博士論文

Studies on the interaction of α/β-hydrolase-type plant hormone receptors with their regulators

（**α/β-**加水分解酵素型植物ホルモン受容体と制御剤の相互作用解析研究）

王　建文

Studies on the interaction of α/β-hydrolase-type plant hormone receptors with their regulators

α/β-加水分解酵素型植物ホルモン受容体と制御剤の相互作用解析研究

王 建文

WANG Jianwen

Supervisor: Prof. Dr. Tadao Asami

A thesis for the degree of Doctor of Philosophy in The Applied Biochemistry

Tokyo, 2022

Contents

List of Figures

List of Tables

Abbreviations

Abbreaviation	Full Name
(*S*)-dCN-deb	Desmethyl-CN-debranone
ABA	Abscisic acid
ABH	α/β-hydrolase
ABHD	α/β hydrolase domain
ABPP	Activity-based protein profiling
AD	Activation domain
AtDSEL	Arabidopsis thaliana DAD1-like Seedling Establishment-related Lipase
BD	DNA-binding domain
BRs	Brassinosteroids
CAST	Combinatorial active-site saturation test
CNB	*p* -chloronitrobenzene
Co-IP	Co-immunoprecipitation
CSM	Combinatorial saturation mutagenesis
D3	DWARF3
DSF	Differential Scanning Fluorimetry
DTT	Dithiothreitol
DWARF14	DWARF14
dYLG	Desmethyl-Yoshimulactone Green
E.coli	Escherichia coli
epPCR	Error-prone polymerase chain reaction
ESTHER	ESTerases and alpha/beta-Hydrolase Enzymes and Relatives
FFAs	Free fatty acids
FP	Fluorophosphonate
GA	Gibberellic acid
GA20ox2	Gibberellin 20 oxidase
GAI	GIBBERELLIN-INSENSITIVE
GAs	Gibberellins
GID1	GIBBERELLIN INSENSITIVE DWARF 1
GID2	GIBBERELLIN-INSENSITIVE DWARF2
GRAS	GAI-RGA-and-SCARECROW
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HNL	Hydroxynitrile lyase
HSLs	Hormone-sensitive lipases
HTL/KAI2	HYPOSENSITIVE TO LIGHT/KARRIKIN INSENSITIVE 2
IAA,auxin	Indole-3-acetic acid
IRM	Iterative random mutagenesis
ISM	Iterative saturation mutagenesis
ITC	Isothermal titration calorimetry

ITF	Intrinsic Tryptophan Fluorescence
JA	Jasmonic acid
KARs	Karrikins
K_d	Dissociation constant
KK094CM	Covalently molecule of KK094
MAX2	MORE AXILLARY GROWTH2
MKS1	Methylketone Synthase1
MKS2	Methylketone Synthase2
N-Ex	N-terminal extension
PA	Phosphatidic acid
PB	Sodium/potassium phosphate buffer
RGA	REPRESSOR OF GA1–3
RGL-1, 2, 3	RGA-like proteins1, 2, 3
RMSD	Root-mean-square deviation
SA	Salicylic acid
SAR	Structure–activity relationship
SCF	SKP1–CULLIN–F box
SC-LW	Synthetic complete dropout medium lacking the amino acids leucine, tryptophan
SC-LWHA	Synthetic complete dropout medium lacking the amino acids leucine, tryptophan, histidine and adenine
SCOP	Structural Classification of Proteins
SL	Strigolactone
SLR1	SLENDER RICE1
SLY1	SLEEPY1
SMXL	SUPPRESSOR-OF-MAX2-1/(SMAX1)-LIKE
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
Y2H	Yeast-two hybrid
YAG	Triacylglycerol
ΔG	Binding free energy

Chapter 1: Introduction

1.1 α/β-hydrolase fold proteins

1.1.1 General description of α/β-hydrolase fold proteins

The α/β-hydrolase (ABH) fold, which was firstly identified by the findings of structural similarities of five functionally diverse hydrolytic enzymes in 1992, are increasingly becoming the largest structural families of protein exist in all three kingdoms of life[1, 2]. It's believed that the members of this superfamily possibly have been undergoing divergent evolution from a common ancestor, eventually presenting poor sequence similarity and remarkable structural versatility. This may partially explain the truth that ABH fold proteins have served with a diverse set of functions as hydrolases, lyases, transferases, hormone receptor, transporters, chaperones or routers of other proteins, and participate in various important biological pathways including biosynthesis, metabolism, signal transduction, and gene regulation[3, 4].

Despite their functional divergence, α/β hydrolase fold proteins have shared a canonical structural framework with high plasticity, that is originally defined with mostly parallel 8-11 stranded β-sheet (ordered as 12435678), with the exception of the antiparallel strand β2, and is flanked by α helices (ordered as A–F), as illustrated in Figure 1-1[2, 5]. The β-strands and the α-helices are organized in an α-turn-β supersecondary structure geometry, where the central β sheets are strongly twisted and bent to form a half-barrel, and the first and the last strands, named as αA and αF, cross each other at an angle of approximately 90° while αB–αE pack onto the concave surface[6]. Typically, a substrate-binding crevice at the cross-over connection on strand β5, and some insertions may occur after one or more of the β strands β3, β4, β7 and β8 (Figure 1-1)[6]. In other cases, additional structure motifs also exist and are inlaid in different positions to comprise a substrate-binding domain, giving rise to the specific functions for catalysis of enzymes or interaction with partner or other non-catalytic roles, for example, the helical lid of lipases (sometimes it's called cap or flap in other proteins) confer essential recognition for substrates at lipid-water interface[5, 7]. However, in spite of substructural diversity, the α/β hydrolase fold is capable of providing a stable scaffold for the similar catalytic machinery, giving members of the family a striking ability for adaptation and evolution.

The catalytic residues, which are always found at identical locations within the ABH fold, constitute a highly conserved triad, Nucleophile-His-Acid (Figure 1-1): a nucleophile (serine, cysteine or aspartic acid) is disposed after strand β5; an acidic residue (aspartic or glutamic acid)

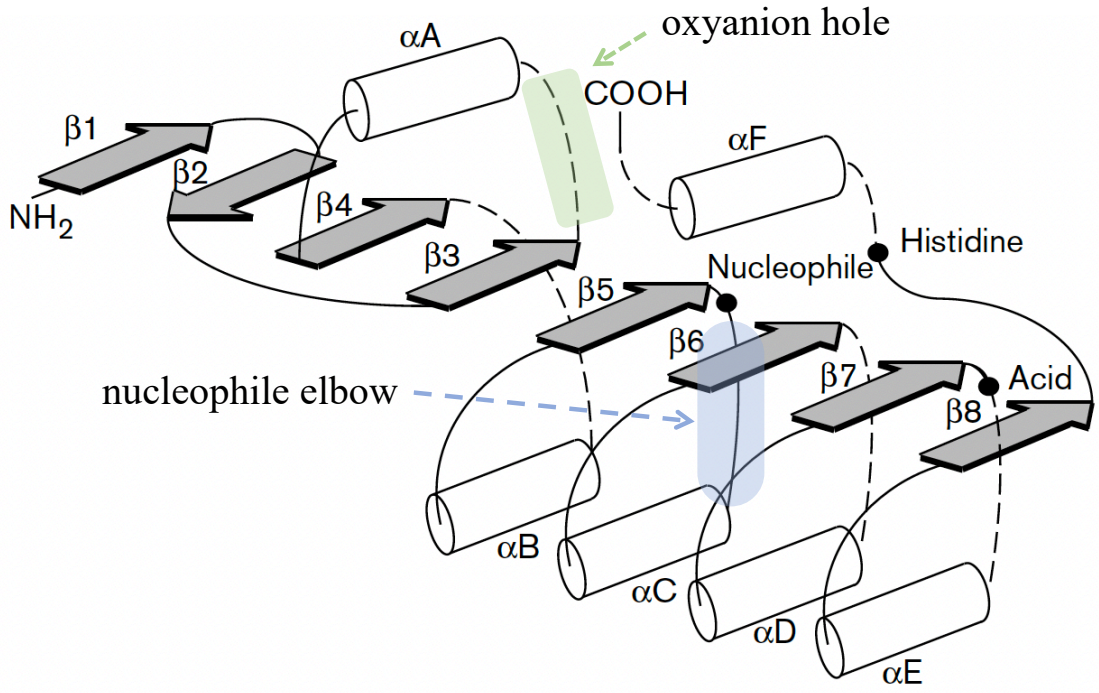


Figure 1-1: Topological diagram of the canonical α/β-hydrolase fold.
The α-helices are indicated by white cylinders and β-strands by gray arrows. The positions of the residues defining the Nucleophile-Histidine-Acid catalytic triad are indicated by black dots. Dashed lines indicate the location of possible insertions. The nucleophile elbow element created by the sharp turn between β5 and αC is shown (blue dashed arrow). The oxyanion hole is usually situated between β3 and αA (green dashed arrow).
Reprinted and modified from "α/β Hydrolase fold enzymes: the family keeps growing. Nardini & DijkstraCurrent Opinion in Structural Biology, 1999, 9:73

positioned after strand β7 or on a tight turn after strand β6; and an absolutely conserved histidine residue located on the loop following the last β strand[6]. In particular, the catalytic nucleophile makes a dominant role in a unique charge-relay system to initiate the covalent reaction through nucleophilic attack. Since the nucleophile is always located at the apex of sharp turn between strand β5 and αC, it can be sufficiently approached by the substrate or ligand. This structural motif is well-known as "nucleophile elbow" and it retains the sequence conservation as Sm-X-Nuc-X-Sm, where Sm is a small residue, X is any residue, and Nu is the nucleophile). The geometry of the nucleophile elbow also leads to form the oxyanion-binding site, known as "oxyanion hole", an important part for the formation of the active site of all ABH fold proteins (Figure 1-1). The oxyanion hole is usually opposite to the catalytic triad and shaped by the main chain nitrogen atoms of two residues, one of which immediately follows the catalytic nucleophile and the second is situated between C-terminus of strand β3 and helix αA, contributing to the spatial arrangement of hydrogen bond donors for the stabilization of the negatively charged transition state formed during hydrolysis[1, 6, 8]. Of the other proteins lacking from one to all of the key catalytic residues, some may lose hydrolytic activity while others are known to be involved in special functions, such as the surface recognition with partner protein in the individual signaling pathways[9].

The conventional description of α/β-hydrolase fold mentioned above has been universally accepted and adopted by several databases: SCOP[10] (http://scop.mrc-lmb.cam.ac.uk/scop/); CATH (http://www.biochem.ucl.ac.uk/bsm/cath/)[11]; and the α/β hydrolase fold database ESTHER[12, 13] (http://www.ensam.inra.fr/cholinesterase/). More recently, some studies, after a comprehensive analysis of 40 ABH protein families from SCOP, have observed an aromatic cluster near the nucleophile and oxyanion zones, which opposite the ligand-binding site. This aromatic cluster is thought to co-ordinate the catalytic histidine loop to approach the catalytic acid and nucleophile through conserved interaction, including hydrogen bonds networks and CH-π contact. This unique type of structural organization shapes the active site of ABH enzymes and influence the properties of the proteins, such as substrate specificity, enantioselectivity, and thermostability[8, 14, 15].

1.1.2 Studies of α/β-hydrolase fold proteins: the biology and chemistry

Powered by the great revolution of complete genome sequencing, an increasing number of genes

are able to be characterized and annotated, providing a framework for understanding complex cellular and physiological processes at a biochemical level[16]. ABH fold, which is the most frequently found in the structural framework of proteins, are served to evolve a unique function of proteins in the life process that ultimately benefits the host organism. In recent years, the study of α/β hydrolase fold proteins has been greatly advanced, especially in the drugs discovery for therapeutics, chemical tools development for probing certain signaling pathways, and activity-based protein profiling through irreversible reaction[4, 17, 18]

As one of the tremendous families covered by the ABH fold, serine hydrolases constitute higher than 1% of the human proteome and work as abounding roles[19]. In mammalian, genes that encode α/β hydrolase fold domain proteins (ABHD1-18) belonging to serine hydrolases family modulate the lipid signaling and metabolism, which are essential as molecules for energy storage, membrane structure, and signal transduction in life cycle[20]. They are consistently functioning to regulate various fundamental cellular processes and correlative with many human diseases including obesity, diabetes, cancer and so on, indicating the enormous potential to be drug targets. Recent studies have highlighted the significance of ABHD as potential targets for therapies. For instance, a high selective inhibitor targeting ABHD2 is accessible to regulate sperm fertility, as well as leading to novel contraceptives[21]; and ABHD6 inhibition may be effective in preventing obesity, nonalcoholic fatty liver disease, and type II diabetes[22]; and a reversible and super-potent inhibitor of ABHD12 help to make improvement for rare neurodegenerative disorder[23].

On the other hand, chemical modification of protein has gained much attention in chemical biology, such as development of small-molecular toolbox and activity-based protein profiling (ABPP), the former of which will be discussed in section 1.3. ABPP, which typically relies on the design of active-site-directed covalent probes for a quantitative readout, has emerged as a powerful technology to investigate the uncharacterized proteins and decipher their function in native physiological environment[24]. This residue-specific covalent modifications desire for at least two elements: 1) the reactive residues, such as serine, cysteine, lysine, tyrosine, threonine, aspartate, and glutamate; 2) the activity-based probes, which contain a reactive group to label mechanistically active residues of proteins and a reporter tag to visualize, enrich, and identify probe-labeled proteins (Figure 1-2)[17, 25]. As the most conserved features in structural organization, the catalytic nucleophiles of ABH folds covering serine, cysteine and aspartic acid have been extensively applied

to study activity-based proteomics of ABH superfamily by judicious selection of an electrophile with the appropriate affinity and selectivity[24]. The annotation of unknown proteins of serine hydrolase family is dramatically advanced by the adoption of fluorophosphonate (FP) as the prototypical activity-based probes with negligible cross-reactivity to other nucleophilic enzymes such as cysteine hydrolases (Figure 1-2c) [26]. It is suggested that the highly efficient serine-FP reaction have facilitated the detection of more than 80% of human/mouse serine hydrolase in ABPP experiments, which, however, indicates that the applicability of FPs as pharmacological probes for specific members of serine hydrolase is restricted[27]. Nowadays, researchers are offered more electrophiles for labeling serine hydrolase. For example, diphenyl phosphonates have been employed to develop selective covalent inhibitors for probing serine protease subfamily[28], and the wide spectrum activity of sulfonyl-fluoride derivatives are capable of covalently labeling multiple serine protease sub-classes[29]. More recently, the 1,2,3-triazole ureas, which covalently carbamoylate serine residues of active-site, display a pharmacologically privileged chemotype with broad activity and tunable selectivity for serine hydrolase inhibition[30]. The other family of α/β hydrolase folds, cysteine hydrolase, possess cysteine as catalytic nucleophile. The thiol group of cysteine residue shows stronger reactivity than hydroxyl of serine residue due to the polarizability and electron-rich nature of sulfur. The developed specific cysteine-reactive electrophiles include ketones (such as acyloxymethyl and chloromethyl ketone)[31], epoxides[32], sulfonate-ester[33], *p*-chloronitrobenzene (CNB)[34], chlorocetamides and iodoacetamides[35]. Of them, the highly tunable electrophile of CNB-based probes is broadly applicable for cysteine hydrolase in ABPP studies[36]. Notably, 1,2-epoxide and sulfonate-ester could be also used as the building block of chemical probes that covalently contact with the nucleophilic carboxylate residue of the acidic trials member that forms hydrogen bond with the basic residue, indicating its promiscuous performance in selectivity[37, 38].

The chemical regulators (or therapeutic drugs) targeting proteins with high selectivity is of great priority followed by sensitivity, but the annoying cross-reactivity of some electrophiles can still be modulated or even overcome by using a tight-binding ligand to direct the probe toward a specific nucleophilic residue proximal to the ligand-binding site. Anyway, since the biological importance of α/β-hydrolase superfamily proteins in life activity, the use of a combination of biochemical characteristics conferred by catalytic triads and chemical biology approaches will definitely expand its application in biology and chemistry.

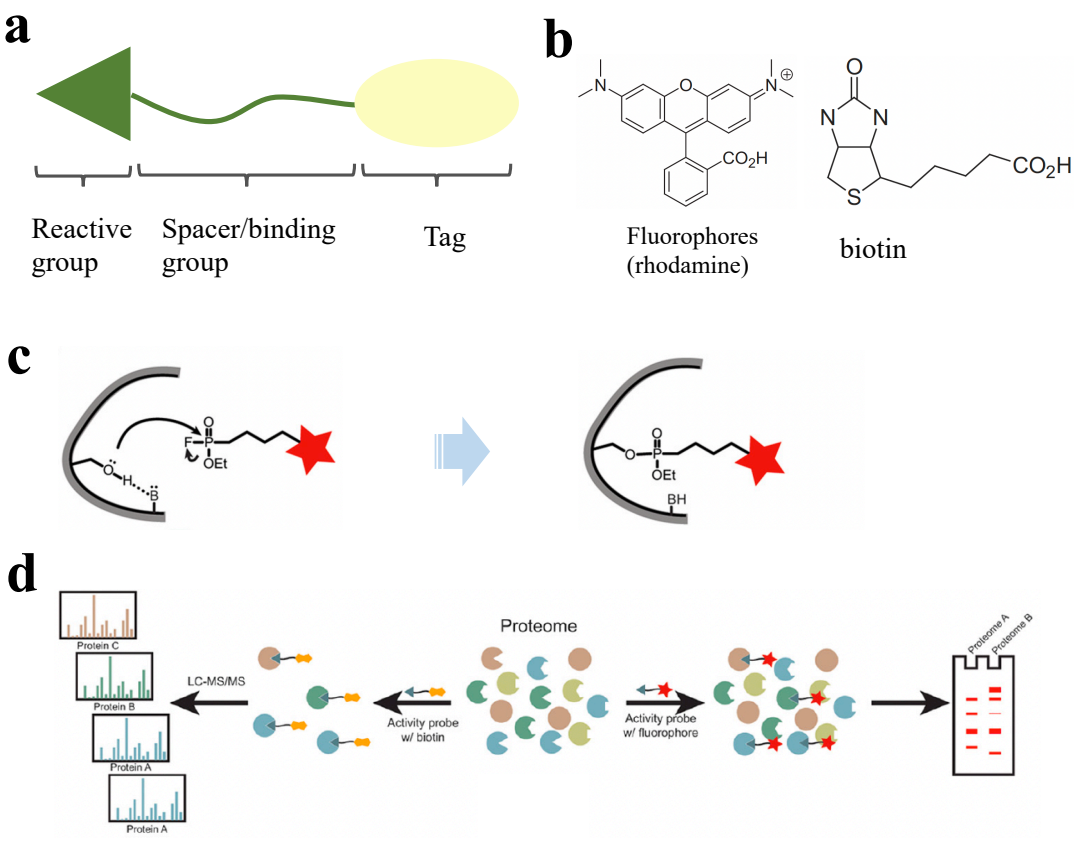


Figure 1-2: The application of activity-based protein profiling technique.
a, A standard structure of ABPP probe should contain a reactive group (green triangle), a spacer or binding group (green line), and a reporter tag (yellow ellipse); **b,** Representative structure of reporter tags used for proteome visualization and enrichment, including fluorophores (rhodamine) and biotin; **c,** an example for the covalent modification of the active-site serine nucleophile by a FP activity-based probe. The red stars indicate a tag, such as a fluorophore or biotin; **d,** Workflow of activity-based protein profiling. The proteome can either be labeled by fluorophore-conjugated probes (right) for the direct visualization by SDS-PAGE and fluorescent gel scanning or biotin-tagged probes (left) for the enrichment and identification by LC/LC-MS/MS.
Reprinted and modified from "Activity-based Proteomics of Enzyme Superfamilies: Serine Hydrolases as a Case Study. Simon & Cravatt. Journal of Biological Chemistry, 2010, 15:11051-11055"

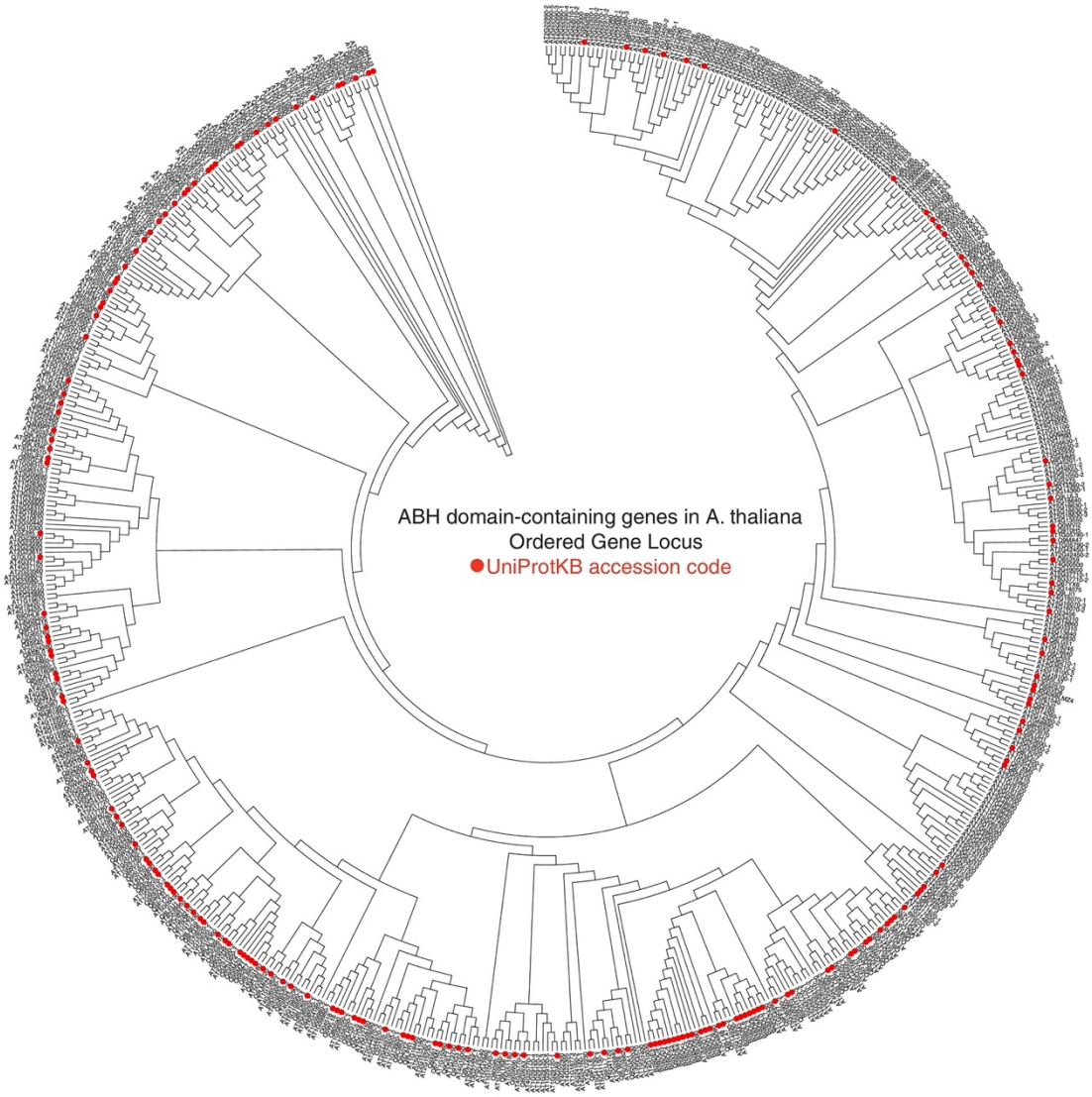


Figure 1-3: Phylogenetic tree analysis of all identifiable ABH-domain encoding genes from *A. thaliana*.

Jones-Taylor-Thornton model was used for the analysis and construction of phylogeny. All sequences were collected from the Interpro database (alpha/beta hydrolase fold: IPR029058). Protein sequences that have been reviewed and entered into the Swiss-Prot have been assigned an Ordered Gene Locus name, while proteins that have not been verified remain in the TrEMBL, have not been assigned an Ordered Gene Locus, and are labeled in the tree based on their UniProtKB accession code (Red dots). Proteins that are splice isoforms are identified with 1, 2, 3 and so on, after the ordered gene locus. Sequence alignments and tree calculation were conducted using MEGA version 5.

Reprinted from "Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. Mindrebo, Nartey, et al. Current Opinion in Structural Biology, 2016, 41:233–246.

1.2 α/β-hydrolase fold superfamily in plant kingdom

As a recurring event in long-term genome evolution, gene duplication plays a major forcing driver for the boom of protein families and the expansion of specialized metabolism, to increase their vulnerability to environmental stresses. The frequent cross-species transmission and high structural adaptability of protein fold provide the possibility of divergent evolution for the re-design or optimization of proteins in gain-of-function under natural selection, intensifying the resilience and adaption to biotic and abiotic factors from the molecular level to phenotypic readout that ultimately benefits the host organism. This strategy is particularly employed for the evolutionary emergence of land plant. Exemplified by *Arabidopsis thaliana*, in the analysis of a total of 35381 sequences that contain 32197 domains, 988 unique superfamilies covering 828 families were found with an incredible coverage of gene duplication at 97%, and 971 pairs of protein domain combinations further result in 3391 unique architectures, according to the database of SUPERFAMILY and SCOP (https://supfam.org/SUPERFAMILY/cgi-bin/gen_list.cgi?genome=at#section_statistics)[39-41].

Specifically, α/β-hydrolases superfamily, where 450 proteins of *Arabidopsis thaliana* are assigned, includes 8 families diversifying 14 different architectures, contributing to the sophisticated network of plant metabolism, growth, development, survival, and overall fitness (Figure 1-3) (https://supfam.org/SUPERFAMILY/cgi-bin/genome.cgi?sf=53474&listtype=sf&cgi_at=yes)[39, 42]. However, hundreds of ABH-containing sequences are uncharacterized, and those genes that have been characterized are almost encoded as enzymes involved in primary and specialized metabolism in the lifecycles of plants. For example, carboxylesterase family enzymes, which are capable of the ester hydrolysis of xenobiotics and endogenous metabolites, are competent as housekeeping roles to control the bioactivity and transport of herbicides, plant signaling agents and secondary metabolites in plants[43, 44]. Plants also make good use of proteases as regulators to degrade dead/nonfunctional proteins for recycling of amino acids. Despite of hundreds of proteases with unknown functions, they are proved to be indispensable in regulating different life processes in response to developmental and environmental cues, such as carboxypeptidase II that mostly use serine as nucleophile residue for the cleavage of negatively charged carboxylate terminus of their peptide substrates, and some of them even act as acyltransferases in the biosynthesis of sinapoyl esters to provide UV-B protection for plants,[45, 46]. More crucially, as stated in 1.1.2 that the lipid

signal and metabolism are regulated by a serial of α/β-hydrolases fold-containing proteins (ABHD1-18) in mammalian, the lipases from plants, like phospholipases, are responsible for the establishment of plant defense reactions through phospholipid-based signal perception, including production of important defense signaling molecules oxylipins and jasmonates, and cell-to-cell transduction of a second messenger phosphatidic acid (PA)[47, 48]. Lipids are also fundamental substances for energy storage, especially for the early stage of seed germination and seedling establishment. AtDSEL, an *Arabidopsis thaliana* DAD1-like Seedling Establishment-related Lipase, is involved in the degradation of storage triacylglycerol (TAG) in oil bodies to yield free fatty acids (FFAs) and glycerol as carbon sources for seed germination[49, 50].

On the other hand, besides the classical performance that hydrolyze peptide, such as oxyester and thioester bonds, plant ABHs also have covered a large diversity of enzymatic reactions by using the common or unique catalytic triads to accommodate different substrates. The atomic structure reveals how hydroxynitrile lyase (HNL) cleaves the C-C bond in cyanohydrins with the concomitant liberation of hydrogen cyanide upon tissue damage, which is suggested as a defense against predation in plants[51]. It is interesting to notice that HNL (*Hevea brasiliensis*) not only maintain the canonical catalytic triad of ABH (Ser80/His235/Asp207), but also employ the lysine residue (Lys236) as the modulator that fine-tunes the p*K*a of His235 and stabilizes negatively charged reaction intermediates through electrostatic interaction, which provides a likely structural explanation for the enantioselectivity of this enzyme[52]. Such a defense strategy is not unique but has its counterpart. To resist the invading of chewing insects, plants make use of the specialized methylketone-containing metabolites as toxic compounds produced by sequential reactions of Methylketone Synthase1 (MKS1) and MKS2, the former of which belongs to α/β-hydrolase fold superfamily. However, unexpectedly, even though two of key catalytic triad of MKS1, serine and aspartic acid, were replaced by alanine and asparagine, respectively, MSK1 still have the capacity to efficiently bind and subsequently decarboxylate 3-keto acid substrates to release methylketones with a narrow distribution of chain lengths (Figure 1-4). The structural determination and biochemical characterization of MSK1 reveals a novel enzymatic mechanism that a unique residue, threonine, is adopted to capture substrates through the interaction network with histidine, the only one conserved residue. And an active site-leading tunnel is also conducive to provide a hydrophobic environment for fatty-acyl chain binding[53]. These exclusive structural characteristics reflect that the

divergent evolution of α/β-hydrolases superfamily makes an immense contribution to the functional versatility of novel proteins to strengthen the environmental adaption of plants.

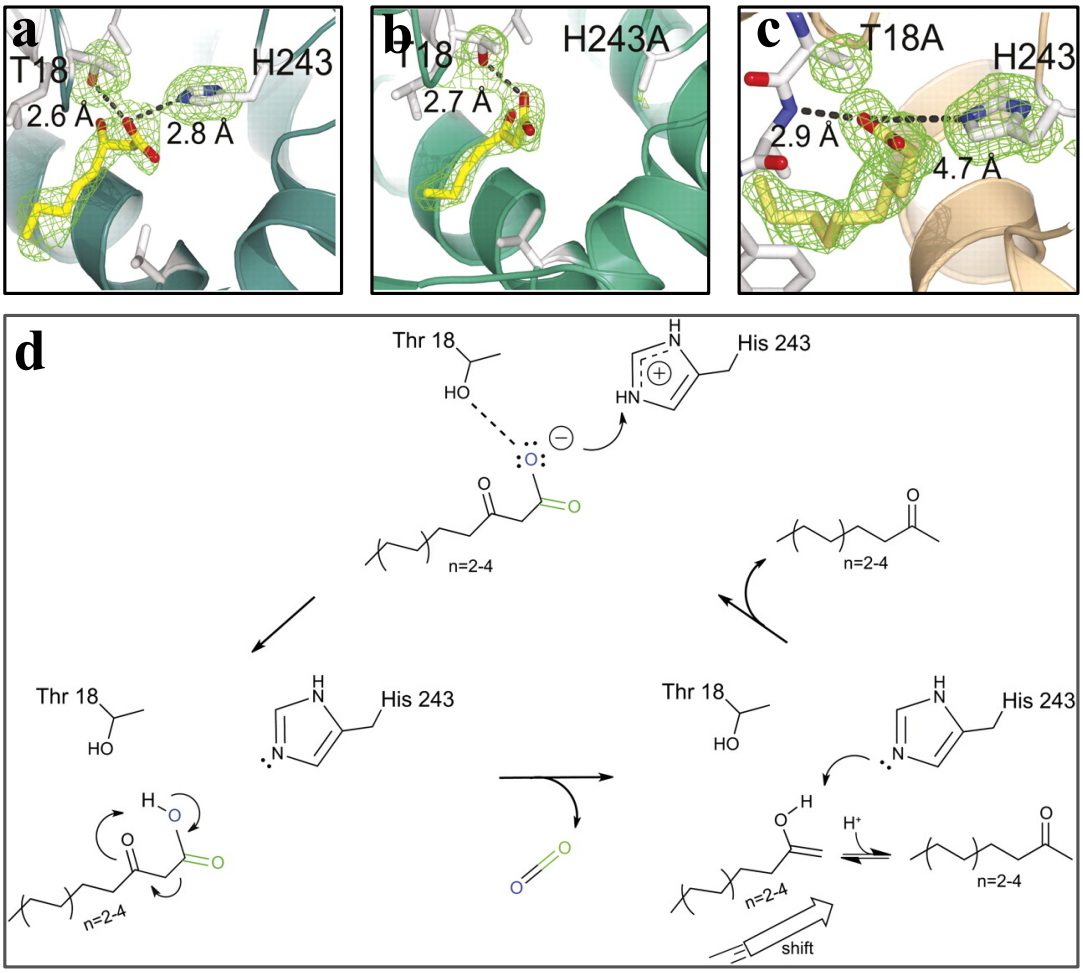


Figure 1-4: Structural basis of MKS1 reaction.

a, the substrate analog 3-hydroxyoctanoate (3HO) lay the acyl tail in the hydrophobic central tunnel and the polar head group point to Thr-18 and/or His-243 to form hydrogen bonds; **b,** Active-site variants H243A of MKS1 complex with substrate 3-ketoheptanoate (3KH); **c,** Active-site variants T18A of MKS1 complex with a fatty acid substrate. d, Proposed MKS1 decarboxylase reaction mechanism. 3-keto acid substrate have contacted with Thr-18 and His-243 within the MKS1 active site. His-243 acts as a proton donor to promote decarboxylation. Block arrow represents the shift in product orientation within the active site in relation to the substrate. *Reprinted and modified from "Emergent Decarboxylase Activity and Attenuation of α/β-Hydrolase Activity during the Evolution of Methylketone Biosynthesis in Tomato. Auldridge, Guo, et al. The Plant Cell, 2012, 24:1596-1607.*

1.3 α/β-hydrolase-type plant hormone receptors

Plant hormones, or phytohormones, covering abscisic acid (ABA), indole-3-acetic acid (IAA or auxin), jasmonic acid (JA), brassinosteroids (BRs), ethylene, cytokinin, gibberellic acid (GA), strigolactone (SL), and salicylic acid (SA), are a structurally unrelated collection of endogenously signaling and regulatory small molecules metabolically derived to regulate all aspects of plant lifestyle, from pattern formation during development to responses to biotic and abiotic stress[54, 55]. A core component for their biological functions is a receptor. With their respective receptors, these simple compounds are allowed to integrate the numerous and complicated internal physiological information and external stimuli to co-ordinate their growth and development accordingly. Notably, the topology versatility of ABH fold has made it possible to serve as an exquisite scaffold for bona fide receptors to perceive different types of plant hormones, such as gibberellin receptor *GIBBERELLIN INSENSITIVE DWARF 1* (GID1), strigolactone receptor DWARF14 (D14) and karrikin receptor *KARRIKIN INSENSITIVE 2* (KAI2) (Figure 1-5)[56, 57].

1.3.1 Gibberellin and its receptor GID1

Gibberellins (GAs) are a large family of tetracyclic diterpenoid plant hormones, and well acknowledged as the essential roles that regulate many developmental processes through the entire life cycle of the plant, including seed germination, cell division and elongation, leaf expansion, pollen maturation, root growth, flowering induction, fruit enlargement, etc[58-60]. So far, although up to 136 different gibberellin molecules have been discovered and characterized in plant, named from GA_1 to GA_{136}, only a few of which have biological activity as regulators of plant growth and development, such as GA_1, GA_3, GA_4, and GA_7 (Figure 1-6). Particularly, all these bioactive GAs possess the *ent*-GA skeleton that containing a C6 carboxyl group, a γ-lactone ring between C4 and C10, a hydroxyl group at C3, and loss of the hydroxyl group at C2 (Figure 1-6), and interestingly, biochemical and genetic experiments have suggested that the major forms of active GAs in most plant species are just GA_1 and GA_4, the latter of which is most active in both Arabidopsis and rice[61-63].

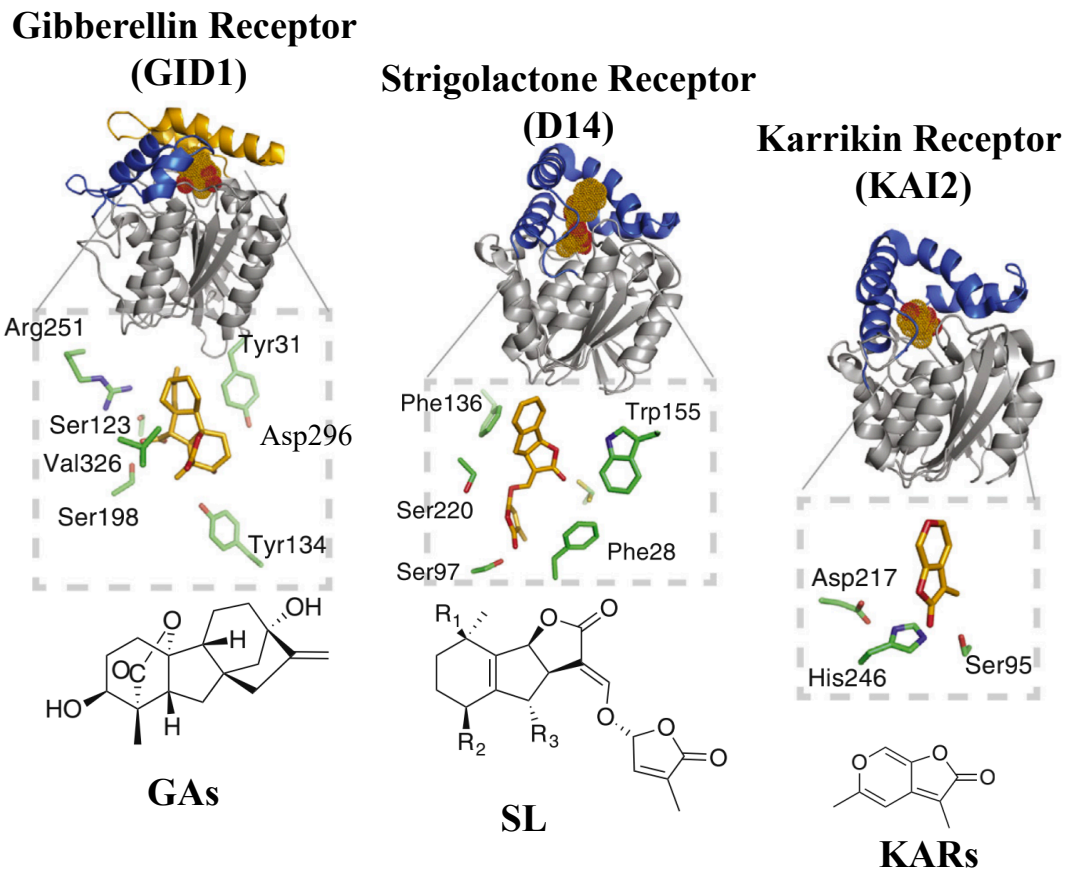


Figure 1-5: Structural insight of α/β-hydrolase-type plant hormone receptors.
These receptors include gibberellin an its receptor *GIBBERELLIN INSENSITIVE DWARF 1* (GID1), strigolactone and its receptor *DWARF14* (D14) and karrikin and its receptor *KARRIKIN INSENSITIVE 2* (KAI2). A close-up of the active site is shown below each overall structure of receptors. The catalytic residues of AtGID1 are S191-D289-V319, that of AtD14 are S97-H247-D218, and that of AtKAI2 are S95-D217-H246.
Reprinted and modified from "Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. Mindrebo, Nartey, et al. Current Opinion in Structural Biology, 2016, 41:233–246.

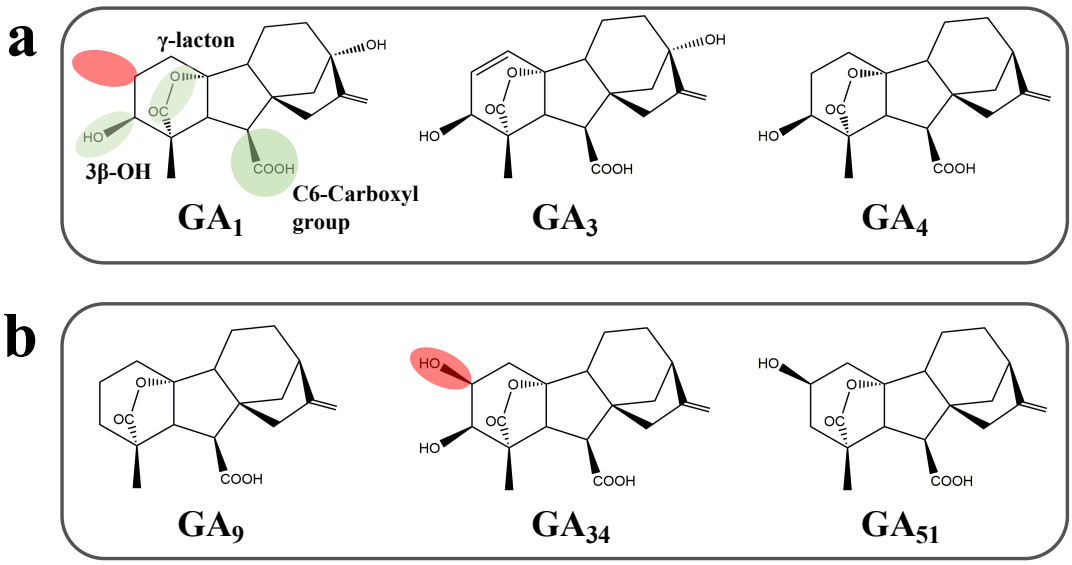


Figure 1-6: Chemical structures of biologically active (a) and inactive (b) GAs.
a, Bioactive GAs should contain a γ-lactone structure in the A-ring, 3β-hydroxylation (3β-OH) of the A-ring, and a C-6 carboxyl group in the B-ring, as illustrated in green ellipse in GA_1. **b,** The existence of 2β-hydroxylation (3β-OH) of the A-ring would abolish its biological activity in plant, red cycles in GA_{34} and GA_{51}.

In the recent decades, the remarkable progress has been made on uncovering the mode of action of gibberellin. Firstly, the genetic screening of rice (*Oryza sativa*) and *Arabidopsis thaliana* mutant contributed to identify a key component, the GA receptor GIBBERELLIN INSENSITIVE DWARF1 (GID1) as an essential gene in the GA insensitive dwarf mutant, *gid1*[56]. *In vitro* experiment of binding kinetics by using GST-GID1 and 16,17-dihydro-GA_4 suggested that the GID1 can directly but only interact with bioactive GAs with ligand saturability and specificity[64]. There exist three GA receptors with overlapping function in *Arabidopsis thaliana*, AtGID1a, AtGID1b, and AtGID1c, however, rice only have one GID1 protein, and all of them are localized to both cytoplasm and nucleus, revealed by the expression of GID1-GFP fusion protein[56, 65]. More crucially, DELLA proteins, which are defined by the presence of a DELLA regulatory domain at the N-terminus and a GRAS functional domain at the C-terminus, were identified as negative regulators of GA responses. Specifically, DELLA protein belongs to the GRAS superfamily of putative transcription factors that are universal in plants. DELLA regulatory domain contains N-terminal DELLA and TVHYNP motifs (Figure 1-7), both of which are indispensable to interact with GID1 proteins in a GA-dependent manner, as suggested by yeast-two hybrid (Y2H) assay[64, 66]. Rice has

only one DELLA protein, SLENDER RICE1 (SLR1), whereas five genes of Arabidopsis genome was encoded as DELLAs, including GA-INSENSITIVE (GAI), REPRESSOR OF GA1–3 (RGA), and three RGA-like proteins, RGL-1, RGL-2, and RGL-3, all of which play specific but also redundant functions to depress GA signaling. Moreover, studies show that GIBBERELLIN-INSENSITIVE DWARF2 (GID2) of rice[67] and SLEEPY1 (SLY1) of Arabidopsis [68] function as the F-box protein to interact with DELLA protein through GGF and LSL motifs, which is for the positive co-ordination of GA signaling (Figure 1-7). Collectively, after the formation of ternary complex of GID1-GA-DELLA, the F-box protein (GID2 in rice or SLY1 in Arabidopsis) was immediately recruited to bind with C-terminus of DELLA and confer the specificity to SKP1-CULLIN-F box protein (SCF) E3 ubiquitin ligase, which will subsequentially promote the addition of a polyubiquitin chain on DELLA for degradation by the 26S proteasome complex[63, 69-71], eventually releasing the expression repression on specific genes (Figure 1-8).

The subsequent effort was paid on the structural determination of OsGID1-GA[72] and AtGID1-GA-GAI[73] complex by X-ray crystallography, which strengthen the in-depth understanding of GA signaling transduction at molecular level. GID1 is a homologue of the hormone-sensitive lipases (HSLs) family, and its overall structure contains a core domain and a lid domain [72, 73]. The core domain adopts the canonical ABH fold that is composed of eight-stranded β-sheet with α-helices packing the sides and retains the conserved HGG and GXSXG motifs characteristic of hormone-sensitive lipases and other carboxylesterases (Figure 1-7). The N-terminal extension of the cap domain, which also exists in the HSLs family for lipid binding, extends up the core to provide a surface for the interaction with the DELLA domain, as well as to promote the recognition of GA on its inside surface. However, GID1 proteins can't show hydrolase activities due to the replacement of most conserved basic residue His within the Ser-His-Asp catalytic triad of the ABH family by other amino acids, such as Val in AtGID1a/c or Ile in AtGID1b. Notably, the reverse mutation of AtGID1a^{V319H} fail to rescue the hydrolytic activity, and other residues of GID1 pocket, as revealed by the results of Ala scanning experiments, are all essential for GA-binding activity and conserved within plant GID1 proteins but not among HSL proteins, indicating the long-term divergent evolution of GID1 from the ancestral HSL[72], as well as co-evolution of chemical and protein. The co-evolution of GA and GID1 can be further explained by the analysis of several GID1 homologs

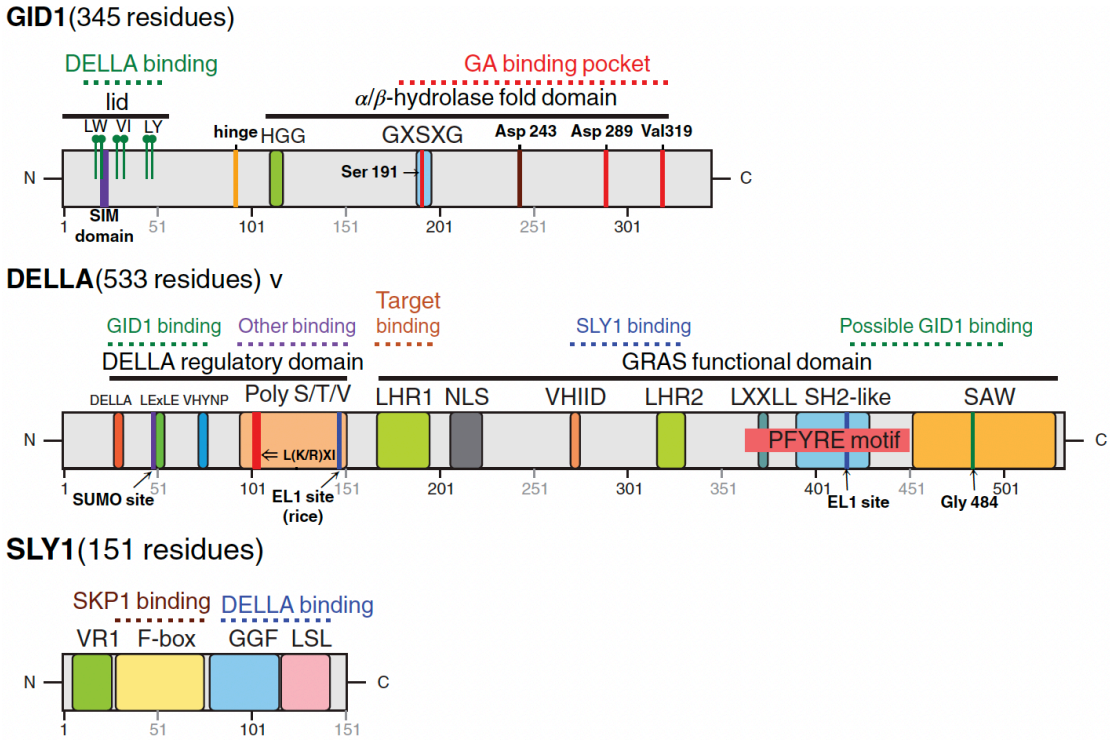


Figure 1-7: Diagrams of the GID1, DELLA and SLY1 domains, motifs and key amino acid residues.

Proteins are drawn to scale based on the amino acid sequences of Arabidopsis GID1a, DELLA GAI and SLY1. The key motifs and significant amino acid residues presented in the diagrams is explained below. 1) GID1: the hinge residue (orange bar for GID1a P92, OsGID1 P99); the catalytic triad (Ser 191, Asp 289, Val 319) involved in GA binding; the six lid hydrophobic residues involved in DELLA-binding (lollipops for L, W, V, I, L, Y); the SUMO-Interaction-Motif (SIM) domain (WVLI); the HGG and GXSXG motifs characteristic of hormone-sensitive lipases. 2) DELLA: The major DELLA regulatory domain and the GRAS functional domain are marked by black bars above the diagram. Within the DELLA regulatory domain the DELLA, LExLE and VHYNP motifs are involved in GID1 binding (shaded boxes), and the poly S/T/V motif contains the L(K/R)XI motif likely involved in binding an undetermined 'other' GA signalling component. The GRAS functional domain contains: two leucine heptad repeats (LHR1 and LHR2), a nuclear localisation signal (NLS), the VHIID, the PFYRE, LXXLL, SH2-like, and SAW motifs (shaded boxes). Significant residues are marked by full-height bars. 3) SLY1: SLY1 and GID2 contain the F-box domain that binds SKP1, the GGF and LSL motifs involved in DELLA-binding, and a variable region (VR1).

Reprinted from "Gibberellin hormone signal perception: down-regulating DELLA repressors of plant growth and development. Nelson, Steber, et al. Annual Review of Plant Biology, 2016, 49:153-188.

from different species and their affinity and specificity to different GAs. For example, Yoshida et al. performed phylogenetic analysis of 169 GID1s from 66 plant species and measure the binding affinity of proteins to active GA and inactive GA, found that two important amino acids (S127 and

I133 of OsGID1) in terms of GID1 evolution are not shared by *Selaginella moellendorffii* GID1 (SmGID1s) that was conferred the ability to perceive GA_{34} and GA_9 as an active form of GA[74, 75].

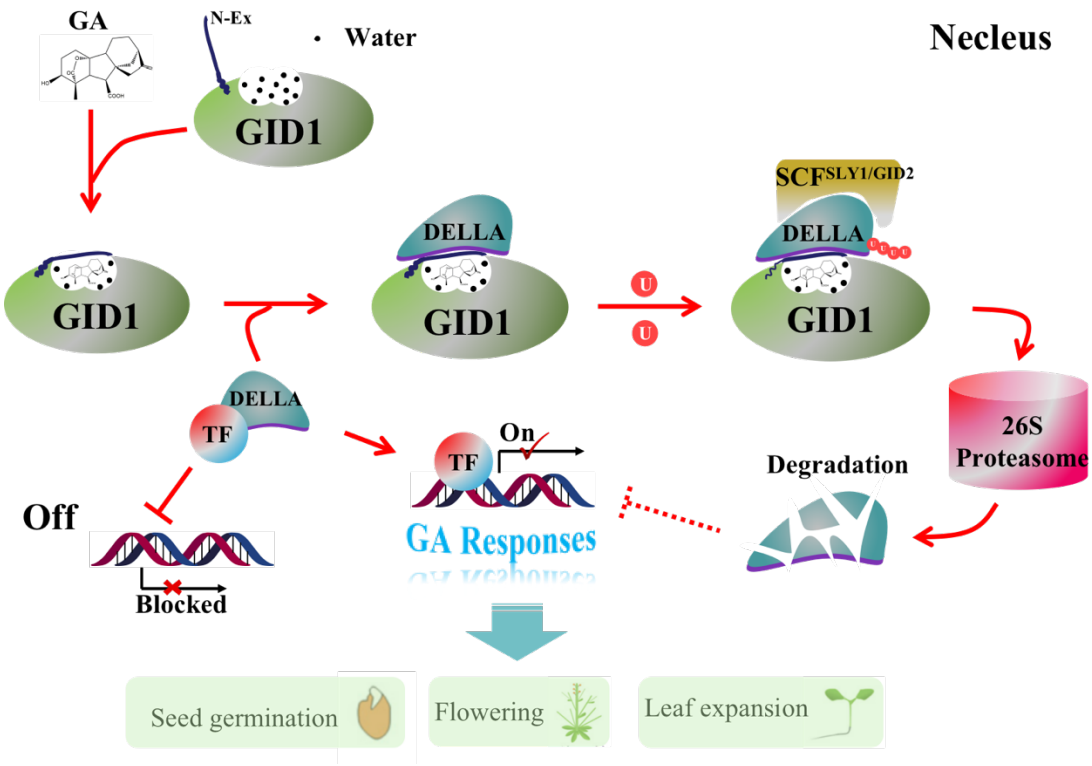


Figure 1-8: Model for GA action through DELLA protein degradation.
The binding of bioactive GA would induce a conformational switch in the N-Ex of GID1 to allow the formation of the GID1-GA-DELLA complex, which in turn force the SLY1/GID2 F-box protein to bind and polyubiquitinate DELLA, thereby targeting DELLA for destruction by the 26S proteasome. The degradation of DELLA protein then lifts the repression of GA-related gene to initiate GA responses, like promoting seed germination, plant flowering and leaf expansion.

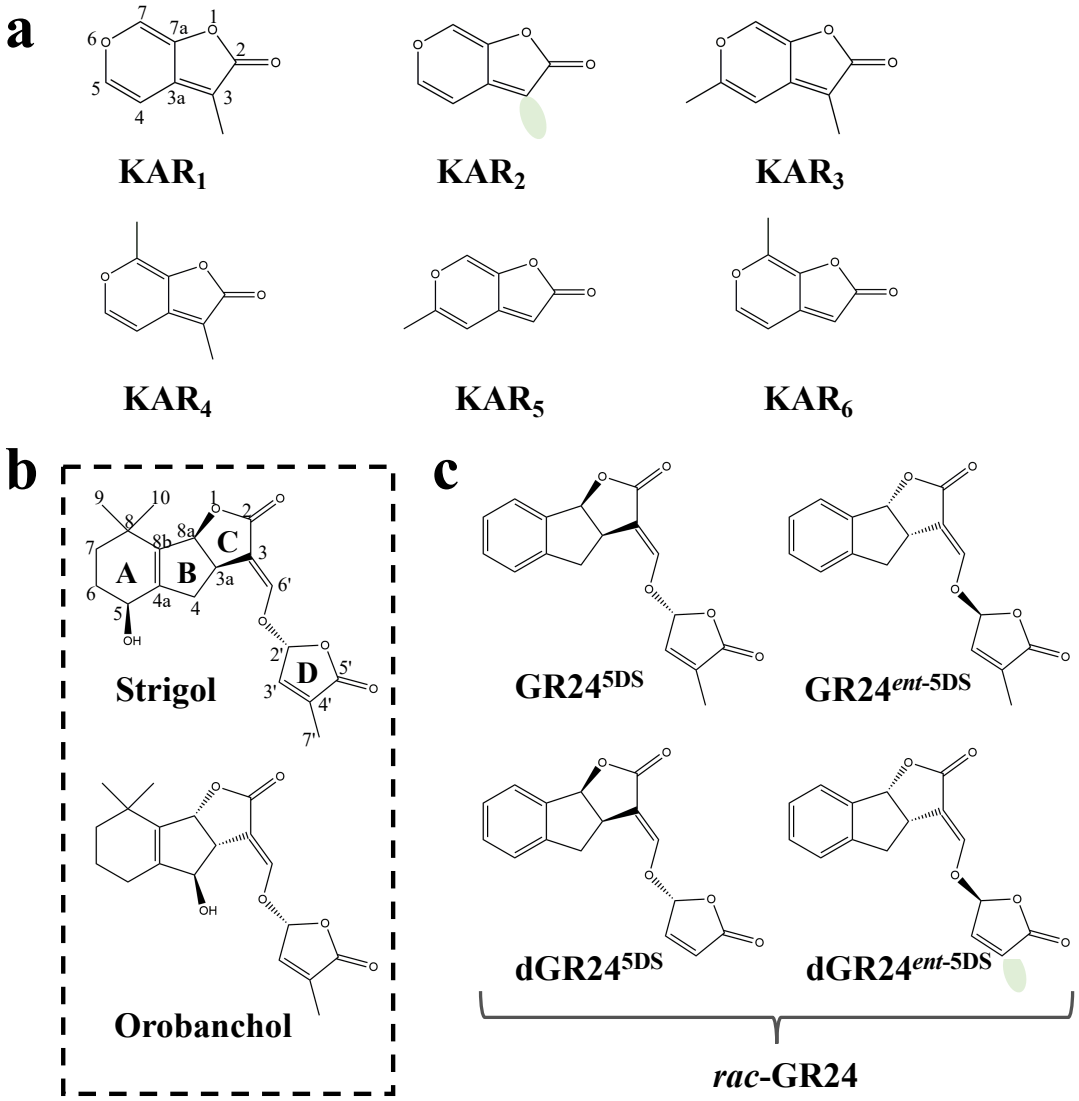


Figure 1-9: Chemical structures of various bioactive butenolides.
a, KAR_1 to KAR_6, six members of the karrikin family. KAR_1 and KAR_2 differ only by a methyl group at the 3 position on the butenolide ring (green ellipse), whereas KAR_2 shows highest biological activity in Arabidopsis. **b,** The strigol (first discovered) and orobanchol define the two major classes of strigolactones with characteristic ABC-D ring structure, they only differ by the stereochemical arrangement between the B and C rings (positions 3a and 8b). For naturally occurring SLs the ABC rings are attached to a 2′R-configured D ring. **c,** GR24 was an artificial strigolactone analogue with the same ABC–D structure. *rac*-GR24, mixture of $GR24^{5DS}$ and its enantiomerthe, is commonly used for research. Arabidopsis KAI2 shows enhanced responses to desmethyl butenolides (green ellipse in $dGR24^{ent\text{-}5DS}$).

1.3.2 Karrikin and its receptor KAI2

Vegetation fires have been seen as an important natural process for the regeneration and reshaping of ecosystems[76, 77]. This has endowed the plants with fire-adapted traits (pyrophytes), and the main drivers enable plants to get thrived after a wildfire are the supplies of nutrient cycling derived from smoke and ash production[78, 79], among which the so-called smoke water generated by soaking char through water was found to promote seed germination[80, 81]. In 2004, This active substance of smoke was isolated and identified as 3-methyl-2H-furo [2,3-c] pyran-2-one[82]. Afterward, other five analogues have been discovered and comprising a family of butenolides, annotated as karrikins, KAR_1 to KAR_6 (Figure 1-9)[83, 84]. Among them, KAR_1 containing a methyl group on its butenolide ring is the most abundant in the pyrolyzed-plant materials, and also indicated to be the most biologically active for seeds of a large proportion of plant species, even presented potent germination activity at concentrations as low as 1 nM for some species[82, 85, 86]. However, Arabidopsis is a notable exception, the desmethyl compound KAR_2 exhibit higher bioactivity than other KARs[87]. Nowadays, both KAR_1 and KAR_2 are commercially available and widely used in research of plant science.

