inc. (Tokyo Japan). Tetraoleoyl cardiolipin (CL) (0.1 µmol; Avanti Polar Lipids) and dipalmitoyl phosphatidylcholine (PC) (0.5 µmol; Avanti Polar Lipids) were dried under nitrogen gas and resuspended in 10 mM HEPES pH 7.5 (0.6 ml) by sonication. The enzyme reaction was performed at 26 °C (DmZuc) or 37 °C (MmZuc and *Actinomadura* PLD) for 2 h with 10 µl of lipid solution, 38 µl of assay buffer (50 mM HEPES pH 7.5, 80 mM KCl, 3 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 1 mM DTT) and purified recombinant DmZuc

(0.4  $\mu\text{g}$ ), MmZuc (0.4  $\mu\text{g}$ ) or *Actinomadura* PLD (0.3 U). The reaction was stopped by adding 450  $\mu\text{l}$  of methanol containing an internal standard (1  $\mu\text{M}$  diheptadecanoyl PC) and centrifuged at 21,500g. The sample solution (5  $\mu\text{l}$ ) was injected into liquid chromatography (LC) and analysed by tandem mass spectrometry (MS/MS). In brief, lipids were separated by Nanospace LC (Shiseido) with a silica column (5  $\mu\text{m}$ , 2.0 mm  $\times$  150 mm; Shiseido), using a gradient of solvent A (5 mM ammonium formate in water) and solvent B (acetonitrile), and then analysed by MS/MS using a Quantum Ultra triple quadrupole mass spectrometer (Thermo Fisher Scientific). Lipids were monitored in negative-ion mode and quantified by multiple reaction monitoring (MRM). Tetraoleoyl CL, dioleoyl phosphatidic acid (PA), dipalmitoyl PA and dipalmitoyl PC were monitored with the MRM transition of  $m/z$  1,456.1 > 281.2,  $m/z$  699.4 > 281.2,  $m/z$  759.5 > 255.2 and  $m/z$  778.6 > 255.2, respectively. The ratio between analyte and internal standard peak area was used for quantification. Calibration curves (0.5–5,000 ng) and LC retention times for each compound were established using synthetic standards.

### 2.2.2.3 Rescue and dimerization assays in OSCs (ovarian somatic cells)

The RNAi-resistant WT and mutants of DmZuc were prepared by a PCR-based method, using a pAc-Zuc-Myc vector (Saito *et al.*, 2010) as the template. Transfection was performed essentially as described previously (Saito *et al.*, 2009). In brief, trypsinized OSCs ( $5 \times 10^6$  cells) were transfected with siRNA duplex (200 pmol), transferred to fresh OSC medium and incubated at 26 °C. Two days after transfection, OSCs ( $3 \times 10^6$  cells) were again transfected with siRNA duplex (200 pmol) and plasmid vector (5 µg). The cells were again incubated at 26 °C for 2 days, and then total RNAs were purified using ISOGEN reagent (Nippon Gene). Total RNA (0.5 µg) was used to reverse transcribe target sequences using oligo (dT) primers. The resulting cDNAs were analysed by quantitative RT-PCR using a LightCycler real-time PCR system (Roche Diagnostics) and SYBR Premix Ex Taq (Takara). Relative steady-state mRNA levels were determined from the threshold cycle for amplification. Ribosomal protein 49 was used as an internal control. The expression levels of Piwi and DmZuc were analysed by western blotting using culture supernatants of anti-Piwi hybridoma cells (P4D2; 1:1 dilution) (Saito *et al.*, 2006) and a mouse monoclonal antibody 9E10 against the Myc

tag (1:1,000 dilution; Sigma), respectively.

The C-terminal Flag-tagged DmZuc (pAc-Zuc-Flag) was prepared, by inserting oligonucleotides encoding a 3 × Flag tag between the *XhoI* and *BamHI* sites of pAc-Zuc-Myc (Saito *et al.*, 2010). OSCs ( $5 \times 10^6$  cells) were co-transfected with pAc-Zuc-Flag and either pAc-Zuc-Myc or pAcM-EGFP. Two days after transfection, the OSCs were homogenized in lysis buffer (30 mM HEPES pH 7.3, 150 mM potassium acetate, 2 mM magnesium acetate, 5 mM DTT, 0.1% Nonidet P40) to prepare OSC lysate. Zuc-Flag was immunopurified using an anti-Flag M2 antibody (Sigma) immobilized on GammaBind beads (GE Healthcare). The mixtures were rocked at 4 °C for 2 h and the beads were washed five times with lysis buffer. After immunoprecipitation, proteins were resolved by SDS-PAGE and then detected by western blotting using anti-Flag M2 and anti-Myc antibodies.

#### **2.2.2.4 RNA immunoprecipitation**

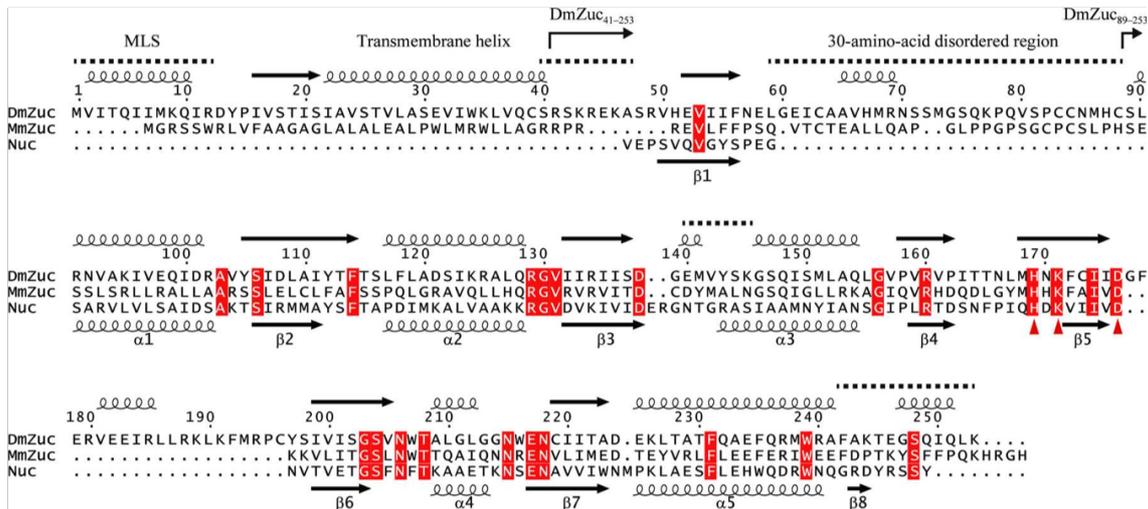
The C-terminal Myc-tagged, full-length DmZuc was expressed in OSCs by transfection. Two days after transfection, the OSCs were homogenized in lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM DTT,  $2 \mu\text{g ml}^{-1}$  pepstatin,

2  $\mu\text{g ml}^{-1}$  leupeptin, 0.5% aprotinin, 20 U  $\mu\text{l}^{-1}$  RNasin Plus (Promega), 1% Triton X-100) to prepare OSC lysate. Zuc-Myc was immunopurified using a monoclonal anti-Myc antibody (9E10) immobilized on Dynabeads Protein G (Invitrogen). The mixtures were rocked at 4 °C for 2 h and the beads were washed five times with lysis buffer. Total RNAs were isolated from the immunoprecipitates with phenol–chloroform and were precipitated with ethanol. Total RNA was reverse transcribed using SuperScript III reverse transcriptase (Invitrogen) and a strand-specific RT primer for a fragment of the *flamenco* transcript. PCR was performed using Ex *Taq* DNA polymerase (Takara) and primers for a fragment of the *flamenco* transcript. The primer sequences were as follows: *flamenco* forward, 5'-AACGTATGCTCAGTCCAGTGAA-3'; *flamenco* RT and reverse, 5'-AAAACCTTCTAGCTTGCCCTCT-3'.

## 2.3 Results

### 2.3.1 Construction

Secondary-structure prediction using the Phyre2 server (Kelley & Sternberg, 2009) suggested that DmZuc and MmZuc have an N-terminal MLS and a transmembrane helix followed by a  $\beta$ -strand (residues 52–56 of DmZuc and residues 38–42 of MmZuc), which corresponds to the first  $\beta$ -strand of Nuc (Fig. 1-1). In contrast to Nuc, DmZuc and MmZuc are predicted to have a disordered region (residues 59–88 of DmZuc and residues 57–72 of MmZuc) between the  $\beta$ 1 strand and the  $\alpha$ 2 helix in the PLD core fold. Therefore, the author prepared DmZuc<sub>41–253</sub> and MmZuc<sub>35–221</sub>, which contain the first  $\beta$ -strand and the disordered region, as well as DmZuc<sub>89–253</sub> and MmZuc<sub>74–221</sub>, which lack these regions.



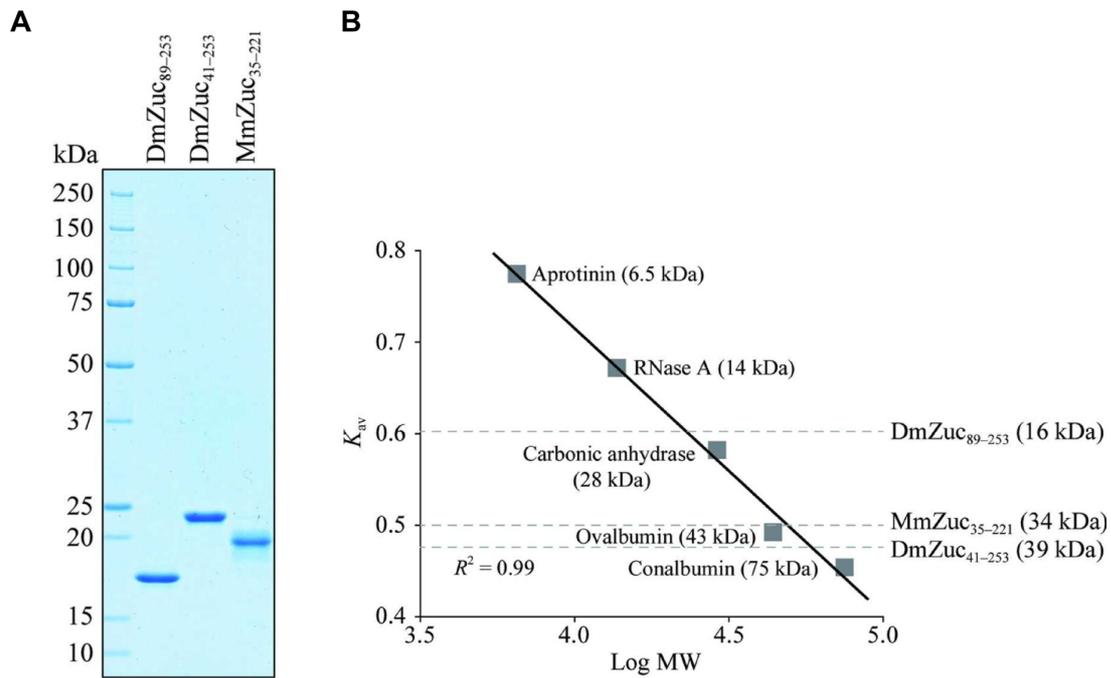
**Figure 1-1. Multiple sequence alignment of DmZuc, MmZuc and Nuc.**

The predicted secondary structure of DmZuc is shown above the sequences and the predicted disordered regions are indicated by dashed lines. The secondary structure of Nuc (PDB entry 1byr) is shown below the sequences. The HKD motif is indicated by red triangles. The alignment was prepared using ClustalW (Larkin *et al.*, 2007) and ESPript (Gouet *et al.*, 1999).

### 2.3.2 Expression and purification

DmZuc<sub>89-253</sub>, DmZuc<sub>41-253</sub> and MmZuc<sub>35-221</sub> were expressed in *E. coli* and purified to homogeneity (Fig. 1-2 A), whereas MmZuc<sub>74-221</sub> precipitated during purification. DmZuc<sub>89-253</sub> (19.4 kDa), DmZuc<sub>41-253</sub> (24.7 kDa) and MmZuc<sub>35-221</sub> (21.8

kDa) eluted from the size-exclusion column at volumes corresponding to molecular weights of 16, 39 and 34 kDa, respectively (Fig. 1-2B). These results suggested that DmZuc<sub>89-253</sub> may be monomeric in solution and DmZuc<sub>41-253</sub> and MmZuc<sub>35-221</sub> may be dimeric in solution, which is consistent with the prediction that DmZuc<sub>89-253</sub> lacks the first  $\beta$ -strand that contributes to dimerization in Nuc (Stuckey & Dixon, 1999). The expression level of DmZuc<sub>41-253</sub> was much lower (about 0.02 mg protein was purified from 1 l culture) than that of DmZuc<sub>89-253</sub> (about 2 mg protein was purified from 1 l culture), making it difficult to perform extensive crystallization screening. During the preparation of several DmZuc<sub>41-253</sub> mutants for functional analyses, we fortuitously found that the DmZuc<sub>41-253</sub> K171A mutant, in which Lys171 in the HKD motif is replaced by alanine, was highly expressed in *E. coli*. The DmZuc<sub>41-253</sub> K171A mutant was also purified to homogeneity (about 1 mg protein was purified from 1 l culture).



**Figure 1-2. Preparation of DmZuc and MmZuc.**

(A) SDS-PAGE analysis of DmZuc<sub>89-253</sub>, DmZuc<sub>41-253</sub> and MmZuc<sub>35-221</sub>. The purified proteins were analyzed by 10–20% SDS-PAGE and stained with SimplyBlue SafeStain (Invitrogen). (B) Calibration curve of the Superdex 200 10/300 column. The column was calibrated with a Gel Filtration Calibration Kit LMW (GE Healthcare).

### 2.3.3 Crystallization, X-ray crystallographic analysis and structure determination

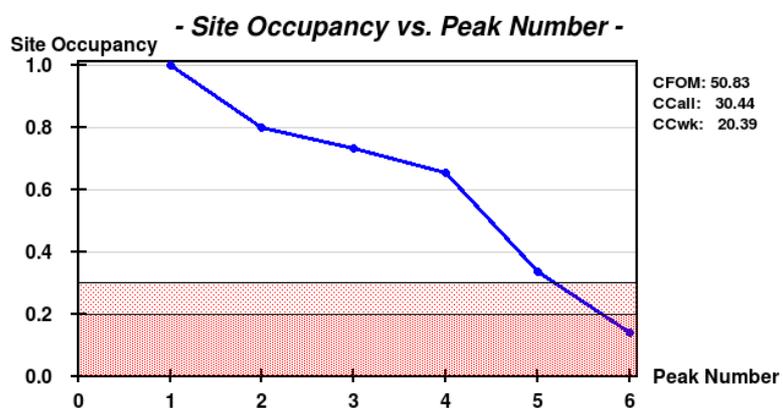
Crystals of DmZuc<sub>89–253</sub> were obtained by mixing 0.1  $\mu\text{l}$  protein solution (6 mg  $\text{ml}^{-1}$ ) with 0.1  $\mu\text{l}$  reservoir solution [condition No. A2 of JBScreen Classic 1; 15%(v/v) PEG 400, 100 mM MES pH 6.5; Fig. 1-3A]. The DmZuc<sub>89–253</sub> crystal diffracted to 2.5  $\text{\AA}$  resolution and belonged to space group  $P3_x21$  (where  $x$  may be 0, 1 or 2), with unit-cell parameters  $a = 60.4 \text{ \AA}$ ,  $b = 60.4 \text{ \AA}$ ,  $c = 254.1 \text{ \AA}$ . Assuming two protein molecules per asymmetric unit, the Matthews coefficient ( $V_M$ ) was estimated to be  $3.5 \text{ \AA}^3 \text{ Da}^{-1}$ , corresponding to a solvent content of 64%. We were unable to determine the structure by molecular replacement using the Nuc structure (PDB entry 1byr; Stuckey & Dixon, 1999) as a search model, even though DmZuc shares 26% sequence identity with Nuc. Neither native nor SeMet-labelled DmZuc<sub>89–253</sub> crystallized reproducibly, possibly owing to the presence of a C-terminal disordered region (residues 242–253; Fig. 1-1). The surface entropy reduction technique was unable to produce diffraction-quality crystals. Therefore, we prepared a series of constructs lacking C-terminal residues, performed crystallization screening and obtained initial crystals of SeMet-labelled DmZuc<sub>89–250</sub> under condition No. A9 of Natrix [10%(v/v) PEG 400, 50

mM MES pH 6.0, 100 mM KCl, 10 mM MgCl<sub>2</sub>]. We obtained diffraction-quality crystals of SeMet-labelled DmZuc<sub>89-250</sub> by mixing 0.4 μl protein solution (2.5 mg ml<sup>-1</sup>) with 0.4 μl reservoir solution consisting of 8%(v/v) PEG 400, 50 mM MES pH 6.0, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 3%(w/v) D -(+)-trehalose (Fig. 1-3 B). The SeMet-labelled crystal diffracted to 2.2 Å resolution and belonged to space group *P*2<sub>1</sub>, with unit-cell parameters  $a = 35.3 \text{ \AA}$ ,  $b = 52.4 \text{ \AA}$ ,  $c = 38.8 \text{ \AA}$ ,  $\beta = 104.9^\circ$ . The structure of DmZuc<sub>89-250</sub> was determined by the MAD method. Four Se atoms were located using SHELXD, and the initial phases were calculated using SHARP, followed by automated model building using RESOLVE. The model was further built manually using COOT and refined using PHENIX (Fig. 1-4, Fig. 1-5, Table 1-1).

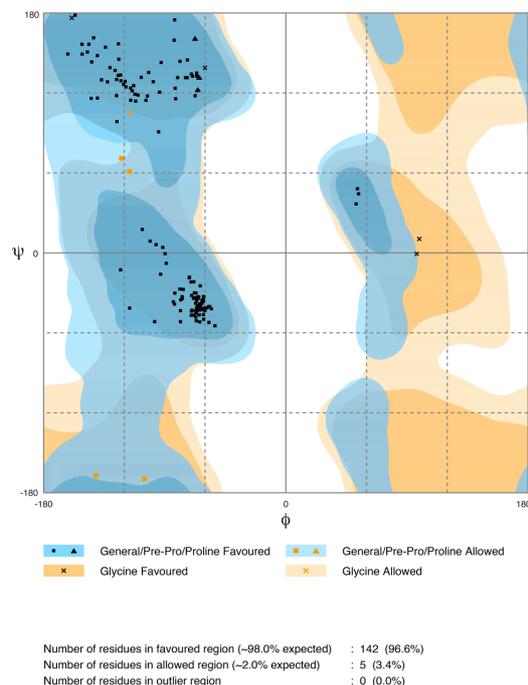


**Figure 1-3. Crystals of DmZuc.**

(A) Crystals of DmZuc<sub>89-253</sub>. (B) Crystals of SeMet-labelled DmZuc<sub>89-250</sub>.



