学位論文

Molecular basis of ligand recognition by cell surface proteins of higher eukaryotes
(高等真核生物における細胞表面タンパク質のリガンド認識の分子基盤)

平成 29 年 12 月博士（理学）申請

東京大学大学院理学系研究科
生物科学専攻
森田　純子

Abstract

Cells in multicellular orgaminsms communicate by means of numerous signaling molecules. In many cases, these molecules act as ligands that bind to the receptor molecule on the surface of the target cells. Elucidating the molecular mechanism of ligand recognition by these cell surface receptors helps us to understand biological phenomena of multicellular organisms. Here in this dissertation, the ligand recognition mechanism of cell surface proteins of higher eukaryotes is described using X-ray crystal structure analysis.

X-ray crystallographic analysis of ENPP6, choline-metabolizing enzyme

Choline is an essential nutrient for all living cells and is produced extracellularly by sequential degradation of phosphatidylcholine (PC). ENPP6, a choline-specific phosphodiesterase, hydrolyzes glycerophosphocholine (GPC), a degradation product of PC, as a physiological substrate and participates in choline metabolism. ENPP6 is highly expressed in liver sinusoidal endothelial cells and developing oligodendrocytes, which actively incorporate choline and synthesize PC. ENPP family proteins are conserved extracellular pyrophosphatase/phosphodiesterase in vertebrate, and involved in different biological processes by hydrolyzing pyrophosphate or phosphodiester bonds in extracellular compounds. To understand the molecular mechanism of how ENPP6 recognizes the choline moiety, I determined the crystal structure of ENPP6 in complex with phosphocholine at 1.8 Å resolution. The resulting structure revealed that the choline moiety of the phosphocholine is recognized by a choline-binding pocket formed by conserved aromatic and acidic residues. The present study provides the molecular basis for ENPP6-mediated choline metabolism at atomic, cellular and tissue levels. Furthermore, a structural comparison of ENPP6 and other ENPPs offers an explanation of why among the family members only ENPP6 specifically recognizes the choline moiety of substrates.

X-ray crystallographic analysis of the plant receptor-like kinase TDR in complex with the peptide hormone TDIF

In plants, leucine-rich repeat receptor-like kinases (LRR-RKs) perceive ligands, including peptides and small molecules, to regulate various physiological processes. TDIF, a member of the CLE peptide family, specifically interacts with the LRR-RK TDR to inhibit vascular meristem differentiation into tracheary elements, and promotes cell proliferation. Here we report the crystal structure of the extracellular domain of TDR in complex with the TDIF peptide. The extracellular domain of TDR adopts a superhelical structure comprising 22 LRRs, and specifically recognizes TDIF by its inner concave surface. Together with the biochemical and sequence analyses, the determined structure reveals a conserved TDIF-recognition mechanism of TDR among plant species. Furthermore, a structural comparison of TDR with other plant LRR-RKs suggested the activation mechanism of TDR by TDIF. The structure of this CLE peptide receptor provides insights into the recognition mechanism of the CLE family peptides.

Contents

List of Abbreviations

ABC	dimethylamine-borane
AMP	Adenosine monophosphate
Ap3A	Diadenosine triphosphate
ATP	Adenosine triphosphate
ATX	Autotaxin
BAK	BRI1-associated kinase
BAM	Barely any meristem
BES	BRI1-EMS suppressor
BIN	Brassinosteroid-insensitive
BRI	Brassinosteroid-insensitive
BSA	Bovine serum albumin
CFP	Cyan fluorescent protein
CLE	CLAVATA3/embryo surrounding region-related
CLV	CLAVATA
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Endo H	Endoglycosidase H
ENPP	Ecto-nucleotide pyrophosphate/phosphodiesterase
ER	Endoplasmic reticulum
ESI	Electrospray
FBS	Fetal bovine serum
FLS	Flagellin sensing
FRET	Fluorescence resonance energy
FUC	Fucose
GPC	Glycerophosphorylcholine
GPCR	G-protein coupled receptor
GPI	Plycosylphosphatidylinositol
GSK	Glycogen synthase kinase

HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRV3C	Human rhinovirus 3C
IR/MAR	Initiation region/matrix attachment region
LC-MS/MS	Liquid chromatography/tandem mass spectrometry
LPA	Lysophosphatidic acid
LPC	Lysophosphatidylcholine
LPDS	Lipoprotein-depleted serum
LRR-RK	Leucine-rich repeat receptor-kinase
MAN	Mannose
MES	2-(*N*-morpholino)ethanesulfonic acid
MWCO	Molecular-weight cutoff
NAG	*N*-Acetyl glucosamine
PAF	Platelet-activating factor
PAMPs	Pathogen-associated molecular pattern
PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDE	Phosphodiesterase
PE	Phosphatidylethanolamine
PEG	Polyethylene glycol
PEPR	PEP receptor
PSKR	Phytosulfokine receptor
PXL	PXY-like
PXY	Phloam intercalated with xylem
RMSD	Root mean square deviation
RTK	Receptor tyrosine kinase
SDS-PAGE	Sodium dodecyl sulfate poly-acrylamide gel electrophoresis
SERK	Somatic embryogenesis receptor kinase
Sf9	Spodoptera frugiperda 9

SKM	Sterility-regulating kinase member
SM	Sphingomyelin
SPC	Sphingosylphosphorylcholine
TDIF	Tracheary element differentiation factor
TDR	TDIF receptor
TEV	Tobacco etch virus
Tris	Tris (hydroxymethyl) aminomethane
UDP-GlcNAc	Uridine diphosphate *N*-acetylglucosamine
WOX	WUSCHEL-related homeobox
WT	Wild-type
YFP	Yellow fluorescent protein

Table of Amino Acid Abbreviations

1-letter	3-letter	Full name
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic aci
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Aspargine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

General Introduction

Biological functions and cell surface proteins in higher eukaryotes

In the course of evolution, multicellular organisms did not appear on the Earth until unicellular organism had already been in existence for about 2.5 billion years. This is accomplished by the ability of cells constituting the whole individual to specialize in different functions and cooperate each other. In multicellular organisms, the cells must communicate with each other, thereby controlling physiological functions as individuals. Such cell-to-cell communication is often mediated by extracellular signal molecules. These molecules include proteins, small peptides, amino acids, nucleotides, retinoids, fatty acid derivatives, and even dissolved gases. Many of these molecules are 'self' origin, but some are 'non-self' derived molecules including PAMPs from pathogens, which triggers host defense reactions such as immune responses. In many cases, these molecules are received by receptor molecules on the cell surface, which in turn activates intracellular signaling pathway to regulate biological events. The signaling systems in higher eukaryotes become much more complicated than those in primitive eukaryotes such as yeasts[1]. For example, the human genome contains more than 1500 genes that encode receptor proteins, as a result of gene duplication and divergence during evolution[2].

Furthermore, multicellular organisms must distribute circulating nutrient throughout the cells constituting the individual, as well as take nutrition from outside. To achieve this, homeostatic responses maintain circulating nutrients levels constant[3]. In these eukaryotic multicellular organisms, cell surface proteins are responsible in many cases for catalysis and uptake of substrates for nutritional metabolism as well as acceptance of extracellular signal molecules in target cells.

X-ray crystal structure analysis of cell surface proteins

Therefore, it is indispensable for maintaining individuals that cell surface proteins recognize extracellular signal molecules and external nutrient molecules. Such cell surface

proteins are often transmembrane proteins or extracellular proteins attached to the cell membrane through electrostatic, hydrophobic, and other non-covalent interactions[4,5]. Some extracellular proteins are post-translationally modified to have fatty acid or prenyl chains, or GPI (glycosylphosphatidylinositol), which will be anchored in the lipid bilayer[6]. Cell surface proteins perceive ligands through their extracellular domains to generate intracellular signal, which enables cell to alter the behavior of the cell or take up nutrients. Thus, understanding the ligand recognition mechanism of cell surface proteins is necessary to understand multicellular orgamisms.

By determining the complex structure of the cell surface protein and its ligand by X-ray crystallography, the structural basis of such a recognition mechanism can be clarified with near-atomic resolution. However, it is difficult to use a prokaryotic expression system, which is easy to handle with, for the expression of proteins for X-ray crystallographic analysis, as eukaryotic cell surface proteins have the following characteristics.

In order for proteins translated within cells to be transported to extracellular regions, it is necessary to have secretory signal sequences in very first amino acids at the N-terminus of the polypeptide chain. The secretory signal peptide directs the polypeptide chain to the translocation apparatus of the endoplasmic reticulum (ER), which enables protein to be located in the correct position outside the cell[7]. Some of the proteins that are exposed to the external side of the membrane are called glycoproteins because they have carbohydrates attached to their outer surfaces. Glycosylation occurs in sequence-specific manner in the ER and Golgi apparatus. Glycosylation makes the protein more polar, which contributes to its stability and solubility[8]. Another structural feature of cell surface protein to survive harsh extracellular condition is disulfide bonds. Apart from peptide bonds, disulfide bonds are the most common covalent bonds between amino acid residues in proteins. Disulfide bonds promote the resistance of the protein to proteases, contribute to the thermodynamic stability of the protein, while protecting the reactive thiol group of Cys[9].

Complex modifications such as glycosylation and disulfide bonds as described above are basically not carried out in prokaryotic expression systems and often eukaryotic expression systems are required for X-ray crystallographic analysis. Thus, in the present dissertation, to elucidate the ligand recognition mechanism of the cell surface proteins of higher eukaryotes, extracellular domain of such proteins are expressed in a secreted form in a eukaryotic expression system, and the X-ray crystallographic analysis of the receptor–ligand complex was performed. In Chapter 1, crystal structure of ENPP6, an extracellular phosphodiesterase preserved in vertebrates was determined in complex with its reaction product, phosphocholine. In Chapter 2, X-ray crystallographic analysis of TDR, a plant hormone receptor in complex with its ligand TDIF, is described. Based on the obtained structure, mutant analysis and structure–sequence comparison, the molecular function was analyzed.

Chapter 1 X-ray crystallography of ENPP6, a choline-metabolizing enzyme

1.1 Metabolism of choline

In 1998, U.S. National Academy of Sciences recognized choline (trimethyl-2-hydroxyl-ethylammonium) as an essential nutrient for humans[10]. Prolonged (weeks to months) ingestion of a diet deficient in choline results in hepatic, renal, pancreatic, memory and growth disorders[11]. Choline and its metabolites, assure the structural integrity and signaling functions of cell membranes. Choline is predominantly used for the synthesis of essential phospholipid components of cell membranes such as phosphatidylcholine (PC) and sphingomyelin (SM)[12]. Choline is also a source of methyl groups in the diet. For example, betaine, a choline metabolite, participates in the methylation of homocysteine to form methionine. Choline directly affects cholinergic neurotransmission and lipid transport from liver. Choline is de novo synthesized from phosphatidylethanolamine (PE) as a form of PC, which is catalyzed by PE *N*-methyltransferase in liver cells[13]. In contrast, in other cells choline is supplied from an outside source through choline-specific transporters expressed on their plasma membrane[14]. Since the amount of choline required is much higher than that synthesized by the liver in vivo, choline must be supplied from dietary sources[11].

Phospholipids such as PC or SM in the diet are digested to choline, which is then assimilated in the small intestine and transferred to the liver, where it is converted to PC[15]. PC is secreted from the liver and circulates in the blood as a form of lipoprotein, which is sequentially degraded to free choline in the circulation[16,17]. In blood, choline is present in the form of choline derivatives such as PC, SM, glycerophosphorylcholine (GPC) and lysophosphatidylcholine (LPC) or a free form (free choline)[18]. PC is the most abundant choline source in blood and its concentration is at the mM level. SM and LPC are also major choline-containing phospholipids in blood, and their concentrations are at the several hundred μ M level[19]. In contrast, free choline and GPC are minor components and their concentrations are only at the several μ M level in the blood and cerebrospinal fluids[18]. Since *de novo* synthesis

of choline occurs specifically in liver cells, it is reasonable that free choline used by most cells in a body originates from choline-containing phospholipids in blood.

1.1 ENPP family proteins

ENPP (Ecto-nucleotide pyrophosphate/phosphodiesterase or NPPs) family proteins are extracellular enzymes conserved among vertebrates and hydrolyze phosphodiesters in biological molecules[20,21]. In mammals, seven members are characterized as ENPP family proteins (ENPP1–ENPP7, Fig 1-1). While ENPP1–3 consists of two N-terminal somatomedin B-like domains, a central phosphodiesterase (PDE) domain and a C-terminal nuclease-like domain, ENPP4–7 have only the PDE domain in common. Although the PDE domains of ENPPs share high sequences similarities over 30%, they hydrolyze various pyrophosphate or phosphodiester bonds in (di)nucleotides and phospholipids, and thereby regulate a range of biological processes. For example, ENPP2 (also known as autotaxin, ATX) has lysophospholipase D activity and hydrolyzes LPC and other lysophospholipids to generate a signaling molecule, lysophosphatidic acid (LPA), which contributes to cell growth and cell motility[22,23]. ENPP1, ENPP3 and ENPP4 hydrolyze nucleotides such as ATP[24], UDP-GlcNAc[25] and diadenosine triphosphate (Ap3A)[26], and have roles in bone mineralization, regulation of sugar chain synthesis and platelet aggregation, respectively. ENPP3, as known as CD203c is highly expressed on basophils and hydrolyzes ATP to negatively regulate allergy and inflammation[27]. ENPP7 specifically hydrolyzes sphingomyelin and contributes to sphingolipid metabolism in the intestine[28].

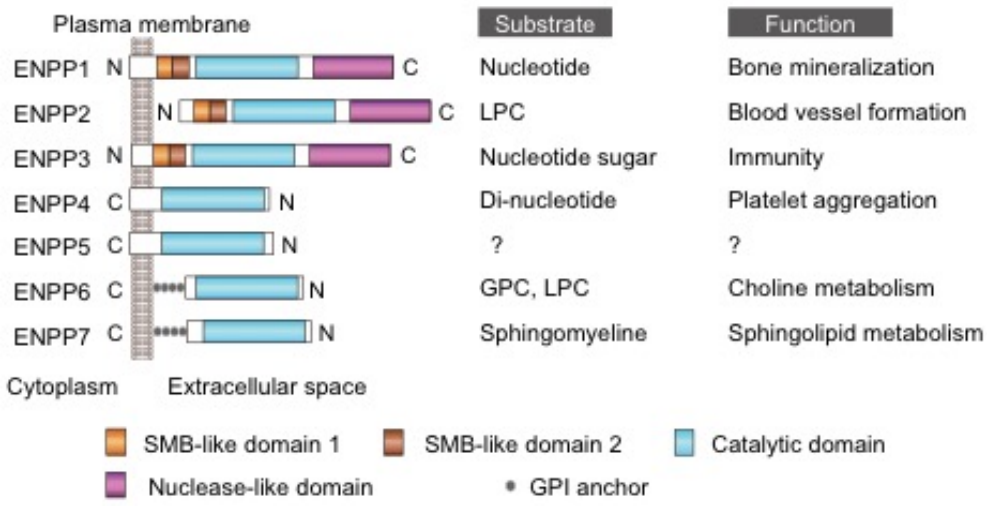

Fig. 1-1 Domain schematic of ENPP family proteins.

ENPP6 has *in vitro* phosphodiesterase activities towards choline-containing compounds such as GPC[29] and *O*-phosphorylcholine *N*-acyl ethanolamine[30] to produce phosphocholine. ENPP6 hydrolyzes only choline-containing lysophospholipids, such as LPC, sphingosylphosphorylcholine (SPC), platelet-activating factor (PAF) and lysoPAF, but not other lysophospholipids[29], suggesting that ENPP6 recognizes the choline moiety of its substrates.

Recently, crystal structures of ENPP1, ENPP2 and ENPP4 have revealed that the insertion loop in the catalytic pocket of the PDE domain determines their substrate specificities (Fig. 1-2 A, B) [31–35]. ENPP1 and ENPP4 have a nucleotide-binding pocket formed by the insertion loop[33–35], whereas ENPP2 has an LPC-binding hydrophobic pocket, since it lacks the insertion loop (Fig. 1-2 B)[31,32]. ENPP6 hydrolyzes LPC to generate phosphocholine and monoacylglycerol (lysoPLC activity), whereas ENPP2 hydrolyzes LPC to generate LPA and choline (lysoPLD activity). This catalytic difference suggested that whereas ENPP2 has a hydrophobic pocket that accommodates the acyl chains of LPC substrates[32], ENPP6 has an active-site pocket that recognizes the choline head group of LPC substrates. However, the choline-recognition mechanism of ENPP6 remains unknown because of a lack of structural information.

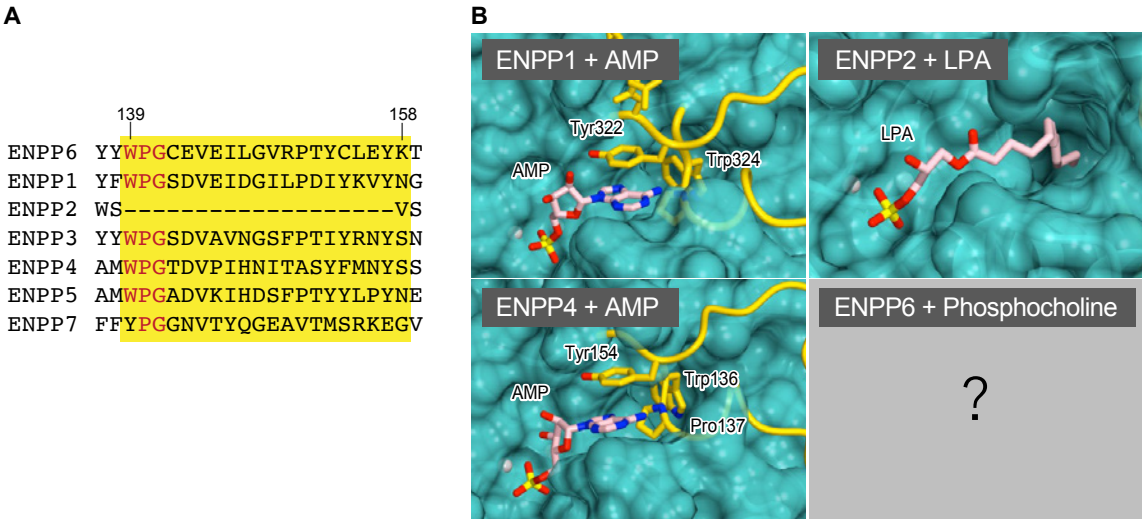

Fig. 1-2 Structures of ENPP family proteins. A. Multiple amino acid sequence alignment of the mouse ENPP family proteins. Insertion sequences are highlighted. **B.** Active sites of the ENPP1 in complex with AMP (PDB ID 4GTW), ENPP2 in complex with 14:0-LPA (PDB ID 3NKN) and ENPP4 in complex with AMP (PDB ID 4LQY). Insertion loops are colored gold.

1.2 Physiological role of ENPP6

1.2.1 Cellular role of ENPP6

ENPP6 is expressed in the kidney, brain and liver, and is bound to the plasma membrane through its C-terminal glycosylphosphatidylinositol (GPI) anchor[29,36]. GPC in various biological fluids such as blood and cerebrospinal fluids at the several μM level, is hydrolyzed to phosphocholine by ENPP6, which is present on the extracellular surface of certain cells such as oligodendrocytes and liver sinusoid endothelial cells. The product phosphocholine is quickly converted to free choline in various cell types and used as a source for synthesis of choline-containing phospholipids such as PC. In conclusion, ENPP6 serves as a choline-supplying enzyme (Fig. 1-3).

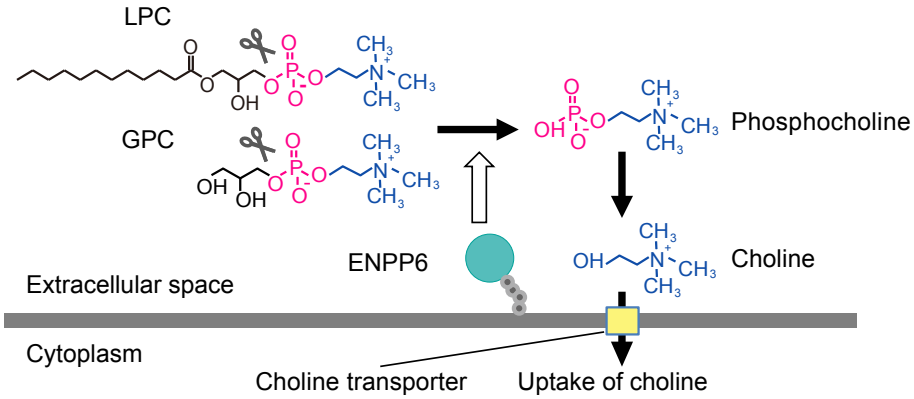

Fig. 1-3 Functional model of ENPP6 in cells.

1.2.2 Physiological role of ENPP6 in brain

ENPP6 is an early marker of oligodendrocyte differentiation. It is reported that the myelin sheath was less developed in ENPP6 knock out mice than in wild-type mice. Notably, myelin is rich in lipids, especially choline-containing phospholipids such as SM[37], suggesting that ENPP6 is involved in choline metabolism in the brain. In the developing brain, oligodendrocytes require large quantities of choline to make lipid-rich myelin. ENPP6 contributes to supplying choline for oligodendrocytes (Fig. 1-4 A).

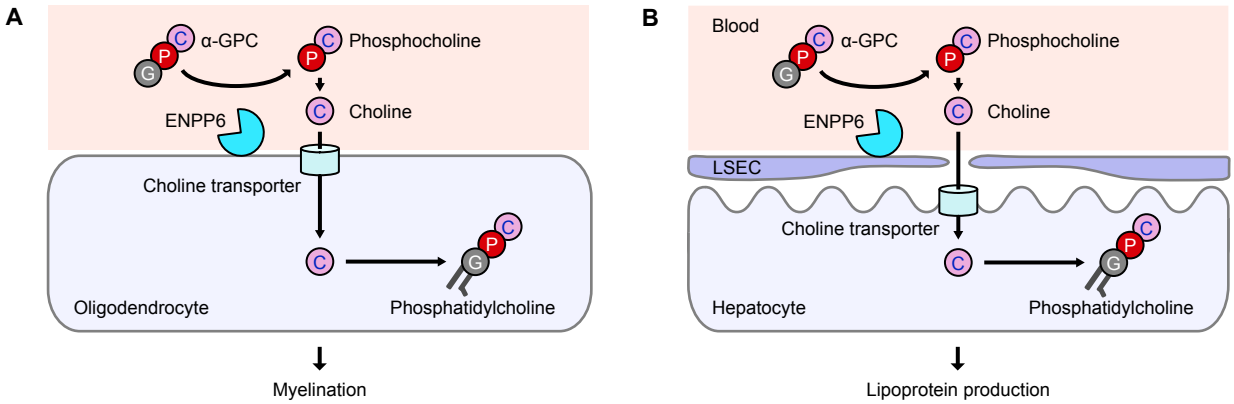

Fig. 1-4 Schematic model of ENPP6-mediated metabolism for GPC in oligodendrocyte (A) and hepatocyte (B).

1.2.3 Physiological role of ENPP6 in liver and kidney

In the liver, ENPP6 is expressed in sinusoidal endothelial cells, which mediate the

transfer of substrates between the sinusoidal blood and hepatocytes, and thus are important for hepatic function. Like the brain, the liver requires large quantities of choline. ENPP6 knock out mice exhibit mild symptoms of fatty liver, the most known symptoms induced by choline-deficiency. Consistently, ENPP6 knock out mice were more susceptible to a choline-deficient diet and exhibited severe fatty liver phenotypes at the early stage, indicating that ENPP6 is involved in choline metabolism in the liver. As liver cells continuously synthesize and secrete PC as a form of lipoprotein, ENPP6 appears to have a role in supplying choline in the liver (Fig. 1-4 B)

Similar ENPP6-mediated choline uptake was observed in the kidney. ENPP6 is expressed in epithelial cells forming the lumen of the proximal tubules, which is where choline in the primary urine is reabsorbed. Thus, it is possible that ENPP6 in the kidney contributes to the reabsorption of choline by hydrolyzing GPC in the primary urine.

1.3 Overview of this study

As described above, ENPP6 is strongly suggested to be responsible for phosphocholine production *in vivo*. In addition, among ENPP family proteins, only ENPP6 specifically recognize choline moiety. However, the molecular structure of ENPP6 is unknown and the molecular mechanism of choline recognition by ENPP6 remains elusive.

In this chapter, the crystal structure of ENPP6 in complex with phosphocholine is determined at high resolution, providing the structural basis for the choline recognition mechanism by ENPP6. Futhermore, a structural comparison of ENPP6 and other ENPPs offers an explanation of why among the family members only ENPP6 specifically recognizes the choline moiety of substrates.

1.4 Material and Methods

1.4.1 Construction

An expression plasmid encoding the catalytic domain (residues 1–421) of mouse ENPP6 fused with the C-terminal TARGET tag, which consists of 21 amino acids (YPGQ×5 + V) and is recognized by the P20.1 antibody was constructed [38]. DNA fragments encoding the catalytic domain of mouse ENPP6 were PCR- amplified by PrimeSTAR MAX DNA polymerase (TakaRa Bio) using pCAG-GS-ENPP6 as the template. The PCR products were inserted into the *XbaI* and *KpnI* sites of pcDNA3.1 (Invitrogen), which had been modified to contain a C-terminal Tobacco etch virus (TEV) protease cleavage site followed by the TARGET tag (referred to as pcD-CW)[38]. The expression vectors encoding ENPP6 mutants were prepared by site-directed mutagenesis using pcD-CW-ENPP6 as the template. The sequences were verified by DNA sequencing.

1.4.2 Cell culture and culture condition

The wild type and mutants of ENPP6 were expressed as secreted forms in stably transfected HEK293S GnT1- cells or transiently transfected HEK293T cells. The cells were cultured in DMEM (Dulbecco's modified Eagle's medium, Sigma) supplemented with 10%(v/v) FBS (Fetal bovine serum, Euro Clone), 1%(v/v) MEM non-essential amino acids (Sigma) and 1%(v/v) sodium pyruvate solution (Gibco) and were incubated at 37 °C in a humidified atmosphere containing 5% CO_2.

1.4.3 Generation of stable cell line and protein expression

HEK293S GnT1$^-$ cells were co-transfected with the expression plasmid and the IR/MAR plasmid (Trans Genic Inc.) using the Lipofectamine 2000 reagent (Invitrogen). Stably transfected cell lines were established in medium containing 1 mg mL^{-1} G418 (Nacalai Tesque) and 100 mg mL^{-1} Blasticidin (InvivoGen), and were cloned by a limiting-dilution procedure in 96-well plates for two weeks. To obtain a single clone secreting a high level of ENPP6, the clones were evaluated by measuring the phosphodiesterase activity in culture supernatants

using *p*-nitrophenylphosphorylcholine as a substrate, as described by Sakagami *et al*[29]. HEK293T cells were transfected with the expression plasmid, using the Lipofectamine 2000 reagent, and were then cultured on 150 mm dishes for 3 days.

1.4.4 Protein purification

All purification operations were performed on ice or at 4 °C. The wild type and mutants of ENPP6 were purified from the culture supernatants using P20.1-Sepharose resin and a Superdex 200 Increase gel-filtration column (GE Healthcare) in essentially the same manner as described for ENPP1[39]. The TARGET tag was cleaved by TEV protease before the gel-filtration step. For crystallization, the C393A/C412S mutant was purified using the P20.1-Sepharose resin followed by TARGET tag cleavage. The protein was further purified by chromatography on a HiLoad 16/60 Superdex 200 gel-filtration column (GE Healthcare) in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, and concentrated to 2 mg mL^{-1} using an Amicon Ultra-4 filter 10 kDa MWCO (molecular-weight cutoff; Millipore) and then stored at -80 °C until use. The purity and oligomeric state of the protein were assessed by SDS–PAGE in the presence (reducing conditions) and absence (nonreducing conditions) of 5% β-mercaptoethanol, and the gels were stained with Simply Blue SafeStain (Invitrogen).