A closely following attention was paid on the structural similarity of the butenolide moiety of KARs with enzymolysis product of strigolactones (SLs) (Figure 1-9). The early structure–activity relationship (SAR) studies of KARs and SLs demonstrated that the butenolide ring are indispensable to retain their bioactivity in plants[84, 88, 89]. However, a growing number of studies reveal that these two chemicals individually possess different physiological function. Strigolactones, which are a collection of carotenoid derived plant hormones, were originally discovered as germination stimulants of parasitic weeds in the Orobanchaceae family[90]. In the past decades, plenty of genetic analysis further uncovered numerous additional roles of SLs in regulation of plant growth and development, such as promotion of symbiotic interactions between roots and arbuscular mycorrhizal fungi[91], influencing shoot branching and root architecture (Figure 1-10)[92, 93]. In addition to being found as seed germination stimulant, KARs also enable to shape seedling photomorphogenesis, including the inhibition of hypocotyl elongation, and the promotion of cotyledon expansion, etc (Figure 1-10)[87, 94]. Recent studies further demonstrated that KARs could increase seedling vigor of some crop species under biotic and abiotic stresses[95-97]. Nevertheless, karrikins fail to break dormancy of parasitic weed seeds and shape shoot architecture[98], which indicate that KARs are not

just structurally SL mimics, but essentially distinct from SL in the aspect of physiological functions.

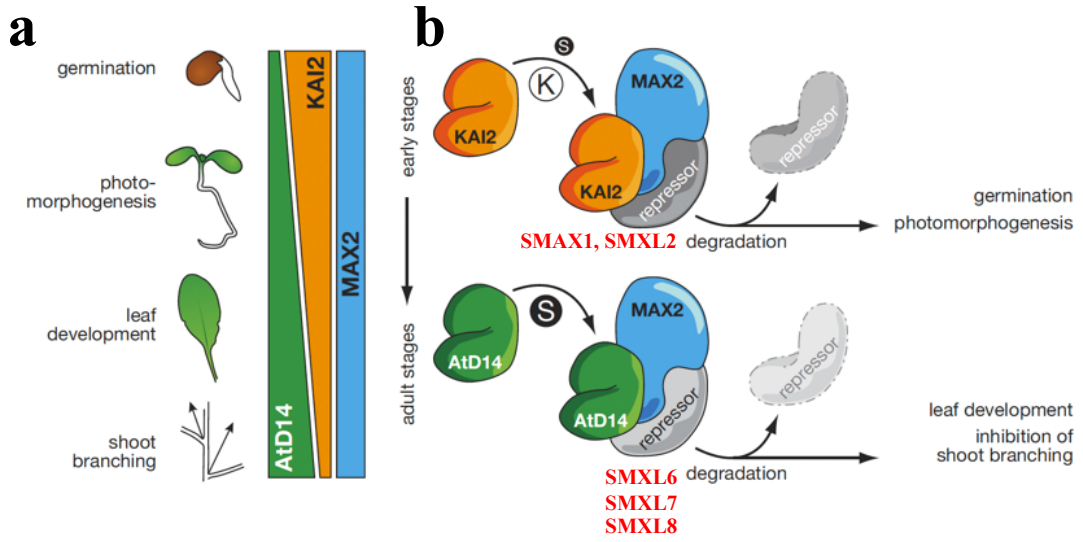


Figure 1-10: Proposed model for butenolide signal discrimination by KAI2 and D14 during plant development.

a, Different phenotype response of KAI2 and D14, which is probably caused by the relative expression level of D14 and KAI2 at different developmental stages. At early stage, KAI2 and karrikin-related signalling play a dominant role for seed germination and seedling establishment, but in later vegetative development, D14 and strigolactone signalling prevail. **b,** Signal transduction of karrikin by KAI2 resembles that of SL by D14 with the MAX2 as the key modulator that both shared. It is postulated that the ligand binding induces a conformational change in their receptor protein, which will extend a interface to interact with MAX2 proteins, later recruiting and polyubiquiting repressor proteins for degradation.

Reprinted from "Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in Arabidopsis. Waters, Nelson, et al. Development, 2012, 139:1285-1295.

Intriguingly, the structural similarities of these two classes of chemicals are reflected by a common signaling transduction process (Figure 1-10). Firstly, the genetic screen of KAR-insensitive mutants leads to the identification of two loss-of-function alleles of MORE AXILLARY GROWTH2 (MAX2, orthologous to DWARF3 in rice), which, however, have been already implicated in regulation of SLs response[92, 98]. Another element in Arabidopsis, HYPOSENSITIVE TO LIGHT/KARRIKIN INSENSITIVE 2 (AtHTL/KAI2), was later identified as a crucial role in KARs-regulated signaling pathway[99]. KAI2, as well as its paralog DWARF14 (D14)/DECREASED APICAL DOMINANCE (DAD2)[100], are all classified to the RsbQ-like family of ABH fold. The X-

ray crystallography studies obviously illustrated that KAI2 possess a canonical structure architecture of ABH fold that contains a seven-stranded β-sheet flanked by five helices at one side and two at the other, and a double layer V-shaped helical fold comprised of four helices lie on top as a gatekeeper of substrate-binding pocket where the catalytic triad residues Serine-Histidine-Aspartic acid was located[57, 101]. Resembling mechanism of action of SLs' receptor D14[100], KAI2 act as a receptor to perceive the ligand KARs directly, and was thought to associate with F-box protein MAX2 to regulate its activity (Figure 1-10)[99, 101]. The third key proteins SUPPRESSOR-OF-MAX2-1/(SMAX1)-LIKE (SMXL) family was specifically targeted by activated MAX2 of SCF-type (SKP1-CULLIN-F-box) E3 ubiquitin ligase complex for ubiquitination and proteasomal degradation (Figure 1-10)[102].

1.4 Chemical regulation of plant hormone function

Working under the Darwinian assumption of biological evolution that all species of organisms have to raise and develop inherited variations that increase the individual's ability to compete, survive, and reproduce as the countermeasures against environmental stresses of biotic and abiotic, one is able to imagine that plant hormone, the collection of small molecular "substances" derived from various metabolic pathways, would be the inevitable heritage of plants evolution that allow themself to get survival in facing the internal or external challenges[103, 104]. It's now well acknowledged that all aspects of plant growth and development are regulated by these endogenous compounds, and thanks to genetic approaches, mainly in *Arabidopsis thaliana*, the biology of various plant hormones are going to be much clear, especially the sparkling findings of various crucial elements in signaling transduction (as stated above). However, some limitations consistently stay in genetic studies. For example, gene duplication is sometimes required for evolution and adaption, and resultant redundant genes always have additional functions that remain unknown. Therefore, it is not effortless to overcome gene redundancy. And more since direct knock-out may result in lethal or have severe defects in growth and development, to know the function of the genes can be challenges.

Chemical biology is a relatively new field that just recently emerged, but we can still trace the history back more than 100 years ago when scientists used chemical pigments as probes to contrast

and visualize biological samples for the first time[105]. *Nature Chemical Biology* defines chemical biology as both the use of chemistry to advance a molecular understanding of biology and the harnessing of biology to advance chemistry[105, 106]. Simply defined, chemical biology is to use drug-like chemicals to answer biological questions at molecular level. Despite the fact that the use of small molecules to perturb biological systems has a rich history, its great potential and broad applicability in the field of plant science have just been recognized and gained increasing attention[107, 108]. Through the application of small molecules in living systems, spatial-temporal, reversible, and conditional manipulation of a phenotype by over-riding the intrinsic drawbacks of classical genetic approaches, including genetic redundancy, lethality, and pleiotropism can be possible[109]. And what matters is that researchers could have convenient access to publicly available small-molecule libraries with tremendous chemical diversity to advance their works in biology. These small molecules can sometime provide additional benefits, e.g., optical control or emission of fluorescence, which promote precise control or dynamic visualization of biological processes[110]. What's more, advanced by biology-related or chemistry-related disciplines, including structural biology, molecular biology, and organic chemistry, the use of chemical probes could offer unique but in-depth insight into the biological processes, such as the genetics-aided elucidation of the mode of action of small molecules of interest (especially the identification of the candidate targets)[111, 112] and structure-based rational design of novel and more potent agrochemicals for the basic research use or even practical application[113-115].

On the other hand, in a recent decade, chemical genetics, which can be regarded as a subset of chemical biology where chemical probes are used in combination with genetic approaches to perturb biological systems as stated above[116], has been broadly applied to investigate the basis of macromolecule complex in the highly dynamic processes to strengthen our understanding of hormone biology, especially in aspect of dissecting the signaling transduction of plant hormone[117]. The striking achievements by using the small-molecule toolbox of strigolactones include having a better understanding of the interaction and communication of plants with parasitic weeds[118], identifying the ShHTL7 from 11 homologs of *Striga* (ShHTL1-11) as the most potent and dominant receptor that senses the SLs from the soil as stimuli for germination[119], combining live imaging to fluorescently probe regulatory dynamics of strigolactone signal transduction in *Striga*[120], and taking global and structural grasp of the overview of macromolecular machinery of D14-D3-ASK1[121, 122].

Most importantly, aided by the chemical biology, strigolactones research in the parasitic plants offers unique perspectives to explore and possibly fix the problem of agricultural destruction caused by parasitic weeds that have caused the loss of $10 billion U.S. dollars' worth of crops every year[123].

1.5 Research objectives and significance

As stated in section 1.3, the receptor of GAs and KARs, GID1 and KAI2, respectively, are all belong to ABH fold superfamily and involved in regulating diverse aspects of plant growth and development, therefore chemical manipulation of their function is of tremendous significance to boost our agricultural system.

1.5.1 Development of orthogonal AtGID1a variant-AC94377 pairs

GAs have the great potential to promote plant growth and shape plant architecture, such as induction of seedless grapes, enlargement of fruit size, which stand its far-reaching roles for practical use in agriculture and horticulture[124, 125]. However, the application of GA was restrained by several factors. For example, the commercial production of GAs, mainly through the chemical synthesis and fungal fermentation, are at high cost; complicated chemical structure shows poor stability; and most fundamentally, overtreatment of GA causes some unfavorable changes in plant development, resulting in plants lodging[126]. The 1st generation of 'Green Revolution' was actually achieved by using the semi-dwarf varieties of wheat Rht mutant insensitive to gibberellin and rice IR8 mutant deficient in gibberellin biosynthesis in the 1960s, which show resistance against lodging[127]. In this context, there is an increasing demand for developing GA mimics with desired functions for basic research use and practical application.

Moreover, caused by gene redundancy, the clear delineation of specific GA response still remains a major challenge. For example, Arabidopsis genome has three genes coded as GA receptors and five as DELLAs, even with slightly overlapping functions, the various combination of receptors and repressors would result in different responses from the molecular level to phenotype. As a rapidly emerged strategy attractive in plant research, the development of an orthogonal ligand-protein pair that employs the collaborative approach between chemistry and biology would be favorable to comprehensively grasp the biology of small molecular regulators, including plant hormones

themself. This strategy has been successfully applied to control plant water use through hijacking abscisic acid signaling using a variant of ABA receptor and a nanomolar-level sensitive agrochemical[128], or to modulate auxin signaling by both engineering of auxin and its receptor[129]. Therefore, it is expected to generate a 'functionally' orthogonal high-affinity ligand-protein variant pair as a controllable switch that flexibly regulates plant growth through the GA signaling pathway, as well as dissecting the biological function of individual GID1 protein without interfering with the endogenous GA (Figure 1-11).

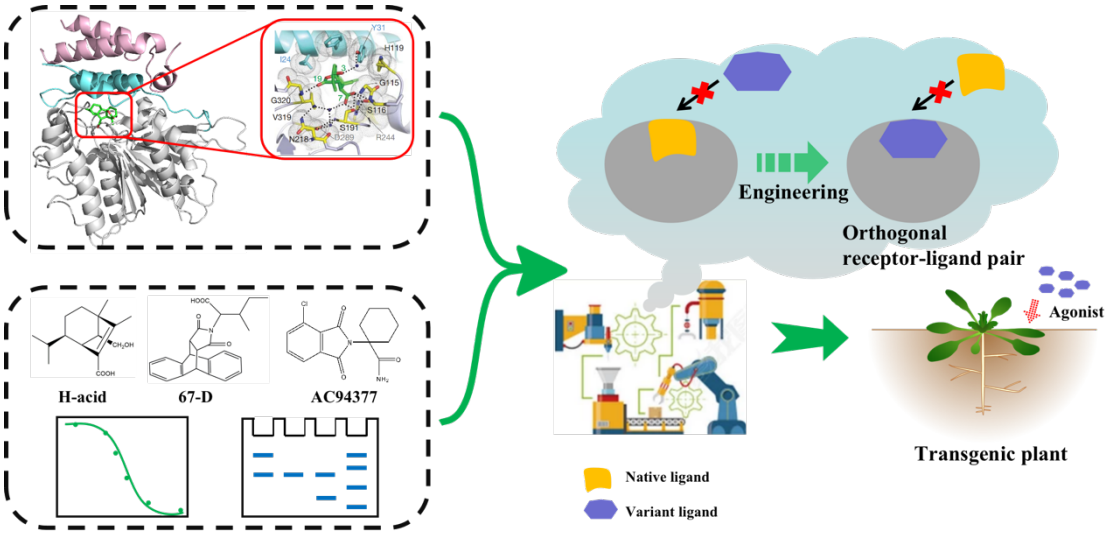


Figure 1-11: Proposed workflow for the generation of an orthogonal high-affinity ligand-protein variant pair to hijack GA signaling.
The structural inspection and biochemical characterization of protein complex will be the first step for the design of orthogonal ligand-protein variant pair. The chemicals considered for using in this study are the agonists of GA previously reported. The biochemical results obtained will inspire to perform protein engineering to get a GID1 variant that lost the affinity to native ligand GAs but shows higher activity to the synthetic small molecules, like AC94377. Finally, the successfully constructed orthogonal ligand-protein pair will be applied to make the transgenic plant to verify its practicability.

In the previous works of our laboratory, we have found that chemical AC94377 shows GA agonistic activity of GA in Arabidopsis and rice, such as induction of seed germination and hypocotyl growth, and increase plant height[130]. Moreover, AC94377 could bind to GA receptor GID1 to promote the interaction of GID1 with DELLA and subsequentially induce the degradation of DELLA through 26S proteasome. However, AC94377 only displays less activity than GA.

In this study, for the one hand, I will focus on unveiling the interaction of AC94377-GID1 complex at atomic level by using X-ray crystallography or docking simulation, making attempts to provide structural information for our future research, including the design of a more potent GA agonist based on AC94377 skeleton. On the other hand, in order to generate an orthogonal high-affinity ligand-protein variant pair to hijack GA signaling by applying AC94377, I will perform the chemical AC94377-driven directed evolution of GID1 to improve its ligand sensitivity and selectivity from native plant hormone GA to artificially synthesized chemical AC94377 or its derivatives. It is expected that this orthogonal system has great potentiality to dig out more clues of GA response pathways undiscovered through bypassing the complexity of genetic redundancies.

1.5.2 Structural basis for the perception of a Karrikin antagonist KK181N1 by receptor AtKAI2

Given the fact that the endogenous native ligand of KAI2 still remains elusive, KARs, especially KAR_1 and KAR_2, are universally used as agonists and show various physiological functions by activating receptors KAI2[131]. Even though KARs and SLs have a common butenolide moiety in their chemical structures and share a key component MAX2/D3 in their signaling transduction, they are first perceived by two receptor proteins, KAI2 and D14, respectively, so that the response could be transduced and initiated by releasing the repression of KARs or SLs signaling mediated by SMAX1/SMXLs family members. Nowadays SMAX1/SMXLs family members that are specifically involved in each of KARs or SLs response were defined[132-135]. In *Arabidopsis*, the activated KAI2 trigger proteolysis of two members of SMAX1/SMXLs family, SMAX1/SMXL2, to mediate plant growth and development (e.g., seed germination, seedling photomorphogenesis, root morphology)[136-138]; and the destruction of three members, SMXL6/7/8, are required for regulating strigolactone-initiated phenotypes (e.g., shoot architecture, leaf morphology)[139, 140]. Nevertheless, more recent studies have found some overlaps occurred in the two chemicals-triggered disruption of certain SMAX1/SMXLs proteins and questioned this dichotomy[135, 141, 142]. For example, not only SMAX1/SMXL2 but also SMXL6,7,8 are required for proteasomal degradation in regulating root skewing through a KAI2-dependent and D14-independent manner[141]; and both SLs and KARs trigger polyubiquitin-mediated degradation of SMXL2 to regulate

hypocotyl elongation[142]. In addition, the confusion of functional overlaps of KAI2 and D14 are exacerbated to a certain extent by using synthetic SL analog GR24 in many studies. Experimentally applied GR24 is actually a racemic mixture of two enantiomers (*rac*-GR24), 2'R-configured butenolide of which could sufficiently activate D14, whereas the non-naturally produced 2'S-configured component could elicit KAI2 signaling as that activated by karrikin (Figure 1-9)[143]. This would cause huge misunderstanding in delineating physiological functions of SLs and KL. Thus, efforts are still in demand to develop a highly selective chemical probe that discriminates D14 and KAI2 response, which will also contribute to unraveling the convergence of these two receptors in the aspect of evolution.

Preserving canonical α/β-hydrolase fold with intact arrangement of catalytic triads Ser-His-Asp for ligand perception, KAI2 are potentially targeted by small molecules that are artificially synthesized to mediate its sophisticated signaling transduction or perturb biological function[42, 101]. This could be preferably accomplished by covalent modification of protein, since the serine nucleophile within catalytic triads of KAI2 would be highly effective to extensive electrophiles, including 1,2,3-triazole-urea probes that have been manifested to carbamoylate the activated serine residue in a covalent manner (reviewed by Alexander Shannon, 2015) with high selectivity towards serine hydrolases family[17, 30, 144].

Our laboratory has previously designed and synthesized a small chemical library based on chemical scaffold of 1,2,3-triazole-urea, aiming to target α/β hydrolase-type macromolecules of plants with tunable selectivity. Recently, from this in-house chemical library, phenotype-based screening has enabled us to find a compound, named KK181N1, exhibiting opposite effect on KAR-initiated inhibition of hypocotyl growth. Our LC-MS/MS results stand an unexpected binding mode of KK181N1 with AtKAI2 that KK181N1 was hydrolyzed and covalently attach to serine214, but not the nucleophile residue serine95.

In this study, I applied X-ray crystallography to get insight into the interaction of KK181N1 with AtKAI2, especially to confirm the location where the hydrolytic product of KK181N1 covalently bound. And later, the biochemical characterization of KK181N1 was performed to study its structure and activity relationship. Lastly, as an antagonist of KAI2, KK181N1 will be used to investigate its mode of action with partner proteins downstream, such as MAX2/D3 and SMAX1. Importantly, our result will not only enlighten us to perform structure-guided optimization of KK181N1 to generate

a more robust inhibitor but also encourage us to apply this antagonist in plants for basic research, such as for the exploration of the native ligand of KAI2 and the solo establishment of SL function that bypass KAI2 signaling.

Chapter 2: Development of orthogonal AtGID1a variant-AC94377 pairs

2.1 Research Summary

AC94377, a substituted phthalimide, has been revealed to mimic the biological functions of GAs through activating the GA receptor GID1s in various plants. Despite that its mode of action is clear, the accurate binding mode of AC94377 with GID1s still remains unknown. GID1 engages in regulating diverse aspects of plant growth and development, but delineation of the precise biological events of a specific GID1 is of great challenge, especially in Arabidopsis because the pleiotropic effects of various receptor-ligand-repressor combinations (AtGID1s-GAs/AC94377-DELLA) lead to the complicated regulatory networks that difficult to be artificially and specifically manipulated. With these in mind, I firstly made attempts to uncover the mode of interaction between GID1 and AC94377 at atomic level, with the expectation of developing more potent GA regulators based on the chemical structure of AC94377 skeleton. Secondly, I tried to design orthogonal GID1^{variant}-AC94377 pairs for probing specific GA response.

In this study, I firstly applied X-ray crystallography to get insight into the binding mode of AC94377 in GID1. However, even though we had paid substantial efforts on crystal screening and optimization of GID1-AC94377 complex, I was unable to get quality-sufficient crystals for X-ray diffraction. Therefore, to understand the perception mechanism of AC94377 by GID1 protein, I performed the structure-activity relationship studies on AC94377 by using a series of its derivatives containing different substituent on chloro and carboxamide groups of AC94377. Combining the results of *in vivo* (Y2H and physiological assessment) and *in vitro* (ITF and DSF) assays, I realized that both carboxamide and chloro groups are fundamentally important to retain the high GA-like activity for AC94377. Moreover, I found that AC94377 also favors the Ile-subtype OsGID1 (V319I), which is particularly consistent with our previous work that AC94377 exerts GA-like activities with a preference for Arabidopsis GA-receptors AtGID1c and AtGID1b that carry Isoleucine 317 (Ile-subtype) rather than AtGID1a that carries Valine 319 (Val-subtype). In order to support the results from SAR studies, I performed the molecular docking simulation of AC94377 with OsGID1 and AtGID1a. Superposition of simulated structures of both OsGID1-AC94377 and AtGID1a-AC94377 complexes shows good agreement in the binding pose of AC94377 in the pocket of GID1s. Notably, based on this simulation, I thought that the reinforced hydrophobic interaction between the aromatic plane of AC94377 and the extended alkyl group of

isoleucine (compared with valine) may account for the improved affinity of AC94377 for the Ile-subtype GID1. However, it is difficult to fully explain the results obtained by current simulating models because two AC94377 derivatives with the modifications of carboxamide group in AC94377 by changing amine into hydroxyl or isopropoxy (ACOH and ACOIp) did not show the preference to Ile-subtype GID1. Here, I proposed that the ligand-induced conformational changes may occur to accommodate the occupation of AC94377 in the pocket of GID1.

Since the absence of a precise structure complex of GID1-AC94377, I performed the AC94377-directed laboratory evolution on AtGID1a by Y2H-based screening system to generate the orthogonal ligand-protein variant pair of GID1-AC94377. To efficiently achieve this goal, I firstly adopted the combinatorial saturation mutagenesis (CSM) on AtGID1a by grouping the 24 pocket-lining residues into 9 sites/sublibraries. This has produced several beneficial AtGID1a variants from 3 sites with 10-folds higher AC94377-binding affinity than wild type. Next, I performed the iterative random mutagenesis (IRM) on these positive variants to maximize the cooperative effects of mutagenesis. Fortunately, these efforts have enabled to generate the AtGID1a variants with greatly enhanced affinity and selectivity to AC94377. Specifically, Y2H assay reveals that the strength of AC94377-induced PPI between these AtGID1a variants and GAI is increased about 16,000-folds than that between AtGID1a wild-type and GAI (at 50 μM AC94377), and the sensitivity of AtGID1a variants to AC94377 is increased 10-30 folds than that of AtGID1a wild-type to GA_4. To explore the direct binding event between ligands and variant receptors, I performed the ITF assay and DSF assay to evaluate the binding affinity. Notably, some of the AtGID1a variants tested show greatly improved binding affinity to AC94377 with about 500-folds higher than wild type, whereas they totally lost the GA_4-binding affinity. The DSF assay also supports these results that AC94377 induces the enhancement of thermostability for AtGID1a variants in the same manner as GA_4 did for wild type. However, GA_4 cannot increase the structural stability of AtGID1a variants. Interestingly, even without the treatment of ligand, thermostability of AtGID1a variants is also 10-12 ºC higher than that of AtGID1a wild type, suggesting the more stable interactions between these mutated residues.

My works on the establishment of orthogonal AtGID1a variant-AC94377 pairs would be served as a powerful tool to delineate specific GID1-mediated responses through bypassing the complexity of genetic redundancies, further advancing our understanding of GA biology.

2.2 Results and discussion

2.2.1 Crystallization trials of GID1-AC94377

In order to unveil the accurate binding mode of AC94377 in the active pocket of GID1, I have made abundant attempts on the crystallization of GID1-AC94377 complex. In the beginning, I have successfully generated the quality-sufficient crystals of AtGID1a-GA_4-AtGAI complex (Figure 2-1c) in a reproducible condition after 2 weeks[73], indicating that the protein complex could be well prepared by the consecutive purification steps, including Ni-NTA affinity chromatography, ion-exchange chromatography and gel filtration chromatography (Figure 2-1a, b).

Employing the same methods, I tried to produce the protein complex AtGID1a2- AC94377-AtGAI5. However, the expression efficiency of AtGID1a2- AC94377-AtGAI5 is dramatically decreased due to the formation of a large number of inclusion bodies. Even though a small amount of soluble portion is available, they are unstable and gradually turn into aggregates after purification. I think that this result is actually reasonable. As we know, GID1 protein contains a highly flexible N-terminal extension (N-Ex) that consists of three α-helices for its unique function involved in GA signaling perception and transduction[72, 73]. N-Ex contains serval hydrophobic residues for the non-polar interaction with the N-terminus of DELLA protein, as well as for recognition of the aliphatic moiety of GA_4. In an unbound status, these hydrophobic residues likely tend to aggregate in an aqueous solution. Only in the presence of gibberellin that has a high affinity to bind with GID1, could this N-Ex steadily cover the core of GID1. However, AC94377 did not have an affinity comparable to GA_4 or GA_3 from the results of physiological experiments. Based on this knowledge, in addition to increasing the culture volume of bacteria expressing the target protein, I also prepared a AtGID1a variant where four hydrophobic residues of N-Ex were replaced by hydrophilic one, including V7T, I10D, L18D, V29T. Even though the expression efficiency is slightly enhanced, it did not make much improvement for protein crystallization. It is noted that the microcrystal seeding and streak seeding of AtGID1a2-GA_4-AtGAI5 is of great success (Figure 2-1d), but that of AtGID1a2-AC94377-AtGAI5 using crystals AtGID1a2-GA_4-AtGAI5 as seeds only result in precipitates, which possibly indicates a large population of inhomogeneous protein complexes caused by the low affinity of AC94377.

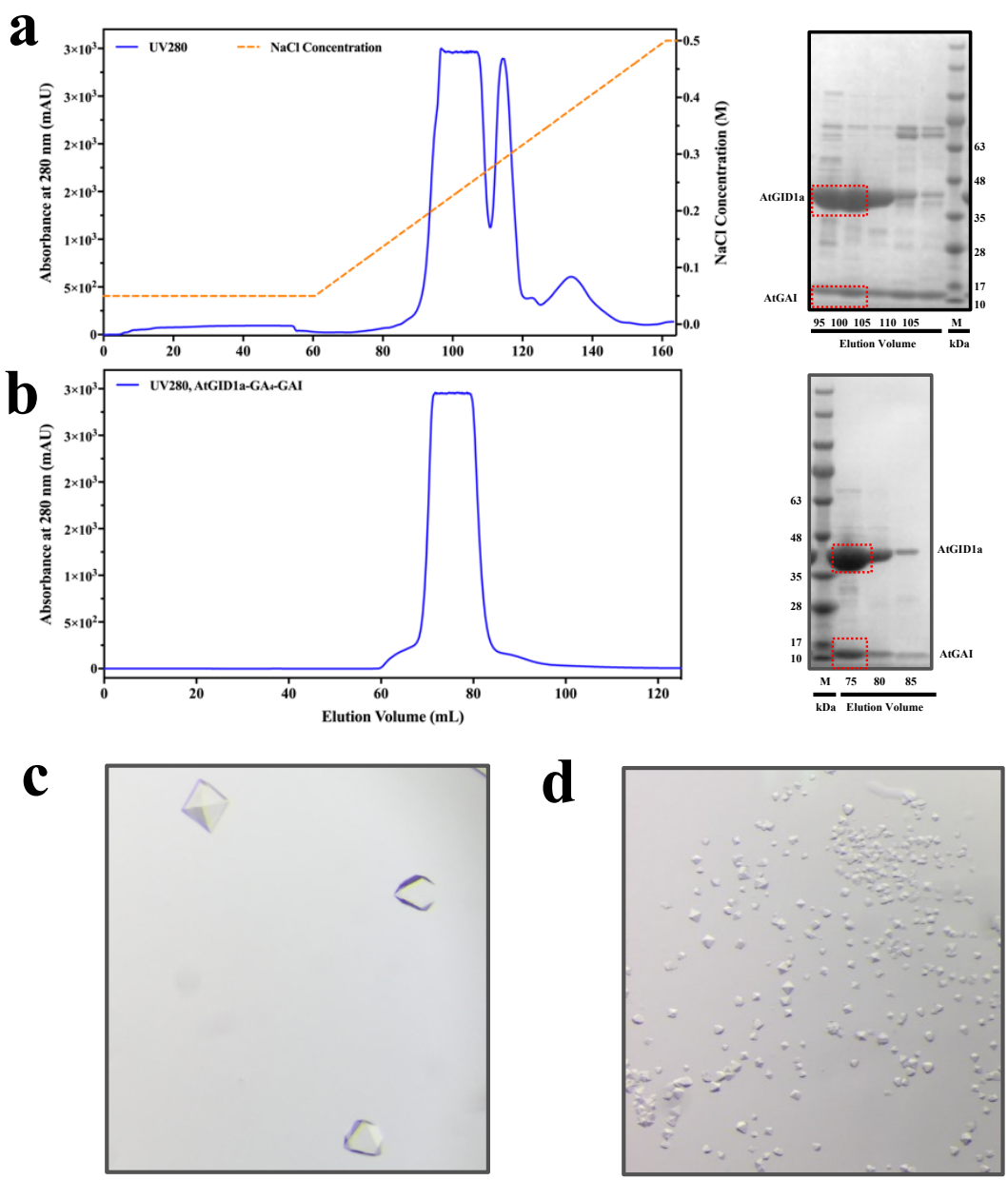


Figure 2-1: Purification and crystallization of AtGID1a-GA$_4$-AtGAI complex.

a, Anion-exchange chromatography of AtGID1a-GA$_4$-AtGAI complex (left panel). The fractions obtained were analyzed by SDS-PAGE (right panel). The red square-indicated fractions were concentrated for purification further. **b,** Gel filtration chromatography of AtGID1a-GA$_4$-AtGAI complex (left panel). The fractions obtained were analyzed by SDS-PAGE (right panel). The red square-indicated fractions were collected and concentrated into 50 mg/mL for crystallization. **c,** Crystals of AtGID1a-GA$_4$-AtGAI complex were obtained under the same condition as reported: 0.1 M Tris-HCl pH 8.0, 0.8 M LiCl, 32% PEG4000, 15 ℃. **c,** Crystals of AtGID1a2-GA$_4$-AtGAI5 were generated by the streak seeding into condition: 0.1 M Tris-HCl pH 8.0, 0.8 M LiCl, 32% PEG4000. The crystals obtained in **c** were used as the seeds.

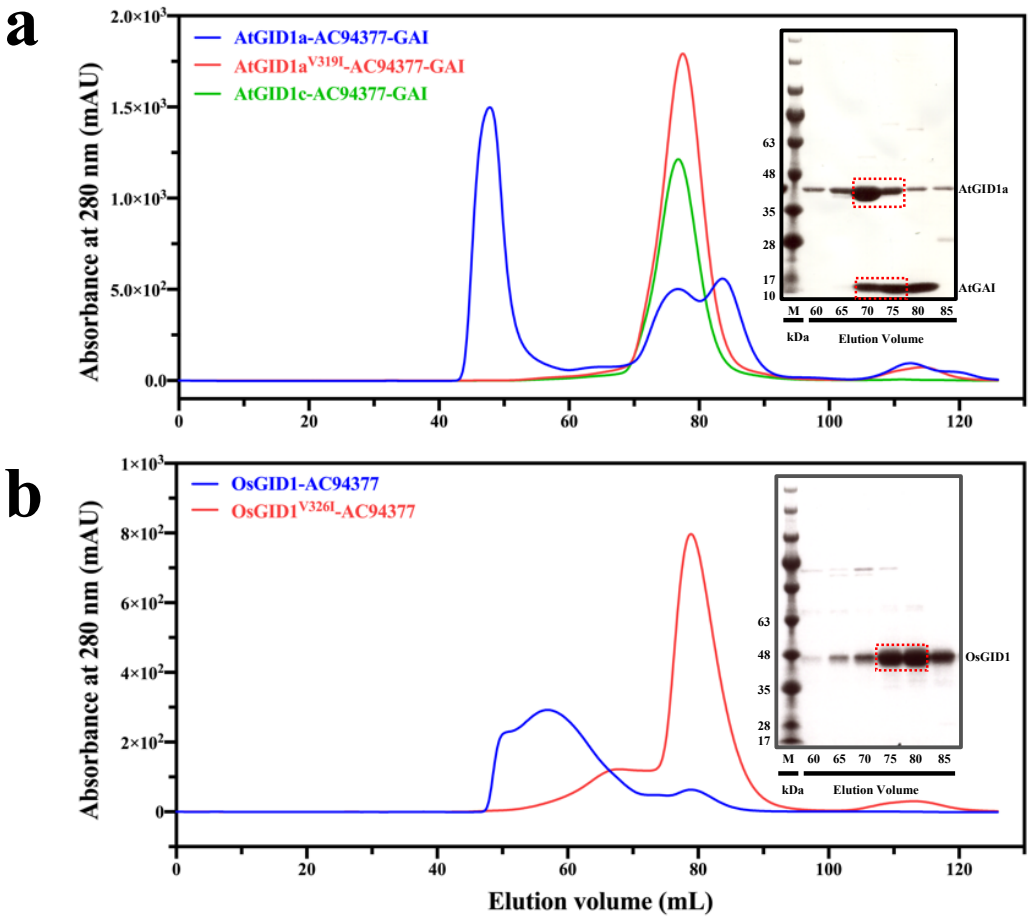


Figure 2-2: Purification of AtGID1a-AC94377-AtGAI (a) and OsGID1-AC94377 (b) complex by gel filtration chromatography.

The fractions obtained were analyzed by SDS-PAGE (right panel). The red square-indicated fractions were collected and concentrated into 50 mg/mL for crystallization. **a,** The blue, red, and green curves respectively indicated the chromatograms of AtGID1a-AC94377-AtGAI, AtGID1a^{V319I}-AC94377-AtGAI, and AtGID1c-AC94377-AtGAI complex. **b,** The blue and red curves respectively indicated the chromatograms of OsGID1-AC94377 and OsGID1^{V326I}-AC94377 complex. As indicated, the mutation V319I in AtGID1a and V326I in OsGID1 strongly increase the solubility of GID1 proteins when incubated with AC94377.

Next, I move to make various truncations on the N-terminus of GID1 and the GAI (Table 2-1) and extend our screening trials of crystallization, such as setting the different protein concentration and equilibrium temperature. On the other hand, inspired by our former work that AC94377 exerts GA-like activities with a preference for Arabidopsis GA-receptors AtGID1c and AtGID1b that carry Isoleucine 317 rather than AtGID1a that carries Valine 319[130], I prepared GID1 homologs

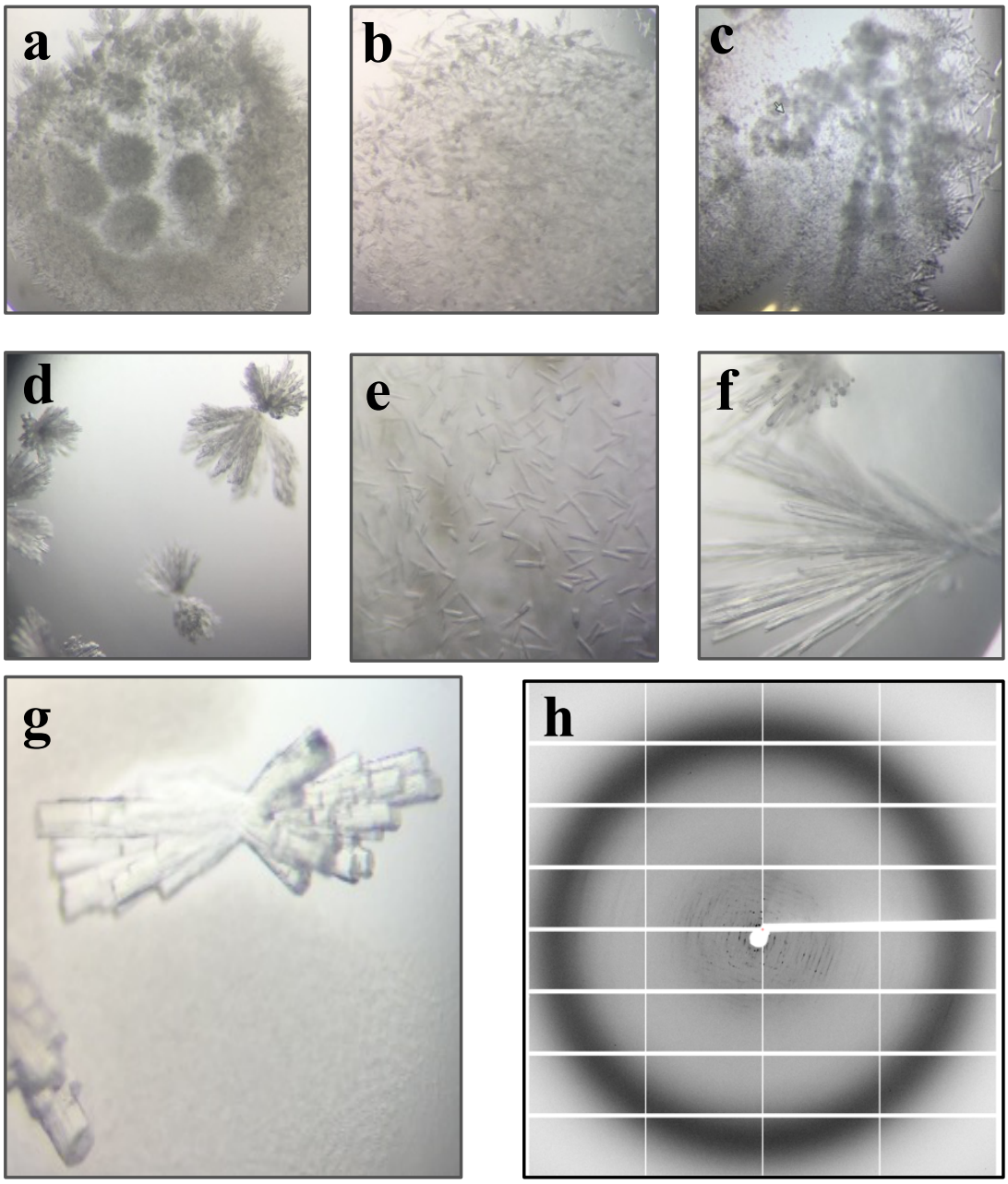


Figure 2-3: Crystallization and preliminary X-ray diffraction of AtGID1a^{V319I}-AC94377-AtGAI complex.

a-c, Co-crystallization of AtGID1a^{V319I}-AC94377-AtGAI complex under 4 °C. Reservoir conditions include JCSG I-1B (**a**, 0.2 M NaCl; 0.1 M HEPES pH 7.5; 20% PEG3000), PACT-9D (**b**, 0.2 M LiCl; 0.1 M Tris pH 8.0; 20% PEG6000), and PACT-5H (**c**, 0.2 M $NaNO_3$; 0.1 M Bis Tris propane pH 8.5; 20% PEG3350). **d-f**, Microcrystal seeding trials. Protein concentration is set around 30-50 mg/mL. Crystals were generated within 2 days. **g**, Crystals of AtGID1a^{V319I}-AC94377-AtGAI were generated by microcrystal seeding using the crystals (**b**) as seeds. **f,** Crystals (**g**) were used for the preliminary x-ray diffraction test. The resolution was revealed at higher than >10 Å.

and their corresponding variants at this site for crystallization (Figure 2-2a, b). I noted that low temperature (4 °C) is specifically favored for the generation of crystals of ternary complex GID1a2^{V319I}-AC94377-AtGAI5 and GID1c2-AC94377-AtGAI5 in serval reservoir solutions (Table 2-2), but the shape of these crystals is really strange and irregular, such as fibrous crystals, whiskers, needle clusters (Figure 2-3a, b, c). Considering the poor quality and massive amounts of small crystals, we thought that the growth rate may be too fast under these conditions. Therefore, the subsequent efforts were paid on the optimization of crystals obtained, especially low down the growth rate. The strategies include a use of decreased concentration gradient of protein, salt, and PEG, a use of the pH gradient, an additives screening, and a test for the microcrystal seeding under those conditions that remain clear. Microcrystal seeding in some conditions could effectively decrease the number of crystals but still make no difference to improve their quality (Figure 2-3d, e, f). The seeding-generated crystals were used for X-ray diffraction test in SPRING-8 (Figure 2-3g), whereas the low resolution (>10 Å) obtained was insufficient for data analysis (Figure 2-3h). The parallel efforts were also taken for the crystallization of OsGID1^{V326I} with AC94377, however, we could not find any good protein crystals, but a large number of salt crystals produced in several conditions (Figure 2-4), as confirmed by SDS-PAGE.

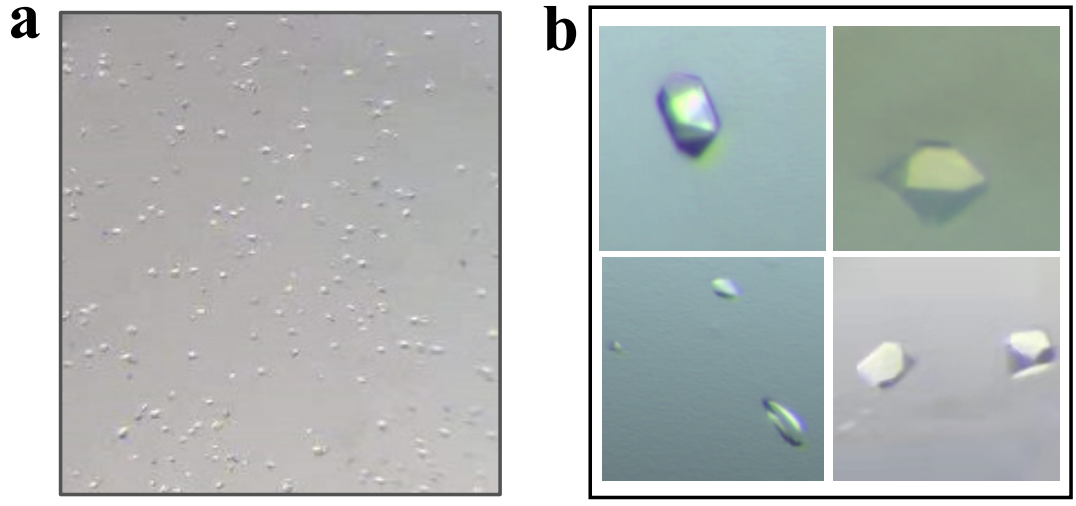


Figure 2-4: Crystallization of OsGID1^{V326I}-AC94377 complex.
a-b, Salt crystals were observed at several conditions of commercial kits (JCSG PACT, I, II).

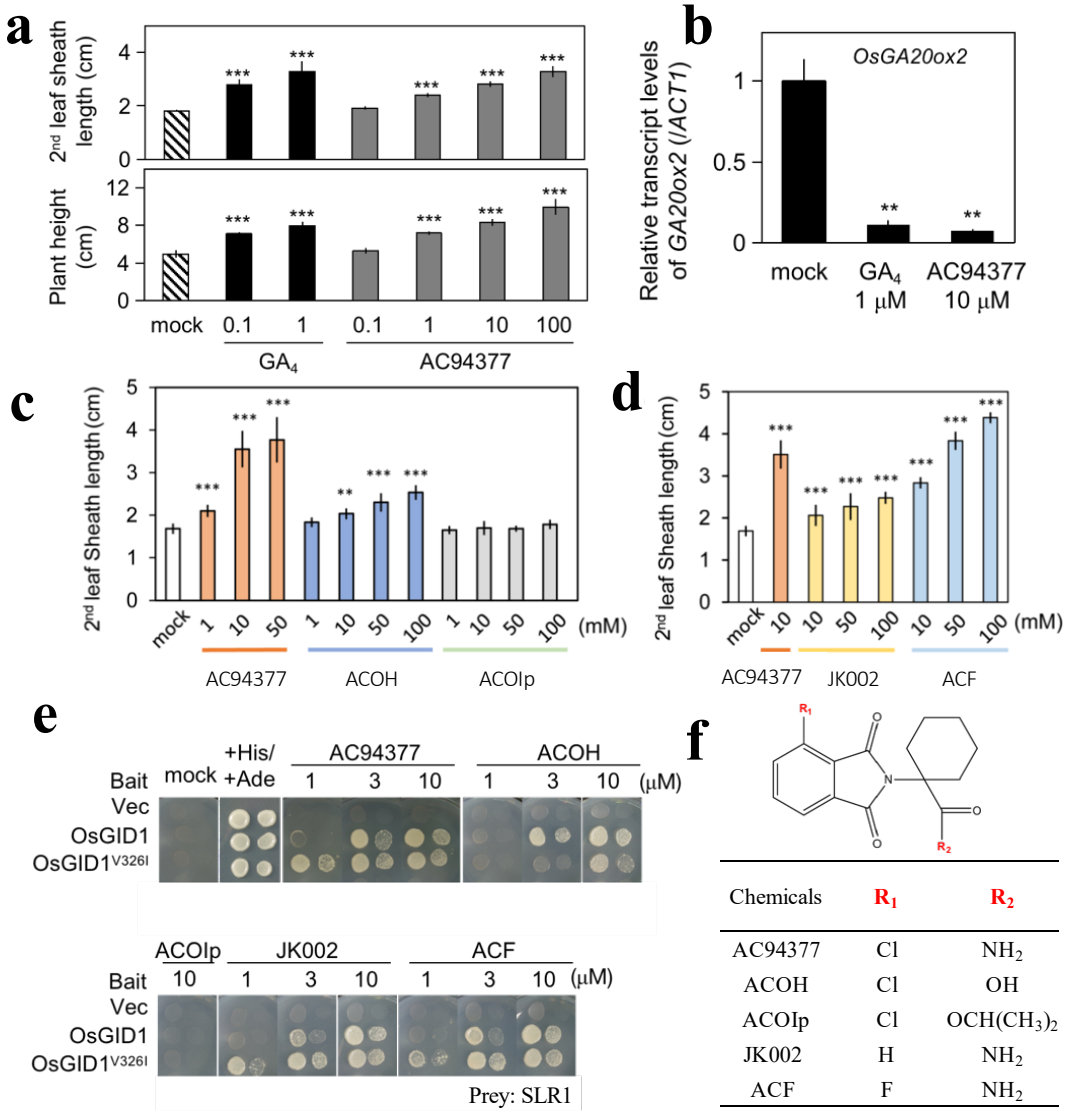
2.2.2 Structure-activity relationship studies of AC94377 and its derivatives

Despite in the absence of structural information of protein complex, I continue to make attempts to uncover the perception mechanism of AC94377 by GID1 protein using a series of AC94377 derivatives that are synthesized under the Briggs' "Rule of Three" for agrochemicals. Compounds ACOH and ACOIp were prepared by the modifications of carboxamide in AC94377 by changing amino group into hydroxy or isopropoxy group, respectively. The substitution of chloro group on the phenyl ring of AC94377 with hydrogen or fluoro group results in chemicals JK002 and ACF, respectively (Figure 2-5f).

All these chemicals were designed and synthesized by Dr. Jiang Kai in our laboratory. He also assessed their biological activity in rice, as well as the ability inducting PPI interaction of OsGID1 with SLR1 in yeast two-hybrid assay. In summary, AC94377 promoted the elongation of 2nd leaf sheath and total plant height in a concentration-dependent manner, as GA_4 did. And the transcript level of *GA20ox2* (GA20-oxidase, an enzyme that is considered as the negative marker in GA signaling) in rice was suppressed by AC94377 in the same manner as GA_4 (Figure 2-5a, b). As for AC94377 derivatives, ACF possessed comparable activity with AC94377 on promoting the 2nd leaf sheath elongation, ACOH and JK002 displayed decreased activity, whereas ACOIp completely lost GA-like activity on the 2nd leaf sheath elongation (Figure 2-5c, d). Similarly, the results in the yeast two-hybrid assay suggested that AC94377 and its derivatives, but not for ACOIp, could promote the interaction of OsGID1 with SLR1. Especially, AC94377, JK002 and ACF significantly promoted the interaction of SLR1 with OsGID1^{V326I} even at low concentration (1 μM) (Figure 2-5e). Taken together, the investigation on the bioactivity of AC94377 *in vivo* stands its great potential for the development of more potent GA-mimics.

To have a clear biochemical characterization of OsGID1 with AC94377 and its derivatives, I carried on measuring the ligand-induced changes in the thermostability of OsGID1 by differential scanning fluorimetry assay (DSF) and also quantify the binding affinity of all compounds by intrinsic tryptophan fluorescence assay (ITF). As observed in the figures, the thermostability of OsGID1 was greatly enhanced by the native ligand GA_4 in a concentration-dependent manner (Figure 2-6). While AC94377, ACOH and ACF only promote a slight increasement in the thermostability of OsGID1, even at a high concentration (300 μM). Compounds ACOIp and JK002 can barely stabilize OsGID1, which is consistent with our physiological assay mentioned

above (Figure 2-5; Figure 2-6; Table 2-3).

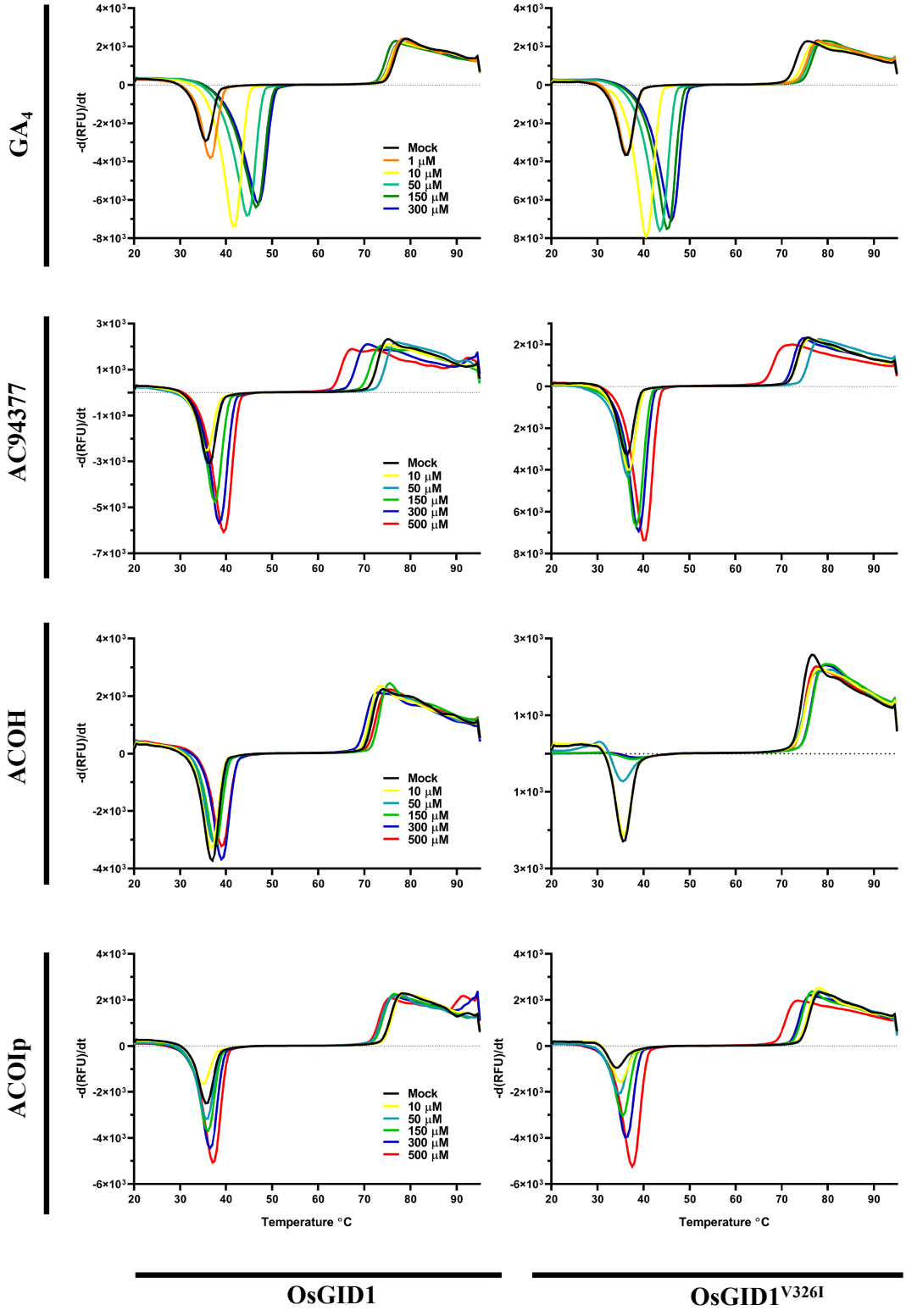


Chemicals	R_1	R_2
AC94377	Cl	NH_2
ACOH	Cl	OH
ACOIp	Cl	$OCH(CH_3)_2$
JK002	H	NH_2
ACF	F	NH_2

Figure 2-5: Investigation of biological activity of AC94377 in rice.

a, Micro-drop assay of semi-dwarf rice cultivar, Tan-ginbozu, in response to GA_4 and AC94377. Error bars indicate ± SD, n ≥ 6. **P < 0.01 (two-tailed Student's t-test). The mock treatment and test compound treatments (GA_4 or AC94377) were compared. **b,** Transcript levels of *OsGA20ox2* relative to those of OsACT1 in 12-day-old Tan-ginbozu seedlings treated with the test compound (or not). Error bars indicate ± SD, n = 3. **P < 0.01 (two-tailed Student's t-test). **c-d,** Micro-drop assay of Tan-ginbozu, in response to treatments with AC94377 and its derivatives. *P < 0.05, **P < 0.01, ***P < 0.001 (two-tailed Student's t-test). **e,** Yeast two-hybrid assay of the interactions between OsGID1/OsGID1^{V326I} and SLR1 in response to treatments with AC94377 and its derivatives. **f,** Chemical structures of AC94377 and its derivatives.

These results (unpublished) are supported by Dr. JIANG Kai (Southern University of Science and Technology).


GA4
AC94377
ACOH
ACOIp
-d(RFU)/dt
Mock
1 μM
10 μM
50 μM
150 μM
300 μM
500 μM
Temperature °C
OsGID1
OsGID1V326I

(Continued)

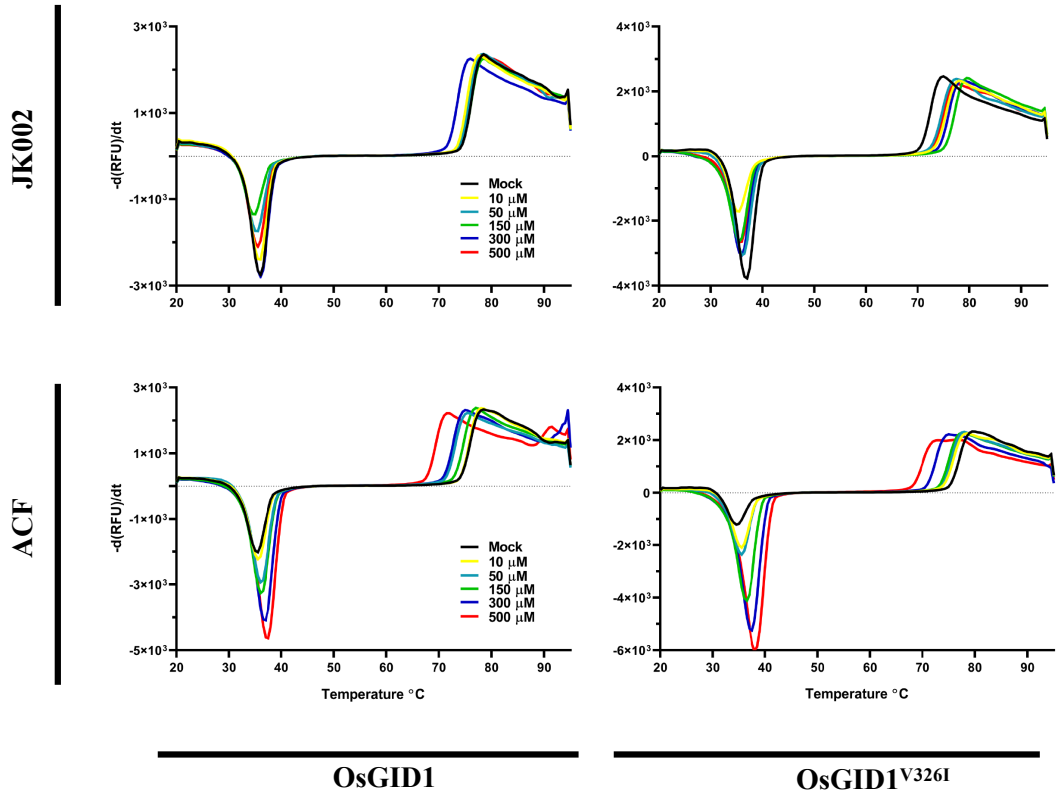
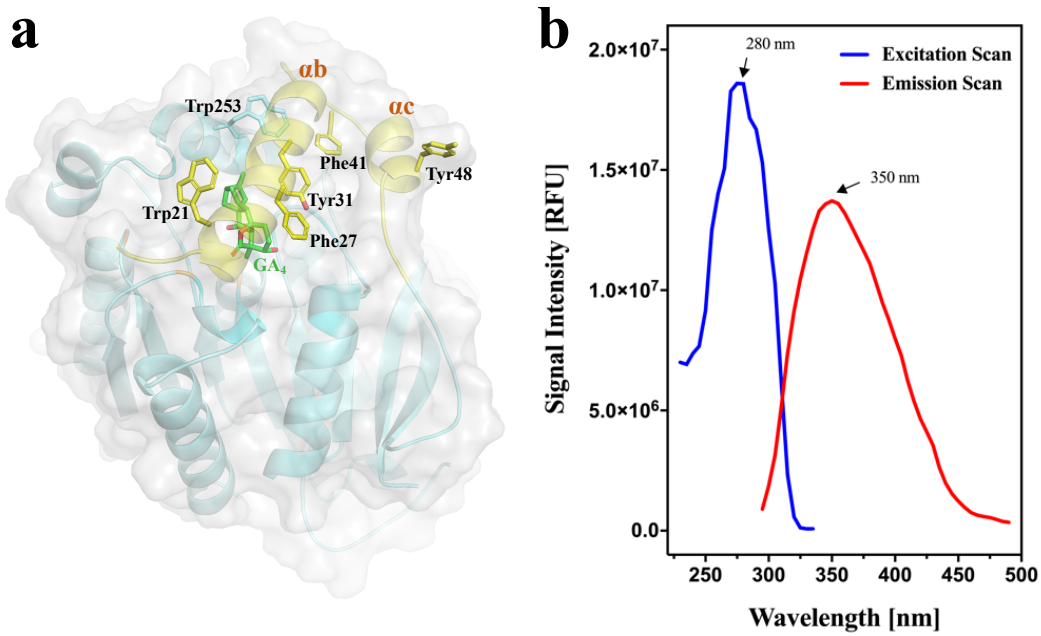


Figure 2-6: Differential scanning fluorometry assay of OsGID1 and OsGID1^{V326I} in response to GA_4, AC94377 and its derivatives.
The concentration of ligand varies: GA_4 (0-300 μM), AC94377 and its derivatives (0-500 μM). The negative first derivative of the relative fluorescence units with respect to temperature (-d(RFU)/dt) was calculated by Precision Melt Analysis^{TM} software (Biorad), and further plotted against the temperature in GraphPad Prism 5.0. The representative graphs are shown. n=3.