**Figure 1-4. Search result of Se atom by SHELXD.**

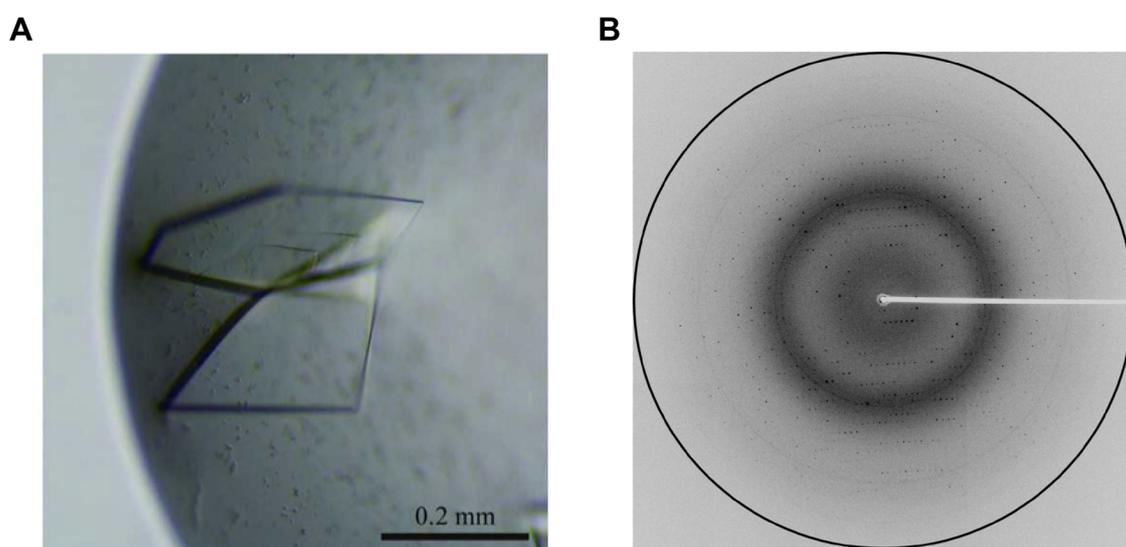


**Figure 1-5. Ramachandran analysis of the DmZuc<sub>89-250</sub>.**

Crystals of DmZuc<sub>41-25</sub> K171A were obtained with condition No. 42 of PEG/Ion screen [20%(w/v) PEG 3350, 200 mM KH<sub>2</sub>PO<sub>4</sub>]. We obtained plate-shaped crystals by mixing 1  $\mu$ l protein solution (3 mg ml<sup>-1</sup>) with 1  $\mu$ l reservoir solution consisting of 16%(w/v) PEG 3350, 200 mM KH<sub>2</sub>PO<sub>4</sub>. The crystal of DmZuc<sub>41-253</sub> K171A diffracted to 2.2 Å resolution and belonged to space group *P*2<sub>1</sub>, with unit-cell parameters  $a = 55.9$ ,  $b = 70.2$ ,  $c = 56.9$  Å,  $\beta = 108.8^\circ$ . The structures of DmZuc<sub>41-253</sub> K171A was solved by molecular replacement with MOLREP, using the structure of DmZuc<sub>89-250</sub> as a search model.

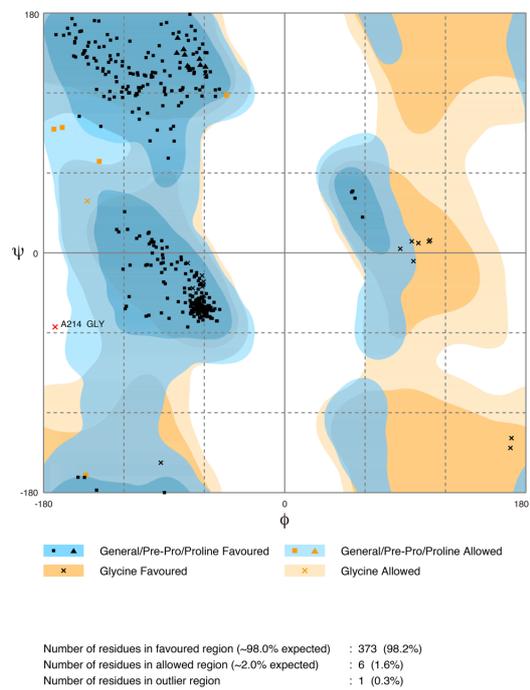
Crystals of WT DmZuc<sub>41-253</sub> were obtained under conditions similar to those

used for the K171A mutant (Fig. 1-6 A). The crystal of WT DmZuc<sub>41-253</sub> diffracted to 1.75 Å resolution (Fig. 1-6 B) and belonged to space group  $P2_1$ , with unit-cell parameters  $a = 55.0$  Å,  $b = 71.2$  Å,  $c = 56.3$  Å,  $\beta = 107.9^\circ$ . Assuming the presence of two protein molecules per asymmetric unit, the  $V_M$  was estimated to be  $2.1$  Å<sup>3</sup> Da<sup>-1</sup>, with a solvent content of 42%. The structures of WT DmZuc<sub>41-253</sub> was solved by molecular replacement with MOLREP, using the structure of DmZuc<sub>89-250</sub> as a search model (Fig. 1-7, Table 1-1). MmZuc<sub>35-221</sub> failed to crystallize, although it eluted as a single monodisperse peak from the size-exclusion column.



**Figure 1-6. Crystals and X-ray diffraction image of the WT DmZuc<sub>41-253</sub>.**

(A) Crystal of the WT DmZuc<sub>41-253</sub>. (B) X-ray diffraction image of the WT DmZuc<sub>41-253</sub>. The ring indicates 1.75 Å resolution.



**Figure 1-7. Ramachandran analysis of the WT DmZuc<sub>41-253</sub>.**

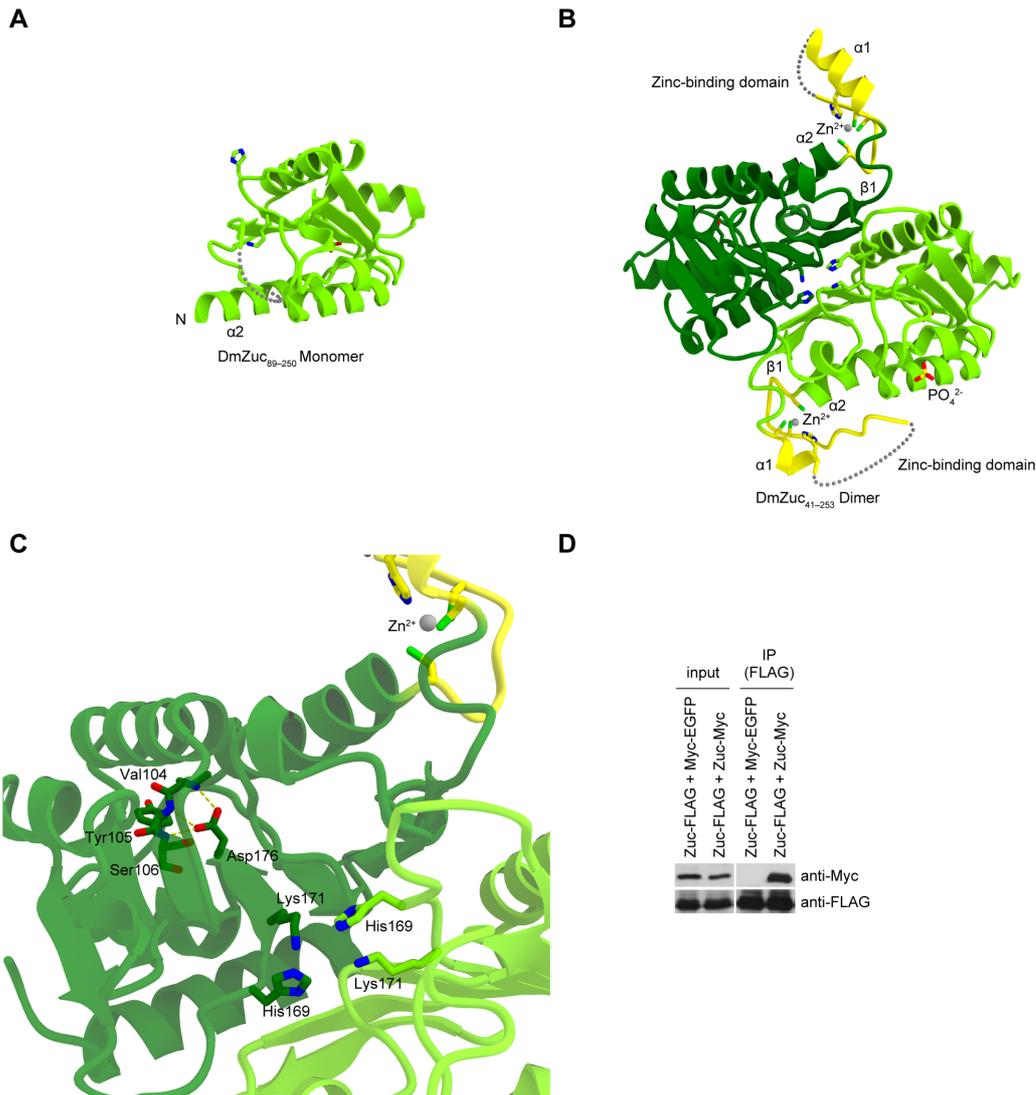
**Table 1-1 Data collection and refinement statistics**

	DmZuc <sub>41-253</sub>	DmZuc <sub>41-253</sub> K171A	SeMet DmZuc <sub>89-250</sub>	
<b>Data collection</b>				
Space group	<i>P</i> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub>	
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	55.0, 71.2, 56.3	55.9, 70.2, 56.9	35.3, 52.4, 38.8	
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 107.9, 90	90, 108.8, 90	90, 104.9, 90	
Wavelength (Å)	1.000	1.000	<i>Peak</i>	<i>Inflection</i>
Resolution (Å)	50.00–1.75 (1.78–1.75)	50.00–2.2 (2.24–2.20)	0.9793	0.9800
<i>R</i> <sub>sym</sub>	0.064 (0.34)	0.083 (0.27)	50.00–2.2 (2.24–2.20)	50.00–2.5 (2.54–2.50)
<i>I</i> / $\sigma$ ( <i>I</i> )	27.5 (2.2)	16.6 (2.6)	0.14 (0.93)	0.13 (0.90)
Completeness (%)	97.2 (91.2)	95.0 (87.7)	22.9 (2.7)	24.5 (3.1)
Redundancy	5.8 (3.7)	2.8 (2.2)	100 (100)	100 (100)
			8.4 (8.2)	8.3 (8.3)
<b>Refinement</b>				
Resolution (Å)	42.8–1.75	45.8–2.2	37.5–2.2	
No. reflections	39,874	19,674	6,836	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.187 / 0.230	0.190 / 0.246	0.223 / 0.270	
No. atoms				
Protein	3,080	2,986	1,173	
Ligand/ion	31	10	1	
Solvent	180	96	16	
<i>B</i> -factors (Å <sup>2</sup> )				
Protein	28.4	40.4	39.0	
Ligand	33.6	35.3	24.0	
Solvent	34.3	38	30.4	
R.m.s. deviations				
Bond lengths (Å)	0.008	0.005	0.002	
Bond angles (°)	1.09	0.95	0.57	
Ramachandran plot				
Favored (%)	98.2	97.3	96.6	
Allowed (%)	1.6	2.7	3.4	
Outlier (%)	0.3	0.0	0.0	

### 2.3.4 Overall structure of DmZuc

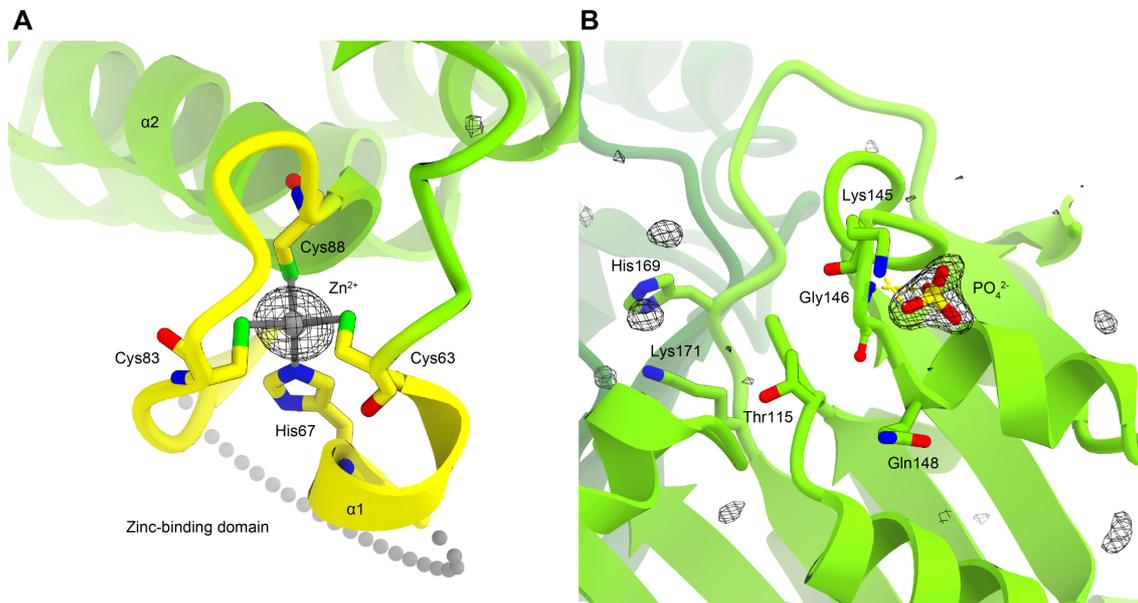
While the crystal structure of a DmZuc<sub>89-250</sub>, lacking the  $\beta$ 1 strand, was solved as a monomer (Fig. 1-8 A), WT DmZuc<sub>41-253</sub> was solved as a dimer (Fig. 1-8 B).

The structure consists of a catalytic domain (residues 49–62 and 89–245) and a zinc-binding domain (residues 63–88), with the two protomers in the asymmetric unit forming a dimer (Fig. 1-8 B). His 169 and Lys 171 in the HKD motif in the two protomers form an active site at the dimer interface, and Asp 176 in the motif helps to maintain the structural integrity of each protomer (Fig. 1-8 C) Co-immunoprecipitation experiments suggested self-interaction of DmZuc *in vivo* (Fig. 1-8 D). The crystal structure revealed that a zinc ion is tetrahedrally coordinated by Cys 63, His 67, Cys 83 and Cys 88 in the zinc-binding domain (Fig. 1-9 A). A phosphate ion derived from the crystallization buffer binds to a positively charged patch near the active-site groove (Fig. 1-9 B) The active-site groove of DmZuc is positively charged (Fig. 1-10).



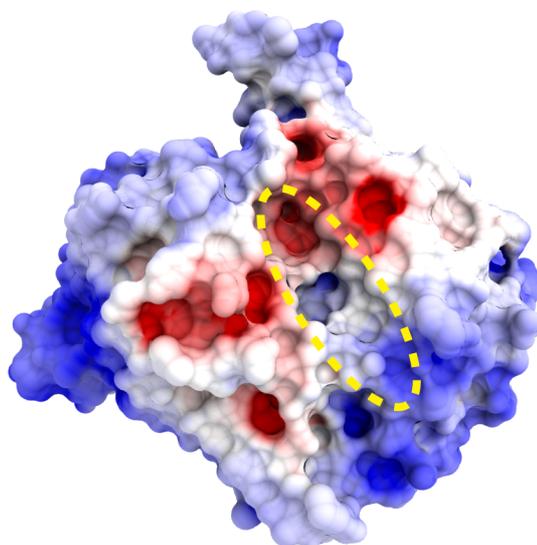
**Figure 1-8. Crystal structures of DmZuc.**

(A) Overall structure of a DmZuc<sub>89-250</sub>. (B) Overall structure of a DmZuc<sub>41-253</sub>. The two protomers are colored green and yellow-green, and the zinc-binding domains are colored yellow. The bound zinc ions are shown as grey spheres. Cys 63, His 67, Cys 83, Cys 88, His 169, Lys 171 and the bound phosphate ion are shown as sticks. Disordered regions are shown as dashed lines. (C) Asp176 in the HKD motif helps to maintain structural integrity. Dashed lines indicate hydrogen bonds between the side chain of Asp176 and the main-chain amide groups of Val104, Tyr105 and Ser106. (D) Self-interaction of DmZuc in OSCs. OSCs were co-transfected with an expression vector encoding C-terminal FLAG-tagged DmZuc and either an expression vector encoding N-terminal Myc-tagged enhanced green fluorescent protein (EGFP) or an expression vector encoding C-terminal Myc-tagged DmZuc. Immunoprecipitation (IP) using an anti-FLAG M2 antibody, followed by western blotting using anti-Myc and anti-FLAG antibodies, revealed that DmZuc-FLAG interacts with DmZuc-Myc but not Myc-EGFP.



**Figure 1-9. The zinc-binding domain and phosphate-binding site.**

(A) The zinc-binding domain. The bound zinc ion is shown as a grey sphere, and a simulated annealing  $F_o - F_c$  omit electron density map (contoured at  $4\sigma$ ) is shown as a black mesh. (B) Phosphate-binding site. Hydrogen bonds are indicated by yellow dashed lines, and simulated annealing  $F_o - F_c$  omit electron density map is shown as a black mesh (contoured at  $4\sigma$ ). The phosphate ion interacts with the side chain of Lys145 and the main-chain amide group of Gly146, and is also surrounded by Thr115 and Gln148. Thr115, Gly146, and Gln148 are conserved in Zuc proteins.



**Figure 1-10. Active site groove of DmZuc.**

Electrostatic surface potential of DmZuc. The active sites are indicated by yellow dashed lines.

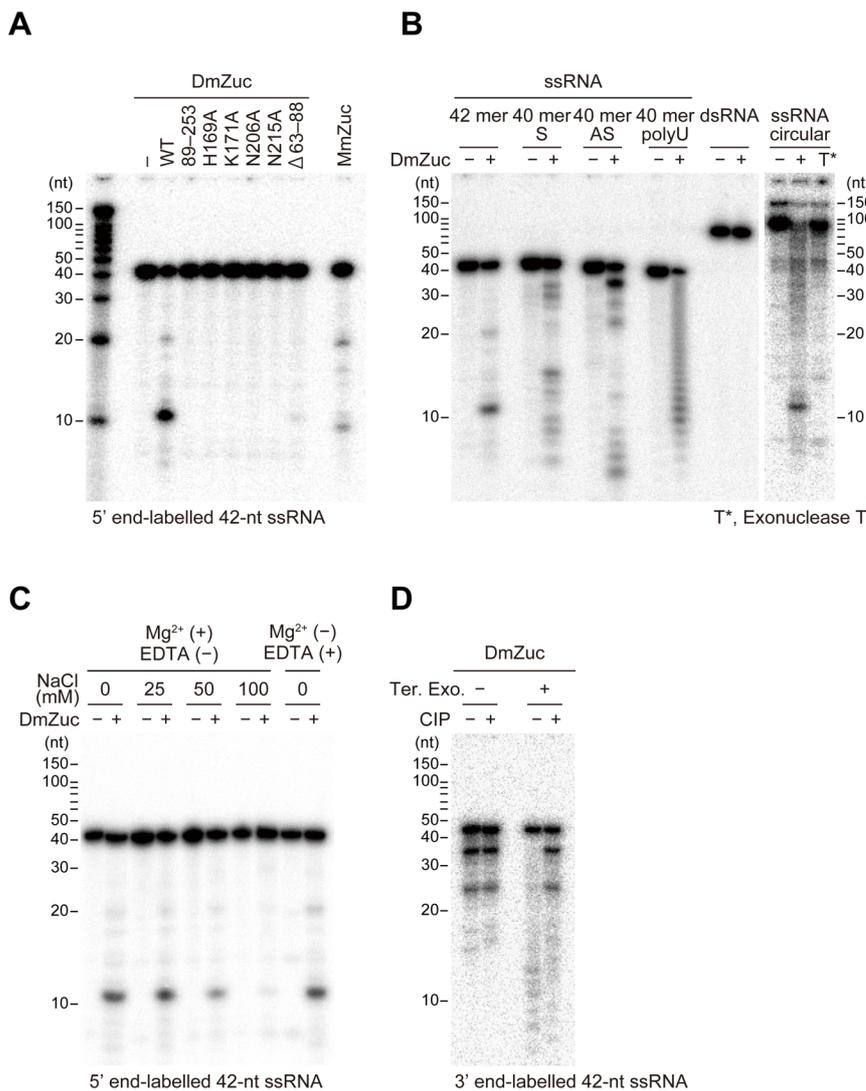
### 2.3.5 Functional analysis

In the functional analyses (2.2.2.1 to 2.2.2.4), the author expressed and purified recombinant DmZuc or MmZuc proteins. Dr. Hirotsugu Ishizu, Dr. Kuniaki Saito, and Ms. Miharu K. Kamatani conducted biochemical and biological assays. Mr. Keita Nakanaga and Dr. Junken Aoki conducted PLD assay. All results of functional analyses were mainly quoted from Nishimasu *et al.*, 2012.