1.4.5 Crystallization

Initial crystallization screening was performed at 293 K by the sitting-drop vapour-diffusion method in an MRC 96-well Crystallization Plate (Molecular Dimensions) using the following screening kits: Crystal Screen (Hampton Research), The PACT Suite and The JCSG+ Suite (Qiagen) and JBScreen Classic 1, 2, 4 and 5 (Jena Bioscience). Although the protein was crystallized in the presence of the inhibitor T11[24], no electron density was observed for the inhibitor. Thus this structure was refered as an apo form. Crystals of the apo form were obtained by mixing 0.1 µl of protein solution (2 mg/mL ENPP6, 2 mM T11, 20 mM Tris-HCl pH 7.5, 150 mM NaCl) and 0.1 µl of reservoir solution (0.2 M ammonium chloride, 0.1 M sodium acetate pH 5.0, 20% PEG 6000). Crystals of ENPP6 in complex with phosphocholine were

obtained by mixing 100 nl protein solution (2 mg/mL ENPP6 in 20 mM Tris–HCl pH 7.5, 150 mM sodium chloride, 0.2 mM zinc sulfate, 10 mM phosphocholine) and 100 nl reservoir solution using a Mosquito crystallization robot (TTP Labtech). Initial hits were optimized at 293 K by the sitting-drop vapour-diffusion method by mixing 200 nl protein solution and 200 nl reservoir solution.

1.4.6 Data collection and structure determination

Crystals were cryoprotected in reservoir solution supplemented with 25%(v/v) ethylene glycol and were flash-cooled in liquid nitrogen. X-ray diffraction data for crystals of the apo form were collected at 100 K on beamline BL41XU at SPring-8 (Hyogo, Japan) and processed using XDS[40]. The structure was determined by molecular replacement with MOLREP[41]. The search model was built by the Phyre 2 server[42], using the PDE domain of mouse ENPP1 (PDB ID 4GTW) as a template. Model building and refinement were performed using COOT[43] and PHENIX[44]. X-ray diffraction data for crystals of ENPP6 in complex with phosphocholine were collected on the beamline PXI at the Swiss Light Source (Villigen, Switzerland) using a Pilatus 6M detector. A data set was collected at a wavelength of 1.278 Å with an oscillation angle of 360° (0.1° per frame), an exposure time of 0.1 s per frame and a transmission of 10%. Diffraction images were integrated and scaled using XDS[40]. Molecular replacement was performed with MOLREP[41], using using the apo form of ENPP6 as a search model. The stereochemical qualities of the models were validated by Ramachandran plot analysis with the program RAMPAGE[45]. The data collection and refinement statistics are summarized in Table 1.1.

1.4.7 α-GPC-hydrolyzing assay for the ENPP6 mutants

Site-directed mutagenesis was performed using PrimeSTAR Max DNA Polymerase (Takara Bio). The mutants of ENPP6 were expressed as secreted forms in transiently transfected HEK293T cells, and then purified as described in 1.4.4. Each ENPP6 protein was mixed with α-GPC and calf intestine alkaline phosphatase (20 U/ml) at 37 °C in the presense

of 100 mM Tris-HCl (pH 9.0), 5 mM $MgCl_2$, 500 mM NaCl, 0.05% Triton X-100. After 1 h, choline produced from α-GPC hydrolysis was detected by an enzymatic photometric method using choline oxidase (Asahi Chemical), horseradish peroxidase (Toyobo), and TOOS reagent (*N*-ethyl-*N*-(2-hydoroxy-3-sulfoproryl)-3-methylaniline; Dojindo Molecular Technologies, Inc.) as a hydrogen donor[46,47].

1.4.8 Choline incorporation assay

Choline incorporation assays were performed by Dr. Kuniyuki Kano and Prof. Junken Aoki (Tohoku University).

ENPP6-expressing cells were generated with a retroviral system as follows: Mouse ENPP6 and ENPP6 mutant cDNAs were introduced into the retrovirus vector pMXs-IG[48]. The obtained vector was transfected into PLAT-E cells[49] (kindly gifted by Dr. Kitamura). 10 mL of culture supernatants of PLAT-E cells containing the viruses were collected and filtered through a 0.45 μm filter. For infection, the virus stock was added into Rh7777 cells with polybrene (Sigma). After twice infection, GFP positive cells were sorted by Cell Sorter SH800 (Sony).

Choline incorporation activity was measured using α-GPC in ENPP6-expressing cells in DMEM supplemented with 1% lipoprotein-depleted serum (LPDS). LPDS was prepared from FBS as described previously[50]. It is noted that the obtained LPDS does not contain free choline or GPC. The cells were incubated with 10 μM D-α-GPC for 24 h and then lipids were extracted by methanol. The level of PC was determined by the liquid chromatography–mass spectroscopy (LC-MS/MS).

The LC-MS/MS system consisting of a NANOSPACE SI-2 HPLC (Shiseido Co., Ltd) and a TSQ Quantum Ultra triple quadropole mass spectrometer (Thermo Fisher Scientific) equipped with a heated ESI source. For analysis of PC and D-PC, they were separated by a SILLICA SG80 column (2.0 mm × 150 mm i.d., 5 μm particle size, Shiseido Co., Ltd.) using a

gradient of solvent A (5 mM ammonium formate in water) and solvent B (acetonitrile). The initial condition was set at 100% B. The following solvent gradient was applied: a 15-min gradient to 82.5% B, a 2-min gradient to 30% B, a 2-min gradient to 100% B and hold for 4 min. The flow rate was set at 400 μL/min. All compounds were analyzed as [M+H]+ in the positive ion mode. The product ions were as follows: m/z 60.2 (choline), m/z 69.2 (D-choline), m/z 104.2 (α-GPC), m/z 113.2 (D-α-GPC), m/z 184.0 (PC and LPC) and m/z 193 (D-PC). The ratio between analyte and internal standard peak area was used for quantification.

1.5 Results

1.5.1 Protein preparation

Since our group successfully expressed ENPP1[39] and ENPP2[32] in HEK293S GnT1$^-$ cells as soluble secreted proteins, the catalytic domain (residues 1–421) of mouse ENPP6, which lacks the C-terminal GPI-anchor signal sequences (residues 422–440), was expressed in HEK293S GnT1$^-$ cells (Fig. 1-5). However, unlike ENPP1 and ENPP2, the affinity-purified ENPP6 protein predominantly eluted in the void volume from the gel-filtration column, with two small peaks corresponding to molecular weights of 50 and 100 kDa (Fig. 1-6 A). Reducing and nonreducing SDS–PAGE analyses revealed that these three peaks are all composed of ENPP6, indicating that ENPP6 exists as a mixture of monomers, dimers and higher-order oligomers in solution (Fig. 1-6 B). In addition, these results indicated that the observed oligomerization is mediated by intermolecular disulfide linkages. In contrast, when ENPP6 was expressed in HEK293T cells, the monomeric population of ENPP6 eluted from the

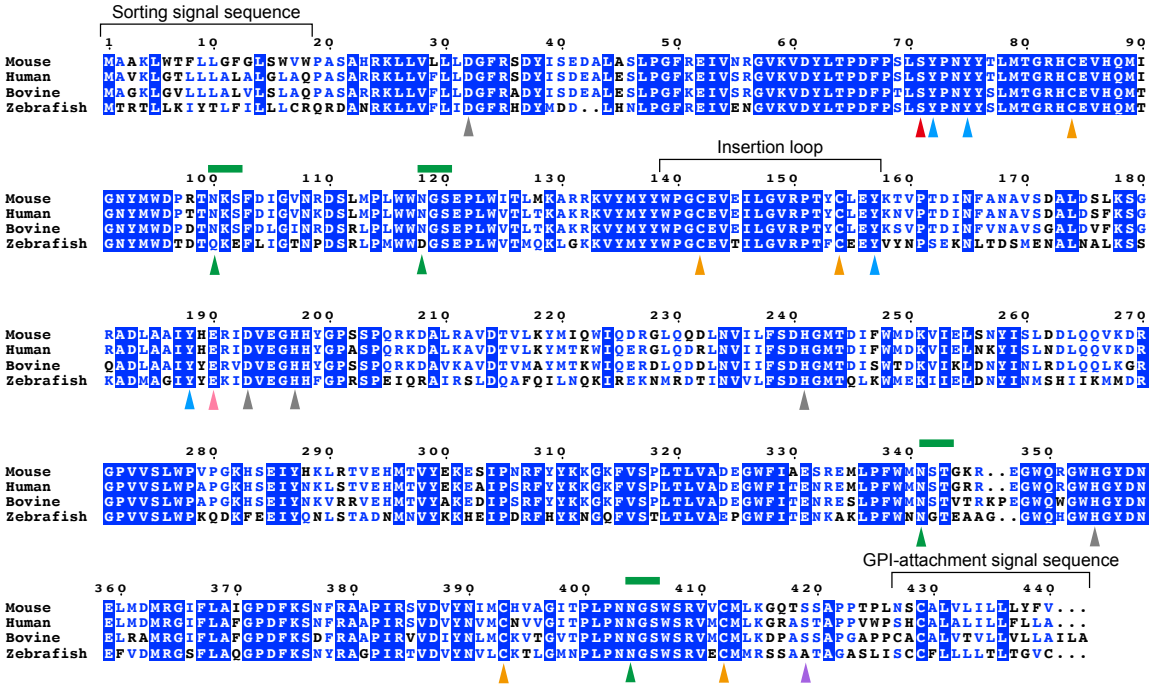

Fig. 1-5 Multiple sequence alignment of ENPP6. The zinc-coordinating residues, gray triangles; the catalytic residue, red triangles; the tyrosine residues in the cholinebinding pocket, light blue triangles; the residues hydrogen bonding with the tyrosine residues in the choline-binding pocket, light pink triangles; cysteine residues, orange triangles; the *N*-glycosylated asparagine residues, green triangles; the consensus sequences for *N*-glycosylation, green lines; the predicted C-terminal GPI-attachment residue, a purple triangle. The sequence alignment was prepared using ClustalW[124] and ESPript[125].

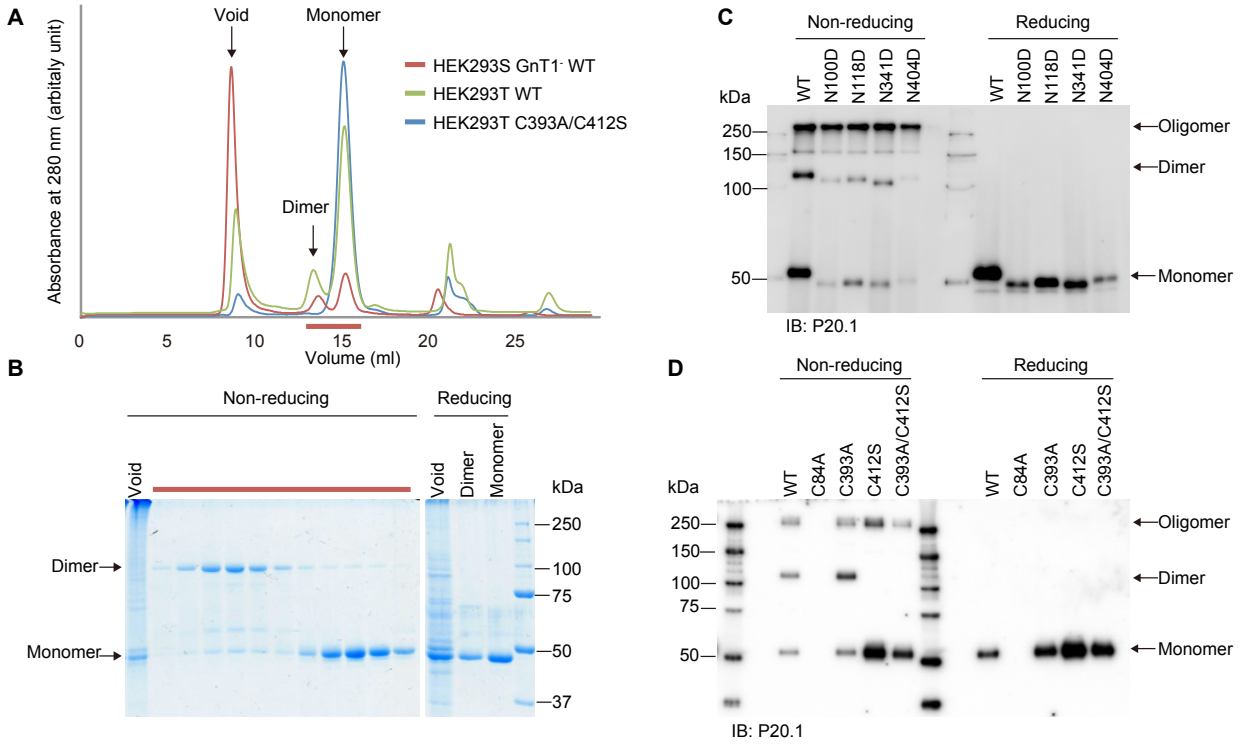

Fig. 1-6 Oligomeric state of the wild type and mutants of ENPP6. A. Gel-filtration analysis of wild-type ENPP6 expressed in HEK293S GnT1$^-$ cells and the wild type and C393A/C412S mutant of ENPP6 expressed in HEK293T cells. The proteins were purified by P20.1-Sepharose and then chromatographed on a Superdex 200 Increase gel-filtration column. WT, wild type. **B**. SDS–PAGE analysis of wild-type ENPP6 expressed in HEK293S GnT1$^-$ cells. The fractions indicated by the red line in A. were analyzed by SDS–PAGE under reducing and nonreducing conditions. **C**. Effect of *N*-glycosylation on the oligomeric state of ENPP6. The wild type and *N*-glycosylation site mutants of ENPP6 were expressed in HEK293T cells, purified by P20.1-Sepharose and analyzed by SDS–PAGE under reducing and nonreducing conditions. The proteins were detected by the P20.1 antibody. **D**. Effect of the cysteine residues on the oligomeric state of ENPP6. The wild type and cysteine mutants of ENPP6 were expressed, purified and then analyzed as in C. The C84A mutant was not expressed in HEK293T cells.

gel-filtration column increased remarkably, with a concomitant decrease in the dimeric and higher-order oligomeric populations (Fig. 1-6 A). Given that HEK293S GnT1$^-$ and HEK293T cells produce proteins with high mannose *N*-linked glycans and complex *N*-linked glycans, respectively[51], these observations indicated that complex *N*-linked glycosylation is important for preventing the disulfide-mediated oligomerization of ENPP6.

ENPP6 has four predicted consensus sequences for *N*-glycosylation [NX(S/T), where X is any amino-acid residue other than Pro; Fig. 1-5], and a mass-spectrometric analysis confirmed that the four *N*-glycosylation sites are indeed modified in bovine ENPP6[36]. To examine the effects of *N*-glycosylation on the oligomeric state of ENPP6, the four mutants of mouse ENPP6 (N100D, N118D, N341D and N404D) were prepared. Western blotting analysis revealed that the monomeric populations of the four mutants were remarkably reduced compared with that of the wild type (Fig. 1-6 C), indicating that all four of the *N*-glycosylation modifications are important for preventing the disulfide-mediated oligomerization of ENPP6.

ENPP6 has five conserved cysteine residues (Cys84, Cys142, Cys154, Cys393 and Cys412; Fig. 1-5). Mass-spectrometric analyses revealed that Cys142 and Cys154 of bovine ENPP6, which correspond to Cys142 and Cys154 of mouse ENPP6, respectively, form an intra-molecular disulfide bond and that Cys414 of bovine ENPP6, which corresponds to Cys412 of mouse ENPP6, forms an intermolecular disulfide bond[36]. These observations suggested that Cys412 of mouse ENPP6 is involved in disulfide-mediated oligomerization. To examine the effects of these cysteine residues on oligomerization, three mutants of mouse ENPP6 (C84A, C393A and C412S) were prepared. Western blotting analysis showed that the C412S mutant predominantly exists as a monomer (Fig. 1-6 D), suggesting that Cys412 is involved in the disulfide-mediated dimerization. The Western blotting analysis further revealed that the expression levels of the C393A and C412S mutants were increased compared with that of the wild type (Fig. 1-6 D). We thus expressed the C393A/C412S double mutant in HEK293T cells and purified it using P20.1-Sepharose. The C393A/C412S mutant predominantly eluted as a monomer from the gel-filtration column (Fig. 1-6 A). The final yields of the wild type and C393A/C412S mutant of ENPP6 were 0.25 and 0.5 mg from 0.5 L culture, respectively.

1.5.2 Crystal structure analysis of ENPP6

To elucidate the choline recognition mechanism of ENPP6, I attempted to solve the crystal structures of ENPP6. The extracellular PDE domain of mouse ENPP6 (residues 24–415) C393A/C412S mutant showed α-GPC-hydrolyzing activity comparable to the activities of

the monomer and dimer of wild-type ENPP6 (Fig. 1-9 C), indicating that neither the oligomeric state nor the C393A/C412S mutation affects the enzymatic activity. Thus, the C393A/C412S mutant was crystallized in complex with phosphocholine, a reaction product.

After the initial screening, clustered crystals were obtained under condition No. 8 of The PACT Screen [0.2M ammonium chloride, 0.1M sodium acetate pH 5.0, 20%(w/v) PEG 6000; Fig. 1-7 A]. The conditions were optimized by adjusting the pH, the concentration of PEG and the types of salt in the reservoir solution. It was found that the addition of zinc sulfate enhanced the growth of single crystals. Finally, single crystals were obtained under crystallization conditions consisting of 0.2 M ammonium chloride, 0.1 M sodium acetate pH 4.5, 20%(w/v) PEG 6000, 0.2 mM zinc sulfate (Fig. 1-7 B).

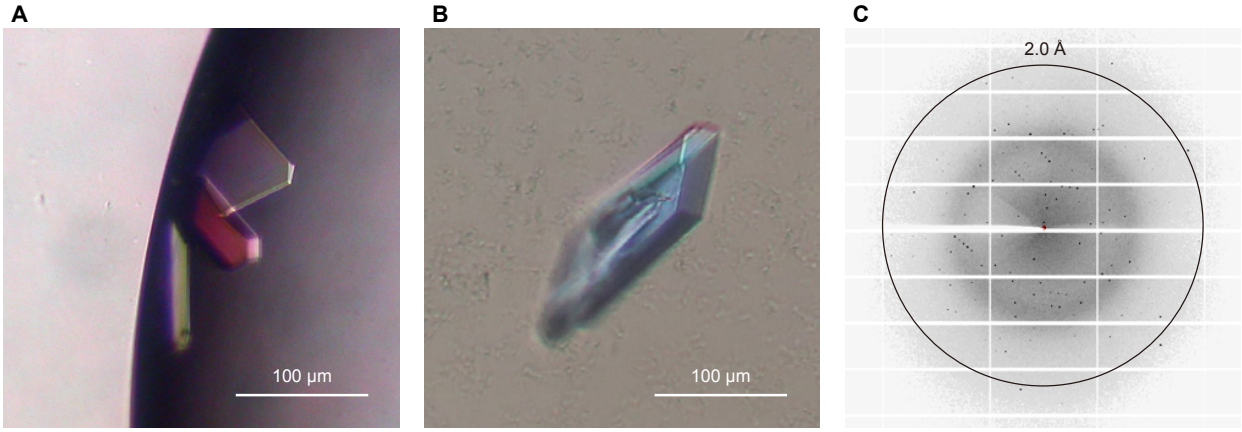

Fig. 1-7 Crystals and X-ray diffraction image of the ENPP6 C393A/C412S mutant. A. Crystals of the ENPP6 C393A/C412S mutant obtained by the initial screening. **B**. Crystals of the ENPP6 C393A/C412S mutant obtained under the optimized conditions. **C**. X-ray diffraction image of the ENPP6 C393A/C412S mutant.

The crystal of the C393A/C412S mutant diffracted to 1.8 Å resolution (Fig. 1-7 C) and belonged to space group *P*1, with unit-cell parameters $a = 63.7$, $b = 68.8$, $c = 69.7$ Å, $\alpha = 60.6$, $\beta = 87.0$, $\gamma = 68.1°$. The data-collection statistics are summarized in Table 1-1. Assuming the presence of two protein molecules (48.5 kDa calculated from the amino-acid sequence) per asymmetric unit, the Matthews coefficient (VM) was estimated to be 2.44 $Å^3\ Da^{-1}$, with a solvent content of 49.5%[52].

The crystal structures of the C393A/C412S mutant in complex with phosphocholine was determined at 1.8 Å resolution (Fig. 1-8, Fig. 1-9 A, B, Table 1-1). Previous crystal structures of ENPP1, ENPP2 and ENPP4 in complex with their reaction products (AMP, LPA and AMP, respectively) provided structural insights into their substrate recognition mechanisms[31–35] (Fig. 1-10). Accordingly, the crystal structure of ENPP6 in the apo form (without substrates) was also determined at 2.0 Å resolution (Fig. 1-8 A, C, Fig. 1-11 A, Table 1-1). Since the two structures are virtually identical (root-mean-square deviation [rmsd] of 0.23 Å for aligned Cα atoms), I will hereafter describe the ENPP6–phosphocholine complex structure, unless otherwise stated. The structure of the PDE domain of ENPP6 is similar to those of ENPP1 (PDB code 4GTW, 32% identity, rmsd of 1.6 Å for aligned 364 Cα atoms)[33], ENPP2 (PDB code 3NKM, 28% identity, rmsd of 1.7 Å for aligned 340 Cα atoms)[32], and ENPP4 (PDB code 4LR2, 33% identity, rmsd of 1.4 Å for aligned 359 Cα atoms)[35] (Fig. 1-10). Although electron densities of four glycans linked to Asn100, Asn118, Asn341 and Asn404 are observed in the apo form, the electron density of the Asn100-linked glycan was not observed in the complex, probably due to the flexibility of the Asn100-linked glycan (Fig. 1-11 A).

Although C393A/C412S mutant predominantly exists as a monomer in solution, both of the ENPP6–PC complex and the apo ENPP6 crystallize as dimers in the same conformations (Fig. 1-12 A). The active site of each promoter is located on the outer surface of the dimer and these two active sites are spatially independent, in agreement with the biochemical data that indicate dimerization of ENPP6 does not affect its enzymatic activity (Figure 1-12 B). Consistent with the fact that ENPP6 forms a dimer via the intermolecular disulfide bond between Cys412 of the respective monomers[53], the two Ser412 in the dimer observed in the crystal structure come close to each other (Fig. 1-12 A). Furthermore, dimerization is mediated by extensive hydrogen bonding networks between the promoters, and the residues involved in this hydrogen bonding and Cys412 are conserved among species (Fig. 1-5). Moreover, the C-termini that lead to GPI anchor signal sequences of the two protomers are close to each other. These observations implicate that ENPP6 forms a dimer structure found in the crystal

structure *in vivo* and binds to the plasma membrane via GPI anchor (Fig. 1-12 A), although the physiological significance of dimerization is unknown.

As observed in the structures of ENPP1, ENPP2 and ENPP4, two zinc ions are coordinated by conserved active-site residues (Fig. 1-9 B). One zinc ion is coordinated by Asp193, His197 and His354, while the other is coordinated by Asp32, His241 and the catalytic nucleophile Ser71 (Fig. 1-9 B).

Table 1-1. Data collection and refinement statistics.

	Apo ENPP6	ENPP6–phosphocholine
Data collection		
Beamline	SPring-8 BL41XU	SLS PXI
Wavelength (Å)	1.282	1.278
Space group	*P*1	*P*1
Cell dimensions		
a, *b*, *c* (Å)	64.3, 78.2, 103.6	63.7, 68.8, 69.7
α, β, γ (°)	90.0, 90.0, 114.6	60.6, 87.0, 68.1
Resolution (Å)	50.0–1.99 (2.02–1.99)	50.0–1.80 (1.91–1.80)
R_{sym}	0.188 (0.835)	0.058 (0.624)
$I/\sigma I$	6.11 (1.40)	12.8 (1.78)
Completeness (%)	96.6 (93.7)	93.6 (88.5)
Redundancy	3.96 (3.99)	3.54 (3.28)
CC(1/2)	0.99 (0.62)	0.99 (0.75)
Refinement		
Resolutions (Å)	38.9–2.00 (2.01–2.00)	45.6–1.80 (1.87–1.80)
No. Reflections	120,105	81,806
R_{work} / R_{free}	0.167 / 0.206	0.175 / 0.209
No. atoms		
Protein	12,807	6,439
Ligand/ion	394	220
Water	2,048	528
B-factors (Å^2)		
Protein	20.4	31.7
Ligand/ion	38.5	49.9
Water	31.1	38.2
R.m.s. deviations		
Bond length (Å)	0.007	0.007
Bond angles (°)	0.890	0.906
Ramachandran plot		
Favored region	97.2%	97.2%
Allowed region	2.8%	2.8%
Outlier region	0.1%	0%
PDB ID	5EGE	5EGH

*Values in parentheses are for the highest-resolution shell.

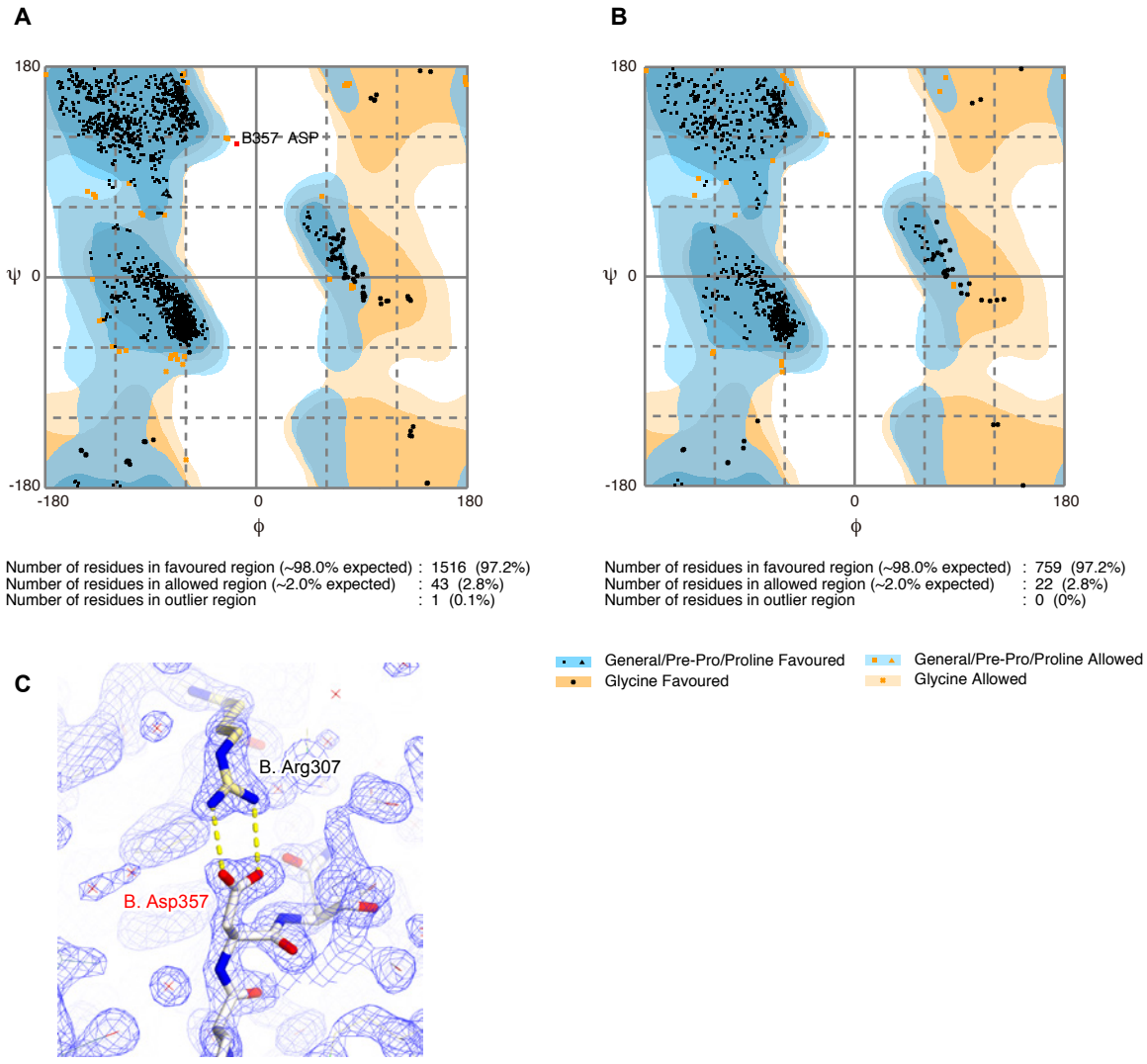

Fig. 1-8 Ramachandran analysis of the ENPP6 structures. A. Ramachandran plots of the apo structure. **B.** Ramachandran plots of the phosphocholine complex structure. **C**. The residue in the outlier region (B. Asp357 in the apo structure) and interacting residue are shown as sticks. The side chain of Asp357 is fixed by a intramolecule salt bridge with the side chain of Arg307. $2F_O - F_C$ electron density map (contoured at 2 σ) is shown in blue mesh.