Next, I adopt the intrinsic tryptophan fluorescence assay to directly visualize the binding ability of chemicals to OsGID1. ITF is an uncomplicated tool for measuring the binding ability and efficiency between protein and ligand by monitoring the fluorescence changes of tryptophan in response to ligand[145]. OsGID1 has five tryptophans and two of them (Trp21 and Trp253), at least, may be sensitive to ligand binding since that Trp21 resides in the lid domain with conformational changes, and Trp253 interacts with the residues (Arg35, Arg36, and Gly39) in the lid domain. In addition, there exist other four aromatic residues in the cap domain (two tyrosines: Tyr31 and Tyr48; two phenylalanines: Phe27 and Phe41) (Figure 2-7a, b). Even though tryptophan dominates

over tyrosine and phenylalanine as a relatively strong fluorescence emitter due to the higher molar absorption coefficient, higher quantum yield, and the possibility of efficient energy transfer mechanisms, tyrosine and phenylalanine may still make more or less contribution to the intrinsic fluorescence change upon ligand binding[146-148] (Figure 2-7c). As show in Figure 2-8a, b and Table 2-3, the titration of OsGID1 with GA_4 or AC94377 resulted in fluorescence quenching in a ligand concentration-dependent manner, indicating that AC94377 is capable of binding to OsGID1 in the same manner as natural ligand GA_4. The binding affinity (K_d) of GA_4 calculated from ITF is 4.73 μM, which is comparable to the K_d determined by isothermal titration calorimetry (ITC)[74, 149], albeit there are big differences from that estimated by surface plasmon resonance (SPR)[75]. The subsequent affinity assay of AC94377 indicates that it shows moderate affinity to OsGID1 (79.0 μM), but stronger to $OsGID1^{V326I}$ (25.2 μM), supporting our examination of AC94377's preference for Ile-subtype of GID1 in rice and Arabidopsis[130]. However, none of AC94377 derivatives exhibit improved affinity to OsGID1 and $OsGID1^{V326I}$ (Figure 2-8). The K_d value of ACOH, JK002, and ACF are all in the range of 200-300 μM, suggesting the weak binding affinity to OsGID1. The variant $OsGID1^{V326I}$ shows slightly higher affinity to JK002 and ACF (both have group substitution in chloro group of AC94377, Figure 2-5f), but not to ACOH and ACOIp (both have group substitution in carboxamide group of AC94377, Figure 2-5f). These results may indicate some unique structural arrangement between the carboxamide group of AC94377 and the side chain of V326 in OsGID1 that remains to be discovered. However, it is unexpected to find that contrary to Y2H assay, ACF shows dramatically decreased affinity to OsGID1 compared with AC94377 (Figure 2-8; Table 2-3).

Taken together, our structure-activity relationship studies on AC94377 and its derivatives demonstrate that both carboxamide and chloro groups are important to retain the GA-like activity in rice and the binding affinity to OsGID1 *in vitro*. Furthermore, as indicated in our previous study that AC94377 exerts GA-like activities with a preference for Arabidopsis GA-receptors AtGID1c and AtGID1b that carry Isoleucine 317 (Ile-subtype) rather than AtGID1a that carries Valine 319 (Val-subtype)[130], AC94377 also favors Ile-subtype of GID1 in rice ($OsGID1^{V326I}$).



c

Amino acid	Lifetime τ (ns)	Absorption λ (nm)	Absorption ε (M^{-1} cm^{-1})	Fluorescence λ (nm)	Fluorescence Φ	Fluorescence *B* at λ_{absmax}	Fluorescence *B* at 280 nm
Tryptophan	0.5, 3.0–8.7	279	5,579	348	0.13	725	722
Tyrosine	3.3–3.8	274	1,405	303	0.14	197	170
Phenylalanine	~7.5	257	195	282	0.024	5	0.11

Figure 2-7: The application of intrinsic tryptophan fluorescence assay for GID1 protein.

a, Overall structure of OsGID1 complex with GA_4 (PDB code: 3ebl). OsGID1 is represented as cartoon and surface model. The cap and core domains of OsGID1 are respectively colored yellow and cyan. GA_4 is indicated as green stick. The aromatic residues (Trp21, Phe27, Tyr31, Phe41, and Tyr48) located in cap domain are indicated as yellow sticks. Trp253 is indicated as cyan stick. **b,** The excitation and emission scan of OsGID1 protein. The highest excitation and emission wavelength were respectively fixed to 280 nm and 350 nm. **c,** Fluorescence lifetime τ, molar absorptivity ε, quantum yield Φ, and effective brightness B of the three aromatic residues in proteins. These values are retrieved from '*Patrik and Liu. Quantitative Prediction of Fluorescence Quantum Yields for Tryptophan in Proteins. The Journal of Physical Chemistry B. 2004, 108, 14, 4248-4259*' and '*Mocz and Ross. Fluorescence Techniques in Analysis of Protei-Ligand Interactions. Protein-ligand interactions. Humana Press, Totowa, NJ, 2013. 169-210*'. Individual data sets may differ to some extent, especially the quantum yield of tryptophan which varies from 0.01 to 0.35 depending on the environment and the temperature.

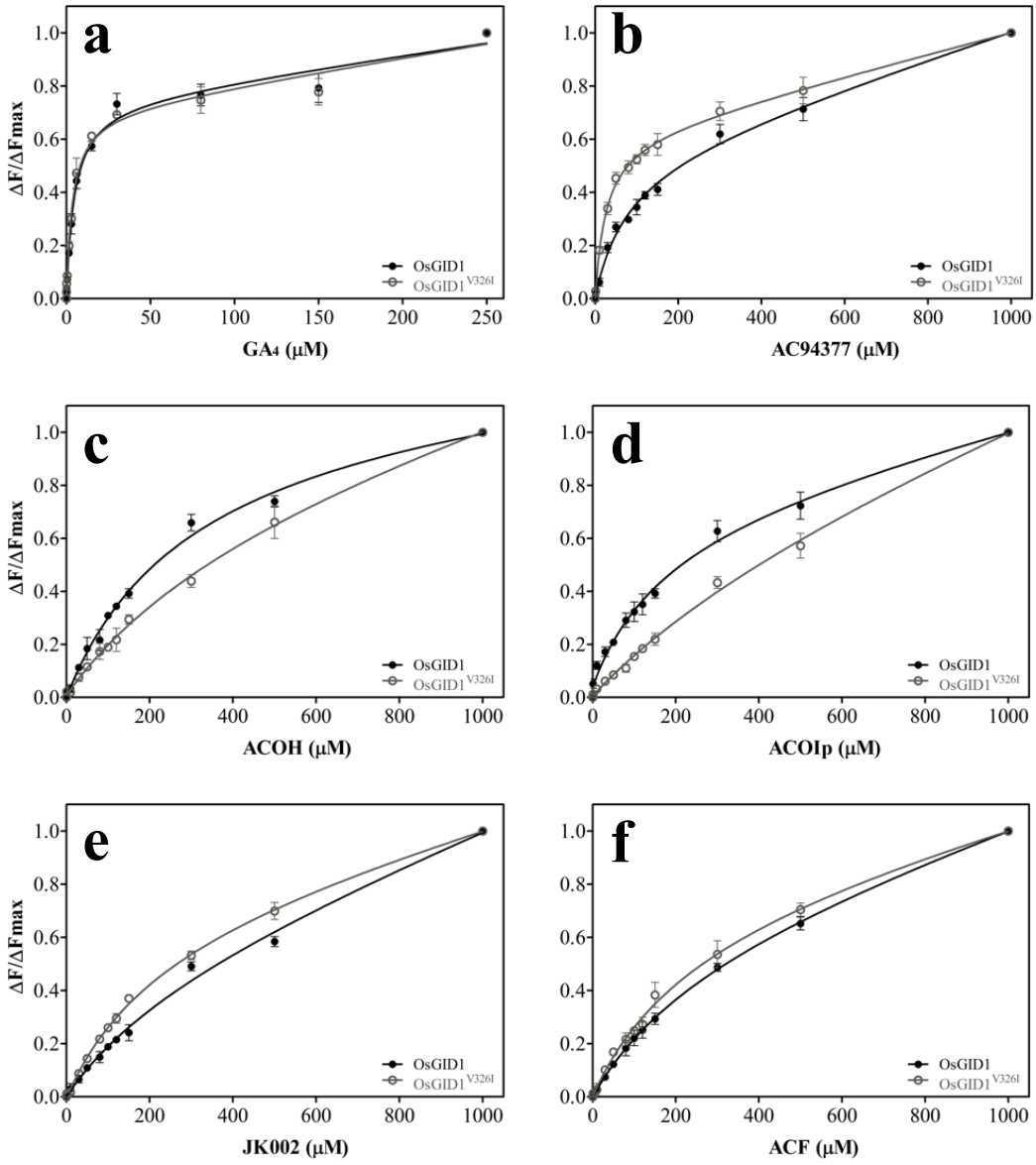


Figure 2-8: Intrinsic tryptophan fluorescence assay of AC94377 derivatives.
The fluorescence changes of OsGID1 or OsGID1^{V326} in response to the titration with chemicals: GA_4 (**a**), AC94377 (**b**), ACOH (**c**), ACOIp (**d**), JK002 (**e**), ACF (**f**). ΔF/ΔFmax is plotted *versus* ligand concentration by fitting to nonlinear regression with '*One Site-Nonspecific Binding*' model, finally giving K_d value. n=3.

2.2.3 Molecular docking simulation of GID1-AC94377

Our physiological assay and biochemical characterization suggested that AC94377 is a functional agonist of GID1, and particularly, it shows improved bioactivity to Ile-subtype of GID1 (AtGID1a^{V319I} in Arabidopsis or OsGID1^{V326I} in rice), which would require the structural information to explain the basis mentioned above. Since the failure of the crystallization trials in GID1-AC94377 as mentioned in section 2.2.1, we decided to perform the molecular docking simulation using the *apo* structure of OsGID1 as the receptor and energy-minimized AC94377 as the ligand, according to the protocol of AutoDock Vina[150]. In order to evaluate the accuracy of AutoDock Vina, I first applied GA_4 as the ligand that is docked into active site of ligand-free OsGID1. As we can see in Figure 2-9b, the overall orientation of GA_4 from docking simulation (binding free energy ΔG = -11.8 kcal/mol) is essentially superposed with that from crystal structure OsGID1-GA_4 (PDB: 3ebl) with negligible deviation on the C6-carboxyl group of GA_4, suggesting the reliability of this molecular docking procedure.

The molecular docking of AC94377 into OsGID1 was subsequently performed with the grid box fixed in the central pocket of OsGID1, deriving the best docked pose in the pocket of OsGID1 with binding free energy ΔG = -9.6 kcal/mol (Figure 2-9c). In this simulating structure complex, chlorobenzene ring of AC94377 directs into the bottom of pocket with the chloro group pointing to the hydrophilic interface constituted by the hydroxyl side chain of Y134 and Y329, the carboxy side chain of D197, and the amide backbone of G327, indicating the potentiality for the formation of halogen bond (Figure 2-9d). This could be the reason why compound JK002 with a deletion of chloro group on the phenyl ring in AC94377 shows less potent performance of GA-like activity than AC94377 (Figure 2-8e, Table 2-3).

On the other hand, the cyclohexane ring of AC94377 heads up to the N-extension cap domain of OsGID1 and involved in the hydrophobic and van der Waals interactions with several residues, including I24, F27, K28, Y31, V246, Y254 (Figure 2-9c, e). This structural pose of AC94377 can be partially verified by the observation that the orientation of cyclohexane ring of AC94377 is almost superposed to where the aromatic C/D-ring of GA_4 point (Figure 2-9f), suggesting that these non-polar residues should have preferences for hydrophobic moiety. Remarkably, this docking structure also present us the rich polar interactions between AC94377 and alkyl side chains of R251 and S123 (Figure 2-9g). One carbonyl group on the succinimide moiety of

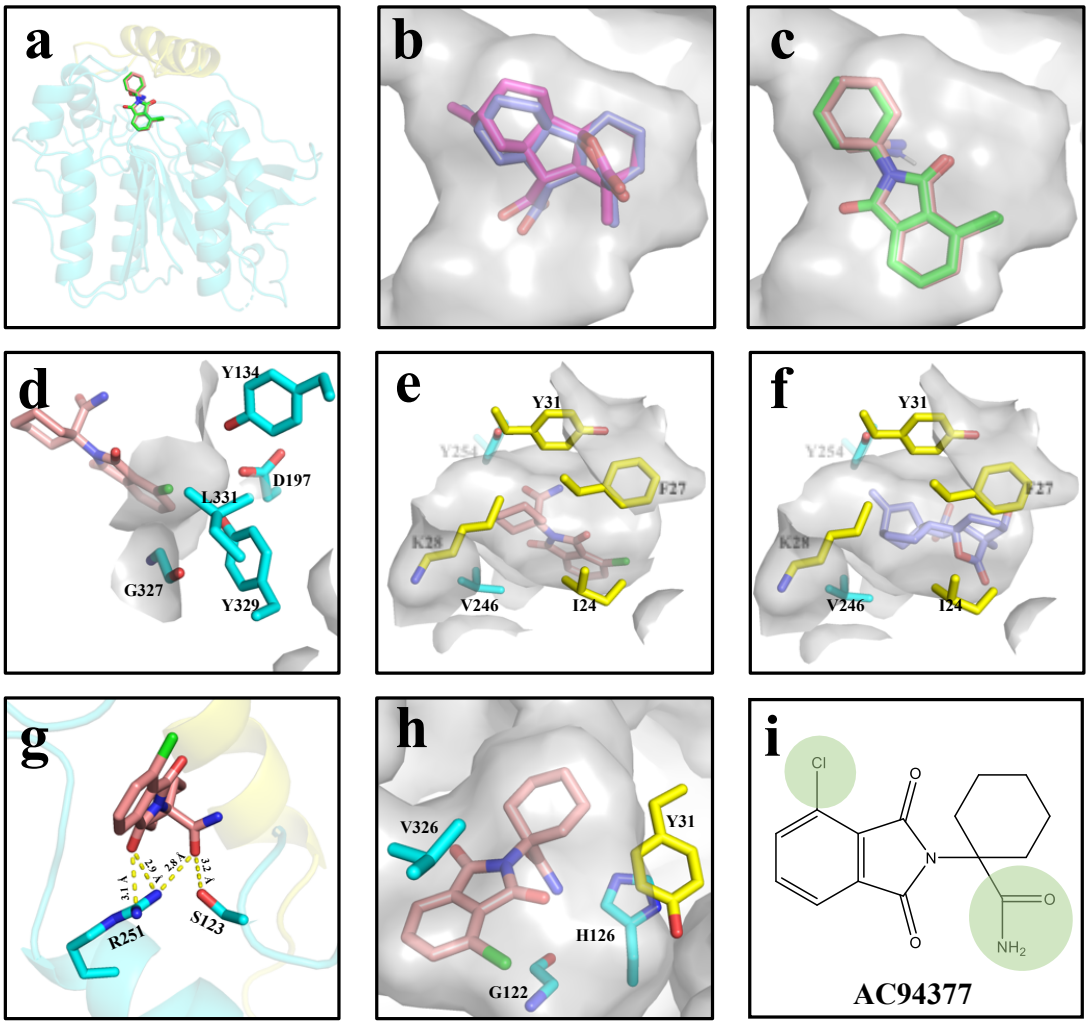


Figure 2-9: Analysis of the docking structure of OsGID1-AC94377 complex.
a, overall view of the docking structure. OsGID1 is represented as the cartoon model with AC94377 as stick. **b,** re-docking OsGID1 with its crystallographic ligand GA_4. GA_4 from docking simulation (magenta stick) is essentially superposed with that (purple stick) from crystal structure OsGID1-GA_4. **c,** Close-up view of the docking structure of OsGID1-AC94377 (orange stick) and AtGID1a-AC94377 (green stick) complex. **d,** Chloro group of AC94377 is surrounded by the resides (cyan sticks): Y134, Y329, D197, G327, and L331. **e,** Cyclohexane ring of AC94377 heads up to the N-extension of OsGID1 and involved in the hydrophobic interactions with residues surrounded (yellow sticks): I24, F27, K28, Y31, V246, and Y254. These residues also provide the hydrophobic environment to recognize the aromatic C/D-ring of GA_4 (**f**). **g,** The hydrogen bonds (indicated as yellow dotted lines with black labels) were detected between OsGID1 and AC94377 in the docking structure. **h,** The hydrophobic interaction between V326 and the aromatic plane of AC94377. The amine group of AC94377 is tightly surrounded by the residues, Y31, G122, and H126. **i,** Chemical structure of AC94377.

AC94377 forms two hydrogen bonds respectively to the N (η1) and N (η2) atoms of R251. Moreover, the carbonyl group on carboxamide of AC94377 also forms two hydrogen bonds, one of which points to the N (η2) atoms of R251 and the other to the O (γ) atom on the hydroxy group of S123. It should be noted that both oxygen atoms of these two carbonyl groups work as the hydrogen bond acceptor by sharing two lone pairs of electrons (Figure 2-9g). The amino group of carboxamide is tightly surrounded by the residues, Y31, G122, and H126 (Figure 2-9h). Moreover, I found the alky side chain of V326 establishes an edge-to-face hydrophobic interaction with the aromatic plane of AC94377, which would possibly be enhanced by the elongation of the alkyl group as shown in the mutant I326 (OsGID1^{V326I}) (Figure 2-9h). I also perform the docking simulation for AC94377 and AtGID1a under the same docking procedure applied for AC94377-OsGID1. As you can see in Figure 2-9c, the best structural pose of AC94377 in AtGID1a (ΔG = -9.4 kcal/mol) is substantially superposed to that in OsGID1, further suggesting the proper application of Autodock program in this work.

Regardless of the fact that docking structure of AC94377-OsGID1 could mostly support the structure and activity relationship studies (SAR) on AC94377, some of the SAR results still remain enigmatic even with this complex structure. For example, both Y2H assay and affinity measurement reveal that, unlike AC94377 and other derivatives, ACOH and ACOIp do not show the preference to OsGID1^{V326I} (Figure 2-8, Table 2-3), but the docking structure indicated that Val326 may not affect the carboxamide of AC94377 since they are structurally flanked by the aromatic plane of AC94377 (Figure 2-9h). It is unlikely that the simple substitution of the hydroxy group with amino moiety of carboxamide would rearrange the orientation of AC94377. Then, how does the variant OsGID1^{V326I} decline its binding affinity to ACOH and ACOIp? Moreover, it is also difficult to explain the reason why ACF containing the fluoro substituent instead of chloro group of AC94377 strongly lost its affinity, while JK002 with hydrogen substituent on this position shows higher affinity to OsGID1^{V326I}. Perhaps, some conformational changes occurred for the adaption of different chemicals' binding, which cannot be observed in docking simulation. In other words, we still need accurate information of the structure complex, which will be particularly useful for the redesign of a more potent agonist based on the chemical skeleton of AC94377.

2.2.4 Orthogonal ligand-protein variant pairs of AtGID1a-AC94377

The use of synthetic small molecules has accelerated our understanding of hormone physiology and signaling in plants by overcoming the obstacles of gene essentiality (lethality) or high redundancy in gene families, which are always urgent problems in classical genetics approaches[108, 151]. In addition, as the newly emerged application of chemical biology, orthogonal plant hormone-receptor pairs, which is a multidisciplinary approach by integrating chemical biology with synthetic biology, is designed as a controllable switch with high sensitivity and selectivity to modulate hormone responses in a tissue/cell-specific manner[152, 153]. There exist three GA receptors (GID1a, GID1b, GID1c) and five GID1 repressors [GAI, RGA, RGA-LIKE1 (RGL1), RGL2, and RGL3] with overlapping functions in Arabidopsis, all of which respond to the native ligand. Therefore, the delineating of specific GID1-mediated signaling in the complex living system of Arabidopsis by using GAs or agonists that show pleiotropic effect is still a challenging task. In this context, the orthogonal system, GID1^{variant}-AC943377 in this study, could erase the pleiotropic interference of the original and different receptor-ligand combinations (AtGID1s-GAs) in the complicated regulatory networks so that boost its practicality for the precise depiction of the physiological role of genes or proteins of interests.

2.2.4.1 The methodology for engineering ligand-receptor pairs

The successful establishment of specifically matched ligand-receptor pairs requires the tasks from either protein engineering approach or synthetic chemistry approach. More specifically, the latter involves the use of existing mutant forms of the receptors that respond weakly to natural ligands to prepare new synthetic ligands for which the mutant receptors have strong affinity[154, 155]. As stated, this approach has been favored by synthetic chemists who specialize in finding ligands that reproduce the interaction with a given mutant receptor. For example, inspired by SAR studies ligands that specifically bind to mutant forms of human estrogen receptors hERα or hERβ[156, 157] were developed by rationally manipulating the ligand. However, despite these achievements, there could be several constraints in this approach, such as the time-consuming trials of chemical synthesis when in the absence of the guidance of structural information, and only moderate affinity of pairs. Conversely, the protein engineering approach, which is to find a directed receptor

mutant on the ligand-binding site that is only specific to a new ligand that poorly activates the wild-type receptor, is more favorable to create an orthogonal ligand-protein pair, since it is much easier to obtain the receptor mutants, even the pools of the mutant libraries, just under the routine procedures of DNA-based mutagenesis[158, 159].

Notably, application of the protein engineering approach can be further advanced by rational re-engineering or random mutagenesis/selection of receptors. The structure-guided rational engineering of receptors requires structural grasp of the interaction mode of ligand with receptor so that generate the efficient schemes for receptors engineering, such as the design of steric complementation, manipulation of charged interactions, ion pair neutralization, polar/charged group exchange, and even creation of *de novo* binding sites at ligand-receptor interfaces[154]. Of these, the commonly used and easiest way to the envision of creating an orthogonal ligand-receptor pair is the steric complementation, which can modify the complementary shape of the ligand-receptor interface. That is, even if the engineered ligand is too large to fit within the binding site of the wild-type receptor, it can bind to a modified form of the receptor whose binding site has been appropriately enlarged by mutagenesis. This is also well known as the 'bumps and holes' strategy and extensively applied to generate ligand-receptor pairs[160-163]. The extraordinary example in chemical regulation of plant hormone function would be the synthetic auxin-receptor pair that works orthogonally to natural auxin signaling with 10,000-fold improvement of sensitivity[129, 164]. Regardless of these achievements, it is still a big challenge to skillfully make effective variant receptors that actively discriminates the binding to a natural ligand and yet can be recognized by synthetic ligands with high affinity and selectivity, even if the rational engineering of proteins could be adopted for designing orthogonal pairs.

The alternative is to *de novo* reshape the pocket of receptors by variants library generation based on combinatorial codon randomization, forcing the mutated receptor to have a specific adaption for new ligand in a rapid fashion. This chemical-directed evolution of receptor usually relies on the random mutagenesis of several/all residues involved in the construction of ligand-binding pocket of receptor[165]. Moreover, coupled with the ligand-dependent interaction with partner protein downstream, the yeast two-hybrid based screening system greatly benefit this approach by saving numerous efforts to rapidly sift through a large library of receptor variants, which is particularly achieved by placing the selection pressure on host cells that can survive and

grow on selective media lacking essential nutrition when harboring a variant that strongly responds to new ligands. Recently, a great success of plant chemical biology by employing this method was made through the engineering of a highly specific ligand-receptor pair for the hijacking of ABA signaling[128]. From another perspective, directional evolution-assisted trial, which shifts the ligand selectivity of a protein, is also feasible, since biologists have already made great achievements in enzyme engineering with *de novo* function or novel preference for other substrates and offer more reasonable strategies and systematic theories for variants library construction and selection[166-172]. Additionally, it should be noted that, α/β-hydrolase fold superfamily, where GID1 belongs, have long been considered as the versatile structural scaffold for protein engineering to enhance some properties, such as the expanding substrate range, increasing thermostability and altering substrate enantioselectivity[173-175].

In this study, I place the focus on the random mutagenesis of GID1 to generate a variant that is highly sensitive to artificially synthesized chemicals, AC94377 and its derivatives, but loses its affinity to its native ligand GAs. More importantly, this variant GID1-AC94377 pairs with higher affinity will be in turn applied for co-crystallization again so that provide the essential clues for the structure-based design of AC94377 analogues with upgraded biological activity in plants.

2.2.4.2 Design, construct, and screening of the variant libraries of AtGID1a

To make the randomized variant libraries of protein of interests, we need to choose a proper and efficient method of gene mutagenesis. In past two decades, the most frequently used gene mutagenesis techniques for the directed evolution of enzymes/proteins are error-prone polymerase chain reaction (epPCR; a shotgun method), DNA shuffling (a recombinant technique), and saturation mutagenesis (usually generate the focused mutant libraries)[176-178]. The first two methods could randomize variant libraries in the random location of the protein sequence, therefore are usually applied for the engineering of biocatalysts when their structural or functional knowledge is not available. It should be noted that Arnold and her coworker have first employed epPCR as a successful strategy for directed evolution to enhance the resistance of a protease to a hostile organic solvent[179]. Her pioneering works in directed evolution have been awarded Nobel prize in Chemistry in 2018[180]. However, such approaches usually require screening of large protein

combinatorial libraries, which is growing to be the bottleneck of directed evolution[181]. Especially, in the case of engineering GID1, the manipulation of residues on the protein surface would potentially influence some interactions with partner proteins downstream in plants, leading to the changes/loss of its physiological functions. Therefore, we decide to take saturation mutagenesis (SM) as the lead strategy to engineering GID1 with the focus on the residues lining the ligand-binding pocket.

The 24 residues shaping the pocket of AtGID1a include I24, F27, Y31, G114, G115, S116, H119, S120, I126, Y127, D190, S191, N218, M220, V239, D243, R244, Y247, D289, L290, I291, G320, Y322, and L323. Previously, Park et al. performed the single-site saturation mutagenesis (SSM) on the 23 pocket-lining residues in ABA receptor PYR1 by using NNK primers[128]. It is really a labor-intensive task both in the library construction and the screening steps, not to mention the following efforts on iterative screening of the hits obtained from each site of 23. Opposed to SSM, saturation mutagenesis at multi-residue randomization sites designed by appropriate residue-grouping, termed combinatorial saturation mutagenesis (CSM), is more preferred[172]. Moreover, the cooperative effects of several residues always affect the interaction with ligand, suggesting the coevolution of adjacent residues (or simultaneous randomization) would be much more favorable for the generation of best variants. Therefore, these 24 residues of AtGID1 would be assigned into several groups for random mutagenesis.

Next, we are going to address the numbers problem, which is also the basic challenge in directed evolution. Traditionally, NNK (N: adenine/cytosine/guanine/thymine; K: guanine/thymine) codon degeneracy covering 32 codon combinations for encoding all 20 canonical amino acids is widely used for SM trials[166, 171]. However, the diversity of library is caused at the codon level, when the number of residues in the randomization site increases, oversampling of transformants (usually need to ensure ≥ 95% coverage of variant library) would dramatically reach astronomical dimensions, impossible to screen in a practical manner (Table 2-4). For example, saturation mutagenesis at a two-residue site would need to sample 3,000 transformants, which increases to 100,000 when choosing a three-residue site (Table 2-4). To reduce the codon redundancy yet retain the full coverage of 20 canonical amino acids at a single site, there did develop a special trick (named 22c-trick) by making a mixture of serval primers, only creating a degeneracy of 22 unique codons. Despite of reducing screening loads, the

application of this method is really at high cost[168]. For example, the combinatorial saturation mutagenesis at two sites will require 9 primers and 27 for three sites. Recently, researchers have generated the 'smarter' randomized libraries by using reduced amino acid alphabets[166, 182]. Among them, the NDT codon degeneracy (D: adenine/guanine/thymine; T: thymine) covering 12 codon combinations have been proved to be one of the most efficient strategies for mutagenesis due to the balanced nature of 12 encoded amino acids: Phe, Leu, Ile, Val, Tyr, His, Asn, Cys, Arg, Ser, Asp, and Gly, a "cocktail" containing hydrophobic/hydrophilic, polar/non-polar, aromatic/aliphatic, and small/large sidechains of residues[172, 183]. Therefore, for the 95% library coverage, it is practically acceptable for sampling 3000 transformants of an NNK-based library on two sites and 5000 of an NDT-based library on three sites (Table 2-4; Table 2-5). Based on this, the 24 residues lining the pocket of AtGID1 were grouped into 9 sites/sublibraries: Site A (G114, G115), Site B (S116, H119, and S120), Site C (G320, Y322, L323), Site D (I126 Y127), Site E (I24, F27, Y31), Site F (V239, D243, R244), Site G (D190, S191), Site H (N218, M220) and Site I (D289, L290, I291), among which Sites B/C/E/F/I containing 2 residues were created by NNK-based SM and Sites A/D/G/H containing 3 residues created by NDT-based SM. Presently, several molecular biological SM methods based on polymerase chain reaction (PCR) are available for variant libraries construction, including QuikChange™/SPRINP[184, 185], MegaPrimer[167, 186], and overlap-extension PCR[187, 188]. In this study, we choose a post-ligation method by using a commercially available kit (KOD-Plus-Mutagenesis kit, TOYOBO) with completely non-overlapping primers because of the high convenience and efficiency. It should be noted that, after getting some beneficial mutants from one or more sites, the iterative saturation/random mutagenesis experiments (ISM/IRM) or combinatorial active-site saturation test (CAST)[189], will subsequently performed to generate the best hit. This is particularly important, since it has been theoretically and practically examined to cause the strong cooperative mutational effects (more than traditional additivity), which will eventually produce a greatly enhanced variant[172, 190].

On the other hand, an efficient screening method would greatly minimize our efforts. The AC94377-dependent interaction between GID1 and DELLA in yeast cells have been proved to be reliable in our previous work[130]. Moreover, AtGAI has been revealed to have the strongest interaction with AtGID1a protein than other four DELLA proteins, I thus employed the Y2H assay for the high-throughput screening of 9 AtGID1 variants sites/libraries using AtGAI as prey fused

to GAL4-AD (activation domain) and all GID1 variants as bait fused to GAL4-BD (DNA-binding domain). The schematic diagram of workflow is shown in Figure 2-10.

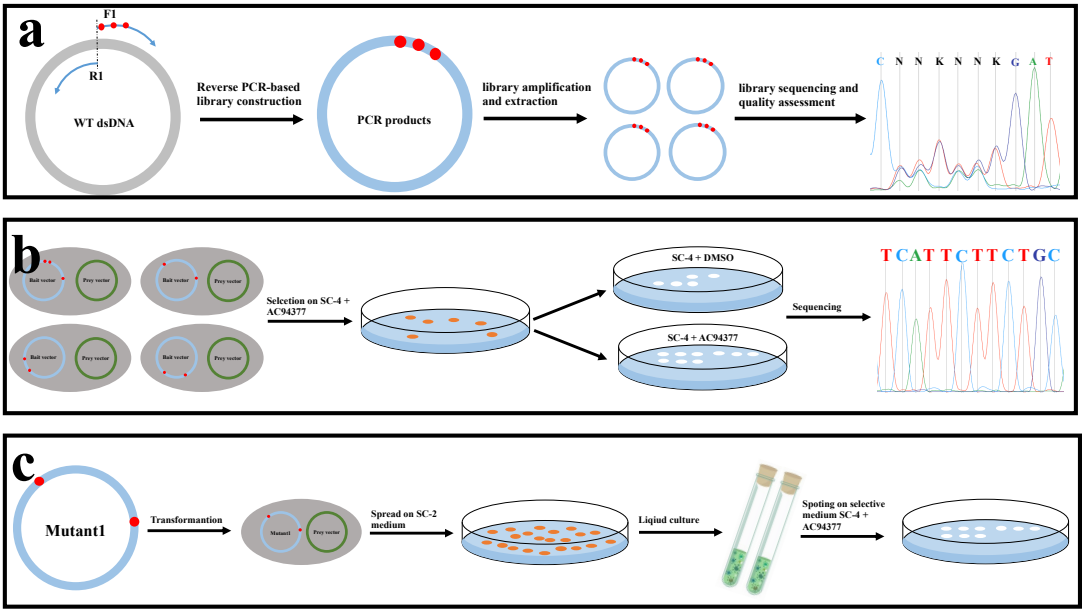


Figure 2-10: The schematic diagram of workflow for the AC94377-driven evolution of GID1. **a,** AtGID1 variants library construction. **b,** Y2H-based screening of variants library. **c,** The verification of beneficial variants. SC-2 means SC-LW. SC-4 means SC-LWHA.

2.2.4.3 Initial screening of 9 pocket-focusing sublibraries (Round 1)

Before starting the SM experiments, the sensitivity of wild-type AtGID1a to GA_4 and AC94377 was semi-quantitatively examined by Y2H assay using AtGAI as prey under the concentration gradient of ligands. As shown in Figure 2-11, the minimum concentration of GA_4 and AC94377 to sufficiently induce the PPI interaction between AtGID1a and AtGAI are 0.03 μM and 50 μM, respectively. The responsiveness of a variant AtGID1a^{V319I} to AC94377 was stronger than that of wild type but comparable to that to GA_4, which was consistent with our previous test[130], confirming the reproducibility of AC94377-induced interaction between AtGID1a and AtGAI. Next, the pools of mutagenized DNA were efficiently transformed into yeast cells (AH109) harboring the prey plasmid (AtGAI), and the cells were then spread on the synthetic complete drop-out medium (SC-LWHA) containing 1-5 μM chemical AC94377 as the evolutionary pressure

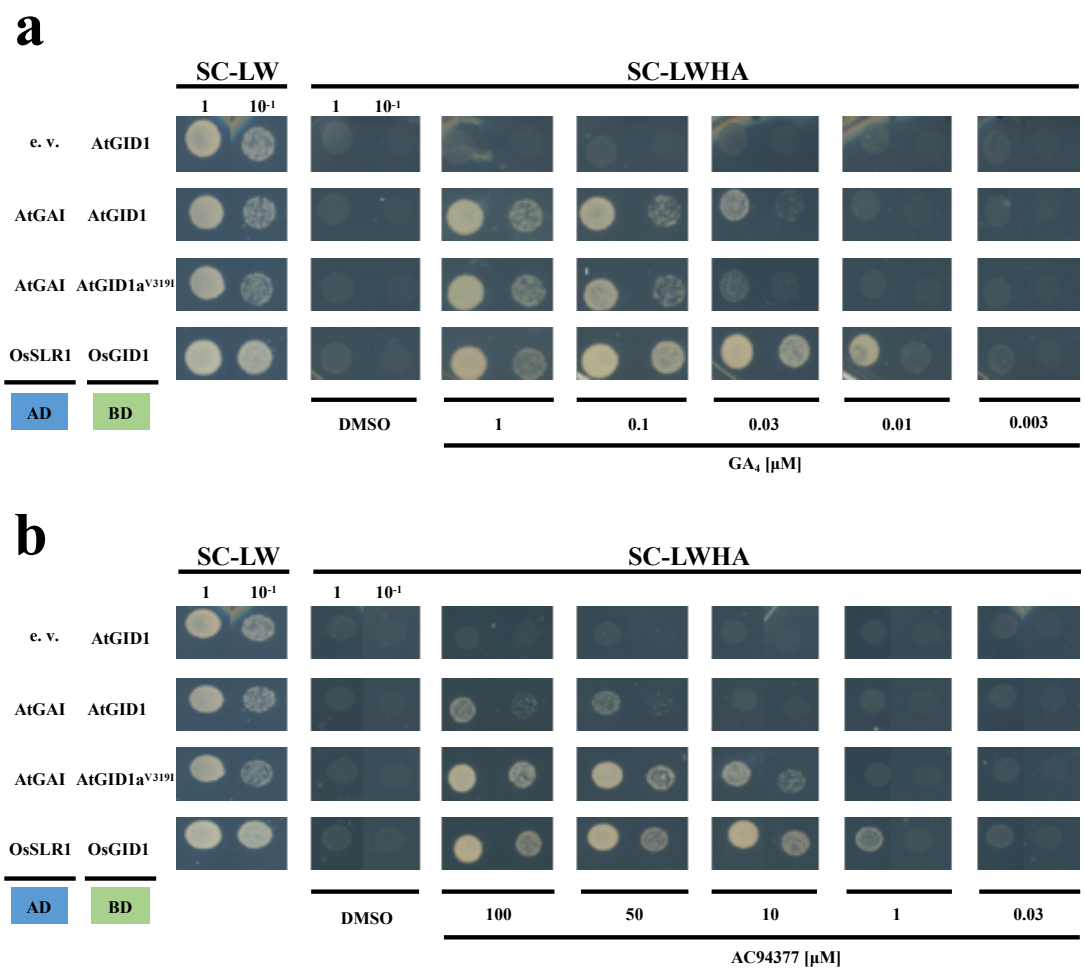


Figure 2-11: The responsiveness of GID1 to GA_4 and AC94377 in the Y2H assay.
a, The sensitivity of GID1 to GA_4 is examined. The minimum concentration of GA_4 for initiating the PPI of AtGID1a-AtGAI and OsGID1-OsSLR1 is 0.03 μM and 0.01 μM, respectively. **b,** The sensitivity of GID1 to AC94377 is examined. The minimum concentration of AC94377 for initiating the PPI of AtGID1a-AtGAI and OsGID1-OsSLR1 is 50 μM and 0.1 μM, respectively. Yeast cells were grown for 2 days in the liquid SC-LW media and 3 μL of each yeast culture was spotted on SC-LWHA plates containing varying concentration of AC94377 and GA_4 or DMSO at 30 ºC for 3 days. n=1.

for selection. To avoid the self-activation and ligand-independent activation (considered as the false-positive colonies), all generated transformants were spotted on SC-LWHA solid medium complemented with DMSO or 1-5 μM AC94377. This step is strongly needed, since so many transformants were found to show the ligand-independent activation of GAL4. Furthermore, if some candidates were found to show improved affinity to AC94377, the plasmid of these transformants will be extracted for DNA sequencing and transformed back to AH109 for re-

examination by standard Y2H assay again (Figure 2-10).

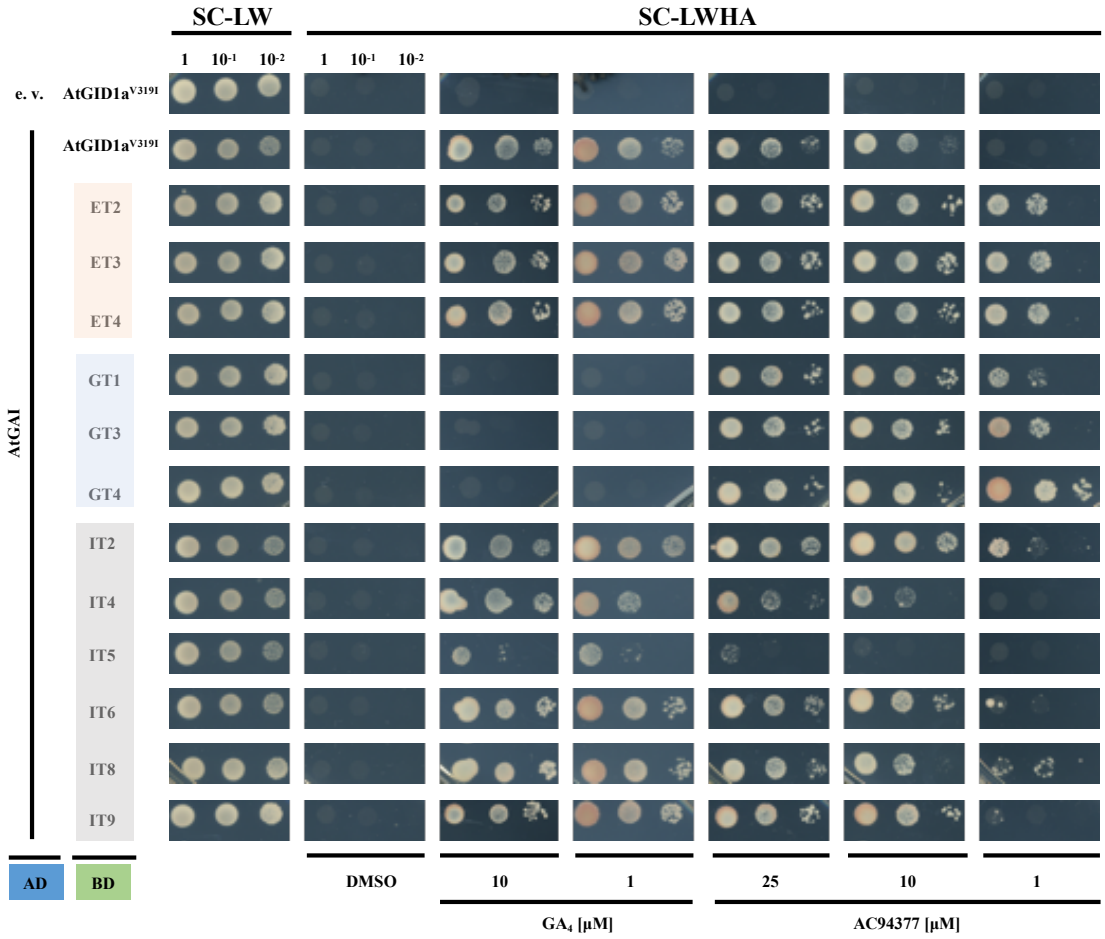


Figure 2-12: Initial screening produces several variants from sites E, G and I with improved affinity.

Variants ET2, ET3, and ET4 are obtained from the screening of sites E. Variants GT1, GT3, and GT4 are obtained from the screening of sites G. Variants IT2, IT4, IT5, IT6, IT8, and ET9 are obtained from the screening of sites I. Yeast cells were grown for 2 days in the liquid SC-LW media and 3 μL of each yeast culture was spotted on SC-LWHA plates containing varying concentration of AC94377 and GA_4 or DMSO at 30 ºC for 3 days. n=3.

From these 9 sublibraries, we found several hits from sites E/G/I that could respond to AC94377 at as low as 1 μM. The sensitivity of these hits to AC94377 was increased 10 and 50 times higher than AtGID1a and AtGID1a^{V319I}, respectively (Figure 2-12; Table 2-6). Of these, variants from sites E/G showed higher affinity than those from sites I, suggesting a more positive contribution of residues in sites E/G in the interaction of AtGID1a with AC94377. I noticed that the variants ET2/3/4 all contains the mutation I24L, whereas residues Tyr/Leu is suitable for the F24 position to enhance the interaction with AC94377. As for the positive variants from site G, it

is interesting to find that the catalytic nucleophile S191 could be replaced by residues Asp/Cys/Glu, among which cysteine and aspartate are well-known to act as the nucleophiles in α/β-hydrolase fold superfamily. More importantly, we believed that the mutation at this position shifted the ligand selectivity of AtGID1a from GA_4 to AC94377. As reported, in the structure of AtGID1a-GA_4 complex, GA_4 is anchored at the bottom of the pocket by forming multiple hydrogen bonds between 7-carboxylate and residues S191. Therefore, the mutation at this site probably disrupts these hydrogen bonds or causes the steric clash and electrostatic repulsion between carboxylate of GA_4 and substituent residues (aspartate or glutamate). However, Y31 of site E and D190 of site G are still retained after SM, suggesting the important roles they play to maintain the AC94377-binding affinity in AtGID1a. Unlike residues in sites E and G, the three residues in site I have more mutational combinations, where variant IT2 could respond to AC94377 at 1 μM, but others show only slightly increased affinity (Table 2-6).

2.2.4.4 Combination of beneficial mutants (Round 2)

My next work is to make the best combination of improved mutants at these three sites. However, this is not as simple as doing basic math, like 1 + 1 = 2, according to the knowledge of directed evolution, that is, the addition of the beneficial mutations is not at all additive, but enormously cooperative (more than additive)[191]. Moreover, except for the mutational additivity, we should never ignore the possible epistatic effects, which means that a combination of two improved mutants would further increase or decrease the directed properties, such as substrates selectivity[190]. The best way to optimize the result is to test all combinations, however, which would place huge burdens on researchers for plasmid constructs and affinity test. For example, 3 good site-directed variants would result in 3! = 6 evolutionary pathways. In this study, the variants ET2/3/4, GT1/3/4 and IT2 would lead to 5040 combinations, which are not practically acceptable. Supported by the knowledge of cooperative effects of residues, I still have a chance to greatly improve the affinity of mutants to AC94377, therefore it is necessary to combine these beneficial variants for further examination. To avoid such labor combination test one by one, I decided to make one more library (Site J) covering the codons encoding all beneficial mutants of site E/G/I, including variants that show a relatively weak response to AC94377 at 1-10 μM (Figure 2-13;

Table 2-6). Considering the epistatic effects, the residues of wild type at corresponding position will be also included in site J.

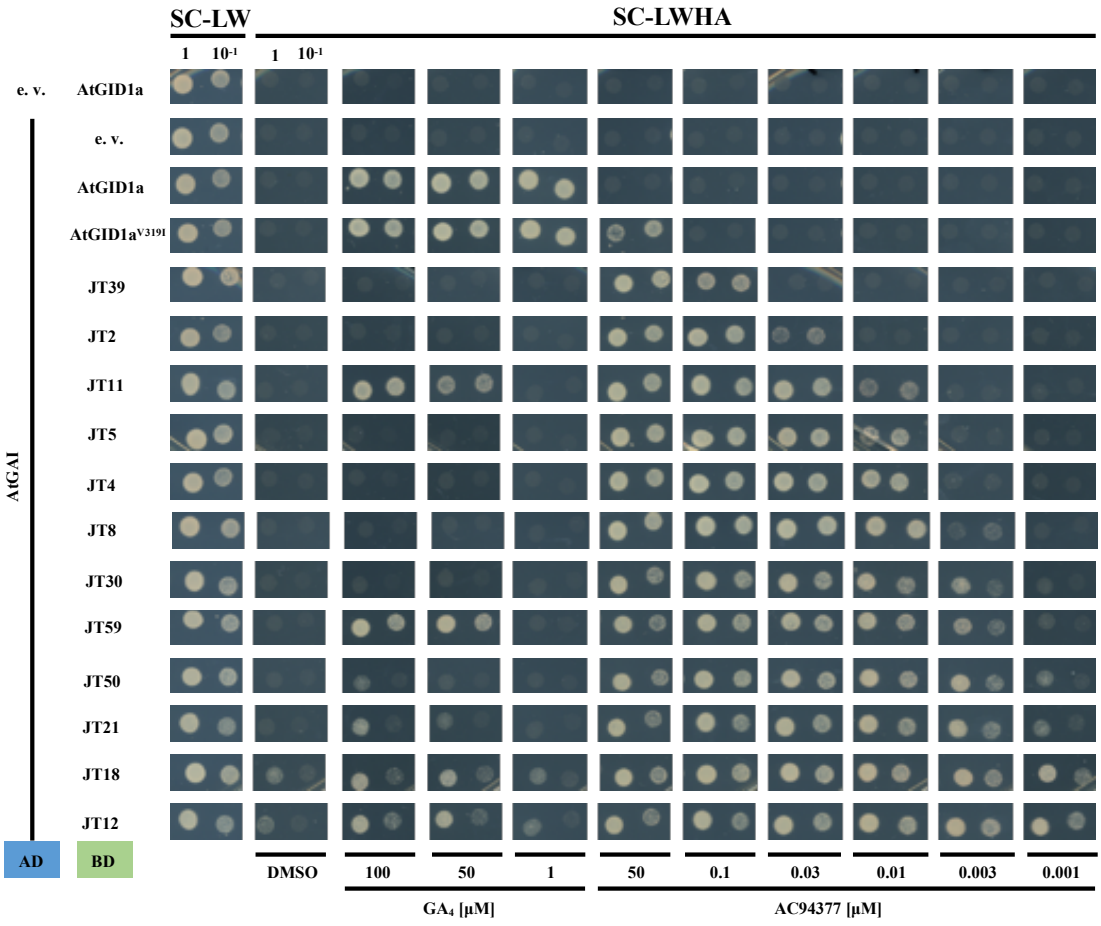


Figure 2-13: The combination of beneficial mutants of site E, G and I produces variants with greatly improved affinity.
The site J covers the codons encoding all beneficial mutants of site E/G/I (shown in Figure 2-11), including those variants making a relatively weak response to AC94377 at 1-10 μM. Yeast cells were grown for 2 days in the liquid SC-LW media and 3 μL of each yeast culture was spotted on SC-LWHA plates containing varying concentration of AC94377 and GA_4 or DMSO at 30 ºC for 3 days. n=3.

With the goal of obtaining the AtGID1a variants showing the obviously improved AC94377-binding affinity, the screening of site J was performed under the stringent condition with 10 nM of AC94377. Afterward, all positive transformants, which had been examined by Y2H once again, were used to compare the sensitivity to AC94377 and GA_4 under the concentration gradient (Figure 2-13). As we can see in the Table 2-6, all of selected variants show strongly enhanced AC94377-binding affinity. Among them, JT8, JT30, JT59, JT50, JT21, JT18, and JT12 could

respond to AC94377 at as low as 1-3 nM, which suggests that the strength of AC94377-induced PPI between these AtGID1a variants and GAI is increased about 16,000-folds than that between AtGID1a wild-type and GAI (at 50 μM AC94377) under Y2H system (Figure 2-13). Moreover, the sensitivity of AtGID1a variants to AC94377 is increased 10-30 folds than that of AtGID1a wild type to GA_4 (Figure 2-11; Figure 2-13). However, I also noticed that some variants could still respond to GA_4. For example, even though variants JT12 and JT18 show the highest responsiveness to AC94377 (1 nM), they could also respond to GA_4 at 1 μM. Notably, variants JT5, JT4 JT8, and JT30 could respond to AC94377 at 3-10 nM, but not to GA_4 even at 100 μM (Table 2-6), suggesting the successful establishment of orthogonal ligand-protein variant pair of AtGID1a and AC94377.

Based on the results of Y2H assay for these variants, I believed that the mutations I24L, S191E, and V319I were very important for AC94377-binding affinity for AtGID1a, among which mutation S191E was a major contributor to shift the ligand selectivity of AtGID1a from GA_4 to AC94377. The mutations F24Y and F24L were all acceptable for the binding to AC94377. It is noteworthy that, after initial screening, most positive transformants of site I contain the mutation D289N, while this mutation in site J is lost (recovered to D289), which may suggest the existence of mutational cooperativity and epistasis (Table 2-6). Since the mutation combinations within site I are variable after selection (Table 2-6), especially in I291, I feel difficult to trace the regular pattern regarding how these three residues of site I make the evolutionary adaption of themselves or co-adaption with other residues of site E and G to strongly bind to AC94377.

Sequence alignments have suggested that all residues lining the active pocket are strongly conserved with exception of L323, N218, and I291 of Arabidopsis GID1a (Figure 2-18). Afterwards, we construct the phylogenetic tree based on the amino acid sequences of these GID1 proteins, which are artificially classified into three different types: A-type, B-type and Monocots type (Figure 2-19). I noticed that the GID1 proteins of A-type and Monocots type contain Leucine and Asparagine in the corresponding position of L323 and N218 in AtGID1a, respectively, whereas GID1 proteins of B-type contain Phenylalanine and Histidine in the corresponding positions, respectively (Figure 2-18). My efforts of laboratory evolution did not find any positive variants of these two residues contributing to the binding affinity to AC94377. However, sequence alignment suggested that I291 (assigned to site I) is poorly conserved among GID1 proteins of A-

type and B-type (could be Ile, Val, Ser, and Leu) but strongly conserved in Monocots type GID1 proteins (only Thr) (Figure 2-18, Figure 2-19). Interestingly, these amino acids are all observed in my positive variants of site J, for example, JT5 and JT21 contain I291; JT59 has V291; JT39 and JT50 have T291; etc (Table 2-6). This may explain that my trials of laboratory evolution on GID1 could approximate the natural selections and all of these residues promote the AC94377-binding affinity in GID1.

2.2.4.5 Biochemical characterization of AtGID1a variants *in vitro*

To characterize the direct interaction between ligands and variant receptors, we choose the AtGID1a wild type and variants (JT8, JT30, JT59, JT50, JT21, JT18, and JT12) for the measurement of ligand-binding affinity to AC94377 and GA_4 and detection of ligand-induced shift of thermostability by AC94377 and GA_4.

Surprisingly, we find that AtGID1a wild type (K_d = 1333.2 μM) have extremely low affinity to AC94377 compared with OsGID1 (K_d = 79 μM), which may partially explain that strength of AC94377-induced PPI between AtGID1a and AtGAI is weaker than that between OsGID1 and OsSLR1 (Table 2-3; Table 2-7; Figure 2-11b). For the AtGID1a variants tested, they all show greatly improved binding affinity to AC94377, which is about 500-folds higher than AtGID1a wild type. Moreover, ligand selectivity of these variants is totally changed from the natural ligand GA_4 to the artificially synthesized ligand AC494377 (Figure 2-14; Figure 2-15; Table 2-7). However, even though Y2H assay suggests that the sensitivity of these variants to AC94377 is stronger than that of AtGID1a wild type to GA_4 (Figure 2-11a, Figure 2-13), the binding affinity of variants to AC94377 is slightly weaker than that of AtGID1a wild type to GA_4 (Table 2-7). This may be caused by the different assay systems between Y2H and ITF since Y2H assay is for the examination of protein-ligand-protein interaction, whereas ITF assay is to inspect the protein-ligand interaction. Moreover, protein amount (or expression efficiency, which could be influenced by the protein stability as discussed below) is the one of main factors affecting the Y2H assay. Interestingly, the titration of GA_4 failed to induce the intrinsic fluorescence changes of these variants, suggesting that there is no binding event between protein and ligand (Figure 2-14).

To further confirm the improved sensitivity and selectivity of these variants to AC94377 and

GA4, I performed the DSF assay to investigate the ligand-induced changes of thermostability in GID1 proteins. As expected, the thermostability of AtGID1a wild type is increased by GA_4 in the concentration-dependent manner, whereas GA_4 cannot (or slightly) induce the thermostability changes in AtGID1a variants (Figure 2-16, Table 2-8). In contrast, AC94377 could strongly enhance the thermostability of AtGID1a variants but not wild type (Figure 2-17, Table 2-8). More importantly, we unexpectedly found that the thermostability of variants is higher than wild type even without the treatment of ligand. For example, the melting temperature (Tm) of JT8, JT30, TT50, JT18, and JT12 is 12 ºC higher than that of AtGID1a wild type when treated with DMSO (Table 2-8). This may result in the difference in protein expression efficiency. For example, for each 2-liter culture, 8-10 mg of AtGID1a wild type and 25-35 mg of AtGID1a variants are available after expression and purification (data are not shown). As reported, the degradation of GID1 proteins, which is negatively regulated by ubiquitination (though GARU, GA receptor RING E3 ubiquitin ligase) and positively regulated by Tyr phosphorylation (through CRK2 TAGK2), is an important event in the regulation of seedlings and seeds by accumulation of DELLA proteins in plants[192]. I would be looking forward to seeing whether and how the increment of thermostability makes changes in the physiological functions of GID1 protein when AtGID1a variants are applied to the plants.

Taken together, the AC94377-directed evolution of AtGID1a has enabled us to find some variants showing greatly improved affinity and selectivity to AC94377. More importantly, the sensitivity of variants to AC94377 is higher than that of AtGID1a wild type to GA_4, stating the successful establishment of orthogonal high-affinity AtGID1a-AC94377 variant pairs.

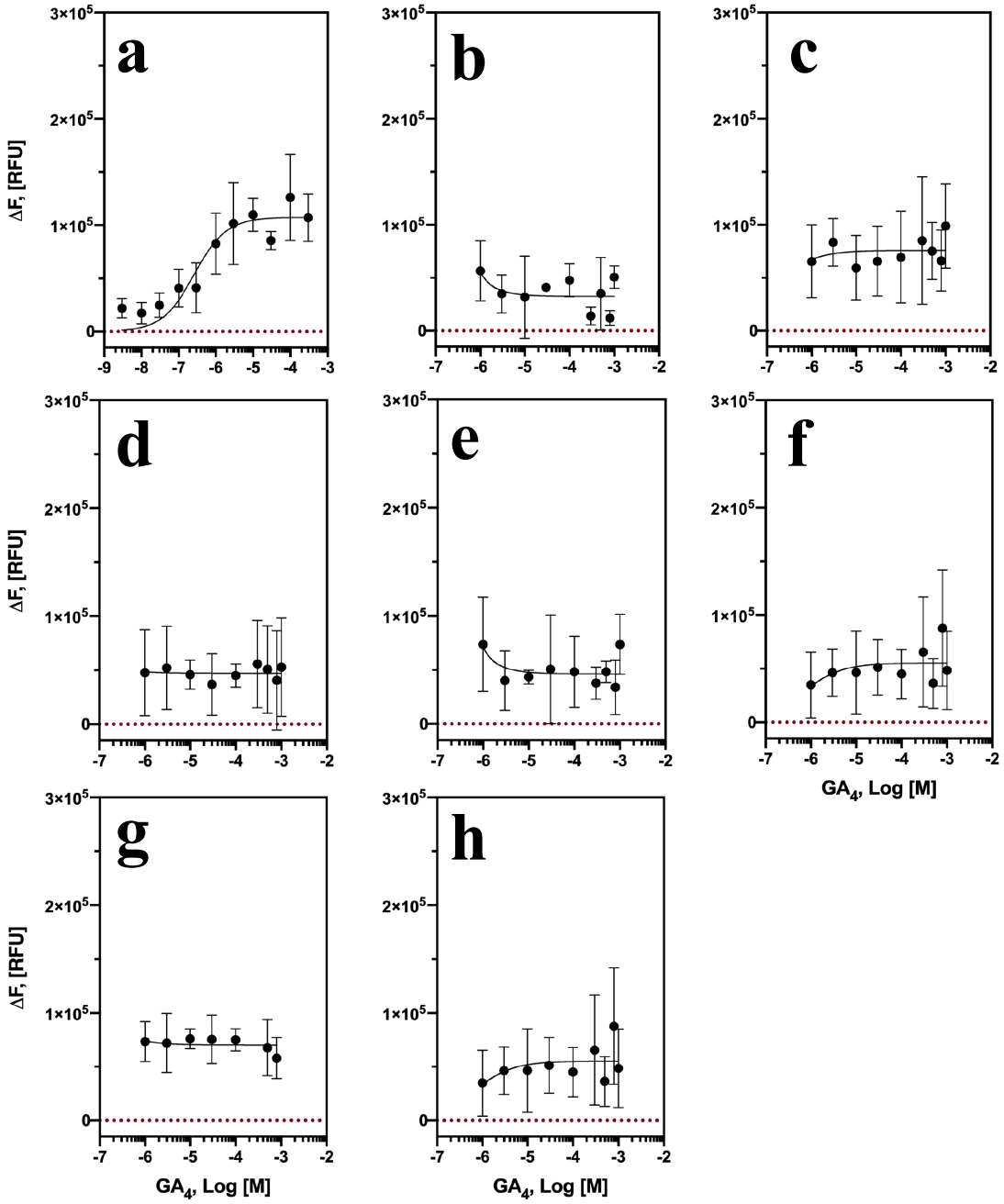


Figure 2-14: Measurement of GA_4-binding affinity of AtGID1a and variants by ITF.
The fluorescence changes of AtGID1a (**a**), JT8 (**b**), JT30 (**c**), JT59 (**d**), JT50 (**e**), JT21 (**f**), JT18 (**g**), and JT12 (**h**) in response to the titration with GA4. ΔF is calculated by taking the difference in protein fluorescence of the DMSO control and each ligand concentration. ΔF is plotted *versus* ligand concentration by fitting to nonlinear regression with '*One Site-Nonspecific Binding*' model. n=3. The dissociation constant/K_d (μM) was summarized in Table 2-7.