Purified DmZuc<sub>41-253</sub> was incubated with a 42-nucleotide (nt) single-stranded RNA (ssRNA) labelled with <sup>32</sup>P at the 5' end. DmZuc cleaved the ssRNA substrate (Fig. 1-11 A). In contrast, the DmZuc mutants H169A, K171A, N206A and N215A failed to cleave the substrate (Fig. 1-11 A). A DmZuc  $\Delta$ 63-88 mutant, lacking the zinc-binding domain, showed decreased ssRNase activity (Fig. 1-11 A). The DmZuc<sub>89-253</sub> failed to cleave the ssRNA substrate (Fig. 1-11 A). Purified MmZuc<sub>35-221</sub> also showed a ssRNase activity (Fig. 1-11 A). DmZuc cleaved poly (U) and circular ssRNA (Fig. 1-11 B), and also ssDNA (Fig. 1-12 A), but not double-stranded RNA (dsRNA) (Fig. 1-11 B). DmZuc showed similar activity in the presence of magnesium ions or EDTA (Fig. 1-11 C). The DmZuc ssRNase activity was inhibited by NaCl at concentrations of 50 mM or

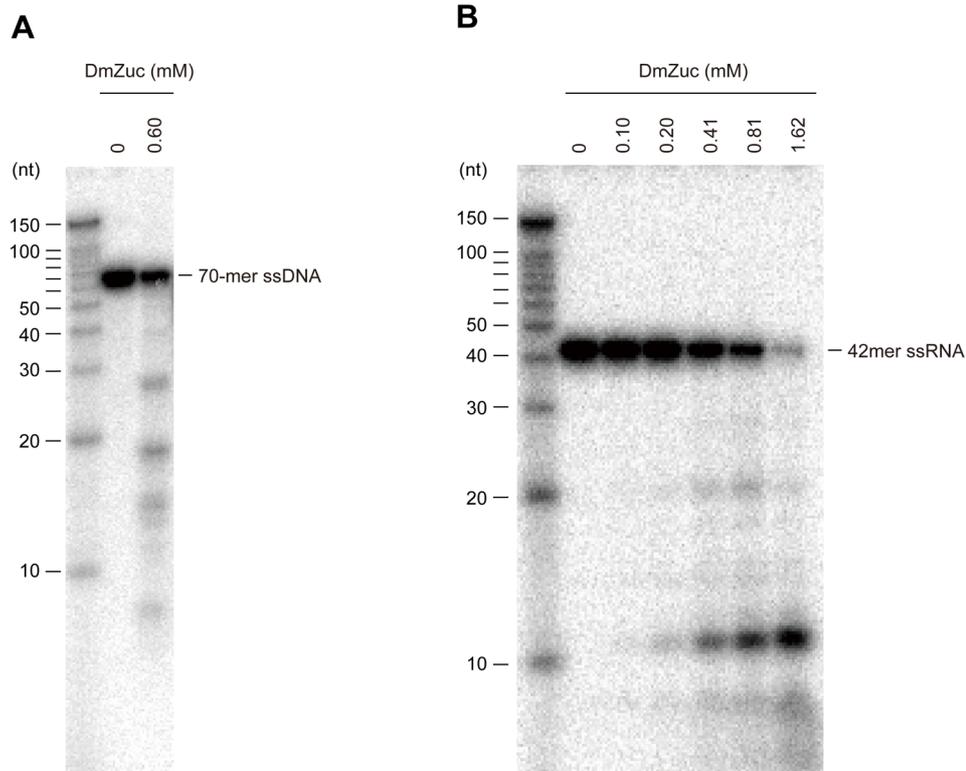
higher (Fig. 1-11 C). The 42-nt ssRNA labelled with  $^{32}\text{P}$  at the 3' end was used to characterize the 5' end structure of the cleavage products.

The cleavage products showed resistance to treatment with 5'-phosphate ssRNA-specific exonuclease (Terminator Exonuclease) after treatment with calf intestinal phosphatase (CIP) (Fig. 1-11 D). The endonuclease activity of DmZuc is modest in vitro, although it showed dose dependence (Fig. 1-12 B). Although MmZuc reportedly hydrolysed the mitochondrial lipid cardiolipin to phosphatidic acid (Choi *et al.*, 2006), neither DmZuc nor MmZuc hydrolysed cardiolipin in vitro (Fig. 1-13).



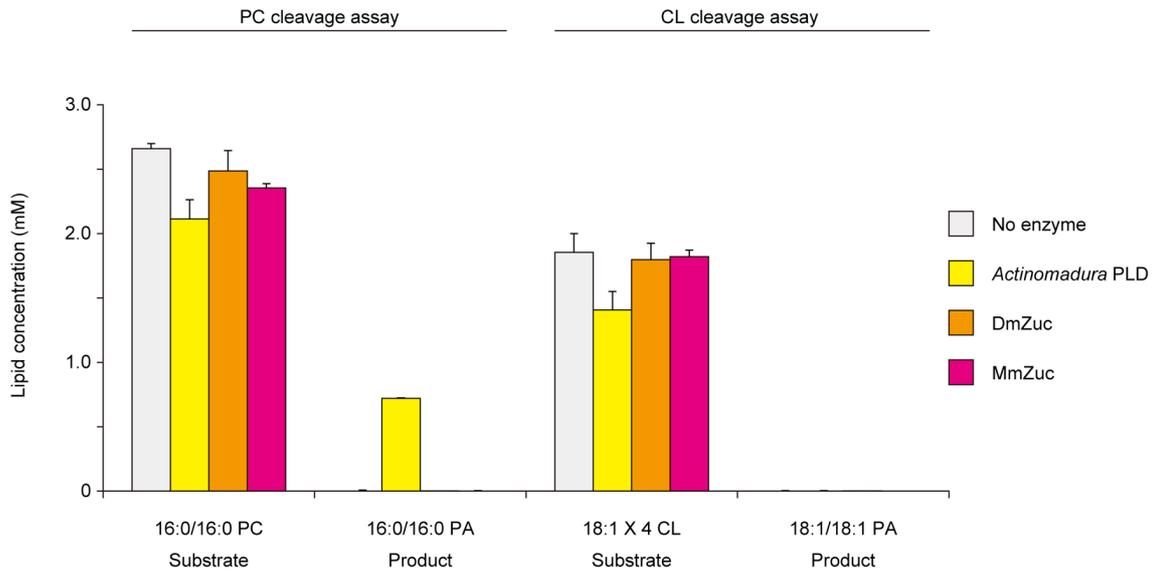
**Figure 1-11. Zuc is an endonuclease for single-stranded nucleic acids.**

(A) DmZuc and MmZuc, but not monomeric DmZuc or the DmZuc mutants, cleave a 42-nt ssRNA substrate. (B) DmZuc cleaves ssRNAs of various sequences and circular ssRNA, but not dsRNA. Exonuclease T failed to cleave the circular ssRNA. S and AS correspond to the sense and antisense strands of dsRNA, respectively. (C) DmZuc endonuclease activity requires no magnesium ions, and is inhibited by NaCl. (D) The ssRNA products cleaved by DmZuc have a 5' monophosphate. The cleavage products were resistant to treatment with Terminator Exonuclease (Ter. Exo.) after treatment with calf intestinal phosphatase (CIP).



**Figure 1-12. Characterization of DmZuc nuclease activity *in vitro*.**

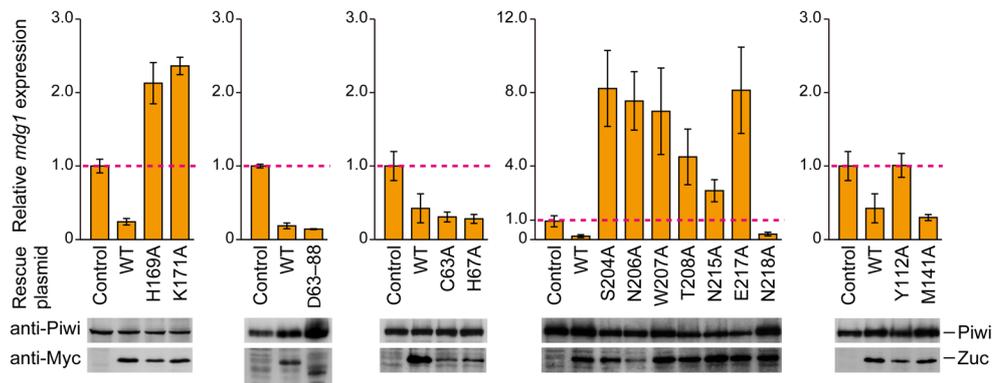
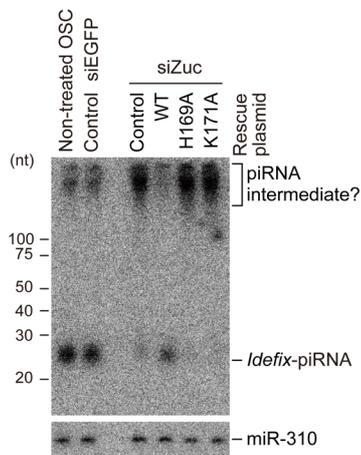
(A) DmZuc cleaves ssDNA. The 70-nt ssDNA labelled with  $^{32}\text{P}$  at the 5' end was incubated with WT DmZuc (0.60  $\mu\text{M}$ ) for 1 h at 26°C, and was then resolved on a 20% denaturing polyacrylamide gel. (B) DmZuc cleaves ssRNA in a dose-dependent manner. The 42-nt ssRNA labelled with  $^{32}\text{P}$  at the 5' end was incubated with WT DmZuc (0.10–1.62  $\mu\text{M}$ ) for 1 h at 26°C, and was then resolved on a 20% denaturing polyacrylamide gel.



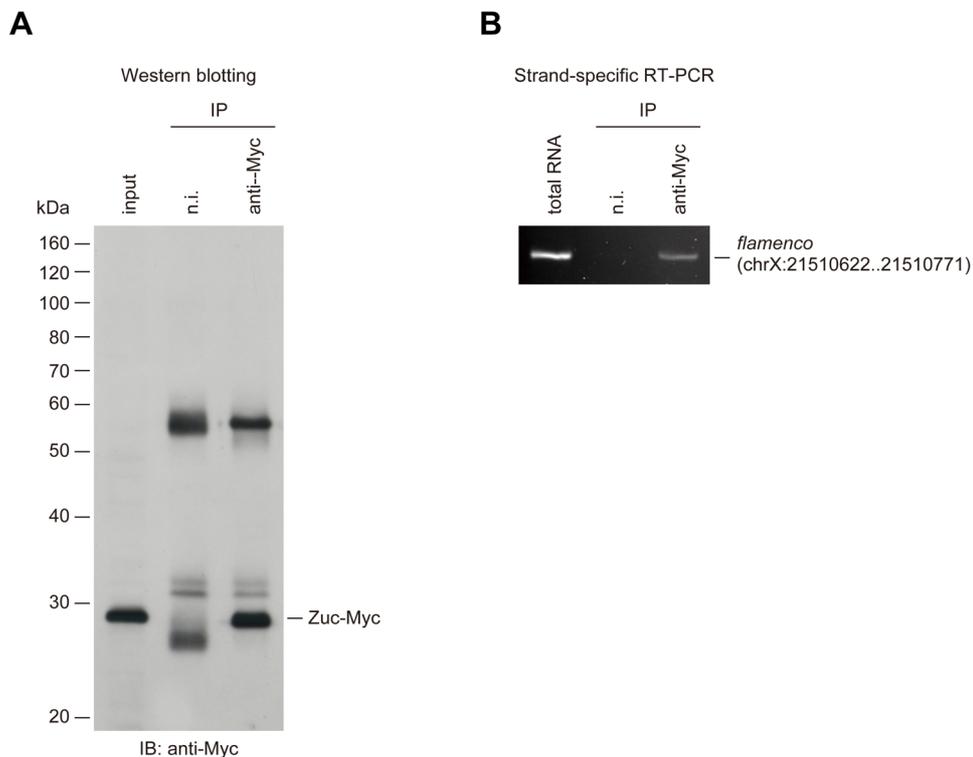
**Figure 1-13. DmZuc and MmZuc displayed no PLD activity.**

Purified DmZuc, MmZuc, or *Actinomadura* PLD (positive control) was incubated with dipalmitoyl (16:0/16:0) PC or tetradioleoyl (18:1 X 4) CL, and then phospholipid products were analyzed by LC/MS/MS. *Actinomadura* PLD failed to hydrolyse tetradioleoyl CL, but hydrolysed dipalmitoyl PC to produce dipalmitoyl PA. In contrast, DmZuc and MmZuc hydrolysed neither dipalmitoyl PC nor tetradioleoyl CL.

To explore the biological relevance of the DmZuc ssRNase activity in transposon silencing, Dr. Kuniaki Saito expressed RNA-mediated interference (RNAi)-resistant WT DmZuc<sub>1-253</sub> and mutants of DmZuc<sub>1-253</sub> in DmZuc-depleted ovarian somatic cells (OSCs), and then monitored the expression levels of the *mdg1* transposon by quantitative PCR with reverse transcription (RT-PCR). Western blotting confirmed similar expression levels of DmZuc WT and mutants (Fig. 1-14 A). WT DmZuc<sub>1-253</sub> rescued *mdg1* derepression, whereas the H169A and K171A mutants failed to rescue it (Fig. 1-14 A). The C63A, H67A and  $\Delta$ 63-88 mutants rescued *mdg1* derepression (Fig. 1-14 A). None of the mutants Y112A, S204A, N206A, W207A, T208A, N215A or E217A rescued *mdg1* derepression (Fig. 1-14 A). In contrast, the N218A mutant rescued *mdg1* derepression (Fig. 1-14 A). The expression of WT DmZuc, but not that of the H169A and K171A mutants, rescued the defects in *Idefix*-piRNA maturation in DmZuc-depleted OSCs (Fig. 1-14 B). RT-PCR after immunoprecipitation of DmZuc from OSCs revealed that DmZuc interacts with the piRNA precursors in OSCs (Fig. 1-15).

**A****B**

**Figure 1-14. DmZuc endoribonuclease activity is required for transposon silencing.** (A) Myc-tagged full-length WT and mutant DmZuc, mutated at a short interfering RNA (siRNA) recognition site and thus resistant to RNAi, were expressed in DmZuc-depleted OSCs. The expression levels of Piwi and DmZuc were examined by western blotting, and the expression levels of the *mdg1* transposon were monitored by quantitative RT-PCR. Results are shown as means  $\pm$  s.e.m. ( $n = 3$ ). As a control, OSCs were transfected with an empty vector. (B) Northern blotting with a piRNA probe, *Idefix*-piR1-R, showed that WT DmZuc, but not H169A or K171A, rescued the defects in piRNA maturation in DmZuc-depleted OSCs. The expression of WT DmZuc increased mature piRNAs with a concomitant decrease in piRNA intermediate like molecules. miR-310, one of the microRNAs in OSCs, was used as an internal loading control. ‘piRNA intermediate?’ corresponds to ‘piRNA- intermediate-like molecules’.



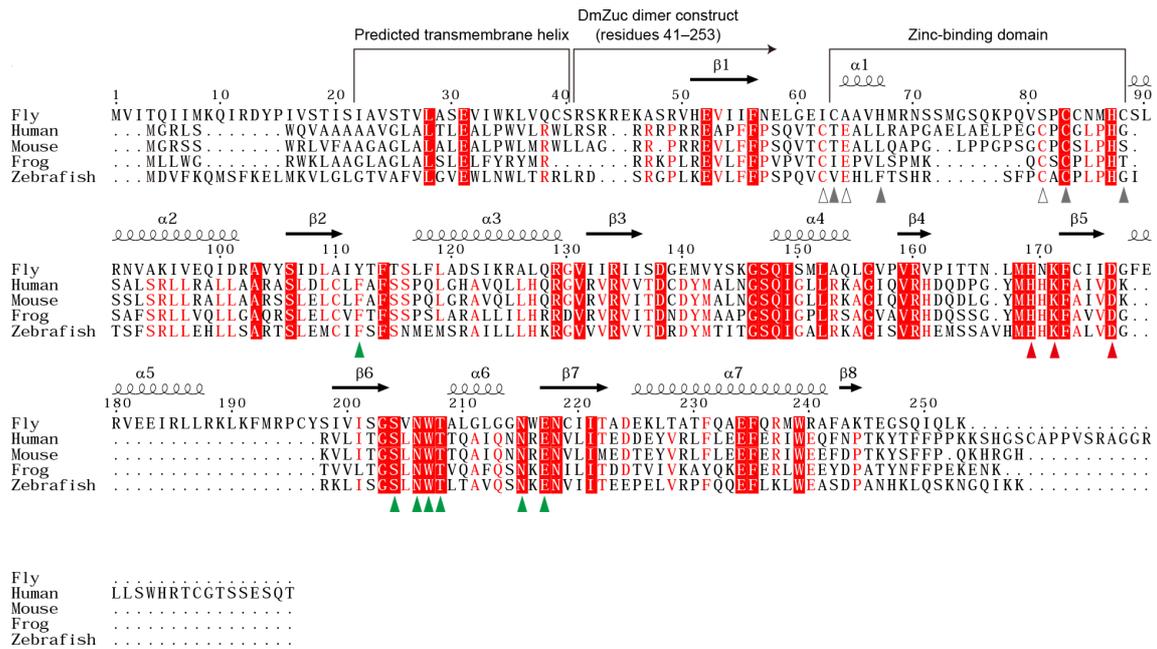
**Figure 1-15. DmZuc associates with piRNA precursors in OSCs.**

(A) C-terminal Myc-tagged, full-length DmZuc was expressed in OSCs by transfection. RT-PCR was performed on RNAs isolated from the material immunoprecipitated from the DmZuc-Myc expressing cells, using an anti-Myc antibody. Western blotting confirmed that DmZuc was efficiently immunoprecipitated from the cells. (B) RT-PCR indicated that DmZuc-Myc associates with a fragment of the *flamenco* transcript. A control experiment using non-immune IgG (n.i.) indicated that the interaction between DmZuc and the *flamenco* piRNA precursor is specific.

## 2.4 Discussion

### 2.4.1 Structure comparison

The catalytic domain consists of an eight-stranded mixed  $\beta$ -sheet flanked by  $\alpha$ -helices on both sides, and shares structural similarity with the catalytic domains of other PLD superfamily members, such as Nuc (Stuckey *et al.*, 1999, Leiros *et al.*, 2000) (PDB 1BYR; 26% sequence identity, root mean squared deviation 1.8 Å for 145 C $\alpha$  atoms) (Fig. 1-8). In the zinc-binding residues of DmZuc, only Cys 83 is conserved in animals (Fig. 1-16). In the corresponding region of MmZuc, the residues Cys 49, Glu 51, Cys 66 and Cys 68 are highly conserved among animals except for flies (Fig. 1-16), suggesting that these residues may coordinate a zinc ion in these species.