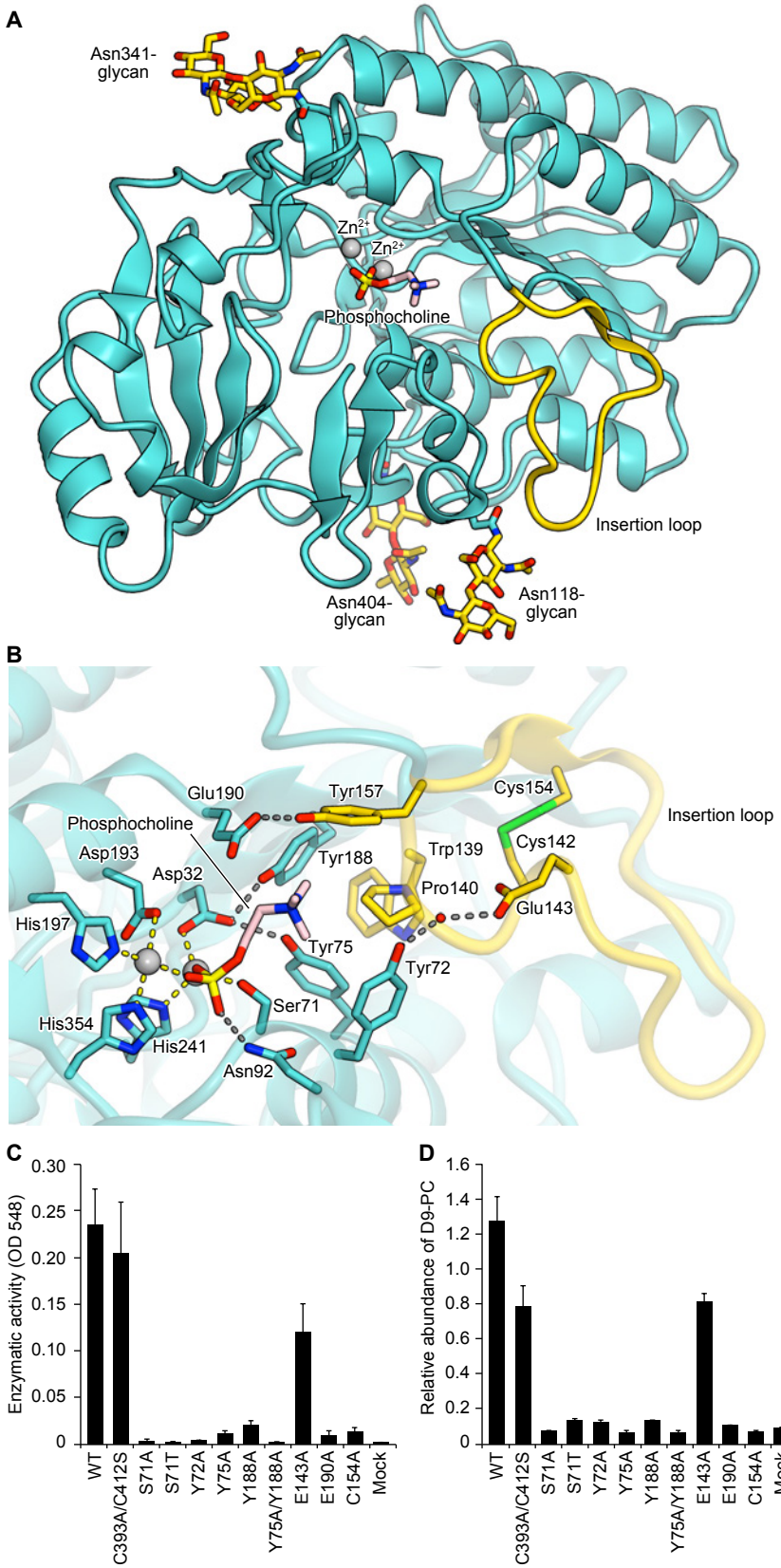

Fig. 1-9 Crystal structure of ENPP6. A. Crystal structure of the PDE domain of ENPP6 in complex with phosphocholine. *N*-linked glycans and the phosphocholine are shown as gold and light pink sticks, respectively. **B**. Active site of ENPP6 in complex with phosphocholine. Hydrogen bonds and coordinate bonds are shown as dashed gray and yellow lines, respectively. A water molecule is shown as a red sphere. The insertion loop is shown in gold and the bound zinc ions are shown as gray spheres in **A**, **B**. **C**. α-GPC-hydrolyzing activities of the ENPP6 mutants. **D**. Choline uptake in cultured cell lines that express wild-type or mutant ENPP6 proteins. Data are shown as mean ± SEM (n = 3).

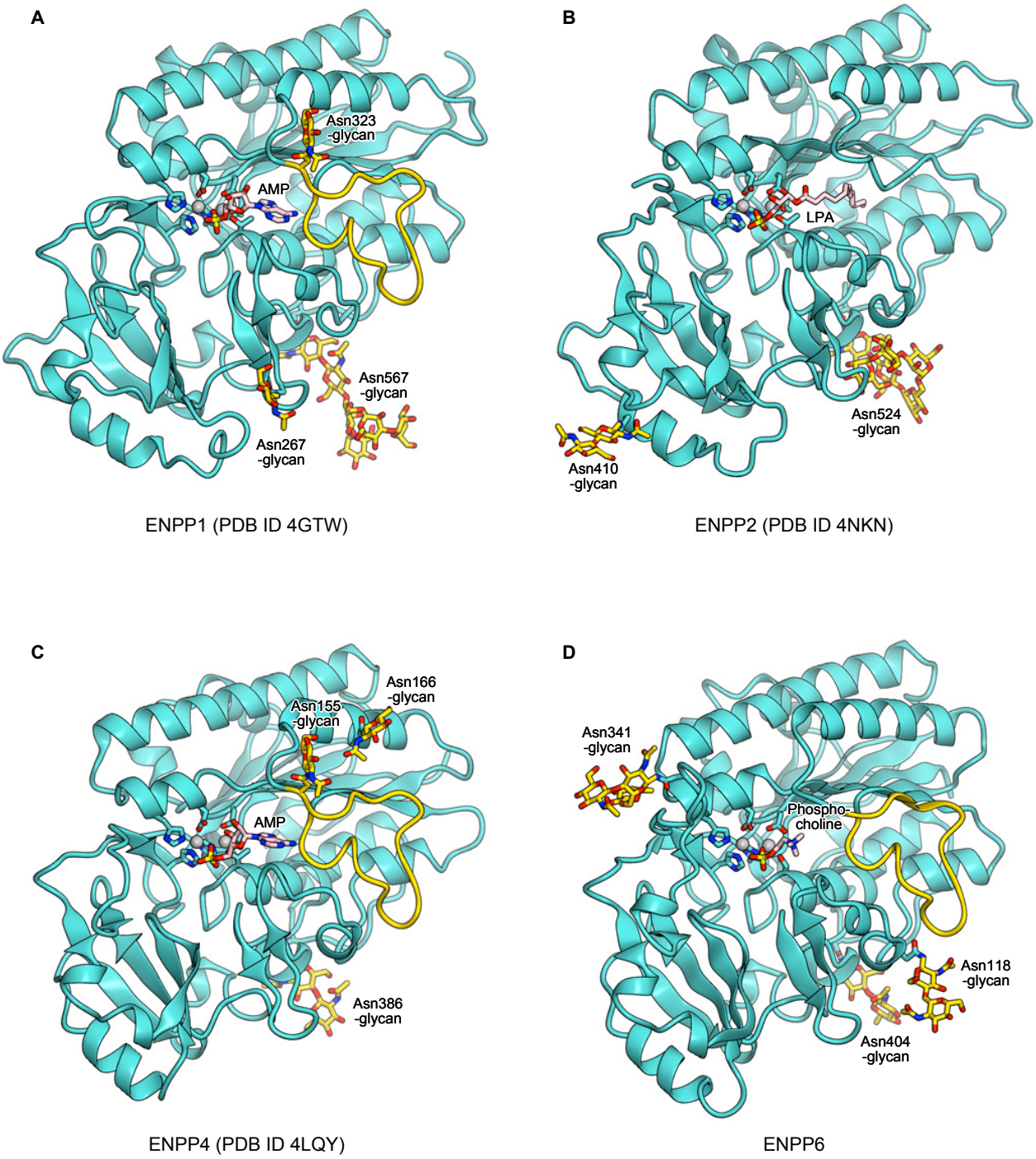

Fig. 1-10 Structural comparison of the PDE catalytic domains of ENPP1, ENPP2, ENPP4 and ENPP6. Crystal structures of the PDE domains of ENPP1 in complex with AMP (PDB ID 4GTW) (**A**), ENPP2 in complex with 14:0-LPA (PDB ID 3NKM) (**B**), ENPP4 in complex with AMP (PDB ID 4LQY) (**C**), and ENPP6 in complex with phosphocholine (**D**). The bound zinc ions are shown as gray spheres, and the active-site residues are shown as sticks. *N*-linked glycans are shown as gold sticks. The "insertion" loops are highlighted in gold.

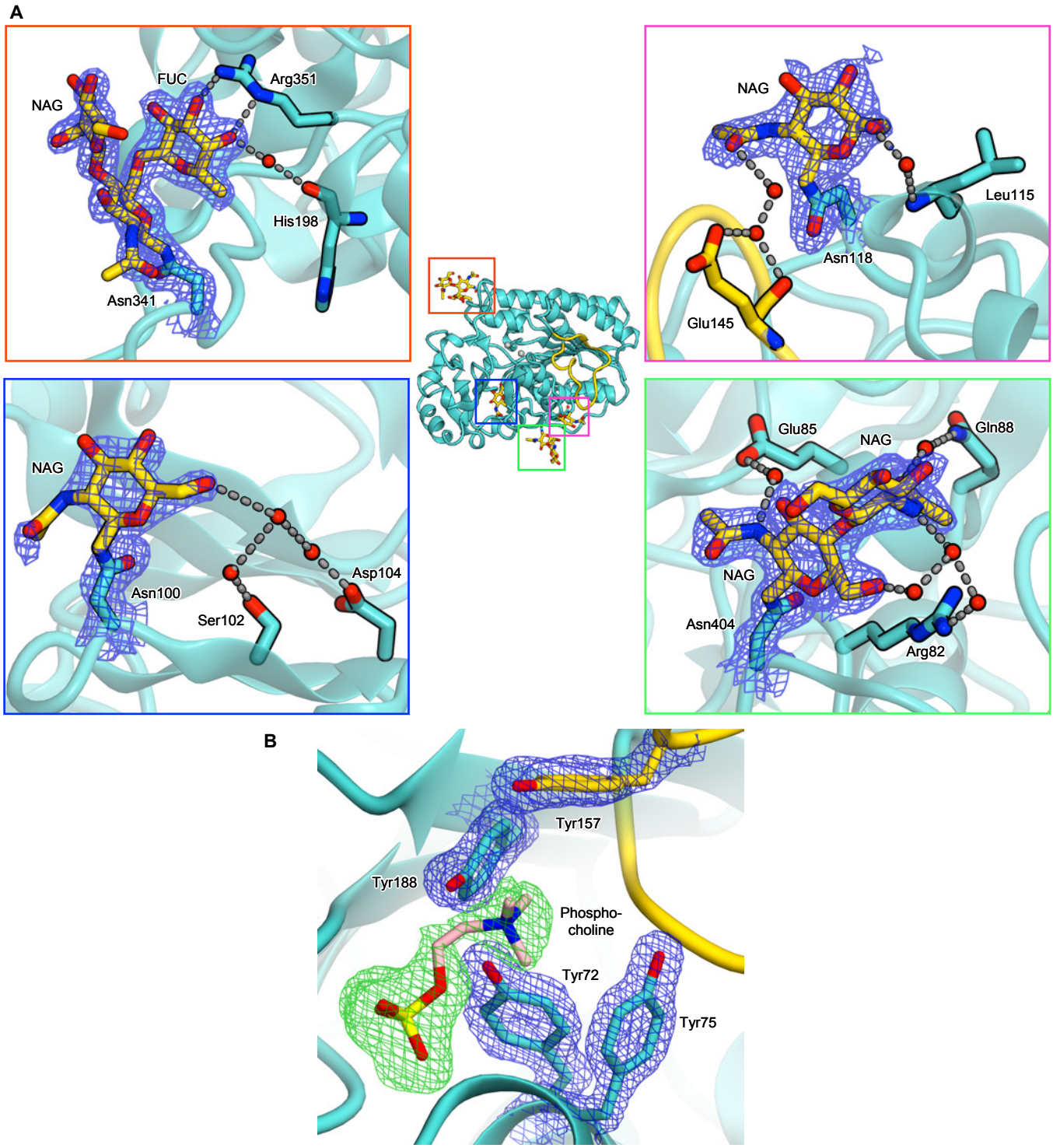

Fig. 1-11 Electron density maps of ENPP6. A. *N*-linked glycans of apo ENPP6. $2Fo - Fc$ electron density maps of *N*-linked glycans (contoured at 1σ) are shown in blue meshes. Water molecules are shown as red spheres. The *N*-acetylglucosamine (NAG) residues of the glycans linked to Asn100, Asn118 and Asn404 form water-mediated hydrogen bonds with the protein, and a fucose (FUC) residue of the Asn341-linked glycan forms direct and water-mediated hydrogen bonds with Arg351 and His198, respectively. These observations indicated that the intramolecular interactions between the *N*-linked glycans and protein surface are important for the stability of ENPP6. **B.** The active site of ENPP6. A simulated annealing $Fo - Fc$ omit electron density map (contoured at 3σ) of phosphocholine is shown in green mesh, and a $2Fo - Fc$ electron density map (contoured at 2σ) of Tyr72, Tyr75, Tyr157 and Tyr188 is shown in blue mesh.

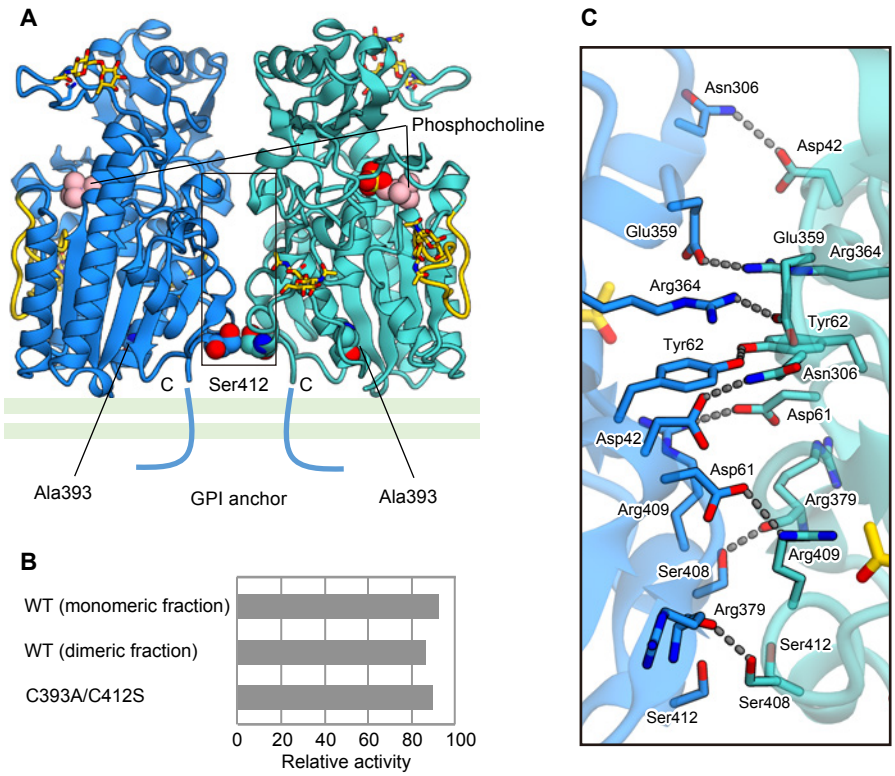

Fig. 1-12 The dimer structure of ENPP6 found in crystals. A. The schematic view of the dimer structure of ENPP6. Two protomers are differently colored. The region shown in C is boxed. B. α-GPC-hydrolyzing activities of ENPP6. The monomeric fraction and the dimeric fraction were collected from the chromatography on a Superdex 200 Increase gel-filtration column. C. A magnified view of the interactions between two protomers. Hydrogen bonds are shown as dashed gray lines.

1.5.3 Active site of ENPP6 in complex with phosphocholine

An electron density corresponding to the bound phosphocholine was observed in the vicinity of the zinc ions in the catalytic pocket (Fig. 1-11 B). The choline moiety of the phosphocholine is accommodated in a choline-binding pocket formed by the side chains of Tyr72, Tyr75, Tyr157 and Tyr188, while the phosphate group of the phosphocholine is located in the vicinity of the two zinc ions (Figs 1-9 B). Tyr72, Tyr75 and Tyr188 provide π-cation interactions with the positively-charged choline moiety. Furthermore, Asp32 and Glu190 electrostatically interact with the choline moiety. Tyr72 forms a water-mediated hydrogen bond with Glu143, while Tyr75/Tyr188 and Tyr157 form direct hydrogen bonds with Asp32 and Glu190, respectively, thereby stabilizing the conformation of the choline-binding pocket. In addition, Trp139 and Pro140 in the insertion loop (residues 139–157) interact with the side

chains of Tyr72, Tyr75, Tyr157 and Tyr188. The conformation of the insertion loop is stabilized by a disulfide linkage between Cys142 and Cys154.

In contrast, in the vicinity of the phosphate group of the phosphocholine, ENPP6 has a spacious region to accommodate substrates (the glycerol moiety of GPC or the acyl chain of LPC, Fig. 1-13). Since this region is spacious, it is difficult to predict the specific binding mode of such substrates and argue for the difference in substrate preference between GPC and LPC. Further studies will be required to address these issues.

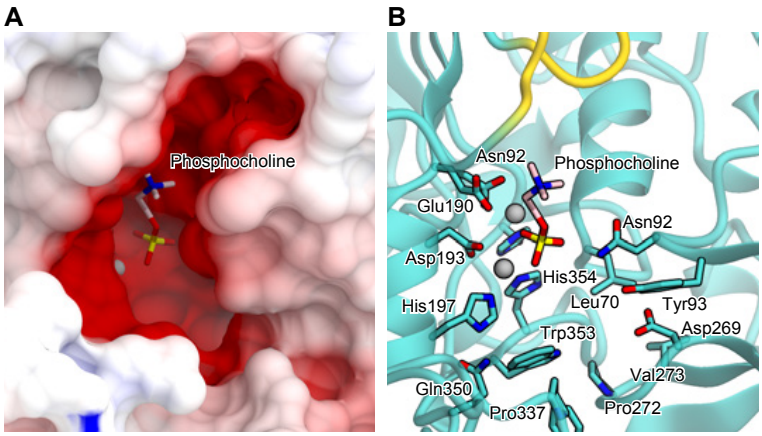

Fig. 1-13 Space around the phosphate group of the phosphocholine is shown in surface representation (A) and ribbon and stick representation (B). The electrostatic surface potentials is colored in in red (contoured from −10 kT/e) to blue (+10 kT/e).

1.5.4 α-GPC-hydrolyzing activities of the choline recognition pocket mutants

To verify the functional significance of the choline-binding pocket, I introduced mutations to the C393A/C412S mutant, purified them, and measured their α-GPC-hydrolyzing activities. Among the mutants tested, the Y157A mutant was not expressed in the culture supernatants of HEK293T cells, which suggests that Tyr157 is essential for correct protein folding. The other mutants eluted as a single peak from the gel-filtration column, confirming their structural integrity (data not shown). The Y72A, Y75A, Y188A and E190A mutants exhibited the remarkably reduced α-GPC-hydrolyzing activities, as compared with the wild-type and C393A/C412S mutant proteins (Fig. 1-9 C), confirming the importance of Tyr72, Tyr75, Tyr188 and Glu190 for choline recognition. The activities of the C154A mutants were

reduced, indicating the importance of the disulfide linkage between Cys142 and Cys154 for the choline-binding pocket formation. The Tyr72, Tyr75, Cys142 and Cys154 are evolutionarily conserved among ENPP6 proteins, but not among other ENPP family members (Fig. 1-5, Fig. 1-14), highlighting the importance of the tyrosine residues and the disulfide linkage in the insertion loop for choline recognition. The catalytic nucleophile is a serine residue in ENPP6 (Ser71 in mouse ENPP6), whereas it is a threonine residue in other ENPP family members (Fig. 1-14 and 1-15 A–D). Indeed, the S71A and S71T mutants of ENPP6 showed almost no α-GPC-hydrolyzing activities (Fig. 1-9 C). In the crystal structure, the side-chain Cβ atom of Ser71 forms van der Waals interactions with the side chain of Tyr75, and contributes to the choline-binding pocket formation (Fig. 1-15 D), suggesting that the side-chain methyl group of a threonine residue at this position would generate steric clashes with Tyr75. These observations provided a structural explanation for why only ENPP6 has a serine residue as a catalytic nucleophile among the ENPP family members.

1.5.5 Choline uptake in cultured cell lines that express wild-type or mutant ENPP6 proteins

To examine the physiological significance of choline recognition by ENPP6 for the choline metabolism, choline uptake in cultured cells that express wild-type or mutant ENPP6 proteins were evaluated by Dr. Kuniyuki Kano and Prof. Junken Aoki (Tohoku University). Wild-type ENPP6 expressed in cultured hepatomas hydrolyzed α-GPC for the biosynthesis of PC (Fig. 1-9 D). In contrast, the mutant ENPP6 proteins, which had impaired α-GPC-hydrolyzing activities (Fig. 1-9 C), failed to support such choline metabolism (Fig. 1-9 D). Together, structural and mutational analyses reveal that specific recognition of the choline moiety of GPC is required for ENPP6-mediated choline metabolism.

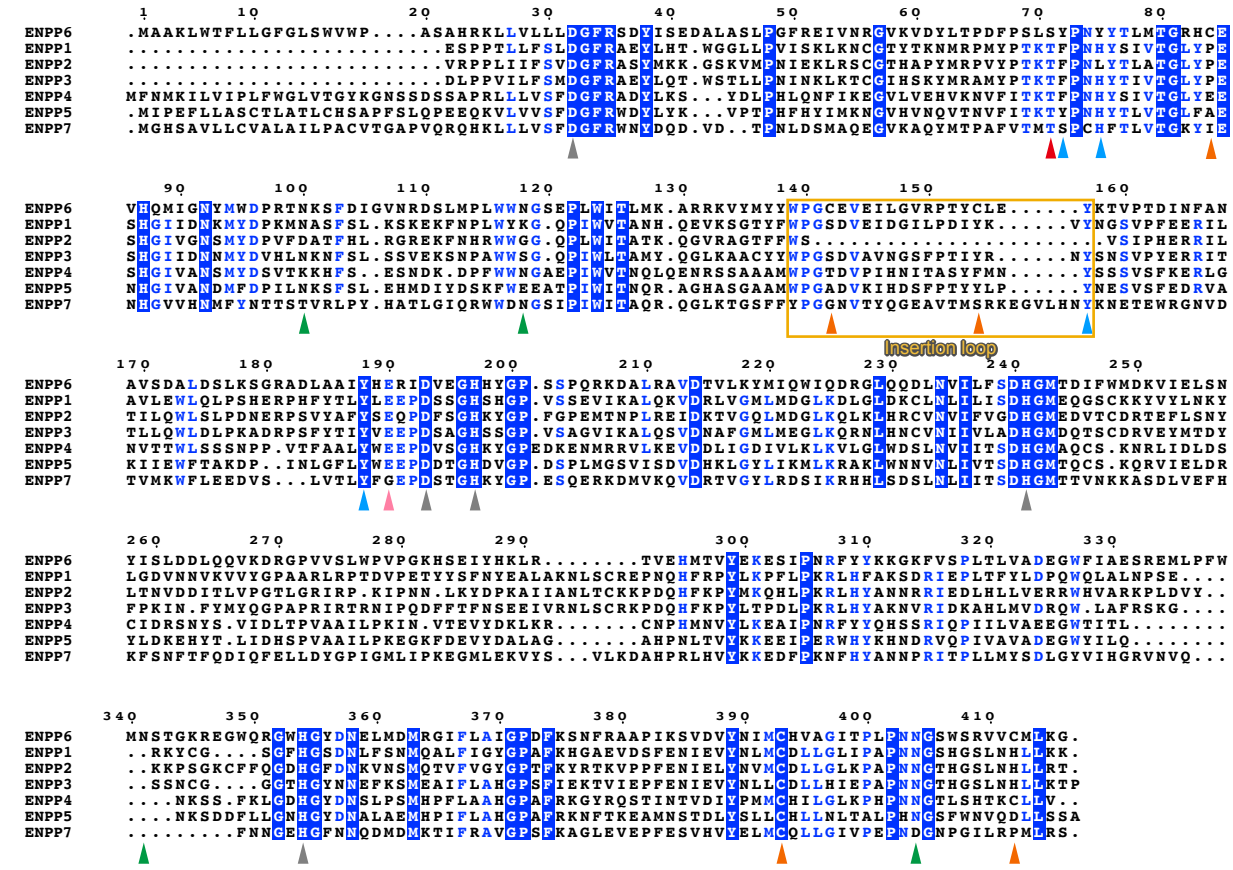

Fig. 1-14 Multiple amino acid sequence alignment of the mouse ENPP family proteins. The zinc-coordinating residues, gray triangles; the catalytic residue, red triangles; the tyrosine residues in the cholinebinding pocket, light blue triangles; the residues hydrogen bonding with the tyrosine residues in the choline-binding pocket, light pink triangles; cysteine residues, orange triangles; the *N*-glycosylated asparagine residues, green triangles; the residues in the insertion loop, orange boxes.

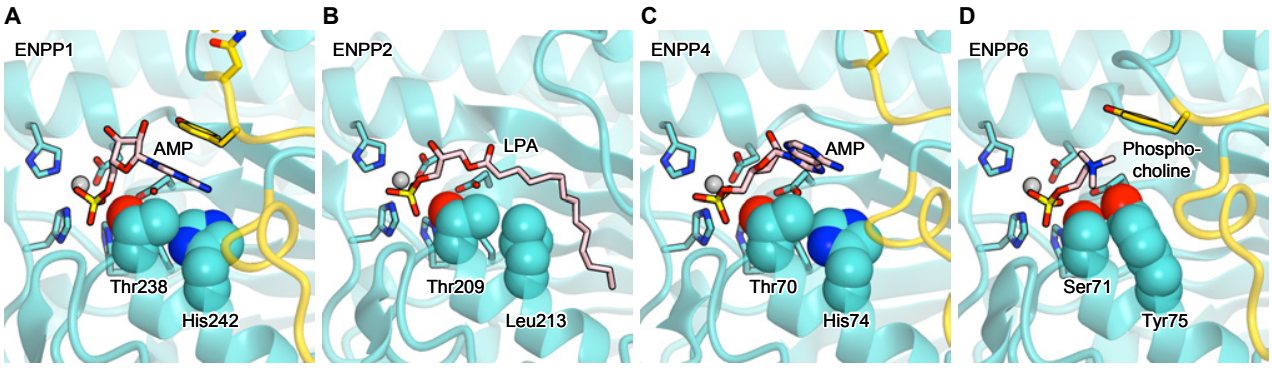

Fig. 1-15 Catalytic nucleophiles of the ENPP family members. Active sites of ENPP1 in complex with AMP (PDB ID 4GTW) (**A**), ENPP2 in complex with 14:0-LPA (PDB ID 3NKM) (**B**), ENPP4 in complex with AMP (PDB ID 4LQY) (**C**), and ENPP6 in complex with phosphocholine (**D**). Ser71 and Tyr75 of ENPP6 and corresponding residues in ENPP1, ENPP2 and ENPP4 are shown as space-filling models.