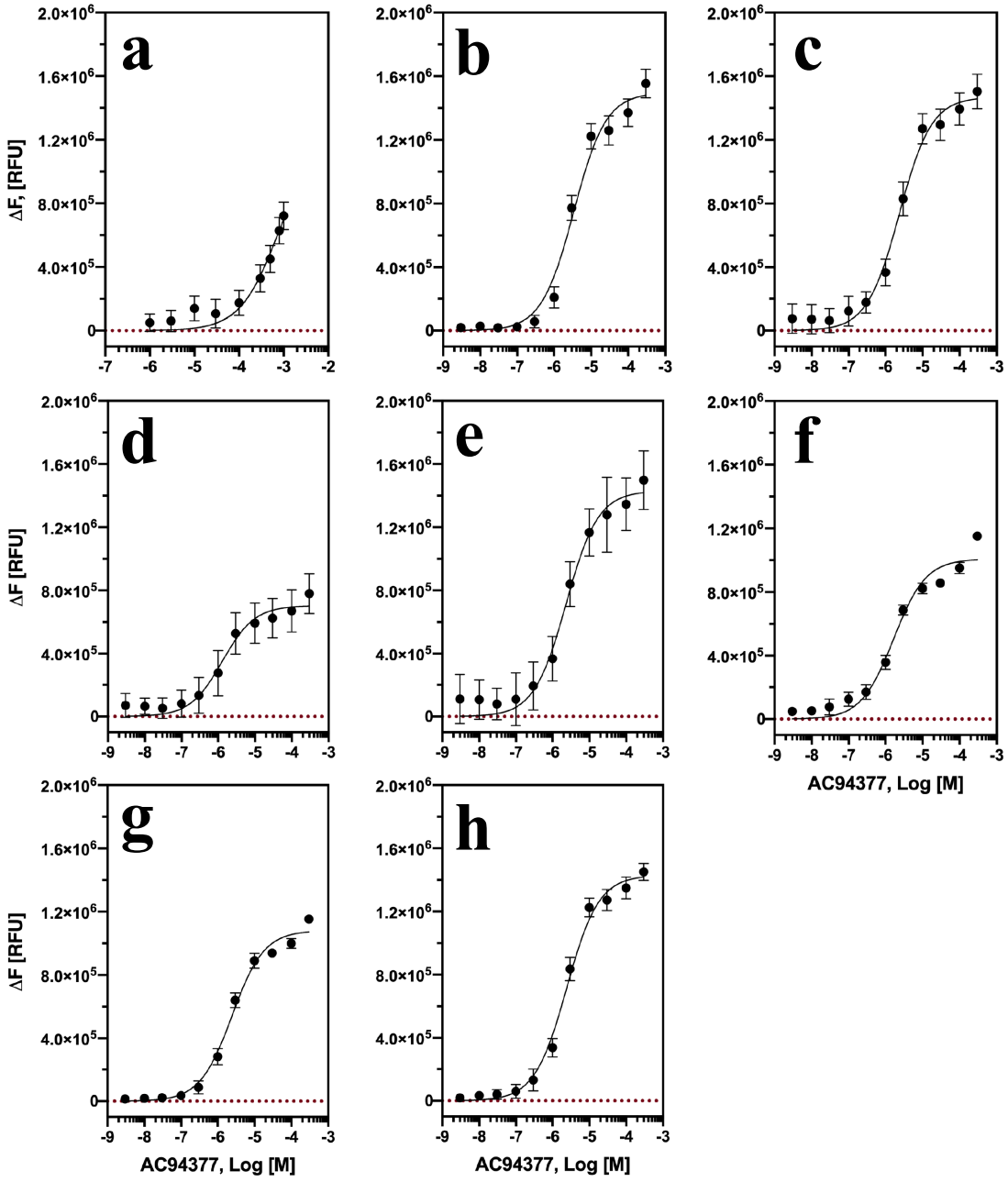


Figure 2-15: Measurement of AC94377-binding affinity of AtGID1a and variants by ITF. The fluorescence changes of AtGID1a (**a**), JT8 (**b**), JT30 (**c**), JT59 (**d**), JT50 (**e**), JT21 (**f**), JT18 (**g**), and JT12 (**h**) in response to the titration with AC94377. ΔF is calculated by taking the difference in protein fluorescence of the DMSO control and each ligand concentration. ΔF is plotted *versus* ligand concentration by fitting to nonlinear regression with '*One Site-Nonspecific Binding*' model. n=3. The dissociation constant/K_d (μM) was summarized in Table 2-7.

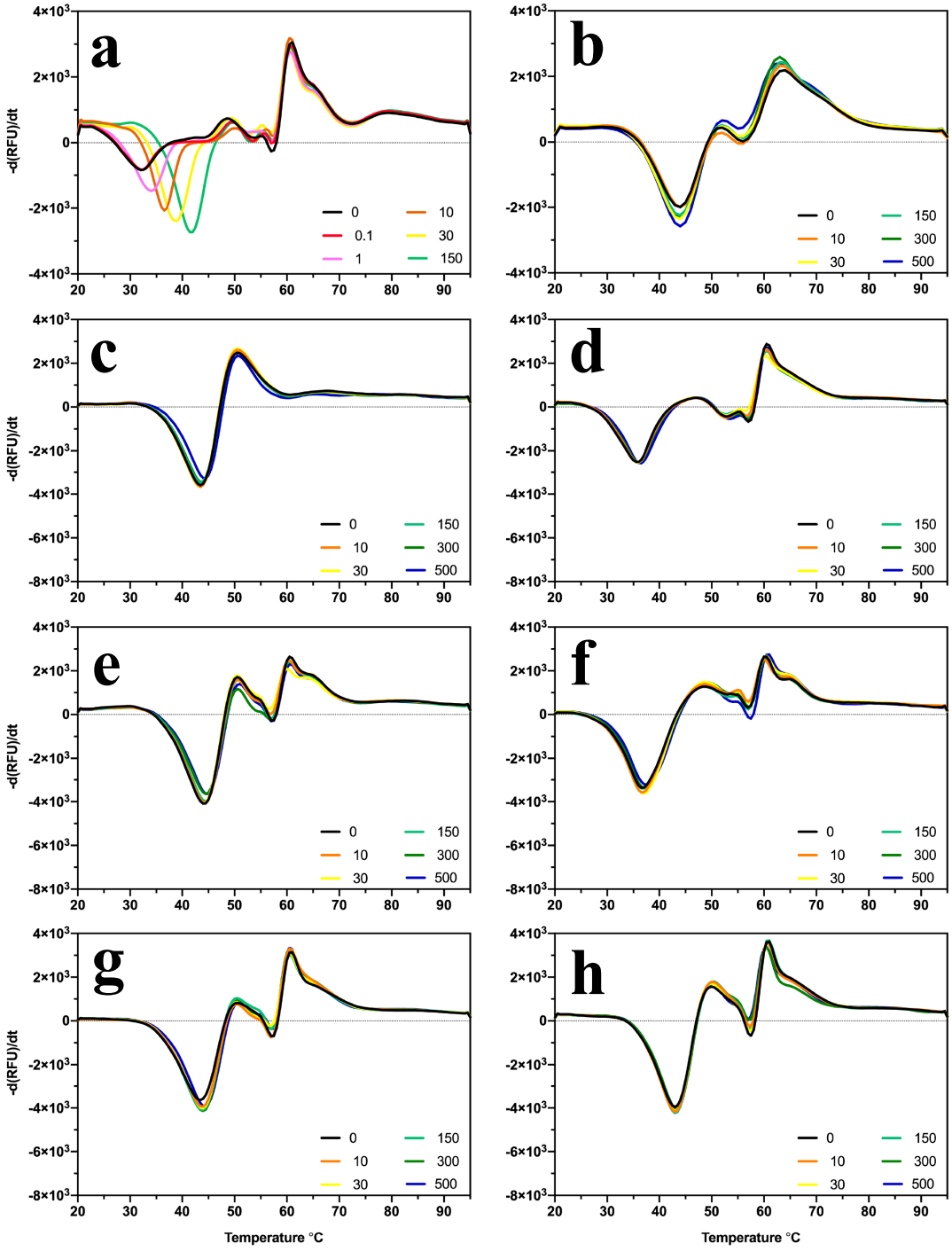


Figure 2-16: Detection of GA_4-induced shift of thermostability of AtGID1a and its variants. Proteins used for assay include: AtGID1a (**a**) and AtGID1a variants [JT8 (**b**), JT30 (**c**), JT59 (**d**), JT50 (**e**), JT21 (**f**), JT18 (**g**), and JT12 (**h**)] by DSF. The negative first derivative of the relative fluorescence units with respect to temperature (-d(RFU)/dt) was calculated by Precision Melt Analysis™ software (Biorad), and further plotted against the temperature in GraphPad Prism 5.0. The representative graphs are shown. n=3. ΔTm are provided in Table 2-8.

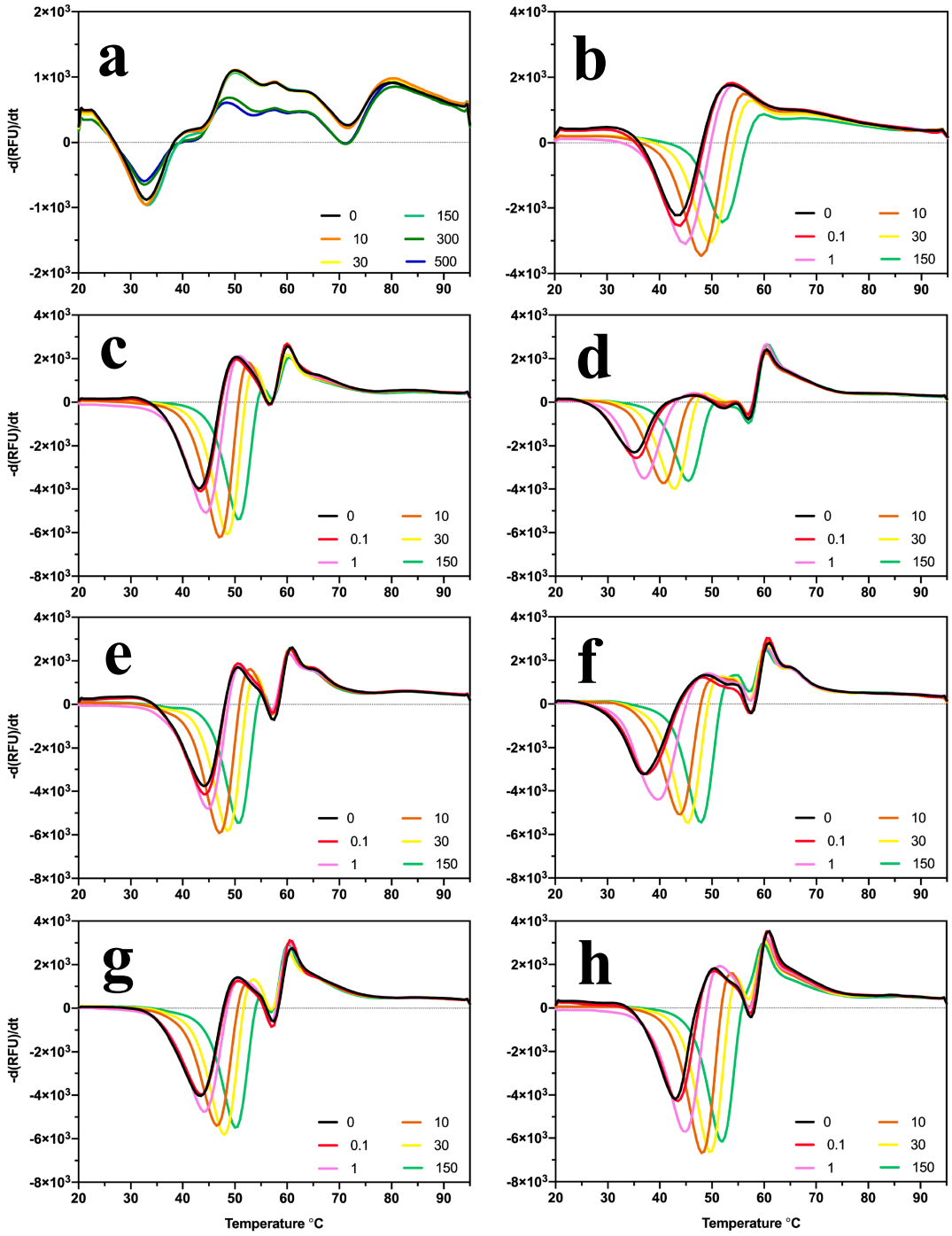
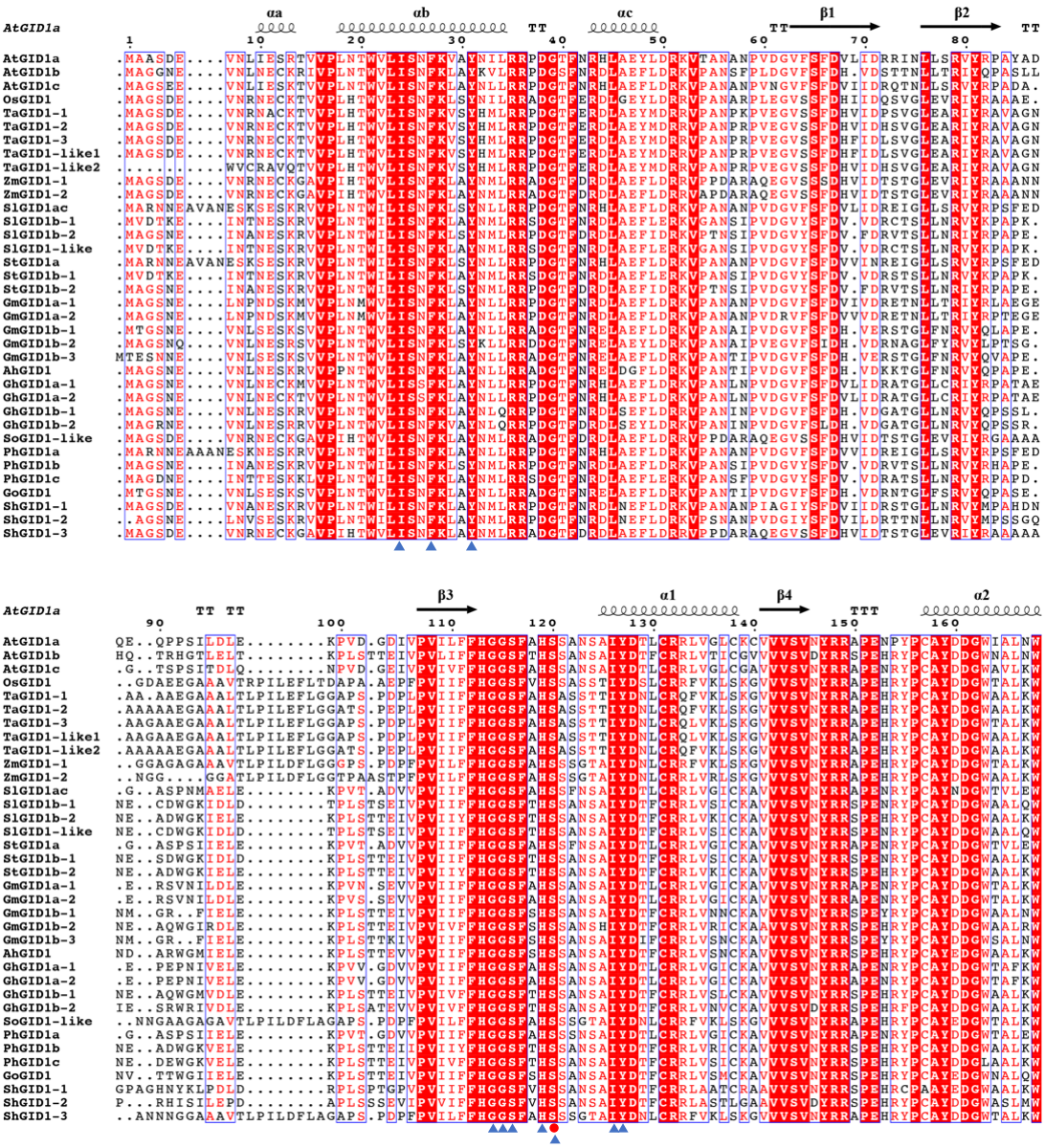


Figure 2-17: Detection of AC94377-induced shift of thermostability of AtGID1a and its variants.

Proteins used for assay include: AtGID1a (**a**) and AtGID1a variants [JT8 (**b**), JT30 (**c**), JT59 (**d**), JT50 (**e**), JT21 (**f**), JT18 (**g**), and JT12 (**h**)] by DSF. The negative first derivative of the relative fluorescence units with respect to temperature (-d(RFU)/dt) was calculated by Precision Melt Analysis^{TM} software (Biorad), and further plotted against the temperature in GraphPad Prism 5.0. The representative graphs are shown. n=3. ΔTm are provided in Table 2-8.


AtGID1a
αa
αb
αc
β1
β2
TT
1 10 20 30 40 50 60 70 80
AtGID1a .MAASDE....VNLIESRTVVPLNTWVLISNFKVAYNILRRPDGTFNTHLAEYLDRKVTANANPVDGVFSFDVLIDRRINLLSRVYRPAYAD
AtGID1b .MAGGNE....VNLNECKRIVPLNTWVLISNFKLAYKVLRRPDGSFNRDLAEFLDRKVPANSFPLDGVFSFDH.VDSTTNLLTRIYQPASLL
AtGID1c .MAGSEE....VNLIESKTVVPLNTWVLISNFKLAYNLLRRPDGTFNTHLAEFLDRKVPANANPVNGVFSFDVIIDRQTNLLSRVYRPADA.
OsGID1 .MAGSDE....VNRNECKTVVPLHTWVLISNFKLSYNILRRADGTFERDLGEYLDRKVPANARPLEGVSSFDHIIDQSVGLEVRIYRAAAE.
TaGID1-1 .MAGSDE....VNRNACKTVVPLHTWVLISNFKVSYHMLRRPDGTFERDLAEYMDRKVPANPKPVEGVSSFDHVIDPSVGLEARIYRAVAGN
TaGID1-2 .MAGSDE....VNRNECKTVVPLHTWVLISNFKVSYHMLRRPDGTFERDLAEYMDRKVPANPRPVEGVSSFDHVIDHSVGLEARIYRAVAGN
TaGID1-3 .MAGSDE....VNRNECKTVVPLHTWVLISNFKVSYHMLRRPDGTFERDLAEYMDRKVPANPRPVEGVSSFDHFIDLSVGLEARIYRAVAGN
TaGID1-like1 .MAGSDE....VNRNECKTVVPLHTWVLISNFKVSYHMLRRPDGTFERDLAEYMDRKVPANPRPVEGVSSFDHFIDLSVGLEARIYRAVAGN
TaGID1-like2 .....WVCRAVQTVVPLHTWVLISNFKVSYHMLRRPDGTFERDLAEYMDRKVPANPRPVEGVSSFDHVIDHSVGLEARIYRAVAGN
ZmGID1-1 .MAGSDE....VNRNECKGAVPIHTWVLISNFKLAYNMLRRADGTFDRDLAEFLDRKVPPDARAQEGVSSSDHVIDTSTGLEVRIYRAAANN
ZmGID1-2 .MAGSDE....VNRNECKGAVPIHTWVLISNFKLAYNMLRRADGTFDRDLAEFLDRKVAPDARAQEGVSSFDHVIDTSTGLEVRIYRAAANN
SlGID1ac .MARNNEAVANESKSESKRVVPLNTWILISNFKLSYNLLRRPDGTFNRHLAEFLDRKVPANANPVDGVFSFDVLIDREIGLLSRVYRPSFED
SlGID1b-1 .MVDTKE....INTNESKRVVPLNTWILISNFKLAYNMLRRSDGTFNRDLAEFLERKVGANSIPVDGVYSFDV.VDRCTSLLNRVYKPAPK.
SlGID1b-2 .MAGSNE....INANESKRVVPLNTWILISNFKLSYNMLRRPDGTFDRDLAEFIDRKVPTNSIPVDGVYSFDV.FDRVTSLLNRIYRPAPE.
SlGID1b-like .MVDTKE....INTNESKRVVPLNTWILISNFKLAYNMLRRSDGTFNRDLAEFLERKVGANSIPVDGVYSFDV.VDRCTSLLNRVYKPAPK.
StGID1a .MARNNEAVANESKSESKRVVPLNTWILISNFKLSYNLLRRPDGTFNRHLAEFLDRKVPANANPVDGVFSFDVVINREIGLLSRVYRPSFED
StGID1b-1 .MVDTKE....INTNESKRVVPLNTWVLISNFKLAYNMLRRSDGTFNRDLAEFLERKVPANSIPVDGVYSFDV.VDRSTSLLNRVYKPAPK.
StGID1b-2 .MAGSNE....INANESKRVVPLNTWILISNFKLSYNMLRRPDGTFDRDLAEFIDRKVPTNSIPVDGVYSFDV.FDRVTSLLNRIYRPAPE.
GmGID1a-1 .MAGSNE....LNPNDSKMVVPLNMWVLISNFKLAYNLLRRPDGTFNRDLAEFLDRKVPANANPVDGVFSFDVIVDRETNLLTRIYRLAEGE
GmGID1a-2 .MAGSNE....LNPNDSKMVVPLNMWVLISNFKLAYNLLRRPDGTFNRDLAEFLDRKVPANANPVDRVFSFDVVVDRETNLLTRIYRPTEGE
GmGID1b-1 .MTGSNE....VNLSESKSVVPLNTWVLISNFKLAYNLLRRADGTFNRDLAEFLDRKVPANAIPVDGVFSFDH.VERSTGLFNRVYQLAPE.
GmGID1b-2 .MAGSNQ....VNLNESRSVVPLNTWVLISNFKLSYKLLRRDDGTFNRELAEYLDRKVPANAIPVEGVFSIDH.VDRNAGLFYRVYLPTSG.
GmGID1b-3 MTESNNE....VNLSESKSVVPLNTWVLISNFKLAYNLLRRADGTFNRELAEFLDRKVPANTIPVDGVFSFDH.VERSTGLFNRVYQVAPE.
AhGID1 .MAGSNE....VNLNESKRVVPPNTWVLISNFKLAYNLLRRADGTFNRELDGFLDRKVPANTIPVDGVFSFDH.VDKKTGLFNRVYQPAPE.
GhGID1a-1 .MAGSNE....VNLNECKMVVPLNTWVLISSFKLAYNLLRRPDGTFNRHLAEFLDRKVPANLNPVDGVFSFDVLIDRATGLLCRIYRPATAE
GhGID1a-2 .MAGSNE....VNLNECKTVVPLNTWVLISSFKLAYNLLRRPDGTFNRHLAEFLDRKVPANLNPVDGVFSFDVLIDRATGLLCRIYRPATAE
GhGID1b-1 .MAGSNE....VNLNESKRVVPLNTWVLISNFKLAYNLQRRPDGTFNRDLSEFLDRKVPANINPVDGVFSFDH.VDGATGLLNRVYQPSSL.
GhGID1b-2 .MAGRNE....VNLNESRRVVPLNTWVLISNFKVAYNLQRRPDGTFDRDLSEYLDRKVPANINPVDGVFSLDH.VDGATGLLNRVYQPSSR.
SoGID1-like .MAGSDE....VNRNECKGAVPIHTWVLISNFKLAYNMLRRADGTFDRDLAEFLDRKVPPDARAQEGVSSFDHVIDTSTGLEVRIYRGAAAA
PhGID1a .MARNNEAAANESKNESKRVVPLNTWILISNFKLAYNLLRRPDGTFNRHLAEFLDRKVPANANPVDGVFSFDVVIDREIGLLSRVYRPSFED
PhGID1b .MAGSNE....INANESKRVVPLNTWILISNFKLSYNMLRRSDGTFNRDLAEFLERKVPANSIPVDGVYSFDV.VDRATSLLNRVYRPAPD.
PhGID1c .MAGDNE....INTTESKKIVPLNTWILISNFKLAYNMLRRSDGTFNRDLAEFLERKVPANSFPVDGVYSFDV.VDRATSLLNRVYRPAPD.
GoGID1 .MTGSNE....VNLSESKSVVPLNTWVLISNFKLAYNLLRRADGTFNRELAEFLDRKVPANTIPVDGVFSFDH.VDRNTGLFSRVYQPASE.
ShGID1-1 .MAGSDE....VNANESKRVVPLNTWILISNFKLAYNMLRRPDGTFNRDLNEFLDRKVPANANPIAGIYSFDVIIDRSTGLLNRVYMPAHDN
ShGID1-2 .MAGSNE....LNVSESKRVVPLNTWILISNFKLAYNMLRRPDGTFNRDLNEFLDRKVPSNANPVDGIYSFDVILDRTTNLLNRVYMPSSGQ
ShGID1-3 .MAGSDE....VNRNECKGAVPIHTWVLISNFKLAYNMLRRADGTFDRDLAEFLDRRVPPDARAQEGVSSFDHVIDTSTGLEVRIYRAAAAA
β3
α1
β4
α2
TT TT
TTT
90 100 110 120 130 140 150 160
AtGID1a QE..QPPSILDLE.....KPVD.GDIVPVILFPHGGSFAHSSANSAIYDTLCRRLVGICKCVVVSVNYRRAPENPYPCAYDDGWIALNW
AtGID1b HQ..TRHGTLELT.....KPLSTTEIVPVLIFPHGGSFTHSSANSAIYDTFCRRLVTICGVVVVSVDYRRSPEHRYPCAYDDGWNALNW
AtGID1c .G..TSPSITDLQ.....NPVD.GEIVPVIVFPHGGSFAHSSANSAIYDTLCRRLVGLCGAVVVSVNYRRAPENRYPCAYDDGWAVLKW
OsGID1 ..GDAEEGAAAVTRPILEFLTDAPA.AEPFPVIIFPHGGSFVHSSASSTIYDSLCRRFVKLSKGVVSVNYRRAPEHRYPCAYDDGWTALKW
TaGID1-1 .AA.AAEGAAALTLPILEFLGGAPS.PDPLPVIIFPHGGSFAHSASSTTIYDNLCRQFVKLSKGVVSVNYRRAPEHRYPCAYDDGWAALKW
TaGID1-2 .AAAAAEGAAALTLPILEFLGGATS.PDPLPVIIFPHGGSFAHSASSTTIYDNLCRQFVKLSKGVVSVNYRRAPEHRYPCAYDDGWAALKW
TaGID1-3 .AAGAAEGAAALTLPILEFLGGAPS.PDPLPVIIFPHGGSFAHSASSTTIYDNLCRQLVKLSKGVVSVNYRRAPEHRYPCAYDDGWTALKW
TaGID1-like1 .AAGAAEGAAALTLPILEFLGGAPS.PDPLPVIIFPHGGSFAHSASSTTIYDNLCRQLVKLSKGVVSVNYRRAPEHRYPCAYDDGWAALKW
TaGID1-like2 .AAAAAEGAAALTLPILEFLGGATS.PEPLPVIIFPHGGSFAHSASSTTIYDNLCRQFVKLSKGVVSVNYRRAPEHRYPCAYDDGWAALKW
ZmGID1-1 ..GGAGAGAAAVTLPILDFLGGGPS.PDPFPVILFPHGGSFAHSSSGTAIYDNLCRFVKLSKGVVSVNYRRAPEHRYPCAYDDGWTALKW
ZmGID1-2 ..NGG....GGATLPILDFLGGTPAASTPFPVILFPHGGSFAHSSSGTAIYDNLCRLVRLSKGVVSVNYRRAPEHRYPCAYDDGWAALKW
SlGID1ac .G..ASPNMAELE.....KPVT.ADVVPVIIFPHGGSFAHSSFNSAIYDTLCRRLVGICKAVVSVNYRRAPENRYPCAYNDGWTVLEW
SlGID1b-1 NE..CDWGKIDLD.....KPLSTSEIVPVIIFPHGGSFTHSSANSAIYDTFCRRLVSICKAVVSVNYRRSPEHRYPCAYDDGWAALQW
SlGID1b-2 NE..ADWGKIELE.....KPLSTTEIVPVIIYPHGGSFTHSSANSAIYDTFCRRLVKICKAVVSVNYRRSPEHRYPCAYDDGWAALKW
SlGID1b-like NE..CDWGKIDLD.....KPLSTSEIVPVIIFPHGGSFTHSSANSAIYDTFCRRLVSICKAVVSVNYRRSPEHRYPCAYDDGWAALQW
StGID1a .G..ASPSIIELE.....KPVT.ADVVPVIIFPHGGSFAHSSFNSAIYDTLCRRLVGICKAVVSVNYRRAPENRYPCAYDDGWTVLEW
StGID1b-1 NE..SDWGKIDLD.....KPLSTTEIVPVIIFPHGGSFTHSSANSAIYDTFCRRLVSICKAVVSVNYRRSPEHRYPCAYDDGWAALKW
StGID1b-2 NE..ADWGKIELE.....KPLSTTEIVPVIIYPHGGSFTHSSANSAIYDTFCRRLVKICKAVVSVNYRRSPEHRYPCAYDDGWAALKW
GmGID1a-1 .E..RSVNILDLE.....KPVN.SEVVPVIIFPHGGSFAHSSANSAIYDTLCRRLVGICKAVVSVNYRRAPENRYPCAYDDGWTALKW
GmGID1a-2 .E..RSVNILDLE.....KPVS.SEVVPVIIFPHGGSFAHSSANSAIYDTLCRRLVGICKAVVSVNYRRAPENRYPCAYDDGWTALKW
GmGID1b-1 NM..GR..FIELE.....KPLSTTEIVPVIIFPHGGSFSHSSANSAIYDTFCRRLVNNCKAVVSVNYRRSPEYRYPCAYDDGWAALKW
GmGID1b-2 NE..AQWGIRDLE.....KPLSTTEIVPVIVFPHGGSFSHSSANSHIYDTFCRRLVRICKAAVSVNYRRSPEHRYPCAYDDGWAALRW
GmGID1b-3 NM..GR..FIELE.....KPLSTTKIVPVIIFPHGGSFSHSSANSAIYDIFCRRLVSNCKAVVSVNYRRSPEYRYPCAYDDGWSALNW
AhGID1 ND..ARWGMIELE.....KPLSTTEVVPVIIFPHGGSFSHSSANSAIYDTFCRRLVSNCKAVVSVNYRRSPEHRYPCAYDDGWAALKW
GhGID1a-1 .E..PEPNIVELE.....KPVV.GDVVPVIIFPHGGSFAHSSANSAIYDTLCRRLVGICKAVVSVNYRRAPENRYPCAYDDGWTAFKW
GhGID1a-2 .E..PEPNIVELE.....KPVV.GDVVPVIIFPHGGSFAHSSANSAIYDTLCRRLVGICKAVVSVNYRRAPENRYPCAYDDGWTAFKW
GhGID1b-1 NE..AQWGMVDLE.....KPLSTTEIVPVIVFPHGGSFTHSSANSAIYDTFCRRLVSLCKAVVSVNYRRSPEHRYPCAYDDGWAALKW
GhGID1b-2 IE..SRWRIVDLE.....KPLSATEVVPVIVFPHGGSFTHSSANSAIYDTFCRRLVNVCKAVVSVDYRRSPEHRFPCAYDDGWAALKW
SoGID1-like ..NNGAAGAGAVTLPILDFLAGAPS.PDPFPVILFPHGGSFAHSSSGTAIYDNLCRFVKLSKGVVSVNYRRAPEHRYPCAYDDGWTALKW
PhGID1a .G..ASPSIIELE.....KPVT.GDVVPVIIFPHGGSFAHSSSNSAIYDTLCRRLVGICKAVVSVNYRRAPENRYPCAYDDGWTALEW
PhGID1b NE..ADWGKVELE.....KPLSTTEIIPVIIYPHGGSFTHSSANSAIYDTFCRRLVRICKAVVSVNYRRSPEHRYPCAYDDGWAALKW
PhGID1c NE..DEWGKVELE.....KPLSTTEIVPVIVFPHGGSFTHSSANSAIYDTFCRRLVSICKAVVSVNYRRSPEHRYPCAYDDGLAALKW
GoGID1 NV..TTWGIIELE.....KPLSTTEIVPVIIFPHGGSFSHSSANSAIYDTFCRRLVSMCKAVVSVNYRRSPEHRYPCAYSDGWNALQW
ShGID1-1 GPAGHNYKLPDLD.....RPLSPTGPVPVIIFPHGGSFVHSSANSAIYDTFCRRLAATCRAAVSVNYRRSPEHRCPAAYSDGWAALKW
ShGID1-2 P...RHISILEPD.....APLSSSEVIPVVIFPHGGSFVHSSANSAIYDTFCRRLASTLGAAVVSVDYRRSPEHRYPCAYDDGWASLMW
ShGID1-3 ..ANNNGGAAAVTLPILDFLAGAPS.PDPFPVILFPHGGSFAHSSSGTAIYDNLCRFVKLSKGVVSVNYRRAPEHRYPCAYDDGWTALKW

(Continued)

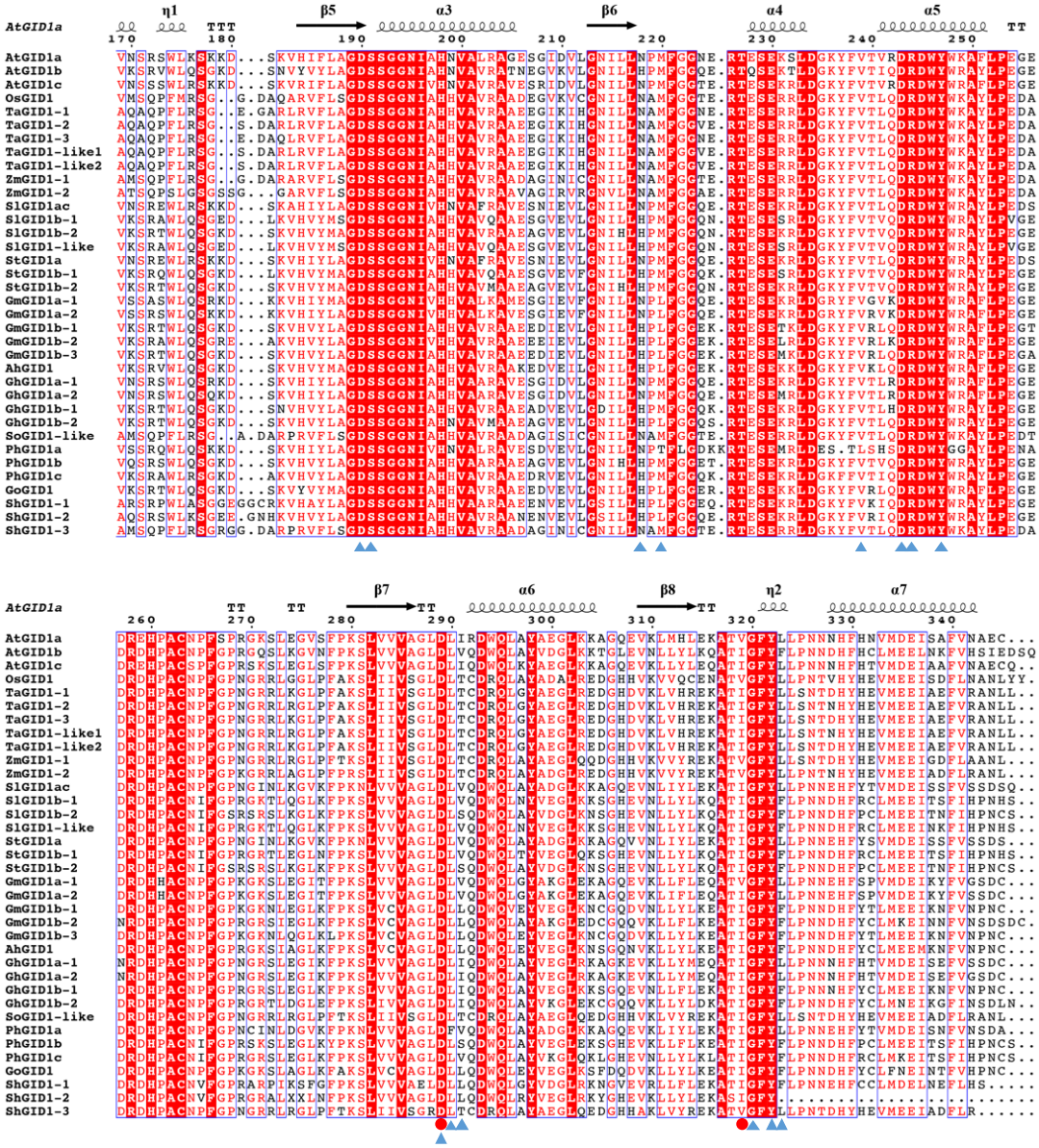


Figure 2-18: The sequence alignment of GID1 proteins.

The sequences of GID1 protein are collected from Arabidopsis (At), Oryza sativa (Os, rice), Triticum aestivum (Ta, wheat), Zea mays (Zm, maize), Solanum lycopersicum (Sl, tomatoes), Solanum tuberosum (St, potatoes), Glycine max (Gm, soybeans), Arachis hypogaea (Ah, peanut), Gossypium hirsutum (Gh, cotton), Saccharum officinarum (So, sugarcane), Petunia x hybrida (Ph), Galega orientalis (Go), and Striga hermonthica (Sh, weeds). The three catalytic residues of AtGID1a, S120, D289 and V319, are indicated by red dots. The residues involved in shaping active pocket of AtGID1a are indicated by blue triangles.

It should be noted that a non-conserved amino acid C-terminal extension of GhGID1a-2 (MNYGFLSSFFQKDLFLVLGFVFECLKERG) and AtGID1b (SKSSPVLLTP) have been omitted in the alignment; three GID1 sequences from Striga are retrieved from the master thesis of Mr. CHEN (University of Tokyo).

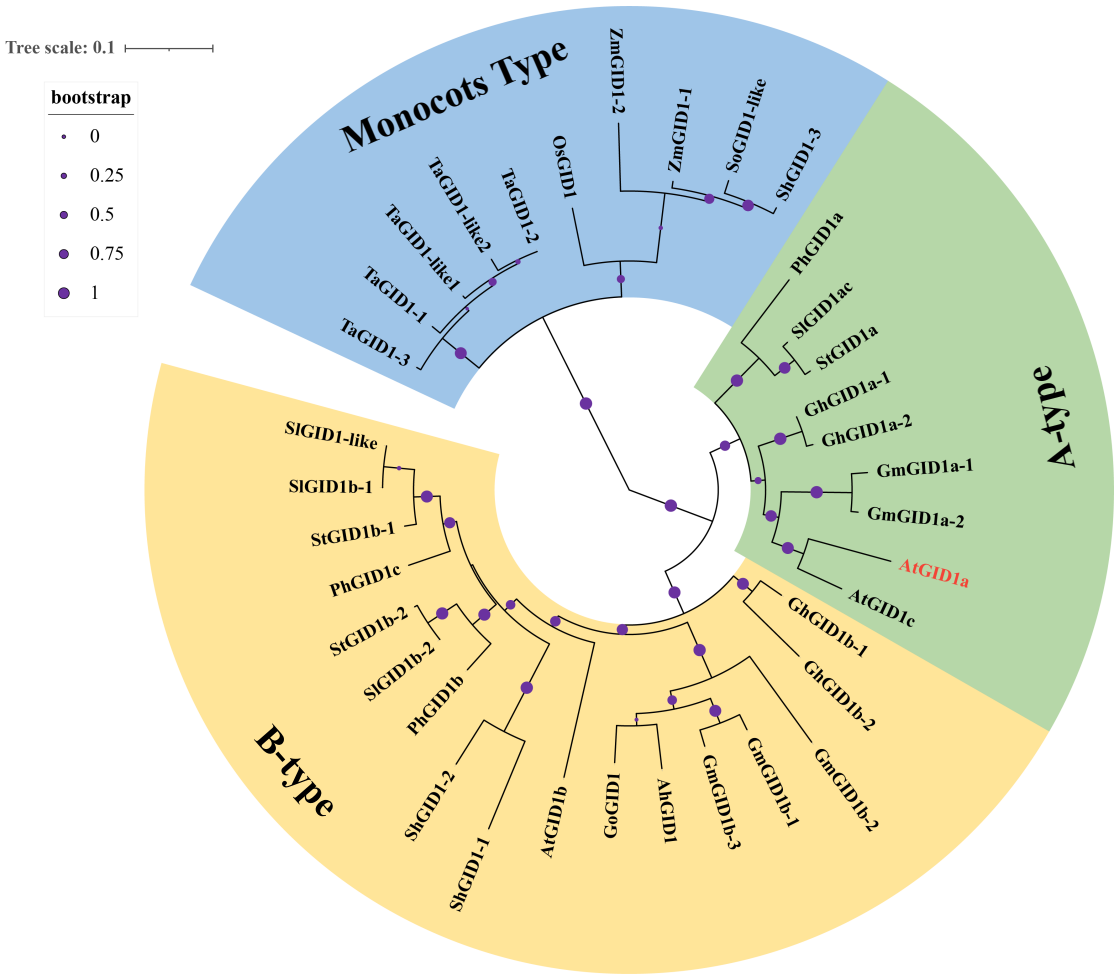


Figure 2-19: Maximum likelihood phylogeny of GID1 proteins.
Phylogenetic tree was constructed by Phylogeny.fr based on the sequence alignment of GID1 proteins from Arabidopsis (At), Oryza sativa (Os, rice), Triticum aestivum (Ta, wheat), Zea mays (Zm, maize), Solanum lycopersicum (Sl, tomatoes), Solanum tuberosum (St, potatoes), Glycine max (Gm, soybeans), Arachis hypogaea (Ah, peanut), Gossypium hirsutum (Gh, cotton), Saccharum officinarum (So, sugarcane), Petunia x hybrida (Ph), Galega orientalis (Go), and Striga hermonthica (Sh, weeds). Bootstrap (purple dots) is applied to indicate the confidence levels in a clade of phylogenetic tree. Three major types of GID1 are generated. A-type GID1 proteins are colored green; B-type GID1 proteins are colored yellow; Monocots type GID1 proteins are colored blue.

2.3 Conclusion and perspectives

Biologically active GAs have been demonstrated to control diverse aspects of plant growth and development, such as seed germination, stem elongation, pollen maturation, and transition from vegetative growth to flowering. In plants, the perception of GA is mediated by its receptor GID1 protein, the resultant formation of GA-bound GID1 complex leads to a key event to initiate GA response through the degradation of its repressor DELLA proteins. There encoded three GA receptors and five repressors in the genome of *Arabidopsis*. In our previous work, we have revealed that these three receptors (AtGID1a, AtGID1b, and AtGID1c) all have the ability to bind with bioactive GA_4, and each activated GID1s could recruit all five DELLA proteins so that define GA response by releasing the repression on the expression of GA-related genes. This complicated regulatory network, contributed by the typical gene duplication event, is of great importance for plant adaptation in response to the internal or external challenges during the long term of evolution. Recently, we have found that chemical AC94377 displays GA-like activity through the canonical GA signaling pathways both in Arabidopsis and rice. However, the functional overlapping of receptors and repressors greatly impedes our knowledge of specific GA responses through the spatiotemporal regulation by using AC94377 in facing different stimuli. It is believed that an orthogonal high-affinity protein-ligand variant pair, which is a chemical genetic approach by integrating chemistry and biology, have great potential to address such problems.

In this study, since the moderate affinity of AC94377 to GID1, I am unable to solve the molecular interactions between them. Using a serial of AC94377 derivatives, I have basically unraveled the structure and activity relationships of AC94377 and its derivatives, and these results can be partially explained by using docking simulation models of AC94377-GID1s. Since the absence of a precise structure complex, I performed the AC94377-directed laboratory evolution of AtGID1a by using the Y2H screening system with the aim of developing the 'functionally' orthogonal high-affinity $GID1^{variant}$-AC94377 pairs to flexibly regulate the specific biological function of GID1 in Arabidopsis. Moreover, this orthogonal pairs with high affinity and selectivity will also benefit our subsequent efforts of co-crystallization on $GID1^{variant}$-AC94377 once more, which will inspire us to design a more potent and broad-spectrum GA agonist for agricultural application and research use.

After the biochemical characterization of these beneficial AtGID1a variants by DSF assay, I am

unexpected to find that their thermostability is also greatly enhanced, compared with wild type (Figure 2-16; Figure 2-17; Table 2-8). Since these mutations all exist in the deep pocket of GID1, I speculate that it is through strengthening the molecular interactions between mutated residues that the structural scaffold of variant proteins becomes more stable. What' more, the heterologous expression level of variants is also about 7-folds higher than wild type in *E.coli* cells. These may explain why Y2H assay suggested the super-sensitivity of AtGID1a variants to AC94377, which is particularly 30 folds-higher than that of wild type to GA_4 (Figure 2-13), whereas affinity assay revealed that binding affinity of AtGID1a variants to AC94377 is not as strong as that of wild type to GA_4 (Figure 2-14; Figure 2-15; Table 2-7). The high and stable expression level in yeast cells would partially or even totally offset the weak protein-protein interaction of the relatively low affinity. Therefore, most importantly, we are going to apply this orthogonal ligand-receptor system in plants, especially in Arabidopsis, to verify their functions in native environment. We are expected to see whether these GID1 variants can be also stably expressed in plant cells, and what these biochemical features offer plants?

The 1st generation of 'Green Revolution' was achieved by creating the semi-dwarf varieties insensitive to GA or deficient in GA biosynthesis, which improved crop yields by reducing the plant height[127, 193]. In this context, we are wondering through applying the GID1 variants in *gid1* mutant plants whether this orthogonal high-affinity pair would uncouple the trade-offs between plant growth/development and crops production. Moreover, this tool may also help to answer the in-depth genetic basis underlying the importance of the differences in spatial-temporal expression of GID1 proteins for the growth and development of Arabidopsis.

To summarize, we combined the approaches of chemical biology and synthetic biology to develop the orthogonal AtGID1a variant-AC94377 pairs with the great potentiality of probing GA biology in plants.

Table 2- 1: The different truncations of GID1 and DELLA proteins used for crystallization.

Protein Constructs	Sequence Range	Mutation	Note
AtGID1a1	1-345	-	Full length
AtGID1a2	1-344	-	Removal of Cys345
AtGID1a2v	1-344	V319I	
AtGID1a4	15-344	-	Removal of Cap domain αa
AtGID1a7	1-344	V7T, I10D, L18D, V29T	Hydeophilic residues
AtGID1c2	1-342	-	Removal of Cys345
AtGID1c4	15-342	-	Removal of Cap domain αa
AtGID1c7	1-342	V7T, I10D, L18D, L29D	Hydeophilic residues
OsGID1-1	1-354	-	Full length
OsGID1-2	15-352	-	Removal of Cap domain αa
OsGID1-2v	15-352	V326I	
AtGAI3	11-97	-	DELLA domain
AtGAI5	1-113	-	
AtGAI7	22-97	-	
OsSLR1-1	28-112	-	

Table 2- 2: Crystals observation of GID1a2^{V319I}-AC94377-AtGAI5 and GID1c2-AC94377-AtGAI5 complex.

Samples	Concentration		Temperature (ºC)	Conditions	
	Protein (mg/mL)	Ligand (mM)		Commercial Kits	Positions
AtGID1a2^{V319I}-AC94377-AtGAI5	40-50	1	4	JCSG I	1A, 1B, 1D, 2D, 1G, 2A, 9B,11B, 12A, 1B
				JCSG II	5C, 7B
				PACT suit	2E, 2H, 3G, 3H, 5H, 6H, 9D, 8D, 5G, 12D, 12G
AtGID1c2-AC94377-AtGAI5	30-50	1	4	PACT suit	1G

Table 2- 3: Biochemical characterization of OsGID1 with AC94377 and its derivatives by ITF and DSF.

K_d value is calculated by fitting ΔF/ΔFmax against ligand concentration using nonlinear regression with '*One Site-Nonspecific Binding*' model in GraphPad Prism 5.0. ΔTm is calculated by subtracting the melting temperature of protein with chemical at 300 μM to that at 0 μM (DMSO). Data are means ± SD (n = 3).

Chemicals	R_1	R_2	Dissociation Constant/K_d (μM)		ΔTm (°C)	
			Trx-OsGID1	Trx-OsGID1^{V326I}	Trx-OsGID1	Trx-OsGID1^{V326I}
GA_4	-	-	4.7±0.5	3.7±0.5	11.50±0.29	10.67±0
AC94377	Cl	NH_2	79.0±14.4	25.2±2.9	2.5±0	3.33±0.29
ACOH	Cl	OH	300.4±65.0	450.6±95.1	1.83±0.29	1.67±0.29
ACOIp	Cl	$OCH(CH_3)_2$	181.7±47.4	572.6±370.0	1.50±0.29	0.83±0
JK002	H	NH_2	295.9±121.1	227.6±29.1	-0.17±0.29	0±0
ACF	F	NH_2	274.2±56.4	245.3±59.1	1.67±0	2.17±0

Table 2- 4: Calculation of screening effort when applying NNK and NDT primers.

Number of residues at one randomization site	NNK		NDT	
	Codons	Transformants needed	Codons	Transformants needed
1	32	96	12	36
2	1024	3068	144	431
3	3.3E+04	9.8E+04	1.7E+03	5.2E+03
4	1.0E+06	3.1E+06	2.1E+04	6.2E+04
5	3.4E+07	1.0E+08	2.5E+05	7.5E+05
6	1.1E+09	3.2E+09	3.0E+06	8.9E+06
7	3.4E+10	1.0E+11	3.6E+07	1.1E+08
8	1.1E+12	3.3E+12	4.3E+08	1.3E+09
9	3.5E+13	1.1E+14	5.2E+09	1.5E+10
10	1.1E+15	3.4E+15	6.2E+10	1.9E+11

NNK: encoding all 20 canonical amino acids;

NDT: encoding 12 amino acids: Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, and Gly;

"Transformants needed" is calculated by the following formula: transformants needed = -codons × ln (1–95%). All values are calculated based on the 95% library coverage.

Table 2- 5: Degenerate nucleotide codes.

UPAC nucleotide code	Degenerated bases	
A	Adenine	A
C	Cytosine	C
G	Guanine	G
T	Thymine (DNA)	T
U	Uracil (RNA)	U
W	Weak	A/T
S	Strong	C/G
M	Amino	A/C
K	Keto	G/T
R	Purine	A/G
Y	Pyrimidine	C/T
B	Not A	C/G/T
D	Not C	A/G/T
H	Not G	A/C/T
V	Not T	A/C/G
N	Any	A/C/G/T

Table 2- 6: Mapping beneficial mutation in sites E, G, I, and J.

The sensitivity of variants to AC94377 and GA_4 were indicated based on the Y2H assay (Figure 2-12, 2-13). n = 3.

Variants code		Mutation									Sensitivity [μM]		
		Site E			Site G			Site I		V319	AC94377	GA_4	
		I24	F27	Y31	D190	S191	D289	L290	I291				
WT		I	F	Y	D	S	D	L	I	V	50	0.03	
Site E	ET2	L	Y	-	-	-	-	-	-	I	1	n.d	Round 1: Combinatorial Saturation Mutagenesis
	ET3	L	L	-	-	-	-	-	-	I	1	n.d	
	ET4	L	-	-	-	-	-	-	-	I	1	n.d	
Site G	GT1	-	-	-	-	D	-	-	-	I	1	n.d	
	GT3	-	-	-	-	C	-	-	-	I	1	n.d	
	GT4	-	-	-	-	E	-	-	-	I	1	n.d	
Site I	IT2	-	-	-	-	-	N	I	V	I	1	n.d	
	IT4	-	-	-	-	-	S	I	V	I	10	n.d	
	IT5	-	-	-	-	-	N	V	S	I	10	n.d	
	IT6	-	-	-	-	-	N	-	V	I	10	n.d	
	IT8	-	-	-	-	-	-	I	L	I	≥1	n.d	
	IT9	-	-	-	-	-	N	I	S	I	≥1	n.d	
Site J	JT39	L	Y	-	-	D	-	-	T	I	0.1	×	Round 2: Iterative Random Mutagenesis
	JT2	L	-	-	-	D	-	V	L	I	≥0.03	×	
	JT11	L	L	-	-	D	N	V	T	I	≥0.01	50	
	JT5	L	L	-	-	E	-	-	-	I	0.01	×	
	JT4	L	-	-	-	E	-	V	L	I	0.01	×	
	JT8	L	-	-	-	E	-	I	L	I	≥0.003	×	
	JT30	L	Y	-	-	E	-	I	L	I	≥0.003	×	
	JT59	L	-	-	-	C	-	I	V	I	0.003	50	
	JT50	L	Y	-	-	E	-	V	T	I	≥0.001	≥100	
	JT21	L	L	-	-	E	-	V	-	I	≥0.001	≥50	
	JT18	L	-	-	-	E	-	V	S	I	0.001	≥1	
	JT12	L	L	-	-	E	-	I	L	I	0.001	≥1	

"-" indicates no mutation at this position.

"n. d." indicates "not determined".

"≥" suggests the weak interaction under the indicated concentration of ligand.

"×" indicates that variants cannot respond to GA_4 even at 100 μM.

Table 2- 7: The dissociation constant/K_d (μM) for GA_4 and AC94377 binding to AtGID1a and its variants.

Affinity assay is determined by ITF. All K_d values are calculated using the nonlinear regression with '*One Site-Nonspecific Binding*' model in GraphPad Prism 5.0. Data are means ± SD (n = 3). Corresponding fitting graphs are provided in Figure 2-14 and 2-15.

Variant Code	Dissociation constants/K_d (μM)	
	GA_4	AC94377
AtGID1a (wt)	0.71 ± 0.48	1333.2 ± 397.79
JT8	N. A.	3.42 ± 0.35
JT30	N. A.	2.39 ± 0.43
JT59	N. A.	1.54 ± 0.95
JT50	N. A.	2.4 ± 0.63
JT21	N. A.	1.66 ± 0.28
JT18	N. A.	2.46 ± 0.41
JT12	N. A.	2.45 ± 0.37

N. A.: Not available (GraphPad Prism cannot provide the reliable dissociation constant (K_d) based on the fitted graphs.

Table 2- 8: The melting temperature of AtGID1a and its variants in respond to GA_4 and AC94377.

ΔTm is calculated by subtracting the melting temperature of protein with chemical at 30 or 150 μM to that at 0 μM (DMSO). Data are means ± SD (n = 3).

Variant Code	Chemical Con. (μM)	GA_4 (ºC)		AC94377 (ºC)	
		Tm	ΔTm	Tm	ΔTm
Wild type	0	31.0 ± 1.3	-	32.0 ± 0	-
	30	38.7 ± 0.3	7.7 ± 1.2	32.0 ± 0	0.0 ± 0
	150	41.7 ± 0.3	10.7 ± 1.6	32.3 ± 0.6	0.3 ± 0.6
JT8	0	44.0 ± 0	-	43.3 ± 0.6	-
	30	44.0 ± 0	0.0 ± 0	50.0 ± 0	6.7 ± 0.6
	150	44.0 ± 0	0.0 ± 0	52.0 ± 0	8.7 ± 0.6
JT30	0	43.5 ± 0	-	43.2 ± 0.3	-
	30	43.7 ± 0.3	0.2 ± 0.3	48.7 ± 0.3	5.5 ± 0
	150	43.7 ± 0.3	0.2 ± 0.3	50.7 ± 0.3	7.5 ± 0
JT59	0	35.7 ± 0.3	-	35.2 ± 0.3	-
	30	35.7 ± 0.3	0.0 ± 0	42.8 ± 0.3	7.7 ± 0.3
	150	35.8 ± 0.3	0.2 ± 0.3	45.5 ± 0	10.3 ± 0.3
JT50	0	44.2 ± 0.3	-	44.2 ± 0.3	-
	30	44.2 ± 0.3	0.0 ± 0	48.8 ± 0.3	4.7 ± 0.3
	150	44.5 ± 0	0.3 ± 0.3	50.8 ± 0.3	6.7 ± 0.3
JT21	0	36.7 ± 0.3	-	37.2 ± 0.3	
	30	37.0 ± 0.5	0.3 ± 0.3	45.5 ± 0	8.3 ± 0.3
	150	37.2 ± 0.3	0.5 ± 0	48.0 ± 0	10.8 ± 0.3
JT18	0	43.7 ± 0.3	-	43.5 ± 0	-
	30	43.8 ± 0.3	0.2 ± 0.3	48.2 ± 0.3	4.7 ± 0.3
	150	43.8 ± 0.3	0.2 ± 0.3	50.2 ± 0.3	6.7 ± 0.3
JT12	0	43.0 ± 0	-	43.0 ± 0	-
	30	43.0 ± 0	0.0 ± 0	49.8 ± 0.3	6.8 ± 0.3
	150	43.0 ± 0	0.0 ± 0	52.0 ± 0	9.0 ± 0

Table 2- 9: The oligonucleotides used to make GID1 and DELLA constructs for crystallization.

The stop codons are highlighted with blue bold. The mutated bases are colored green. All constructs using *Bam*HI and *Sac*I as the restriction sites with the exception of SLRI, which use *Bam*HI and *Hin*dIII as the restriction sites. The restriction site sequences are colored red.

Proteins	Primer Name	Oligo Sequence (5' → 3')	Sequence range
AtGID1a	AGa2_F	CTAGGGATCCGATGGCTGCGAGCGATGAAACT	1-344
	AGa2_R	CATGGAGCTCTTATTCCGCGTTTACAAACGCCG	
	AGa4_F	CTAGGGATCCGGTGGTTCCTCTCAATACATGGGTTTTAATATCCA	15-344
	AGa4_R	AGa2_R	
	AGaV319I_F	ATTGGGTTTTACCTCTTGCCTAATAACAATCAT	V319I
	AGaV319I_R	AGTTGCTTTCTCTAAATGCATAAGCTTAACCTCT	
AtGID1c	AGc2_F	CTAGGGATCCATGGCTGGAAGTGAAGAAGTTAATCTTATTGAG	1-342
	AGc2_R	CCGGAGCTCTTATTCTGCGTTTACAAATGCAGCTATCTCATCC	
	AGc4_F	CTAGGGATCCGGTGGTTCCTCTCAATACATGGGTTC	15-342
	AGc4_R	AGc2_R	
OsGID1	OsG2_F	CTAGGGATCCGGTGGTGCCGCTCCACACATGGG	15-352
	OsG2_R	CATGGAGCTCTTAGAGGTTAGCGTTGAGGAAGTCGGAGATC	
	OsG2V326I_F	ATCGGGTTCTACCTGTTGCCC	V326I
	OsG2V326I_R	CGTGGCGTTCTCGCATTGGA	
	OsG3_F	GCTAGTGGATCCATGGCCGGCAGCGACGAGGTCAAC	1-352
	OsG3_R	CATGGAGCTCTTAGAGGTTAGCGTTGAGGAAGTCGGAGATCTC	
DELLA	GAI3_F	CTAGGGATCCGCAAGATAAGAAGACTATGATGATGAATG	11-97
	GAI3_R	CATGGAGCTCTTAGTTAGACGACGGAGGATTAAGG	
	GAI7_F	CTAGGGATCCGGACGACGGTAACGGCATGGATG	22-97
	GAI7_R	CATGGAGCTCTTAATTAAGGTCGGTGAGCATAGAATCAAGC	
	SLR1-1_F	CTAGGGATCCGATGGCGGGGGCGGCGGGG	28-112
	SLR1-1_R	CATGAAGCTTTTACGCGTTGAGCTCGGAAAGCATGCTC	

Table 2- 10: The oligonucleotides used to make variant libraries of AtGID1a.