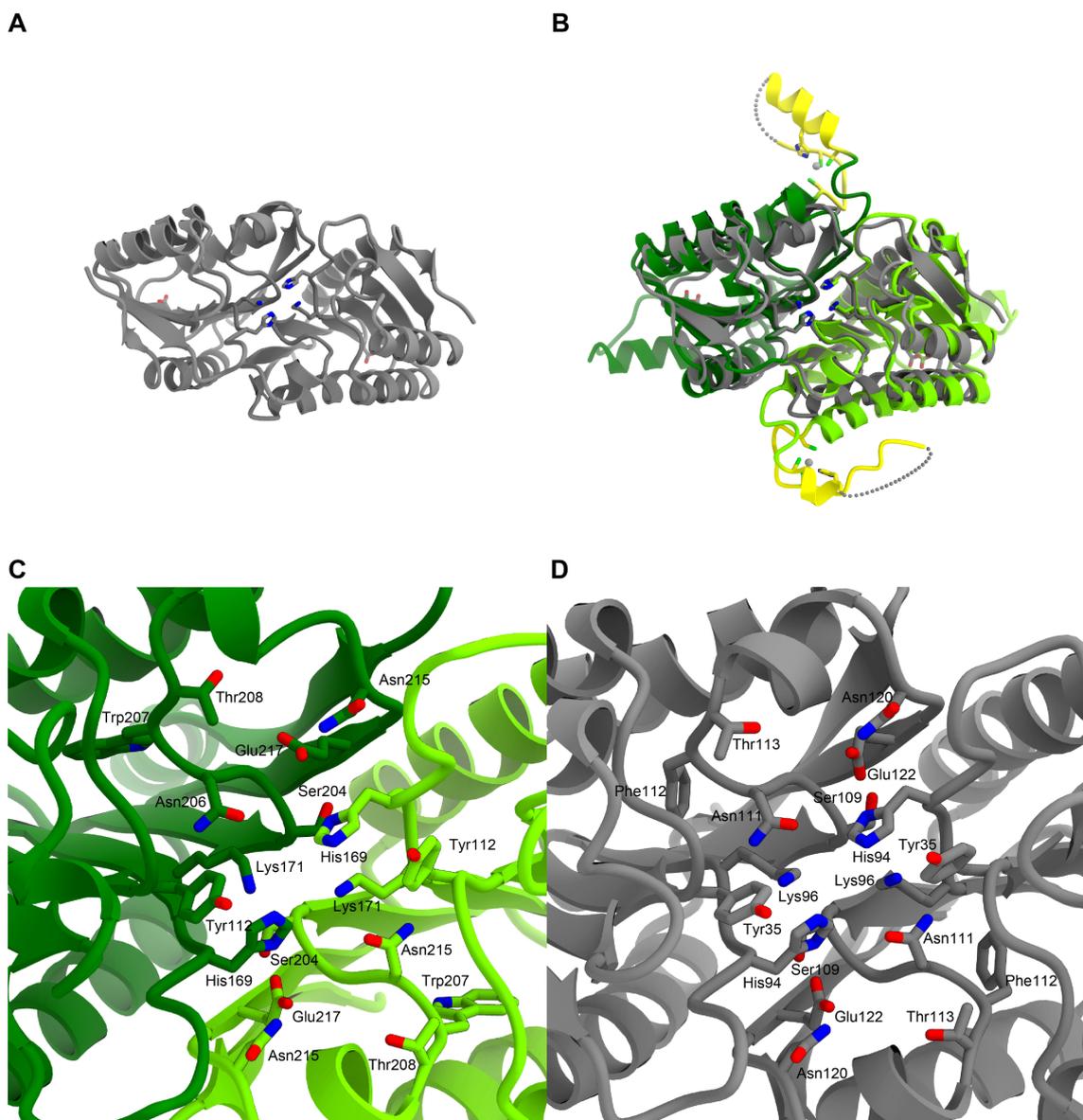
**Figure 1-16. Multiple sequence alignment of Zuc proteins from different animal species.**

The secondary structure of DmZuc is indicated above the sequences. The zinc-binding residues in DmZuc are indicated by grey triangles, and the CEC motif is indicated by white triangles. The active-site residues are indicated by green triangles.

The active-site groove is composed of conserved residues from the two protomers, including Tyr 112, His 169, Lys 171, Ser 204, Asn 206, Trp 207, Thr 208, Asn 215 and Glu 217, which correspond to Tyr 35, His94, Lys96, Ser109, Asn111, Phe112, Thr113, Asn120 and Glu 122, respectively, in Nuc (Fig. 1-17). Nuc cleaves substrates through a two-step catalytic mechanism, in which His94 in one protomer forms a phosphoenzyme intermediate with a substrate phosphorus atom, followed by the hydrolysis of this intermediate by a water molecule, activated by His94 in the other

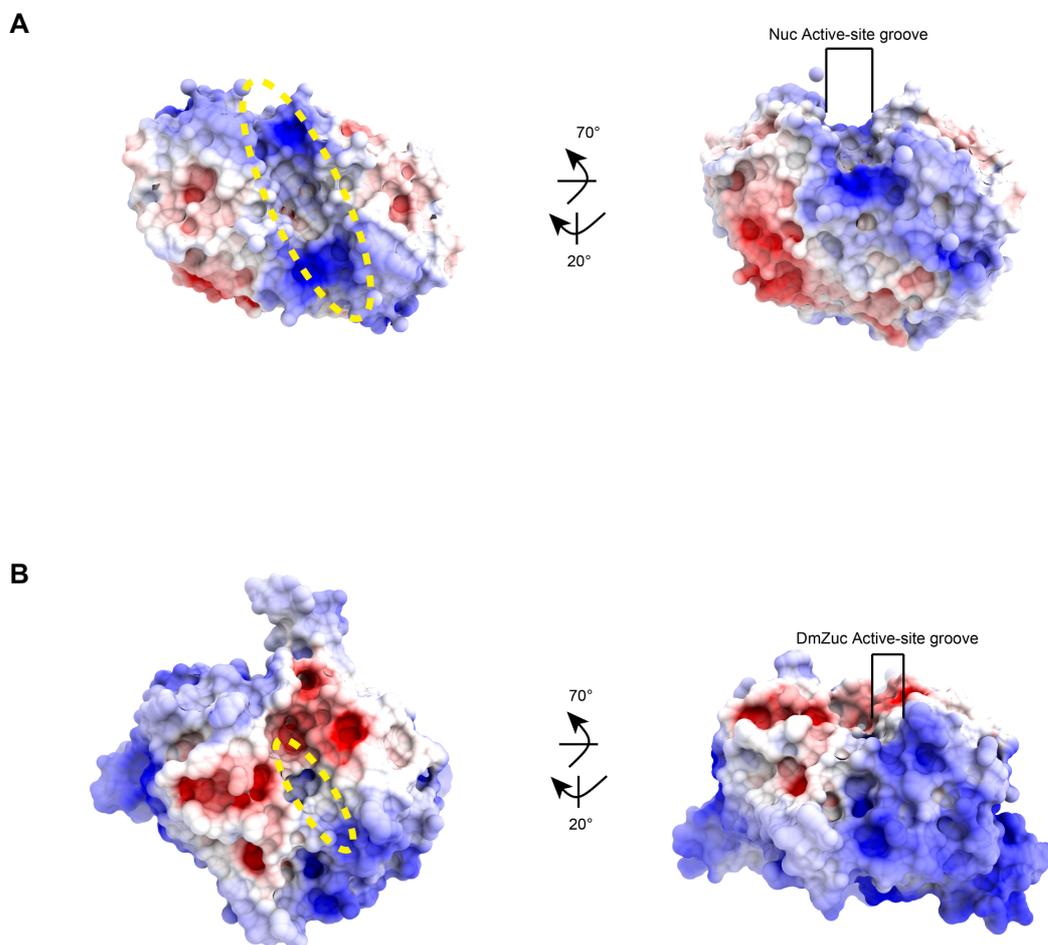
protomer (Stuckey *et al.*, 1999, Gottlin *et al.*, 1998). Lys96, Ser109, Asn111 and Glu122 in Nuc form a hydrogen-bonding network and participate in catalysis (Stuckey *et al.*, 1999). The active-site residues are similarly arranged in DmZuc and Nuc, suggesting that Zuc cleaves a substrate phosphodiester linkage through a similar catalytic mechanism. However, a structural comparison revealed a notable difference in their active-site architectures (Fig. 1-18). Nuc cleaves single-stranded and double-stranded nucleic acids *in vitro*, and has a wide, positively charged groove that can readily accommodate double-stranded nucleic acids (Pohlman *et al.*, 1993, Stuckey *et al.*, 1999, Zhao *et al.*, 1997) (Fig. 1-18 A). In contrast, the active-site groove of DmZuc is narrower than that of Nuc, and can apparently accommodate single-stranded, but not double-stranded, nucleic acids (Fig. 1-18 B). This structural difference is partly due to the replacements of Gly 64 and Ala 115 in Nuc with the bulkier Met141 and Leu210 residues in DmZuc, respectively (Fig. 1-19). A phosphate ion derived from the crystallization buffer binds to a positively charged patch near the active-site groove (Fig. 1-9 B), suggesting that this patch could interact with the phosphate group of nucleic acid substrates. The molecular surface on the side opposite the active-site groove is

positively charged (Fig. 1-20). The N termini of the two protomers are located on the same face, and the N-terminal residues 41–48 of one protomer form a positively charged helix, which would follow the transmembrane helix (residues 41–48 are disordered in the other protomer). This structural feature supports a previous model in which Zuc localizes on the outer mitochondrial membrane (Fig. 1-20) (Choi *et al.*, 2006, Watanabe *et al.*, 2011, Huang *et al.*, 2011, Saito *et al.*, 2010). Taken together, the structural features of DmZuc suggested that Zuc is a single-stranded nuclease that functions on the mitochondrial surface.



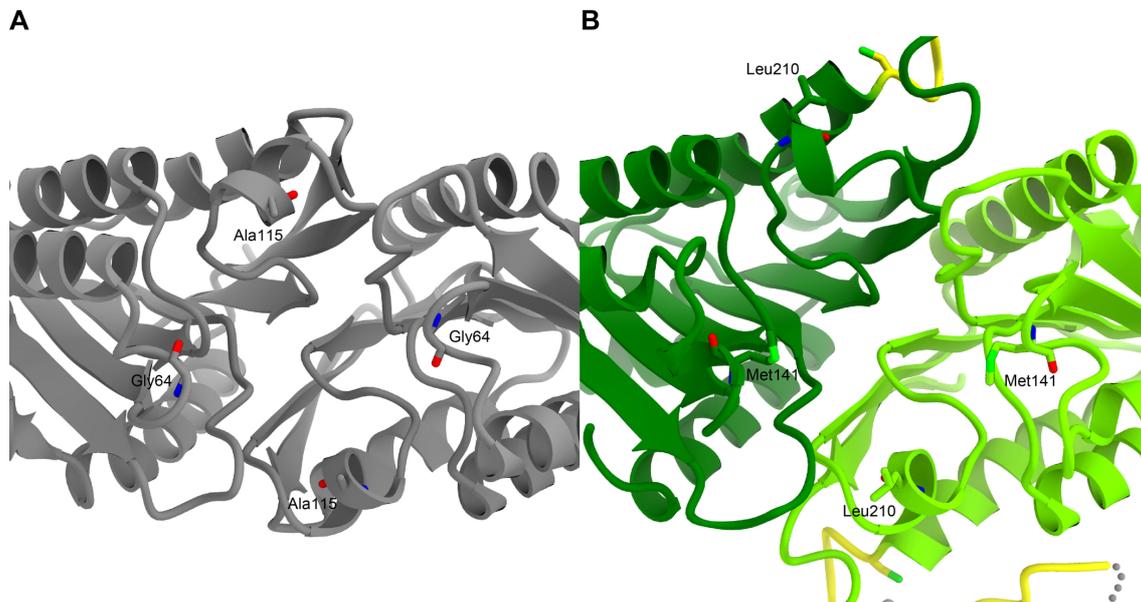
**Figure 1-17. Structural comparison of DmZuc and Nuc.**

(A) Overall structure of *Salmonella typhimurium* Nuc (PDB: 1BYR). His94, Lys96, and Asp101 are shown as sticks. (B) Superposition of the overall structures of DmZuc (coloured) and Nuc (PDB: 1BYR) (grey). (C) Active site of DmZuc. (D) Active site of Nuc (PDB 1BYR).



**Figure 1-18. Active-site groove.**

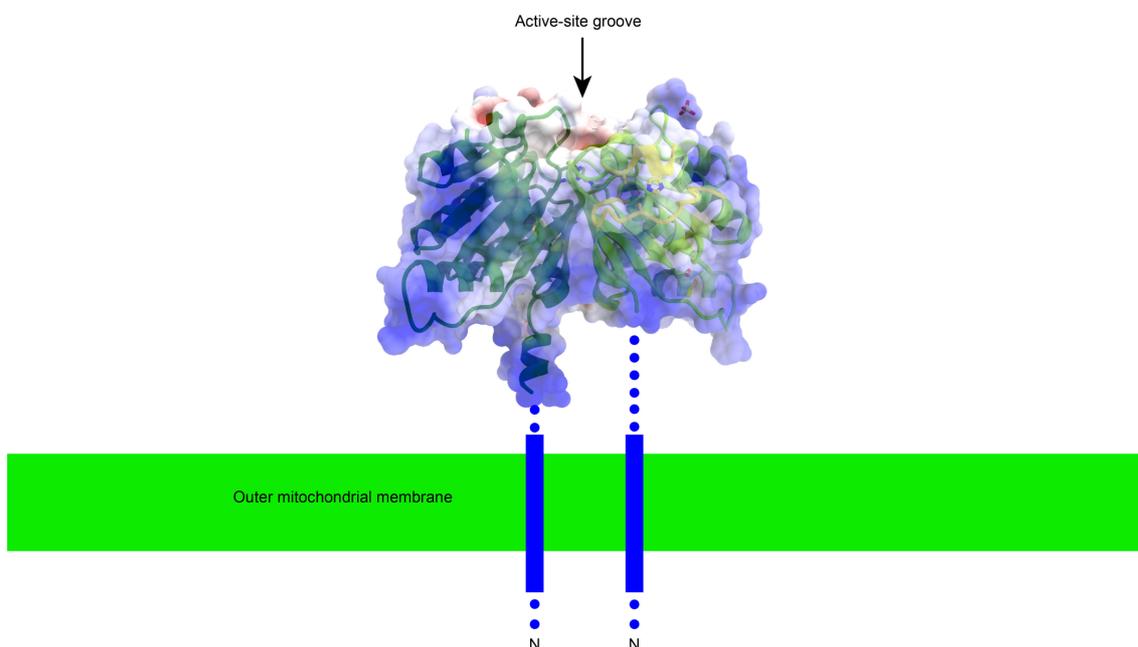
(A) Electrostatic surface potential of Nuc (PDB: 1BYR). (B) Electrostatic surface potential of DmZuc. The active sites are indicated by yellow dashed lines in (A) and (B).



**Figure 1-19. Structural comparison of active-site groove.**

(A) Active-site groove of Nuc (PDB: 1BYR). (B) Active-site groove of DmZuc.

Gly64 and Ala115 in Nuc (left) are not conserved, and are replaced with the bulkier Met141 and Leu210 residues in DmZuc (right), respectively.



**Figure 1-20. Model of DmZuc localization on the outer mitochondrial membrane.**

Electrostatic surface potential suggests that DmZuc is anchored on the mitochondrial surface through its N-terminal transmembrane helices, with the positively charged surface facing the outer mitochondrial membrane.

## 2.4.2 Functional analysis

DmZuc mutants (H169A, K171A, N206A, and N215A) and DmZuc<sub>89–253</sub> failed to cleave the ssRNA substrate (Fig. 1-11 A), highlighting the importance of the conserved residues and the requirement of dimerization, respectively. A DmZuc  $\Delta$ 63–88 mutant showed decreased ssRNase activity (Fig. 1-11 A), indicating the importance of the zinc-binding domain for the nuclease activity. These results affirmed the importance of designing crystallization constructs based not only on bioinformatics predictions but also on functional experiments.

Both purified MmZuc<sub>35–221</sub> and DmZuc showed a ssRNase activity (Fig. 1-11 A). Thus, Zuc is an evolutionarily conserved RNase. As DmZuc cleaved ssRNA, and also ssDNA (Fig. 1-12 A), but not dsRNA (Fig. 1-11 B), DmZuc is a single-strand-specific endonuclease. Although the poly(U) cleavage products showed an even distribution, DmZuc did not always cleave the ssRNA substrates at specific nucleotides (Fig. 1-11 B), suggesting that DmZuc does not have strict sequence specificity. As DmZuc showed similar activity in the presence of magnesium ions or EDTA (Fig. 1-11 C), DmZuc is a metal-independent nuclease. Sodium ions may inhibit

substrate binding (Fig. 1-11 C), as observed for T4 DNA ligase (Raae, *et al.*, 1975) and the dsRNA-specific RNase Pac1 (Rotondo *et al.*, 1996). The cleavage products showed resistance to treatment with 5'-phosphate ssRNA-specific exonuclease after treatment with CIP (Fig. 1-11 D), suggesting that the cleavage products contain a 5' monophosphate. Since the DmZuc cleaves ssRNA in a dose-dependent manner (Fig. 1-12 B), the DmZuc activity may be stimulated *in vivo* by unknown cofactors. Neither DmZuc nor MmZuc hydrolysed cardiolipin *in vitro* (Fig. 1-13). According to previous studies that Zuc hydrolyzed cardiolipin, there was a possibility that contaminants were mixed in purified Zuc, so it may be considered that the hydrolytic activity of lipids by contaminants was measured (Choi *et al.*, 2006). Together, these results indicated that Zuc is an endonuclease specific for single-stranded nucleic acids.

The *mdg1* expression was higher in OSCs expressing the H169A or K171A mutant than in control OSCs (Fig. 1-14 A). The mutants may have dominant-negative effects, although the underlying mechanism is unknown. While the  $\Delta 63-88$  mutant had a lower ssRNase activity (about 5% of the WT activity) (Fig. 1-11 A), the C63A, H67A and  $\Delta 63-88$  mutants rescued *mdg1* derepression (Fig. 1-14 A). These mutants may

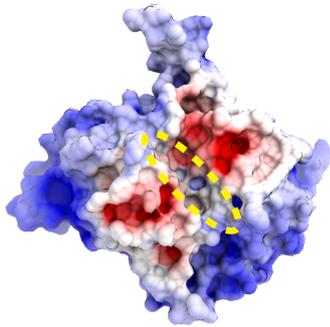
suffice for transposon silencing. The residues, which are conserved (Fig. 1-16) and form the active-site groove (Fig. 1-17) are necessary for rescue DmZuc-depleted OSCs from *mdg1* derepression (Fig. 1-14 A). Although the N218A mutant rescued *mdg1* derepression (Fig. 1-14 A), Asn 218 is buried within the protein, and thus would not contribute to substrate binding. The H169A, K171A, N206A and N215A mutants lacked ssRNase activity in vitro (Fig. 1-11 A) and the ability to repress transposons in vivo (Fig. 1-14 A), indicating that the DmZuc ssRNase activity is critical for transposon silencing. The expression of WT DmZuc rescued the defects in *Idefix*-piRNA maturation in DmZuc-depleted OSCs (Fig. 1-14 B), suggesting that the DmZuc ssRNase activity is required for primary piRNA maturation. Together, these results indicate that Zuc is an endoribonuclease essential for primary piRNA processing.

#### **2.4.3 Comparison of DmZuc with other solved phospholipase D (PLD) family**

At the same time as the structure of DmZuc was published, a research group on the US published a paper on the structural and functional analysis of MmZuc, a mouse Zuc homolog (Ipsaro *et al.*, 2012). This group also performed biochemical analysis, indicating that MmZuc cleaves single-stranded nucleic acids, but does not

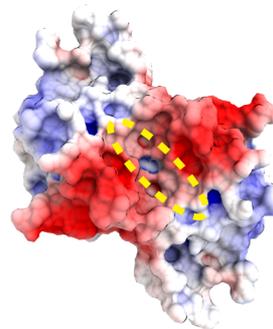
cleave cardiolipin (Ipsaro *et al.*, 2012). In addition, this research group determined the crystal structure of MmZuc (Ipsaro *et al.*, 2012). Focusing on the fact that the active site of MmZuc is narrower than that of Nuc, and that the active site of MmZuc is different in the structure of other PLDs with phospholipase activity, this research group reported that Zuc is a ribonuclease specific for single stranded RNA (Ipsaro *et al.*, 2012). From the fact that two independent studies have reached similar conclusions, Zuc was thought to be directly involved in piRNA production as a ribonuclease (Fig. 1-21).

**A**



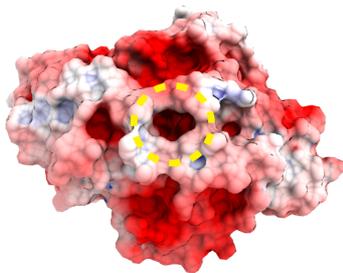
DmZuc

**B**



MmZuc (PDB: 4GGJ)

**C**



phospholipase D (PDB: 2ZE9)

**Figure 1-21. Comparative analysis of PLD-family proteins.**

(A) Electrostatic surface potential of DmZuc. (B) Electrostatic surface potential of MmZuc (PDB: 4GGJ). (C) Electrostatic surface potential of phospholipase D (PDB: 2ZE9). The active sites are indicated by yellow dashed lines. The active site structures of DmZuc and MmZuc are different from that of phospholipase D. DmZuc and MmZuc have distinct binding surfaces for single stranded RNA, while phospholipase D has distinct binding surfaces for phospholipids.