1.6 Discussion

1.6.1 Comparison of choline-binding pocket in ENPP6 and other choline binding proteins' structures

The aromatic and acidic residues forming the choline-binding pockets are conserved completely in ENPP6 proteins from various species (Fig. 1-5) and are also conserved in other choline-binding proteins[54] (Fig. 1-16). In the crystal structures of ENPP6 and other choline-binding proteins, the choline moiety of ligands form π–cation interactions with surrounding multiple aromatic residues, with an acidic residue serving as a counterion to the positively charged choline moiety. Thus, the present structure of ENPP6 provides a new example of the choline recognition mechanism of choline-binding proteins, and reinforces the notion that protein of diverse sequences has evolved convergently to recognize the choline moiety of their substrates, using choline-binding pockets of a similar structural feature.

1.6.2 Structural insight into the substrate specificity of ENPP family proteins

Previously determined crystal structures of ENPP1, ENPP2 and ENPP4 provided insights into their substrate recognition mechanisms[31–35]. A structural comparison of ENPP6 and other ENPPs offers an explanation of why among the family members only ENPP6 specifically recognizes the choline moiety of substrates. ENPP1 and ENPP4 hydrolyze nucleotide substrates to produce AMP and pyrophosphate[26,55]. In the crystal structures of mouse ENPP1, the adenine base of AMP is recognized by the conserved phenylalanine and tyrosine residues (Phe239 and Try322) through stacking interactions (Fig. 1-17 A). Similarly, in the crystal structure of mouse ENPP4[35], the adenine base is recognized by the conserved aromatic residues (Phe71 and Try154) through stacking interactions (Fig. 1-17 C). ENPP6 Tyr72/Tyr157 correspond to ENPP1 Phe239/Tyr322 and ENPP4 Phe71/Tyr154, respectively (Fig. 1-17 D). These observations suggested that ENPP6 cannot hydrolyze nucleotide substrates, due to steric clashes between the hydroxyl group of Tyr72 and the base moiety of nucleotide substrates.

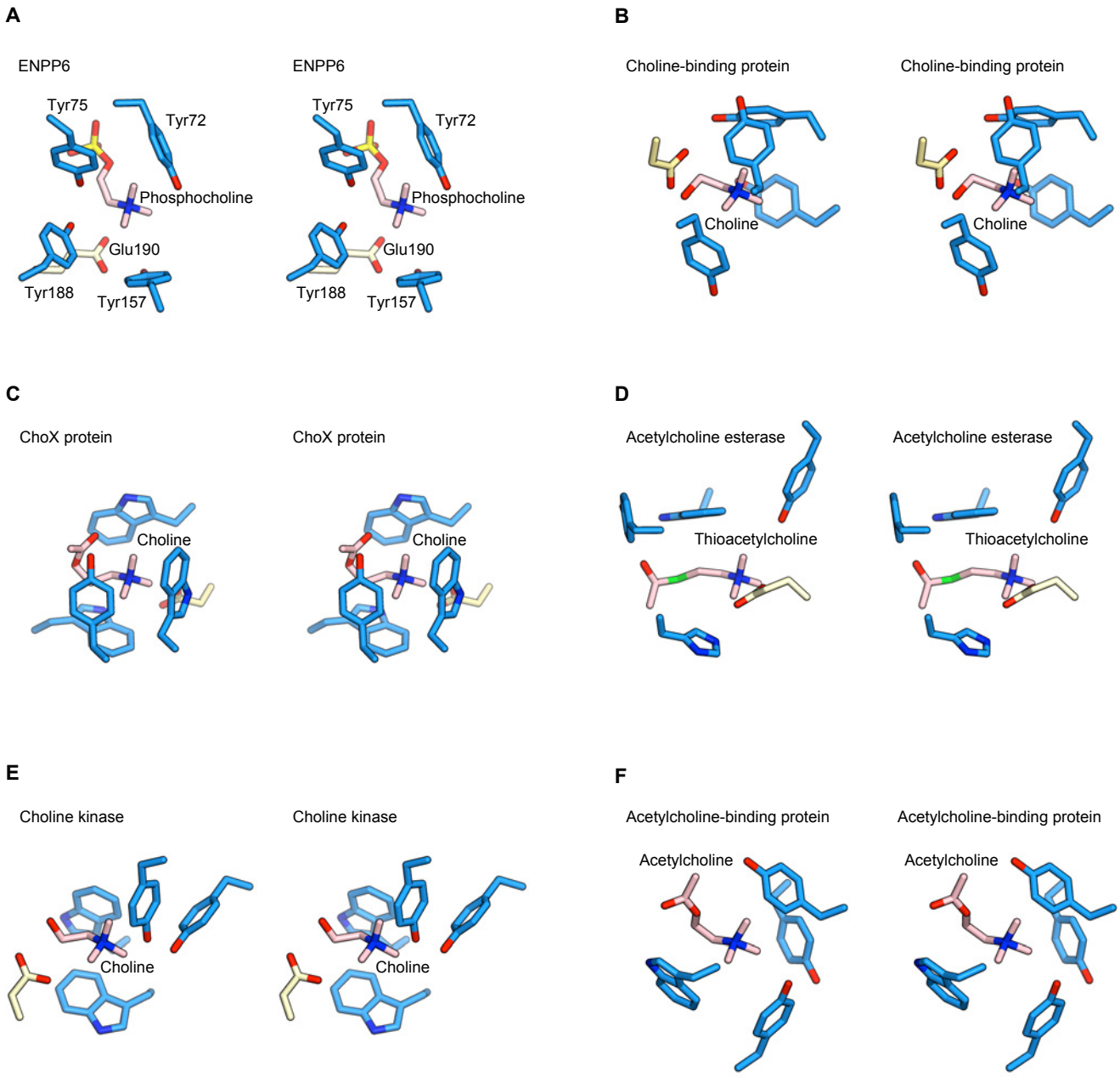

Fig. 1-16 Choline recognition by choline-binding proteins (stereo view). **A**. ENPP6. **B**. Bacteria choline-binding protein (PDB ID 3R6U). **C**. Bacteria ChoX (PDB ID 2RIN). **D**. Electric ray acetylcholine esterase (PDB ID 2C4H). **E**. Protist choline kinaes (PDB ID 3C5I). **F**. Sea hare acetylcholine-binding protein (PDB ID 2XZ5). The residues in the choline-binding pocket and the choline-containing compounds are shown as blue and light pink sticks, respectively.

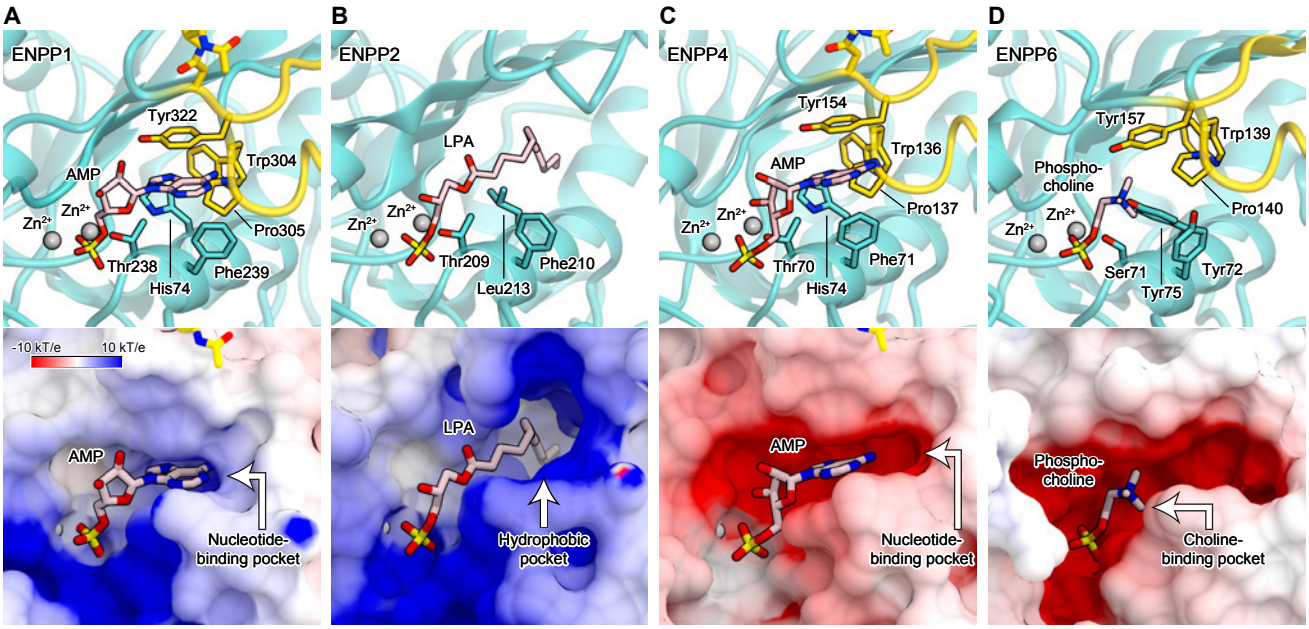

Fig. 1-17 Active sites of the ENPP family members. Active sites of ENPP1 in complex with AMP (PDB ID 4GTW) (**A**), ENPP2 in complex with 14:0-LPA (PDB ID 3NKN) (**B**), ENPP4 in complex with AMP (PDB ID 4LQY) (**C**), and ENPP6 in complex with phosphocholine (**D**). The structures are shown in ribbon representations (top panels) and in electrostatic surface potentials (contoured from −10 kT/e [red] to +10 kT/e [blue]) (bottom panels). The bound products are shown as light pink sticks. The insertion loops are shown in gold in (**A**, **C**, **D**).

1.6.3 Structural comparison of ENPP6 and ENPP2

ENPP2 hydrolyzes LPC into LPA and choline by its lysoPLD activity[22], whereas ENPP6 hydrolyzes LPC into monoacylglycerol and phosphocholine by its lysoPLC activity. Unlike the other ENPP family members, ENPP2 lacks the insertion loop and instead has a hydrophobic pocket, which accommodates the acyl chain of LPC substrates[32] (Fig. 1-17 B). In contrast, ENPP6 has the insertion loop, which occludes a hydrophobic pocket and contributes to the formation of the negatively charged choline-binding pocket, which accommodates the positively charged choline moiety of LPC (Fig. 1-17 D). These observations indicated that LPC binds to the catalytic pockets of ENPP2 and ENPP6 in the reverse orientation, thereby explaining their distinct activities for the LPC substrates.

1.7 Conclusion

In this chapter, the crystal structure of the ENPP6–phosphocholine complex was determined at 1.8 Å resolution. The structure showed that ENPP6 recognizes the choline moiety of phosphocholine, which strongly supports the idea that ENPP6 recognizes the choline moiety of its substrates, such as GPC and LPC, and hydrolyzes them to produce phosphocholine in a similar manner. Hence, the molecular basis by which ENPP6 is involved in the metabolism of choline by is provided (Fig. 1-18). In the crystal structure, the choline moiety is recognized by four conserved tyrosine residues and one conserved glutamic acid residue of ENPP6 (Fig. 1-9 B). Biochemical and biological analyses confirmed the functional significance of the choline-binding pocket for choline recognition (Fig. 1-9 C, D). Also, structural and sequence comparison revealed that differences in the substrate specificity of ENPP proteins may caused by differences in the structures of substrate-binding pocket, therby exerting diverse biological functions.

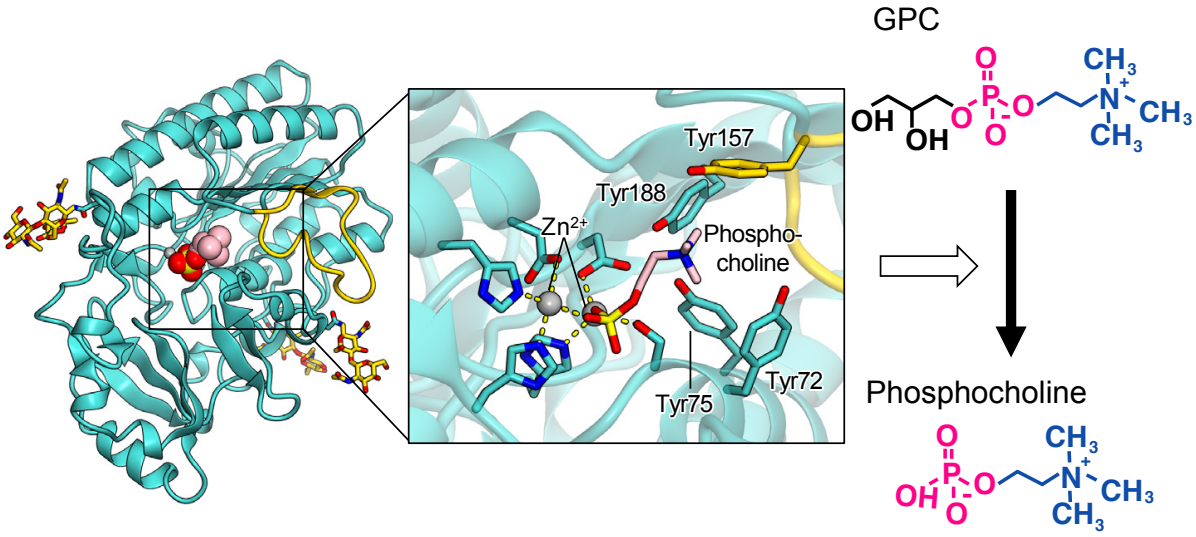

Fig. 1-18 Schematic model for catalityc activity of ENPP6.

Chapter 2 X-ray crystallographic analysis of the plant receptor-like kinase TDR in complex with the peptide hormone TDIF

2.1 CLE peptide signaling in plants

In plants, small secreted polypeptide hormones mediate the signalling pathways betIen cells, and regulate various physiological functions, including the immune response, differentiation and growth control[56]. The CLAVATA3 (CLV3)/ endosperm surrounding region-related (CLE) family peptides, encoded by *CLE* genes, are the most thoroughly studied family among these small peptide hormones. The *CLE* family genes have been found in diverse plant species, including dicots, monocots, conifers, mosses and green algae[57–59]. They are expressed in various tissues by developmental and environmental cues and regulate numerous biological processes, including the proliferation and differentiation of stem cells, and the development of the meristem, the vascular system and the embryo[60,61]. In *Arabidopsis*, 32 CLE family genes have been identified, and all of them encode pre-pro-proteins with an N-terminal signal peptide, a central variable domain, and 14 amino-acid peptides with the conserved CLE domain at their C-termini[57,62]. Then, the C-terminal 12 or 13 amino acids are cleaved by processing to yield the mature CLE peptide. As well as native mature CLE peptides, some of chemically synthesized CLE peptides are sufficient to have CLE activities.[63–65] The CLE peptides are secreted into the apoplast, after proteolytic processing and post-translational modifications. They mainly interact with the extracellular regions of membrane associated leucine-rich receptor kinases (LRR-RKs), thereby triggering intracellular signalling cascades[57,66]. Several ligand–receptor pairs of the CLE family peptides have been identified, including CLV3–CLV1 and CLE9/10–BAM1[61]. CLE peptides, corresponding receptors and their functions are summarized in Table 2-1.

Table 2-1. Known CLE peptides, corresponding putative receptors and their functions.

CLE	Amino acid sequence	Receptors	Functions	Selected references
CLV3	RTVPSGHDPLHH	CLV1, 2, CRN, RPK2, BAM1, 2	Maintenance of the SAM	Clark et al. (1993, 1995,1997)[67–69] Kayes and Clark (1998)[70] Brand et al. (2000)[71] Schoof et al. (2000)[72] Kondo et al. (2006)[65] Muller et al. (2008)[73] Ogawa et al. (2008)[74] Ohyama et al. (2009)[62] Guo et al. (2010)[75] Kinoshita et al (2010)[76] Yadav et al. (2011)[77] Betsuyaku et al. (2011)[78] Shinohara & Matsubayashi (2013, 2015)[79,80] Ishida et al. (2014)[81]
CLE2	RLSPGGPDPQHH	CLV1	?	Ogawa et al. (2008)[74] Ohyama et al. (2009)[62]
CLE9	RLVPSGPNPLHN	BAM1, 2, 3	?	Shinohara et al. (2012)[82]
CLE10	RLVPSGPNPLHN	CLV2	Inhibition of protoxylem vessel formation	Kondo et al. (2011)[83]
CLE19	RVIPTGPNPLHN	CLV2-CRN	Maintenance of the RAM	Casamitjana-Martinez et al. (2003)[84] Fiers et al. (2005)[85] Tamaki et al. (2013)[86] Xu et al. (2015)[87]
CLE40	RQVPTGSDPLHH	ACR4-CLV1	Maintenance of the RAM, stem cell niche	Hobe et al. (2003)[88] Stahl et al. (2009, 2013)[89,90]
CLE41/ 44/ TDIF	HEVPSGPNPISN	TDR/PXY	Vascular stem cells maintenance	Ito et al. (2006)[64] Hirakawa et al. (2008, 2010)[66,91] Etchells & Turner (2010)[92] Whitford et al. (2008)[93] Kondo et al. (2014)[94]
CLE42	HGVPSGPNPISN	TDR/PXY	Vascular bundle maintenance	Yaginuma et al. (2011)[95]
CLE45	RRVRRGSDPIHN	SKM1, 2z, BAM3	Pollen tube growth, root phloem development	Endo et al. (2013)[96] Depuydt et al. (2013)[97] Rodriguez-Vialon et al. (2014, 2015)[98,99]
FON2/4	RSVPAGPDPMHH	FON1	Maintenance of the floral meristems	Nagasawa et al. (1996)[100] Chu et al. (2006)[101] Suzaki et al. (2006)[102]

2.2 CLE peptide TDIF and its receptor TDR

Tracheary element differentiation factor (TDIF), a CLE family peptide, was first isolated from the xylogenic cell culture of *Zinnia elegans*. TDIF consists of 12 amino acids including two hydroxyproline residues, and is encoded by the *CLE41* and *CLE44* genes in *Arabidopsis*[64]. TDIF is secreted from the phloem and is perceived by its specific receptor, TDR/PXY (TDIF receptor/phloem intercalated with xylem), which is located on the procambial cell membrane[66]. TDR was originally identified as the component that maintains the cell polarity required for the orientation of cell division, during vascular development in the vascular meristem[103]. TDR belongs to the LRR-RK class XI subfamily, and consists of the extracellular LRR domain and the cytoplasmic kinase domain[103]. TDR recognizes TDIF by its extracellular LRR domain, and activates the intracellular GSK3 pathway to suppress the transcription factor BES1, which promotes cellular differentiation into tracheary elements (Fig. 2-1)[94]. Simultaneously, TDR activation enhances the transcription of the homeobox genes *WOX4* and *WOX14*, which promote procambial cell proliferation[91,104]. Furthermore, the tissue-specific co-overexpression of the TDR–TDIF pair increases wood formation by twofold in the aspen tree, suggesting that the modulation of the TDR–TDIF pathway can be applied for biomass engineering[105].

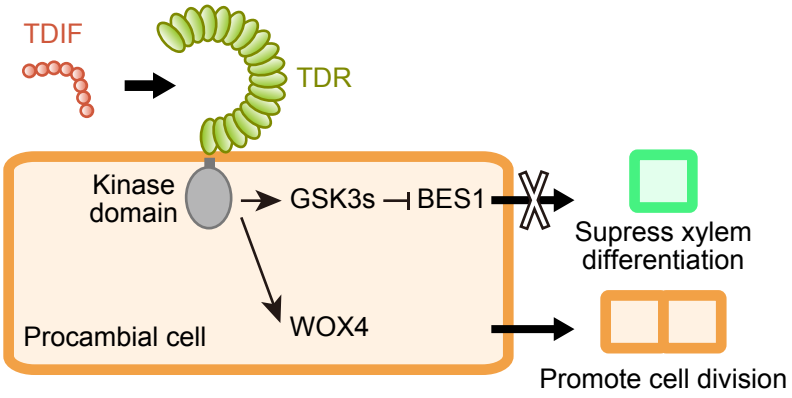

Fig. 2-1 Schematic model for TDIF–TDR signaling pathway that controls cell fate determination in procambial cells.

2.3 Structures of plant LRR-RKs

Recently, several crystal structures of plant LRR-RKs in complex with their ligands Ire reported. The crystal structures of BRI1 and PSKR1 in complex with brassinolide and phytosulfokine, respectively, revealed the recognition mechanism of phytohormones by LRR-RKs (Fig. 2-2)[106–108]. Furthermore, the crystal structures of FLS2 and PEPR1 in complex with the bacterial flagella-derived peptide flg22 and the self-derived danger signal Pep1, respectively, demonstrated the epitope-recognition mechanism by LRR-RKs involved in the immune systems of plants[109,110]. HoIver, the molecular mechanism of the CLE family peptide recognition by LRR-RKs has remained elusive, due to the lack of the crystal structures of LRR-RKs in complex with CLE peptides.

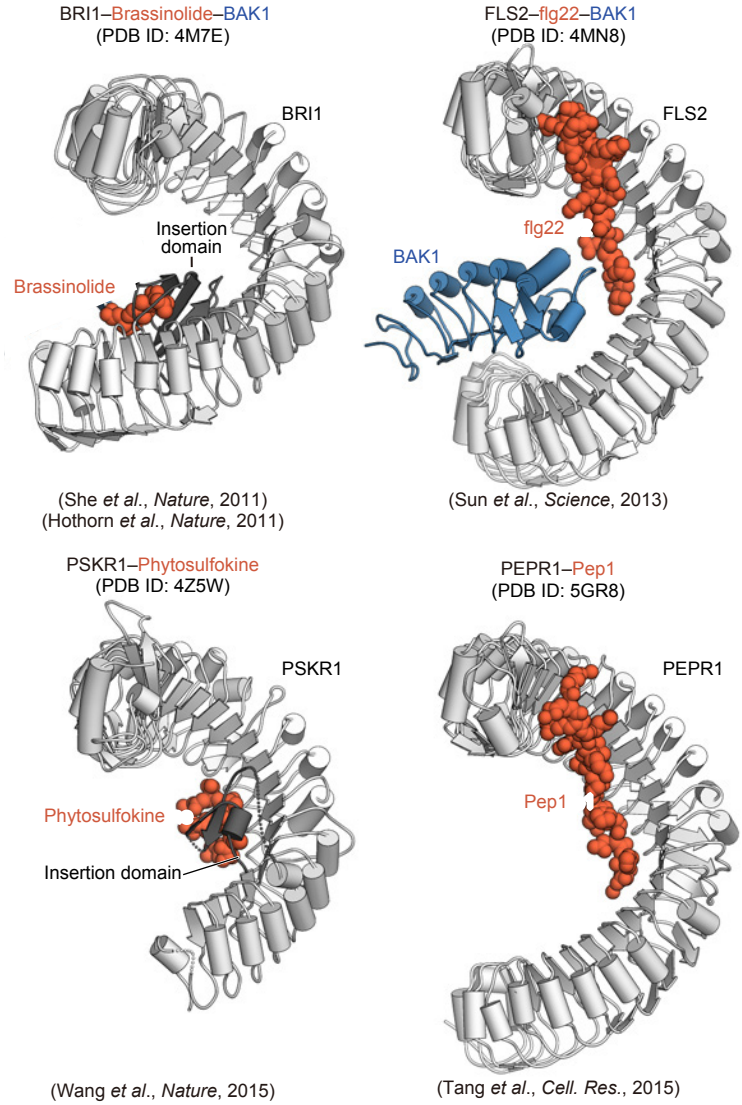

Fig. 2-2. Structures of plant LRR-RKs. LRR-RKs, co-receptors and ligands are colored white, blue and orange, respectively. The insertion domain of BRI1 and PSKR1 are colored dark grey.

2.4 Overview of this study

Here, I report the crystal structure of the TDR–TDIF complex. In conjunction with the structure-guided functional analysis, structural insights into the mechanism of TDIF recognition by TDR are provided. By comparing the structure of TDR–TDIF complex and known plant LRR-RK–ligand complex, the conserved structural features and variable structural features are revealed. Structure and sequence comparisons provided insights into the recognition mechanism of the CLE family peptides by their cognate receptors.

2.5 Material and Methods

2.5.1 Plasmid construction

DNA fragments encoding the extracellular LRR domain of TDR from *Arabidopsis thaliana* (residues 31–631) are amplified by PCR using pDONR-221-TDR plasmid (gifted from Dr. Yuki Kondo and Prof. Hiroo Fukuda, The University of Tokyo) as templates, PrimeSTAR MAX DNA polymerase (Takara Bio) and the following primers: 5′-ATTGTGGGTTCAGCGAAGTTTTCACCTCAACTCTTGTCTCTC-3′ and 5′-AACTTCCAGGCCGCTATCAGAATTGCAAGGTTTTCCGAC-3′. The PCR products were inserted into pFastBac1 vector (Invitrogen), which had been modified to contain a N-terminal Hemolin peptide from *Hyalophora cecropia* (residues 1–18) and a C-terminal HRV3C protease cleavage site (LEVLFQGP) followed by a 10×His tag (pHem-HRV3C- His10-TDR; Fig. 2-3). Plasmids were generated using In-Fusion methods (Clontech). To improve the solution behaviour of TDR, two less-conserved cysteine residues were mutated (C259A/C540S) by a PCR-based method using pHem-HRV3C-His10-TDR as templates, PrimeSTAR MAX DNA polymerase (Takara Bio) and the following primers: C259A, 5′-CTTTGACGTTTCCAATGCCAGCCTCTCTGGTTC-3′ and 5′-GAACCAGAGAGGCTGGCATTGGAAACGTCAAAG-3′; C540S, 5′-CTTCTCTCTTTGAATCTCAGCCAAAATCATC-3′ and 5′-CTTCTCGCAATGTCCGATG-3′. The DNA sequences were verified by DNA sequencing.

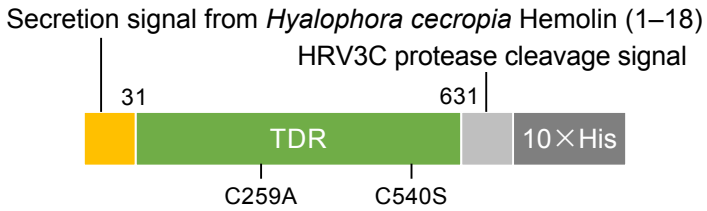

Fig. 2-3 Schematic diagram of the TDR crystallization construct

2.5.2 Cell culture and protein expression

The wild type and mutants of TDR were expressed as secreted forms in Sf9 cells using Bac-to-Bac baculovirus expression system methods (Invitrogen). In detail,

pHem-HRV3C-His10-TDR (Wild type or C259A/C540S) plasmids generated in 2.5.1 were transformed into DH10Bac competent cells (Thermo fisher scientific) to produce bacmids, which were subsequently transfected into Sf9 cells using FuGENE HD transfection reagent (Promega) to generate baculoviruses. Sf9 cells were infected with P3 vaculoviruses at a density of approximately 2×10^6 cells/mL. After a 72 h incubation at 27 °C, the culture supernatant was collected by centrifugation (6,000g, 10 min, 4 °C) and then was supplemented with 1/100 volume of 1 M Tris-HCl pH 7.5.

2.5.3 Protein purification

All purification operations were performed on ice or at 4 °C. The culture supernatant was filtrated with bottle top filter (Corning) and then incubated with Ni Sepharose excel resin (GE Healthcare) overnight. The resin was then washed with buffer consisting of 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 20 mM imidazole on an Econo-Column (Bio-Rad). The bound protein was eluted with buffer consisting of 20 mM Tris-HCl, pH 7.5, 500 mM NaCl and 500 mM imidazole, and then dialyzed against 20 mM Tris-HCl, pH 7.5 and 150 mM NaCl to remove the imidazole. The protein was incubated with Talon cobalt affinity resin (Clontech). The resin was washed with buffer supplemented with 20 mM imidazole, and then the protein was eluted with buffer supplemented with 300 mM imidazole. The eluted protein was dialyzed against 20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 20 mM imidazole with HRV3C protease to cleave the C-terminal His tag, and then was passed through the Talon column again. The eluted protein was dialyzed against 20 mM sodium citrate, pH 5.5, 500 mM NaCl with Endo H (NEB) to digest *N*-linked glycans at 20 °C overnight. To facilitate crystallization, the lysine residues of the deglycosylated TDR protein were methylated, as previously described[111]. In brief, the protein was incubated with the dimethylamine-borane complex (ABC) and formaldehyde in 50 mM HEPES, pH 7.5, at 4 °C overnight. The methylated TDR protein was further purified by gel filtration chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare) in 20 mM Tris-HCl pH 7.5, 500 mM NaCl, and concentrated to 3 mg/mL using an Amicon Ultra-4 filter 30 kDa MWCO (Millipore). The purity of the protein was

assessed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) under nonreducing conditions, and the gels were stained with Simply Blue SafeStain (Invitrogen).