The degenerate bases are highlighted with red bold. For the amino acids encoded by degenerate primers of each sites, "X" indicates all 20 canonical amino acids; Z indicates 12 amino acids with balanced nature (F, L, I, V, Y, H, N, D, C, R, S, and G). "Transformants needed" is calculated by the following formula: -codons × ln (1–95%).

Sites/Sublibraries	Primer Name	Oligo Sequence (5' → 3')	Amino acids encoded	Transformats needed	
Site A	AGaSiteA_F	CAT**NNKNNK**AGCTTTGCTCATTCTTCTGCAAAC	G114X, G115X	3068	
	AGaSiteA_R	GAAGAACAATATAACAGGGACAATGTCGCCATC			
Site B	AGaSiteB_F	**NDT**TTTGCT**NDTNDT**TCTGCAAACAGTGCCATCTACG	S116Z, H119Z, S120Z	5177	
	AGaSiteB_R	ACCTCCATGGAAGAACAATATAACAGGGAC			
Site C	AGaSiteC_F	**NDT**TTT**NTDNDT**TTGCCTAATAACAATCATTTCCATAATG TTATGGATGAGATTTCGG	G320Z, Y322Z, L323Z	5177	
	AGaSiteC_R	AACAGTTGCTTTCTCTAAATGCATAAGCTTAACCT			
Site D	AGaSiteD_F	**NNKNNK**GATACTCTTTGTCGCAGGCTTGTTGGTTTGTG	I126X, Y127X	3068	
	AGaSiteD_R	GGCACTGTTTGCAGAAGAATGAGCAAAG			
Site E	AGaSiteE_F	**NDT**TCCAAC**NDT**AAAGTAGCC**NDT**AATATCCTTCGTCGC CCTG	I24Z, F27Z, Y31Z	5177	
	AGaSiteE_R	TAAAACCCATGTATTGAGAGGAACCACTGT			
Site F	AGaSiteF_F	**NDT**ACGGTTAGA**NDTNDT**GATTGGTACTGGAAAGCGT	V239Z, D243Z, R244Z	5177	
	AGaSiteF_R	AAAGTATTTCCCATCCAAACTTTTCTCAGACTCCG			
Site G	AGaSiteG_F	**NNKNNK**TCTGGAGGTAACATCGCGCATAATGT	D190X, S191X	3068	
	AGaSiteG_R	ACCAGCCAAGAAAATATGGACCTTTGAGTC			
Site H	AGaSiteH_F	**NNK**CCT**NNK**TTTGGTGGGAATGAGAGAACG	N218X, M220X	3068	
	AGaSiteH_R	CAGCAGAATGTTCCCCAAAACATCGATTC			
Site I	AGaSiteI_F	**NDTNDTNDT**AGAGATTGGCAGTTGGCATACGCGGAAGG	D289Z, L290Z, I291Z	5177	
	AGaSiteI_R	CAAACCCGCGACAACCACAAGAC			
Site J	AGaSiteE_R2_F	**MTT**TCCAAC**YWC**AAAGTAGCCTACAATATCCTTCGTC	I24L/I, F27L/Y/F, Y31Y	24	27609
	AGaSiteE_R2_R	AGaSiteE_R			
	AGaSiteG_R2_F	GAT**DRS**TCTGGAGGTAACATCGCGCATAAT	D190D, S191D/C/E	36	
	AGaSiteG_R2_R	AGaSiteG_R			
	AGaSiteI_R2_F	**VTANYC**AGAGATTGGCAGTTGGCATACGC	D289/N/D/S, L290/I/V/L, I291/V/L/S/I	288	
	AGaSiteI_R2_R	**GYY**CAAACCCGCGACAACCACAAG			

2.4 Materials and methods

2.4.1 Chemicals

Chemical GA_4 and AC94377 were offered by Prof. Tadao Asami, University of Tokyo. All AC94377 derivatives were provided by Dr. Kai Jiang, University of Tokyo.

2.4.1 Plasmid constructs

The gene encoding full length *Arabidopsis thaliana* AtGID1a, AtGID1c, AtGAI and *Oryza sativa* OsGID1, OsSLR1 were provided by Prof. Kohji Murase, University of Tokyo. All of protein constructs used for crystallization in this study were summarized in Table 2-1, and the primers are shown in Table 2-8. Specifically, the genes of variants AtGID1a7 and AtGID1c7 were synthesized by GENEWIZ (Tokyo, Japan), and that of variants AtGID1a2v and OsGID1-2v were prepared by site-directed mutagenesis according to the protocol of KOD-Plus-Mutagenesis Kit (TOYOBO, Japan). All truncated proteins were amplified with primers containing the *Bam*HI site at 5' end and *Sac*I site at 3' end by the regular PCR using KOD-Plus-neo (TOYOBO, Japan).

For crystallization, AtGID1a2 and AtGID1c2 were cloned into pCDFDuet-1 (Novagen), respectively generating pCDFa2 and pCDFc2, and AtGAI fragments encoding different length of DELLA domain were constructed into to pET-47b+ (Novagen), resulting in pET47-GAI. OsGID1-2 were cloned into pET-47b+ (Novagen) with 6×His tag at N-terminus. For intrinsic tryptophan fluorescence assay and differential scanning fluorimetry, AtGID1a2, AtGID1c2 and OsGID1-1 were cloned into pET-32+ (Novagen) with N-terminal Trx-6×His tag. For Y2H assay, AtGID1a2 and OsGID1-1 were inserted into pGBKT7 (Addgene), and full length of GAI and SLR1 were inserted into pGADT7 (Addgene).

2.4.2 Protein preparation and purification

The all proteins were expressed in *Escherichia coli* Rosetta 2(DE3) stains (Novagen). Particularly, for the purpose of co-crystallization, the AtGID1 proteins (pCDFa2 or pCDFc2) and AtGAI fragments (pET47-GAI) were co-transformed and co-expressed in *E.coli* Rosetta 2. For protein production, 20 mL of overnight culture was inoculated into 4 liters of sterile Luria Bertani (LB) medium supplemented with kanamycin (working con. 25 μg/mL) and streptomycin (working

con. 50 μg/mL). The cells were grown at 37 °C until the optical density at 600 nm (OD_{600}) reached to 0.8-01. Afterwards, protein expression was induced by adding 0.1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) (SigmaAldrich) at 16 °C for 20 h (110 rpm). Cells were then harvested and resuspended in lysis buffer [20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM imidazole, 1 mM DTT]. After the cells were lysed by sonication at 4 °C, the cell pellets were removed by ultracentrifugation at 3,5000g for 30 min. The resultant supernatants were incubated with 5 mL Ni-NTA agarose (Fujifilm) on the ice for 40 min. The resin-supernatant mixture was then filled back to the open column and washed by 100 mL of lysis buffer. 30 mL of elution buffer [20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 200 mM imidazole, 1 mM DTT] was used to elute the target protein, which was then added with 100 mg HRV 3C protease solution for tag removal at 4 °C overnight. After 12 hours later, the protein sample was further purified by ion-exchange chromatography to remove 6×His-tag and impurities using HiTrap™ Q HP column (GE Healthcare). The fractions containing the target protein were collected for purification once more by gel filtration chromatography using HiLoad™ 16/60 Superdex™ 200 column (GE Healthcare). SDS-PAGE was used to check the purity of proteins obtained after each step. Finally, protein samples were concentrated and stored at -80 °C.

It's important to note that, for the purpose of crystallization, the soluble expression of target protein at high efficiency was achieved by the extra addition of 0.1 mM AC94377; For the biochemical characterization of GID1 proteins, there is unnecessary to remove the fusion tag Trx·6×His, otherwise, the protein solubility may be dramatically decreased.

2.4.3 Crystallization trials

All crystallization experiments were conducted by standard vapour diffusion in sitting drop plates. Crystals of AtGID1a-GA_4-GAI were obtained under the condition that previously reported[73]. Crystals of AtGID1a^{V319I}-AC94377-GAI were observed by a 2:1 mixture solution comprising 50 mg/ml of protein solution (in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM DTT with 1 mM AC94377) and several different reservoir solutions (listed in Table 2-2). The reservoir solutions JCSG I-1B (0.2 M NaCl, 0.1 M HEPES pH 7.5, 20% PEG3000), PACT-9D (0.2 M LiCl, 0.1 M Tris pH 8.0, 20% PEG6000), and PACT-5H (0.2 M $NaNO_3$, 0.1 M Bis Tris propane pH 8.5, 20% PEG3350) were selected for subsequent optimization by making gradient of PEG *vs* pH,

PEG *vs* salt concentration, etc. The microseeding and streak seeding were performed according to the protocol (Crystal Growth 101) from HAMPTON research. Briefly, for microseeding, macro crystals were crushed into micro crystals with a probe, which was then collected several times by injecting reservoir solution in wells and subsequently transferred into tube that have already contained a seed bead. The seed bead tube was vortex three times on the ice, and it's now ready for use accompanied by the protein solution and reservoir solution. Streak seeding was conducted by using a seeding tool (HAMPTON). This tool should be assembled properly, and then touch the end of the probe to the donor crystal, and quickly run the tip of the probe in a straight line across the middle of the recipient drop containing the sample and reagent. Crystals should appear along the streak line.

For X-ray diffraction test, crystals were picked up and transferred into cryoprotectant solution containing mother liquor with 10% polyethylene glycol 200 and then flash-cooled in liquid nitrogen. Diffraction attempts were made at 100 K using a Rigaku R-AXIS VII detector equipped with a Rigaku FR-E X-ray generator at the SPring-8 synchrotron facility. However, the diffraction data was not collected due to the low resolution (>10 Å).

2.4.4 Molecular docking simulation

The ligand-free crystal structure of OsGID1 (PDB ID: 3ebl) was used as the receptor for docking simulation by Autodock Vina (The Scripps Research Institute, La Jolla, CA, USA) according to the protocol[150].

The 2D structure of AC94377 was obtained by using ChemDraw, and the energy-minimized conformer was obtained by using Chem3D. Next, AutoDockTools-1.5.6 was used to further prepare the ligand and receptor, including assigning partial charges to every atom of the ligand and the receptor, removing water atoms and adding polar hydrogen atoms of the protein structure, detecting rotatable bonds of ligand, setting grid box with the focus on the centra pocket of OsGID1, etc. The center of the grid box was -17.767, 63.951, and 20.439 Å in x, y, and z axis, respectively, whereas the dimension (Å) was all 28.11 for x, y, and z axis. The generated PDBQT files will be used for docking trials. The best binding mode of OsGID1-AC94377 was used for structural analysis based on the rank of the negative free-binding energy. Notably, re-docking

OsGID1 with its crystallographic ligand GA_4 was performed to validate docking studies.

2.4.5 Intrinsic tryptophan fluorescence assay (ITF)

Fluorescence from individual aromatic residues (Trp, Phe, Tyr) is influenced by the conformation of protein, especially in the case that the conformational change induced by ligand binding makes it possible to be dynamically monitored in real-time[145, 148]. For GID1 protein, the highly flexible cap domain containing several aromatic residues (Figure 2-7a) is conformationally changed by ligand binding, reflecting on the fluctuation of intrinsic fluorescence, which is availably quantified by binding affinity between protein and ligand.

For the fusion protein Trx·6×His-GID1, the excitation of 280 nm and emission wavelength of 350 nm were fixed by performing the excitation (range: 230-340 nm) and emission (range: 300-490 nm) scanning with a slit width of 1 nm. In a standard assay, the ligand stock of varying concentration (0.1-100 mM) was prepared by dissolving solid compound in DMSO. The protein-ligand binding reaction was initiated at 25 °C on the flat-bottomed, black 96-well plate with 100 μL of mixture in each well containing 10 μM protein and 1% ligand stock. After 30 min, the fluorescence intensity was recorded by EnSightTM (PerkinElmer, USA) with the number of flashes fixed to 100 and measurement height at 5 nm. The mean from three independent experiments was used to calculate the ΔF by taking the difference in protein fluorescence of the DMSO control and each ligand concentration. The dividing of each ΔF by the maximum change of protein fluorescence ΔFmax was performed to calculate the ΔF/ΔFmax, which was then plotted against ligand concentration by fitting to nonlinear regression with '*One Site-Nonspecifc Binding*' model in GraphPad Prism 5, finally giving K_d value.

2.4.6 Differential scanning fluorimetry (DSF)

The DSF experiment was monitored using 96-well plate in the FRET channel of CFX Connect Real-Time PCR Detection System (Bio-Rad, CA, USA). The reactions mixture (25 μL) consists of 2 μM Trx·6×His-GID1, 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 × SYPRO orange and 1% (v/v) DMSO that containing varying concentrations of ligand stock. A typical denaturation program was set up with minor modifications[194]. Briefly, the mixture was pre-incubated at 20 °C for 5 min, and then ramp up to 95 °C at a rate of 0.5 °C /s. The experiments were repeated three times.

2.4.7 Variants library construction

All sites/sublibraries of AtGID1 variants were constructed by reverse PCR-based site-directed mutagenesis. Before starting PCR, the phosphorylation of degenerate primers (Table 2-9) was performed using T4 polynucleotide kinase (New England Biolabs) under the standard protocol. Each PCR reaction mixture (50 μL) contained 33 μL water, 5 μL KOD-plusneo buffer (10×), 3 μL 25mM $MgSO_4$, 5 μL 2 mM dNTPs, 1 μL (50~100 ng) template DNA, 10 μM primers mix 2 μL, and 1 μL KOD-plus-neo polymerase (TOYOBO, Japan). PCR program is as follows: 98 °C 3 min, (94 °C 20 s, 62 °C 20 s, 72 °C 8 min 30 s) × 30 cycles, 72 °C 10 min, 16 °C 10 min. After confirmed by 1% gel electrophoresis, the PCR products were digested with *Dpn* I at 37 °C for 6 h. And then, the purified products were self-ligated using T4 DNA ligase (TAKARA, Japan) at 16 °C for 12 h. Afterwards, the ligation products were purified and transformed into TOP10 competent cells (ThermoFisher Scientific) by electroporation (Bio-Rad). The cells harvest for plasmid extraction of DNA libraries were performed under the protocol recommended (https://www.addgene.org/protocols/pooled-library-amplification/). The sequence of pooled libraries was confirmed by DNA sequencing and the quality of each libraries was briefly assessed by calculating the Q_{pool} values[195]. The concentration of each libraries was fixed around 2 μg/μL, and it is now ready for transformation into yeast cells.

2.4.8 Y2H-based screening

The high transformation efficiency of libraries was achieved by using Fast Yeast Transformation™ Kit (G-Biosciences) under the standard protocol provided. The screening of sites A-I and J were performed on the SC-LWHA plates respectively containing 1 μM and 10 nM AC94377. The transformation efficiency was calculated by spreading the transformation solution on SC-LW plates. Transformants from individual site/sublibrary were picked and spots on the SC-LW plates at 30°C. Two days later, each transformant generated was spots on SC-LWHA plates containing the same concentration of AC94377 as used in screening. Scraping the positive transformants in the tubes containing 300 μL zymolyase solution (10 mg/mL, 50 mM PB buffer, pH7.5) for cell lysis at 35 °C for 1 h. The plasmids were extracted and amplified by PCR, and later

used for DNA sequencing. The obtained variant plasmids were transformed back to AH109 harboring AtGAI for Y2H assay on the SC-LWHA plates containing the concentration gradient of AC94377.

Chapter 3: Structural basis for the perception of a Karrikin antagonist KK181N1 by receptor AtKAI2

3.1 Research Summary

Through the recent works in our laboratory, a compound (named KK181N1) that shows antagonistic effect on karrikin-induced suppression of hypocotyl growth in Arabidopsis has been screened. This chemical binds directly and alleviate the hydrolysis activity of karrikin receptor AtKAI2. However, the precise binding mode of KK181N1 in AtKAI2 still remains unknow. In this study, the structure of AtKAI2-KK181N1 complex was solved at 1.90 Å resolution with a clear electron-density map of KK181N1. As reported, the overall structure of AtKAI2 adopts canonic α/β hydrolase fold with S95, H246 and D217 as the catalytic triads. KK181N1 belongs to N-heterocyclic ureas and designed to covalently modify α/β hydrolase fold-containing proteins through the catalytic nucleophile. However, unexpectedly, I found that this binding was caused by noncovalent interaction and KK181N1 was firmly embedded in the pocket of AtKAI2. This binding mode of KK181N1 is totally different from that of OsD14-KK094, although OsD14 is highly homologous to AtKAI2 and KK094 only differs from KK181N1 in one methyl group. To thoroughly understand the structural basis of KK181N1's binding mode in AtKAI2, we divide the binding pocket into three regions: the bottom, the middle region, and the entrance of the pocket.

At the bottom of the pocket, I found a four water molecules-mediated hydrogen-bonding network connecting the 1,2,3-triazole moiety of KK181N1 to the residues surrounded, including the catalytic nucleophile S95, catalytic base H247, Y124, S119, and the backbone of F26 and V96. Our structure and activity relationship studies have indicated that this hydrogen-bonding network is of great importance for KK181N1-binding affinity in AtKAI2. It is notable that the variants S95C and S95D respectively retain 97% and 65% of wild type's affinity to this chemical, suggesting the broad-spectrum applicability of our in-house chemical library for α/β hydrolase fold superfamily proteins. Intriguingly, the structure comparison between OsD14-KK094CM and AtKAI2-KK181N1 complexes has shown that, even though these two chemicals have different binding mode in the pocket, they all participate in this water molecules-mediated hydrogen-bonding network with greatly similar topology, suggesting this hydrogen-bonding network, which seems to be conserved among KAI2/D14 family proteins, would be favored for the binding of 1,2,3-triazole compounds.

At the middle region of the pocket, we found extensive hydrophobic interactions between the aromatic plane of KK181N1 with non-polar residues shaping the active pocket. These residues

include F26, Y124, F157, A161, F174, I193, F194, A219, and H246. Alanine scanning experiment reveals that the hydrophobic interactions are essential for the KK181N1-binding in the pocket. It's interesting to notice that a variant A219S shows two-folds higher affinity than wild type for the binding to KK181N1. Previous work of X-ray crystallography has suggested that the variant of A219V may narrow the entrance gate for the ligand entry. However, the structure comparison between AtKAI2-KK181N1 complex and *apo*-OsD14 (corresponding residue: S270) indicates that the side chain of serine at this site would not cause the steric clash with ligands, including KAR_1 and KK181N1. Moreover, I compared the structures of AtKAI2-KAR_1 and *apo*-AtKAI2 with that of AtKAI2-KK181N1 complex and found that F134, F157, L142, F194 and L218 have undergone the conformational changes for the perception of KK181N1.

At the entrance of pocket, I notice that three residues, including F134, L142 and L218, account for the recognition of ligand KK181N1. It is interesting to find that each of the alanine substituent at these positions all possesses increased affinity to KK181N1, which may suggest that the releasing of steric effect at the entrance will promote the recognition of KK181N1. This is particularly verifying my structural information where all of the residues (F134, L142 and L218) occur conformational shift to accommodate the KK181N1.

To study the ligand selectivity, I place my focus on the three residues (F157, L218, and A219), which interact with the methyl group of KK181N1 by hydrophobic interaction but are different from the residues at the corresponding location in OsD14. The structural comparison and affinity assay have suggested that although S270 (A219 in AtKAI2) and V269 (L218 in AtKAI2) in OsD14 have the potentiality to elevate the binding affinity of OsD14 to KK181N1, the Y209 (F157 in AtKAI2) would cause sever steric clash with carbonyl and the methyl group of KK181N1. This may partially explain the binding preference of KK181N1 to AtKAI2 but not OsD14 and is consistent with our physiological result that KK181N1 exhibit the inhibitory effect on hypocotyl growth through karrikin signaling pathway, but not SL. On the other hand, to further confirm the significance of this methyl group for ligand recognition by AtKAI2, we prepared an array of derivatives with different substitution at this group, including KK094N1 (hydrogen group), KK167N1 (nitro group), KK175N1 (chloro group), and KK187N1 (fluoro group). However, all of them shows decreased affinity to receptor AtKAI2, suggesting that methyl group is most favored by this hydrophobic environment.

In the process of the chemical synthesis of KK compounds, two forms of isomers were prepared. In one compound, carbonyl group is connected to nitrogen atom at 1-position of 1,2,3- triazole moiety (annotated as N1), whereas in another carbonyl group is connected to nitrogen atom at 2-positon of 1,2,3-triazole moiety (annotated as N2). Since I have found the great contribution of hydrogen-bonding network to anchoring the KK181N1 in the deep site of pocket, I am wondering whether this network also contributes to the specificity of KK compounds in AtKAI2. The affinity comparison between KK181N1 and KK181N2 or KK094N1 and KK094N2 suggested that AtKAI2 only shows higher affinity toN1-type ureas, which is consistent with the structural details with respect to the topology of hydrogen-bonding nework where N1-type 1,2,3-triazole ureas could form three hydrogen bonds to join but N2-type 1,2,3-triazole ureas may only have one. This result was further examined by applying the chemical KK122 for affinity assay. Even in the absence of one nitrogen atom (N5, indicated by green shadow in Figure 3-1g) forming one hydrogen bond, the presence of another nitrogen atom (N6, indicated by blue shadow in Figure 3-1g) harboring two hydrogen bonds could reinforce the binding affinity of KK122 to approximately 3 folds higher than KK094N2, or 2 folds than KK181N2, stating that N6 atom-formed hydrogen bonds are vitally important for the binding affinity of 1,2,3-triazole-containing chemicals toward AtKAI2. What's more, by using other N-heterocyclic ureas with divergent moieties, I also found that N1-type1,2,3-triazole is greatly favored by AtKAI2 proteins, which is contributed by the unique hydrogen-bonding network. Taken together, the structural analysis and SAR studies have suggested that the water molecules-mediated hydrogen-bonding network not only promotes the tight binding of KK181N1, but also confers the ligand selectivity of AtKAI2 for different isomers, and the residues involved in the recognition of the methyl group of KK181N1 greatly contribute to the selectivity of KK181N1 to AtKAI2 but not D14.

Finally, we made attempts to unveil the mode of action of KK181N1 by Y2H assay. Although we have found the sufficient fragments of SMAX1 and MAX2/D3 for the establishment Y2H assay, we fail to detect the inhibitory effect of KK181N1 on the KL agonists-induced PPI interaction between AtKAI2 and SMAX1 or AtKAI2 and MAX2/D3. In our future work, we would try other methods to elucidate the mode of action of KK181N1, such as Co-IP assay.

This study of uncovering the structural basis of the recognition of KK181N1 by AtKAI2 will not only enlighten us to perform structure-guided optimization of KK181N1 to generate a more robust

inhibitor but also encourage us to apply this antagonist in plants for basic research, such as for the exploration of the native ligand of KAI2 and the solo establishment of SL function that bypass KAI2 signaling.

3.2 Results and discussion

3.2.1 Discovery of a Karrikin antagonist KK181N1 ***(Results from the Master thesis by Toshihiko Sakai)***

Both KAI2 and D14 are plant hormone receptors that belong to α/β hydrolase superfamily and share MAX2 as a common signal component for their activity exhibitions. Previous phylogenetic analyses have suggested that the KAI2 is likely the ancestral lineage where D14 arose[196, 197], and in the aspect of ligand perception, they all respond to the similar butenolide compounds. However, the failure in the KAI2 and D14 promoter-swapping experiments demonstrate the distinct roles of KAI2 and D14 in regulation of plant growth. Therefore, the different outcomes of MAX2-dependent signaling would be reflected by the specific protein-protein interactions between receptors and SMXLs family proteins[136, 198]. Recent works in the establishment of SL biology have been extensively advanced by using the small molecule tools including strigolactone analogs, such as GR24[118]. Nevertheless, it still remains a challenge to have the intact elucidation of the mechanism of KAI2-mediated signal transduction due to the shortage of karrikin-related chemical probes. Therefore, an available KAI2 antagonists would enlighten our studies on tracing the essential clues for the identification of native KAI2 ligand and its biosynthetic origins. With this in mind, we performed the phenotype-based screening from an in-house chemical library for compounds that antagonize karrikin signaling in the model plant Arabidopsis thaliana, with the goal of identifying potential inhibitors of KAI2-dependent reduction of hypocotyl growth.

Hypocotyl growth of *Arabidopsis thaliana* is regulated through both KAI2 and D14 signaling pathways. Only 2'-(*R*) SLs can reduce hypocotyl length in a AtD14 dependent manner, but both 2'-(*R*) and 2'-(*S*) SLs can reduce AtKAI2-dependent hypocotyl length[199] and this effect is significantly remarkable under red light conditions[200]. Based on this result, a hypocotyl-based screen system for KAI2 inhibitors was developed as shown in Figure 3-1a. After first screening, it is important to rule out the possibility that screened chemicals restoring hypocotyl length shortened by KAR1 is due to antagonizing D14, not due to KAI2[201]. To achieve this goal, *d*14 and *kai2* signaling-deficient mutants were used. That is, chemicals insensitive to *kai2* and sensitive to *d14* should be our targets.

Substantial screening efforts enable us to find two candidates compounds that show an inhibitory effect on hypocotyl length of *Arabidopsis thaliana* (Columbia) (Figure 3-1b). This result is further

confirmed by the experiment using *Arabidopsis thaliana* (Landsberg erecta) because Landsberg erecta is more sensitive to KAR_2 than Columbia[134] (Figure 3-1c). It is well known that AtKAI2 also involved in promoting the seed germination. Therefore, in order to further confirm that these two inhibitors are directly targeting AtKAI2, we try to check if inhibitors could suppress the seed germination rate. Figure 3-1d demonstrate that KK181N1 could strongly repress breaking dormancy of seeds in single treatment or co-treatment with KAR_2, the most potent agonist of AtKAI2[87]. Furthermore, the inhibitory effect of KK181N1 was confirmed by measuring the remaining hydrolytic activity of AtKAI2 using the desmethyl-Yoshimulactone Green (dYLG), a fluorescent probe that is specifically hydrolyzed by KAI2 (Figure 3-1e, f) [202].

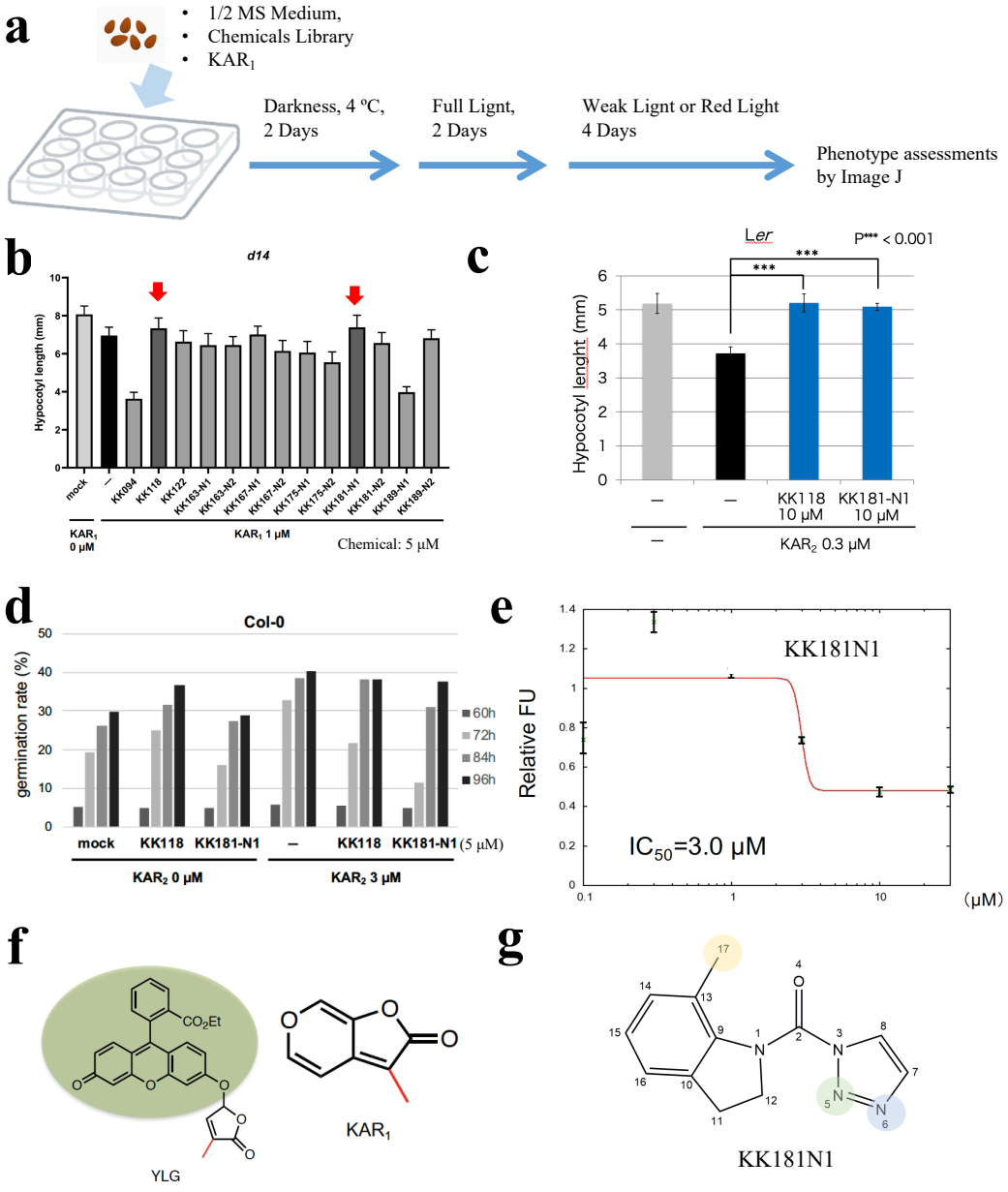


Figure 3-1: Previous work for the discovery of a Karrikin antagonist KK181N1.
a, A schematic diagram of the phenotype-based screening. **b,** Effect of the co-treatment of KAR_1 (1 μM) and compounds (5 μM) on hypocotyl length of Arabidopsis *d14* mutant under red light conditions. diagram of the canonical α/β-hydrolase fold with α-helices indicated by white cylinders and β-strands by gray arrows. Compounds KK118 and KK181N1 were considered as the candidates antagonizing KAR_1 response. hypocotyl length was measured by Image J, n = 12-22, Error bar = s.e.m. **c,** Landsberg erecta was used in this experiment in the same condition as in **b. d,** Effect of KK118 and KK181N1 (10 μM) on seed germination rate in *Arabidopsis*. The seed germination rate was measured every 12 hours after sowing. Number of seeds 52-88, n = 1. **e,** Competitive inhibition of KAI2-mediated dYLG hydrolysis. The fluorescent values (FU) relative to the hydrolysis with inhibitor are shown. n=3, Error bar=s.e.m. IC50 values were calculated using the website http://www.ic50.tk/index.html. **f,** Fluorescent probe desmethyl-Yoshimulactone Green (dYLG), the KAR_1 agonist and shows specific affinity to AtKAI2. **g,** Chemical structure of antagonist KK181N1 used in this study. Each atom of KK181N1 was numbered by Chemdraw.
Reprinted and modified from "Master thesis of Mr. Toshihiko Sakai (酒井 寿彦), 2020"

3.2.2 Crystallization of AtKAI2-KK181N1 complex

Our in-house chemical library is specially designed to target the α/β-hydrolase fold proteins that contain the canonical catalytic residues for the hydrolysis of N-heterocyclic ureas. To elucidate the structural basis of KK181N1's binding in AtKAI2, I performed the co-crystallization of AtKAI2 with compound KK181N1. Initially, except for the operation of crystallization screens, I also made attempts to crystallize this complex under those condition that have previously reported to grow crystals of *apo*-AtKAI2 or AtKAI2-KAR_1[57, 101, 203]. I observed the generation of crystals on the reservoir solution (1.4 M sodium/potassium phosphate pH 8.0-8.4)[101], and the single crystal was picked up for X-ray diffraction in Photon Factory, Japan. However, even though this complex structure was solved at 2.1 Å, the poor electron density fails to present us a clear ligand map.

It is well known that a homogeneous protein sample is one of the most important factors, especially for co-crystallization, where the inhomogeneity may occur between ligand-bound protein and *apo*-protein. This unliganded protein sometimes act as impurities to interfere with crystal lattice formation, yet leading to the failure in the observation of crystallographic occupancy in the generated crystals[204]. Moreover, my subsequential efforts in ligand soaking experiments were also inadequate to make electron-density maps of ligand clear, indicating that ligand, even at high concentration, may not enter the pocket site. This may be caused either by ligands that induce large conformational changes that are not tolerated by crystal packing in the *apo*-protein or by the fact that the interaction interface between two protein molecules covers pocket entrance[205]. In order to satisfy at least 90% occupancy so that the protein-ligand complex is well-populated in co-crystallization trials, the practical experiments suggest amount of ligand added must be greater than that of protein, that is, the unbound ligand at equilibrium should be not exhausted to less than about $10\times K_d$[206]. Generally, ratios of ligand to protein of up to 10:1 or more are used. For AtKAI2-KK181N1 complex, since the affinity of KK181N1 was estimated at micromolar level, high concentration of purified protein AtKAI2 obtained from ion-exchange chromatography (Figure 3-2a, b) was first diluted into 0.7 mg/mL (about 30 μM) and incubated with 1-2% (vol/vol) ligand stock (500 mM) for 1 hours at room temperature. Next, the initial reservoir solution was further optimized by making the gradient of salt concentration and pH value, and using additives termed silver bullets. Several quality-sufficient crystals were observed in the reservoir solution containing 1.7 M sodium phosphate (pH 7.0-7.2) and 1% 1,2-butanediol (Figure 3-2c, d). Through these

attempts, I have succeeded in solving the crystal structure of the AtKAI2-K181N1 complex with a clear electron-density map of ligand from a single crystal generated in 1.7 M sodium phosphate pH 7.0 and 1% 1,2-butanediol (Figure 3-2e, f).

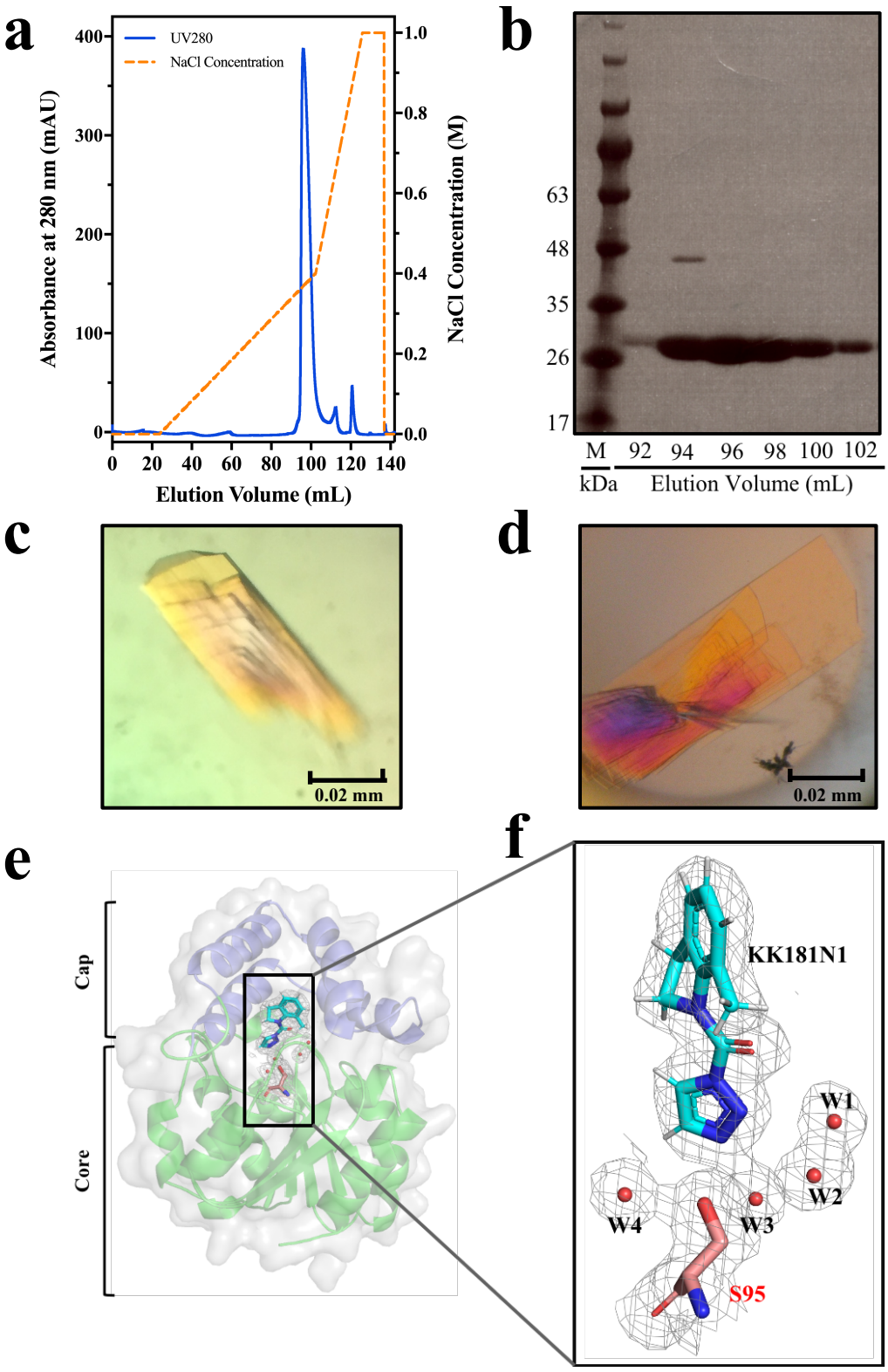


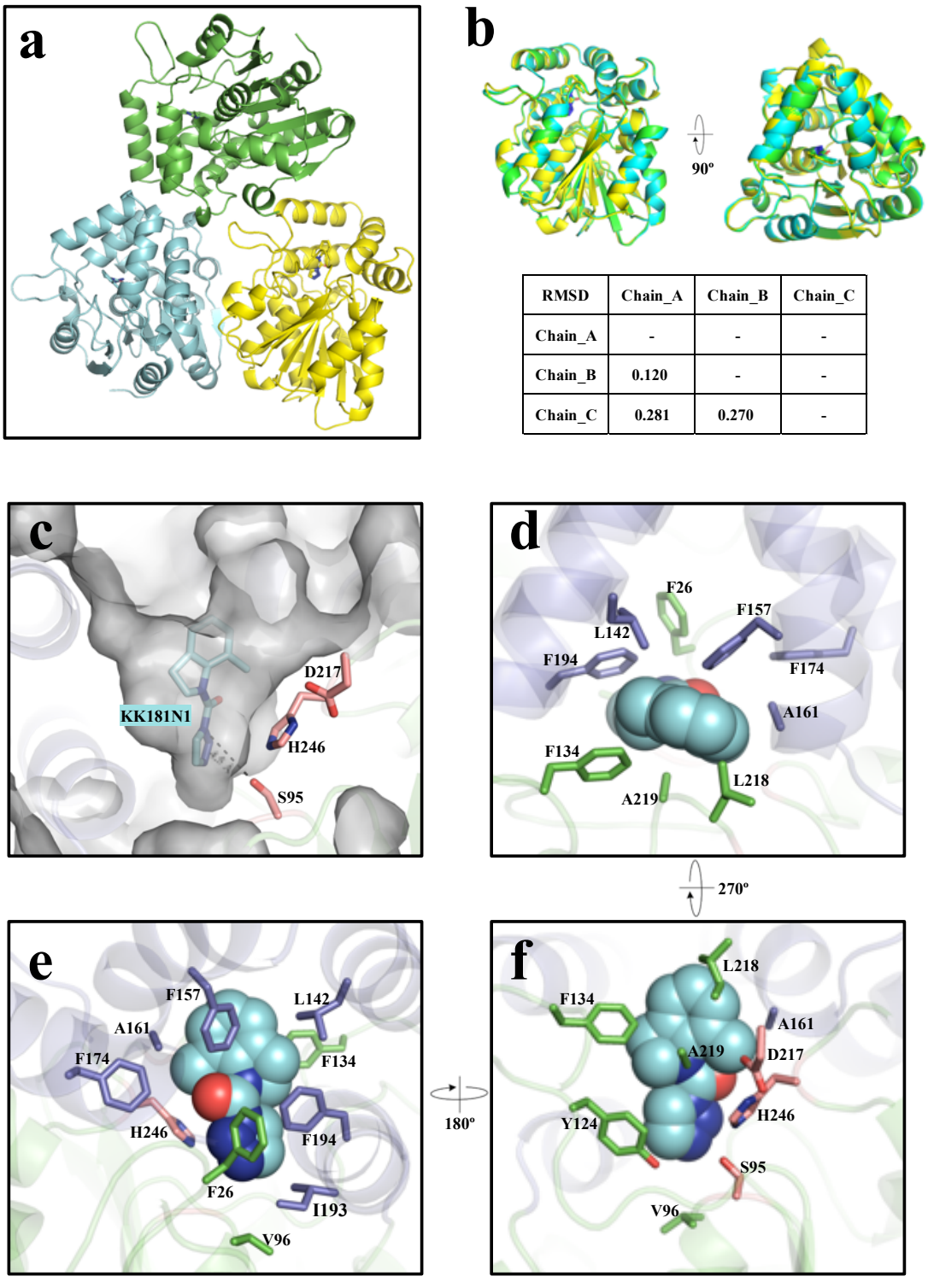
Figure 3-2: Purification and crystallization of AtKAI2-KK181N1 complex.
a, Chromatogram of ion-exchange chromatography for the purification of AtKAI2. **b,** The purity of AtKAI2 was confirmed by SDS-PAGE. Quality-sufficient crystals of AtKAI2 were observed in the reservoir solution containing 1.7 M sodium phosphate pH 7.0 (**c**) or 7.2 (**d**) and 1% 1,2-butanediol. **e,** The stereoscopic cartoon and surface diagram of AtKAI2-KK181N1 complex. The cap domain is colored purple, and the core domain is green. **f,** a focused view of the active site of AtKAI2. KK181N1 (shown as the cyan stick) was non-covalently embedded in the pocket. The relative spatial position of catalytic nucleophile S95 (shown as the pink stick) and four water molecules (named as W1-W4, shown as pink nb_spheres) situated at bottom of the pocket to KK181N1 was presented, along with 2*Fo-Fc* electron density map contoured at 1.0 σ (gray).

3.2.3 Recognition of KK181N1 by AtKAI2

The subsequent data analysis and refinement led to the determination of complex structure at 1.90 Å resolution (Table 3-1). Three protein molecules were detected within each unit cell of the AtKAI2-KK181N1 crystal in the P 1 2_1 1 space group, and these three molecules are essentially identical with root-mean-square deviation (RMSD) values of 0.12-0.281 Å when the overall structures are superimposed pairwise (Figure 3-3a, b). As reported previously, the overall 3-D structure of AtKAI2 contains a core domain that is subject to the α/β hydrolase fold, and a cap domain formed by a double layer V-shaped helical fold comprised of four helices (Figure 3-2e; Figure 3-3b). The catalytic triad residues of AtKAI2 are S95, D217 and, H246 all of which are deposited at the bottom of the hydrophobic residues-constituted active site and on the loops after the β5, β7 and β8 strands, respectively (Figure 3-9). This canonical spatial arrangement indicates a conserved catalytic mechanism as the hydrolases, that S95 works as highly reactive nucleophile to attack the electrophiles, while the H246 and D217 contribute to the formation of oxyanion hole, which stabilize the negatively charged transition state of the catalyzed reaction (Figure 3-4c). N-heterocyclic urea compounds, as one of electrophiles with tunable reactivity, could be covalently approached by active-site serine residues through carbamoylation[17, 36]. Therefore, before determining the complex structure, I have anticipated that KK181N1 would be hydrolyzed by AtKAI2 and leaving carbamoyl moiety bound to S95 as similar way how SL antagonist KK094 bound to D14 in our previous work[194], KK181N1 would be hydrolyzed by AtKAI2 as well, and leaving carbamoyl moiety bound with S95 (Figure 3-4c). Unexpectedly, an unambiguous electron density map of the intact structure of KK181N1 was detected (Figure 3-3; Figure 3-4).

As shown in Figure 3-3a, the ligand is embedded completely in the cavity, but the distance of its carbamoyl group to the hydroxy group of the catalytic nucleophile S95 is as far as 4.5 Å, impossibly initiating the covalent carbamoylation (Figure 3-4c). Conversely, it is obvious to see that the active pocket, which is shaped by many hydrophilic and hydrophobic residues, brings multiple polar and non-polar interactions (within 5 Å) to firmly envelope the entire KK181N1 in the active site of AtKAI2, including F26, S95, V96, Y124, F134, L142, F157, A161, F174, I193, F194, D217, L218, A219, H246 (Figure 3-3). Specifically, the acceptance for ligand entry was governed by the alkyl side chains of A161, L142 and L218 and aromatic side chain of F134 and F157, which all located around the entrance of pocket (Figure 3-3c, d); And the spatial orientation of KK181N1 in the

middle of the pocket is anchored by the robust π-stacking interactions, which are in the geometrically defined edge-to-face and face-to-face ways, between the aromatic side chain of F26, Y124, F174, F194, and H246, and alkyl side chains of A161 and A219 with the aromatic and 1,2,3-triazole ring of KK181N1 (Figure 3-3c, d); Finally, at the bottom of pocket, the recognition of KK181N1 was dominantly contributed by the four water molecule-mediated hydrogen-bonding networks, which connected to the 1, 2, 3-triazole ring of KK181N1, amide backbone of F26 and V96, and the hydroxy side chain of catalytic nucleophile S95 and Y124, further clarifying a non-covalent, but not covalent, binding mode of KK181N1 in AtKAI2 (Figure 3-2f; Figure 3-4a, b).



RMSD	Chain_A	Chain_B	Chain_C
Chain_A	-	-	-
Chain_B	0.120	-	-
Chain_C	0.281	0.270	-

Figure 3-3: Structural overview of AtKAI2-KK181N1 complex.

a, Structures of the three AtKAI2-KK181N1 molecules in each asymmetric unit are nearly identical. **b,** Three molecules in one asymmetric unit are superimposed with root-mean-square deviation (RMSD) values of 0.120-0.281 Å (Table below). **c,** Ligand-binding pockets are indicated as gray surface. The relative spatial position of catalytic residues (S95, D217, H246) to KK181N1 are presented. The distances of carbamoyl group of KK181N1 to the hydroxy group of

catalytic nucleophile S95 are indicated as black dash line, about 4.5 Å. The hydrophobic and/or van der Waals interactions between residues (F26, S95, V96, Y124, F134, L142, F157, A161, F174, I193, F194, D217, L218, A219, H246) and ligand (cyan spheres) are shown in top view (**d**) and side view (**e, f**). All residues involved in shaping the pocket of AtKAI2 and having closed contact with KK181N1 within 5 Å are shown as color-coded sticks in which nitrogen is blue, oxygen is red, and the color of carbon depends on the residues. Those residues situated in the cap domain are colored purple and in core domain are green, and catalytic residues are colored pink.

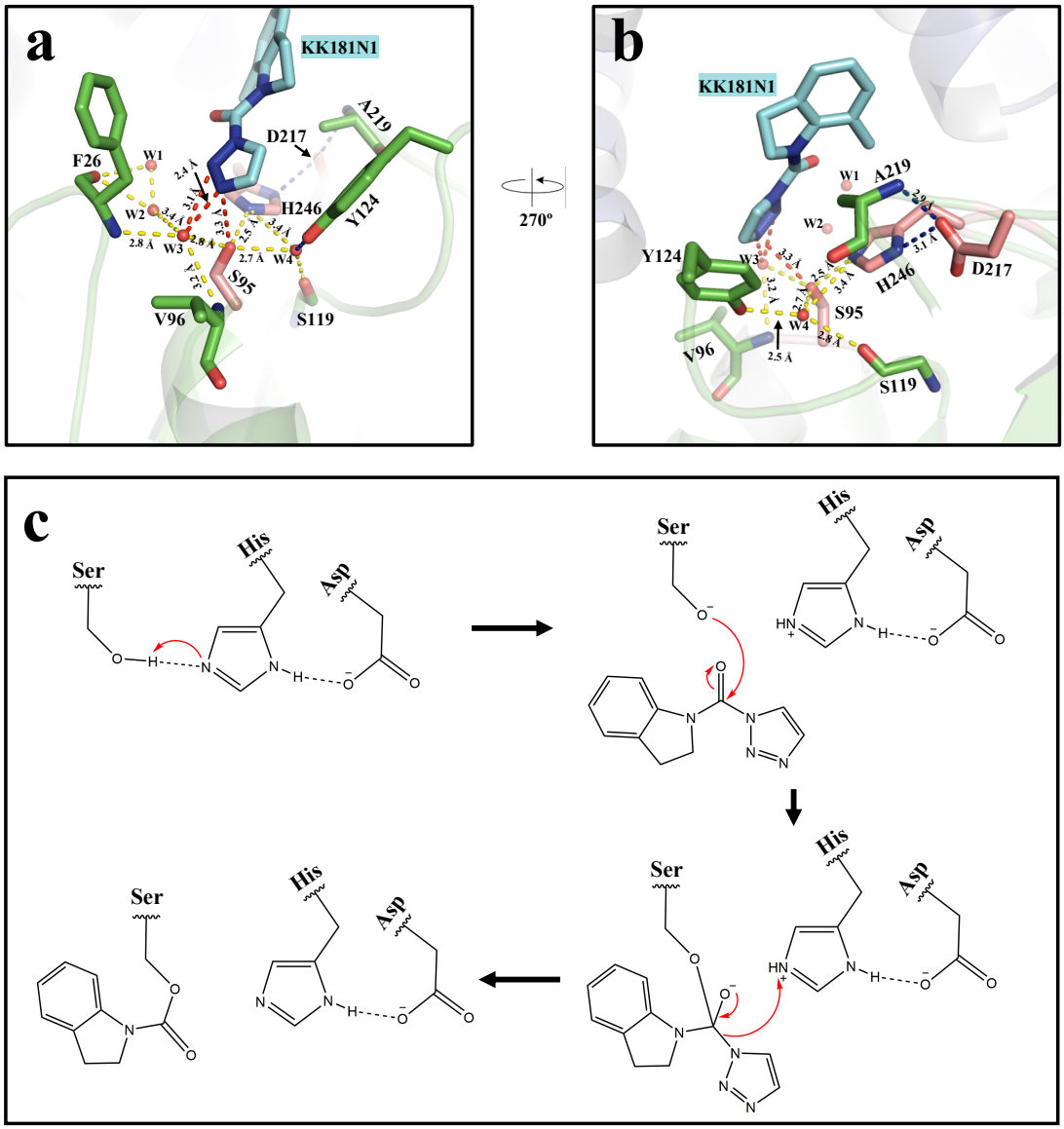


Figure 3-4: The hydrogen-bonding network and putative catalytic mechanism of covalent bond formation.

a, The four water molecules-mediated hydrogen-bonding network are colored as yellow and red dashes, the latter of which are to highlight the hydrogen bonds built from ligand KK181N1 (shown as cyan stick). Four water molecules (named from W1 to W4) are shown as red nb_spheres. The atoms numbering in KK181N1 see in Figure 3-1g. **b,** The additional interaction network covering two hydrogen bonds between A219, D217 and H246 are indicated as blue dashes. The lengths of all hydrogen bonds are shown with black values (Å). The catalytic residues are colored as pink sticks, and other residues located in the core domain to build up the hydrogen-bonding network are represented as green sticks. **c,** Proposed mechanisms of the covalent attachment of KK compound (taking KK094N1 as a example) by ABH proteins. A pair of electrons on the histidine nitrogen absorbs the hydrogen from the hydroxyl group of catalytic serine (Upper left panel), coordinating the subsequent serine-initiated nucleophilic attack on the carbonyl carbon of KK094N1 (Upper right panel). As a result, a pair of electrons from the double bond of the carbonyl oxygen moves to the oxygen so that a tetrahedral intermediate is generated stabilized by oxyanion hole (Lower right

panel). The bond joining the nitrogen (triazole moiety) and the carbon (carbamoyl moiety) in KK094N1 is now broken. The covalent electrons creating this bond move to attack the hydrogen of the histidine. The electrons that previously moved from the carbonyl oxygen double bond move back from the negative oxygen to recreate the bond, resulting in protein-KK094CM complex (Lower left panel).

3.2.4 Structure and activity relationship studies of AtKAI2-KK181N1

In order to verify this structural assignment, we decided to perform the structure and activity relationship studies between ligand and protein by producing AtKAI2 variants that substitute the residues involved in the interactions with KK181N1 with alanine or others that show similar chemical characteristics and examined their remaining affinity to KK181N1 by ITC experiments *in vitro*.

My spotlight is first shifted to assess the unique contribution that catalytic triads made to binding to KK181N1(Figure 3-5, Table 3-2). Although the recognition of KK181N1 by AtKAI2 did not rely on the catalytic triads-regulated hydrolysis followed by the covalent contact with S95, further structural analysis indicate that these three catalytic residues are all indispensable for the tight binding to KK181N1. For example, the affinity assay of S95A toward KK181N1 can barely be detected. It is reasonable because S95 plays a vital role to participate in this hydrogen-bonding network at the bottom of the pocket through a direct hydrogen bond and two indirect hydrogen bonds that are mediated by two water molecules (W3 and W4) to 1, 2, 3-triazole ring of KK181N1 (Figure 3-4a), but alanine is incompetent to build up such exquisite hydrogen-bonding network. In addition, I noticed that the expression efficiency of S95A is extremely low, which may indicate that the substitution of alanine at this position would destroy the conserved structural architecture of α/β-hydrolase fold. As reported, the catalytic nucleophile of ABH fold could be serine, cysteine or aspartic acid[6]. This knowledge can be further strengthened by two mutants S95C and S95D of AtKAI2, the former of which almost maintain all affinity of wild type to KK181N1 and the latter one still holds 65 % affinity compared with wild type, suggesting protein stability is retained (Table 3-2).

Even though D217 does not directly interact with KK181N1, it constructs an additional interaction network to stabilize the environment of the pocket site of AtKAI2, including the formation of hydrogen bonds between the carboxyl moiety of itself with the imidazole ring of H246 and the amide backbone of A219, among which H246 and A219 are fundamentally involved in the establishment of the hydrophobic wall to recognize KK181N1 (Figure 3-4 b). However, both variants D217A and D217E almost lost entire affinity, implicating this interaction network is of significance for the affinity for KK181N1-binding. Finally, as the only invariant residue among catalytic triads, histidine, along with acidic residues aspartate, forms a negatively charged

environment to intensify the nucleophilicity of serine in ABH fold proteins. In our complex structure, I found that the imidazole ring of H246 not only provides the aromatic interaction with KK181N1 (Figure 3-3c) but also directly engages in this hydrogen-bonding network by forming a hydrogen-bond with S95 (Figure 3-4). Specifically, the H-N1 (δ^1) of H246 works as the donor of hydrogen bond to the carboxylate of D217 and the N3 (ε^2) works as the acceptor to the hydroxyl side chain of S95 (Figure 3-4b). However, I noticed that variants H246A and H246F still show about 30 % binding affinity to KK181N1. This may suggest that even though such polar interactions in two variants may collapse, the side chain of alanine and phenylalanine may still partially retain the hydrophobic interaction with KK181N1. Moreover, I noticed that the expression efficiency of H246A and H246F is extremely low, which may be caused by the poor protein stability of two variants, since the hydrogen-bonding network of catalytic triads S-H-D is conserved among α/β-hydrolase fold superfamily.[14, 15] However, for residues Y124 and its variant Y124F, SAR studies suggested that maintaining the non-polar interaction (π-π stacking interaction) with KK181N1 while eliminating the polar interaction (particularly regarding building up the hydrogen-bonding network) does not sustain the binding affinity of the proteins to KK181N1 (Table 3-2). At the pocket bottom of complex structure, W4-mediated hydrogen-bonding network was formed between the hydroxy groups of S95, S119, and Y124, and the imidazole ring of H246 (Figure 3-4a). The benzene ring of F124 may still support this edge-to-face hydrophobic interaction with 1, 2, 3-triazole ring of KK181N1 but crush the core hydrogen-bonding network due to the loss of polar hydroxyl group (Figure 3-4a), eventually resulting in the undetectable affinity to KK181N1 (Figure 3-6, Table 3-2). Taken together, the results combining affinity assay of AtKAI2 variants and structural information demonstrate that the four water molecules-mediated hydrogen-bonding network with residues surrounded is of vital importance to anchor KK181N1 at the bottom of active pocket.

On the other hand, as discussed above, those non-polar residues with aromatic and alkyl side chains form numerous favorable hydrophobic and/or van der Waals interactions with KK181N1 (Figure 3-3a, b; Figure 3-5a, b). Our alanine-scanning experiment also confirmed that these residues are important for the binding affinity to KK181N1 (Figure 3-6, Table 3-2). The variants F157A, F174A, and F194A show very little (< 10 %) or no binding activity, and variants F26A, Y124A, and I193A only maintain low (10 - 36 %) or moderate (71 %) affinity to KK181N1 (Figure 3-6, Table 3-2).