## 2.5 Conclusion

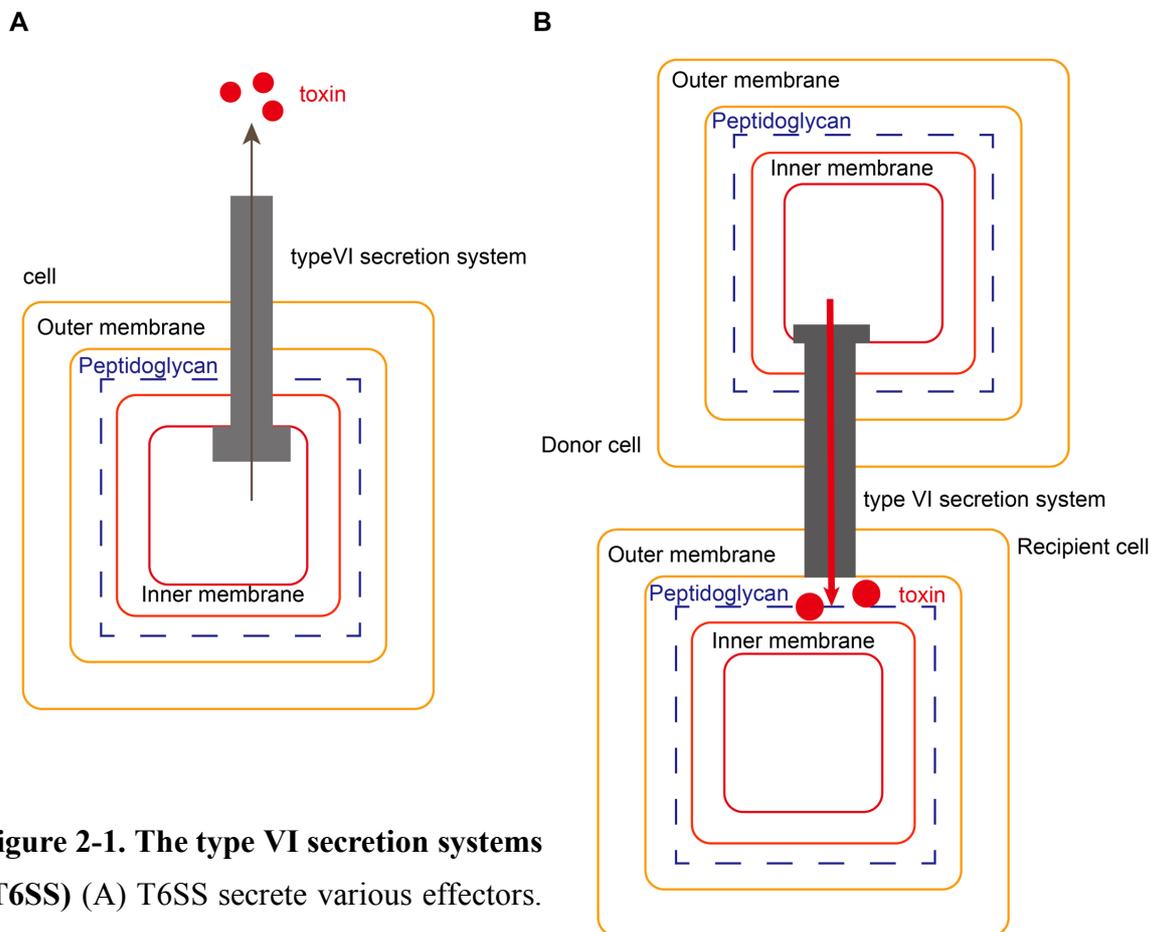
In this work, the author determined the crystal structures of DmZuc. The structure of DmZuc shares structural similarity with the catalytic domains of other PLD superfamily members. The structure revealed that the dimer interface of DmZuc has a positively charged, narrow catalytic groove, which could interact a single-stranded RNA. In addition, functional analysis based on the crystal structure showed that the DmZuc and the mouse homologue MmZuc have enzymatic activity to cleave single-stranded RNA. The conserved active-site residues of DmZuc are critical for the ribonuclease activity, for piRNA maturation, and for transposon silencing. These structural and functional analyses revealed the Zuc has a key role in piRNA biogenesis.

## **Chapter3: Crystal structure of Type VI Effector-Immunity complex from**

### ***Agrobacterium tumefaciens***

#### **3.1 Introduction**

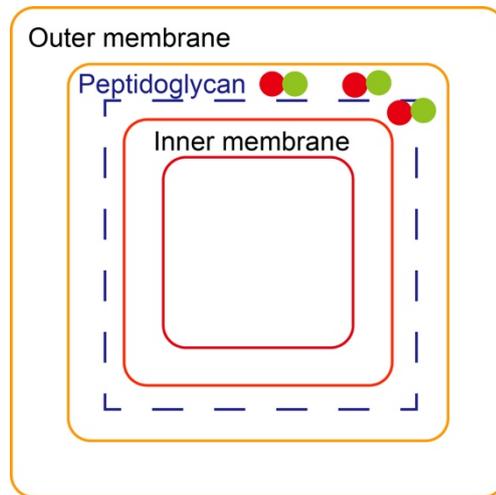
The type VI secretion systems (T6SS) of Gram-negative bacteria inject various toxic effectors into the periplasmic or cytoplasmic space of their target cells and induce cell lysis of enemy cells (Fig. 2-1). T6SS is composed of various proteins, that are usually encoded by clustered genes. In previous studies, T6SS is expected to deliver toxic effectors to target cells triggered by a signaling cascade that occurs when bacteria come in contact with other bacteria. For the transport of toxic effectors via T6SS, proteins such as Hcp and VgrG, which are co-expressed by the same operon as toxic effectors act as adapters for toxic effectors. Proposed model of secretion mechanism is that a complex of toxic effector and adapter protein is ultimately secreted by incorporation of the complex into T6SS (Hood *et al.*, 2010; MacIntyre *et al.*, 2010; Schwarz *et al.*, 2010; Murdoch *et al.*, 2011; Russell *et al.*, 2011; Russell *et al.*, 2012; Durand *et al.*, 2014; Russell *et al.*, 2014).



**Figure 2-1. The type VI secretion systems (T6SS)** (A) T6SS secrete various effectors. (B) T6SS inject toxic effectors into the periplasmic or cytoplasmic space of their target cells.

There are many T6SS-related effector proteins, for example, T6SS-related amidase effector proteins (Tae), T6SS-related DNase effectors (Tde), T6SS-related lipase effectors (Tle), and T6SS-related glycoside hydrolase effectors (Tge) (Durand *et al.*, 2014; Russell *et al.*, 2014). One of the most studied effectors are Tae. The various Tae are classified into four families (Tae1, Tae2, Tae3, and Tae4), based on their

cleavage specificities (Russell *et al.*, 2012). These effectors and the unique bacterial secretion system, T6SS, which responds to enemy bacteria, enable Gram-negative bacteria to attack targeted heterologous cells (Russell *et al.*, 2012). In addition to these toxic effectors, Gram-negative bacteria have four amidase immunity proteins (Tai1, Tai2, Tai3, and Tai4). Tai1, Tai2, Tai3, and Tai4 neutralize the endogenous toxic effectors Tae1, Tae2, Tae3, and Tae4, respectively. These effector-immunity pairs (Tae1-Tai1, Tae2-Tai2, Tae3-Tai3, and Tae4-Tai4) originate from the same operons in general. The presence of cognate effector-immunity pairs suggests that the self-protection systems with the co-expression of effector proteins and immunity proteins are a common feature in the Gram-negative bacteria possessing T6SS (Russell *et al.*, 2012) (Fig. 2-2).



**Figure 2-2. Effector-immunity pairs.**

Immunity proteins (green) neutralize the toxic effectors (red) and prevent cell lysis.

Tae4-Tai4 is the fourth family member of the T6SS-related effector-immunity pairs. Tae4 cleaves the  $\gamma$ -D-glutamyl-*m*DAP bond, which presents in the bacteria's cell wall peptidoglycan (Russell *et al.*, 2012). In contrast, Tai 4, which generally has a signal peptide at the N terminus is present in the periplasm, preventing cell-destruction by neutralizing Tae 4 derived from bacteria of the same species and Tae 4 derived from a close kind of bacteria (Russell *et al.*, 2012; Zhang *et al.*, 2013). Previous studies reported the Tae4-Tai4 complex structures from *Enterobacter cloacae*, *Salmonella typhimurium*, and *Serratia marcescens* (Zhang, Gao, Wang *et al.*, 2013; Zhang, Gao, Wei *et al.*, 2013; Benz *et al.*, 2013; Srikannathan *et al.*, 2013). Comparisons of these

structures revealed that the *S. marcescens* Tai4 is structurally different from *E. cloacae* Tai4 and *S. typhimurium* Tai4, whereas the Tae4 proteins from different species are highly conserved (Srikannathasan *et al.*, 2013). However, the crystal structures of the Tae4-Tai4 complexes from the other species have remained unknown.

*Agrobacterium tumefaciens* has T6SS and secretes Tae4 and Tde. *A. tumefaciens* infects plants and uses these effectors for interbacterial competition (Ma *et al.*, 2014). *A. tumefaciens* also has Tai4 and T6SS-related DNase immunity (Tdi) and neutralize the toxic Tae4 and Tde, respectively (Ma *et al.*, 2014). Due to amino acid sequence homology, the structure of Tae4-Tai4 complex of *A. tumefaciens* was expected to have similar structure to Tae4-Tai4 complex derived from *S. marcescens* (Srikannathasan *et al.*, 2013). Therefore, structural studies of Tae4 and Tai4 of *A. tumefaciens* were expected to enhance our understanding of the catalytic and inhibitory mechanisms of the Tae4 and Tai4 family proteins. In addition, in the case of Tde and Tdi of *A. tumefaciens*, since similar structures or their secretion mechanism has not been elucidated, their structures and secretion systems are intriguing (Durand *et al.*, 2014).

To gain insights into the T6SS-related effector-immunity pairs, the author

tried structural analysis of Tae4-Tai4 complex and Tde-Tdi complex derived from *A. tumefaciens*. As a result, the author determined the crystal structures of Tai4 and the Tae4-Tai4 complex from *A. tumefaciens* at 1.55 Å and 1.9 Å resolutions, respectively.

## 3.2 Materials and methods

### 3.2.1 Macromolecule production

The gene encoding the Tai4 and Tae4 proteins from *A. tumefaciens* (ATU4346 and ATU4347) were codon optimized for *Escherichia coli* and synthesized by Invitrogen. The SignalP 4.1 server (Petersen *et al.*, 2011) was used to predict the signal peptide of *A. tumefaciens* Tai4 (*AtTai4*). The *AtTai4* gene segment (residues 26–129) without the putative signal sequence was PCR-amplified and inserted into the pCold-GST vector. The plasmid was transformed into *E. coli* Rosetta 2 (DE3) cells for protein expression.

The cells were grown in Luria-Bertani (LB) medium at 310 K until the OD<sub>600</sub> reached 0.8, and the gene expression was induced with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG), following a temperature reduction to 277 K. The growth was continued for 24 h at 288 K. The N-terminally His<sub>6</sub>-GST-tagged *AtTai4* was affinity-purified using a Ni-NTA column (QIAGEN). The N-terminally His<sub>6</sub>-GST-tag was removed by an incubation with Turbo3c protease (Nacalai Tesque) for 16 h at 277 K. After re-chromatography on the Ni-NTA column, further purifications were

conducted by ion exchange chromatography on a Resource Q column (GE Healthcare) and gel filtration chromatography on a HiLoad Superdex 75 column (GE Healthcare). The purified samples were concentrated to 8.6 mg ml<sup>-1</sup> for crystallization.

For the co-expression of the *AtTae4-AtTai4* complex, the *AtTai4* gene segment (residues 26–129) was cloned into the first multiple cloning site of the pETDuet-1 vector (Novagen), and the *AtTae4* gene segments (residues 1–166 and 1–163) were subsequently cloned into the second multiple cloning site. A tobacco etch virus (TEV) protease recognition sequence was introduced between the His<sub>6</sub> tag and the *AtTai4* sequence by a PCR-based method. The plasmid was transformed into *E. coli* Rosetta 2 (DE3) cells for overexpression. The cells were cultured in LB medium at 310 K until the OD<sub>600</sub> reached 0.8, and the gene expression was induced with 0.5 mM IPTG following a temperature reduction to 277 K. The cells were further cultured at 291 K for 24 h. The *AtTae4-AtTai4* complex was affinity-purified using a Ni-NTA column (QIAGEN). The N-terminal His<sub>6</sub>-tag was removed by an incubation with TEV protease for 24 h at 277 K. After re-chromatography on the Ni-NTA column, the complex was further purified by ion-exchange chromatography on a Resource Q column and

subsequent gel-filtration chromatography on a HiLoad Superdex 75 column. The purified complex was concentrated to 13 mg ml<sup>-1</sup> for crystallization trials.

The DNA fragments encoding the Tde and Tdi proteins from *A. tumefaciens* (ATU4350 and ATU4351) were codon optimized for *Escherichia coli* and synthesized by Invitrogen. The *AtTdi* gene segment (residues 1–224) was cloned into the first multiple cloning site of the pETDuet-1 vector (Novagen). The *AtTde* gene segments (residues 1–278) were cloned into pBAD vector. The plasmids were transformed into *E. coli* Rosetta 2 (DE3) cells for overexpression. For the co-expression of the *AtTde-AtTdi* complex, the cells were grown in LB medium at 310 K until the OD<sub>600</sub> reached 0.5, and the *AtTdi* gene expression was induced with 0.5 mM IPTG. The growth was continued for 2 h at 310 K. Thereafter, *AtTde* gene expression induced with 0.0002% to 0.2% L-arabinose.

### **3.2.2 Crystallization**

Initial crystallization trials were performed at 293 K by the sitting-drop vapor diffusion method in a 96-well crystallization plate, using various commercially available screening kits. Crystallization drops were prepared by mixing 200 nl of

purified protein solution and 200 nl of reservoir solution using a Mosquito crystallization robot (TTP LabTech). Initial crystals of *AtTai4* were optimized at 293 K, by varying the concentrations of PEG and salt in the reservoir solutions using an Additive Screen kit (Hampton Research). The plate-shaped crystals of *AtTai4* were obtained in 33% PEG 6000, 1.5 M lithium chloride, and 100 mM sodium acetate. The *AtTae4-AtTai4* complexes formed thick plate-shaped crystals, under Mem GOLD E11 reservoir conditions consisting of 35% PEG 400, 0.05 M Tris pH 8.5, 0.05 M sodium sulfate, and 0.05 M lithium sulfate.

### **3.2.3 Data collection and processing**

All crystals were cryoprotected in reservoir solution supplemented with 25% ethylene glycol, and flash-cooled in a nitrogen gas stream. X-ray diffraction data of *AtTai4* and the *AtTae4-AtTai4* complex were collected on beamlines BL41XU and BL32XU at SPring-8 (Hyogo, Japan), using a PILATUS3 6M detector (DECTRIS Ltd.) and an MX225HS detector (Rayonix, LLC), respectively. The continuous helical data collection scheme was applied using  $12 \times 8 \mu\text{m}^2$  (*AtTai4*) and  $18 \times 1 \mu\text{m}^2$  (*AtTae4-AtTai4* complex) beam. Diffraction data were integrated with DIALS (Waterman *et al.*, 2013)

and scaled with AIMLESS (Evans & Murshudov, 2013).

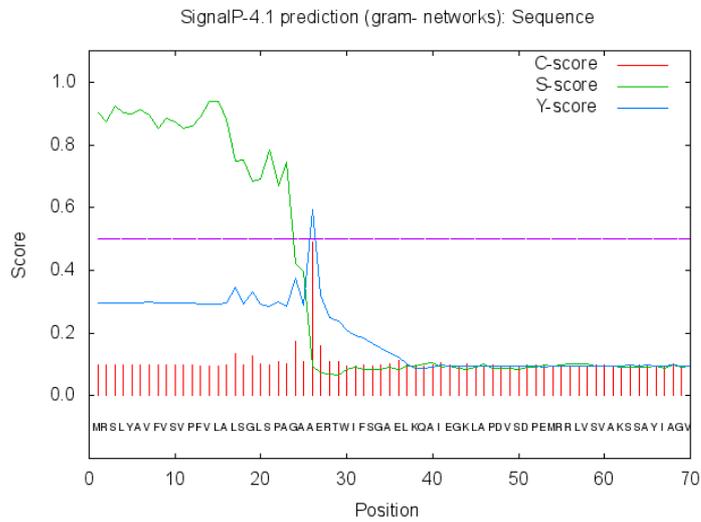
### **3.2.4 Structure determination**

The structures of *AtTai4* and the *AtTae4-AtTai4* complex were solved by molecular replacement with MOLREP (Vagin & Teplyakov, 2010), using the structure of Tai4 from *S. marcescens* (PDB code 3ZFI, Srikannathasan *et al.*, 2013) and the Tae4-Tai4 complex from *S. marcescens* (PDB code 4BI8, Srikannathasan *et al.*, 2013), respectively, as the search models. Model building and structure refinement were performed using Coot (Emsley *et al.*, 2010) and REFMAC5 (Murshudov *et al.*, 2011), respectively. Ramachandran plot analysis was performed using RAMPAGE (Lovell *et al.*, 2003). The atomic coordinates and structure factors of *AtTai4* and the *AtTae4-AtTai4* complex have been deposited in the Protein Data Bank (PDB), under the accession codes 6IJE and 6IJF, respectively. X-ray diffraction images have been also deposited in the Zenodo data repository (<https://doi.org/10.5281/zenodo.1453302>).

### 3.3 Results

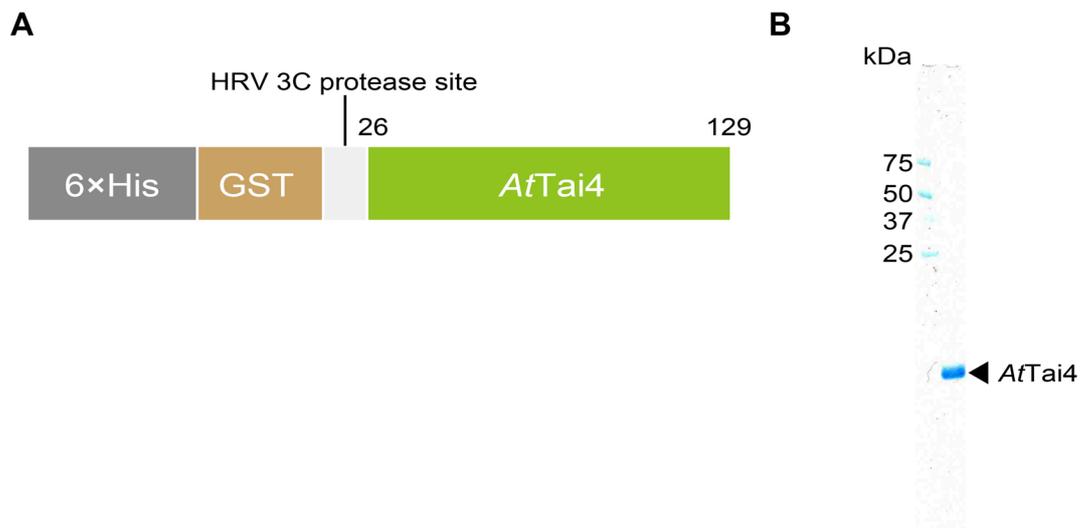
#### 3.3.1 Structure determination

The signal peptide (residues 1–25) of *AtTai4* was predicted by the SignalP 4.1 server (Petersen *et al.*, 2011) (Fig. 2-3). The domain (residues 26–129) of *AtTai4* was prepared, using *E. coli* expression system and purified (Fig. 2-4). After the initial crystallization screening, the author performed crystallization, using hanging drop vapor diffusion method. The crystal of *AtTai4* was obtained, which diffracted to 1.55 Å resolution (Fig. 2-5 A, B). The structure was determined by molecular replacement, using Tai4 from *S. marcescens* (PDB code 3ZFI, Srikannathasan *et al.*, 2013) as a search model (Fig. 2-6, Table 2-1).