2.5.4 Crystallization

A mixture of TDR and a chemically synthesized TDIF peptide (HEV(Hyp)SG(Hyp)NPISN, Eurofins Genomics) dissolved in DMSO, at a molar ratio of 1:1.5, was crystallized at 20 °C by the vapour diffusion method. Initial crystallization screening was performed by the sitting-drop vapour-diffusion method in an MRC 96-well Crystallization Plate (Molecular Dimensions) using the following screening kits: Crystal Screen (Hampton Research), The PACT Suite, The JCSG+ Suite (Qiagen), JBScreen Classic 1, 2, 4 and 5 (Jena Bioscience) Memgold (Molecular Demensions) and SaltRX (Hampton Research). Crystals of TDR in complex with TDIF peptide were obtained by mixing 100 nl protein solution (3 mg/mL TDR, 75 μM TDIF, 20 mM Tris-HCl, pH 7.5 and 150mM NaCl) and 100 nl reservoir solution using a Mosquito crystallization robot (TTP Labtech). Initial hit was optimized by the hanging-drop vapour-diffusion method by mixing 1 μl of protein solution and 1 μl of reservoir solution (4 M sodium nitrate, 0.1 M sodium acetate trihydrate, pH 4.8 and 200 mM ammonium sulfate). Crystallization conditions were also optimized with the Additive Screen kit (Hampton Research). Crystals were cryoprotected in reservoir solution supplemented with 6 M sodium nitrate, and were flash-cooled in liquid nitrogen.

2.5.5 Data collection and structure determination

X-ray diffraction data were collected at 100 K on the beamline PXI at the Swiss Light Source using a Pilatus 6M detector, and processed using DIALS[112]. The structure was determined by molecular replacement with MOLREP[41] using truncated FLS2 (residues 56–611, PDB ID 4MNA), in which the residues that did not align with TDR were removed, as the initial search model. Automated model building was performed with Buccaneer[113], and the resulting model was manually completed with COOT[43]. The refinement was performed using REFMAC[114] and PHENIX[44]. The stereochemical qualities of the models were validated by

Ramachandran plot analysis with the program RAMPAGE[45]. The data-collection and refinement statistics are summarized in Table 2.2.

2.5.6 Mutant protein preparation

The TDR mutants were prepared by a PCR-based method, using the expression vector encoding the TDR C259A/C540S mutant as the template PrimeSTAR MAX DNA polymerase (Takara Bio) and the following primers: F161A, 5′-CTTAAAAGTCTTCAATGCGGCCAGCAACAACTTCG-3′ and 5′-CGAAGTTGTTGCTGGCCGCATTGAAGACTTTTAAG-3′; S162A, 5′-GTCTTCAATGCGTTCGCCAACAACTTCGAAGG-3′ and 5′-CCTTCGAAGTTGTTGGCGAACGCATTGAAGAC-3′; D255E, 5′-CAAATCTCAAGTACTTTGAAGTTTCCAATGCCAGC-3′ and 5′-GCTGGCATTGGAAACTTCAAAGTACTTGAGATTTG-3′; R421A, 5′-GCGAATCTCTATGGGCGTTTCGGAGTCAAAAC-3′ and 5′-GTTTTGACTCCGAAACGCCCATAGAGATTCGC-3′; R423A, 5′-CTCTATGGCGGTTTGCGAGTCAAAACAATCG-3′ and 5′-CGATTGTTTTGACTCGCAAACCGCCATAGAG-3′; R421A/R423A, 5′-CTCTATGGGCGTTTGCGAGTCAAAACAATCG-3′ and 5′-CGATTGTTTTGACTCGCAAACGCCCATAGAG-3′, and the sequences were verified by DNA sequencing. Baculovirus generation and protein expression were performed in the same manner as that for the TDR C259A/C540S mutant. The mutant proteins were purified using Ni Sepharose excel resin, in a similar manner to that for the TDR C259A/C540S mutant.

2.5.7 Pull-down assay

The biotinylated TDIF peptide (2 μg) (HEV(Hyp)SG(Hyp)NPISN-GSGS-(Ahx)-Biotin, Eurofins Genomics) was mixed with 10 μl of Dynabeads M-280 Streptavidin (Life Technologies) beads at room temperature for 30 min. The beads were washed with 200 μL of wash buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl) three times. To test the interaction between TDR and

TDIF, the purified TDR protein was mixed with the biotinylated TDIF-bound beads in the presence of 0.1 mg/mL BSA, and then incubated at 4 °C for 2 h. The beads were then washed three times with 200 μL wash buffer. The bound proteins were eluted by boiling in SDS sample buffer and analysed by SDS–PAGE.

2.5.8 FRET analysis

FRET analysis was performed by Dr. Yuki Kond and Prof. Hiroo Fukuda (The University of Tokyo).

Mutated TDR variants were produced by site-directed mutagenesis with the pDONR221 vector harbouring TDR. The coding sequences of BIN2 and the mutated TDR were transferred from the entry clones into the destination vectors by the LR recombination reaction, to generate 35 S-BIN2–YFP and pER8-TDR(mut)-CFP. These expression vectors were transformed into the *Rhizobium radiobacter* strain GV3101 MP90. The cultured agrobacterium suspended in infiltration buffer (10 mM MES, 10 mM $MgCl_2$ and 150 μM acetosyringone, pH 5.7) was injected into *Nicotiana benthamiana* leaves with a 1 mL syringe (Terumo). After 2 days of infiltration, the tobacco leaf disks were treated with 10 μM estradiol for 18 h. After confirming YFP and CFP expression in the leaf epidermis, the FRET analysis was performed according to the acceptor photobleaching method, using an LSM-510 META confocal microscope (Carl Zeiss). First, the CFP fluorescent signal intensities at the plasma membrane was measured. Subsequently, BIN2–YFP was photobleached with 514 nm laser irradiation at the plasma membrane, and then the CFP signal intensities were re-measured. For the detection of the CFP fluorescent signal, a 458 nm excitation laser and a 477–520 nm emission filter were used. FRET efficiencies were calculated by comparing CFP fluorescent intensities before and after photobleaching.

2.6 Results

2.6.1 Structure determination

Initially, the extracellular domain (residues 31–631) of TDR from *Arabidopsis thaliana* was prepared, using the Sf9-baculovirus expression system and purified (Fig. 2-4). To improve the solution behaviour of the TDR extracellular domain, mutations were introduced to two less-conserved cysteine residues (C259A/C540S, Fig. 2-3) and performed the crystallization screening using the C259A/C540S TDR mutant. I attempted the crystallization screening of the complex with a synthesized TDIF peptide (HEV(Hyp)SG(Hyp)NPISN), but no crystals were obtained.

Prediction by the NetNGlyc 1.0 Server[115] revealed that 11 *N*-glycosylation motifs (NxS/T) were highly likely to be modified in the extracellular domain of TDR (Fig. 2-5). To increase the chance of the quality of crystallization, *N*-linked glycans of TDR were enzymatically digested (Fig. 2-6 A). The resulting change in molecular weight after the deglycosylation was detected by as shifts in gel mobility (Fig. 2-6 B). The deglycosylated TDR and TDIF yielded crystals that diffracted to around 5 Å, which was not of sufficient quality for data collection (Fig. 2-6 C, D).

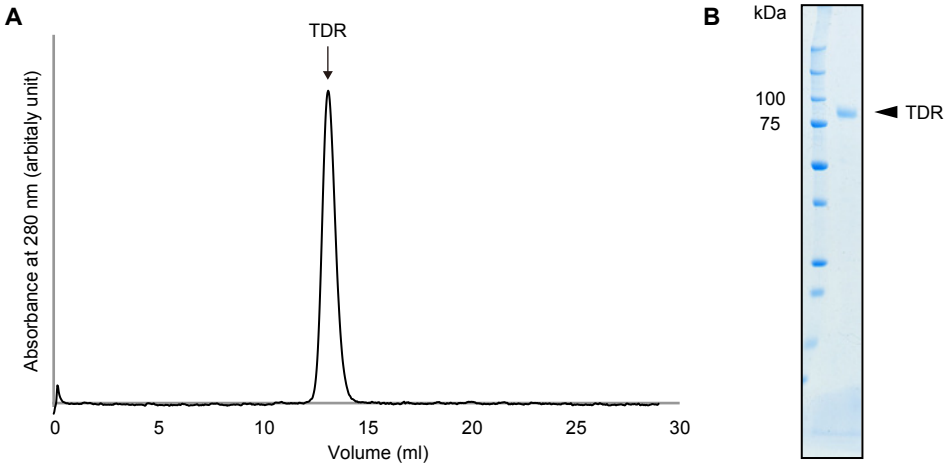

Fig. 2-4 Purification of the extracellular domain of TDR. A. Gel-filtration analysis of the extracellular domain of TDR. The proteins were chromatographed on a Superdex 200 Increase 10/300 GL column. **B**. SDS–PAGE analysis of purified TDR.

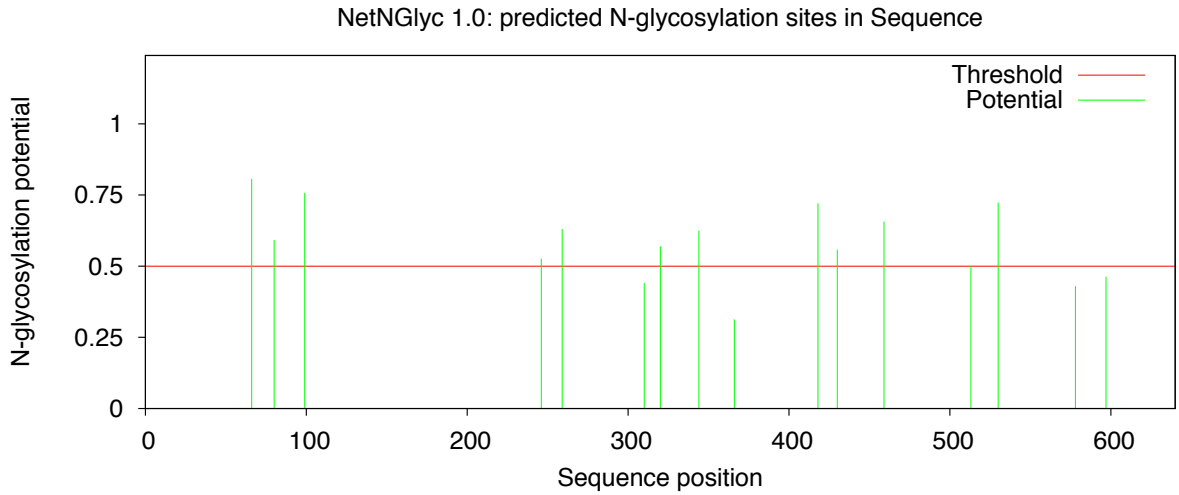

Fig. 2-5 Potential *N*-glycosylation sites of the extracellular domain of TDR predicted by NetNGlyc 1.0 server.

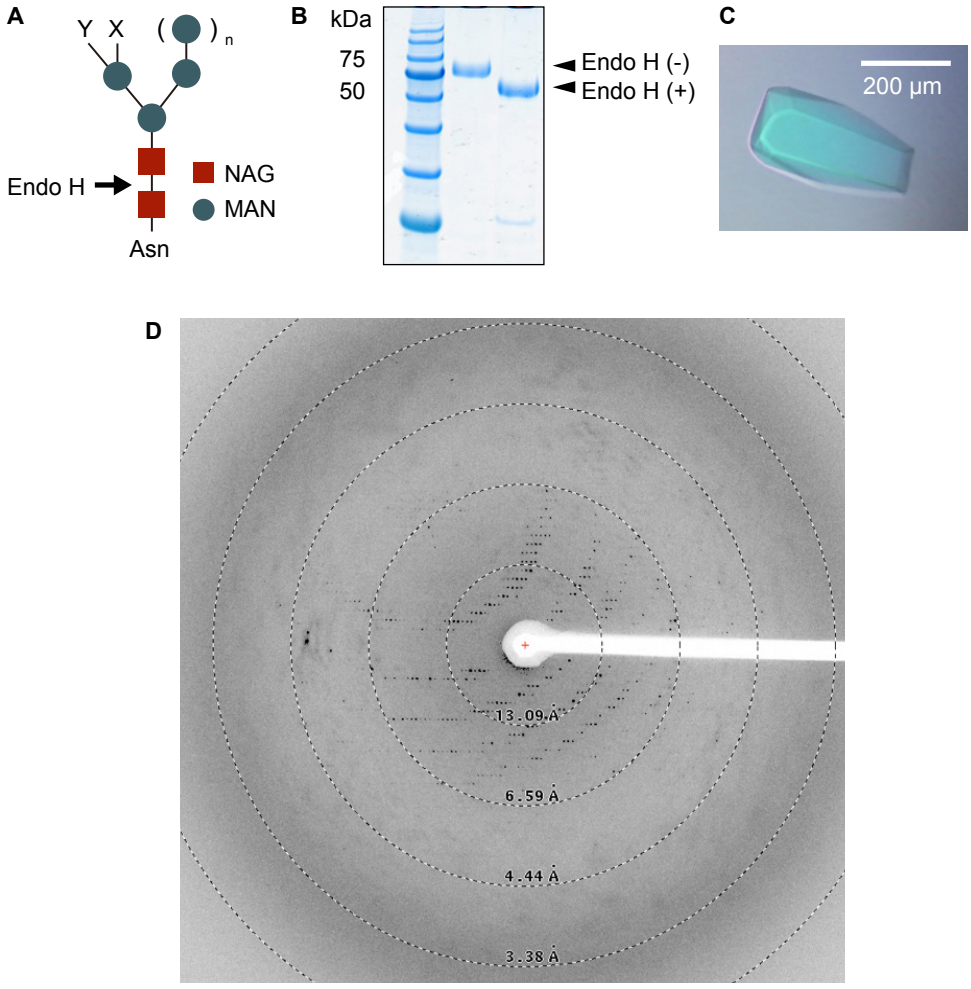

Fig. 2-6 Enzymatic cleavage of *N*-linked glycans. A. Cleavage site for Endo H in *N*-linked glycans. NAG, red square; MAN, blue circle. High mannose glycan: (Man)1-2, X, Y.; n=2~150 Hybrid glycan: AcNeu-Gal-GlcNAc, X and/or Y; n=2. **B.** Mobility shift analysis of TDR before and after Endo H treatment. **C.** A Crystal of the Endo H-treated TDR C259A/C540S mutant and TDIF. **D.** X-ray diffraction image of the Endo H-treated TDR C259A/C540S mutant and TDIF.

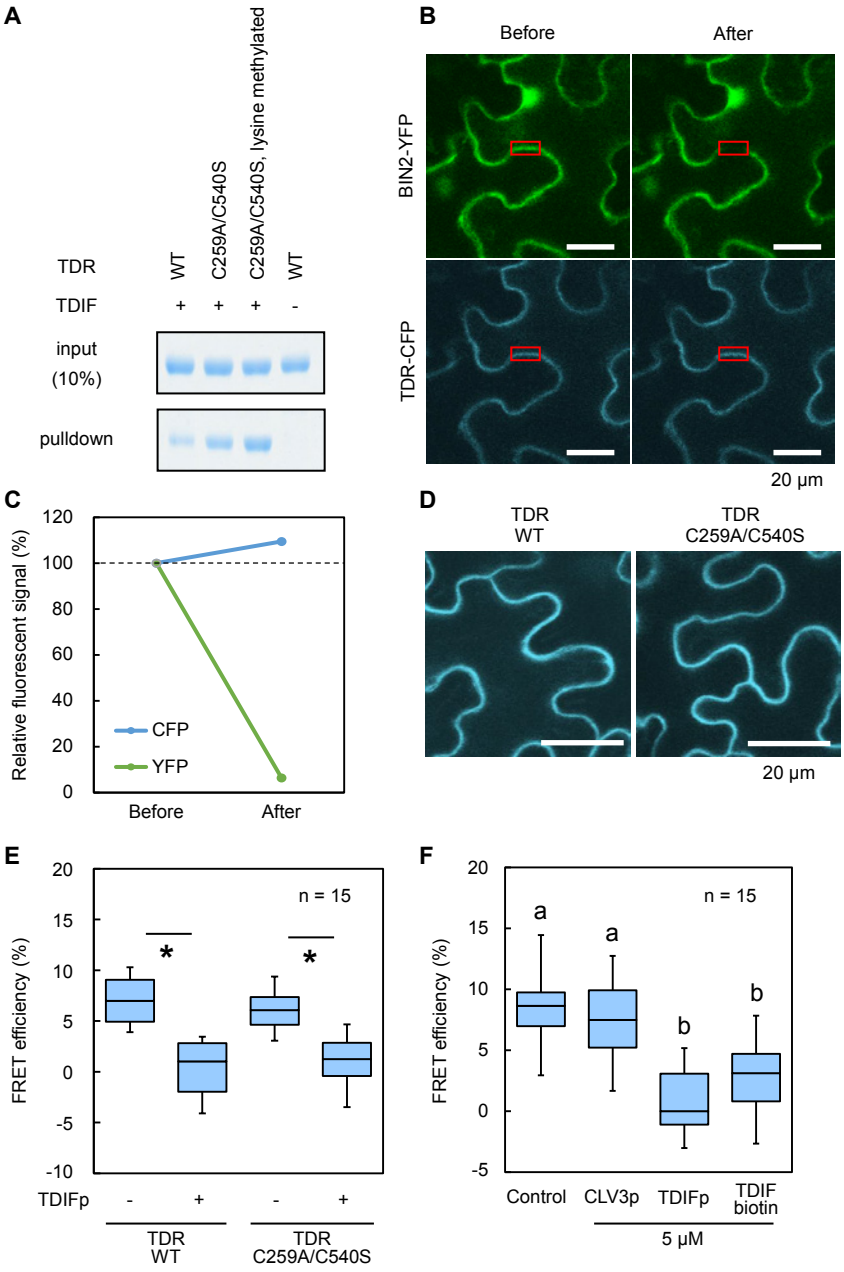

Fig. 2-7 Activities of TDR construct used for crystallization. A. Pull-down experiments between biotinylated-TDIF and TDR WT and the lysine methylated C259A/C540S TDR mutant. The wild type and C259A/C540S mutant of TDR were expressed and purified from Sf9 insect cells, and mixed with Streptavidin beads in the presence or absence of biotinylated TDIF. Bound proteins were eluted with SDS sample buffer and analyzed by SDS-PAGE. **B**. Subcellular localization of BIN2–YFP and TDR–CFP in *N. benthamiana*, before and after photobleaching. Areas surrounded by red rectangles indicate the regions photobleached by the 514 nm laser. Scale bars indicate 20 µm. **C**. Quantification of the fluorescent signals of BIN2–YFP and TDR–CFP in the red rectangle areas in **B**. The relative fluorescent signals were calculated by comparison with the signal before photobleaching and are shown as a percentage (%). **D**. Localization of TDR–CFP variants in *Nicotiana benthamiana* epidermis. Scale bars indicate 20 µm. **E**. TDIF responses in the TDR C259A/C540S mutant. TDIF responses were evaluated by measuring FRET efficiencies between TDR–CFP and BIN2–YFP. The boxplot diagram displays the FRET efficiencies upon incubations with or without 5 µM TDIF for 30 min. Significant differences according to the Student's t-test are indicated by asterisks ($P < 0.01$; $n = 15$). **F**. Bioactivities of synthetic peptides. TDIF activities of synthetic peptides were evaluated by measuring FRET efficiencies between TDR–CFP and BIN2–YFP. The boxplot diagram displays the FRET efficiencies with 5 µM synthetic peptides applied for 30 min. Significant differences ($P < 0.05$) are indicated by distinct letters (Tukey's test; $n = 15$).

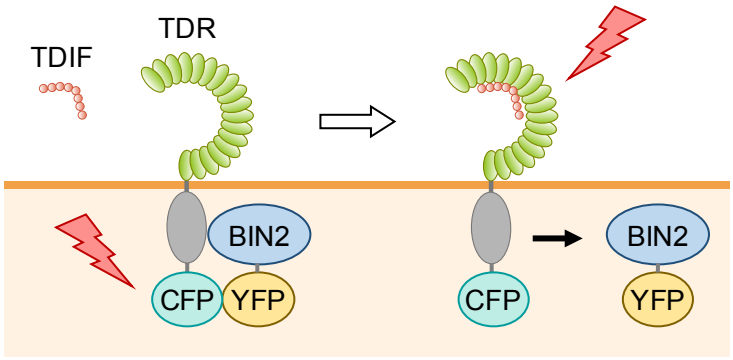

Fig. 2-8. Schematic model of the reconstituted TDR–TDIF pathway in *Nicotiana benthamiana* leaves.

In order to improve the quality of the crystals, lysine residues of TDR were methylated as previously described[111]. A pull-down assay using the biotinylated TDIF peptide confirmed that the deglycosylated and lysine methylated C259A/C540S mutant has TDIF-binding activity comparable to that of the wild-type protein (Fig. 2-7 A). Furthermore, to measure the signalling activity of the C259A/C540S mutant, the TDIF–TDR pathway was reconstituted in *Nicotiana benthamiana* leaves with TDR or its mutants fused with the cyan fluorescent protein (TDR–CFP), together with BIN2, a member of GSK3s, fused with the yellow fluorescent protein (BIN2–YFP), as previously described[94] (Fig. 2-7 B, C, Fig. 2-8, performed by Dr. Yuki Kondo and Prof. Hiroo Fukuda). It was confirmed that the C259A/C540S mutant TDR localizes to the plasma membrane, similarly to wild-type TDR (Fig. 2-7 D). Both the wild-type TDR and C259A/C540S mutant showed reduced fluorescence resonance energy transfer (FRET) efficiency of TDR–CFP with BIN2–YFP, on the TDIF peptide application, indicating that C259A/C540S mutant TDR possesses TDIF responsiveness comparable to that of the wild-type protein (Fig. 2-7 E). In addition, the biotinylated TDIF peptide used for pull-down assays also reduced the FRET efficiency, indicating that the biotinylated TDIF peptide is biologically active (Fig. 2-7 F).

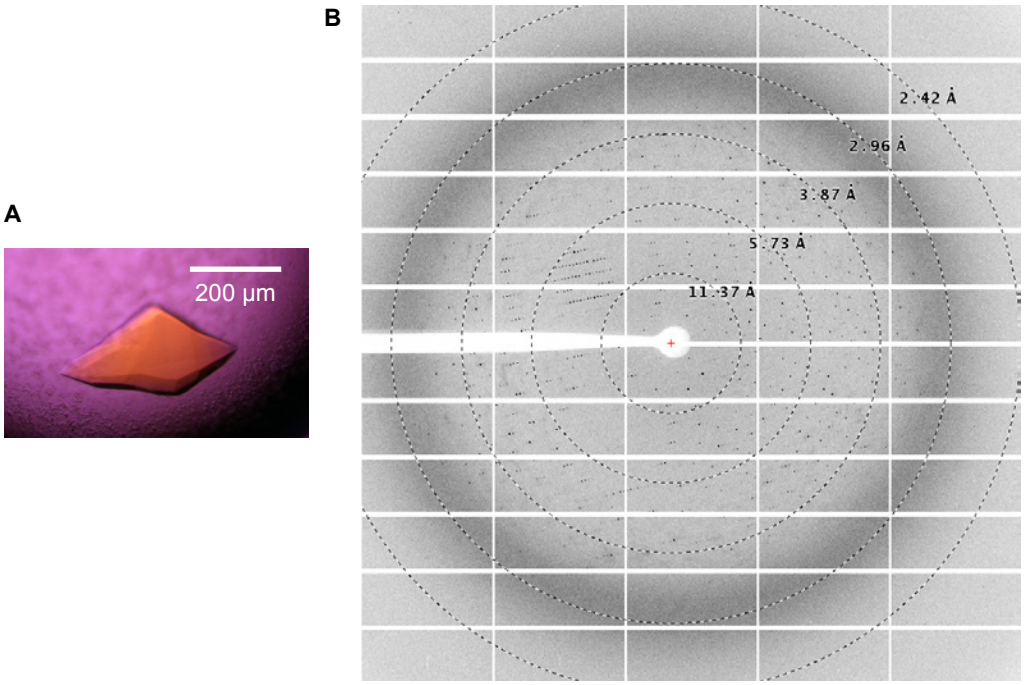

Fig. 2-9 Crystals and X-ray diffraction image of the lysine-methylated TDR. A. Crystals obtained from methylated TDR in complex with TDIF. **B.** X-ray diffraction image of the TDR–TDIF complex.

Thus the C259A/C540S mutant was crystallized TDR in complex with TDIF. After the initial screening, crystals were obtained under condition No. 46 of SaltRX Screen [4 M sodium nitrate, 0.1 M sodium acetate trihydrate pH 4.6]. Finally, the crystals of the TDR–TDIF complex were obtained, which diffracted to 3.0 Å resolution (Fig. 2-9 A, B) and belonged to space group $P3_121$, with unit-cell parameters $a = b = 132.9$, $c = 229.8$ Å, $\alpha = \beta = 90.0°$, $\gamma = 120.0°$. The crystal structure was determined at 3.0 Å resolution by molecular replacement, using the structure of the partial extracellular LRR domain of FLS2 (PDB ID: 4MN8) as the search model (Fig. 2-10, Table 2-2).

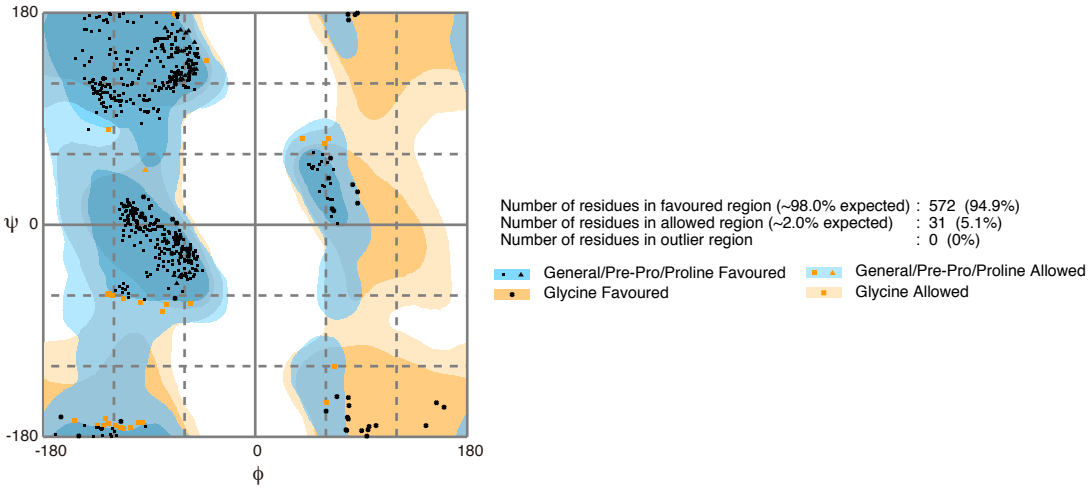

Fig. 2-10. Ramachandran plots of the TDR–TDIF complex structure.