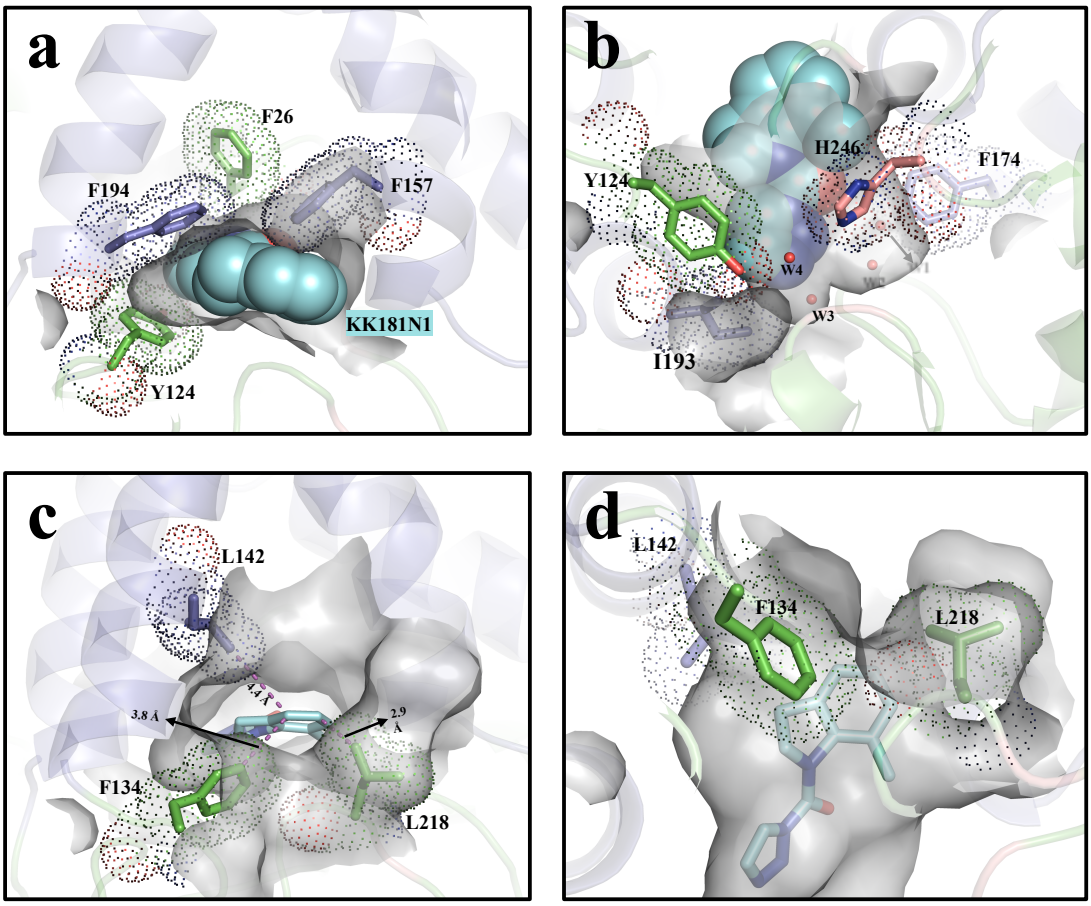
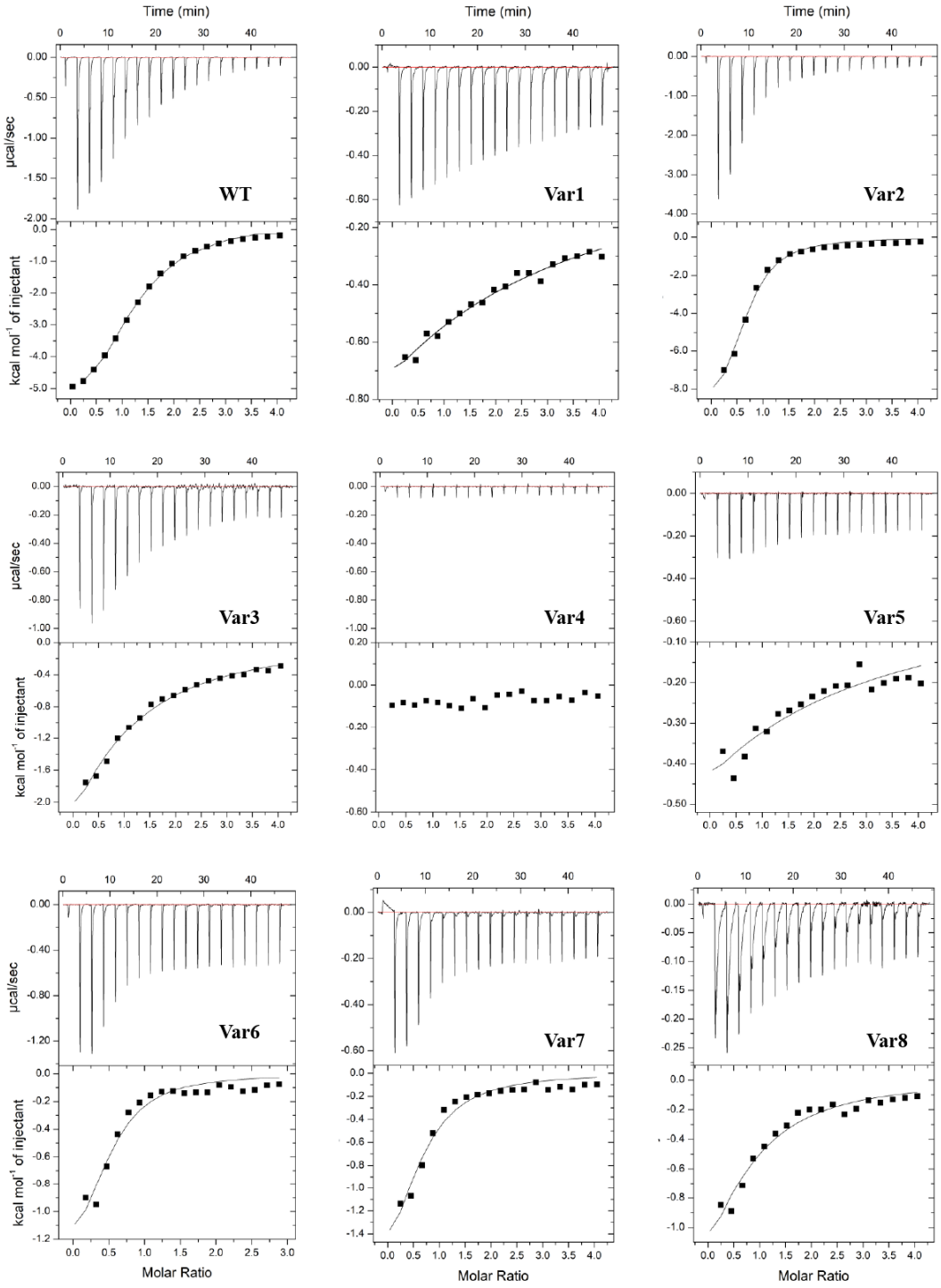


Figure 3- 5: The hydrophobic and van der Waals interactions between KK181N1 and residues lining the pocket of AtKAI2.

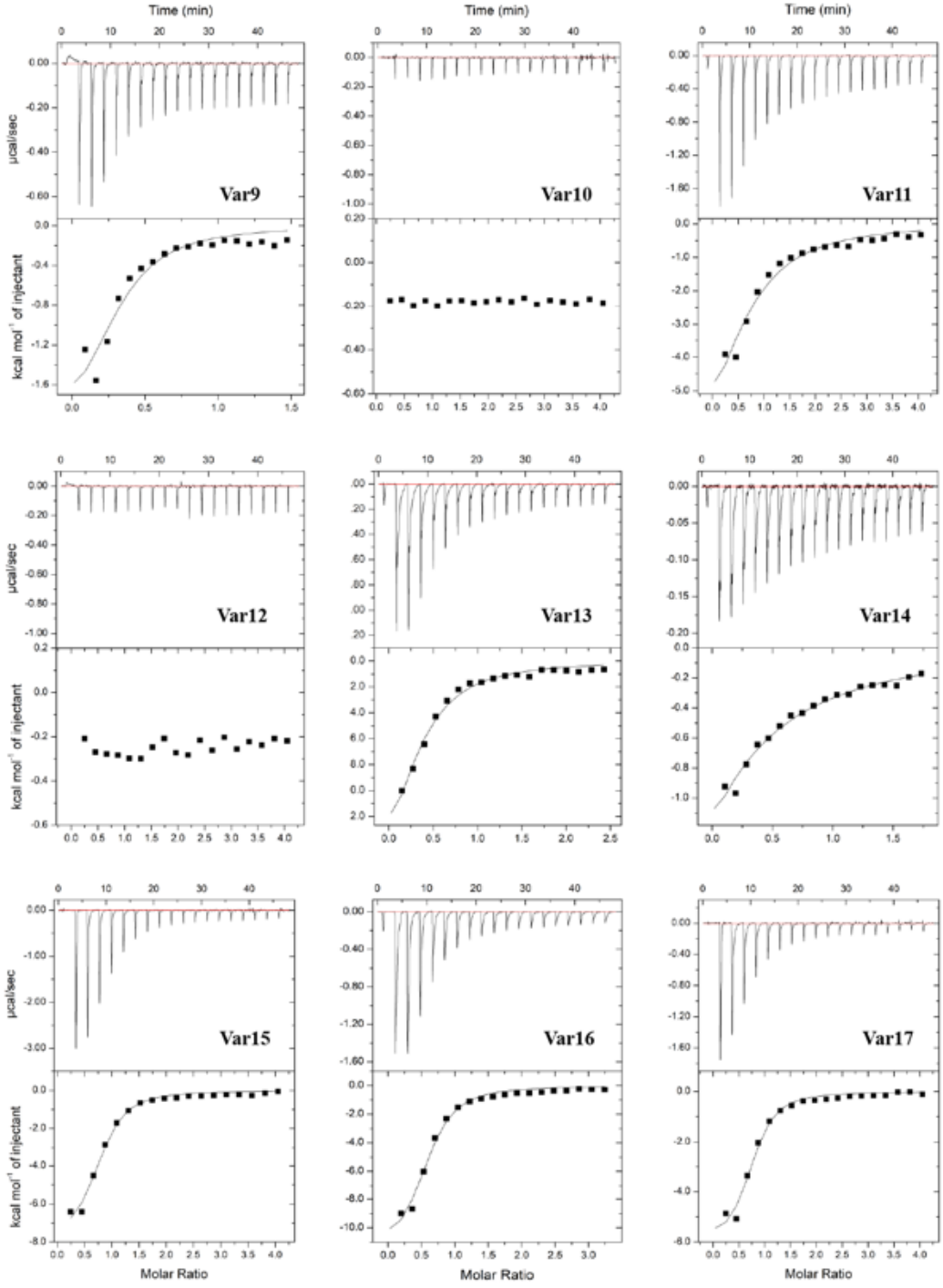
The top (**a**) and side (**b**) view of the middle part of pocket in the AtKAI2-KK181N1 complex structure. Those residues situated in the cap domain are colored purple and in core domain are green, and catalytic residues are colored pink. Four water molecules (named from W1 to W4) are shown as red nb_spheres. All residues are presented as sticks and further covered with dots to indicate the van der Waal's surface relative to KK181N1 (shown as cyan spheres). The top (**c**) and side (**d**) view of the entrance of the pocket in the AtKAI2-KK181N1 complex structure. Three residues (F134A, L142A, and L218A) along the boundary of pocket was indicated. The closest distance of the side chain of three residues to the edge of KK181N1 (shown as cyan sticks) are measured and shown as the purple dashes with black values (Å).

Intriguingly, the binding affinity of variants F134A, L142A, and L218A to KK181N1 is dramatically increased (Table 3-2). Given that KK181N1 has already been tightly embedded in the pocket by the forces contributed by the polar hydrogen-bonding network mediated by water molecules between the 1, 2, 3-triazole ring and nearby residues and the non-polar stacking between the 7-methylindoline moiety and hydrophobic residues (Figure 3-5a, b), I inferred that these three residues situated around the entrance of the pocket only play the guardians for ligand entry.

Therefore, as the KK181N1 has already been closely encased, the mutants containing alanine substituent would enlarge the entrance of pocket to accelerate the association/dissociation of KK181N1 without interfering with ligand recognition (Figure 3-5c, d). To support this hypothesis, I attempted to examine the affinity outputs of different variants on a single position. In this case, the attention was paid on L218 for three reasons. Firstly, of these three residues, leucine has the moderate size of hydrophobic side chain compared with phenylalanine (F134), I can produce variants of L218 that have either longer (F/Y) or shorter (A/V/I) alkyl side chains. Secondly, compared with L142, the alkyl side chain of L218 directs to KK181N1 more closely, and the closest distance of L218 to the aromatic ring of KK181N1 is 2.9 Å from the C6, whereas L142 is 4.4 Å (Figure 3-5c). In the third point, which is also a result of the second point, the side chain of L218 changes its conformation significantly to accommodate KK181N1(discussed in section 3.2.4). Thus, the functional investigation of L218 is more conform to our purpose. We prepared four more variants of L218 that is substituted with V/I/F/Y and measured their residual KK181N1-binding affinity by ITC (Figure 3-6, Table 3-2). As presumed above, compared with L218A, L218V and L218I possess slightly decreased activity, which, however, is still higher than wild type (Table 3-2). And both L218F and L218Y failed to exhibit the binding affinity to KK181N1 anymore (Table 3-2), which is probably caused by the partial or full coverage of pocket entrance by the bulky aromatic side chain, leading to the rejection for ligand entry. Collectively, consistent with structural information, our results of affinity assay stand that, followed by the hydrogen-bonding network at the bottom of the pocket, the non-polar residues-built hydrophobic walls also make a great contribution to tightly fix the orientation of KK181N1, and additionally, the releasing of hydrophobic effect at the entrance will promote the recognition of KK181N1.


Time (min)
µcal/sec
kcal mol^{-1} of injectant
Molar Ratio
WT
Var1
Var2
Var3
Var4
Var5
Var6
Var7
Var8

(Continued)


Time (min)
µcal/sec
kcal mol⁻¹ of injectant
Var9
Var10
Var11
Var12
Var13
Var14
Var15
Var16
Var17
Molar Ratio

(Continued)

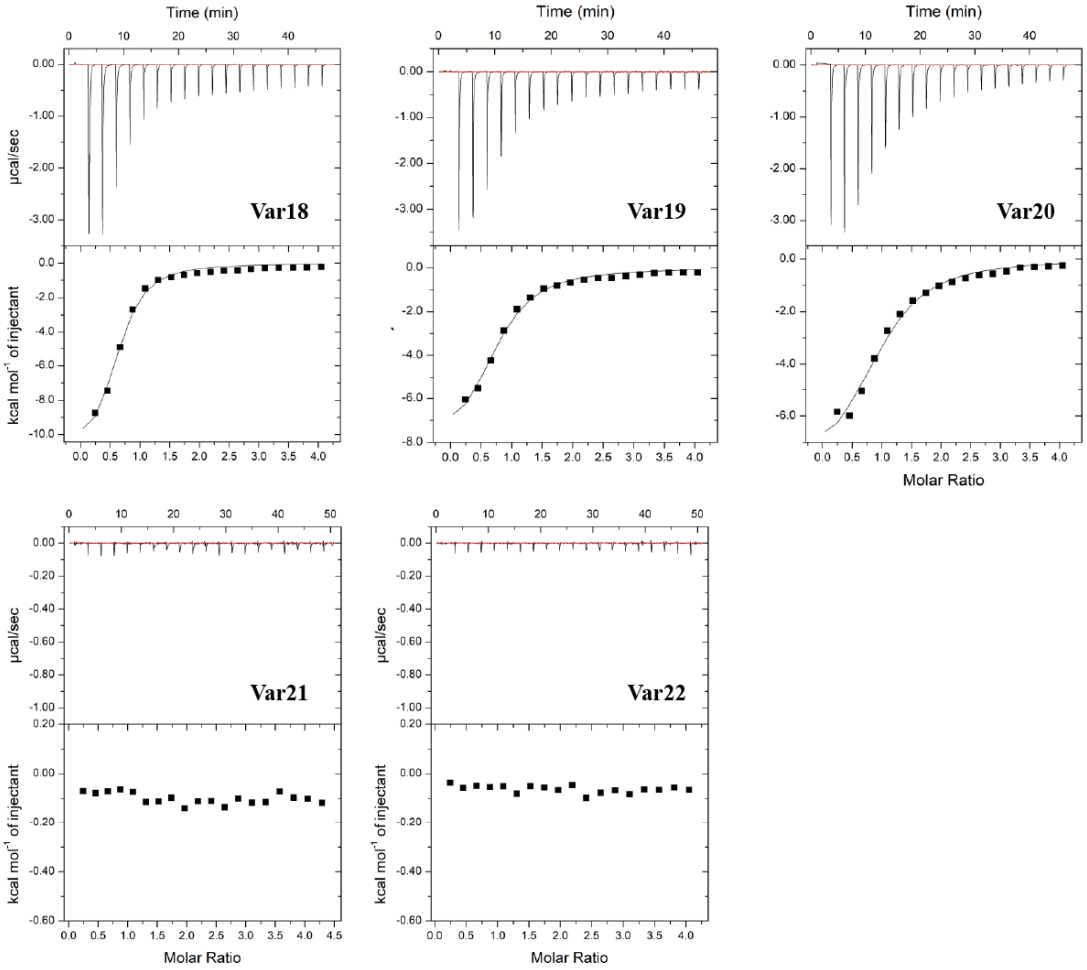


Figure 3-6: Isothermal titration calorimetry of KK181N1 binding to AtKAI2 variants. Representative ITC thermograms titrating AtKAI2 variants with KK181N1. The data for ITC were fitted using the 'one binding site' model by the MicroCal origin software. The panels show the ITC titration curves (µcal s^{-1}) and the integrated heats of injection (kcal mol^{-1}). n=>3.

3.2.5 Structural comparison

In order to further understand the structural basis for the perception of KK181N1, I compared the structure of AtKAI2-KK181N1 with that of *apo*-AtKAI2[101] and AtKAI2-KAR_1[101], respectively, by generating the 0.188 Å and 0.196 Å RMSD for the main chain Cα atoms. It is noted that the major deviation of protein backbone was observed only in the two loops, whereas structural comparison between *apo*-AtKAI2 and AtKAI2-KAR_1 results in 0.075 Å RMSD with no deviation on these two loops (Figure 3-7a).

Through the structural superpose of compounds KAR_1 and KK181N1, I found that they have possessed a different spatial position to have a better accommodation in the pocket (Figure 3-7b). For KAR_1, as previously reported, its methyl group at the position 3 points to the catalytic nucleophile Ser95 in the general direction, leaving the oxygen-bearing edge of KAR_1 exposed to solvent[101]. For KK181N1, the 1, 2, 3-triazole ring directs to the deeper site of the pocket to establish the aforementioned hydrogen-bonding networks (Figure 3-4), and the hydrophobic edge of 7-methylindoline moiety points out of the pocket to have the access to solvent (Figure 3-7b). However, the KAR_1 and KK181N1 did share the commons in the orientation of oxygen atom that composes the lactone carbonyl groups of KAR_1 and the carbamoyl group of KK181N1, respectively (Figure 3-7b), implying some structural features for ligand recognition remained to be future analyzed. I observed some residues at the pocket undergo conformational shifts on their side chains upon KK181N1 binding when superposed the structures of *apo*-AtKAI2 and AtKAI2-KK181N1. The minor structural differences occurred in the side chains of residues F134, F157, L142, among which the benzene ring of F134 and F157 respectively flip 4.1° and 5.4° outward the aromatic face of KK181N1 (Figure 3-7c, d), and the alkyl side chain of L142 horizontally deviated 0.8 Å away from pyrrolidine plane of KK181N1 (Figure 3-7e). Additionally, the slight shifts of benzene ring of F157 seem to contribute the edge tilted-T aromatic interaction with 7-methylindoline moiety of KK181N1 (Figure 3-7d). Nevertheless, the binding of KAR_1 did not induce any significant structural shifts on these three residues compared with ligand-free status. The most pronounced conformational changes were detected in the side chain of F194 and L218. For F194, the binding of KK181N1 and KAR_1 all cause conformational fluctuation at almost the same degree where the benzene ring of F194 flip outward 79.0° to escape the steric hindrance with ligand body, and to further establish π-π stacking interaction with the *4H*-pyran of KAR_1 in the face-to-face (sandwich) geometry or with

pyrrolidine of KK181N1 in the parallel displaced orientations (Figure 3-7f). Notably, the $C_{(\gamma)}$-$C_{(\delta)}$ bond of L218 makes a self-rotation for 134.3° upon KAR_1 binding (Figure 3-7h) but causes dihedral turns over 163.0° along with 1.7 Å shift of $C_{(\gamma)}$ upon KK181N1 binding (Figure 3-7g). This is particularly consistent with our affinity assay of three L218 variants (have the substitution with A/V/I) that show stronger binding affinity to KK181N1 than the wild type (Figure 6, Table 3-2), and also examine our hypothesis with regard to accelerating the recognition of KK181N1 via modulating residue size at the position 218. Collectively, these structural details inform us that some conformational changes are required for a better adaption for KK181N1.

Previously, the in-house chemical library enabled us to discover an N-heterocyclic urea compound, named KK094, that makes an antagonistic effect on SL response through targeting D14, which is the paralogue of AtKAI2 and also belongs to α/β-hydrolase fold superfamily[194]. However, the structural comparison of OsD14-KK094 and AtKAI2-KK181N1 (RMSD: 0.699) demonstrates that KK094 and KK181N1 show different binding modes in the active pocket, despite that they only differ in one methyl group in indole ring (Figure 3-8a). In the structure of OsD14-KK094 complex, KK094 was covalently attaching to the hydroxy group of the nucleophile S147 (OsD14) (Figure 3-8b), and in order to initiate the nucleophile attack on the carbamoyl moiety of KK094, the side chain of S147 has flipped toward ligand by 26.8°, whereas spatial position of S95 in *apo*-AtKAI2 essentially resembled that in AtKAI2-KAR_1 and AtKAI2-KK181N1 (Figure 3-8b, c). Through structural analysis, I further found similarities in the ligand recognition of both receptors. Upon the KK094CM-binding to OsD14, the side chain of F245 moved 1 Å away from the KK094CM, making a space to accommodate KK094 binding. Similar event (but more violently, as stated above) occurred in F194 when KK181N1 is accommodated in the active pocket of AtKAI2 (Figure 7f, Figure 3-8b, c). As reported, the hydrolytic product KK094CM extend its carbonyl group to form a hydrogen bond network with the amide backbone of F78 and V148, the imidazole ring of H297, and two water molecules (Figure 3-8d). Interestingly, even though KK181N1 is noncovalently embedded in the pocket of AtKAI2, I also detected this hydrogen bond network mediated by four water molecules (as mentioned in section 3.3.2), connecting to the 1,2,3-triazole moiety of KK181N1, the amide backbone of F26 (F78 in OsD14) and V96 (V148 in OsD14), the imidazole ring of H246 (H297 in OsD14), and the hydroxy group of the nucleophile S95 (S147 in OsD14) (Figure 3-4, Figure 3-8e). Furthermore, it is clear to see the spatial location of two water molecules

in OsD14-KK094 is closed to that in AtKAI2-KK181N1. In OsD14-KK094, these two water molecules constitute the hydrogen-bonding network in a triangle formed between W1, W2, and the amide backbone of F78 and further extend to the hydroxyl group of Y209 (F156 in AtKAI2) (Figure 3-8d). This triangle-shaped hydrogen-bonding network is conserved in AtKAI2-KK181N1 yet lacking the extension due to residue difference (F156) (Figure 3-8e). More importantly, I unexpectedly noticed that the oxygen in the carbonyl of KK094CM is almost superposed with that in one water molecule (W3) in AtKAI2, and another triangle-shaped hydrogen-bonding network triggered by the carbonyl group of KK094CM is the same as that initiated by W3 in AtKAI2-KK181N1 (Figure 3-8d, e). Notably, one part of hydrogen bond network, mediated by a water molecule (W4), stretches to the hydroxy group of S119 and Y124 in AtKAI2-KK181N1 complex, but not to the corresponding residues G171 and F176 in OsD14-KK094 complex due to the lack of hydroxy group and water molecule (Figure 3-8d, e).

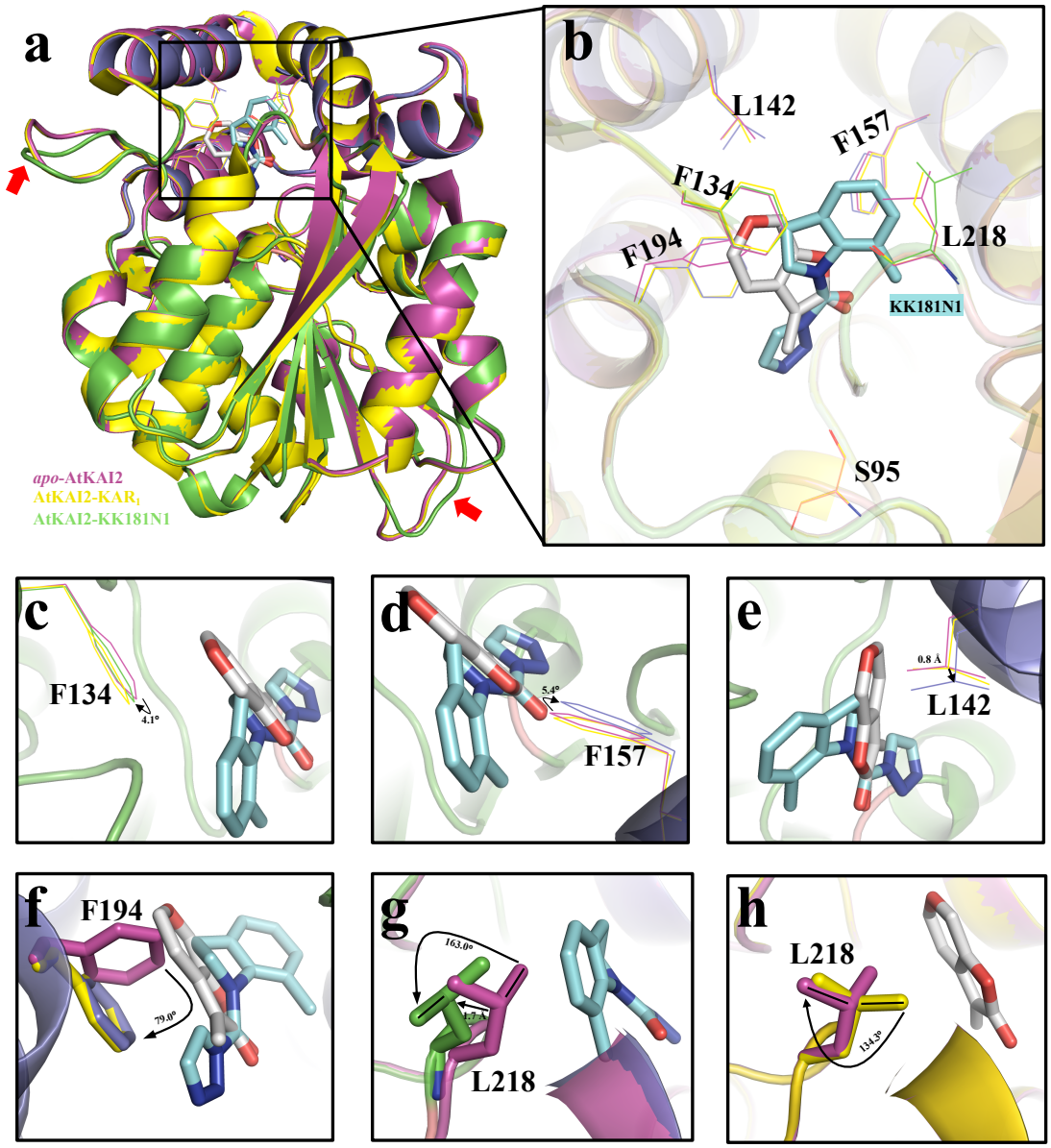


Figure 3-7: Comparison of ligand-free and ligand-bound structure of AtKAI2.
a, Structural overlay of *apo*-AtKAI2 (PDB code: 4JYP, magentas), AtKAI2-KAR$_1$ (PDB code: 4JYM, yellow), and AtKAI2-KK181N1 (in this study, green for those residues located in core domain and purple for that in cap domain). Two major deviations of protein backbone are indicated as red arrows. **b,** close-up view of ligand-binding pocket. KAR$_1$ (white) and KK181N1 (cyan) are represented as sticks. **c-g,** Five residues [F134 (**c,** green), F157 (**d,** purple), L142 (**e,** purple), F194 (**f,** purple) and L218 (**g,** green)] of AtKAI2-KK181N1 showing structural differences against that of *apo*-AtKAI2 are represented, among which F134, F157, and L142 that only make minor conformational shift are indicated as lines, and F194 and L218 that strongly undergo conformational changes upon ligand binding are indicated as sticks. **h,** Conformational changes of L218 between *apo*-AtKAI2 and AtKAI2-KAR$_1$.

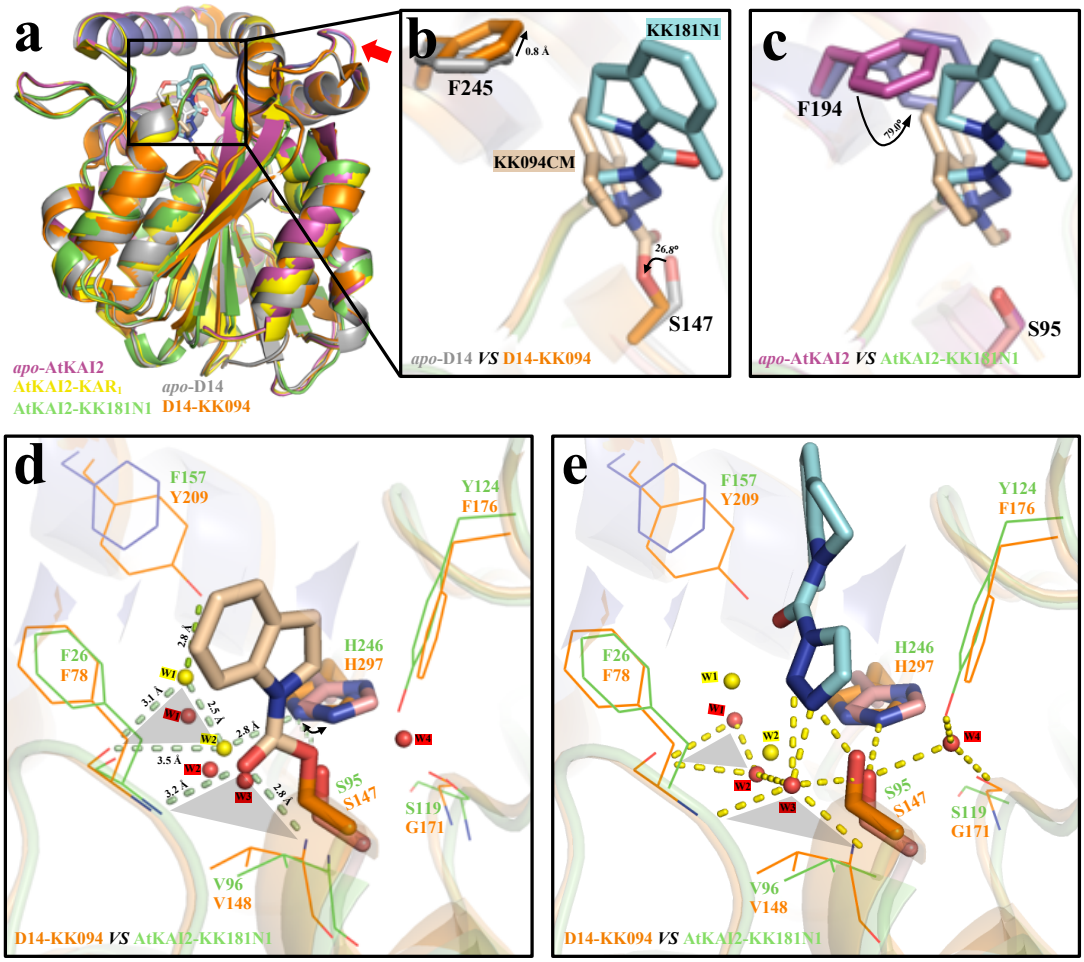


Figure 3-8: Comparison of ligand-free and ligand-bound structure of AtKAI2 and OsD14. **a,** Structural overlay of *apo*-AtKAI2 (PDB code: 4JYP, magentas), AtKAI2-KAR$_1$ (PDB code: 4JYM, yellow), AtKAI2-KK181N1 (in this study, green for those residues located in core domain and purple for that in cap domain), *apo*-OsD14 (PDB code: 3VXK, gray), and OsD14-KK094CM (PDB code: 5ZHR, orange). One major deviation of protein backbone between AtKAI2 and OsD14 is indicated as red arrows. **b-c,** close-up view of ligand-binding pocket. The side chains of F245 and catalytic residue S147 undergo conformational shift for KK094 (wheat) binding in OsD14, whereas KK181N1 (cyan) binding in AtKAI2 only cause major structural change on the corresponding residue F194 not on catalytic residue S95. These residues are shown as sticks without main chain. **d-e,** Differences of the hydrogen bonding networks in OsD14-KK094CM (**c**) and AtKAI2-KK181N1 (**d**). Catalytic residues (pink in AtKAI2, orange in OsD14) are indicated as sticks, others as lines model. Two water molecules from OsD14-KK094CM are labeled as W1 and W2 in yellow, and four from AtKAI2-KK181N1 as W1 to W4 in red. Hydrogen bonds labeled with measured length in black values (Å) are colored pale green in OsD14-KK094CM and yellow in AtKAI2-KK181N1. The triangle-shaped hydrogen-bonding networks are indicated as the gray shadow.

3.2.6 Selectivity studies of KK compounds for AtKAI2

The sequence alignment (Figure 3-9, Figure 3-10) and structural architecture enlighten us that, even KAI2 and D14 share high sequence identity and structural arrangement, the different side chains of residues involved in assembling the pocket may define ligand selectivity. These residues, including Y124 (F176 in OsD14), L139 (I191 in OsD14), L142 (V194 in OsD14), F157 (Y209 in OsD14), I193 (V244 in OsD14), L218 (V269 in OsD14), and A219 (S270 in OsD14) (Figure 3-9, Figure 3-11a), build the hydrophobic walls to recognize the different parts of ligand body. Here, since the only structural difference between KK094 and KK181N1 is the methyl group, our following studies were paid to the residues involved in the interaction with methyl group of KK181N1. As we can see in Figure 3-11b, residues A161, F157, L218, A219, and H246 provide rich aromatic interaction with methyl group of KK181N1, but only three residues (F157, L218, and A219) are divergent in the identical position of OsD14, corresponding to Y209, V269, and S270 (Figure 3-9, 3-11b). As we have discussed in the last section, I concluded that the major role of L218 is to recognize ligand KK181N1 at the entrance of pocket based on the observation that the variant L218V promotes the ligand entry (Figure 3-7g, 3-11c). The superposed structure shows that the hydroxyl group of Y209 (in OsD14) causes strong steric effects on the carbonyl group and the methyl group of KK181N1, which may define ligand selectivity for D14 (Figure 3-11d). While, interestingly, KK181N1 and S270 (ligand-free structure of D14) do not have any steric interaction (Figure 3-11e), which, however, is observed between the methyl group of KK181N1 and the hydroxyl group of S270 (KK094-bound structure of D14) since the binding of KK094 induces the conformational change of S270 by flipping its hydroxyl group over 130.1°outwards pocket (Figure 3-11d). Thus, we prepared a variant of AtKAI2 containing the serine substituent on A219 to see whether ligand-binding would also cause such conformational change on its side chain, leading to steric effects on the KK181N1. Unexpectedly, the mutant A219S shows 2 folds higher binding affinity to KK181N1 than wild type (Table 3-2).

Previously, Lee et al. identified a light-signaling mutant deficient in KL-signaling (designated as *ply2*) thorough genetic screening. Further work confirmed that this phenotype is caused by a single amino acid substitution in AtKAI2, A219V[207]. X-ray crystallography studies supported their idea that Valine would not only result in a narrowed entrance gate for the ligand but also alters the structural flexibility of the helical lid domains[207]. To explain the improved affinity of AtKAI2^{A219S}

for KK181N1, I superposed the structures of AtKAI2^{ply2} with AtKAI2-KK181N1, *apo*-OsD14, and OsD14-KK094. It is obvious to see that KK181N1 and KAR_1 are all well accommodated in the entrance tunnel built by the residues, including A219 and F157 (Figure 3-12a). However, for mutant A219V, one ($C_{\gamma1}$) of two C_γ atoms extends to narrow down the width of the tunnel, causing slight steric clash to KAR_1 but severe to KK181N1 on the aromatic plane, whereas $C_{\gamma2}$ did not make any steric effect on ligands (Figure 3-12c). In the aspect of chemical structure, Ser differs from Val by lacking two methyl substituents ($C_{\gamma1}$ and $C_{\gamma2}$) on the β carbon but having a hydroxyl substituent (O_γ) in place of one of the hydrogens on the β carbon of Ala (Figure 3-12b). Suggested by the superposed structures where the $C_{\gamma2}$ of V219 (in AtKAI2^{ply2}) almost overlays with the O_γ of S270 (in *apo*-OsD14) Figure 3-12d), I thus assumed, when marginally conformational shift is allowed, S219 (in AtKAI2) would not cause any steric hindrance on ligand-binding, but probably promote some additional interaction with ligands instead, including KK181N1 and KAR_1, so that increase the ligand-binding affinity. Taken together, I conclude that the recognition of methyl group of KK181N1 relies on the contribution made by each residue surrounded, including A161, F157, L218, A219, and H246. Even though residues S270 and V269 in OsD14 (corresponding to the variants A219S and L218V in AtKAI2) may promote ligand recognition for KK181N1, the steric clash between the hydroxyl group of Y209 (OsD14) and the carbonyl group and the methyl group of KK181N1 will not allow the acceptance of KK181N1 in OsD14 (Figure 3-11a, d), leading to the selectivity of KK181N1 toward AtKAI2 but not OsD14. What's more, considering the differences of residues lining the pocket (Figure 3-11a) and the hydrogen-bonding network (Figure 3-8d, c) of AtKAI2 and OsD14, I speculated that the perception of different ligands is a process that needs the co-movement of all residues, just as the proposed models that the ligand selectivity (between KARs, and 2'*R* and 2'*S* configured SLs) of HTL/KAI2 and D14 are conferred by the residues defining the volume of pocket[143, 208-212]. A recent study even suggested that co-modulation of residues at three positions are indispensable to repurpose karrikin receptor to a strigolactone receptor [213]. In this study, however, referring to my physiological results and affinity assay that KK181N1 only antagonizes the AtKAI2 signaling and KK094 is less active to bind with AtKAI2, I still cannot explain the ligand selectivity of AtKAI2 toward KK181N1 but not KK094 using the current structural model.

To confirm that this methyl substituent is most favored by the hydrophobic environment built by A161, F157, L218, A219, and H246 in AtKAI2, I tested an array of derivatives (Figure 3-13) that

are newly prepared by Mr. Kikuzato in our laboratory and have different substituents on C13 of KK181N1. I measured their binding affinity to AtKAI2 by ITC experiment *in vitro*. These derivatives include KK094N1 (hydrogen group), KK167N1 (nitro group), KK175N1 (chloro group), and KK187N1 (fluoro group) (Figure 3-13), all of which displays the alleviated affinity to AtKAI2 (Table 3-3). It is expected that the binding affinity of AtKAI2 to KK167N1 is barely detectable due to the bulky size of the nitro group (Figure 3-11b). However, KK094N1 having less-occupied space of hydrogen is also not preferred in this hydrophobic area of AtKAI2, leading to K_d of 200 μM.

Halogens are present in around 25% of drugs, and frequently used as bioisosteric replacements for -H, -CH_3, -OH and -NH_2 (according to 'Guide to Pharmacology Database', https://www.guidetopharmacology.org/). However, in this study, the halogen substituents, chloro or fluoro, only remain 15% of the binding affinity of AtKAI2 to KK181N1 (Table 3-3). Combining the structural information, I speculated that even though halogens may have a great capacity to form halogen-π interactions with the electron-rich π-systems[214], such as pointing to the π-cloud of the imidazole ring of H246 (Figure 3-11b), the CH-π interaction seems to be more favorable under some circumstances when the contact distances, angle, and non-polar environment are favored[215, 216]. More details regarding 'Molecular Interactions' can refer to the webservice (https://www.cambridgemedchemconsulting.com/resources/).

Next, I inspected the contribution of the hydrogen-bonding network formed on the 1,2,3-triazole moiety of KK compounds to the ligand specificity. A serial of compounds constituted by divergent moieties, including the N1- and N2-carbamoylated regioisomers of triazole, were prepared again by Mr. Kikuzato. The affinity of these compounds toward AtKAI2 were measured by ITC machine *in vitro*. Firstly, by comparing the affinity of KK200 and KK199 with that of KK094, I noticed that furan moiety is tolerated in the pocket of AtKAI2, but this will depend on the size of the carbamoyl group since KK199 did not have any affinity toward AtKAI2. Those compounds (KK199, KK078, KK118, and KK2028) containing a bulky moiety on the other side will also not be recognized by AtKAI2. Therefore, we can imagine that KK200 still have the chance to engage in the hydrogen bonding network at the bottom of the pocket through its furan moiety. Based on the knowledge that the strength of hydrogen bond O–H···N (29 kJ/mol) is slightly stronger than that of O–H···O (21 kJ/mol), but ITC results suggested that the affinity of KK094N2 is actually weaker than that of KK200, possibly implicating a different binding pose for furan-containing compounds.

And then, the affinity comparison between KK181N1 and KK181N2 or KK094N1 and KK094N2 apparently reveals that AtKAI2 have a higher affinity to 1,2,3-triazole ureas with N1 regioisomers. This result is consistent with the structural insight with respect to the construction of hydrogen-bonding network. As we can see from the clear details on the formation of hydrogen bonds (Figure 3-4), N1-configured 1,2,3-triazole enables one of its three nitrogen atoms (N5, indicated by green shadow in Figure 3-1g), acting as the hydrogen bond acceptor, to attract one hydrogen atom from a water molecule (W3), and another nitrogen atom (N6, indicated by blue shadow in Figure 3-1g) to seize two hydrogen atoms respectively donated by a water molecule (W3) and a hydroxyl group of S95. However, N2-configured 1,2,3-triazole can only form one hydrogen bond between N5 atom with water molecule (W3), which is weaker than the two hydrogen bonds built by N6 atom. As the consequence, N1-configured 1,2,3-triazole has extremely higher affinity than N2-configured compounds. This result can be further explained by comparing KK122 and KK094N2. Even in the absence of N5 atom, the harboring of N6 could reinforce the binding affinity of KK122 to approximately 3 folds higher than KK094N2, or 2 folds than KK181N2, stating that N6 atom-formed hydrogen bonds are vitally important for the binding affinity of 1,2,3-triazole-containing chemicals toward AtKAI2.

Taken together, chemical selectivity studies toward AtKAI2 suggested that, for one hand, the methyl substituent in indoline group of KK181N1 is much better accommodated to the hydrophobic environment formed by non-polar residues, and for another, N1-configured 1,2,3-triazole would strengthen the water molecules-mediated hydrogen-bonding network so that increase ligand-binding affinity.

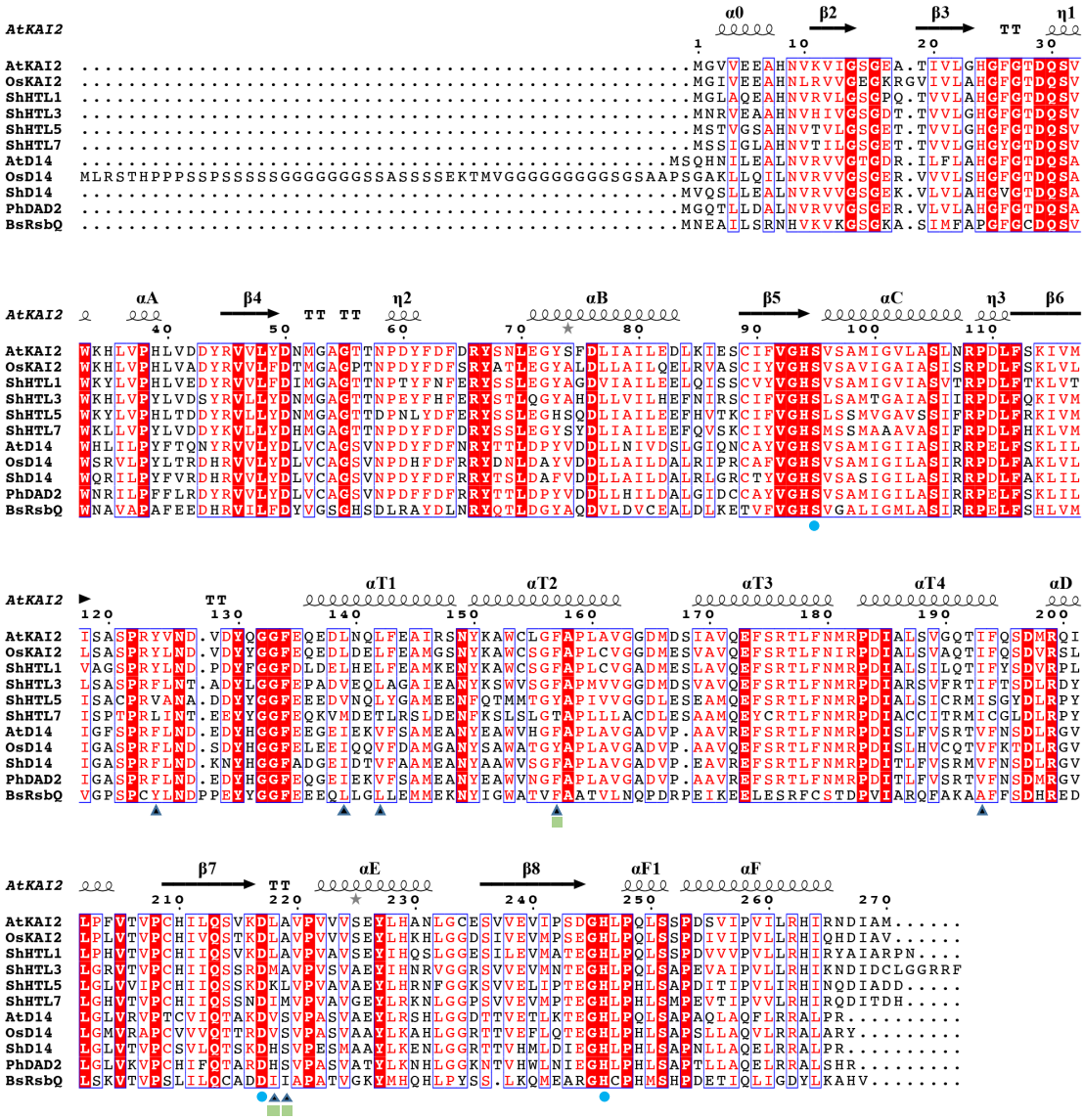


Figure 3-9: The sequence alignment of KAI2/HTL and D14 proteins.

Proteins are from rice (Os), Arabidopsis (At), striga hermonthica (Sh), petunia (Ph, D14 ortholog DAD2), Bacillus subtilis (Bs) RbsQ. The secondary structure assignment is based on the crystal structure of AtKAI2 and the labels of the major strands and helices are based on those of the canonical α/β hydrolase fold. The three catalytic residues in AtKAI2, S95, D217 and H246, are indicated by light-blue dots. The residues involved in shaping active pocket of AtKAI2 but differed in OsD14 are indicated by black triangles. Three residues involved in recognizing the methyl of KK181N1 in AtKAI2-KK181N1 complex but differed in OsD14 are indicated by green squares.

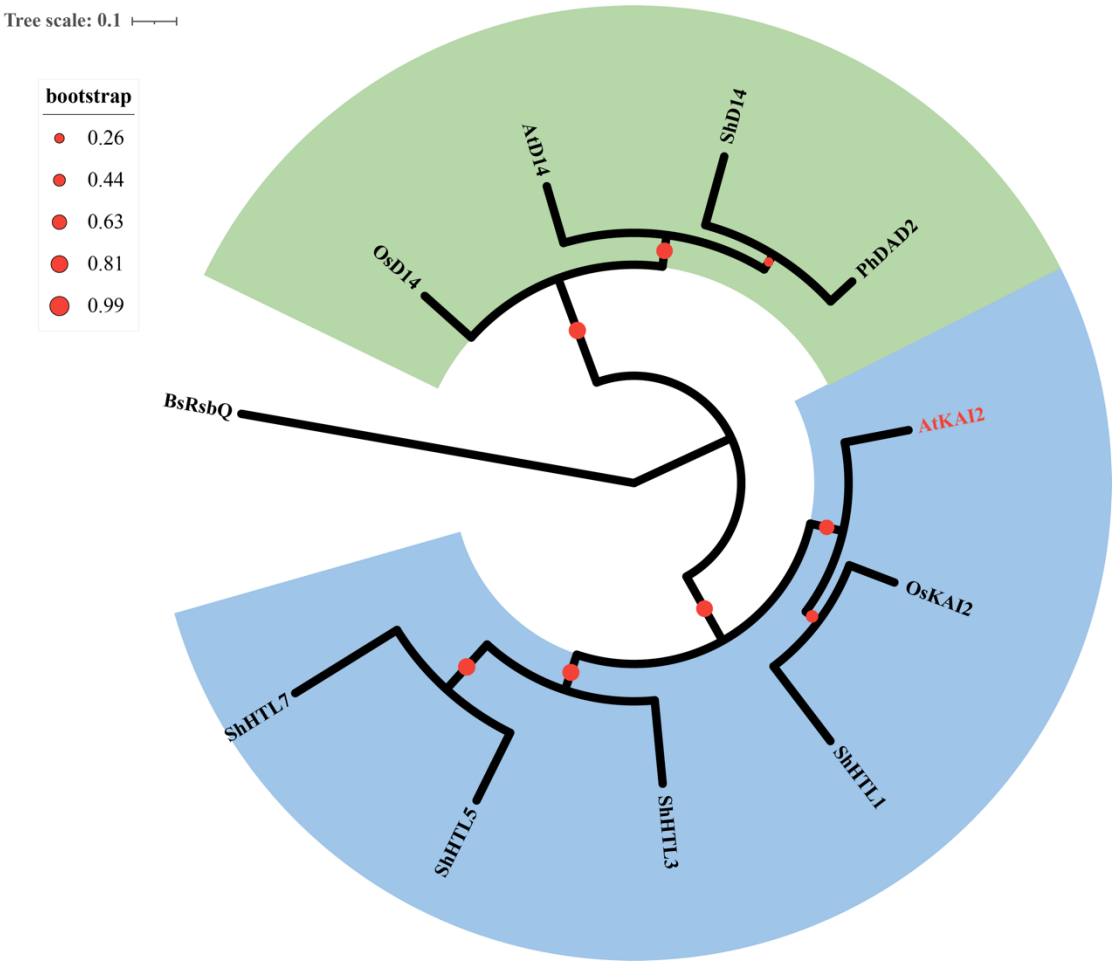


Figure 3-10: Maximum likelihood phylogeny of KAI2/HTL and D14 proteins.
Phylogenetic tree was constructed by Phylogeny.fr based on the sequence alignment of KAI2/HTL and D14 proteins from rice (Os), Arabidopsis (At), striga hermonthica (Sh), petunia (Ph, D14 ortholog DAD2), Bacillus subtilis (Bs) RbsQ. Bootstrap (red dots) is applied to indicate the confidence levels in a clade of phylogenetic tree. Two major clades are generated, one of which is colored green for the indication of D14 family proteins, the other one colored bule for KAI2/HTL family proteins.

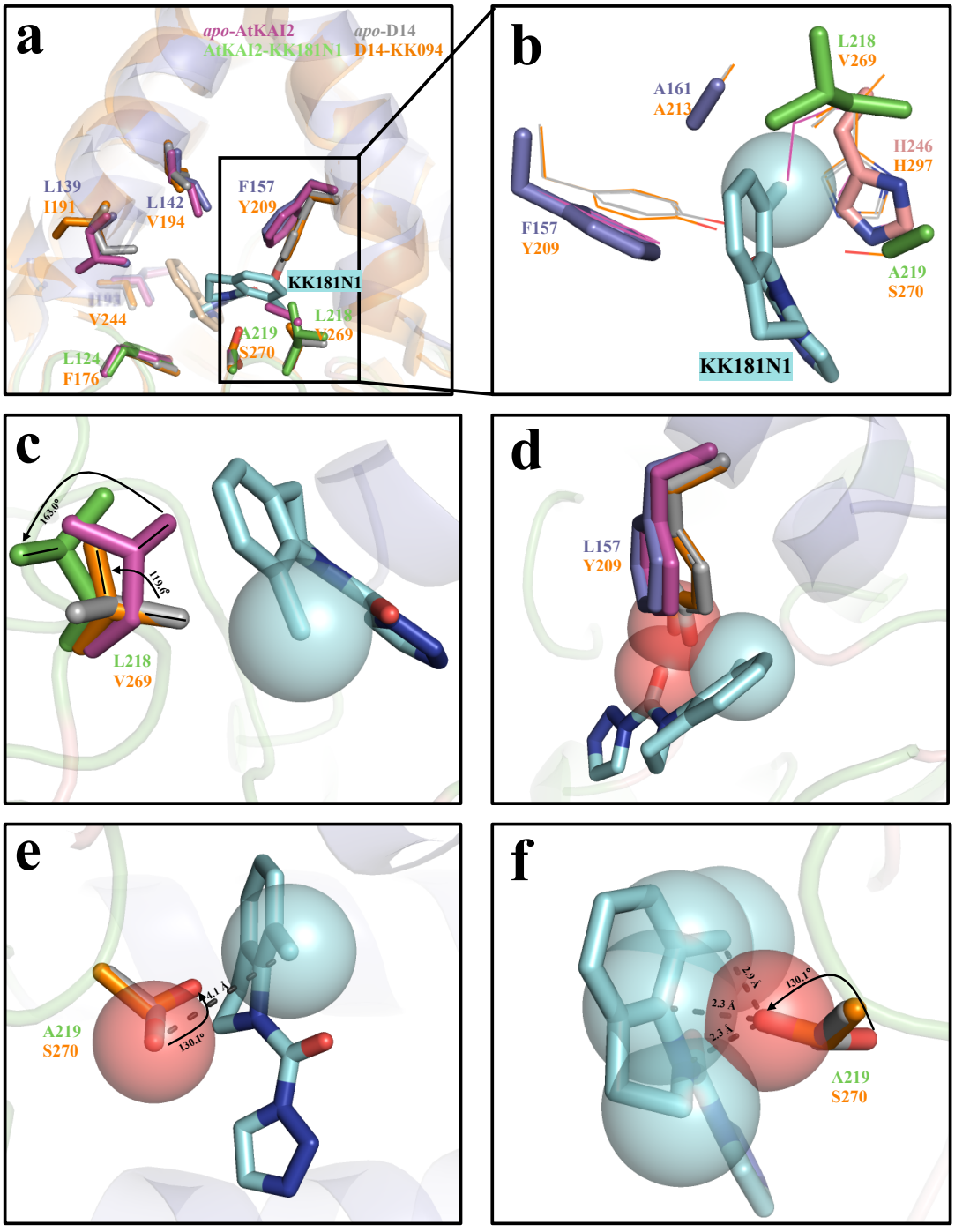

a
apo-AtKAI2
AtKAI2-KK181N1
apo-D14
D14-KK094
L139
I191
L142
V194
F157
Y209
KK181N1
I193
V244
L218
V269
A219
S270
L124
F176
b
L218
V269
A161
A213
H246
H297
F157
Y209
A219
S270
KK181N1
c
163.0°
119.6°
L218
V269
d
L157
Y209
e
2.1 Å
130.1°
A219
S270
f
2.9 Å
130.1°
2.3 Å
2.3 Å
A219
S270

Figure 3-11: Structural basis of ligand selectivity in AtKAI2.
In this figure, the residues from AtKAI2-KK181N1 and OsD14-KK094CM complexes are labeled only. Sphere model is used to indicate the van der Waal's radius. **a,** Structural overlay of *apo*-AtKAI2 (PDB code: 4JYP, magentas), AtKAI2-KK181N1 (in this study, green for those residues located in core domain and purple for that in cap domain), *apo*-OsD14 (PDB code: 3VXK, gray), and OsD14-KK094CM (PDB code: 5ZHR, orange). The residues involved in building the active pocket yet different among AtKAI2 and OsD14 are presented as sticks. **b,** A close-up view of the area built by hydrophobic residues for the recognition on methyl group of KK181N1. Residues of AtKAI2-KK181N1 complex are shown as sticks and others as lined model. Catalytic residues of AtKAI2-KK181N1 colored pink. Close-up views of relative spactial location of KK181N1 to L218/V269 (**c**), F157/Y209 (**d**), and A219/S270 (**e, f**) of AtKAI2-KK181N1/OsD14-KK094CM complex. The ligand-induced conformational changes of residues (sticks) are indicated. **c,** Variant L218V would promote the ligand entry. **d,** The hydroxyl group of Y209 may prevent the binding of KK181N1 because of the formation of steric clash with the carbonyl group and the methyl group of KK181N1. **e-f**, Conformational shift of S270 is occurred by flipping its hydroxyl group over 130.1º outwards pocket upon ligand-binding, so that causing the steric hindrance between the methyl group, and the edge of aromatic plane in KK181N1, and the hydroxyl group of S270 (KK094-bound structure of D14). The carbon atom of methyl group in KK181N1 and oxygen atom of hydroxyl group in S270 [in ligand-free structure (**e**) and KK094-bound structure (**f**) of D14] are shown as sphere, and the distance between them is measured in black dash with values (Å).

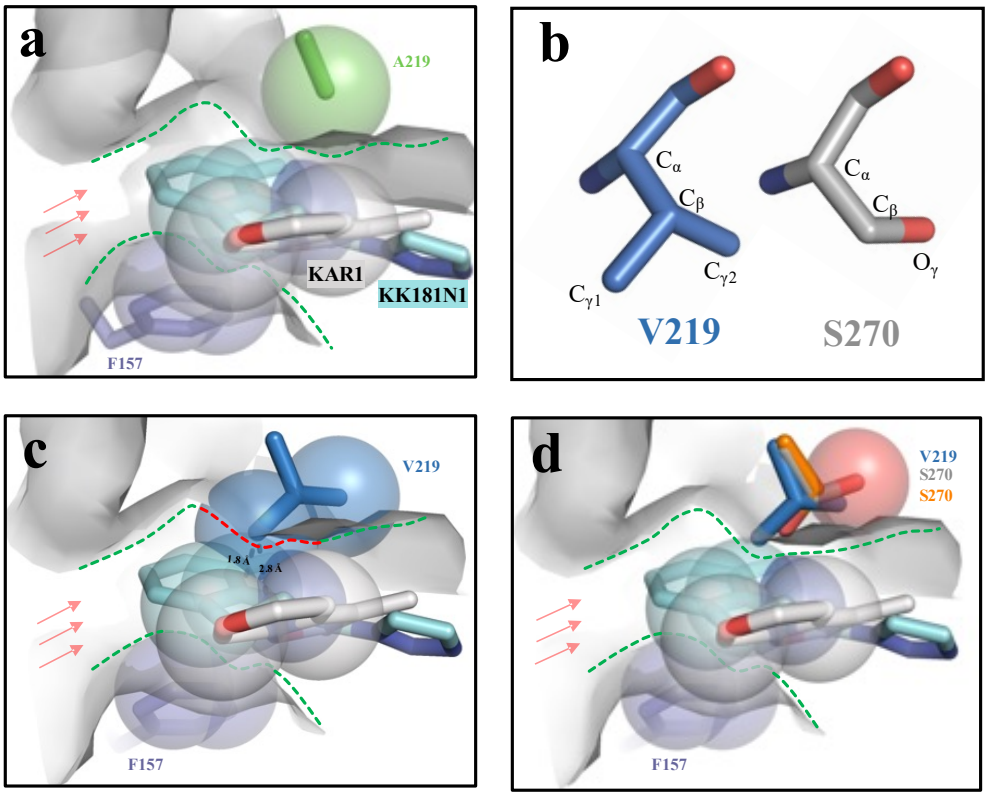
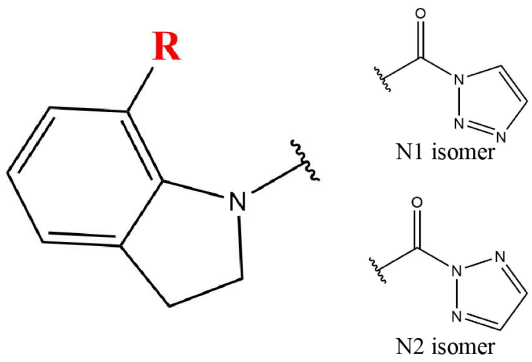


Figure 3-12: Structural basis for AtKAI2 variant A219S showing improved affinity to KK181N1.

a, A close-up view of pocket tunnel (indicated as gray surface and green dotted line) shaped by F157 (purple stick) and A219 (green stick) at Y-axis in AtKAI2-KK181N1 complex. Red arrows indicated the entrance of pocket. KAR_1 (white) and KK181N1 are shown as stick. Carbon atoms in KAR_1 (at the edge of aromatic plane), KK181N1 (at the edge and methyl group of aromatic plane), A219 (at C_β), F157 (at the top edge of aromatic plane) are covered with spheres to indicated the the van der Waal's radius. **b,** 3-D structures of residues Valine and Serine. The configuration of them is fixed respectively according to the V219 in $AtKAI2^{ply2}$ crystal structure and S270 in *apo*-OsD14 crystal structure. Carbon atoms are labeled. **c,** The steric hindrance (red dotted line) between V219 (blue stick) and ligands, including KAR_1 and KK181N1, is observed. Both $C_{\gamma1}$ and $C_{\gamma2}$ are indicated as transparent spheres. The two closest distances of $C_{\gamma1}$ in V219 to the aromatic plane of KK181N1 are measured with black values, suggesting steric effects. **d,** Residues overlay at the position 219 in AtKAI2 and 270 in OsD14. The O_γ (indicated as transparent spheres) of S270 (indicated as gray sticks) in *apo*-D14 will not cause steric clash with ligands in the pocket tunnel. The S270 in D14-KK094 complex is shown as orange stick.