**Figure 2-3. SignalP 4.1 Server prediction result of *AtTai4*.**

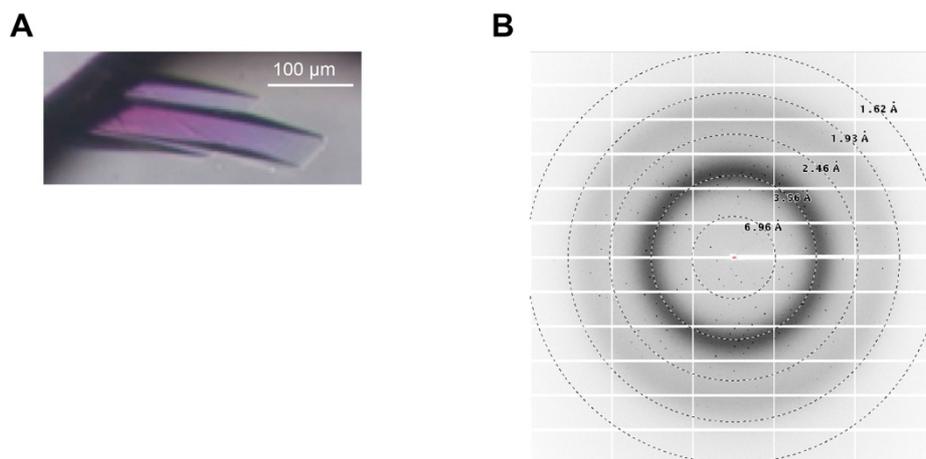
The predicted cleavage position of the signal peptide was between Ala25 and Ala26.



**Figure 2-4. Expression and purification of *AtTai4*.**

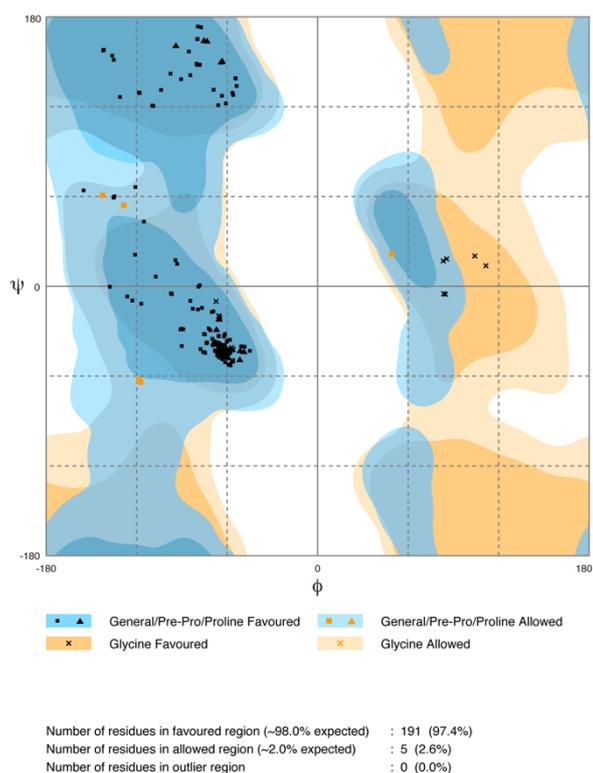
(A) Schematic diagram of the *AtTai4* crystallization construct.

(B) SDS-PAGE analysis of purified *AtTai4* with molecular-weight markers (labelled in kDa).



**Figure 2-5. The crystal and X-ray diffraction image of the *AfTai4*.**

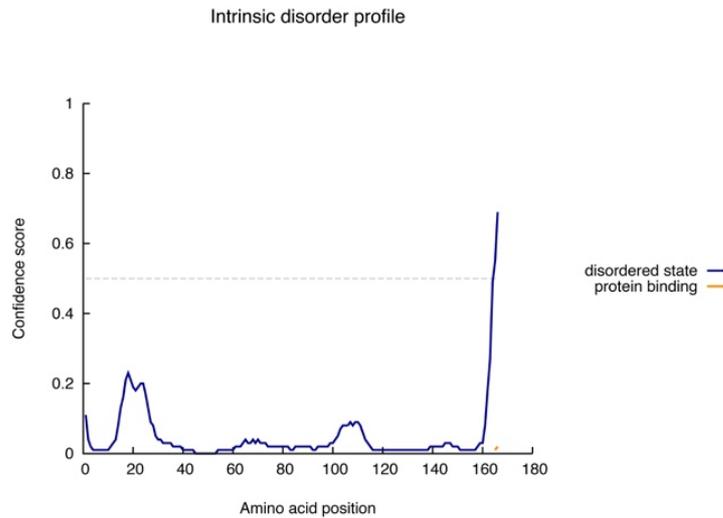
(A) The crystal of the *AfTai4*. (B) X-ray diffraction image of the *AfTai4*.



**Figure 2-6. Ramachandran plots of the *AfTai4* structure.**

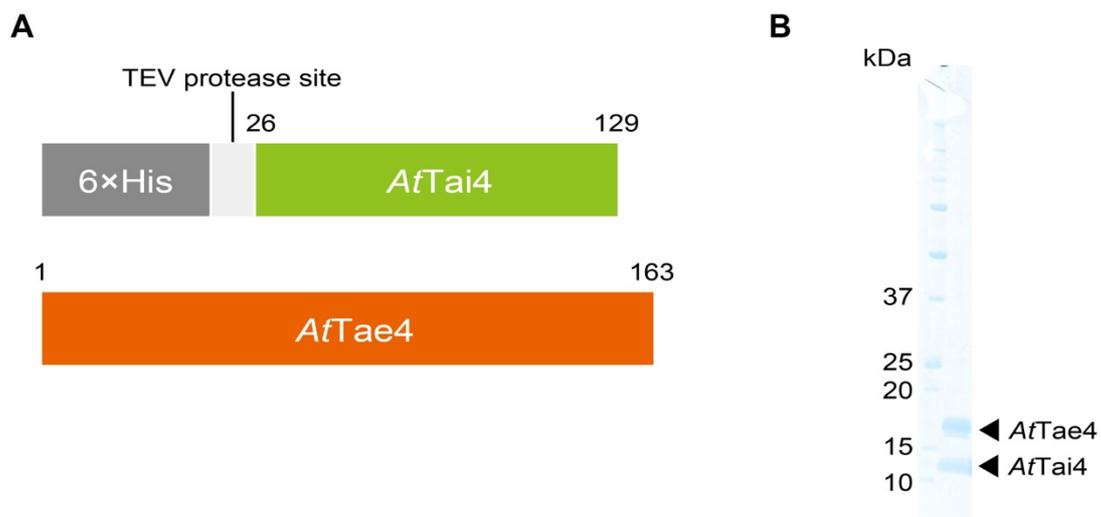
For the co-expression of the *AtTae4-AtTai4* complex, the domain (residues 26–129) of *AtTai4* and the domain (residues 1–166) of *AtTae4* were co-expressed, using *E. coli* expression system and purified. After the initial crystallization screening, the author performed crystallization, using hanging drop vapor diffusion method. However, as the crystals were poorly diffracted, the crystals were not sufficient for data collection. To design the constructs of *AtTae4* suitable for crystallization, the author performed the secondary-structure prediction and disordered prediction, using Psipred (Bryson *et al.*, 2005) and Disopred (Ward *et al.*, 2004), respectively (Fig. 2-7). The domain (residues 26–129) of *AtTai4* and the C- terminus deletion construct (residues 1-163) of *AtTae4* were co-expressed, using *E. coli* expression system and purified (Fig. 2-8). After the initial screening, crystals were obtained under Mem GOLD E11 reservoir conditions consisting of 35% PEG 400, 0.05 M Tris pH 8.5, 0.05 M sodium sulfate, and 0.05 M lithium sulfate (Fig. 2-9 A). The crystal of *AtTae4-AtTai4* complex diffracted to 1.9 Å resolution (Fig. 2-9 B) The structure was determined by molecular replacement, using *Tae4-Tai4* complex from *S. marcescens* (PDB code 4BI8, Srikannathasan *et al.*, 2013) as a search model (Fig. 2-10, Table 2-1).

For the co-expression of the *AtTde4-AtTdi4* complex, the author tried to clone the *AtTde* gene segments (residues 1–278) into the second multiple cloning site of pETDuet-1 vector (Novagen). However, transformed competent *E. coli* cells did not form colonies. Thus, the author tried to use pBAD vector for cloning. The *AtTde* gene segments (residues 1–278) were successfully cloned into pBAD vector. However, co-expression of the *AtTde4-AtTdi4* complex was failed, since *E. coli* Rosetta 2 (DE3) cells for overexpression died almost immediately after induction of *AtTde* expression.



**Figure 2-7. Characteristics of *AtTae4*.**

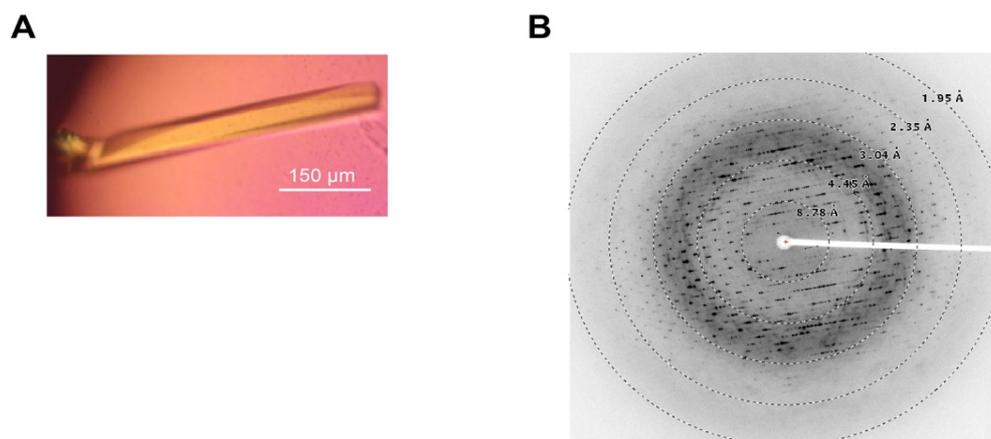
The disorder profile of full-length *AtTae4*. The disorder prediction suggested that C-terminal region of *AtTae4* is predicted to be disordered.



**Figure 2-8. Expression and purification of *AtTae4-AtTai4* complex.**

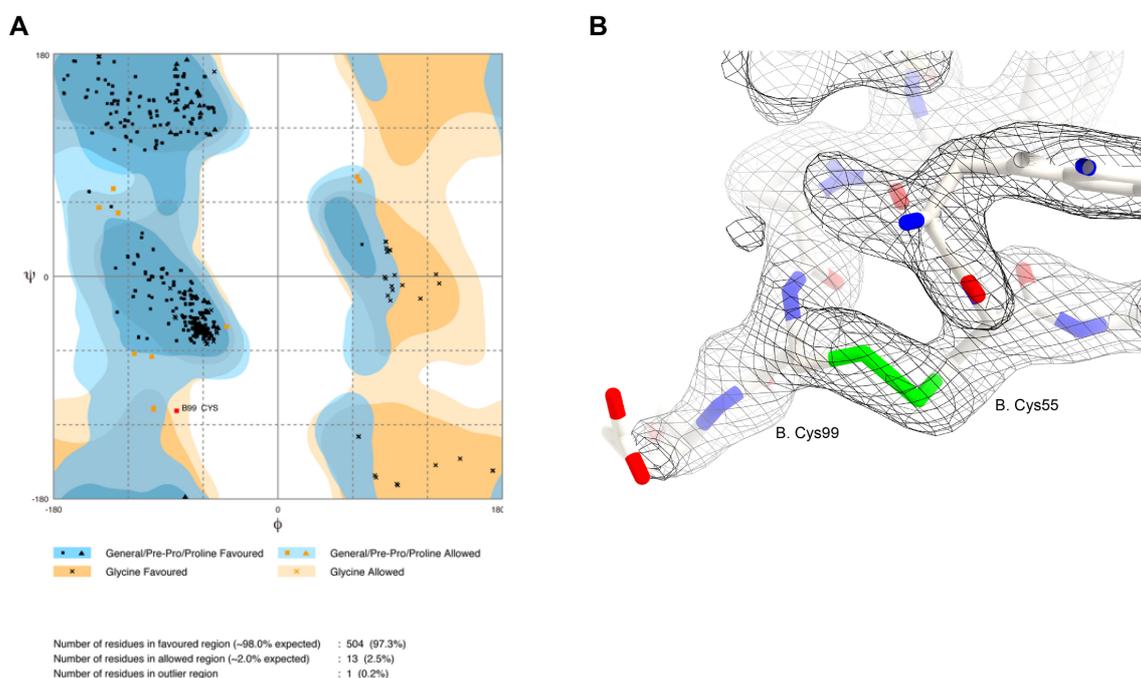
(A) The *AtTae4-AtTai4* crystallization construct.

(B) SDS-PAGE analysis of purified *AtTae4-AtTai4* with molecular-weight markers (labelled in kDa).



**Figure 2-9. The crystal and X-ray diffraction image of the *AtTae4-AtTai4* complex.**

(A) The crystal of the *AtTae4-AtTai4* complex. (B) X-ray diffraction image of the *AtTae4-AtTai4* complex.



**Figure 2-10. Ramachandran analysis of the *AtTae4-AtTai4* complex structure.**

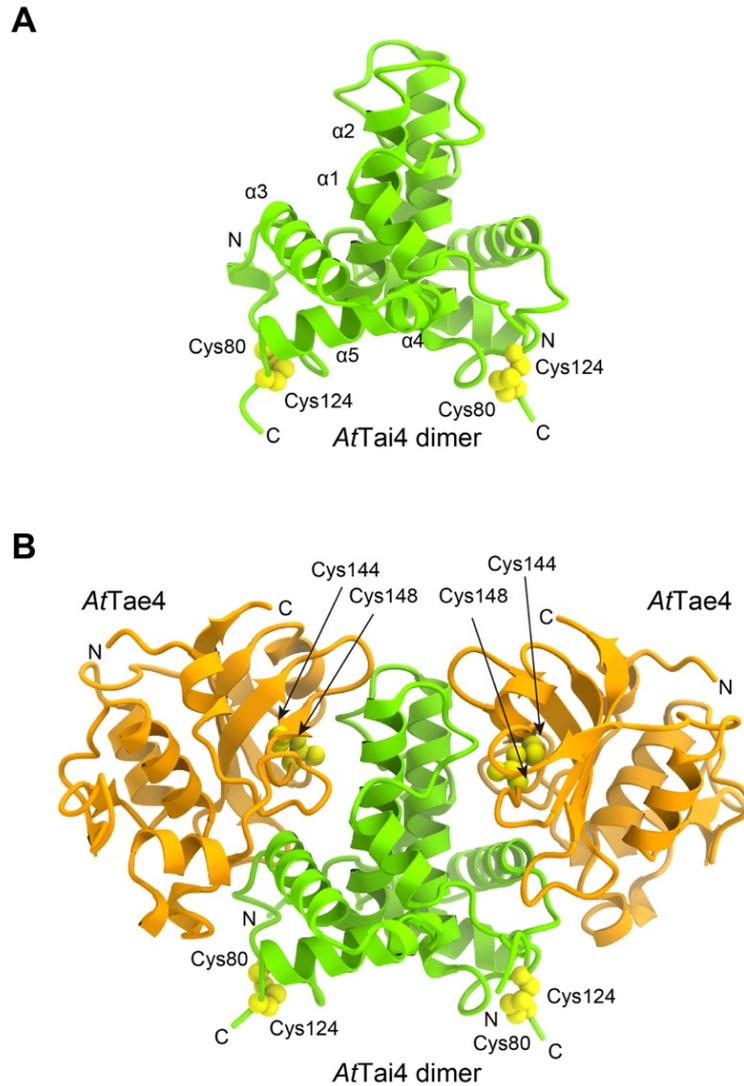
(A) Ramachandran plots of the *AtTae4-AtTai4* complex structure. (B) The residue in the outlier region (B. Cys99) and interacting residue (B. Cys 55) are shown as sticks. A disulfide bond is formed between B. Cys 55 and B. Cys99.  $2F_o - F_c$  electron density map (contoured at  $1\sigma$ ) is shown in black mesh. B. Cys 55 and B. Cys99 correspond to Cys80 and Cys124 in Figure 2-11, respectively.

**Table 2-1 Data collection and refinement statistics**

	<i>AtTai4</i>	<i>AtTae4-AtTai4</i> complex
<b>Data collection</b>		
Beamline	SPring-8 BL41XU	SPring-8 BL32XU
Wavelength (Å)	1.0000	1.0000
Crystal-detector distance (mm)	300	200
Rotation range per image (°)	0.5	0.5
Exposure time per image (s)	0.5	1.0
Oscillation range per crystal (°)	180	180
Helical translation step (µm)	0.5	0.3
Number of crystals	1	1
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 6 <sub>1</sub>
Cell dimensions (Å)	<i>a</i> =53.92, <i>b</i> =57.76, <i>c</i> =71.47	<i>a</i> = <i>b</i> =72.03, <i>c</i> =194.35
Resolution (Å)	53.92–1.55 (1.58–1.55)	97.18–1.90 (1.94–1.90)
<i>R</i> <sub>pim</sub>	0.029 (0.313)	0.022 (0.544)
< <i>I</i> /σ( <i>I</i> )>	13.0 (2.3)	15.0 (1.3)
Completeness (%)	99.5 (94.7)	99.6 (96.3)
Multiplicity	6.2 (4.2)	10.1 (7.3)
CC <sub>1/2</sub>	0.998 (0.807)	0.999 (0.528)
Mosaicity (°)	0.12	0.24
<b>Refinement</b>		
Resolution (Å)	53.92–1.55	97.18–1.90
No. reflections	33,087	44,759
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.1742/0.1975	0.1903/0.2125
No. atoms		
Protein	1546	4050
Ligand	28	21
Solvent	141	122
Average <i>B</i> -factors (Å <sup>2</sup> )		
Protein	26.9	51.5
Ligand	48.4	81.2
Solvent	35.2	51.3
R.m.s. deviations		
Bond lengths (Å)	0.013	0.0089
Bond angles (°)	1.79	1.57
Ramachandran plot		
Favored (%)	97.45	96.91
Allowed (%)	2.55	2.90
Outlier (%)	0	0.19

### 3.3.2 Overall structure of *AtTai4* and the *AtTae4-AtTai4* complex

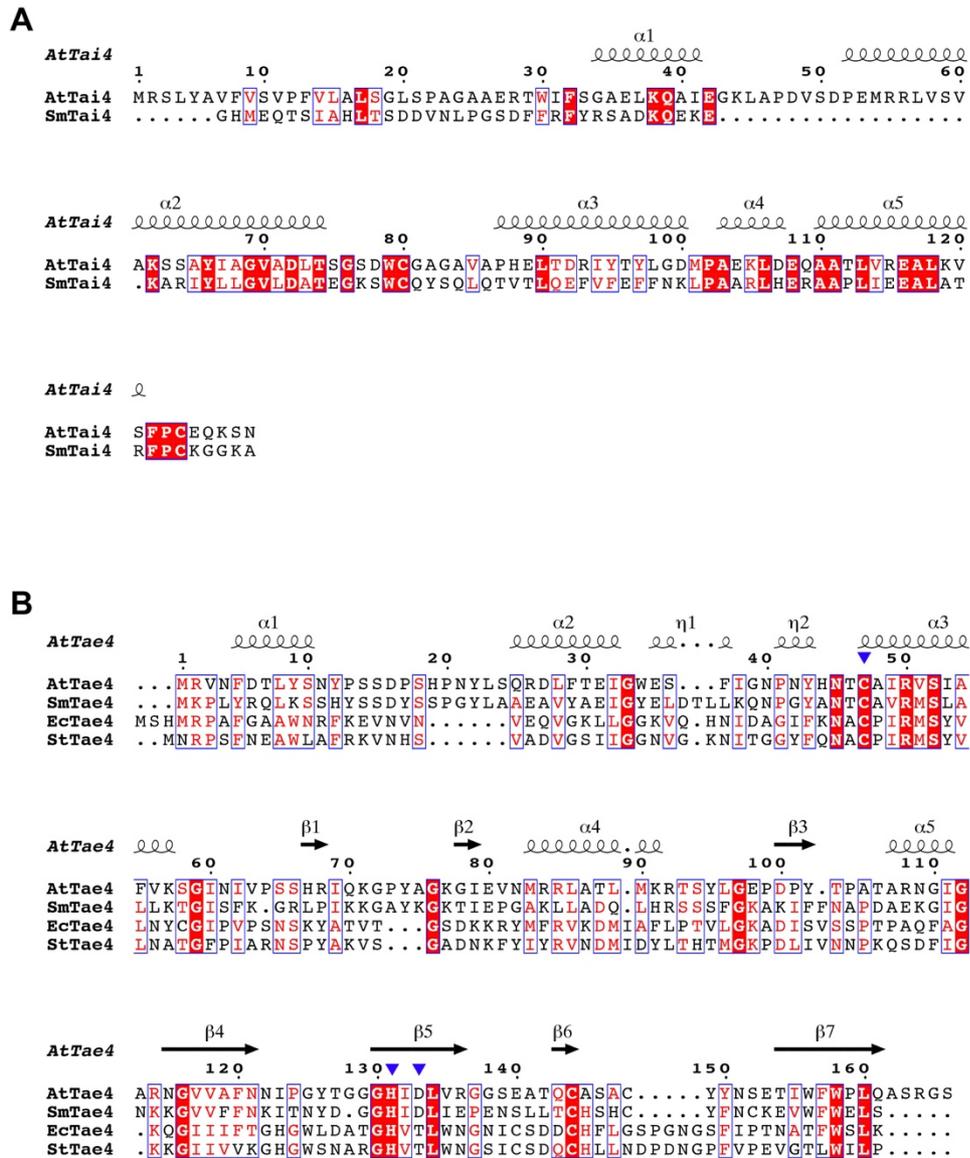
The crystal structure of *AtTai4* was determined at 1.55 Å resolution. *AtTai4* forms a homo-dimer composed of five  $\alpha$ -helices ( $\alpha$ 1–5) (Fig. 2-11A). The  $\alpha$ 2 helix (residues 52–74) contributes to dimer formation in the asymmetric unit. A disulfide bond is formed between Cys80 and Cys124 in each protomer (Fig. 2-11A). In addition, the author determined the crystal structure of the *AtTae4-AtTai4* complex at 1.9 Å resolution (Fig. 2-11B). The structure revealed that the *AtTai4* dimer binds two *AtTae4* molecules to form a hetero-tetramer in the asymmetric unit. The crystal structure revealed that *AtTae4* forms an intra-molecular disulfide bond between Cys144 and Cys148 (Fig. 2-11B). The superimposition of *AtTai4* alone and *AtTai4* bound to *AtTae4* resulted in a root-mean-square deviation (r.m.s.d.) value of 0.8 Å.