Table 2-2. Data collection and refinement statistics

Data collection	
Beamline	SLS PX I
Wavelength (Å)	1.00
Space group	$P3_121$
Cell dimensions	
a, *b*, *c* (Å)	132.9, 132.9, 229.8
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	103–3.00 (3.11–3.00)
R_{sym}	0.053 (0.436)
$I/\sigma I$	8.8 (1.4)
Completeness (%)	99.8 (99.9)
Redundancy	9.7 (10.2)
CC(1/2)	0.995 (0.800)
Refinement	
Resolutions (Å)	57.6–3.00 (3.06–3.00)
No. Reflections	47,591
R_{work} / R_{free}	0.220 / 0.238
No. atoms	
Protein	5,089
Solvent	0
B-factors (Å^2)	
Protein	80.9
R.m.s. deviations	
Bond length (Å)	0.004
Bond angles (°)	1.210
Ramachandran plot	
Favored region	94.9%
Allowed region	5.1%
Outlier region	0.0%
PDB ID	5GIJ

*Values in parentheses are for the highest-resolution shell.

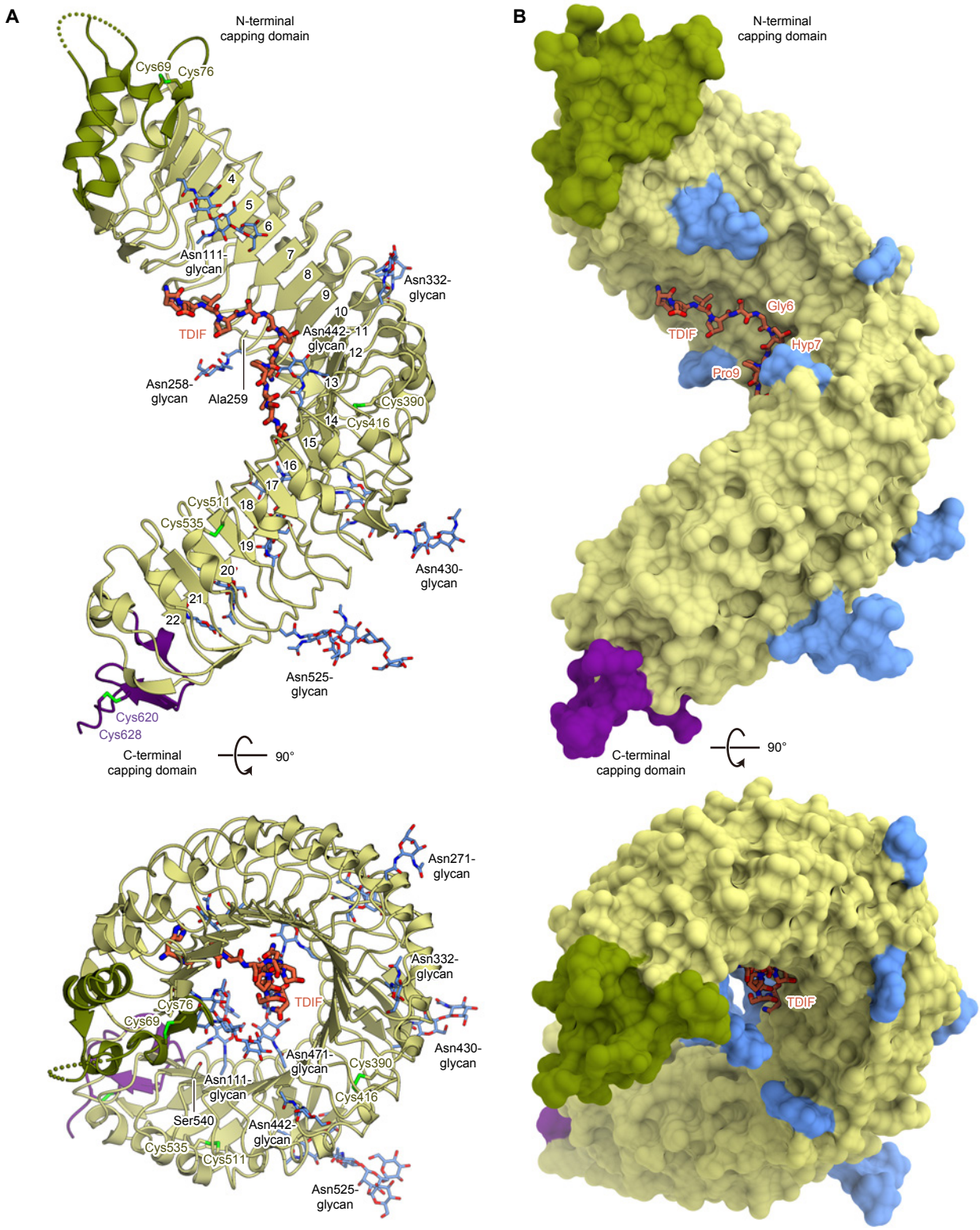

Fig. 2-11 Overall structure of the TDR–TDIF complex. A, B. Ribbon representation (**A**) and surface representation (**B**) of the crystal structure of the extracellular LRR domain of TDR in complex with TDIF. TDIF, N-linked glycans and Cys residues that form disulfide bonds are shown as orange, blue and yellow-green sticks, respectively. Two residues mutated from cysteines (Ala259 and Ser540) are also shown as sticks. The N-terminal capping domain is coloured dark green. The positions of the LRRs are indicated.

2.6.2 Overall structure

The extracellular domain of TDR forms a twisted right-handed superhelix composed of 22 LRRs and N-terminal (residues 34–81) and C-terminal (residues 609–637) capping domains (Fig. 2-11 A, B). The N-terminal capping domain is assembled from α helices, a β strand and a disordered loop (residues 61–64), and is stabilized by a disulfide bond (Cys69–Cys76; Fig. 2-11 A, B). The C-terminal capping domain is assembled from a 3_{10} helix and a β strand, and is stabilized by a disulfide bond (Cys620–Cys629; Fig. 2-11 A). Plant LRR proteins typically contain repeat sequences of 20–29 amino acids with an LxxLxLxxNxGxIP consensus motif, where X represents any amino acid, and Leu is sometimes substituted by other hydrophobic residues, such as Phe or Val[116]. The known plant LRR-RK structures share the inner concave surface, consisting of parallel β strands formed by the LxxLxL motif, while the outer convex surface consists of various secondary structures, including α helices, 3_{10} helices and short additional β strands[106–110,117]. The present crystal structure of TDR revealed that the repetition of 23–24 amino acids forms the 22 LRR domains (LRR1–LRR22), which compose the inner concave surface with the parallel β strands formed by the LxxLxL motif, and the outer convex surface with helices or loops formed by variable sequences following the GxIP motif (Fig. 2-12). Two loops on the convex surface are stabilized through disulfide bonds formed

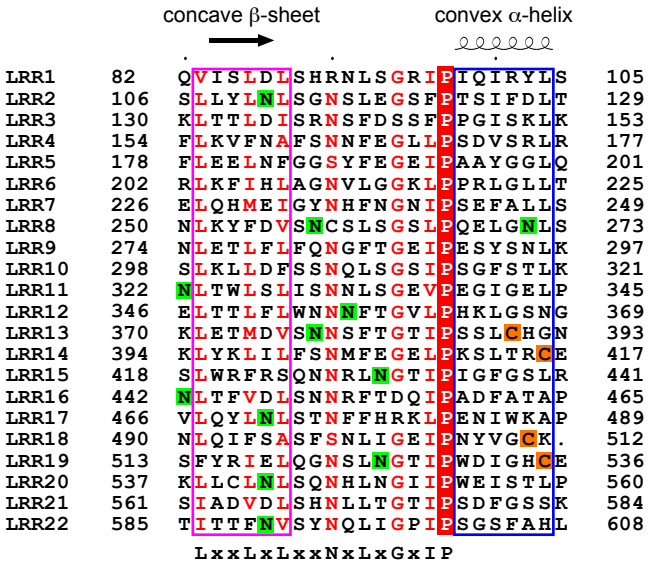

Fig. 2-12 Sequence alignment of LRRs of TDR. The Cys residues that form disulfide bonds are indicated with an orange background, and the glycosylated Asn residues are indicated with a green background. The residues that form the concave β sheets and the convex α helices are enclosed in magenta and blue boxes, respectively.

between adjacent repeats (LRR13–LRR14 and LRR18–LRR19). In the crystal structure, the electron densities were observed for the 11 glycan chains attached to the Asn side chains in the LRR domains (Fig. 2-11 A, B and Fig. 2-12), although their biological functions are unknown.

2.6.3 TDIF kink recognition by the TDR LRR domains

The electron density of TDIF is clearly defined on the concave surface of LRR4–LRR15 of TDR (Fig. 2-11, Fig. 2-13). TDIF adopts an extended conformation, with a kink around Gly6–Hyp7 at the middle of the peptide. The backbone of the TDIF peptide forms van der Waals interactions with the bulky side chains of TDR, such as Phe and Trp (Fig. 2-14 A–D). The interactions between TDIF and TDR can be divided into three parts: the N-terminal part (His1–Hyp4), the middle part (Ser5–Hyp7) and the C-terminal part (Asn8–Asn12; Fig. 2-14 A–D).

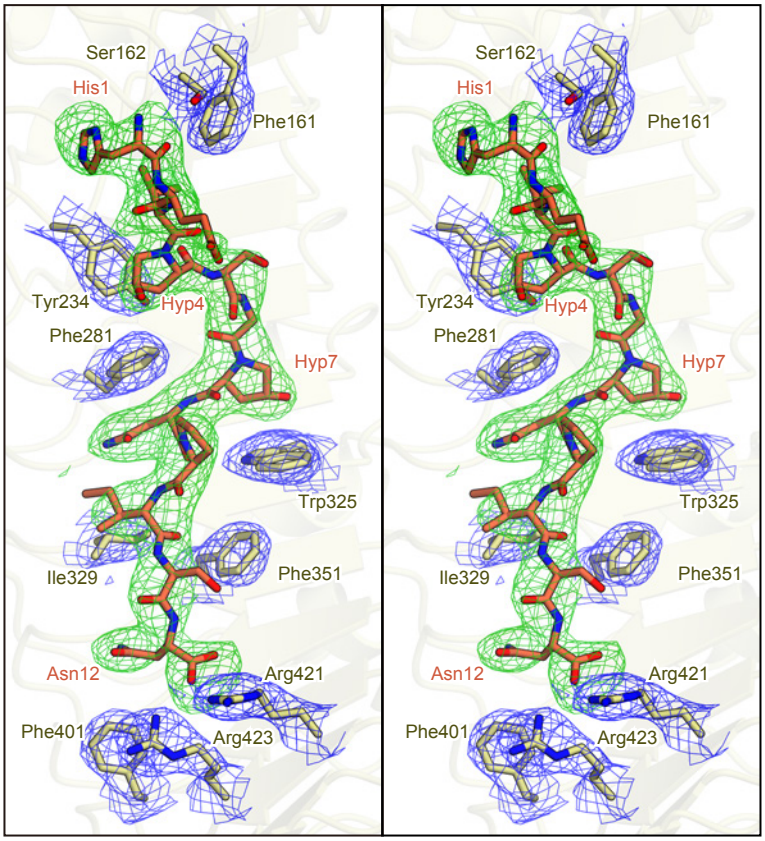

Fig. 2-13 Electron density maps of the TDIF-binding site. Stereo views of the electron density maps of the TDIF-binding site of TDR. The $F_O - F_C$ omit map calculated without TDIF (contoured at 3.0 σ) is shown as a green mesh, while the $2F_O - F_C$ electron density map (contoured at 1.5 σ) is shown as a blue mesh. TDIF and TDIF-interacting TDR residues are shown as orange and green sticks, respectively.

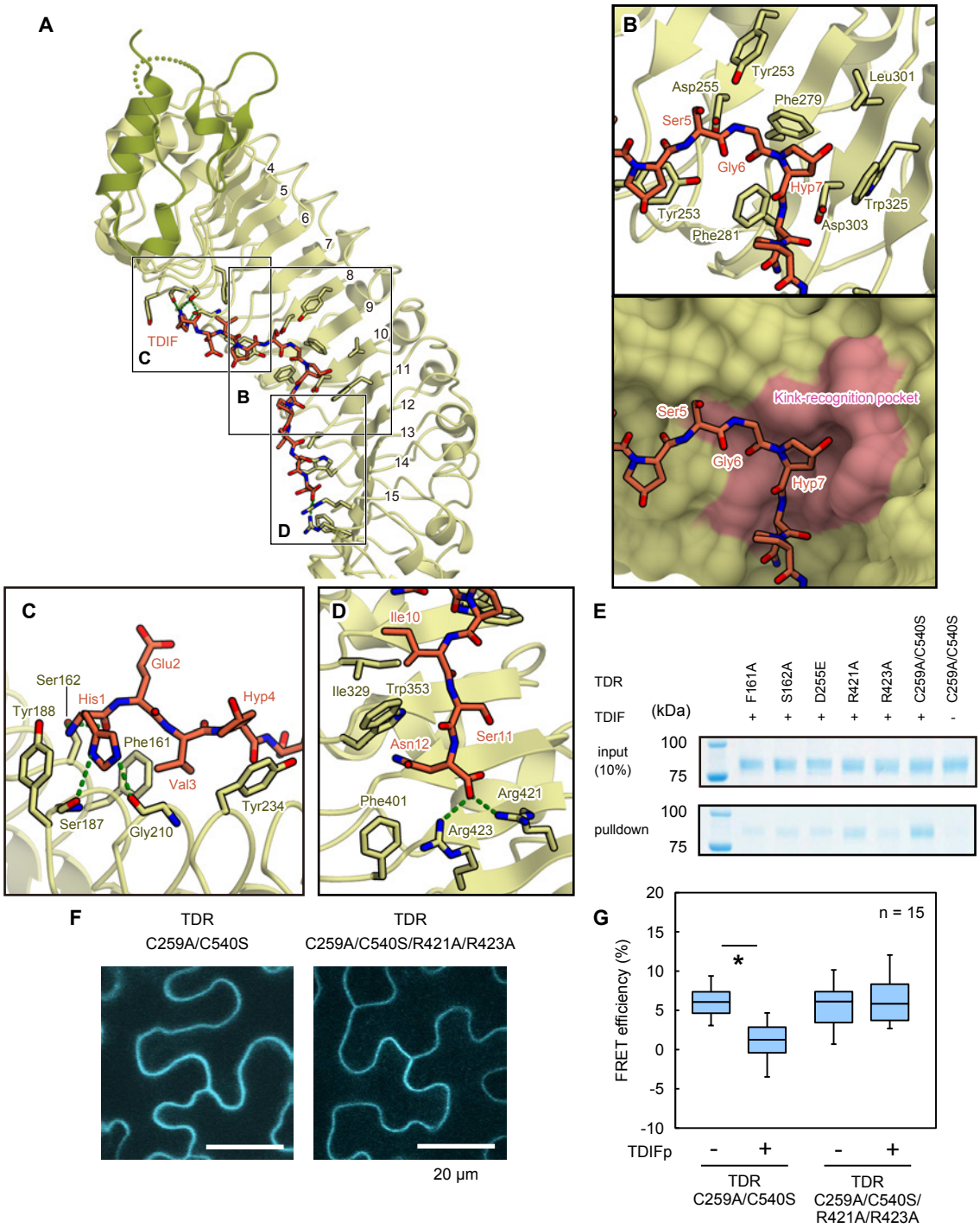

Fig. 2-14 TDIF recognition by TDR. A. The TDIF-interacting residues of TDR are shown in stick representations. TDIF, orange; TDR green. Hydrogen bonds, dashed green lines. **B**–**D**. Recognition of the middle part (residues 5–7) (**B**), the N-terminal part (residues 1–4) (**C**), and the C-terminal part (residues 10–12) (**D**) of TDIF by TDR. In **B**, TDR is shown in a ribbon (upper) and a surface representation (lower). The kink-recognition pocket, pink. **E.** Pull-down experiment using the biotinylated TDIF and TDR mutants. The C259A/C540S TDR mutant was used for crystallization, and other mutations were introduced to this construct. Proteins were mixed with Streptavidin beads in the presence or absence of biotinylated TDIF. The bound proteins were eluted with SDS sample buffer and analysed by SDS–PAGE. See also Fig. 2.5.A. **F**. Localization of TDR–CFP variants in *Nicotiana benthamiana* epidermis. Scale bar indicates 20 µm. **G.** TDIF responses in TDR mutated variants. TDIF responses were evaluated by measuring FRET efficiencies between TDR–CFP and BIN2–YFP. The boxplot diagram displays FRET efficiencies in incubations with or without 5 µM TDIF for 30 min. Significant differences according to the Student's t-test are indicated by asterisks ($P < 0.01$; $n = 15$).

In the middle part, the kink around Gly6–Hyp7 is recognized by the pocket formed by Asp255 (LRR8), Phe279 (LRR9), Phe281 (LRR9) and Trp325 (LRR11) of TDR, in a shape-complementary manner (Fig. 2-14 B). To verify the importance of the TDIF kink recognition by TDR, pull-down assays were performed using biotinylated TDIF peptides. The results showed that the D255E mutation of TDR, which may change the size of the kink-recognition pocket, reduced its TDIF binding, as compared with those of wild-type TDR and the C259A/C540S mutant (Fig. 2-14 E, Fig. 2-7 A). Furthermore, a previous study showed that the G6A mutation abolished the TDIF activity[64]. These results suggest that the kink formation in TDIF and the kink-recognition pocket of TDR are critical for the TDR–TDIF interaction. A previous study reported that the post-transcriptional hydroxylation of Hyp7 is dispensable for the TDIF activity[64]. Consistently, the hydroxyl group of Hyp7 is not recognized by TDR, although the other moieties of Hyp7 form van der Waals interactions with the side chains of Phe279, Leu301 and Trp325 of TDR. Taken together, the central kink of the TDIF peptide is recognized by the kink-recognition pocket formed by the TDR LRR domains, through shape complementarity.

2.6.4 Sequence-specific recognition of TDIF by TDR

In the N-terminal part, the imidazole nitrogen of His1 of TDIF hydrogen bonds with the main-chain carbonyl groups of Ser187 (LRR5) and Gly210 (LRR6), and the N-terminal amino group and the carbonyl group of His1 of TDIF form further hydrogen bonds with the hydroxyl group of Ser162 (LRR4) of TDR. The Val3 side chain of TDIF interacts with the hydrophobic surface around Phe161 (LRR4) of TDR (Fig. 2-14 C). The results of the pull-down assay demonstrated that the F161A and S162A mutations of TDR decreased the TDIF binding, suggesting the importance of these interactions with the N-terminal part (Fig. 2-14 E, Fig. 2-7 A). Moreover, a previous study reported that the H1A and V3A mutations in TDIF diminished the TDIF activity *in planta*[64], suggesting the biological importance of the interactions observed in the crystal structure. In contrast, the side chains of Glu2 and Hyp4 are not recognized by TDR; the Glu2 side chain is disordered, while Hyp4 is solvent-exposed in the present crystal

structure. These observations are consistent with the previous study showing that both the E2A mutation and the Hyp4 to Pro substitution in TDIF do not affect the inhibitory activity against the tracheary element differentiation[64].

At the C-terminal part, the Asn12 carbonyl group of TDIF hydrogen bonds with the side chains of Arg421 and Arg423 of TDR (Fig. 2-14 D). The pull-down assay revealed that the R423A mutation reduced the TDIF-binding activity, while the R421A mutation slightly reduced it (Fig. 2-14 E). Consistent with the results from the *in vitro* assay, the R421A/R423A mutation abolished the responsiveness to TDIF in the reconstructed TDIF–TDR–BIN2 pathway in *Nicotiana benthamiana* leaves, confirming the importance of these TDR–TDIF interactions in the cellular context (Fig. 2-14 F, G). Moreover, a previous study showed that the N12A mutation of TDIF diminishes the TDIF activity[64]. These results suggest that the recognition of the C-terminal Asn residue of TDIF by Arg423 of TDR is critical for the TDR–TDIF signalling pathway *in planta*. The Ser162, Arg421 and Arg423 residues are conserved among the TDRs from other plants, implying that the TDIF-recognition mechanisms are also conserved among plant species (Fig. 2-15). Taken together, the N and C termini of TDIF are both recognized by TDR through sequence-specific hydrogen bonding interactions.

2.7 Discussion

2.7.1 Insight into the dimer formation of TDR

In this chapter the crystal structure of the extracellular domain of TDR in complex with its ligand, TDIF, was determined. In conjunction with the functional analyses, it was revealed that the recognition mechanism of TDIF by TDR. Zhang *et al.*[118] reported the crystal structure of TDR–TDIF, which is essentially identical to the structure determined here (root-mean-square deviation of 1.23 Å for 585 aligned Cα atoms; Fig. 2-16), confirming the functional relevance of the TDIF-recognition mechanism of TDR. In addition, their genetic complementation assays in *Arabidopsis* affirmed that the C-terminal recognition of TDIF by the positive charges of Arg421 and Arg423 in TDR is required for the *in vivo* function of TDR, in agreement with the results of our *in vitro* and *in vivo* assays. Moreover, Zhang *et al.* showed

that TDR forms a TDIF-dependent heterodimer with the extracellular LRR domain of SERK1, SERK2 and BAK1 (also known as SERK3) in a gel-filtration analysis, thereby indicating that SERK family proteins redundantly act as co-receptors of TDR[119].

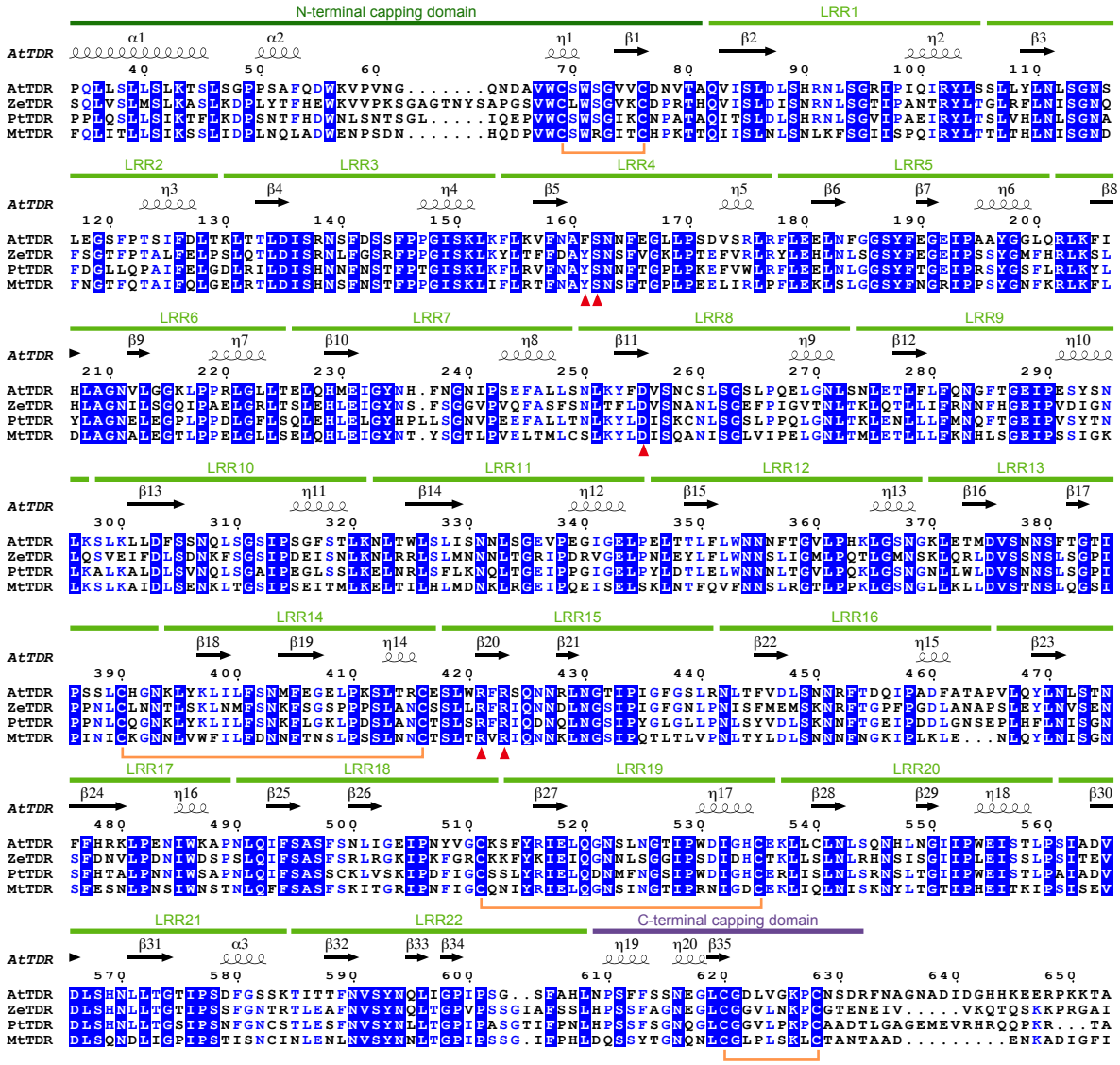

Fig. 2-15 Multiple sequence alignment of the extracellular domains of TDRs from different plant species. "At", "Ze", "Pt" and "Mt" represent *Arabidopsis thaliana*, *Zinnia elegans*, *Populus trichocarpa* and *Medicago truncatula*, respectively. Residues involved in TDIF recognition are indicated by red triangles. Cys residues that form disulfide bonds are indicated by orange lines.

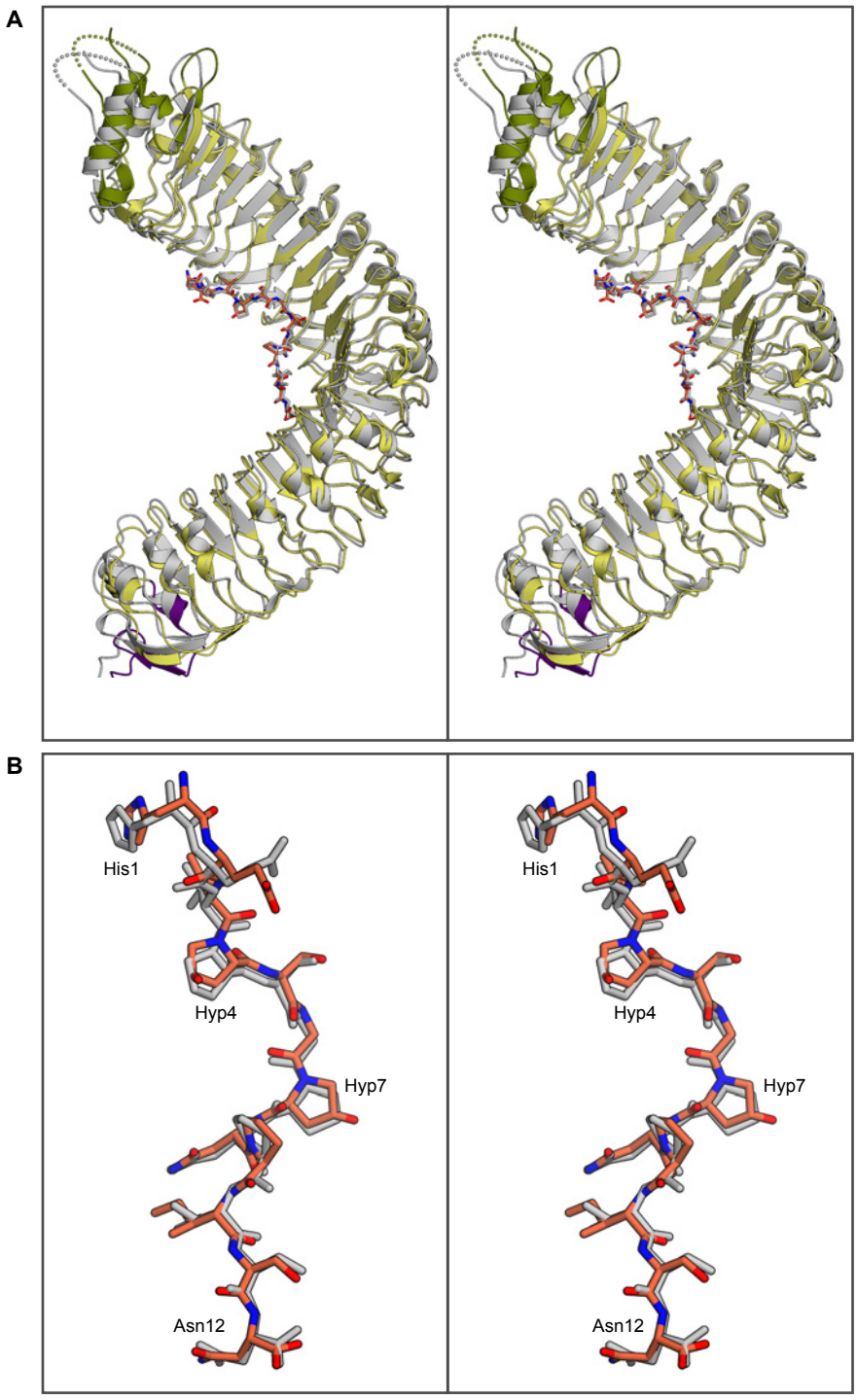

Fig. 2-16. Stereo views of the superimpositions of the current TDR structure (colored) and the reported TDR structure (gray). A. Superimpositions of the TDR–TDIF complex structures. **B.** Superimpositions of the TDIF peptide structures.