Compounds	R
KK181N1	CH_3
KK094N1	H
KK167N1	NO_2
KK175N1	Cl
KK189N1	F

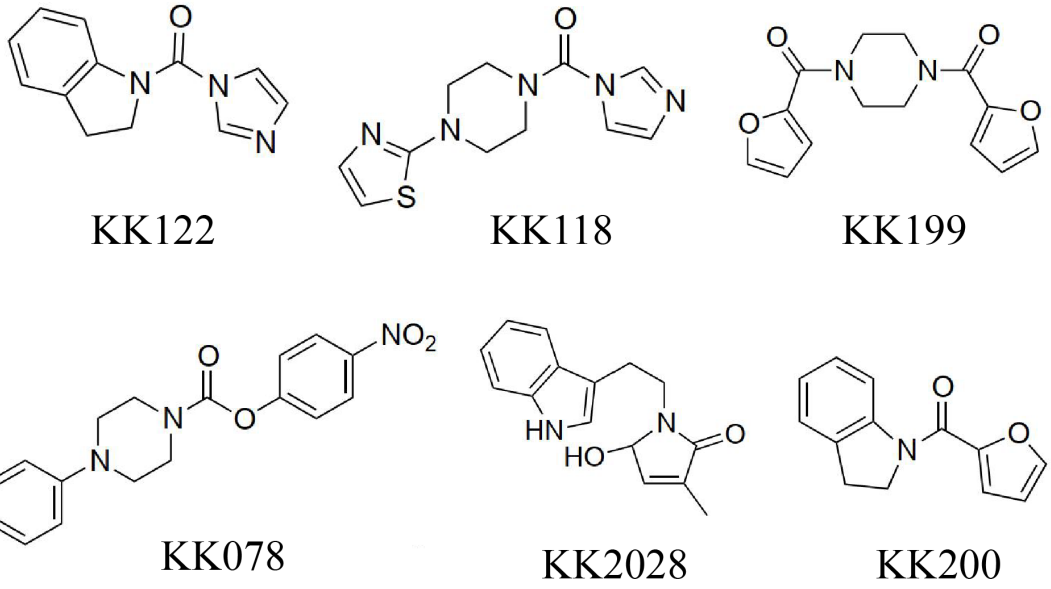
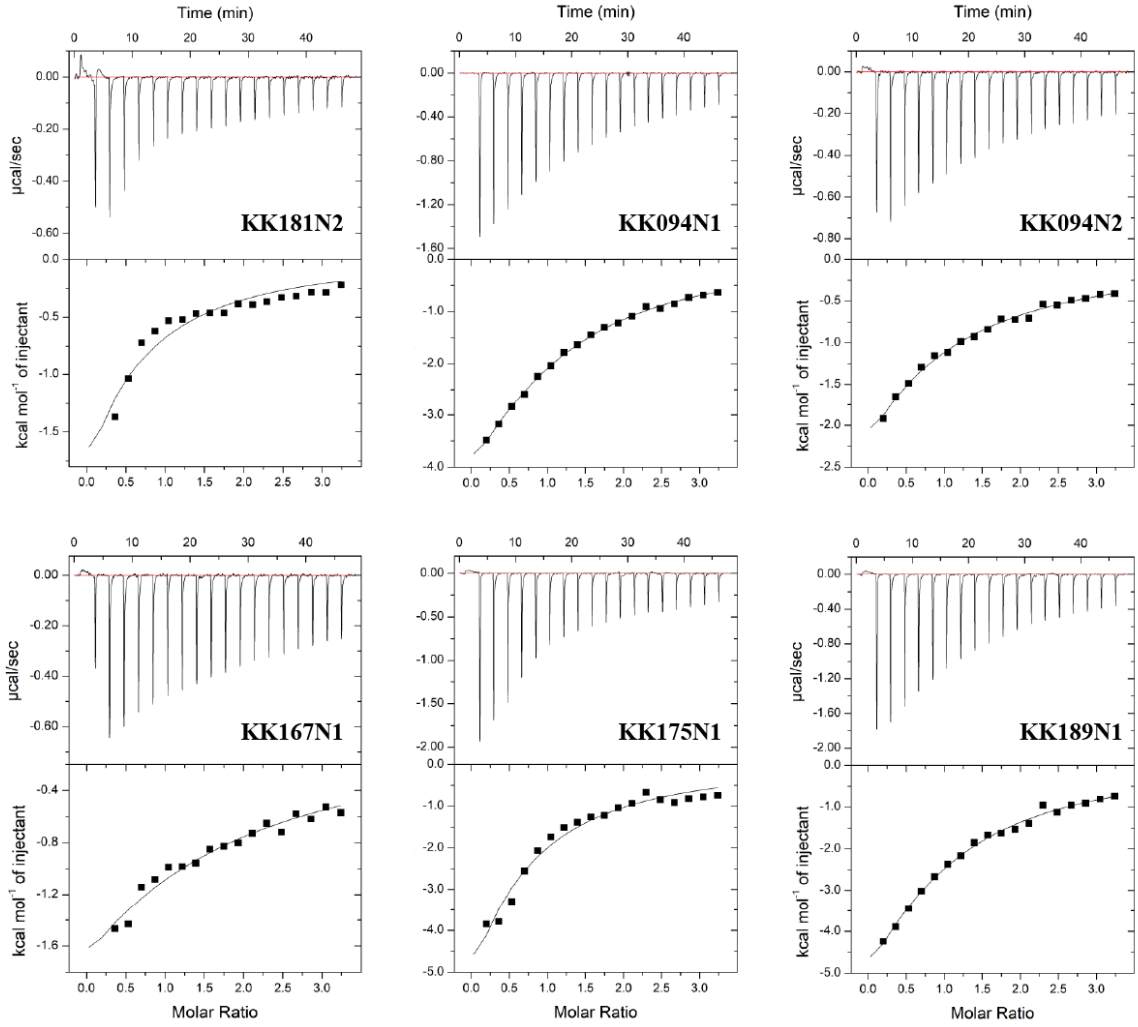


Figure 3-13: 2-D structure of KK compounds used in this study.


Time (min)
µcal/sec
kcal mol^{-1} of injectant
Molar Ratio
KK181N2
KK094N1
KK094N2
KK167N1
KK175N1
KK189N1

(Continued)

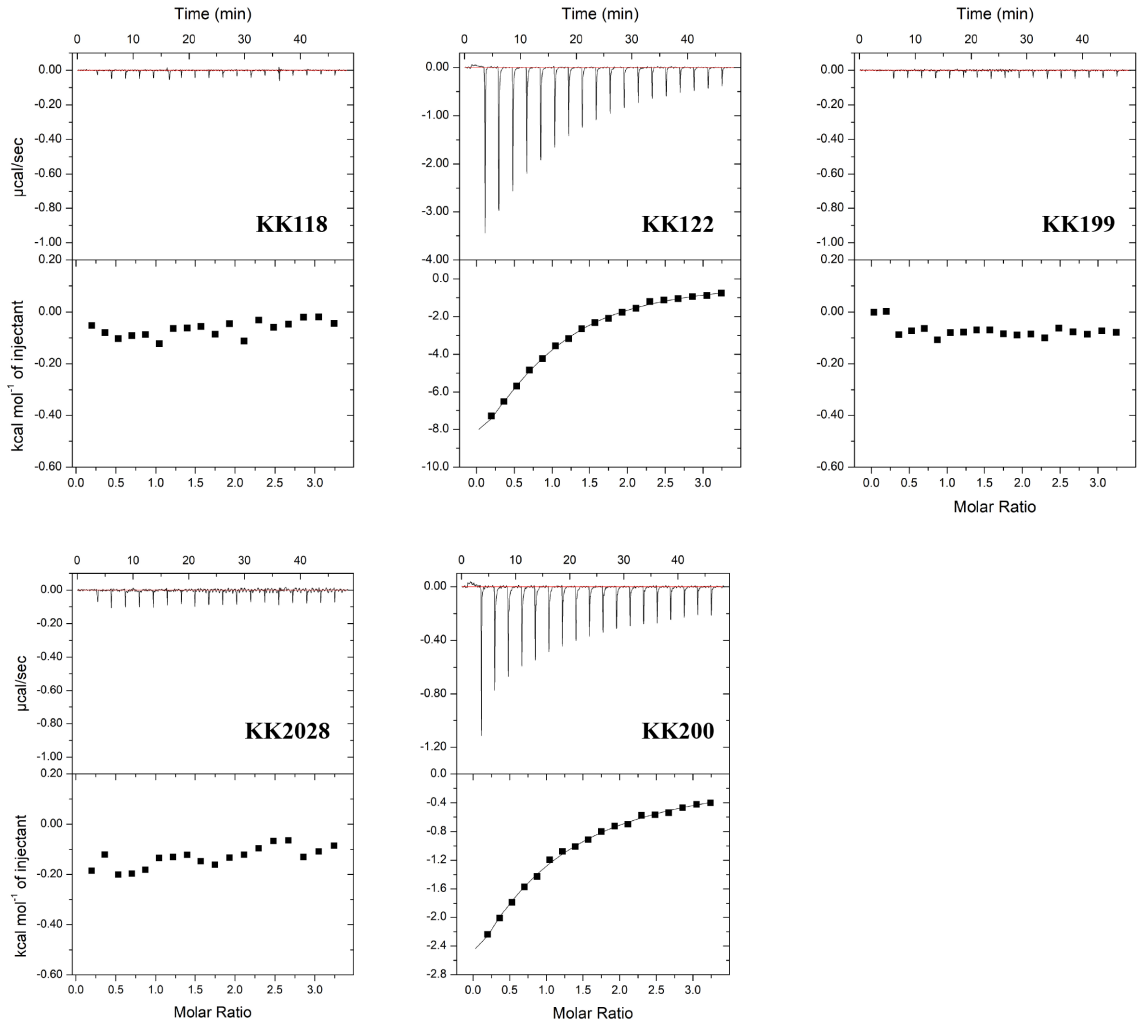


Figure 3-14: Isothermal titration calorimetry of wild type AtKAI2 to different chemicals. Representative ITC thermograms titrating AtKAI2 with N-heterocyclic ureas. The data for ITC were fitted using the 'one binding site' model by the MicroCal origin software. The panels show the ITC titration curves (µcal s^{-1}) and the integrated heats of injection (kcal mol^{-1}). n=>2.

3.2.7 Preliminary investigation of the mode of action of KK181N1

Although some details are still under debate, the general pathways of KARs/KL response are postulated to resemble the well-defined mode of action of SL-signaling, where activated receptor KAI2 regulated the activity of SCF-type E3 ubiquitin ligase complex by binding with one of its components F-box protein MAX2, so that targets members of the SMXL family for polyubiquitination and proteasomal degradation[122, 141, 217-220]. Remarkably, the *rac*-GR24-induced association of KAI2 with SMAX1 or SMXL2 has been comprehensively investigated by genetic manipulation[141, 221], luminescence assays[137], co-immunoprecipitation (Co-IP) assay[137, 142], and yeast two-hybrid (Y2H) system[137]. However, the direct evidence and systematical analysis of KAI2-MAX2/D3 interaction are still in demand. Y2H assay is an effortless but persuasive method that has been extensively applied to study protein-protein, protein-ligand, and protein-DNA interaction[222, 223]. Here, I used the Y2H assay to explore the AtKAI2-SMAX1 and AtKAI2-MAX2 interaction stimulated by KL agonists and try to examine whether these two interactions would be blocked by the KL antagonist KK181N1.

As the ortholog of OsD53, SMAX1 could be accordingly designated as five main domains based on sequence homology, including N-terminal Double Clp-N motif (N), putative ATPase domain 1 (D1), middle region (M), and C-terminal putative ATPase domain 2 (D2a and D2b) domain (Figure 3-15a)[122, 139]. Recent Y2H study has reveals that, induced by *rac*-GR24, AtKAI2 could weakly interact with the full length of SMAX1 or the truncation $SMAX1_{D1MD2}$, whereas sufficiently interact with $SMAX1_{D1M}$[137]. More interestingly, D1M domain of SMAX1 and SMAX7 is responsible for defining their specific response to KAR_1- or SL-induced degradation *in vivo*[137]. However, in my experiment, I failed to neither detect the interaction of AtKAI2 with SMAX1 (Full length) nor $SMAX1_{D1M}$ under stringent selection (-His-Ade media; -LTHA) (Figure 3-16a). Previous work suggested that although the M and D2 domains of D53 would diminish the GR24-dependent interaction between D14 and D53 D1 domain, the addition of the N domain expeditiously overcome this adverse effect[139]. Therefore, I made the construct $SMAX1_{ND1M}$ and found that the interaction between $SMAX1_{ND1M}$ and AtKAI2 was somewhat independent of KL agonists but was enhanced by applying KL agonists (Figure 3-16b), which is consistent with D14 and $D53_{ND1M}$[139].

However, the establishment of Y2H system for AtKAI2 and MAX2 is relatively struggled. Especially, the full-length MAX2 seems to be difficultly expressed and unstable in yeast cells[209, 213].

Some researchers suggested using specific fragments of $MAX2_{N3}$[213] for Y2H or $D3_{CTH}$[122] for pull-down assay, but I failed to detect all interaction of AtKAI2 with $MAX2_{full}$/$MAX2_{N3}$/$MAX2_{CTH}$ or its corresponding length in D3 (Figure 3-17). Inspired by the complex structure of D14-D3-CLIM-ASK1, I made the constructs $MAX2_{X1}$ and $D3_{X1}$, corresponding to the interaction interface of D3 and AtD14 (Figure 3-15b). As shown in (Figure 3-17), I was able to detect the chemical-dependent interaction of them, which is particularly strengthened by germinone A, a potent KL agonist[224].

In response to this result, I applied the AtKAI2-$MAX2_{X1}$ or AtKAI2-$SMAX1_{ND1M}$ to investigate the mode of action of KK181N1, it was strange to see that the treatment of KK181N1 will not inhibit, but promote the interaction of them, and the single treatment of KK181N1 can also induce the protein-protein interaction (Figure 3-17). To further confirm the role of KK181N1, I prepared some AtKAI2 mutants at the key sites, which have been shown no binding affinity to KK181N1 in our ITC measurement, including S95A, H246A, D217A and a positive control S95C which still possesses the binding affinity to KK181N1. However, the results implied that these variants are prone to promiscuous interactions. For example, variants S95A, H246A, and D217A cause severe auto-activation on *His*-negative medium only containing DMSO.

Y2H system has been successfully applied to detect the interaction of AtKAI2 and MAX2[225]. Nevertheless, some group insist that this specific association is difficult to be distinguished from nonspecific KAI2 interactions[137, 202]. Yao et al. reported that some mutants of AtKAI2 are expressed in a sticky form that is capable of autoactivating GAL4 without partner protein, or in an irregular form that makes similar performance as the GR24-activated wild-type protein did, so that activate GAL4 even in the absence of *rac*-GR24[202]. Regardless, these Y2H results (Figure 3-16, Figure 3-17, Figure 3-18) are inadequate to confidently illustrate the functional roles of KK181N1 between AtKAI2 and its downstream partner proteins. And we reckoned that the remarkable evidence from the Co-IP assay would be quietly required for the systematical investigation of the mode of action of KK181N1.

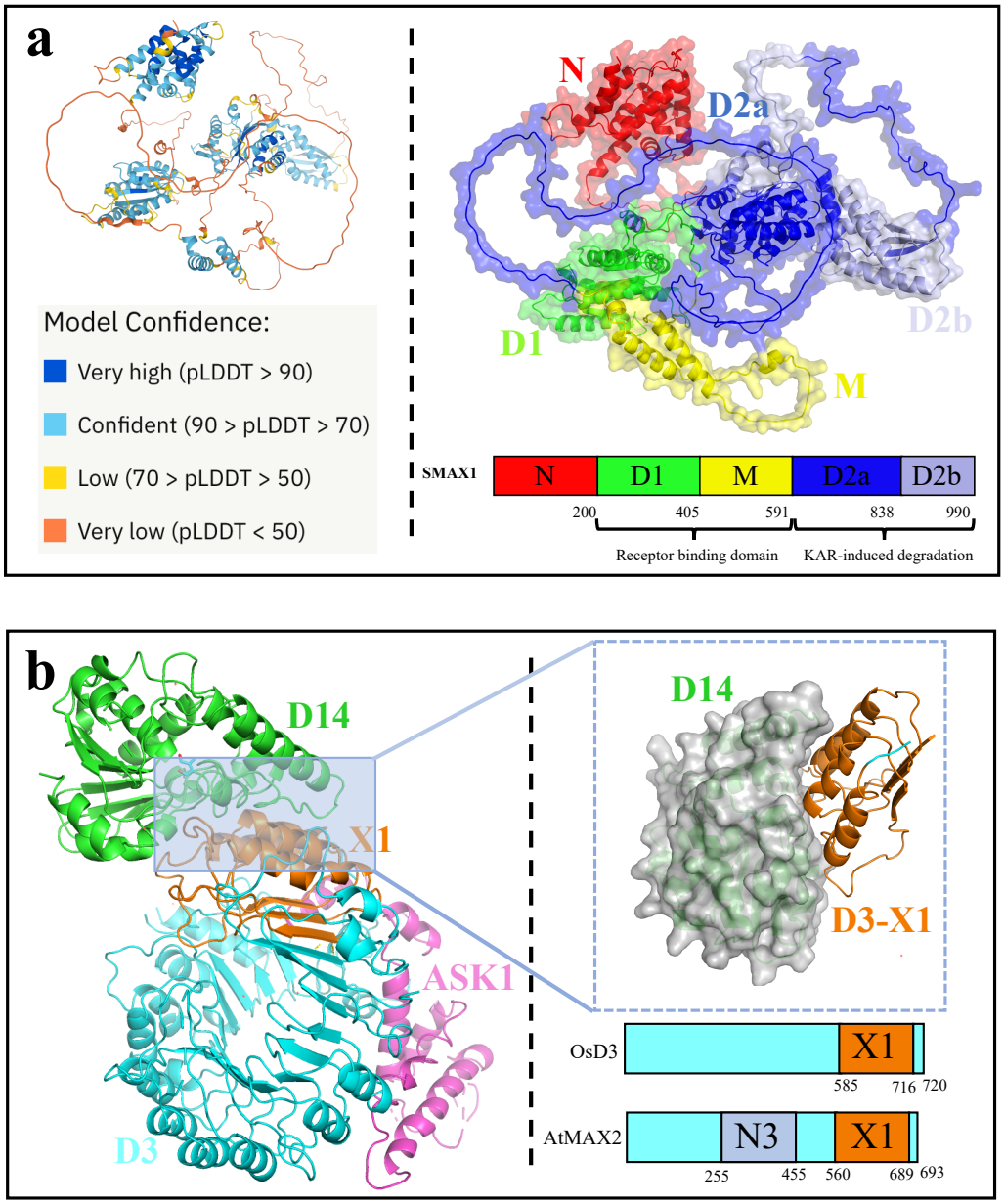


Figure 3-15: Diagram of the major domains of the SMAX1 and MAX2/D3 proteins.
a, Model of the three-dimensional structure of the SMAX1 protein retrieved from AlphaFold database (UniProt Code: Q9FHH2). Left panel: A per-residue confidence score (pLDDT, 0-100) is produced to map the model confidence. Right panel: Four major domains of SMAX1 are colored, and the amino acid numbers at the boundaries of each domain are indicated: N (red, 1-200), D1 (green, 201-405), M (yellow, 406-59), and D2 (blue, 592-990). D2 is further divided into D2a (dark blue, 592-838) and D2b (light blue, 839-990) subdomains. **b,** Left panel: An overview of the AtD14-D3-ASK1 complex structure (Cartoon model, PDB code: 5HZG). D14 is colored green; D3 is colored cyan; and ASK1 is colored pink. The interaction interface between AtD14 and D3 is colored orange. Top-right panel: The close-up view of AtD14-D3 interface. Lower-right panel: The indication of MAX2/D3 proteins for Y2H assay. The amino acid numbers at the boundaries of each constructs are indicated. For D3: X1 (orange, 585-716), CTH (cyan, 693-720); For MAX2: N3 (light blue, 255-455), X1 (orange, 560-689), CTH (cyan, 667-693).

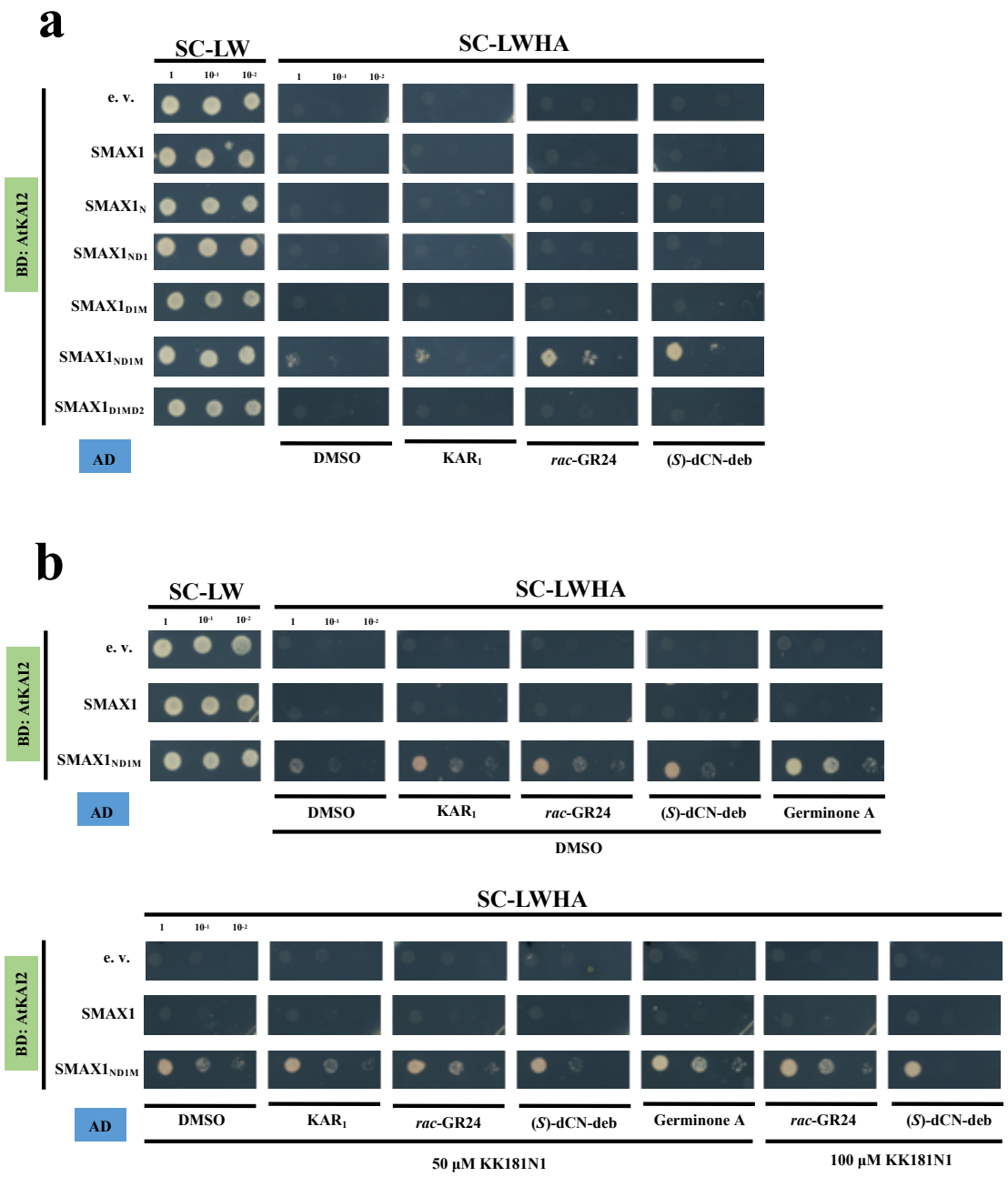


Figure 3- 16: Preliminary investigation of the mode of action of KK181N1 involved in AtKAI2-SMAX1 interaction by Y2H assay.

a, Inspection of ligand-dependent interaction of full-length and various domain deletion variants of SMXL1 with AtKAI2 in the presence or absence of 10 µM KL agonists, including KAR_1, *rac*-GR24, (*S*)-dCN-deb, and Germinone A. **b,** Investigation of the inhibitory effect of KK181N1 on the interaction of $SMXL1_{ND1M}$ with AtKAI2. Yeast cells were grown for 2 days in the liquid SC-LW media for 2 days at 30 ºC and 5 µL of each yeast culture was spotted on SC-LWHA plates containing 10 µM DMSO or KL agonists without/with 50 µM or 100 µM KK181N1. n = 2.

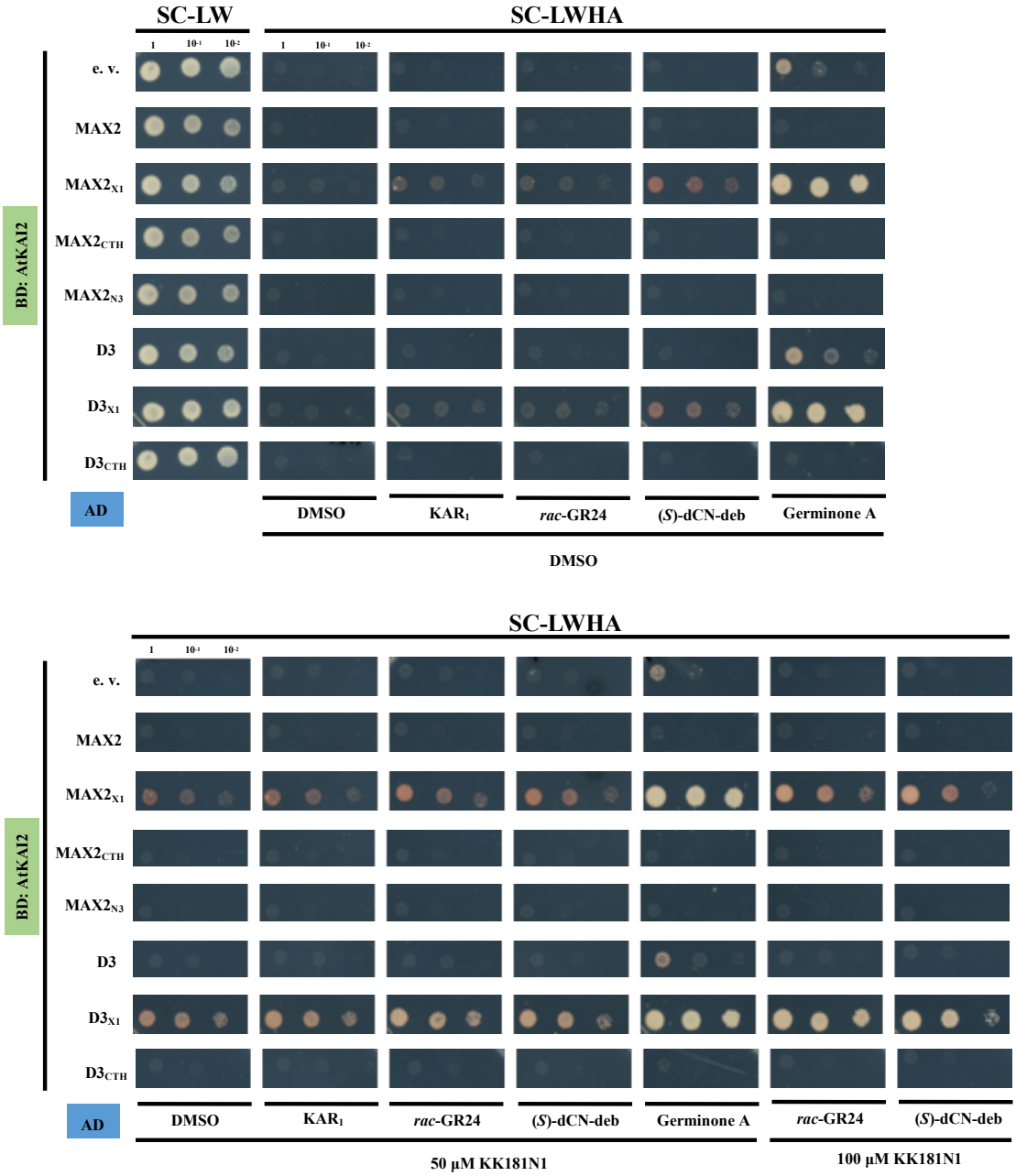
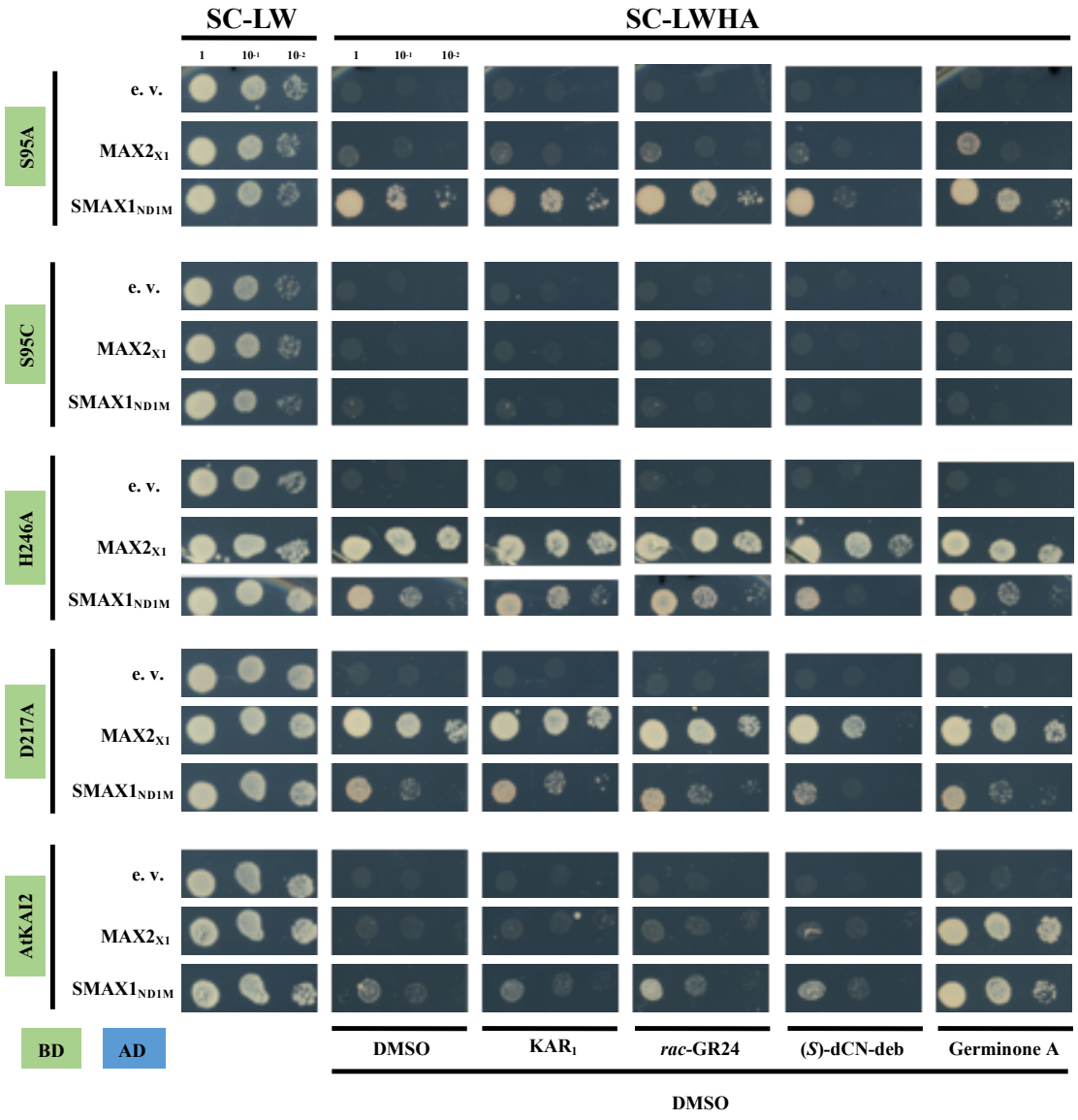


Figure 3-17: Preliminary investigation of the mode of action of KK181N1 involved in AtKAI2-MAX2/D3 interaction by Y2H assay.

Investigation of the inhibitory effect of KK181N1 on the interaction of various constructs of MAX2/D3 proteins with AtKAI2. Yeast cells were grown for 2 days in the liquid SC-LW media for 2 days at 30 ºC and 5 µL of each yeast culture was spotted on SC-LWHA plates containing 10 µM DMSO or KL agonists (KAR_1, *rac*-GR24, (*S*)-dCN-deb, and Germinone A) with/without 50 µM or 100 µM KK181N1. n = 2.


SC-LW
SC-LWHA
1
10-1
10-2
S95A
S95C
H246A
D217A
AtKAI2
e. v.
MAX2X1
SMAX1ND1M
BD
AD
DMSO
KAR1
rac-GR24
(S)-dCN-deb
Germinone A
DMSO

(Continued)

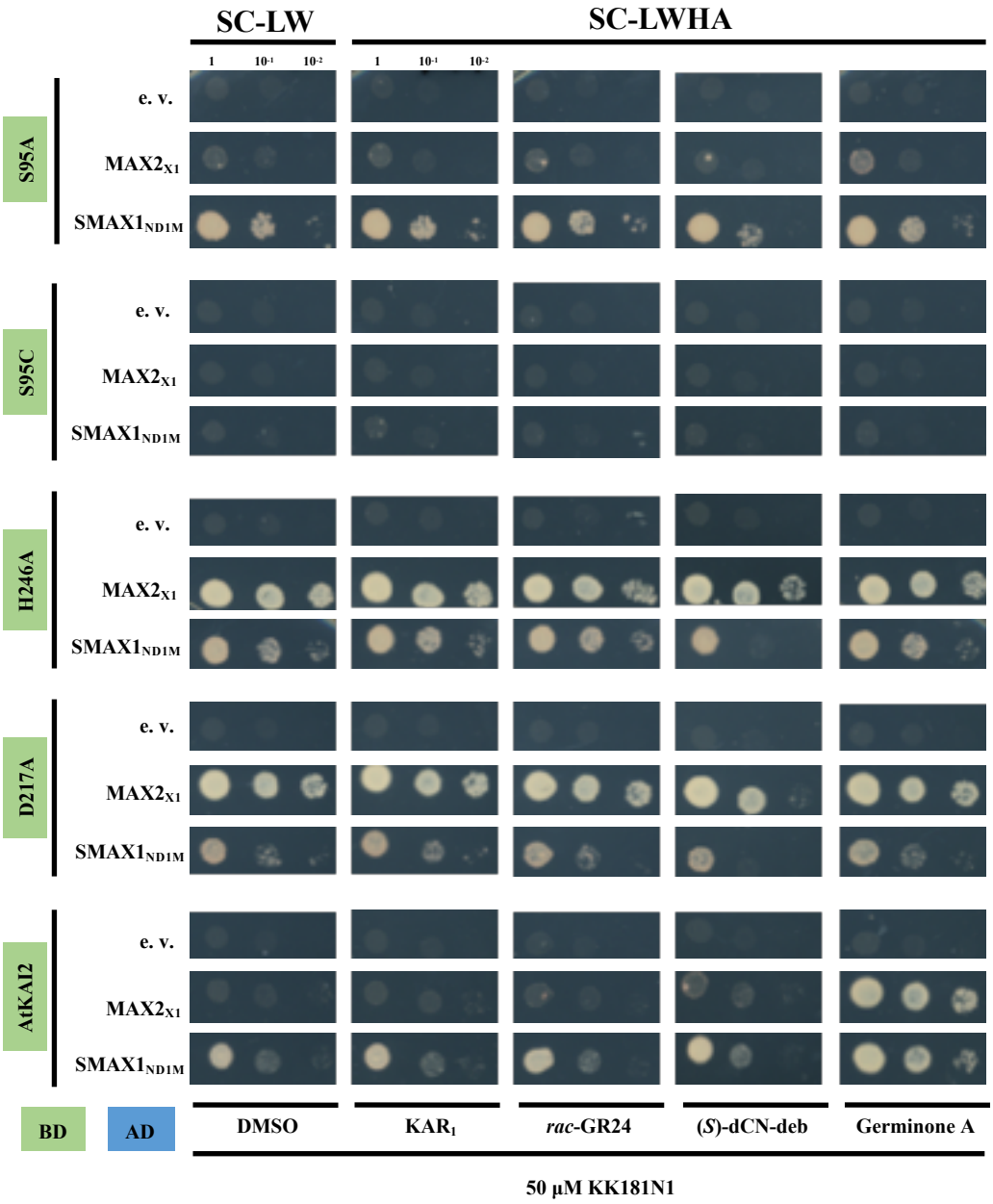


Figure 3-18: Interaction analysis of AtKAI2 variants with $MAX2_{X1}$ and $SMAX1_{ND1M}$ by Y2H assay.

Yeast cells were grown for 2 days in the liquid SC-LW media for 2 days at 30 ºC and 5 µL of each yeast culture was spotted on SC-LWHA plates containing 10 µM DMSO or KL agonists [KAR_1, *rac*-GR24, (*S*)-dCN-deb, and Germinone A] with/without 50 µM or 100 µM KK181N1. n = 1.

3.3 Conclusion and perspectives

In recent years, using chemical tools to modulate target protein functions has emerged as a powerful strategy to study hormone biology in plant science[151]. This approach is advanced tremendously by the tight collaboration between chemical biology and structural biology and has made extraordinary achievements for our understanding of the molecular mechanisms governing phytohormone regulation. KK181N1 is the N-heterocyclic ureas and previously had been revealed to antagonize the karrikin response by recovering the hypocotyl growth in *Arabidopsis thaliana*. In this study, I have clearly uncovered the binding mode and perception mechanism of KK181N1 by AtKAI2 through solving the crystal structure of the AtKAI2-KK181N1 complex. Moreover, by superposing the structure complex of AtKAI2-KK181N1 with OsD14-KK094, we unraveled the structural basis underlying KK181N1 discriminates response of KAI2 from that of D14.

KK181N1 and KK094 (both N1 and N2) structurally differ in one methyl group and are respectively antagonists of AtKAI2 and OsD14. The addition of methyl group in KK181N1 confer the selectivity for AtKAI2 but not OsD14. Previously, it has been revealed that *Arabidopsis thaliana* favors the desmethyl compound KAR_2 but not KAR_1[87]. Recent studies also indicated that desmethyl butenolides (desmethyl-GR24 and desmethyl-CN-debranone) are optimal ligands for AtKAI2 but not AtD14[217]. However, as I compared the structure complex of AtKAI2-KK181N1 and AtKAI2-KAR_1, it is clear to see that the binding poses of the methyl group of KK181N1 and KAR_1 are very different (Figure 3-7). Moreover, I find that this methyl group of KK181N1 is well accommodated by the hydrophobic interaction in AtKAI2 but causing the severe steric interfere in OsD14 (Figure 3-11). Although these two types of methyl group between KK compounds and butenolides may not have closed relationship for discriminating the perception by KAI2 and D14, it possibly suggests that the differences in active site of KAI2 and its paralog D14 are very useful to design the specific chemical probes without causing the cross-reactivity. Therefore, considering the moderate biological activity of KK181N1 and aided by the structural information of AtKAI2-KK181N1, we are dedicated to refining the structure of KK181N1 to produce a more robust inhibitor with higher sensitivity and selectivity in the future.

On the other hand, previously, it is believed that the hydrolyzed product (a covalently linked intermediate molecule, CLIM) of strigolactone initiates an open-to-closed state transition of D14 so

that trigger the interaction with D3[121], where the closed state is actually extending the interaction interface with D3 by narrowing down the pocket size from 420 $Å^3$ to 80 $Å^3$. Therefore, it can image that KK094CM with the bulky size in the pocket of D14 functions as an irreversible covalent inhibitor by not only preventing the entry of SLs agonist but also refusing the essential structural collapse. However, dissimilar to KK094, KK181N1, which would be the first inhibitor of AtKAI2 to our knowledge, stands for the applicability for the design of non-covalently bound chemicals as the antagonist targeting AtKAI2 proteins. This non-covalent inhibitor may mostly rely on the strong binding affinity with target proteins.

To summarize, this work uncovers the structural basis of the perception mechanism of a karrikin antagonist KK181N1 by AtKAI2. And it has great potentiality to be used as a chemical tool to probe SL biology without the interfering caused by AtKAI2.

Table 3-1: X-ray data collection and refinement statistics for AtKAI2-KK181N1 complex.

Data collection statistics	**AtKAI2-KK181N1 Complex**
Space group	*P 1 2 1 1*
Cell parameters	
a,b, c (Å)	85.31, 64.04, 114.96
α, β, γ (°)	90, 104.79, 90
Resolution (Å)	(43.54-1.90) (1.93-1.90)
No. of observation	326477 (16846)
No. of unique reflections	94190 (4662)
R_{merge} (%)	10.8 (79.7)
R_{means} (%)	12.8 (93.6)
R_{pim} (%)	6.8 (48.8)
$CC_{1/2}$ (%)	0.993 (0.496)
Average $I/\sigma I$	7.8 (2.9)
Completeness (%)	99.4 (99.9)
Multiplicity	3.5 (3.6)
Refinement statistics	
Resolution (Å)	1.90
No. of reflections used	94153
R work (%)	23.3
R free (%)	26.5
No. of atoms	
Protein	6231
Ligand	87
Water	732
Average B-value (Å2)	
Protein	32.0
Ligand	39.9
Water	35.6
R. m. s. deviatations	
Bond distance (Å)	0.017
Bond angle (°)	2.182
Ramachandran plot (%)	
Favored region	96.23
Allowed region	3.52
Outliers	0.25

$R_{merge} = \Sigma_h\Sigma_i|I_{h,i} - \langle I_h\rangle|/\Sigma_h\Sigma_i\langle I_{h,i}\rangle$, where $\langle I_h\rangle$ is the mean intensity of the $I_{h,i}$ observations of symmetry related reflections of *h*.

$R_{work} = \Sigma(||F_p(obs)| - |F_p(calc)||)/\Sigma|F_p(obs)|$; R_{free} is an *R* factor for a pre-selected subset (5%) of reflections that was not included in refinement. F_p, structure factor of protein.

Values in parentheses are for the highest resolution shell.

Table 3-2: The KK181N1-binding affinity of AtKAI2 mutations.

The binding affinity, measured by ITC assays, is indicated by the dissociation constants (K_d). Values are means ± SD (n = >3). The relative affinity is calculated for the comparison between each other.

Code	Variants	Roles/Location	Dissociation constants K_d (μM)	Relative affinity
WT	Wild-type	-	27.7 ± 0.7	100%
Var1	S95A	Catalytic residues	2210 ± 331	1.3%
Var2	S95C		28.6 ± 6.4	97.0%
Var3	S95D		42.4 ± 20.0	65.3%
Var4	D217A		n. d.	n. d.
Var5	D217E		1170 ± 67.4	2.4%
Var6	H246A		90.8 ± 32.4	30.5%
Var7	H246F		99.9 ± 54.6	27.8%
Var8	F26A	At the bottom or middle region of pocket	193 ± 51.8	14.3%
Var9	Y124A		76.7 ± 32.6	36.1%
Var10	Y124F		n. d.	n. d.
Var11	F157A		323 ± 90.7	8.6%
Var12	F174A		n. d.	n. d.
Var13	I193A		38.9 ± 8.6	71.2%
Var14	F194A		442 ± 160	6.3%
Var15	A219S		14.1 ± 4.2	197%
Var16	F134A	Lining the entrance of pocket	10.9 ± 2.3	254%
Var17	L142A		10.4 ± 1.2	266%
Var18	L218A		10.9 ± 1.6	255%
Var19	L218V		22.8 ± 6.8	121%
Var20	L218I		26.3 ± 4.7	105%
Var21	L218F		n. d.	n. d.
Var22	L218Y		n. d.	n. d.

n. d.: Not detected.

Table 3-3: The binding affinity of various compounds to AtKAI2 wild type.
The binding affinity, measured by ITC assays, is indicated by the dissociation constants (K_d). Values are means ± SD (n = >3). The relative affinity is calculated for the comparison between each other.

Compounds	Dissociation constants K_d (μM)	Relative affinity
KK181N1	27.7 ± 0.7	100.0%
KK181N2	209.8 ± 39.9	13.2%
KK094N1	199.9 ± 10.7	13.9%
KK094N2	316 ± 76.9	8.8%
KK167N1	711 ± 45.5	3.9%
KK175N1	179 ± 17.8	15.5%
KK189N1	164 ± 24.4	16.9%
KK118	n. d.	n. d.
KK122	106 ± 11.9	26.1%
KK199	n. d.	n. d.
KK2028	n. d.	n. d.
KK200	240 ± 14.0	11.5%

n. d.: Not detected.

Table 3-4: The oligonucleotides used to make AtKAI2 variants in this study.
The mutated bases are colored in red bold.

Primer Name	Variant code	Oligo Sequence (5' → 3')
KAI2_S95A_F	Var1	**G**CTGTTTCTGCCATGATTGGTGTC
KAI2_S95A_R		GTGGCCAACAAAGATACAAGACTCAATC
KAI2_S95C_F	Var2	T**G**TGTTTCTGCCATGATTGGTGTC
KAI2_S95C_R		KAI2_S95A_R
KAI2_S95D_F	Var3	**GA**TGTTTCTGCCATGATTGGTGTCT
KAI2_S95D_R		KAI2_S95A_R
KAI2_D217A_F	Var4	G**C**CTTAGCTGTACCAGTCGTTGTCTCC
KAI2_D217A_R		CTTAACACTTTGGAGAATGTGACACG
KAI2_D217E_F	Var5	GA**G**TTAGCTGTACCAGTCGTTGTCTCC
KAI2_D217E_R		KAI2_D217A_R
KAI2_H246A_F	Var6	**GC**TCTTCCTCAGCTTAGCTCACCAG
KAI2_H246A_R		ACCATCAGAAGGAATAACCTCGAC
KAI2_H246F_F	Var7	**TT**TCTTCCTCAGCTTAGCTCACCAG
KAI2_H246F_R		KAI2_H246A_R
KAI2_F26A_F	Var8	**GC**CGGCACGGACCAGTCAGTATGG
KAI2_F26A_R		CCCGTGACCTAACACGATCGTGG
KAI2_Y124A_F	Var9	**GC**CGTAAACGATGTTGATTACCAAGGT
KAI2_Y124A_R		TCTCGGAGAAGCAGAGATCATG
KAI2_Y124F_F	Var10	T**T**CGTAAACGATGTTGATTACCAAGGT
KAI2_Y124F_R		KAI2_Y124A_R
KAI2_F157A_F	Var11	**GC**CGCTCCACTCGCCGTCGGTGG
KAI2_F157A_R		ACCTAAGCACCACGCTTTGTAGTTGCTTCGG
KAI2_F174A_F	Var12	**GC**CAGCAGAACACTCTTCAATATGCGT
KAI2_F174A_R		TTCTTGAACGGCGATGGAGTC
KAI2_I193A_F	Var13	**GC**TTTCCAAAGTGACATGAGACAGATCTTACCTT
KAI2_I193A_R		GGTCTGGCCGACGGAGAGAGC
KAI2_F194A_F	Var14	**GC**CCAAAGTGACATGAGACAGATCTTACC
KAI2_F194A_R		AATGGTCTGGCCGACGGAG
KAI2_A219S_F	Var15	**AG**TGTACCAGTCGTTGTCTCCGAGTAT
KAI2_A219S_R		TAAGTCCTTAACACTTTGGAGAATGTGAC

(Continued)

Primer Name	Variant code	Oligo Sequence (5' → 3')
KAI2_F134A_F	Var16	GCCGAACAAGAAGACTTAAACCAACTATTCG
KAI2_F134A_R		TCCACCTTGGTAATCAACATCGTTTAC
KAI2_L142A_F	Var17	GCATTCGAAGCCATCCGAAGCAACTAC
KAI2_L142A_R		TTGGTTTAAGTCTTCTTGTTCGAATCCAC
KAI2_L218A_F	Var18	GCAGCTGTACCAGTCGTTGTCTCCG
KAI2_L218A_R		GTCCTTAACACTTTGGAGAATGTGACAC
KAI2_L218V_F	Var19	GTAGCTGTACCAGTCGTTGTCTCCG
KAI2_L218V_R		KAI2_L218A_R
KAI2_L218I_F	Var20	ATAGCTGTACCAGTCGTTGTCTCCG
KAI2_L218I_R		KAI2_L218A_R
KAI2_L218F_F	Var21	TTCGCTGTACCAGTCGTTGTCTCC
KAI2_L218F_R		KAI2_L218A_R
KAI2_L218Y_F	Var22	TATGCTGTACCAGTCGTTGTCTCC
KAI2_L218Y_R		KAI2_L218A_R

Table 3-5: The oligonucleotides used to make various constructs for Y2H assay in this study. The constructs were made by three methods as indicated in the column "Constructs Method". Firstly, the target gene were inserted between two restriction sites of vector. The restriction enzymes used are indicated in the table, and the recognition sequences are labeled red with underline. Secondly, the target gene were inserted after one restriction site of vector (*Bam*HI in this study) by using In-Fusion HD Cloning Kit (TAKARA). The sequences with black underline indicate the 15 bp extensions homologous to vector ends. The oligonucleotides of start/stop are highlighted with blue bold. Primers are designed by In-Fusion Cloning Primer Design Tool. Thirdly, the target constructs are made by reverse PCR using KOD-mutagenesis method (TOYOBO).

Primer Name	Sequence range	Oligo Sequence (5' → 3')	Constructs Method
AtKAI2-F	Full Length: 1-270	GGAATTCCATATG**ATG**GGTGTGGTAGAAGAAGCTCACAAC	Two restriction sites: *Nde* I…*Bam* HI
AtKAI2-R		CTAGGGATCC**TCA**CATAGCAATGTCATTACGAATGTGACGG	
SMAX1-FL_F	Full Length: 1-990	GGAATTCCATATGAGAGCTGGTTTAAGTACGATTCAACAGACT	Two restriction sites: *Nde* I…*Cla* I
SMAX-FL_R		CCATCGAT**TCA**TACTGCCAAAGTAATAGTTGTCGGCAGTA	
SMAX1-N_F	N: 1-200	CATCGATACGGGATC**ATG**AGAGCTGGTTTAAGTACGATTCAACAGAC	In-fusion: after *Bam* HI Site
SMAX-N_R		CGAGCTCGATGGATC**TTA**TGGATTAAGATACGAATTCCGGGTCATCG	
SMAX1-ND1_F	ND1: 1-405	SMAX1-N_F	In-fusion: after *Bam* HI Site
SMAX-ND1_R		CGAGCTCGATGGATC**TTA**CAATGGCGTGAATGATTCCAAATTGTTCG	
SMAX1-D1M_F	D1M: 201-591	CATCGATACGGGATCCGGTTACAACAGAACGCTTCGTCG	In-fusion: after *Bam* HI Site
SMAX-D1M_R		CGAGCTCGATGGATC**TTA**GTTATTATTGTTCTGCACTGATTCAGACGATATGC	
SMAX1-ND1M_F	ND1M: 1-591	SMAX1-N_F	In-fusion: after *Bam* HI Site
SMAX-ND1M_R		SMAX-D1M_R	
SMAX1-D1MD2_F	D1MD2: 201-990	SMAX1-D1M_F	In-fusion: after *Bam* HI Site
SMAX-D1MD2_R		CGAGCTCGATGGATC**TTA**TCATACTGCCAAAGTAATAGTTGTCGGCAG	
MAX2-FL_F	Full Length: 1-693	GGAATTCCATATGGCTTCCACTACTCTCTCCGAC	Two restriction sites: *Nde* I…*Bam* HI
MAX2-FL_R		CTAGGGATCCTCAGTCAATGATGTTGCGGCTGT	
MAX2-1_F	X1: 560-689	CGCGGATCC**AT**AAGATGCAGTTAGATTGCGGGGACAC	Two restriction sites: *Bam* HI…*Sac* I
MAX2-1_R		TGACGAGCTC**TTA**GCGGCTGTTCAATTGGTCCTCG	
MAX2-2-CTH_F	CTH: 667-693	GAGAACGATATGAGCACAGAGATGAGAGTTGGT	KOD Mutagenesis
MAX2-2-CTH_R		CTGGTACCCGGGTCCCTGAAAGAG	
MAX2-3-N3_F	N3: 255-455	CATCGATACGGGATCATGGTGGACACAGCTTCGTTGG	In-fusion: after *Bam* HI Site
MAX2-3-N3_R		CGAGCTCGATGGATC**CTA**AGCCTCACTAGTTTCCACTCTTCC	
D3-FL_F	Full Length: 1-720	CATCGATACGGGATCATGGCGGAAGAGGAGGAGGTG	In-fusion: after *Bam* HI Site
D3-FL_R		CGAGCTCGATGGATC**CTA**ATCATCAATTTGCCGGCTGTTCAG	
D3-1_F	X1: 585-716	CGCGGATCC**ATA**AGATGAAATTGGACCTCAGTGAAGCTGTG	Two restriction sites: *Bam* HI…*Sac* I
D3-1_R		TGACGAGCTC**TTA**CCGGCTGTTCAGTTGCACCTCAAAC	
D3-2-CTH_F	CTH: 693-720	GAGAATGATCTGATGTTCACAGAGATGCGG	KOD Mutagenesis
D3-2-CTH_R		CTGGTACCCGGGTCCCTGAAAG	

3.4 Materials and methods

3.4.1 Chemicals

All KK compounds were provided by Mr. Ko KIKUZATO, University of Tokyo. All karrikin agonists [KAR_1, *rac*-GR24, (*S*)-dCN-deb, and Germinone A] were provided by Dr. Ikuo TAKAHASHI, University of Tokyo

3.4.2 Molecular cloning and site-directed mutagenesis.

The gene encoding full length *Arabidopsis thaliana* AtKAI2 was provided by Prof. Takuya MIYAKAWA, University of Tokyo. Full length *Arabidopsis thaliana* AtMAX2, AtSMAX1 and *Oryza sativa* OsD3 were provided by Dr. Ikuo TAKAHASHI, University of Tokyo.

For protein expression, full length of AtKAI2 (1-270) was amplified by PCR and cloned into the expression vector pET-48b (+) (Novagen) between the restriction enzyme sites of *Bam*HI and *Hin*dIII with Trx·6×His tag and HRV 3C cleavage site at N-terminus, named as pET48-AtKAI2. The reverse PCR-based site-directed mutagenesis was employed to prepare the all variants of AtKAI2 using KOD-Plus-Mutagenesis Kit (TOYOBO, Japan), according to manual. All primers used to make AtKAI2 variants were summarized in Table 3-4. Mutagenesis was confirmed by systematic DNA sequencing.

For Y2H assay, full length of AtKAI2 (1-270) were inserted into pGBKT7 (Addgene) between the restriction enzyme sites of *Nde*I and *Bam*HI; Full length of AtSMAX1 (1-990) was cloned into pGADT7 (Addgene) between the restriction enzyme sites of *Nde*I and *Cla*I; Full length of AtMAX2 (1-693) was cloned into pGADT7 between the restriction enzyme sites of *Nde*I and *Bam*HI; Full length of OsD3 (1-720) was cloned into pGADT7 after restriction enzyme sites *Bam*HI using the In-fusion method (Takara). The different truncations of AD proteins, including AtMAX2 truncations (X1, N3, and CTH), OsD3 truncations (X1 and CTH) and AtSMAX1 truncations (N, ND1, D1M, ND1M, and D1MD2) were amplified and inserted into pGADT7 using the In-fusion method (Takara). All primers used were summarized in Table 3-5, including the information of AD proteins used in this study. All sequences were confirmed by the DNA sequencing service (Eurofins Genomics Co., Ltd).

3.4.3 Protein preparation and purification

The recombinant proteins of AtKAI2 wild-type and variants were prepared in the same manner as described below. The recombinant vector pET48-AtKAI2 was transformed into competent cell *Escherichia coli* BL21 (DE3) (Novagen) for protein expression. 20 mL of overnight seeding cultures were inoculated to fresh 2 liters Luria-Bertani broth containing 25 μg/mL kanamycin at 37°C until OD_{600} reached 0.6-0.8. The protein overexpression was then induced by 0.1 mM Isopropyl-β-D-thiogalactoside (IPTG) at 18°C for 20 h. The cells were harvested and resuspended in sonication buffer (20mM Tris–HCl (pH 8.0), 300 mM NaCl, 10 mM imidazole and 1mM dithiothreitol (DTT)) complemented with 0.1% (vol/vol) protein inhibitor (Nacalai tesque) for cell lysis. After the cells were disrupted by sonication (Power = 59%, In = 1.0 s; Out = 1.0 s; Time = 10 min; Tpulse=4 °C; Tmax=4 °C), the ultracentrifugation (Avanti J-E rotor, Beckman Coulter) at 40,000 × *g* was performed at 4 °C for 30 min to remove cell debris. The supernatant fractions were collected and gently incubated with 2 mL of pre-equilibrated Ni-NTA Agarose (Qiagen, USA) at 4°C. After 30 min later, the supernatant-resin mixture was subjected to a disposable polypropylene column, and then washed by 100 mL of buffer sonication buffer. The on-column removal of N-terminal Trx-His×6 tag was performed overnight at 4°C by adding 40 μg HRV3C protease. Afterwards, certain amount of elution buffer [20mM Tris–HCl (pH 8.0), 500 mM NaCl, 5% glycerol, 10 mM imidazole and 1mM c (DTT)] was applied to column to elute untagged target proteins. A cation exchange chromatographic column Mono Q 10/100 GL (GE Healthcare) was used to further purify the elutes, and resultant fraction containing AtKAI2 was collected after a linear gradient elution of 0-1.0 M NaCl. SDS-PAGE was used to check the purity of proteins obtained, and protein concentration was adjusted as needed by nanodrop 2000 Spectrophotometers (Thermo Scientific). Protein samples can be stored at -80 °C if not use immediately.

3.4.4 Crystallization, data collection and structure determination

The protein-ligand solution for co-crystallization trials was made by incubating 0.7 mg/mL of purified AtKAI2 solution with 5 mM KK181N1 at room temperature for 0.5-1 hours, and then concentrated into 7 mg/mL. Crystals of the AtKAI2–KK181N1 complex were generated by the 1:1 mixture of protein-ligand solution and reservoir solution (1.7 M sodium phosphate (pH 7.0) and 1% 1,2-butanediol) at 20°C.

A single crystal was picked up and soaked in reservoir solution containing 30% sucrose as a cryoprotectant, and then flash-cooled with a nitrogen-gas stream at 100 K. All X-ray diffraction data were collected on the BL-1A beamline at the Photon Factory (Tsukuba, Japan) and processed using the XDS package[226]. Molecular replacement was performed using Phaser[227] in PHENIX[228] with the *apo*-KAI2 structure (PDB: 4JYP)[101] as the initial model. The building of modified electron density and fitting of ligand into the pocket was performed by COOT[229] and further refined through iterative refinement cycles in PHENIX. Data collection and refinement statistics are presented in Table 3-1. All structures were depicted by using PyMOL 2.4[230, 231]. Sequence alignment was performed using CLUSTAL W55 with default parameters, and the results were displayed via ESPript 3.0[232]. Phylogenetic tree was constructed by Phylogeny.fr based on the results of sequence alignment[233, 234].