**Figure 2-11. Crystal structures of the *AtTai4* homo-dimer and the *AtTae4-AtTai4* complex.** (A) Structural overview of the *AtTai4* homo-dimer. The intramolecular disulfide bonds between Cys80 and Cys124 are shown as spheres (yellow). (B) Overall structure of the *AtTae4-AtTai4* complex. The intramolecular disulfide bonds between Cys144 and Cys148 are shown as spheres (yellow).

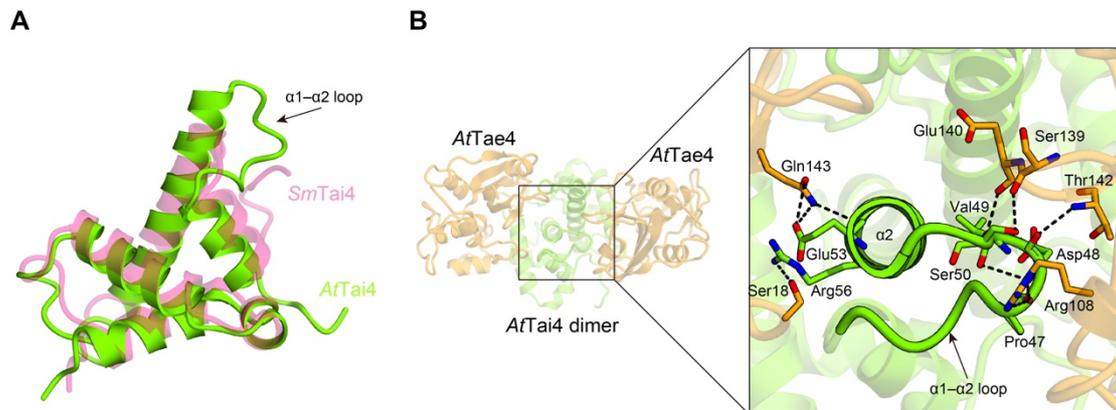
### 3.3.3 Structure comparison

A search for structural homologues was conducted using the DALI server (Holm & Laakso, 2016). The top-scoring structural homologue of *AtTai4* was the Rap1a protein from *S. marcescens* (*SmTai4*) (PDB code 3ZFI, Srikannathasan *et al.*, 2013), with a Z-score of 13.5, and an r.m.s.d. of 1.7 Å. The structural homologues of *AtTae4* are the following proteins: the Ssp1 protein from *S. marcescens* (*SmTae4*) (PDB code 4BI3, Srikannathasan *et al.*, 2013), the Tae4 protein from *E. cloacae* (*EcTae4*) (PDB code 4HFK, Zhang, Gao, Wang *et al.*, 2013), and the Tae4 protein from *S. typhimurium* (*StTae4*) (PDB code 4J30, Benz *et al.*, 2013). The most similar structural homologue was the *SmTae4* protein, with a Z-score of 25.4 and an r.m.s.d. of 1.6 Å. The amino-acid sequence alignments of *AtTai4* and *AtTae4* with their homologues are shown in Fig. 2-12 A and B. *AtTai4* shares 32.3% amino acid sequence identity with *SmTai4*. *AtTae4* shares 41.5%, 21.9%, and 20.5% sequence identities with *SmTae4*, *StTae4*, and *EcTae4*, respectively.



**Figure 2-12. Structure-based sequence alignments of *AtTai4* and *AtTae4* with their homologues, performed with Clustal Omega and ESPrIPT3. (A) Sequence alignment of *AtTai4* with *SmTai4* (PDB code 3ZFI, Srikannathasan *et al.*, 2013). (B) Sequence alignment of *AtTae4* with *SmTae4* (PDB code 4BI3, Srikannathasan *et al.*, 2013), *EcTae4* (PDB code 4HFK, Zhang, Gao, Wang *et al.*, 2013), and *StTae4* (PDB code 4J30, Benz *et al.*, 2013). The potentially catalytic triads, Cys47, His131, and Asp133, are indicated by blue triangles.**

A comparison of *AtTai4* with *SmTai4* revealed that *AtTai4* contains a longer  $\alpha 2$  helix and a longer loop between the  $\alpha 1$  and  $\alpha 2$  helices (Fig. 2-13A). As the longer  $\alpha 2$  helix and the loop between the  $\alpha 1$  and  $\alpha 2$  helices interact with two *AtTae4* molecules in the asymmetric unit (Fig. 2-13B), *AtTae4* and *AtTai4* have structurally distinct interactions as compared to *SmTae4-SmTai4* complex. In addition, neither *StTae4-StTai4* complex nor *EcTae4-EcTai4* complex have these structurally distinct interactions. Glu53 and Arg56 in the  $\alpha 2$  helix form hydrogen bonds with Gln143 and Ser18 in one of the *AtTae4* molecules in the asymmetric unit, respectively (Fig. 2-13B). Pro47, Asp48, and Val49 in the loop of *AtTai4* form hydrogen bonds with Arg108, Thr142, and Arg108 in the other *AtTae4* protomer in the asymmetric unit, respectively. In addition, Ser50 in the loop of *AtTai4* interacts with Ser139 and Glu140 in *AtTae4* (Fig. 2-13B).

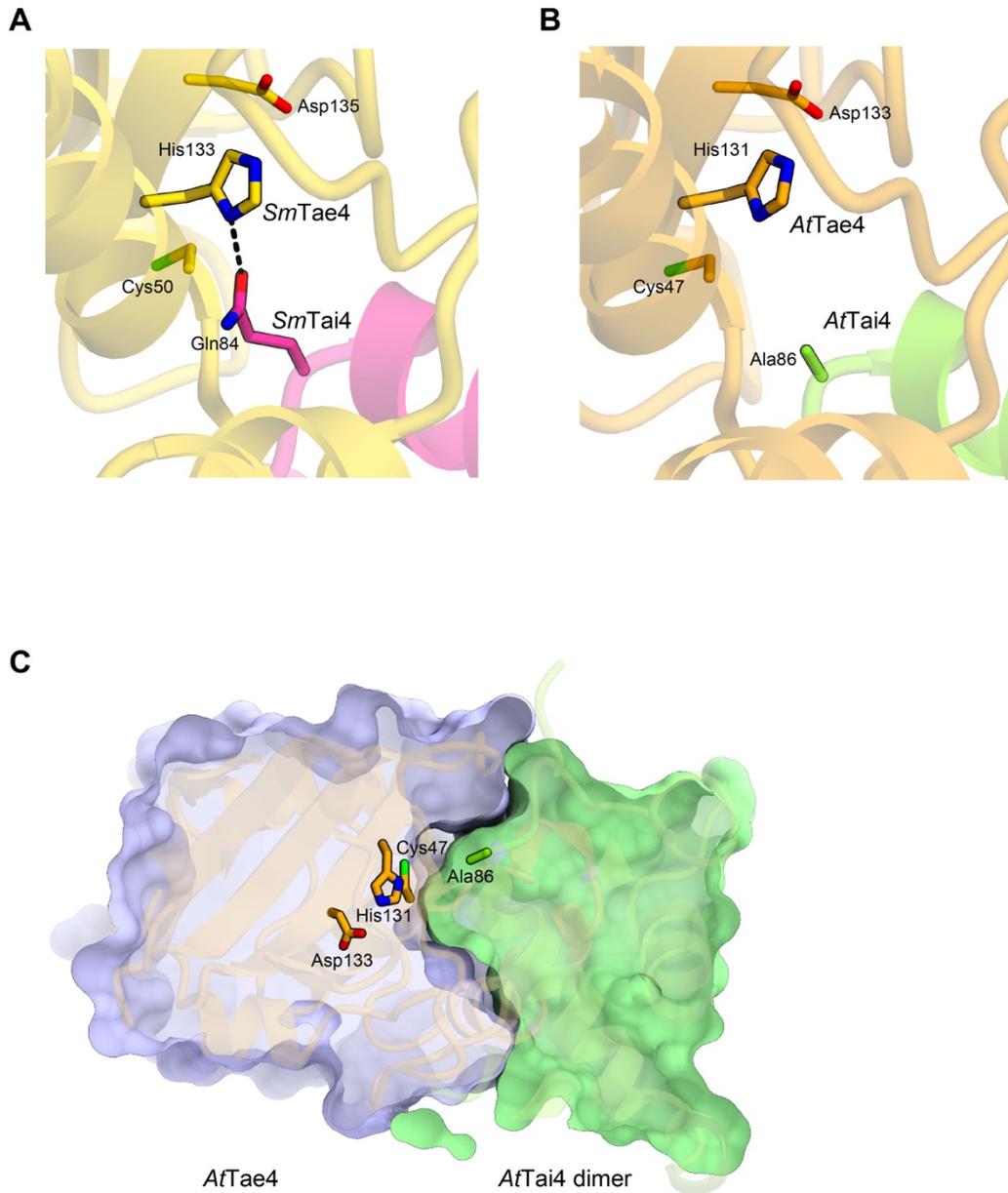


**Figure 2-13. Structural comparison and interactions**

(A) Structural comparison of the *AtTai4* with the *SmTai4*. A superimposition of *AtTai4* and *SmTai4* indicated that *AtTai4* contains extensions in the  $\alpha 2$  helix and the loop between the  $\alpha 1$  and  $\alpha 2$  helices. (B) Interacting residues between *AtTai4* and *AtTae4*.

The *SmTae4-SmTai4* complex structure revealed that *SmTai4* locates at the entrance of the active site of *SmTae4* and blocks substrate access to the active site (Srikannathasan *et al.*, 2013). The catalytic Gln84 of *SmTai4* forms a hydrogen bond with His133 of *SmTae4* and blocks the active site (Fig. 2-14A) (Srikannathasan *et al.*, 2013). The amino-acid sequence alignment of *SmTai4* and *AtTai4* showed that Gln84 of *SmTai4* is not conserved, and is replaced with Ala86 in *AtTai4* (Fig. 2-12A). In the present structure of the *AtTae4-AtTai4* complex, although the potentially catalytic His131 residue of *AtTae4* does not interact with any residues of *AtTai4* (Fig. 2-14B),

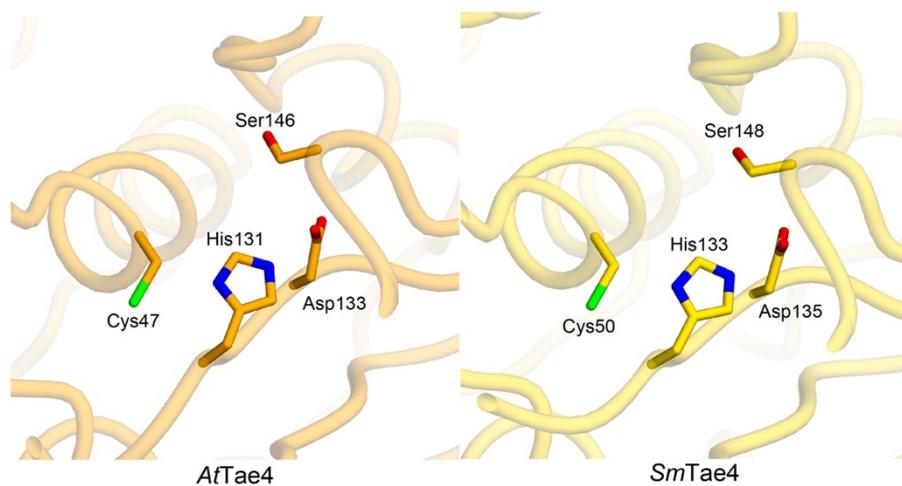
*AfTai4* blocks the entrance of the substrate-binding pocket of *AfTae4* and prevents substrate access to the active site (Fig. 2-14C).



**Figure 2-14. Structural comparison of the *AtTae4-AtTai4* complex with the *SmTae4-SmTai4* complex.** (A) The crystal structure of the *SmTae4-SmTai4* complex revealed that Gln84 of *SmTai4* interacts with His133 of *SmTae4* and blocks the active site (PDB code 4BI8, Srikannathan *et al.*, 2013). (B) The crystal structure of the *AtTae4-AtTai4* complex lacks the interaction between the expected catalytic His131 of *AtTae4* and the corresponding residue of *AtTai4*. The Gln is not conserved in *AtTai4* and is replaced by Ala in *AtTai4*. (C) *AtTai4* homo-dimer is positioned close to the *AtTae4* active site surfaces and may block the substrate binding.

### 3.3.4 Catalytic site

The Tae4 family proteins have the conserved catalytic residues (Cys-His-Asp) responsible for their peptidoglycan amidase activities. In the present structure of the *AfTae4-AfTai4* complex, Cys47, His131, and Asp133 also form a catalytic triad, as in the *SmTae4-SmTai4* complex structure (Fig. 2-15).



**Figure 2-15. Structural comparison in the catalytic triad.**

### 3.4 Discussion

#### 3.4.1. Protein expression

The *AtTai4* and *AtTae4-AtTai4* complex successfully expressed in the cytoplasm of *E. coli* Rosetta 2 (DE3) cells (Fig. 2-4B, Fig. 2-8B). The reason for this success is that *AtTae4* cleaves the  $\gamma$ -D-glutamyl-*m*DAP bond, which presents in the bacteria's cell wall peptidoglycan (Russell *et al.*, 2012) and does not demonstrate toxicity in the cytoplasm of *E. coli* cells. In contrast, the protein expression of *AtTde-AtTdi* complex had failed. Possible reason for this failure is that the interaction between *AtTde* and *AtTdi* was unstable and the collapse of the interaction occurred. As a result, *AtTde* might be digested *E. coli* DNA and induce cell death.

#### 3.4.2. Overall structure

In the crystal structure of *AtTai4*, the  $\alpha 2$  helix (residues 52–74) contributes to dimer formation in the asymmetric unit, consistent with the size-exclusion chromatography results indicating that *AtTai4* exists as a dimer in solution. The *AtTae4-AtTai4* complex forms a hetero-tetramer in the asymmetric unit, consistent with the size-exclusion chromatography results indicating that the *AtTae4-AtTai4* complex exists

in a hetero-tetrameric form in solution. The *AtTae4* forms an intra-molecular disulfide bond between Cys144 and Cys148, which may confer structural stability (Fig. 2-11B). The superimposition of *AtTai4* alone and *AtTai4* bound to *AtTae4* resulted in a root-mean-square deviation (r.m.s.d.) value of 0.8 Å, indicating that no major structural changes occur upon complex formation.

### 3.4.3 Structure comparison

The *AtTai4* forms homodimer and has similar structure to that of *SmTai4* (Fig. 2-11A, Fig. 2-13A) (Srikannathasan *et al.*, 2013). In the *AtTae4-AtTai4* complex structure, one of the *AtTai4* in the asymmetric unit blocks the entrance of the substrate-binding pocket of the *AtTae4*, the other of the *AtTai4* in the asymmetric unit stabilizes interaction between one of the *AtTai4* and the *AtTae4* in the asymmetric unit (Fig. 2-13B) (Fig. 2-14C). As *AtTai4* neutralizes and stabilizes two opposing *AtTae4*s, *AtTai4* binds *AtTae4* with 2:2 stoichiometry.

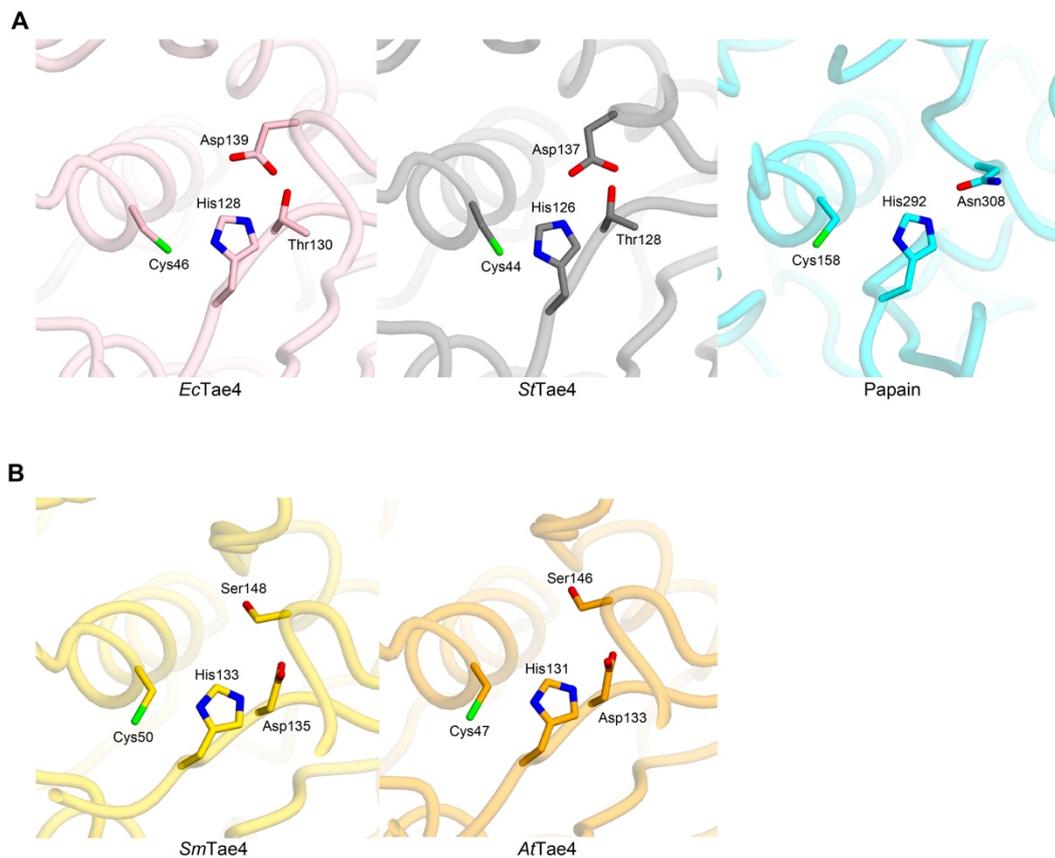
Previous studies suggest that both *AtTai4* and *SmTai4* can neutralize the activity of *AtTae4* and *SmTae4*, respectively. The morphological abnormality mediated by *SmTae4* was neutralized by *SmTai4* (Srikannathasan *et al.*, 2013). The growth inhibition

of *E. coli* DH10B exerted by expression of *AtTae4* was rescued by co-expression of the *AtTai4* (Ma *et al.*, 2014). While there is no direct interaction between the potentially catalytic His131 of *AtTae4* and Ala86 of *AtTai4*, which corresponds to Gln84 of *SmTai4*, the structural comparison, combined with previous studies, suggest that *AtTai4* effectively neutralize the activity of *AtTae4* by blocking the entrance of the substrate-binding pocket of *AtTae4* (Fig. 2-14).