2.7.2 Comparison with known LRR-RKs structures

A comparison of the TDR–TDIF complex structure with those of known plant LRR-RKs and their ligand complexes revealed the conserved structural features among the plant LRR-RK family members. The twisted superhelical structures and the N-terminal capping domain of the LRR domain of TDR are commonly observed in the other plant LRR-RK structures, including BRI1 and FLS2 (Fig. 2-17 A–C)[106,107,109]. In the LRR domains of these structures, the LxxLxL motif forms parallel β sheets, with their Leu residues facing towards the inner side of the molecule, to form the hydrophobic core in the proteins. Furthermore, the N- and C-terminal capping domains contribute to structural stabilization, by preventing the exposure of the hydrophobic core to the solvent[116]. Nonetheless, there are also notable structural differences between TDR and the other LRR-RKs. The LRR-RKs FLS2 and BRI1 form ligand-dependent heterodimers with the co-receptor BAK1, mediated by their ligands, flg22 and brassinolide, respectively. In these structures, the ligands are sandwiched between the receptor and the co-receptor, thereby serving as molecular glue to stabilize the interaction (Fig. 2-17 B, C, Fig 2-18)[109,120]. Subsequently, the kinase domains of LRR-RK (BRI1 or FLS2) and a co-receptor (BAK1) trans-phosphoyrlate each other and consequently enable to activate the downstream signaling events. In FLS2, LRR23–LRR26 are involved in the interaction with BAK1, while in BRI1, both LRR23–LRR25 and insertion domain between LRR21 and LRR22 are involved in the ligand-mediated heterodimerization with BAK1. Notably, these structural features are absent in the LRR domains of TDR; the LRR domains of TDR are shorter than those of FLS2 and BRI1, and lack the insertion domain (Fig. 2-17 E, F). Moreover, the LRR domains of TDR and BRI1 have different curvatures. These structural differences suggest that TDR–TDIF form a heterodimer with BAK1 and also with SERK1 and SERK2, which shares about 70% sequence identity with BAK1 in *Arabidopsis thaliana*, in a manner distinct from those in the other BAK1-mediated receptors (that is, BRI1–brassinolide–BAK1 and FLS2–flg22–BAK1; Fig. 2-17).

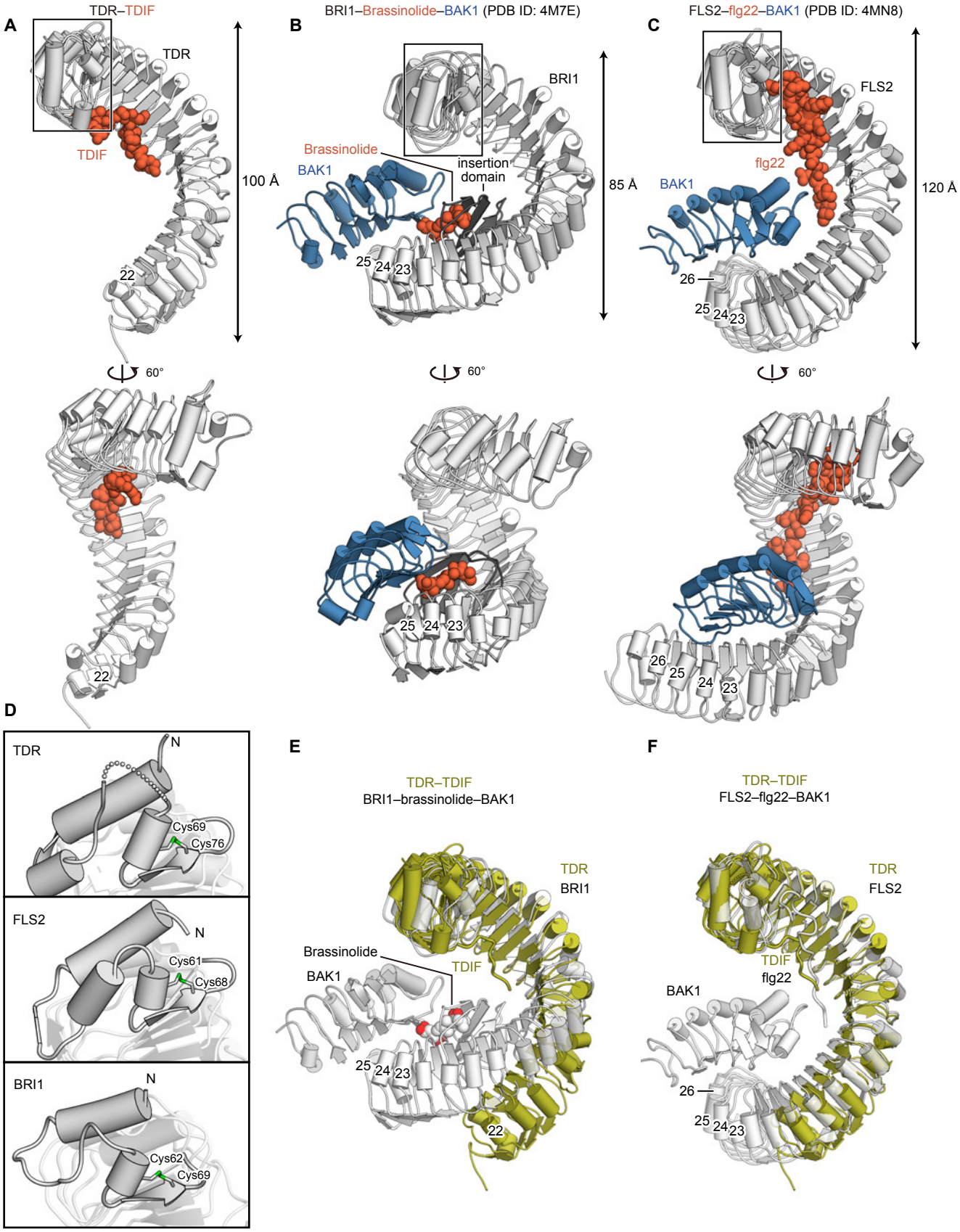

Fig. 2-17 Structural comparison of the plant LRR-RKs. A–C, Crystal structures of TDR–TDIF (**A**), BRI1–brassinolide–BAK1 (**B**) and FLS2–flg22–BAK1 (**C**). LRR-RKs, co-receptors and ligands are colored white, blue and orange, respectively. The insertion domain of BRI1, dark grey. **D.** N-terminal capping domains of TDR, FLS2 and BRI1. The disordered region (residues 61–64) is shown as a dashed line. Cys residues that form disulfide bonds, sticks. **E**, **F**, Superimposition of TDR onto the BRI1–brassinolide–BAK1 (**E**) and the FLS2–flg22–BAK1 (**F**). TDR and the LRR-RK–ligand–BAK1 complexes are shown in olive and white, respectively.

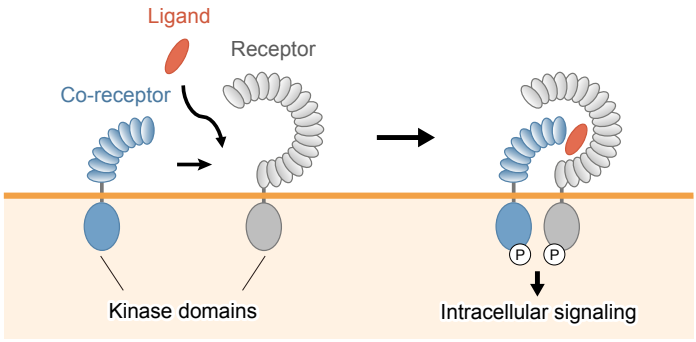

Fig. 2-18 Schematic model of the activation of plant LRR-RKs. LRR-RKs, co-receptors and ligands are colored gray, blue and red, respectively.

2.7.3 Insight into the recognition mechanism of CLE family peptides

The present crystal structure provides structural insights into the common recognition mechanism of the CLE family peptide hormones by their receptors. The kink-forming residues of TDIF (Gly6 and Hyp7) are highly conserved among the CLE family peptides (Table 2-1, Fig. 2-19 A). Previous studies reported that the G6A mutations in other CLE peptides also abrogate their activities, as observed in the case of TDIF[64,121]. Moreover, the residues in the kink-recognition pocket of TDR (Asp255 and Phe279) are well conserved among the CLE family peptide receptors, such as CLV1, BAM1/2/3 and SKM1 (Fig. 2-19 B). These observations suggest that the accommodation of the kinked peptide by the receptor pocket is the general interaction mechanism for the CLE family ligands and their receptors. In contrast, a previous study reported that several CLE family peptides other than TDIF do not interact with TDR[66]. The present crystal structure explains the mechanism by which TDR discriminates TDIF from other CLE peptides. At the C-terminal part of the molecular interface between TDR and TDIF, the Arg421 side chain of TDR hydrogen bonds with the Asn12 of TDIF. In the CLV3 peptide, this Asn12 residue is replaced with His, while Arg421 of TDR corresponds to Lys413 of CLV1, an LRR-RK receptor for the CLV3 peptide. Thus, it is possible that the interaction at the C-terminal part is important for the discrimination of the orthogonal ligand–receptor pairs. Consistently, Arg421 is also conserved in other LRR-RKs, such as BAM1/2/3 and SKM1, which recognize the CLE peptides that possess Asn at position 12 (Table 2-1, Fig. 2-19 A, B)[82,96,97]. Concordant with these observations, a very recent study showed that synthetic

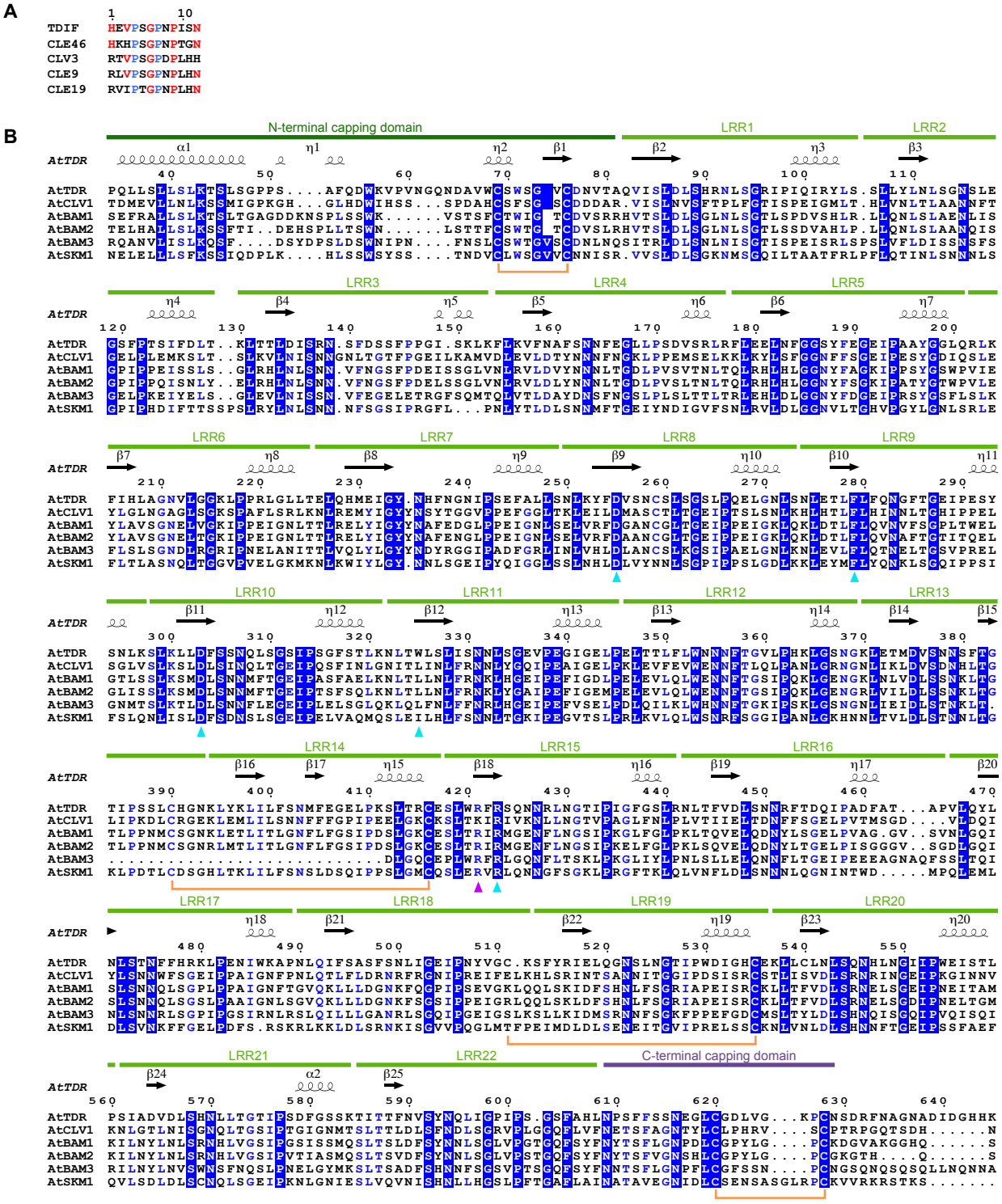

Fig. 2-19 Multiple sequence alignment of CLE family peptides and receptors in *Arabidopsis thaliana*. A. Sequence alignment of CLE family peptides. TDIF residues recognized by TDR are colored red. Conserved Pro/Hyp residues are colored blue. **B**. Sequence alignment of the extracellular domains of LRR-RKs that perceive CLE peptides. Arg421 and residues involved in peptide backbone recognition are indicated by purple and cyan triangles, respectively.

unnatural CLE peptide with amino acid substitution, KIN, can bind and activate both of TDR and CLV1 (Fig. 2-20)[122]. Furthermore, at the N-terminal part of the interface, the His1 side chain of TDIF is recognized by the main-chain carbonyl groups of Ser187 and Gly210 of TDR, and the Val3 side chain of TDIF interacts with the surface around Phe161 of TDR. These interactions can also explain the substrate specificity of TDR. A previous study reported that CLV3, CLE9, CLE19 and CLE46 do not interact with TDR[66]. In CLE46, Val3 of TDIF is replaced with His, showing that the bulky side chain of His cannot interact with the TDR surface around Phe161. In contrast, a previous study reported that CLE42 can act as a ligand of TDR in *Arabidopsis*[92]. Glu2 of TDIF is replaced with Gly in CLE42, consistent with the fact that the Glu2 side chain is not recognized by TDR. Collectively, the present crystal structure suggests that the central kink of the CLE peptides, including TDIF, offers a common platform for the recognition by the CLE peptide receptors, while the N- and C-terminal parts provide the sequence-specific readout for the cognate CLE receptors.

Peptide	Sequence	Receptor
CLV3	RTVPSGPDPLHH	CLV1
KIN	RKVPSGPDPIHN	CLV1, TDR
TDIF	HEVPSGPNPISN	TDR

(Hirakawa *et al.*, 2017)

Fig. 2-20 Sequence and specificity of the synthetic bifunctional CLE peptide, KIN.

2.8 Conclusion

In conclusion, the structural and functional data revealed the recognition mechanism of TDIF by TDR, and provided insights into the recognition mechanism of the CLE family peptides by their cognate receptors. These findings may help pave the way for the rational engineering of the TDR–TDIF axis, to improve biomass production. Actually, the aforementioned example of the designed synthetic CLE peptide KIN[122] agreed well with the sequence–structure relationship, which was proposed in the current study. The KIN peptide is shown to have the bioactivity beyond the original subfamily by interacting with multiple

receptors. These observations suggest the feasibility of the structure-based rational engineering of the TDR–TDIF axis.

The crystal structure of TDR–TDIF–SERK2 was reported by in 2016 by Zhang *et al*[119]. In their study, it is shown that TDR physically interacts SERK2 as well as SERK1 to form a functional complex, expanding the discussion in 2.7.1. Genetic analysis using *Arabidopsis* revealed that SERK1, SERK2 and BAK1 redundantly regulate procambial cell formation as co-receptors of TDR. In the crystal structure of TDR–TDIF–SERK2 complex, the structure of TDR–TDIF (rmsd of 0.81 Å for 589 aligned C α atoms; Fig. 2-21 A, B), suggesting that SERK2 binding does not induce significant structural change in TDR–TDIF. In the structure, TDIF mediates the binding between TDR and SERK2 with its C-terminal sandwiched between them. These observations indicate that binding between TDR and TDIF is prerequisite for the formation of TDR–TDIF–SERK2 ternary complex, similar in flg22-mediated FLS2–BAK1 interaction and brassinolide-mediated BRI1–BAK1 interaction. Taken together, it is proposed that TDIF works as a molecular glue to mediate the connection between the extracellular domains of TDR–co-receptor, and subsequently mediate the connection interaction between intracellular kinase domains to somehow trigger the downstream signaling.

More recently, the crystal structure of PXL2 (PXY-like 2)–CLE42 complex was reported[123]. PXL2 is a close homolog of TDR which shares 62% sequence similarity with TDR, and there is only one difference in amino acid sequence between TDIF and CLE42 (TDIF, HEVPSGPNPISN and CLE42, HGVPSGPNPISN). Concordant with these observations, structural features found in PXL2–CLE42 interaction are apparently common with those in TDR–TDIF, although close inspection can not be performed because the structure data has not been released yet. Taken together, observations in this study are well supported by the latest studies.

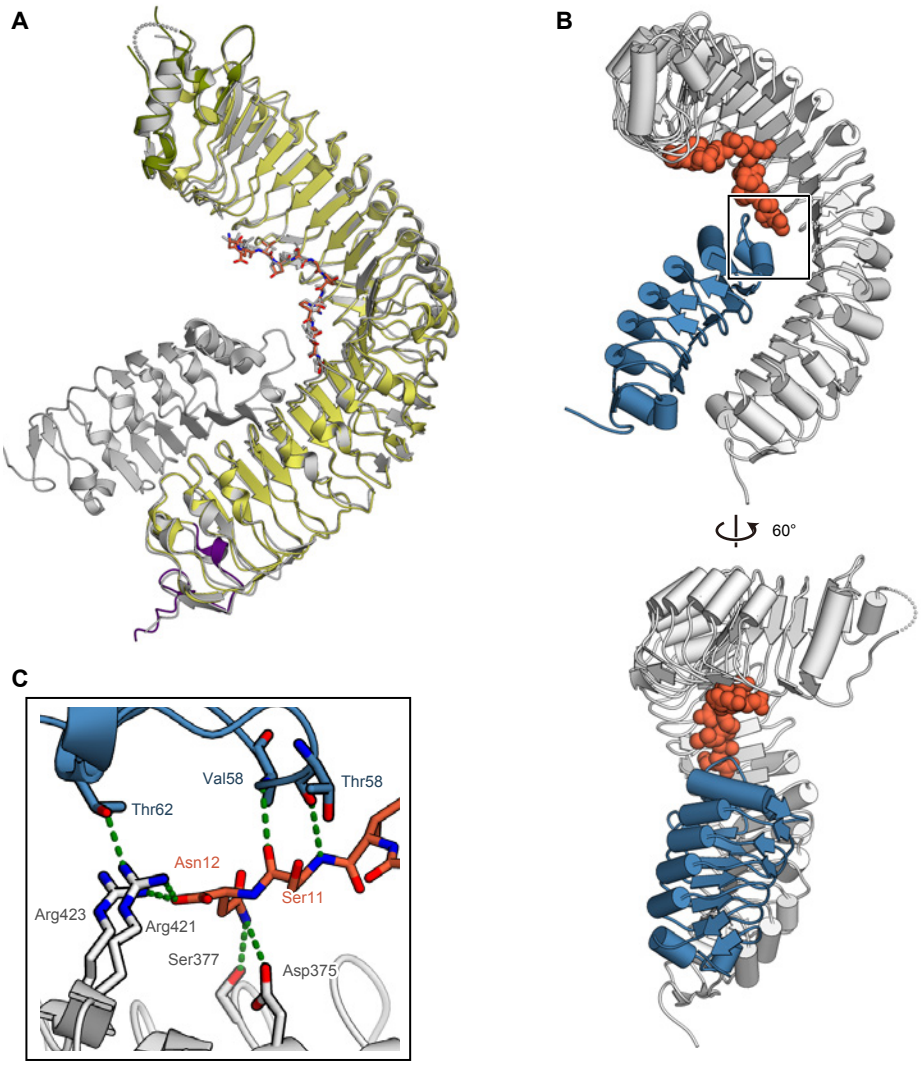

Fig. 2-21 Structure of the TDR–TDIF–SERK2 complex. A. Superimpositions of the current TDR–TDIF complex structure (colored) and the reported TDR–TDIF–SERK2 structure (gray, PDB ID 5GQR). **B.** TDR, SERK2 and TDIF are colored white, blue and orange, respectively. **C.** The TDR–TDIF–SERK2 interactions of the boxed region in **(B).** The TDIF-interacting residues of TDR and SERK2 are shown in stick representations. TDIF, orange; TDR, gray; SERK2, blue. Hydrogen bonds, dashed green lines.

References

1. Pires-daSilva, A. & Sommer, R. J. The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* **4,** 39–49 (2003).
2. Ben-Shlomo, I., Yu Hsu, S., Rauch, R., Kowalski, H. W. & Hsueh, A. J. W. Signaling Receptome: A Genomic and Evolutionary Perspective of Plasma Membrane Receptors Involved in Signal Transduction. *Sci. Signal.* **2003,** re9-re9 (2003).
3. Efeyan, A., Comb, W. C. & Sabatini, D. M. Nutrient-sensing mechanisms and pathways. *Nature* **517,** 302–310 (2015).
4. Von Heijne, G. Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.* **7,** 909–918 (2006).
5. Johnson, J. E. & Cornell, R. B. Amphitropic proteins: Regulation by reversible membrane interactions. *Mol. Membr. Biol.* **16,** 217–235 (1999).
6. Resh, M. D. Covalent lipid modifications of proteins. *Curr. Biol.* **23,** R431–R435 (2013).
7. Walter, P., Gilmore, R. & Blobel, G. Protein translocation across the endoplasmic reticulum. *Cell* **38,** 5–8 (1984).
8. Tams, J. W., Vind, J. & Welinder, K. G. Adapting protein solubility by glycosylation. N-Glycosylation mutants of Coprinus cinereus peroxidase in salt and organic solutions. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1432,** 214–221 (1999).
9. Fass, D. Disulfide Bonding in Protein Biophysics. *Annu. Rev. Biophys.* **41,** 63–79 (2012).
10. Blusztajn, J. K. Choline, a vital amine. *Science* **281,** 794–5 (1998).
11. Zeisel, S. H. & Blusztajn, J. K. Choline and Human Nutrition. *Annu. Rev. Nutr.* **14,** 269–296 (1994).
12. Sheard, N. F. & Zeisel, S. H. Choline: an essential dietary nutrient? *Nutrition* **5,** 1–5
13. Vance, D. E. Physiological roles of phosphatidylethanolamine N-methyltransferase. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1831,** 626–632 (2013).
14. Michel, V., Yuan, Z., Ramsubir, S. & Bakovic, M. Choline transport for phospholipid synthesis. *Exp. Biol. Med. (Maywood).* **231,** 490–504 (2006).
15. Sugimoto, H., Banchio, C. & Vance, D. E. Transcriptional regulation of phosphatidylcholine biosynthesis. *Prog. Lipid Res.* **47,** 204–220 (2008).
16. Vance, J. E. & Vance, D. E. Metabolic insights into phospholipid function using gene-targeted mice. *J. Biol. Chem.* **280,** 10877–80 (2005).
17. Jacobs, R. L., Stead, L. M., Devlin, C., Tabas, I., Brosnan, M. E., Brosnan, J. T. & Vance, D. E. Physiological Regulation of Phospholipid Methylation Alters Plasma Homocysteine in Mice. *J. Biol. Chem.* **280,** 28299–28305 (2005).
18. Klein, J., Gonzalez, R., Köppen, A. & Löffelholz, K. Free choline and choline metabolites in rat brain and body fluids: sensitive determination and implications for choline supply to the brain. *Neurochem. Int.* **22,** 293–300 (1993).
19. Suzuki, A. & Kawakami, M. A hemolytic lipoprotein containing lysophosphatidylcholine produced in incubated mouse plasma. *Biochim. Biophys. Acta* **753,** 236–43 (1983).

20. Stefan, C., Jansen, S. & Bollen, M. NPP-type ectophosphodiesterases: unity in diversity. *Trends Biochem. Sci.* **30,** 542–50 (2005).
21. Bollen, M., Gijsbers, R., Ceulemans, H., Stalmans, W. & Stefan, C. Nucleotide Pyrophosphatases/Phosphodiesterases on the Move. *Crit. Rev. Biochem. Mol. Biol.* **35,** 393–432 (2000).
22. Umezu-Goto, M., Kishi, Y., Taira, A., Hama, K., Dohmae, N., Takio, K., Yamori, T., Mills, G. B., Inoue, K., Aoki, J. & Arai, H. Autotaxin has lysophospholipase D activity leading to tumor cell growth and motility by lysophosphatidic acid production. *J. Cell Biol.* **158,** 227–33 (2002).
23. Tokumura, A., Majima, E., Kariya, Y., Tominaga, K., Kogure, K., Yasuda, K. & Fukuzawa, K. Identification of Human Plasma Lysophospholipase D, a Lysophosphatidic Acid-producing Enzyme, as Autotaxin, a Multifunctional Phosphodiesterase. *J. Biol. Chem.* **277,** 39436–39442 (2002).
24. Kawaguchi, M., Okabe, T., Okudaira, S., Hanaoka, K., Fujikawa, Y., Terai, T., Komatsu, T., Kojima, H., Aoki, J. & Nagano, T. Fluorescence probe for lysophospholipase C/NPP6 activity and a potent NPP6 inhibitor. *J. Am. Chem. Soc.* **133,** 12021–30 (2011).
25. Korekane, H., Park, J. Y., Matsumoto, A., Nakajima, K., Takamatsu, S., Ohtsubo, K., Miyamoto, Y., Hanashima, S., Kanekiyo, K., Kitazume, S., Yamaguchi, Y., Matsuo, I. & Taniguchi, N. Identification of ectonucleotide pyrophosphatase/phosphodiesterase 3 (ENPP3) as a regulator of N-acetylglucosaminyltransferase GnT-IX (GnT-Vb). *J. Biol. Chem.* **288,** 27912–26 (2013).
26. Albright, R. A., Chang, W. C., Robert, D., Ornstein, D. L., Cao, W., Liu, L., Redick, M. E., Young, J. I., De La Cruz, E. M. & Braddock, D. T. NPP4 is a procoagulant enzyme on the surface of vascular endothelium. *Blood* **120,** 4432–40 (2012).
27. Tsai, S. H., Kinoshita, M., Kusu, T., Kayama, H., Okumura, R., Ikeda, K., Shimada, Y., Takeda, A., Yoshikawa, S., Obata-Ninomiya, K., Kurashima, Y., Sato, S., Umemoto, E., Kiyono, H., Karasuyama, H. & Takeda, K. The Ectoenzyme E-NPP3 Negatively Regulates ATP-Dependent Chronic Allergic Responses by Basophils and Mast Cells. *Immunity* **42,** 279–293 (2015).
28. Duan, R.-D., Bergman, T., Xu, N., Wu, J., Cheng, Y., Duan, J., Nelander, S., Palmberg, C. & Nilsson, Å. Identification of Human Intestinal Alkaline Sphingomyelinase as a Novel Ecto-enzyme Related to the Nucleotide Phosphodiesterase Family. *J. Biol. Chem.* **278,** 38528–38536 (2003).
29. Sakagami, H., Aoki, J., Natori, Y., Nishikawa, K., Kakehi, Y., Natori, Y. & Arai, H. Biochemical and molecular characterization of a novel choline-specific glycerophosphodiester phosphodiesterase belonging to the nucleotide pyrophosphatase/phosphodiesterase family. *J. Biol. Chem.* **280,** 23084–23093 (2005).
30. Mulder, A. M. & Cravatt, B. F. Endocannabinoid Metabolism in the Absence of Fatty

Acid Amide Hydrolase (FAAH): Discovery of Phosphorylcholine Derivatives of *N* -Acyl Ethanolamines †. *Biochemistry* **45,** 11267–11277 (2006).