3.4.5 ITC experiments

Isothermal titration calorimetry (ITC) experiments for the binding assays of AtKAI2 and its variants with compounds were performed with a MicroCal iTC200 (GE Healthcare). Prior to the measurements, the protein solution was dialyzed against ITC buffer (20 mM HEPES (pH 8.0) and 150 mM NaCl) for 20 h and adjust to 150 μM. The dialyzed buffer was used to prepare ligand solution containing 1% DMSO at 2-5 mM. The sample cell and syringe were respectively filled with 204 μL of protein solution and 40 μL of ligand solution. For a given experiment, nineteen consecutive 2.0-μL aliquots of ligand solution were injected into sample cells at 20 °C with a stirring speed of 750 rpm, reference power 5 μcal s^{-1} and spacing of 150 s between injections. A negative control was made by titrating ligand solution into the dialyzed buffer containing 1% DMSO. For data analysis, all peaks, among which the first one resulted from initial 0.4 μL injection was excluded, were integrated corresponding to each injection, and the correction for the baseline and the fitting were performed using the Origin 7.0 software in the "*one set of sites*" mode, to obtain the dissociation constant (K_d). Three replicates were made for each measurement.

3.4.6 Y2H assay

AH109 yeast cells were transformed with the constructs BD: pGBK-AtKAI2 and different AD (pGAD-MAX2, pGAD-D3 and pGAD-SMAX1) according to the standard lithium acetate-

mediated method [235]. For the plate assay, the transformed AH109 cells harboring both AD and BD recombinant plasmids were precultured in liquid synthetic complete medium lacking Leu and Trp (synthetic defined/–Leu/–Trp) at 30°C with shaking. After 1 days later, the cells were plated onto synthetic defined/–Leu/–Trp/–His/–Ade medium containing the test compound (KAR1, *rac*-GR24, (*S*)-dCN-deb, Germinone A or KK181N1) or 0.1% DMSO as mock control or containing His and Ade (+His/+Ade) as the positive control for yeast growth.

論文の内容の要旨

応用生命化学　　専攻
令和元年度博士課程　　入学
氏 名　　王 建文
指導教員名　　浅見 忠男

論文題目

Studies on the interaction of α/β-hydrolase-type plant hormone receptors with their regulators
(α/β-加水分解酵素型植物ホルモン受容体と制御剤の相互作用解析研究)

Chapter 1. Introduction

α/β-hydrolase (ABH) fold superfamily is ubiquitously distributed throughout all living organisms in three kingdoms of life. Sharing the canonical framework of the α-turn-β supersecondary geometry with high plasticity in substructure, ABH proteins have served as diverse roles to participate in various biological pathways, including biosynthesis, signal communication and transduction, and individual growth and development. Therefore, an increasing number of studies were paid on the drugs discovery for therapeutics and the chemical tools development for probing biological processes through modulation of ABH proteins activity. Plant hormones are endogenously signal and regulatory small molecules metabolically derived to regulate all aspects of plant lifestyle. Plants possess three plant hormone receptors belonging to the ABH superfamily, two of which are the receptor of gibberellin (GID1) and karrikin (KAI2). The chemical regulation of their function and signal is of great importance and benefits to boost our agricultural system, as well as basic research.

Chapter 2. Development of orthogonal AtGID1a variant-AC94377 pairs

Gibberellins (GAs) are a large family of tetracyclic diterpenoid plant hormones, and well acknowledged as the essential roles that regulate many developmental processes through the entire life cycle of the plant, including seed germination, cell division and elongation, leaf expansion, pollen maturation, root growth, flowering induction, fruit enlargement, etc. Previously, we have found that chemical AC94377 shows GA agonistic activity in Arabidopsis and rice. Despite that its mode of action has been clarified, the accurate binding mode of AC94377 with GID1s still remains

unknown. What' more, GA-activated GID1s proteins have engaged in regulating diverse aspects of plant growth and development, but delineation of the precise biological events of a specific GID1 is of great challenge, especially in Arabidopsis because the pleiotropic effects of various receptor-ligand-repressor combinations (AtGID1s-GAs/AC94377-DELLA) lead to the complicated regulatory networks that difficult to be artificially and specifically manipulated. With these in mind, I firstly made attempts to uncover the mode of interaction between GID1 and AC94377 at atomic level, with the expectation of developing more potent GA regulators based on the chemical structure of AC94377 skeleton. Secondly, I tried to design orthogonal ligand-receptor pairs between GID1-AC94377 for probing specific GA response **(Figure 1)**.

In this study, I firstly applied X-ray crystallography to get insight into the binding mode of AC94377 in GID1. However, even though we had paid substantial efforts on crystal screening and optimization of GID1-AC94377, I was unable to get quality-sufficient crystals for X-ray diffraction. Therefore, to understand the perception mechanism of AC94377 by GID1 protein, I performed the structure-activity relationship studies on AC94377 by using a series of its derivatives containing different substituent on chloro and carboxamide groups of AC94377. Combining the results of *in vivo* (Y2H and physiological assessment) and *in vitro* (ITF and DSF) assays, I realized that both carboxamide and chloro groups are important to retain the high GA-like activity for AC94377. Moreover, I found that AC94377 also favors the Ile-subtype OsGID1 (V319I), which is particularly consistent with our previous work. In order to support the results from SAR studies, I performed the molecular docking simulation of AC94377 with OsGID1 and AtGID1a. Superposition of simulated structures of both OsGID1-AC94377 and AtGID1a-AC94377 complexes shows good agreement in the binding pose of AC94377 in the pocket of GID1s. Notably, based on this simulation, I thought that the reinforced hydrophobic interaction between the aromatic plane of AC94377 and the extended alkyl group of isoleucine (compared with valine) may account for the improved affinity of AC94377 for the Ile-subtype GID1. However, it is difficult to fully explain the results using current simulating models because two AC94377 derivatives with the modifications of carboxamide group in AC94377 by changing amine into hydroxyl or isopropoxy (ACOH and ACOIp) did not show the preference to Ile-subtype GID1. Here, I proposed that the ligand-induced conformational changes may occur to accommodate the occupation of AC94377 in the pocket of GID1.

Since the absence of a precise structure complex of GID1-AC94377, I performed the AC94377-directed laboratory evolution on AtGID1a by Y2H-based screening system to generate the orthogonal ligand-protein variant pair of GID1-AC94377. To efficiently achieve this goal, I firstly adopted the combinatorial saturation mutagenesis on AtGID1a (CSM) by grouping the 24 pocket-lining residues into 9 sites/sublibraries. This has produced several beneficial AtGID1a variants from 3 sites with 10-folds higher AC94377-binding affinity than wild type. Next, I performed the iterative random mutagenesis (IRM) on these positive variants to maximize the cooperative effects of mutagenesis. Fortunately, these efforts have enabled to generate the AtGID1a variants with greatly enhanced affinity and selectivity to AC94377. Specifically, Y2H assay reveals that the strength of AC94377-induced PPI between these AtGID1a variants and GAI is

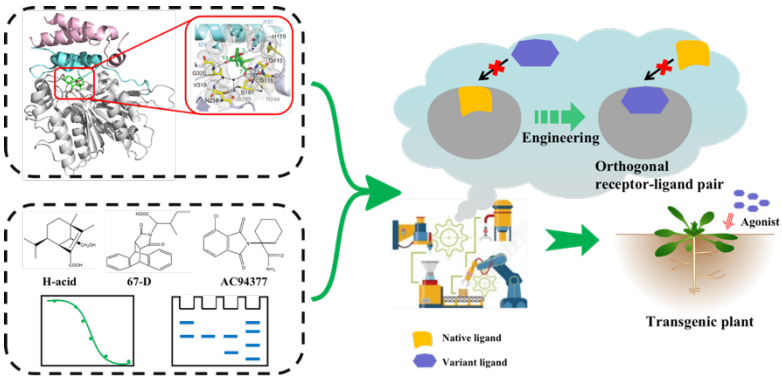


Figure 1. Generation of an orthogonal ligand-receptor pair pairs

increased about 16,000-folds than that between AtGID1a wild-type and GAI (at 50 μM AC94377), and the sensitivity of AtGID1a variants to AC94377 is increased 10-30 folds than that of AtGID1a wild-type to GA_4.

To explore the direct binding event between ligands and variant receptors, I performed the ITF and DSF assay to evaluate the binding affinity. Notably, some of the AtGID1a variants show greatly improved affinity to AC94377 with about 500-folds higher than wild type, whereas they totally lost the GA_4-binding affinity. The DSF assay also supports these results that AC94377 induces the enhancement of thermostability for AtGID1a variants in the same manner as GA_4 did for wild type. However, GA_4 cannot increase the structural stability of AtGID1a variants. Interestingly, even without the treatment of ligand, thermostability of AtGID1a variants is also 10-12 ºC higher than that of AtGID1a wild type, suggesting the more stable interactions between these mutated residues.

My works on the establishment of orthogonal AtGID1a variant-AC94377 pairs could be served as a powerful tool to delineate specific GID1-mediated responses through bypassing the complexity of genetic redundancies, further advancing our understanding of GA biology.

Chapter 3. Structural basis for the perception of a Karrikin antagonist KK181N1 by receptor AtKAI2

Karrikins (KARs) are a family of butenolides derived from smoke and ash of burning materials. KARs are identified as the seed germination stimulant in the early stage of plant development, and also enable to shape seedling photomorphogenesis. Through the recent works in our laboratory, a compound (named KK181N1) that shows antagonistic effect on karrikin induced suppression of hypocotyl growth in Arabidopsis has been screened. This chemical binds directly and alleviate the hydrolysis activity of karrikin receptor AtKAI2. However, the precise binding mode of KK181N1 in AtKAI2 still remains unknow. In this study, the structure of AtKAI2-KK181N1 complex was solved at 1.90 Å resolution with a clear electron-density map of KK181N1. KK181N1 belongs to N-heterocyclic ureas and designed to covalently modify α/β hydrolase fold-containing proteins through the catalytic nucleophile. However, unexpectedly, I found that this binding was caused by noncovalent interaction and KK181N1 was firmly embedded in the pocket of AtKAI2. This binding mode of KK181N1 is totally different from that of OsD14-KK094, although OsD14 is highly homologous to AtKAI2 and KK094 only differs from KK181N1 in one methyl group. To thoroughly understand the structural basis of KK181N1's binding mode in AtKAI2, we divide the binding pocket into three regions: the bottom, the middle region, and the entrance of the pocket.

At the bottom of the pocket, I found a four water molecules-mediated hydrogen-bonding network connecting the 1,2,3-triazole moiety of KK181N1 to the residues surrounded, including the catalytic nucleophile S95, catalytic base H247, Y124, S119, and the backbone of F26 and V96. Our structure and activity relationship studies have indicated that this hydrogen-bonding network is of great importance for KK181N1-binding affinity in AtKAI2. It is notable that the variants S95C and S95D respectively retain 97% and 65% of wild type's affinity to this chemical, suggesting the broad-spectrum applicability of our in-house chemical library for α/β hydrolase fold superfamily proteins. Intriguingly, the structure comparison between OsD14-KK094 and AtKAI2-KK181N1 complexes has shown that, even though these two chemicals have different binding mode in the pocket, they all participate in this water molecules-mediated hydrogen-bonding network with greatly similar topology, suggesting this hydrogen-bonding network, which seems to be conserved among KAI2/D14 family proteins, would be favored for the binding of 1,2,3-triazole compounds.

At the middle region of the pocket, we found extensive hydrophobic interactions between the

aromatic plane of KK181N1 with polar residues shaping the active pocket. Alanine scanning experiment reveals that the hydrophobic interactions of these residues are essential for the KK181N1-binding in the pocket. It's interesting to notice that a variant A219S shows two-folds higher activity than wild type for the binding to KK181N1. Previously, A219V was revealed to narrow the entrance gate for the ligand entry. However, the structure comparison between AtKAI2-KK181N1 complex and *apo*-OsD14 (corresponding residue: S270) indicates that the side chain of serine at this site would not cause the steric clash with ligands, including KAR_1 and KK181N1. Moreover, I compared the structures of AtKAI2-KAR_1 and *apo*-AtKAI2 with that of AtKAI2-KK181N1 complex and found that F134, F157, L142, F194 and L218 have undergone the conformational changes for the perception of KK181N1.

At the entrance of pocket, I notice that three residues, including F134, L142 and L218, account for the recognition of ligand KK181N1. It is interesting to find that each of the alanine substituent at these positions all possesses increased affinity to KK181N1, which may suggest that the releasing of steric effect at the entrance will promote the recognition of KK181N1. This is particularly verifying my structural information where all of the residues (F134, L142 and L218) occur conformational shift to accommodate the KK181N1.

To study the ligand selectivity, I place my focus on the three residues (F157, L218, and A219), which interact with the methyl group of KK181N1 by hydrophobic interaction but are different from the residues at the corresponding location in OsD14. The structural comparison and affinity assay suggested that although S270 (A219 in AtKAI2) and V269 (L218 in AtKAI2) in OsD14 may elevate the binding affinity of OsD14 to KK181N1, the Y209 (F157 in AtKAI2) would cause sever steric clash with carbonyl and the methyl group of KK181N1. This may partially explain the binding preference of KK181N1 to AtKAI2 but not OsD14 and is consistent with our physiological result that KK181N1 exhibit the inhibitory effect on hypocotyl growth through karrikin signaling pathway, but not SL. On the other hand, to further confirm the significance of this methyl group for ligand recognition by AtKAI2, we prepared an array of derivatives with different substitution at this group, including hydrogen, nitro, chloro, and fluoro group. However, all of them shows decreased affinity to AtKAI2, suggesting that methyl group is most favored by this hydrophobic environment.

The chemical synthesis of KK compounds produces two forms of isomers. In one compound, carbonyl group is connected to nitrogen atom at 1-position of 1,2,3- triazole moiety (annotated as

N1), whereas in another carbonyl group is connected to nitrogen atom at 2-positon of 1,2,3-triazole moiety (annotated as N2). Since I have found the great contribution of hydrogen-bonding network to anchoring the KK181N1 in the deep site of pocket, I am wondering whether this network also contributes to the specificity of KK compounds in AtKAI2. The affinity comparison between KK181N1 and KK181N2 or KK094N1 and KK094N2 suggested that AtKAI2 only shows higher affinity toN1-type ureas, which is consistent with the structural details with respect to the topology of hydrogen-bonding network where N1-type 1,2,3-triazole ureas form three hydrogen bonds to join but N2-type 1,2,3-triazole ureas only have one. This result was further examined by applying the chemical KK122 for affinity assay. Even in the absence of one nitrogen atom (indicated by N5) forming one hydrogen bond, the presence of another nitrogen (indicated by N6) harboring two hydrogen bonds could reinforce the KK122-binding affinity to about 3 folds higher than KK094N2, or 2 folds than KK181N2, stating that N6 atom-formed hydrogen bonds are essential for the binding affinity of 1,2,3-triazole-containing chemicals toward AtKAI2. What's more, by using other N-heterocyclic ureas with divergent moieties, I also found that N1-type1,2,3-triazole is greatly favored by AtKAI2 proteins, which is contributed by the unique hydrogen-bonding network. Taken together, the structural analysis and SAR studies suggested that the water molecules-mediated hydrogen-bonding network not only promotes the tight binding of KK181N1, but also confers the ligand selectivity of AtKAI2 for different isomers, and the residues involved in the recognition of the methyl group of KK181N1 greatly account for the selectivity of KK181N1 to AtKAI2 but not D14.

The study of uncovering the interaction mechanism of AtKAI2-KK181N1 will not only enlighten us to perform structure-aided optimization of KK181N1 to generate a more robust inhibitor but also encourage us to apply this antagonist in plants for basic research use, such as for the exploration of the native ligand of KAI2 and the solo establishment of SL function that bypass KAI2 signaling.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Tadao ASAMI for his invaluable guidance, continuous support, and kind patience in my Ph.D. study and research. I want to thank him for accepting me as the ph.D. student to join the research project three years ago. His dynamism, vision, sincerity, and wisdom have deeply encouraged me in all the time of my academic research and daily life. Working under his guidance is the best start to open and broaden my academic career. Besides my supervisor, I would like to thank the rest of my thesis committee: Prof. Seiji TAKAYAMA, Prof. Koji NAGATA, Dr. Masatoshi Nakajima, and Dr. Takuya MIYAKAWA, for their insightful comments and critical reviews.

I would like to give special and heartfelt thanks to Dr. Kohji Murase sensei and Dr. Takuya MIYAKAWA sensei for their guidance on the research of gibberellin and karrikin, respectively. Without their immense knowledge and plentiful experience, this research project wouldn't be possible. I also want to give the warmest thanks to Prof. Dr. Seiji TAKAYAMA. The stay in his laboratory (the bioorganic chemistry laboratory) is comfortable and fruitful for my research and study in the past three years. I would like to extend my thanks to my lab mates, colleagues, and research team for a cherished time spent together in the lab of chemical biology. It was a great privilege and honor to work and study with them. I would also thank Dr. Kai JIANG, Dr. Ikuo TAKAHASHI and Dr. Zhangliang ZHU. The stimulating discussions with them always inspire me a lot. I truly appreciate all assistance offered by our lab secretary Ms. Kyoko SHIMAZU. Thanks for her patience and warm heart in daily life.

I want to express the depth of my gratitude to all my friends who help me during my stay in Japan. I am extremely grateful to Ms. Run WEI. It's her encouragement and unwavering support that make my Ph.D. study possible from the beginning to the end. Thanks for all you have done for me. My gratitude knows no bounds. I truly wish you have brightness and happiness in every moment of day ever. I would also like to thank Dr. Kun CHEN for his friendship and empathy. Thanks for his time to play video games with me. My sincere thanks also go to Ms. Juan TAN for

her companionship in each sleepless night of the past times. Her words always strengthen my faith and carry me on the roads of hope. I am also truly indebted to the financial support from China Scholarship Council (CSC) and SPRING GX program of UTokyo for my study in Japan.

Finally, I greatly appreciate all the support I received from my family. Big thanks to my mom for her love, prayers, caring, and countless sacrifices that she made throughout my life. Thanks and loves to my sisters and brother for supporting me spiritually as always. Praises and thanks to the God for His blessings and provision on our family.

Sincere appreciation to all of you mentioned above.

Reference

1. Ollis, D.L. et al. The alpha/beta hydrolase fold. *Protein Eng* **5**, 197-211 (1992).
2. Carr, P.D. & Ollis, D.L. Alpha/beta hydrolase fold: an update. *Protein Pept Lett* **16**, 1137-1148 (2009).
3. Lenfant, N., Hotelier, T., Bourne, Y., Marchot, P. & Chatonnet, A. Proteins with an alpha/beta hydrolase fold: Relationships between subfamilies in an ever-growing superfamily. *Chem-Biol Interact* **203**, 266-268 (2013).
4. Lord, C.C., Thomas, G. & Brown, J.M. Mammalian alpha beta hydrolase domain (ABHD) proteins: Lipid metabolizing enzymes at the interface of cell signaling and energy metabolism. *Bba-Mol Cell Biol L* **1831**, 792-802 (2013).
5. Holmquist, M. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr Protein Pept Sci* **1**, 209-235 (2000).
6. Nardini, M. & Dijkstra, B.W. Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* **9**, 732-737 (1999).
7. Cambillau, C., Longhi, S., Nicolas, A. & Martinez, C. Acyl glycerol hydrolases: Inhibitors, interface and catalysis. *Curr Opin Struc Biol* **6**, 449-455 (1996).
8. Dimitriou, P.S., Denesyuk, A.I., Nakayama, T., Johnson, M.S. & Denessiouk, K. Distinctive structural motifs co-ordinate the catalytic nucleophile and the residues of the oxyanion hole in the alpha/beta-hydrolase fold enzymes. *Protein Science* **28**, 344-364 (2019).
9. Schrag, J.D. & Cygler, M. Lipases and alpha/beta hydrolase fold. *Methods Enzymol* **284**, 85-107 (1997).
10. Hubbard, T.J.P., Murzin, A.G., Brenner, S.E. & Chothia, C. SCOP: A structural classification of proteins database. *Nucleic Acids Research* **25**, 236-239 (1997).
11. Orengo, C.A. et al. CATH - a hierarchic classification of protein domain structures. *Structure* **5**, 1093-1108 (1997).
12. Chatonnet, A., Marchot, P., Bourne, Y., Lenfant, N. & Hotelier, T. ESTHER database: update on the alpha/beta hydrolase fold superfamily of proteins. *J Neurochem* **142**, 209-209 (2017).
13. Lenfant, N. et al. ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Research* **41**, D423-D429 (2013).
14. Dimitriou, P.S. et al. Alpha/beta-hydrolases: A unique structural motif coordinates catalytic acid residue in 40 protein fold families. *Proteins* **85**, 1845-1855 (2017).
15. Denesyuk, A., Dimitriou, P.S., Johnson, M.S., Nakayama, T. & Denessiouk, K. The acid-base-nucleophile catalytic triad in ABH-fold enzymes is coordinated by a set of structural elements. *PLoS One* **15**, e0229376 (2020).
16. Pareek, C.S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *J Appl Genet* **52**, 413-435 (2011).
17. Shannon, D.A. & Weerapana, E. Covalent protein modification: the current landscape of residue-specific electrophiles. *Curr Opin Chem Biol* **24**, 18-26 (2015).
18. Simon, G.M. & Cravatt, B.F. Activity-based Proteomics of Enzyme Superfamilies: Serine Hydrolases as a Case Study. *J Biol Chem* **285**, 11051-11055 (2010).

19. Rawlings, N.D., Barrett, A.J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Research* **38**, D227-D233 (2010).
20. Bononi, G., Tuccinardi, T., Rizzolio, F. & Granchi, C. alpha/beta-Hydrolase Domain (ABHD) Inhibitors as New Potential Therapeutic Options against Lipid-Related Diseases. *J Med Chem* **64**, 9759-9785 (2021).
21. Baggelaar, M.P. et al. ABHD2 Inhibitor Identified by Activity-Based Protein Profiling Reduces Acrosome Reaction. *Acs Chem Biol* **14**, 2943 (2019).
22. Thomas, G., Brown, A.L. & Brown, J.M. In vivo metabolite profiling as a means to identify uncharacterized lipase function: Recent success stories within the alpha beta hydrolase domain, (ABHD) enzyme family. *Bba-Mol Cell Biol L* **1841**, 1097-1101 (2014).
23. Parkkari, T. et al. Discovery of Triterpenoids as Reversible Inhibitors of alpha/beta-hydrolase Domain Containing 12 (ABHD12). *Plos One* **9** (2014).
24. Cravatt, B.F., Wright, A.T. & Kozarich, J.W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu Rev Biochem* **77**, 383-414 (2008).
25. Basle, E., Joubert, N. & Pucheault, M. Protein chemical modification on endogenous amino acids. *Chem Biol* **17**, 213-227 (2010).
26. Liu, Y., Patricelli, M.P. & Cravatt, B.F. Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A* **96**, 14694-14699 (1999).
27. Bachovchin, D.A. et al. Superfamily-wide portrait of serine hydrolase inhibition achieved by library-versus-library screening. *Proc Natl Acad Sci U S A* **107**, 20941-20946 (2010).
28. Mahrus, S. & Craik, C.S. Selective chemical functional probes of granzymes A and B reveal granzyme B is a major effector of natural killer cell-mediated lysis of target cells. *Chem Biol* **12**, 567-577 (2005).
29. Shannon, D.A. et al. Sulfonyl Fluoride Analogues as Activity-Based Probes for Serine Proteases. *Chembiochem* **13**, 2327-2330 (2012).
30. Adibekian, A. et al. Click-generated triazole ureas as ultrapotent in vivo-active serine hydrolase inhibitors. *Nat Chem Biol* **7**, 469-478 (2011).
31. Berger, A.B. et al. Identification of early intermediates of caspase activation using selective inhibitors and activity-based probes. *Mol Cell* **23**, 509-521 (2006).
32. Greenbaum, D., Medzihradszky, K.F., Burlingame, A. & Bogyo, M. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chemistry & Biology* **7**, 569-581 (2000).
33. Adam, G.C., Cravatt, B.F. & Sorensen, E.J. Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chemistry & Biology* **8**, 81-95 (2001).
34. Shannon, D.A. et al. Investigating the proteome reactivity and selectivity of aryl halides. *J Am Chem Soc* **136**, 3330-3333 (2014).
35. Barglow, K.T. & Cravatt, B.F. Substrate mimicry in an activity-based probe that targets the nitrilase family of enzymes. *Angew Chem Int Edit* **45**, 7408-7411 (2006).
36. Banerjee, R., Pace, N.J., Brown, D.R. & Weerapana, E. 1,3,5-Triazine as a Modular Scaffold for Covalent Inhibitors with Streamlined Target Identification. *Journal of the American Chemical Society* **135**, 2497-2500 (2013).
37. Witte, M.D. et al. Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat Chem Biol* **6**, 907-913 (2010).
38. Kallemeijn, W.W. et al. Novel Activity-Based Probes for Broad-Spectrum Profiling of

Retaining beta-Exoglucosidases In Situ and In Vivo. *Angew Chem Int Edit* **51**, 12529-12533 (2012).

39. Wilson, D. et al. SUPERFAMILY-sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Research* **37**, D380-D386 (2009).
40. Wilson, D., Madera, M., Vogel, C., Chothia, C. & Gough, J. The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Research* **35**, D308-D313 (2007).
41. Fox, N.K., Brenner, S.E. & Chandonia, J.M. SCOPe: Structural Classification of Proteins-extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research* **42**, D304-D309 (2014).
42. Mindrebo, J.T., Nartey, C.M., Seto, Y., Burkartl, M.D. & Noel, J.P. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom (vol 41, pg 233, 2016). *Curr Opin Struc Biol* **41**, 256-257 (2016).
43. Gershater, M., Sharples, K. & Edwards, R. Carboxylesterase activities toward pesticide esters in crops and weeds. *Phytochemistry* **67**, 2561-2567 (2006).
44. Gershater, M.C. & Edwards, R. Regulating biological activity in plants with carboxylesterases. *Plant Sci* **173**, 579-588 (2007).
45. Fraser, C.M., Rider, L.W. & Chapple, C. An expression and bioinformatics analysis of the Arabidopsis serine carboxypeptidase-like gene family. *Plant Physiol* **138**, 1136-1148 (2005).
46. van der Hoorn, R.A.L. Plant proteases: From phenotypes to molecular mechanisms. *Annu Rev Plant Biol* **59**, 191-223 (2008).
47. Jones, J.D.G. & Dangl, J.L. The plant immune system. *Nature* **444**, 323-329 (2006).
48. Canonne, J., Froidure-Nicolas, S. & Rivas, S. Phospholipases in action during plant defense signaling. *Plant Signal Behav* **6**, 13-18 (2011).
49. Kim, E.Y., Seo, Y.S. & Kim, W.T. AtDSEL, an Arabidopsis cytosolic DAD1-like acylhydrolase, is involved in negative regulation of storage oil mobilization during seedling establishment. *J Plant Physiol* **168**, 1705-1709 (2011).
50. Matsui, K., Fukutomi, S., Ishii, M. & Kajiwara, T. A tomato lipase homologous to DAD1 (LeLID1) is induced in post-germinative growing stage and encodes a triacylglycerol lipase. *FEBS Lett* **569**, 195-200 (2004).
51. Gruber, K., Gugganig, M., Wagner, U.G. & Kratky, C. Atomic resolution crystal structure of hydroxynitrile lyase from Hevea brasiliensis. *Biol Chem* **380**, 993-1000 (1999).
52. Gartler, G., Kratky, C. & Gruber, K. Structural determinants of the enantioselectivity of the hydroxynitrile lyase from Hevea brasiliensis. *Journal of Biotechnology* **129**, 87-97 (2007).
53. Auldridge, M.E. et al. Emergent Decarboxylase Activity and Attenuation of alpha/beta-Hydrolase Activity during the Evolution of Methylketone Biosynthesis in Tomato. *Plant Cell* **24**, 1596-1607 (2012).
54. Santner, A. & Estelle, M. Recent advances and emerging trends in plant hormone signalling. *Nature* **459**, 1071-1078 (2009).
55. Santner, A., Calderon-Villalobos, L.I.A. & Estelle, M. Plant hormones are versatile chemical regulators of plant growth. *Nat Chem Biol* **5**, 301-307 (2009).
56. Ueguchi-Tanaka, M. et al. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature* **437**, 693-698 (2005).
57. Zhao, L.H. et al. Crystal structures of two phytohormone signal-transducing alpha/beta

hydrolases: karrikin-signaling KAI2 and strigolactone-signaling DWARF14. *Cell Res* **23**, 436-439 (2013).
58. Sun, T.P. & Gubler, F. Molecular mechanism of gibberellin signaling in plants. *Annu Rev Plant Biol* **55**, 197-223 (2004).
59. Hedden, P. & Thomas, S.G. Gibberellin biosynthesis and its regulation. *Biochem J* **444**, 11-25 (2012).
60. Yamaguchi, S. Gibberellin metabolism and its regulation. *Annu Rev Plant Biol* **59**, 225-251 (2008).
61. Magome, H. et al. CYP714B1 and CYP714B2 encode gibberellin 13-oxidases that reduce gibberellin activity in rice. *P Natl Acad Sci USA* **110**, 1947-1952 (2013).
62. Nomura, T. et al. Functional Analysis of Arabidopsis CYP714A1 and CYP714A2 Reveals That They are Distinct Gibberellin Modification Enzymes. *Plant Cell Physiol* **54**, 1837-1851 (2013).
63. Ueguchi-Tanaka, M., Nakajima, M., Motoyuki, A. & Matsuoka, M. Gibberellin receptor and its role in gibberellin signaling in plants. *Annu Rev Plant Biol* **58**, 183-198 (2007).
64. Ueguchi-Tanaka, M. et al. Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. *Plant Cell* **19**, 2140-2155 (2007).
65. Nakajima, M. et al. Identification and characterization of Arabidopsis gibberellin receptors. *Plant J* **46**, 880-889 (2006).
66. Sun, T.P. The Molecular Mechanism and Evolution of the GA-GID1-DELLA Signaling Module in Plants. *Curr Biol* **21**, R338-R345 (2011).
67. Sasaki, A. et al. Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. *Science* **299**, 1896-1898 (2003).
68. McGinnis, K.M. et al. The Arabidopsis SLEEPY1 gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. *Plant Cell* **15**, 1120-1130 (2003).
69. Achard, P. & Genschik, P. Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. *J Exp Bot* **60**, 1085-1092 (2009).
70. Dill, A., Thomas, S.G., Hu, J.H., Steber, C.M. & Sun, T.P. The Arabidopsis F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. *Plant Cell* **16**, 1392-1405 (2004).
71. Dill, A. & Sun, T.P. Synergistic derepression of gibberellin signaling by removing RGA and GAI function in Arabidopsis thaliana. *Genetics* **159**, 777-785 (2001).
72. Shimada, A. et al. Structural basis for gibberellin recognition by its receptor GID1. *Nature* **456**, 520-U544 (2008).
73. Murase, K., Hirano, Y., Sun, T.P. & Hakoshima, T. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **456**, 459-U415 (2008).
74. Hirano, K. et al. The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte Selaginella moellendorffii but not in the bryophyte Physcomitrella patens. *Plant Cell* **19**, 3058-3079 (2007).
75. Yoshida, H. et al. Evolution and diversification of the plant gibberellin receptor GID1. *P Natl Acad Sci USA* **115**, E7844-E7853 (2018).
76. Bodi, M.B. et al. Wildland fire ash: Production, composition and eco-hydro-geomorphic effects (vol 130, pg 103, 2014). *Earth-Sci Rev* **138**, 503-503 (2014).
77. Pausas, J.G. & Keeley, J.E. Wildfires as an ecosystem service. *Front Ecol Environ* **17**, 289-

295 (2019).
78. Certini, G. Fire as a Soil-Forming Factor. *Ambio* **43**, 191-195 (2014).
79. Downer, P.S. & Harter, R.D. Effects of Fire on Forest Soil and Nutrient Cycling - Annotated-Bibliography. *New Hamp Aes Bull*, 1-31 (1979).
80. Roche, S., Dixon, K.W. & Pate, J.S. Seed ageing and smoke: Partner cues in the amelioration of seed dormancy in selected Australian native species. *Aust J Bot* **45**, 783-815 (1997).
81. Baskin, C.C. & Baskin, J.M. Seeds Ecology, Biogeography, and Evolution of Dormancy and Germination Second Edition Introduction. *Seeds: Ecology, Biogeography, and Evolution of Dormancy and Germination, 2nd Edition*, 1 (2014).
82. Flematti, G.R., Ghisalberti, E.L., Dixon, K.W. & Trengove, R.D. A compound from smoke that promotes seed germination. *Science* **305**, 977-977 (2004).
83. Flematti, G.R., Ghisalberti, E.L., Dixon, K.W. & Trengove, R.D. Identification of Alkyl Substituted 2H-Furo[2,3-c]pyran-2-ones as Germination Stimulants Present in Smoke. *J Agr Food Chem* **57**, 9475-9480 (2009).
84. Flematti, G.R. et al. Structure-Activity Relationship of Karrikin Germination Stimulants. *J Agr Food Chem* **58**, 8612-8617 (2010).
85. Nelson, D.C. et al. Karrikins Discovered in Smoke Trigger Arabidopsis Seed Germination by a Mechanism Requiring Gibberellic Acid Synthesis and Light. *Plant Physiol* **149**, 863-873 (2009).
86. Nelson, D.C., Flematti, G.R., Ghisalberti, E.L., Dixon, K. & Smith, S.M. Regulation of Seed Germination and Seedling Growth by Chemical Signals from Burning Vegetation. *Annual Review of Plant Biology, Vol 63* **63**, 107-130 (2012).
87. Waters, M.T., Scaffidi, A., Sun, Y.M.K., Flematti, G.R. & Smith, S.M. The karrikin response system of Arabidopsis. *Plant J* **79**, 623-631 (2014).
88. Zwanenburg, B. & Pospisil, T. Structure and activity of strigolactones: new plant hormones with a rich future. *Mol Plant* **6**, 38-62 (2013).
89. Boyer, F.D. et al. Structure-activity relationship studies of strigolactone-related molecules for branching inhibition in garden pea: molecule design for shoot branching. *Plant Physiol* **159**, 1524-1544 (2012).
90. Cook, C.E., Whichard, L.P., Turner, B., Wall, M.E. & Egley, G.H. Germination of Witchweed (Striga lutea Lour.): Isolation and Properties of a Potent Stimulant. *Science* **154**, 1189-1190 (1966).
91. Akiyama, K., Matsuzaki, K. & Hayashi, H. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435**, 824-827 (2005).
92. Gomez-Roldan, V. et al. Strigolactone inhibition of shoot branching. *Nature* **455**, 189-194 (2008).
93. Kapulnik, Y. et al. Strigolactones affect lateral root formation and root-hair elongation in Arabidopsis. *Planta* **233**, 209-216 (2011).
94. Nelson, D.C. et al. Karrikins enhance light responses during germination and seedling development in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **107**, 7095-7100 (2010).
95. Kulkarni, M.G., Sparg, S.G., Light, M.E. & van Staden, J. Stimulation of rice (Oryza sativa L.) seedling vigour by smoke-water and butenolide. *J Agron Crop Sci* **192**, 395-398 (2006).
96. Jain, N., Kulkarni, M.G. & van Staden, J. A butenolide, isolated from smoke, can overcome

the detrimental effects of extreme temperatures during tomato seed germination. *Plant Growth Regul* **49**, 263-267 (2006).

97. Waheed, M.A., Jamil, M., Khan, M.D., Shakir, S.K. & Shafiq-Ur-Rehman Effect of Plant-Derived Smoke Solutions Onphysiological and Biochemical Attributes of Maize (Zea Mays L.) under Salt Stress. *Pak J Bot* **48**, 1763-1774 (2016).
98. Nelson, D.C. et al. F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in Arabidopsis thaliana. *P Natl Acad Sci USA* **108**, 8897-8902 (2011).
99. Sun, X.D. & Ni, M. HYPOSENSITIVE TO LIGHT, an Alpha/Beta Fold Protein, Acts Downstream of ELONGATED HYPOCOTYL 5 to Regulate Seedling De-Etiolation. *Molecular Plant* **4**, 116-126 (2011).
100. Hamiaux, C. et al. DAD2 Is an alpha/beta Hydrolase Likely to Be Involved in the Perception of the Plant Branching Hormone, Strigolactone. *Curr Biol* **22**, 2032-2036 (2012).
101. Guo, Y.X., Zheng, Z.Y., La Clair, J.J., Chory, J. & Noel, J.P. Smoke-derived karrikin perception by the alpha/beta-hydrolase KAI2 from Arabidopsis. *P Natl Acad Sci USA* **110**, 8284-8289 (2013).
102. Smith, S.M. & Li, J. Signalling and responses to strigolactones and karrikins. *Curr Opin Plant Biol* **21**, 23-29 (2014).
103. Huxley, T.H. Darwin on the Origin of Species. (Westminster Review; 1860).
104. Darwin, C.F., D. The power of movement in plants. (John Murray; 1880).
105. Morrison, K.L. & Weiss, G.A. The origins of chemical biology. *Nat Chem Biol* **2**, 3-6 (2006).
106. Prescher, J.A. & Bertozzi, C.R. Chemistry in living systems. *Nat Chem Biol* **1**, 13-21 (2005).
107. Surpin, M. & Raikhel, N. Traffic jams affect plant development and signal transduction. *Nat Rev Mol Cell Biol* **5**, 100-109 (2004).
108. Hicks, G.R. & Raikhel, N.V. Opportunities and challenges in plant chemical biology. *Nat Chem Biol* **5**, 268-272 (2009).
109. Toth, R. & van der Hoorn, R.A. Emerging principles in plant chemical genetics. *Trends Plant Sci* **15**, 81-88 (2010).
110. Chan, J., Dodani, S.C. & Chang, C.J. Reaction-based small-molecule fluorescent probes for chemoselective bioimaging. *Nat Chem* **4**, 973-984 (2012).
111. Park, S.Y. et al. Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins. *Science* **324**, 1068-1071 (2009).
112. Zhao, Y. A chemical genetics method to uncover small molecules for dissecting the mechanism of ABA responses in Arabidopsis seed germination. *Methods Mol Biol* **876**, 107-116 (2012).
113. Hayashi, K. et al. Rational Design of an Auxin Antagonist of the SCFTIR1 Auxin Receptor Complex. *Acs Chem Biol* **7**, 590-598 (2012).
114. Takeuchi, J. et al. Rationally Designed Strigolactone Analogs as Antagonists of the D14 Receptor. *Plant Cell Physiol* **59**, 1545-1554 (2018).
115. Uraguchi, D. et al. A femtomolar-range suicide germination stimulant for the parasitic plant Striga hermonthica. *Science* **362**, 1301-1305 (2018).
116. Stockwell, B.R. Chemical genetics: ligand-based discovery of gene function. *Nat Rev Genet* **1**, 116-125 (2000).
117. Jiang, K. & Asami, T. Chemical regulators of plant hormones and their applications in basic

research and agriculture. *Biosci Biotech Bioch* **82**, 1265-1300 (2018).
118. Tsuchiya, Y. Small Molecule Toolbox for Strigolactone Biology. *Plant Cell Physiol* **59**, 1511-1519 (2018).
119. Conn, C.E. et al. Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* **349**, 540-543 (2015).
120. Tsuchiya, Y. et al. PARASITIC PLANTS. Probing strigolactone receptors in Striga hermonthica with fluorescence. *Science* **349**, 864-868 (2015).
121. Yao, R.F. et al. DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature* **536**, 469-473 (2016).
122. Shabek, N. et al. Structural plasticity of D3-D14 ubiquitin ligase in strigolactone signalling. *Nature* **563**, 652-656 (2018).
123. Rich, P.J. & Ejeta, G. Towards effective resistance to Striga in African maize. *Plant Signal Behav* **3**, 618-621 (2008).
124. Chai, L.J. et al. RNA sequencing reveals high resolution expression change of major plant hormone pathway genes after young seedless grape berries treated with gibberellin. *Plant Sci* **229**, 215-224 (2014).
125. Colak, A.M. Effect of melatonin and gibberellic acid foliar application on the yield and quality of Jumbo blackberry species. *Saudi J Biol Sci* **25**, 1242-1246 (2018).
126. Villedieu-Percheron, E. et al. Chemicals Inducing Seed Germination and Early Seedling Development. *Chimia* **68**, 654-663 (2014).
127. Sasaki, A. et al. Green revolution: A mutant gibberellin-synthesis gene in rice - New insight into the rice variant that helped to avert famine over thirty years ago. *Nature* **416**, 701-702 (2002).
128. Park, S.Y. et al. Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature* **520**, 545-548 (2015).
129. Uchida, N. et al. Chemical hijacking of auxin signaling with an engineered auxin-TIR1 pair. *Nat Chem Biol* **14**, 299-305 (2018).
130. Jiang, K. et al. Substituted Phthalimide AC94377 Is a Selective Agonist of the Gibberellin Receptor GID1. *Plant Physiol* **173**, 825-835 (2017).
131. Conn, C.E. & Nelson, D.C. Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Front Plant Sci* **6** (2016).
132. Morffy, N., Faure, L. & Nelson, D.C. Smoke and Hormone Mirrors: Action and Evolution of Karrikin and Strigolactone Signaling. *Trends Genet* **32**, 176-188 (2016).
133. Swarbreck, S.M. Phytohormones Interplay: Karrikin Signalling Promotes Ethylene Synthesis to Modulate Roots. *Trends Plant Sci* **26**, 308-311 (2021).
134. Waters, M.T. et al. Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in Arabidopsis. *Development* **139**, 1285-1295 (2012).
135. Swarbreck, S.M., Mohammad-Sidik, A. & Davies, J.M. Common Components of the Strigolactone and Karrikin Signaling Pathways Suppress Root Branching in Arabidopsis. *Plant Physiol* **184**, 18-22 (2020).
136. Soundappan, I. et al. SMAX1-LIKE/D53 Family Members Enable Distinct MAX2-Dependent Responses to Strigolactones and Karrikins in Arabidopsis. *Plant Cell* **27**, 3143-

3159 (2015).
137. Khosla, A. et al. Structure-Function Analysis of SMAX1 Reveals Domains That Mediate Its Karrikin-Induced Proteolysis and Interaction with the Receptor KAI2. *Plant Cell* **32**, 2639-2659 (2020).
138. Stanga, J.P., Smith, S.M., Briggs, W.R. & Nelson, D.C. SUPPRESSOR OF MORE AXILLARY GROWTH2 1 Controls Seed Germination and Seedling Development in Arabidopsis. *Plant Physiol* **163**, 318-330 (2013).
139. Zhou, F. et al. D14-SCFD3-dependent degradation of D53 regulates strigolactone signalling. *Nature* **504**, 406-410 (2013).
140. Wang, L. et al. Strigolactone Signaling in Arabidopsis Regulates Shoot Development by Targeting D53-Like SMXL Repressor Proteins for Ubiquitination and Degradation. *Plant Cell* **27**, 3128-3142 (2015).
141. Swarbreck, S.M., Guerringue, Y., Matthus, E., Jamieson, F.J.C. & Davies, J.M. Impairment in karrikin but not strigolactone sensing enhances root skewing in Arabidopsis thaliana. *Plant J* **98**, 607-621 (2019).
142. Wang, L. et al. Strigolactone and Karrikin Signaling Pathways Elicit Ubiquitination and Proteolysis of SMXL2 to Regulate Hypocotyl Elongation in Arabidopsis. *Plant Cell* **32**, 2251-2270 (2020).
143. Scaffidi, A. et al. Strigolactone Hormones and Their Stereoisomers Signal through Two Related Receptor Proteins to Induce Different Physiological Responses in Arabidopsis. *Plant Physiol* **165**, 1221-1232 (2014).
144. Ebdrup, S., Sorensen, L.G., Olsen, O.H. & Jacobsen, P. Synthesis and structure-activity relationship for a novel class of potent and selective carbamoyl-triazole based inhibitors of hormone sensitive lipase. *J Med Chem* **47**, 400-410 (2004).
145. Yammine, A., Gao, J.L. & Kwan, A.H. Tryptophan Fluorescence Quenching Assays for Measuring Protein-ligand Binding Affinities: Principles and a Practical Guide. *Bio-Protocol* **9** (2019).
146. Ghisaidoobe, A.B. & Chung, S.J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Forster resonance energy transfer techniques. *Int J Mol Sci* **15**, 22518-22538 (2014).
147. Callis, P.R. & Liu, T.Q. Quantitative prediction of fluorescence quantum yields for tryptophan in proteins. *J Phys Chem B* **108**, 4248-4259 (2004).
148. Mocz, G. & Ross, J.A. Fluorescence techniques in analysis of protein-ligand interactions. *Methods Mol Biol* **1008**, 169-210 (2013).
149. Xiang, H.Y. et al. Thermodynamic characterization of OsGID1-gibberellin binding using calorimetry and docking simulations. *J Mol Recognit* **24**, 275-282 (2011).
150. Forli, S. et al. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* **11**, 905-919 (2016).
151. Kinoshita, T., McCourt, P., Asami, T. & Torii, K.U. Plant Chemical Biology. *Plant Cell Physiol* **59**, 1483-1486 (2018).
152. Hagihara, S., Yamada, R., Itami, K. & Torii, K.U. Dissecting plant hormone signaling with synthetic molecules: perspective from the chemists. *Curr Opin Plant Biol* **47**, 32-37 (2019).
153. Fonseca, S., Rosado, A., Vaughan-Hirsch, J., Bishopp, A. & Chini, A. Molecular locks and keys: the role of small molecules in phytohormone research. *Front Plant Sci* **5**, 709 (2014).

154. Koh, J.T. Engineering selectivity and discrimination into ligand-receptor interfaces. *Chemistry & Biology* **9**, 17-23 (2002).
155. Chockalingam, K. & Zhao, H.M. Creating new specific ligand-receptor pairs for transgene regulation. *Trends Biotechnol* **23**, 333-335 (2005).
156. Shi, Y.H. & Koh, J.T. Selective regulation of gene expression by an orthogonal estrogen receptor-ligand pair created by polar-group exchange. *Chemistry & Biology* **8**, 501-510 (2001).
157. Shi, Y.H. & Koh, J.T. Functionally orthogonal ligand-receptor pairs for the selective regulation of gene expression generated by manipulation of charged residues at the ligand-receptor interface of ER alpha and ER beta. *Journal of the American Chemical Society* **124**, 6921-6928 (2002).
158. Doyle, D.F. et al. Engineering orthogonal ligand-receptor pairs from "near drugs". *Journal of the American Chemical Society* **123**, 11367-11371 (2001).
159. Palli, S.R., Kapitskaya, M.Z., Kumar, M.B. & Cress, D.E. Improved ecdysone receptor-based inducible gene regulation system. *European Journal of Biochemistry* **270**, 1308-1315 (2003).
160. Hou, J.L. et al. Targeted delivery of nitric oxide via a 'bump-and-hole'-based enzyme-prodrug pair. *Nat Chem Biol* **15**, 151-160 (2019).
161. Belshaw, P.J., Schoepfer, J.G., Liu, K.Q., Morrison, K.L. & Schreiber, S.L. Rational Design of Orthogonal Receptor-Ligand Combinations. *Angewandte Chemie-International Edition in English* **34**, 2129-2132 (1995).
162. Runcie, A.C. et al. Optimization of a "bump-and-hole" approach to allele-selective BET bromodomain inhibition. *Chem Sci* **9**, 2452-2468 (2018).
163. Baud, M.G.J. et al. A bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes. *Science* **346**, 638-641 (2014).
164. Yamada, R. et al. A Super Strong Engineered Auxin-TIR1 Pair. *Plant Cell Physiol* **59**, 1538-1544 (2018).
165. Chockalingam, K., Chen, Z.L., Katzenellenbogen, J.A. & Zhao, H.M. Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *P Natl Acad Sci USA* **102**, 5691-5696 (2005).
166. Reetz, M.T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *Chembiochem* **9**, 1797-1804 (2008).
167. Li, A., Acevedo-Rocha, C.G. & Reetz, M.T. Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. *Appl Microbiol Biotechnol* **102**, 6095-6103 (2018).
168. Kille, S. et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth Biol* **2**, 83-92 (2013).
169. Li, D., Wu, Q. & Reetz, M.T. Focused rational iterative site-specific mutagenesis (FRISM). *Methods Enzymol* **643**, 225-242 (2020).
170. Reetz, M.T. & Carballeira, J.D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* **2**, 891-903 (2007).
171. Li, A.T., Qu, G., Sun, Z.T. & Reetz, M.T. Statistical Analysis of the Benefits of Focused Saturation Mutagenesis in Directed Evolution Based on Reduced Amino Acid Alphabets. *Acs Catalysis* **9**, 7769-7778 (2019).

172. Qu, G., Li, A., Acevedo-Rocha, C.G., Sun, Z. & Reetz, M.T. The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes. *Angew Chem Int Ed Engl* **59**, 13204-13231 (2020).
173. Jochens, H. et al. Protein engineering of alpha/beta-hydrolase fold enzymes. *Chembiochem* **12**, 1508-1517 (2011).
174. Jones, B.J., Lim, H.Y., Huang, J. & Kazlauskas, R.J. Comparison of Five Protein Engineering Strategies for Stabilizing an alpha/beta-Hydrolase. *Biochemistry* **56**, 6521-6532 (2017).
175. Kourist, R. et al. The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *Chembiochem* **11**, 1635-1643 (2010).
176. Feng, H., Wang, H.Y. & Zhao, H.Y. Novel Random Mutagenesis Method for Directed Evolution. *Methods Mol Biol* **1498**, 483-490 (2017).
177. Acevedo-Rocha, C.G. & Reetz, M.T. Assembly of Designed Oligonucleotides: a useful tool in synthetic biology for creating high-quality combinatorial DNA libraries. *Methods Mol Biol* **1179**, 189-206 (2014).
178. Zeymer, C. & Hilvert, D. Directed Evolution of Protein Catalysts. *Annual Review of Biochemistry, Vol 87* **87**, 131-157 (2018).
179. Chen, K.Q. & Arnold, F.H. Tuning the Activity of an Enzyme for Unusual Environments - Sequential Random Mutagenesis of Subtilisin-E for Catalysis in Dimethylformamide. *P Natl Acad Sci USA* **90**, 5618-5622 (1993).
180. Arnold, F.H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew Chem Int Edit* **58**, 14420-14426 (2019).
181. Lutz, S. Beyond directed evolution-semi-rational protein engineering and design. *Curr Opin Biotech* **21**, 734-743 (2010).
182. Pines, G. et al. Codon Compression Algorithms for Saturation Mutagenesis. *Acs Synthetic Biology* **4**, 604-614 (2015).
183. Sayous, V., Lubrano, P., Li, Y.Y. & Acevedo-Rocha, C.G. Unbiased libraries in protein directed evolution. *Bba-Proteins Proteom* **1868** (2020).
184. Edelheit, O., Hanukoglu, A. & Hanukoglu, I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *Bmc Biotechnol* **9** (2009).
185. Hogrefe, H.H., Cline, J., Youngblood, G.L. & Allen, R.M. Creating randomized amino acid libraries with the QuikChange (R) Multi Site-Directed Mutagenesis Kit. *Biotechniques* **33**, 1158-1165 (2002).
186. Miyazaki, K. & Takenouchi, M. Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. *Biotechniques* **33**, 1033-1038 (2002).
187. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain-Reaction. *Gene* **77**, 51-59 (1989).
188. Liu, H.T. & Naismith, J.H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *Bmc Biotechnol* **8** (2008).
189. Qu, G., Li, A.T., Acevedo-Rocha, C.G., Sun, Z.T. & Reetz, M.T. The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes. *Angew Chem Int Edit* **59**, 13204-13231 (2020).
190. Reetz, M.T. The Importance of Additive and Non-Additive Mutational Effects in Protein

Engineering. *Angew Chem Int Edit* **52**, 2658-2666 (2013).
191. Reetz, M.T., Prasad, S., Carballeira, J.D., Gumulya, Y. & Bocola, M. Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity: Rigorous Comparison with Traditional Methods. *Journal of the American Chemical Society* **132**, 9144-9152 (2010).
192. Nemoto, K. et al. Tyrosine phosphorylation of the GARU E3 ubiquitin ligase promotes gibberellin signalling by preventing GID1 degradation. *Nature Communications* **8** (2017).
193. Monna, L. et al. Positional cloning of rice semidwarfing gene, sd-1: Rice "Green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res* **9**, 11-17 (2002).
194. Nakamura, H. et al. Triazole Ureas Covalently Bind to Strigolactone Receptor and Antagonize Strigolactone Responses. *Molecular Plant* **12**, 44-58 (2019).
195. Sullivan, B., Walton, A.Z. & Stewart, J.D. Library construction and evaluation for site saturation mutagenesis. *Enzyme Microb Technol* **53**, 70-77 (2013).
196. Bythell-Douglas, R. et al. Evolution of strigolactone receptors by gradual neo-functionalization of KAI2 paralogues. *BMC Biol* **15**, 52 (2017).
197. Walker, C.H., Siu-Ting, K., Taylor, A., O'Connell, M.J. & Bennett, T. Strigolactone synthesis is ancestral in land plants, but canonical strigolactone signalling is a flowering plant innovation. *Bmc Biology* **17** (2019).
198. Waters, M.T. et al. A Selaginella moellendorffii Ortholog of KARRIKIN INSENSITIVE2 Functions in Arabidopsis Development but Cannot Mediate Responses to Karrikins or Strigolactones. *Plant Cell* **27**, 1925-1944 (2015).
199. Umehara, M. et al. Structural Requirements of Strigolactones for Shoot Branching Inhibition in Rice and Arabidopsis. *Plant Cell Physiol* **56**, 1059-1072 (2015).
200. Waters, M.T. & Smith, S.M. KAI2- and MAX2-Mediated Responses to Karrikins and Strigolactones Are Largely Independent of HY5 in Arabidopsis Seedlings. *Molecular Plant* **6**, 63-75 (2013).
201. Holbrook-Smith, D., Toh, S., Tsuchiya, Y. & McCourt, P. Small-molecule antagonists of germination of the parasitic plant Striga hermonthica. *Nat Chem Biol* **12**, 724-729 (2016).
202. Yao, J.R. et al. An allelic series at the KARRIKIN INSENSITIVE 2 locus of Arabidopsis thaliana decouples ligand hydrolysis and receptor degradation from downstream signalling. *Plant J* **96**, 75-89 (2018).
203. Bythell-Douglas, R. et al. The Structure of the Karrikin-Insensitive Protein (KAI2) in Arabidopsis thaliana. *Plos One* **8** (2013).
204. Danley, D.E. Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallogr D* **62**, 569-575 (2006).
205. Hassell, A.M. et al. Crystallization of protein-ligand complexes. *Acta Crystallogr D* **63**, 72-79 (2007).
206. Mueller, I. Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallographica Section D-Structural Biology* **73**, 79-92 (2017).
207. Lee, I. et al. A missense allele of KARRIKIN-INSENSITIVE2 impairs ligand-binding and downstream signaling in Arabidopsis thaliana. *J Exp Bot* **69**, 3609-3623 (2018).
208. Xu, Y. et al. Structural basis of unique ligand specificity of KAI2-like protein from parasitic weed Striga hermonthica. *Sci Rep* **6**, 31386 (2016).
209. Xu, Y.Q. et al. Structural analysis of HTL and D14 proteins reveals the basis for ligand

selectivity in Striga. *Nature Communications* **9** (2018).
210. Burger, M. & Chory, J. In-silico analysis of the strigolactone ligand-receptor system. *Plant Direct* **4** (2020).
211. Chen, J.M., White, A., Nelson, D.C. & Shukla, D. Role of substrate recognition in modulating strigolactone receptor selectivity in witchweed. *J Biol Chem* **297** (2021).
212. Hu, F. et al. MD Simulation Investigation on the Binding Process of Smoke-Derived Germination Stimulants to Its Receptor. *J Chem Inf Model* **59**, 1554-1562 (2019).
213. Arellano-Saab, A. et al. Three mutations repurpose a plant karrikin receptor to a strigolactone receptor. *P Natl Acad Sci USA* **118** (2021).
214. Imai, Y.N., Inoue, Y., Nakanishi, I. & Kitaura, K. Cl-pi interactions in protein-ligand complexes. *Protein Sci* **17**, 1129-1137 (2008).
215. Jian, J. et al. Probing Halogen-pi versus CH-pi Interactions in Molecular Balance. *Org Lett* **22**, 7870-7873 (2020).
216. Platzer, G. et al. PI by NMR: Probing CH-pi Interactions in Protein-Ligand Complexes by NMR Spectroscopy. *Angew Chem Int Ed Engl* **59**, 14861-14868 (2020).
217. Yao, J. et al. Desmethyl butenolides are optimal ligands for karrikin receptor proteins. *New Phytol* **230**, 1003-1016 (2021).
218. Carlsson, G.H., Hasse, D., Cardinale, F., Prandi, C. & Andersson, I. The elusive ligand complexes of the DWARF14 strigolactone receptor. *J Exp Bot* **69**, 2345-2354 (2018).
219. Seto, Y. et al. Strigolactone perception and deactivation by a hydrolase receptor DWARF14. *Nat Commun* **10**, 191 (2019).
220. Marzec, M. & Brewer, P. Binding or Hydrolysis? How Does the Strigolactone Receptor Work? *Trends in Plant Science* **24**, 571-574 (2019).
221. Stanga, J.P., Morffy, N. & Nelson, D.C. Functional redundancy in the control of seedling growth by the karrikin signaling pathway. *Planta* **243**, 1397-1406 (2016).
222. Wong, J.H. et al. A yeast two-hybrid system for the screening and characterization of small-molecule inhibitors of protein-protein interactions identifies a novel putative Mdm2-binding site in p53. *Bmc Biology* **15** (2017).
223. Tanaka, S. An efficient method for the isolation of interaction-null/impaired mutants using the yeast two-hybrid technique. *Genes Cells* **24**, 781-788 (2019).
224. Fukui, K., Arai, K., Kasahara, H., Asami, T. & Hayashi, K. Synthetic agonist of HTL/KAI2 shows potent stimulating activity for Arabidopsis seed germination. *Bioorg Med Chem Lett* **29**, 2487-2492 (2019).
225. Toh, S., Holbrook-Smith, D., Stokes, M.E., Tsuchiya, Y. & McCourt, P. Detection of Parasitic Plant Suicide Germination Compounds Using a High-Throughput Arabidopsis HTL/KAI2 Strigolactone Perception System (vol 21, pg 988, 2014). *Chemistry & Biology* **21**, 1253-1253 (2014).
226. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
227. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
228. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
229. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
230. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface

protein sequences with the hydrophobic moment plot. *J Mol Biol* **179**, 125-142 (1984).

231. Lashkov, A.A., Tolmachev, I.V., Eistrikh-Heller, P.A. & Rubinsky, S.V. PyFepRestr: Plugin to PyMOL Molecular Graphics System for Calculating the Free Energy of Ligand-Receptor Binding. *Crystallogr Rep* **66**, 861-865 (2021).
232. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324 (2014).
233. Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* **36**, W465-W469 (2008).
234. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293-W296 (2021).
235. Gietz, R.D. & Schiestl, R.H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2**, 38-41 (2007).