#### 3.4.4 Catalytic site

The *Tae4* family proteins have the conserved catalytic residues (Cys-His-Asp) responsible for their peptidoglycan amidase activities. Cys46, His128, and Asp139 of *EcTae4* and Cys44, His126, and Asp137 of *StTae4* form the catalytic triads, which are similar to the canonical catalytic triad in the papain-like cysteine peptidase (PDB code 1BP4, LaLonde *et al.*, 1998) (Zhang, Gao, Wang *et al.*, 2013; Zhang, Gao, Wei *et al.*, 2013; Benz *et al.*, 2013) (Fig. 2-16A). *SmTae4* also has the catalytic triad formed by Cys50, His133, and Asp135. While Asp139 of *EcTae4* and Asp137 of *StTae4* are replaced with Ser148 in *SmTae4*, Asp135 of *SmTae4*, which corresponds to Thr130 of *EcTae4* and Thr128 of *StTae4*, is located at a position similar to those of Asp139 of

*EcTae4* and Asp137 of *StTae4* in the *SmTae4* structure (Fig. 2-16A and B). These observations suggested that Asp135 serves as the 3rd Asp residue in the catalytic triad in *SmTae4* (Srikannathasan *et al.*, 2013). In the present structure of the *AtTae4-AtTai4* complex, Cys47, His131, and Asp133 also form a catalytic triad, as in the *SmTae4-SmTai4* complex structure (Fig. 2-16B). Thus, the present structure reinforces the idea that the *Tae4* family proteins have two types of structurally distinct catalytic triads.



**Figure 2-16. Structural differences in the catalytic triad.** (A) *EcTae4* and *StTae4* have the conserved catalytic active center containing the catalytic residues (Cys-His-Asp), which have a similar arrangement to the catalytic triad of papain (PDB code 1BP4, LaLonde *et al.*, 1998). (B) *SmTae4* and *AtTae4* have the conserved catalytic residues (Cys-His-Asp), but the 3rd Asp in the catalytic triad has a distinct spatial arrangement.

### 3.5 Conclusion

In this work, the author determined the crystal structures of *AtTai4* and the *AtTae4-AtTai4* complex. Comparisons of these structures with the homologous proteins revealed that the *AtTae4-AtTai4* complex shares structural similarity with the *SmTae4-SmTai4* complex. A structural comparison of *AtTai4* with *SmTai4* showed that *AtTai4* contains more extended helices and loops, which may enforce the interaction between *AtTai4* and the adjacent *AtTae4*. A structural superimposition highlighted the differences in the spatial arrangement of the Asp residue in the catalytic triad residues (Cys-His-Asp), among the Tae4 family proteins. The present structures enhance our understanding of the catalytic and inhibitory mechanisms of the Tae4 and Tai4 family proteins.

## Chapter4: General Discussion

In the chapter2, DmZuc was suggested to be an endonuclease that randomly cleaves single-stranded RNA. In contrast, Zuc has been implicated in phased piRNA biogenesis in *D. melanogaster* and may preferentially cleave uridine (Han *et al.*, 2015; Mohn *et al.*, 2015). Although purified recombinant Zuc has no strict base specificity, it is quite interesting that Zuc has uridine preference *in vivo*. The major difference between purified recombinant Zuc and Zuc present in cells is the absence or presence of N-terminal mitochondrial localization sequence (MLS). Given this situation, the localization of Zuc to mitochondria may possibly influence the sequence preference of Zuc. Recent paper suggested that Armitage (Armi), an RNA-binding ATPase, shuttles back and forth between nuage and mitochondria to supply precursor piRNAs to Zuc located in the mitochondria (Ge *et al.*, 2018). Thus, purified recombinant Zuc, which has no MLS does not prefer uridine as the substrate nucleotide; however, when Zuc localizes in mitochondria and interacts with other proteins such as Armi, it may preferentially cleave uridine as the substrate nucleotide.

The development of cryo-electron microscope and single particle analysis

method have been remarkable in the field of structural biology. However, as the molecular weight of the Zuc dimer is only around 56 kDa, it is expected that the structure analysis of Zuc dimer by cryo-electron microscopy is difficult with current technology. However, as mentioned above, if Zuc interacts with other proteins to form protein complexes with large molecular weights, the complexes may be suitable as a target for structural analysis using cryo-electron microscope. It is expected that future studies will elucidate how Zuc interacts with other proteins and how sequence specificity of Zuc are generated.

In the chapter3, there are also unclear points about the Tae4-Tai4 complex from *A. tumefaciens*. First, it is unknown how the mutations of residues affect the complex stability. In the *StTae4-StTai4* complex, the mutations in the *StTai* decreases *StTae4-StTai4* interaction (Benz *et al.*, 2013). Therefore, mutational analysis is necessary for the *AtTae4-AtTai4* complex. Second, previous studies showed that *EcTai4* interacts with *StTae4*. In addition, the structure of *StTae4-EcTai4* complex was elucidated (Zhang *et al*, 2013). Therefore, *EcTai4* neutralizes not only endogenous *EcTae4* but also the toxicity of *StTae4*, which has structural similarity to *EcTae4*. This

result suggested that *E. cloacae* could protect against attack from other bacteria such as *S. typhimurium* (Zhang *et al.*, 2013). However, it is unknown whether cross-immunity mechanism between *S. typhimurium* and *E. cloacae* can also exist between *A. tumefaciens* and *S. marcescens*. There is room for further study as to whether immunities derived from *A. tumefaciens* and *S. marcescens* can neutralize effectors derived from *S. marcescens* and *A. tumefaciens*, respectively. When complexes such as *AtTae4-SmTai4* or *SmTae4-AtTai4* can be formed, the author believes that the structural analysis reveals the molecular basis of these complexes.

The strict management of enzymatic activity is necessary for living cells. In the chapter 2, purified recombinant DmZuc has no strict base specificity for single stranded RNA *in vitro*, and even has DNase activity (Fig. 1-11 B, Fig. 1-12 A). However, as described above, it is expected that enzymatic activity of Zuc *in vivo* is controlled so as to have ribonuclease activity specific only to piRNA, due to localization in mitochondria and interaction with other proteins. In the chapter 3, although recombinant Tae4 has enzymatic activity to cleave peptidoglycan *in vitro*, its amidase activity is properly controlled in *A. tumefaciens* cells (Durand *et al.*, 2014; Russell *et al.*, 2014; Ma

*et al.*, 2014). The enzymatic activity of Tae4 is suppressed by interacting with Tai4 in the periplasm where peptidoglycan exists, though Tae4 exists alone in the cytoplasm in which peptidoglycan does not exist (Durand *et al.*, 2014; Russell *et al.*, 2014; Ma *et al.*, 2014). Similar to Tae4, the *A. tumefaciens* is also expected to have a mechanism of suppressing the *AtTde*'s nuclease activity such as interacting with *AtTdi*. However, it was difficult to co-express *AtTde* and *AtTdi* in the cytoplasm of *E. coli*, suggesting that the interaction between *AtTde* and *AtTdi* is weak. Therefore, the *AtTde-AtTdi* complex is expected to be stabilized by other unknown proteins in the *A. tumefaciens* cytoplasm. In conclusion, through the research conducted in the chapter 2 and the chapter 3, the author reaffirmed the importance of adjustment of the enzymatic activity in the cell.

In summary, the author revealed structures of protein Zuc and *AtTae4-AtTai4* complex in this dissertation. As mentioned in the first general introduction, research on self-defense mechanisms can contribute to advances in research in various fields. I would like to hope that the findings of this dissertation contribute to the development of research in related fields.

## References

- Adams, P. D. *et al.* (2002). *Acta Crystallogr. D* **58**, 1948–1954.
- Beck, G. & Habitat, G. S. (1996). *Scientific American*. **275** (5), 60–66.
- Belancio, V. P., Hedges, D. J. & Deininger, P. (2008). *Genome Research*. **18**(3), 343-358.
- Benz, J., Reinstein, J. & Meinhart, A. (2013). *PLoS One*, **8**(6), e67362.
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. & Hannon, G. J. (2007). *Cell*, **128**, 1089–1103.
- Bryson, K. *et al.* (2005) *Nucleic Acids Res.* **33**, 36–38.
- Choi, S.-Y., Huang, P., Jenkins, G. M., Chan, D. C., Schiller, J. & Frohman, M. A. (2006). *Nature Cell Biol.* **8**, 1255–1262.
- Chung, W. J., Okamura, K., Martin, R., & Lai, E. C. (2008). *Current Biology*. **18** (11), 795–802.
- de La Fortelle, E. & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Durand, E., Cambillau, C., Cascalas, E. & Journet, L. (2014). *Trends in Microbiology*. **22**, 498-507.

Ge, D. T., Wang, W., Tipping, C., Gainetdinov, I., Weng, Z. & Zamore, P. D. (2018).

*bioRxiv*. doi: <https://doi.org/10.1101/445825>

Gouet, P., Courcelle, E., Stuart, D. I. & Me'toz, F. (1999). *Bioinformatics*, **15**, 305–308.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr. D* **60**, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D62**, 72–82.

Gottlin, E. B., Rudolph, A. E., Zhao, Y., Matthews, H. R. & Dixon, J. E. (1998). *Proc. Natl Acad. Sci. USA* **95**, 9202–9207.

Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y.,

Nagami, T., Siomi, H. & Siomi, M. C. (2007). *Science*, **315**, 1587–1590.

Haase, A. D., Fenoglio, S., Muerdter, F., Guzzardo, P. M., Czech, B., Pappin,

D. J., Chen, C., Gordon, A. & Hannon, G. J. (2010). *Genes Dev.* **24**, 2499–2504.

Han, B. W., Wang, W., Li, C., Weng, Z. & Zamore, P. D. (2015). *Science*. **348**, 817-821.

Hayashi, K. & Kojima, C. (2008). *Protein Expr. Purif.* **62**, 120–127.

- Holm, L. & Laakso, L. M. (2016). *Nucleic Acids Res*, **44**, 351-355.
- Hood, R. D., Singh, P., Hsu, F. S., Güvener, T., Carl, M. A., Trinidad, R. R. S., Silverman, J. M., Ohlson, B. B., Hicks, K. G., Plemel, R. L., Li, M., Schwarz, S., Wang, W. Y., Merz, A. J., Goodlett, D. R. & Mougous, J. D. (2010). *Cell Host Microbe*, **7**, 25–37.
- Huang, H., Gao, Q., Peng, X., Choi, S.-Y., Sarma, K., Ren, H., Morris, A. J. & Frohman, M. A. (2011). *Dev. Cell*, **20**, 376–387.
- Ipsaro, J., Haase, A. D., Knott, S. R., Joshua-Tor, L. & Hannon, G. J. (2012). *Nature*, **491**, 279–283.
- Kazazian, H. H. & Goodier, J. L. (2002). *Cell*, **110**, 277–280.
- Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G. & Antonarakis, S. E. (1988). *Nature*, **332**, 164–166.
- Kelley, L. A. & Sternberg, M. J. E. (2009). *Nature Protoc.* **4**, 363–371.
- LaLonde, J. M., Zhao, B., Smith, W. W., Janson, C. A., DesJarlais, R. L., Tomaszek, T. A., Carr, T. J., Thompson, S. K., Oh, H. J., Yamashita, D. S., Veber, D. F. & Abdel-Meguid, S. S. (1998). *J. Med. Chem*, **41**, 4567-4576.

- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
- Leiros, I., Secundo, F., Zambonelli, C., Servi, S. & Hough, E. (2000). *Structure* **8**, 655–667.
- Li, C. *et al.* (2009). *Cell*, **137**, 509–521.
- Litman, G., Cannon, J. & Dishaw, L. (2005). *Nat Rev Immunol.* **5**(11), 866-879.
- Lovell, S. C., Davis, I. W., Bryan, W., III, A., De Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437-440.
- Ma, L. S., Hachani, A., Lin, J. S., Filloux, A. & Lai, E. M. (2014). *Cell Host Microbe*. **16**, 94–104.
- MacIntyre, D. L., Miyata, S. T., Kitaoka, M. & Pukatzki, S. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 19520–19524.
- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R. & Hannon, G. J. (2009). *Cell*, **137**, 522–535.
- Malone, C. D. & Hannon, G. J. (2009). *Cell*, **136**, 656–668.

- McCLINTOCK, B. (1950). *Proc. Natl Acad. Sci. USA*, **36**(6), 344–355.
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., Vogelstein, B. & Nakamura, Y. (1992). *Cancer Research*, **52**, 643–645.
- Mohn, F., Handler, D. & Brennecke, J. (2015). *Science*. **348**, 812-817.
- Murdoch, S. L., Trunk, K., English, G., Fritsch, M. J., Pourkarimi, E. & Coulthurst, S. J. (2011). *J. Bacteriol.* **193** 6057-6069.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355-367.
- Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K. & Brennecke, J. (2010). *EMBO J.* **29**, 3301–3317.
- Pancer, Z. & Cooper, M. (2006). *Nat Rev Immunol*, **5**(11), 866-879.
- Pane, A., Wehr, K. & Schüpbach, T. (2007). *Dev. Cell*, **12**, 851–862.
- Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nat. Methods*, **8**, 785-786.
- Pillai, R. S. & Chuma, S. (2012). *Dev. Growth Differ.* **54**, 78–92.
- Pohlman, R. F., Franklin, L., Lu, W., More, I. M., & Stephen, C. W. (1993). *Nucleic*

*Acids Res.* **21**, 4867–4872.

Ponting, C. P. & Kerr, I. D. (1996). *Protein Sci.* **5**, 914–922.

Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D. & Mekalanos, J. J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 15508–15513.

Qi, H., Watanabe, T., Ku, H.-Y., Liu, N., Zhong, M. & Lin, H. (2011). *J. Biol.*

*Chem.* **286**, 3789–3797.

Raae, A. J., Kleppe, R. K. & Kleppe, K. (1975). *Eur. J. Biochem.* **60**, 437–443.

Rotondo, G. & Frendewey, D. (1996). *Nucleic Acids Res.* **24**, 2377–2386.

Russell, A. B., Hood, R. D., Bui, N. K., Leroux, M., Vollmer, W. & Mougous, J. D. (2011). *Nature (London)*, **109**, 19804–19809.

Russell, A. B., Singh, P., Brittnacher, M., Bui, N. K., Hood, R. D., Carl, M. A., Agnello, D. M., Schwarz, S., Goodlett, D. R., Vollmer, W. & Mougous, J. D. (2012). *Cell Host Microbe*, **11**, 538-549.

Russell, A. B., Peterson, S. B. & Mougous, J. D (2014). *Nature Rev. Microbiology.* **12**, 137-148.

Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani,

- H., Asai, K., Siomi, H. & Siomi, M. C. (2009). *Nature (London)*, **461**, 1296–1299.
- Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., Siomi, H. & Siomi, M. C. (2010). *Genes Dev.* **24**, 2493–2498.
- Schwarz, S., Hood, R. D. & Mougous, J. D. (2010). *Trends Microbiol*, **18**, 531–537.
- Senti, K. A. & Brennecke, J. (2010). *Trends Genet.* **26**, 499–509.
- Sheldrick, G. M. (2008). *Acta Crystallogr. A* **64**, 112–122.
- Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. (2011). *Nature Rev. Mol. Cell Biol.* **12**, 246–258.
- Srikannathasan, V., English, G., Bui, N. K., Trunk, K., O'Rourke, P. E. F., Rao, V. A., Vollmer, W., Coulthurst, S. J. & Hunter, W. N. (2013). *Acta Cryst.* **D69**, 2468–2482.
- Stuckey, J. A. & Dixon, J. E. (1999). *Nature Struct. Biol.* **6**, 278–284.
- Sun, W., Samimi, H., Gamez, M., Zare, H. & Frost, B. (2018). *Nature Neuroscience*, **21** (8), 1038–1048.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Crystallogr. D* **55**, 849–861.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.

- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. & Zamore, P. D. (2006).  
*Science*, **313**, 320–324.
- Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004)  
*Bioinformatics*. **20**, 2138–2139.
- Watanabe, T., Chuma, S., Yamamoto, Y., Kuramochi-Miyagawa, S., Totoki, Y.,  
Toyoda, A., Hoki, Y., Fujiyama, A., Shibata, T., Sado, T., Noce, T., Nakano,  
T., Nakatsuji, N., Lin, H. & Sasaki, H. (2011). *Dev. Cell*, **20**, 364–375.
- Waterman, D. G., Winter, G., Parkhurst, J. M., Fuentes-Montero, L., Hattne, J.,  
Brewster, A., Sauter, N. K. & Evans, G. (2013). CCP4 Newsletter on Protein  
Crystallography 16–19.
- Zhang, H., Gao, Z.-Q., Wei, Y., Xu, J.-H. & Dong, Y.-H. (2013). *PLoS One*, **8**(9),  
e73782.
- Zhang, H., Gao, Z. Q., Wang, W. J., Liu, G. F., Xu, J. H., Su, X. D. & Dong, Y. H.  
(2013). *J. Biol. Chem*, **288**, 5928–5939.
- Zhao, Y., Studkey, J. A., Lohse, D. L. & Dixon, J. E. (1997). *Protein Sci.* **6**, 2655–2658.

## Original papers

- 1) Fukuhara, S., Nishimasu, H., Bonnefond, L., Matsumoto, N., Ishitani, R. & Nureki, O. Expression, purification, crystallization and preliminary X-ray crystallographic analysis of Zucchini from *Drosophila melanogaster*. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **68**, 1346–1350 (2012).
- 2) Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M. K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., Ishitani, R., Siomi, H., Siomi, M. C. & Nureki, O. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491**, 284–287 (2012).
- 3) Fukuhara, S., Nakane, T., Yamashita, K., Ishii, R., Ishitani, R. & Nureki, O. Crystal structure of *Agrobacterium tumefaciens* Type VI effector-immunity complex. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **74**, 810–816 (2018).

## **Acknowledgements**

First, I would like to express my great appreciation to Professor Osamu Nureki for his supervision of my Ph.D. study and related research. Without his precious support it would not be possible to conduct this research. Besides Prof. Nureki, I would like to thank Dr. Ryuichiro Ishitani, Dr. Hiroshi Nishimasu, Dr. Keitaro Yamashita, Dr. Takanori Nakane, Dr. Ryohei Ishii, Dr. Tomohiro Nishizawa, Dr. Luc Bonnefond, and Dr. Naoki Matsumoto for technical advices and experimental supports. I would also thank to the other members in Nureki laboratory for critical comments and supports.

This study was performed in collaboration with many scientists. Notably, I would like to thank Professor Mikiko C. Siomi, Professor Haruhiko Siomi, and Professor Junken Aoki for the collaboration in functional analysis. I would also like to thank the beamline staff at BL41XU and BL32XU of SPring-8 for support during the data collection.

I am also grateful for the Department of Professional Development, Tokyo Medical and Dental University Hospital. My sincere thanks especially go to Dr. Makoto Takahashi, Dr. Eriko Okada, and Dr. Yasuhiro Itsui. They always supported me so that

both my junior resident program and graduate student life can be balanced.

The time I spent at The University of Tokyo and Tokyo Medical and Dental University was very fulfilling and enjoyable. I am grateful for all of my friends and former colleagues. In particular, I would like to thank Dr. Mitsugu Yanagidaira for encouraging me to strive towards my goal.

Last but not the least, I would like to thank my family. I would like to say a heartfelt thank you to my parents: Dr. Kiyoshi Fukuhara and Mrs. Ayano Fukuhara, for all of the supports and sacrifices throughout my life. I am deeply thankful to my dear wife Dr. Mari Fukuhara. Thank you for supporting me for everything. To my beloved daughter Emma Fukuhara, I would like to express my thanks for being a good girl and always cheering me up.

Thank you.

Satoshi Fukuhara

The University of Tokyo

December 2018