31. Hausmann, J., Kamtekar, S., Christodoulou, E., Day, J. E., Wu, T., Fulkerson, Z., Albers, H. M. H. G., van Meeteren, L. a, Houben, A. J. S., van Zeijl, L., Jansen, S., Andries, M., Hall, T., Pegg, L. E., Benson, T. E., Kasiem, M., Harlos, K., Kooi, C. W. Vander, Smyth, S. S., Ovaa, H., Bollen, M., Morris, A. J., Moolenaar, W. H. & Perrakis, A. Structural basis of substrate discrimination and integrin binding by autotaxin. *Nat. Struct. Mol. Biol.* **18,** 198–204 (2011).
32. Nishimasu, H., Okudaira, S., Hama, K., Mihara, E., Dohmae, N., Inoue, A., Ishitani, R., Takagi, J., Aoki, J. & Nureki, O. Crystal structure of autotaxin and insight into GPCR activation by lipid mediators. *Nat. Struct. Mol. Biol.* **18,** 205–12 (2011).
33. Kato, K., Nishimasu, H., Okudaira, S., Mihara, E., Ishitani, R., Takagi, J., Aoki, J. & Nureki, O. Crystal structure of Enpp1, an extracellular glycoprotein involved in bone mineralization and insulin signaling. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 16876–81 (2012).
34. Jansen, S., Perrakis, A., Ulens, C., Winkler, C., Andries, M., Joosten, R. P., Acker, M. Van, Luyten, F. P., Moolenaar, W. H. & Bollen, M. Article Structure of NPP1 , an Ectonucleotide Pyrophosphatase / Phosphodiesterase Involved in Tissue Calcification. *Structure* **20,** 1–12 (2012).
35. Albright, R. A., Ornstein, D. L., Cao, W., Chang, W. C., Robert, D., Tehan, M., Hoyer, D., Liu, L., Stabach, P., Yang, G., De La Cruz, E. M. & Braddock, D. T. Molecular basis of Purinergic Signal Metabolism by Ectonucleotide Pyrophosphatase/Phosphodiesterases 4 and 1 and implications in Stroke. *J. Biol. Chem.* **289,** 3294–3306 (2014).
36. Greiner-Tollersrud, L., Berg, T., Stensland, H. M. F. R., Evjen, G. & Greiner-Tollersrud, O. K. Bovine Brain Myelin Glycerophosphocholine Choline Phosphodiesterase is an Alkaline Lysosphingomyelinase of the eNPP-Family, Regulated by Lysosomal Sorting. *Neurochem. Res.* **38,** 300–310 (2013).
37. Jana, A. & Pahan, K. Sphingolipids in multiple sclerosis. *Neuromolecular Med.* **12,** 351–61 (2010).
38. Tabata, S., Nampo, M., Mihara, E. & Tamura-kawakami, K. Technical note A rapid screening method for cell lines producing singly-tagged recombinant proteins using the ' TARGET tag ' system. **3,** 0–8 (2010).
39. Kato, K., Nishimasu, H., Mihara, E., Ishitani, R., Takagi, J., Aoki, J. & Nureki, O. Expression, purification, crystallization and preliminary X-ray crystallographic analysis of Enpp1. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **68,** 778–82 (2012).
40. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 125–32 (2010).
41. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 22–5 (2010).
42. Kelley, L. a & Sternberg, M. J. E. Protein structure prediction on the Web: a case study

using the Phyre server. *Nat. Protoc.* **4,** 363–71 (2009).

43. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).
44. Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–21 (2010).
45. Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. Structure validation by Cα geometry: φ,ψ and Cβ deviation. *Proteins Struct. Funct. Bioinforma.* **50,** 437–450 (2003).
46. Imamura, S. & Horiuti, Y. Enzymatic determination of phospholipase D activity with choline oxidase. *J. Biochem.* **83,** 677–80 (1978).
47. Tamaoku, K., Ueno, K., Akiura, K. & Ohkura, Y. New water-soluble hydrogen donors for the enzymatic photometric determination of hydrogen peroxide. II. N-ethyl-N-(2-hydroxy-3-sulfopropyl)aniline derivatives. *Chem. Pharm. Bull. (Tokyo).* **30,** 2492–2497 (1982).
48. Kitamura, T., Koshino, Y., Shibata, F., Oki, T., Nakajima, H., Nosaka, T. & Kumagai, H. Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Exp. Hematol.* **31,** 1007–14 (2003).
49. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7,** 1063–1066 (2000).
50. Goldstein, J. L., Basu, S. K. & Brown, M. S. Receptor-mediated endocytosis of low-density lipoprotein in cultured cells. *Methods Enzymol.* **98,** 241–60 (1983).
51. de Vries, R. P., de Vries, E., Bosch, B. J., de Groot, R. J., Rottier, P. J. M. & de Haan, C. a M. The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology* **403,** 17–25 (2010).
52. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491–7 (1968).
53. Morita, J., Kato, K., Mihara, E., Ishitani, R., Takagi, J., Nishimasu, H., Aoki, J. & Nureki, O. Expression, purification, crystallization and preliminary X-ray crystallographic analysis of Enpp6. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **70,** 1–6 (2014).
54. Zacharias, N. & Dougherty, D. a. cation-π interactions in ligand recognition and catalysis. *Trends Pharmacol. Sci.* **23,** 281–7 (2002).
55. Hessle, L., Johnson, K. A., Anderson, H. C., Narisawa, S., Sali, A., Goding, J. W., Terkeltaub, R. & Millan, J. L. Tissue-nonspecific alkaline phosphatase and plasma cell membrane glycoprotein-1 are central antagonistic regulators of bone mineralization. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9445–9 (2002).
56. Murphy, E., Smith, S. & De Smet, I. Small Signaling Peptides in Arabidopsis

Development: How Cells Communicate Over a Short Distance. *Plant Cell* **24,** 3198–3217 (2012).

57. Oelkers, K., Goffard, N., Weiller, G. F., Gresshoff, P. M., Mathesius, U. & Frickey, T. Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biol.* **8,** 1 (2008).
58. Miwa, H., Tamaki, T., Fukuda, H. & Sawa, S. Evolution of CLE signaling. *Plant Signal. Behav.* **4,** 477–481 (2009).
59. Strabala, T. J., Phillips, L., West, M. & Stanbra, L. Bioinformatic and phylogenetic analysis of the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and the CLE-LIKE signal peptide genes in the Pinophyta. *BMC Plant Biol.* **14,** 47 (2014).
60. Kucukoglu, M. & Nilsson, O. CLE peptide signaling in plants - the power of moving around. *Physiol. Plant.* **155,** 74–87 (2015).
61. Endo, S., Betsuyaku, S. & Fukuda, H. Endogenous peptide ligand-receptor systems for diverse signaling networks in plants. *Curr. Opin. Plant Biol.* **21,** 140–146 (2014).
62. Ohyama, K., Shinohara, H., Ogawa-Ohnishi, M. & Matsubayashi, Y. A glycopeptide regulating stem cell fate in Arabidopsis thaliana. *Nat. Chem. Biol.* **5,** 578–580 (2009).
63. Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W. & Liu, C.-M. The 14-Amino Acid CLV3, CLE19, and CLE40 Peptides Trigger Consumption of the Root Meristem in Arabidopsis through a CLAVATA2-Dependent Pathway. *PLANT CELL ONLINE* **17,** 2542–2553 (2005).
64. Ito, Y., Ikuko, N., Motose, H., Iwamoto, K., Sawa, S., Dohmae, N. & Fukuda, H. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science (80-.).* **313,** 842–845 (2006).
65. Kondo, T., Sawa, S., Kinoshita, A., Mizuno, S., Kakimoto, T., Fukuda, H. & Sakagami, Y. A Plant Peptide Encoded by CLV3 Identified by in Situ MALDI-TOF MS Analysis. *Science (80-.).* **313,** 845–848 (2006).
66. Hirakawa, Y., Shinohara, H., Kondo, Y., Inoue, A., Nakanomyo, I., Ogawa, M., Sawa, S., Ohashi-Ito, K., Matsubayashi, Y. & Fukuda, H. Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 15208–13 (2008).
67. Clark, S. E., Running, M. P. & Meyerowitz, E. M. CLAVATA1, a regulator of meristem and flower development in Arabidopsis. *Development* **119,** 397–418 (1993).
68. Clark, S. E., Running, M. P. & Meyerowitz, E. M. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* **121,** 2057–2067 (1995).
69. Clark, S. E., Williams, R. W. & Meyerowitz, E. M. The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. *Cell* **89,** 575–85 (1997).
70. Kayes, J. M. & Clark, S. E. CLAVATA2, a regulator of meristem and organ development

in Arabidopsis. *Development* **125,** 3843–51 (1998).

71. Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M. & Simon, R. Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. *Science* **289,** 617–9 (2000).
72. Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F., Jürgens, G. & Laux, T. The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* **100,** 635–44 (2000).
73. Müller, R., Bleckmann, A. & Simon, R. The receptor kinase CORYNE of Arabidopsis transmits the stem cell-limiting signal CLAVATA3 independently of CLAVATA1. *Plant Cell* **20,** 934–46 (2008).
74. Ogawa, M., Shinohara, H., Sakagami, Y. & Matsubayashi, Y. Arabidopsis CLV3 Peptide Directly Binds CLV1 Ectodomain. *Science (80-.).* **319,** 294–294 (2008).
75. Guo, Y., Han, L., Hymes, M., Denver, R. & Clark, S. E. CLAVATA2 forms a distinct CLE-binding receptor complex regulating Arabidopsis stem cell specification. *Plant J.* **63,** 889–900 (2010).
76. Kinoshita, A., Betsuyaku, S., Osakabe, Y., Mizuno, S., Nagawa, S., Stahl, Y., Simon, R., Yamaguchi-Shinozaki, K., Fukuda, H. & Sawa, S. RPK2 is an essential receptor-like kinase that transmits the CLV3 signal in Arabidopsis. *Development* **137,** 3911–3920 (2010).
77. Yadav, R. K., Perales, M., Gruel, J., Girke, T., Jonsson, H. & Reddy, G. V. WUSCHEL protein movement mediates stem cell homeostasis in the Arabidopsis shoot apex. *Genes Dev.* **25,** 2025–2030 (2011).
78. Betsuyaku, S., Takahashi, F., Kinoshita, A., Miwa, H., Shinozaki, K., Fukuda, H. & Sawa, S. Mitogen-Activated Protein Kinase Regulated by the CLAVATA Receptors Contributes to Shoot Apical Meristem Homeostasis. *Plant Cell Physiol.* **52,** 14–29 (2011).
79. Shinohara, H. & Matsubayashi, Y. Chemical Synthesis of Arabidopsis CLV3 Glycopeptide Reveals the Impact of Hydroxyproline Arabinosylation on Peptide Conformation and Activity. *Plant Cell Physiol.* **54,** 369–374 (2013).
80. Shinohara, H. & Matsubayashi, Y. Reevaluation of the CLV3-receptor interaction in the shoot apical meristem: dissection of the CLV3 signaling pathway from a direct ligand-binding point of view. *Plant J.* **82,** 328–336 (2015).
81. Ishida, T., Tabata, R., Yamada, M., Aida, M., Mitsumasu, K., Fujiwara, M., Yamaguchi, K., Shigenobu, S., Higuchi, M., Tsuji, H., Shimamoto, K., Hasebe, M., Fukuda, H. & Sawa, S. Heterotrimeric G proteins control stem cell proliferation through CLAVATA signaling in Arabidopsis. *EMBO Rep.* **15,** 1202–9 (2014).
82. Shinohara, H., Moriyama, Y., Ohyama, K. & Matsubayashi, Y. Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1 reveals diversified ligand recognition mechanisms of plant LRR-RKs. *Plant J.* **70,** 845–854 (2012).

83. Kondo, Y., Hirakawa, Y., Kieber, J. J. & Fukuda, H. CLE Peptides can Negatively Regulate Protoxylem Vessel Formation via Cytokinin Signaling. *Plant Cell Physiol.* **52,** 37–48 (2011).
84. Casamitjana-Martínez, E., Hofhuis, H. F., Xu, J., Liu, C.-M., Heidstra, R. & Scheres, B. Root-specific CLE19 overexpression and the sol1/2 suppressors implicate a CLV-like pathway in the control of Arabidopsis root meristem maintenance. *Curr. Biol.* **13,** 1435–41 (2003).
85. Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W. & Liu, C.-M. The 14-amino acid CLV3, CLE19, and CLE40 peptides trigger consumption of the root meristem in Arabidopsis through a CLAVATA2-dependent pathway. *Plant Cell* **17,** 2542–53 (2005).
86. Tamaki, T., Betsuyaku, S., Fujiwara, M., Fukao, Y., Fukuda, H. & Sawa, S. SUPPRESSOR OF LLP1 1-mediated C-terminal processing is critical for CLE19 peptide activity. *Plant J.* **76,** 970–981 (2013).
87. Xu, T.-T., Ren, S.-C., Song, X.-F. & Liu, C.-M. CLE19 expressed in the embryo regulates both cotyledon establishment and endosperm development in Arabidopsis. *J. Exp. Bot.* **66,** 5217–27 (2015).
88. Hobe, M., Müller, R., Grünewald, M., Brand, U. & Simon, R. Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in Arabidopsis. *Dev. Genes Evol.* **213,** 371–81 (2003).
89. Stahl, Y., Wink, R. H., Ingram, G. C. & Simon, R. A Signaling Module Controlling the Stem Cell Niche in Arabidopsis Root Meristems. *Curr. Biol.* **19,** 909–914 (2009).
90. Stahl, Y., Grabowski, S., Bleckmann, A., Kühnemuth, R., Weidtkamp-Peters, S., Pinto, K. G., Kirschner, G. K., Schmid, J. B., Wink, R. H., Hülsewede, A., Felekyan, S., Seidel, C. A. M. & Simon, R. Moderation of Arabidopsis Root Stemness by CLAVATA1 and ARABIDOPSIS CRINKLY4 Receptor Kinase Complexes. *Curr. Biol.* **23,** 362–371 (2013).
91. Hirakawa, Y., Kondo, Y. & Fukuda, H. Regulation of vascular development by CLE peptide-receptor systems. *J. Integr. Plant Biol.* **52,** 8–16 (2010).
92. Etchells, J. P. & Turner, S. R. The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. *Development* **137,** 767–774 (2010).
93. Whitford, R., Fernandez, A., De Groodt, R., Ortega, E. & Hilson, P. Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. *Proc. Natl. Acad. Sci.* **105,** 18625–18630 (2008).
94. Kondo, Y., Ito, T., Nakagami, H., Hirakawa, Y., Saito, M., Tamaki, T., Shirasu, K. & Fukuda, H. Plant GSK3 proteins regulate xylem cell differentiation downstream of TDIF-TDR signalling. *Nat. Commun.* **5,** 3504 (2014).
95. Yaginuma, H., Hirakawa, Y., Kondo, Y., Ohashi-Ito, K. & Fukuda, H. A novel function of

TDIF-related peptides: Promotion of axillary bud formation. *Plant Cell Physiol.* **52,** 1354–1364 (2011).

96. Endo, S., Shinohara, H., Matsubayashi, Y. & Fukuda, H. A Novel Pollen-Pistil Interaction Conferring High-Temperature Tolerance during Reproduction via CLE45 Signaling. *Curr. Biol.* **23,** 1670–1676 (2013).
97. Depuydt, S., Rodriguez-Villalon, A., Santuari, L., Wyser-Rmili, C., Ragni, L. & Hardtke, C. S. Suppression of Arabidopsis protophloem differentiation and root meristem growth by CLE45 requires the receptor-like kinase BAM3. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 7074–9 (2013).
98. Rodriguez-Villalon, A., Gujas, B., Kang, Y. H., Breda, A. S., Cattaneo, P., Depuydt, S. & Hardtke, C. S. Molecular genetic framework for protophloem formation. *Proc. Natl. Acad. Sci.* **111,** 11551–11556 (2014).
99. Rodriguez-Villalon, A., Gujas, B., van Wijk, R., Munnik, T. & Hardtke, C. S. Primary root protophloem differentiation requires balanced phosphatidylinositol-4,5-biphosphate levels and systemically affects root branching. *Development* **142,** 1437–1446 (2015).
100. Nagasawa, N., Miyoshi, M., Kitano, H., Satoh, H. & Nagato, Y. Mutations associated with floral organ number in rice. *Planta* **198,** 627–633 (1996).
101. Chu, H., Qian, Q., Liang, W., Yin, C., Tan, H., Yao, X., Yuan, Z., Yang, J., Huang, H., Luo, D., Ma, H. & Zhang, D. The FLORAL ORGAN NUMBER4 Gene Encoding a Putative Ortholog of Arabidopsis CLAVATA3 Regulates Apical Meristem Size in Rice. *PLANT Physiol.* **142,** 1039–1052 (2006).
102. Suzaki, T., Toriba, T., Fujimoto, M., Tsutsumi, N., Kitano, H. & Hirano, H.-Y. Conservation and Diversification of Meristem Maintenance Mechanism in Oryza sativa : Function of the FLORAL ORGAN NUMBER2 Gene. *Plant Cell Physiol.* **47,** 1591–1602 (2006).
103. Fisher, K. & Turner, S. PXY, a Receptor-like Kinase Essential for Maintaining Polarity during Plant Vascular-Tissue Development. *Curr. Biol.* **17,** 1061–1066 (2007).
104. Etchells, J. P., Provost, C. M., Mishra, L. & Turner, S. R. WOX4 and WOX14 act downstream of the PXY receptor kinase to regulate plant vascular proliferation independently of any role in vascular organisation. *Development* **140,** 2224–2234 (2013).
105. Etchells, J. P., Mishra, L. S., Kumar, M., Campbell, L. & Turner, S. R. Wood Formation in Trees Is Increased by Manipulating PXY-Regulated Cell Division. *Curr. Biol.* **25,** 1050–1055 (2015).
106. Hothorn, M., Belkhadir, Y., Dreux, M., Dabi, T., Noel, J. P., Wilson, I. a & Chory, J. Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474,** 467–71 (2011).
107. She, J., Han, Z., Kim, T.-W., Wang, J. J., Cheng, W., Chang, J., Shi, S., Wang, J. J., Yang, M., Wang, Z.-Y. & Chai, J. Structural insight into brassinosteroid perception by BRI1.

Nature **474,** 472–6 (2011).

108. Wang, J., Li, H., Han, Z., Zhang, H., Wang, T., Lin, G., Chang, J., Yang, W. & Chai, J. Allosteric receptor activation by the plant peptide hormone phytosulfokine. *Nature* **525,** 265–268 (2015).
109. Sun, Y., Li, L., Macho, A. P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.-M. J. & Chai, J. Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science* **342,** 624–8 (2013).
110. Tang, J., Han, Z., Sun, Y., Zhang, H., Gong, X. & Chai, J. Structural basis for recognition of an endogenous peptide by the plant receptor kinase PEPR1. *Cell Res.* **25,** 110–20 (2015).
111. Walter, T. S., Meier, C., Assenberg, R., Au, K.-F. F., Ren, J., Verma, A., Nettleship, J. E. E., Owens, R. J., Stuart, D. I. I. & Grimes, J. M. Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure* **14,** 1617–1622 (2006).
112. Waterman, D. G., Winterb, G., Parkhurst, J. M., Fuentes-Monterob, L., Hattne, J., Brewster, A., Sauter, N. K. & Evans, G. The DIALS framework for integration software. *CCP4 Newsl. PROTEIN Crystallogr.* 16–19 (2013).
113. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 1002–1011 (2006).
114. 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. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).
115. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4,** 1633–1649 (2004).
116. Zhang, Z. & Thomma, B. P. H. J. Structure-function aspects of extracellular leucine-rich repeat-containing cell surface receptors in plants. *J. Integr. Plant Biol.* **55,** 1212–23 (2013).
117. Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K. a, Savino, C., De Lorenzo, G., Tsernoglou, D. & Cervone, F. The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl. Acad. Sci.* **100,** 10124–10128 (2003).
118. Zhang, H., Lin, X., Han, Z., Qu, L.-J. & Chai, J. Crystal structure of PXY-TDIF complex reveals a conserved recognition mechanism among CLE peptide-receptor pairs. *Cell Res.* **26,** 543–555 (2016).
119. Zhang, H., Lin, X., Han, Z., Wang, J., Qu, L.-J. & Chai, J. SERK Family Receptor-Like Kinases Function as a Co-receptor with PXY for Plant Vascular Development. *Mol. Plant* **9,** 1406–1414 (2016).
120. Sun, Y., Han, Z., Tang, J., Hu, Z., Chai, C., Zhou, B. & Chai, J. Structure reveals that

BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. *Cell Res.* **23,** 1326–1329 (2013).

121. Song, X.-F., Guo, P., Ren, S.-C., Xu, T. & Liu, C. Antagonistic peptide technology for functional dissection of CLV3/ESR genes in Arabidopsis. *Plant Physiol.* **161,** 1076–85 (2013).
122. Hirakawa, Y., Shinohara, H., Welke, K., Irle, S., Matsubayashi, Y., Torii, K. U. & Uchida, N. Cryptic bioactivity capacitated by synthetic hybrid plant peptides. *Nat. Commun.* **8,** 14318 (2017).
123. Mou, S., Zhang, X., Han, Z., Wang, J., Gong, X. & Chai, J. CLE42 binding induces PXL2 interaction with SERK2. *Protein Cell* **8,** 612–617 (2017).
124. 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. Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–8 (2007).
125. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31,** 3320–3 (2003).

General Discussion

In this dissertation, I focused on cell surface proteins of higher eukaryotes, and analyzed the X-ray crystal structure of the complex with their ligands in order to elucidate their ligand recognition mechanism and substrate specificity. Expression of these proteins in eukaryotic cells in the secreted form in this study enabled us to observe the common features of these proteins such as *N*-linked glycans and disulfide bonds to keep them stable. Cell surface proteins are sometimes targeted for drug discovery, and clarifying the molecular mechanisms for the recognition of foreign body recognition by them is also indispensable for understanding essential biological phenomena. It is expected that deepening the understanding of the ligand recognition mechanism of these proteins will lead to an understanding of the structural basis in which each protein is involved in its own physiological functions.

In Chapter 1, I determined the crystal structure of the extracellular phosphodiesterase ENPP6, a choline metabolizing enzyme of vertebrate animals, in complex with a reaction product, phosphocholine. The obtained structure, biochemical analysis and cell physiological analysis revealed that ENPP6 metabolizes choline-containing phosphodiester to produce phosphocholine, so that choline can be taken up into cells. In addition, structural comparison provided insights into the characteristics of the pocket structure that determines the substrate specificity of the ENPP family protein.

In Chapter 2, the crystal structure of a plant hormone receptor TDR was determined in complex with its ligand TDIF at a resolution of 3.0 Å and the structure of the side chain of the amino acid residue of TDIF was cleary defined. The obtained structure and mutant analysis revealed that TDIF forms a kink, which is recognized by the pocket of TDR in a shape-complementary manner. Furthermore, TDR specifically interacts with amino acid residue of TDIF at the N-terminal and C-terminal. Sequence comparison suggested that TDR distinguishes TDIF from other CLE family peptides by this difference in the terminal part. Structural comparison between TDR and known LRR-RKs gave suggestions for both universal features and diversity features in plant LRR-RK structures.

While most of the cell surface receptors in animals are G-protein-coupled receptors

(GPCR), most of them in plants are enzyme-coupled recepors. In animals, most of enzyme-coupled cell surface receptors are receptor tyrosine kinase (RTK), which are rarely found in plants. Instead, the largest class of enzyme-coupled cell surcface receptor in plants is transmembrane receptor serine/threonine kinase, among which LRR-RKs with a tandem array of extracellular LRR are most abundant types. Whereas the identification, functional analysis, and characterization of ligands of them are being advanced, their structural characterization is relatively delayed. The structures obtained in this work and the findings derived from them are expected to help structural and functional analysis of the plant LRR-RK in the future.

In this study, I revealed the substrate recognition mechanism of cell surface proteins of higher eukaryotes at sub-nanometer resolution, revealing the structural basis that each protein exerts its physiological function. In addition, the various methods used in the course of this research may help research to elucidate the structure of other extracellular proteins in the future.

Original papers

1) Junko Morita, Kazuki Kato, Emiko Mihara, Ryuichiro Ishitani, Junichi Takagi, Hiroshi Nishimasu, Junken Aoki and Osamu Nureki "Expression, purification, crystallization and preliminary X-ray crystallographic analysis of Enpp6" ***Acta Crystallographica Section F***, 70; 794–799 (2014)

2) Junko Morita, Kuniyuki Kano, Kazuki Kato, Hiroyuki Takita, Hideki Sakagami, Yasuo Yamamoto, Emiko Mihara, Hirofumi Ueda, Takanao Sato, Hidetoshi Tokuyama, Hiroyuki Arai, Hiroaki Asou, Junichi Takagi, Ryuichiro Ishitani, Hiroshi Nishimasu, Osamu Nureki and Junken Aoki, " Structure and biological function of ENPP6, a choline-specific glycerophosphodiester-phosphodiesterase" ***Scientific Reports***, 6; Article number: 20995 (2016)

3) Kazuki Kato, Hisako Ikeda, Shin Miyakawa, Satoshi Futakawa, Yosuke Nonaka, Masatoshi Fujiwara, Shinichi Okudaira, Kuniyuki Kano, Junken Aoki, Junko Morita, Ryuichiro Ishitani, Hiroshi Nishimasu, Yoshikazu Nakamura and Osamu Nureki, "Structural basis for specific inhibition of Autotaxin by a DNA aptamer" ***Nature Structural & Molecular Biology.***, 23; 395–401 (2016)

4) Kazuki Kato, Yuhkoh Satouh, Hiroshi Nishimasu, Arisa Kurabayashi, Junko Morita, Yoshitaka Fujihara, Asami Oji, Ryuichiro Ishitani, Masahito Ikawa and Osamu Nureki, "Structural and functional insights into IZUMO1 recognition by JUNO in mammalian fertilization" ***Nature Communications***, 7, Article number: 12198 (2016)

5) Junko Morita, Kazuki Kato, Takanori Nakane, Yuki Kondo, Hiroo Fukuda, Hiroshi Nishimasu, Ryuichiro Ishitani and Osamu Nureki, "Crystal structure of the plant receptor-like kinase TDR in complex with the TDIF peptide" ***Nature Communications***., 7, Article number: 12383 (2016)

Acknowledgements

First and foremost, I owe my deepest gratitude to my supervisor Prof. Osamu Nureki for his guidance and advice throughout the years. Thank you for giving me the opportunity to undertake this study. Furthermore, I am grateful for the invaluable support, guidance and advice that I received from my supervisors Dr. Kazuki Kato, Dr. Hiroshi Nishimasu and Dr. Ryuichiro Ishitani.

I would like to express my sincere gratitude to the following collaborators for their contributions and assistance towards this dissertation: Prof. Junken Aoki, Dr. Kuniyuki kano, Mr. Hiroyuki Takita (Tohoku University), Prof. Hiroo Fukuda and Dr. Yuki Kondo (The University of Tokyo). I would like to thank beamline staff of BL32XU and BL41XU in SPring-8 and PXI of Swiss Light Source for the technical support with data collection. Moreover, I would like to thank the people who kindly assisted me during this PhD project including Ms. Arisa Kurabayashi, Ms. Ogomori, Ms. Masae Miyazaki and Ms. Sanae Okazaki for technical assistance, Ms. Rieko Yamasaki for secretarial assistance and Mr. Takanori Nakane for assistance in data processing and technical support. Many thanks go to my contemporaries whom always had my back when I needed them.

Last but not the least, my family for all the support and the loving environment that has allowed me to achieve all that I have today.