Jan Hejátko · Toshio Hakoshima *Editors*

Plant Structural Biology: Hormonal Regulations

Springer

Plant Structural Biology: Hormonal Regulations

Jan Hejátko • Toshio Hakoshima
Editors

Plant Structural Biology: Hormonal Regulations

Springer

Editors
Jan Hejátko
CEITEC - Central European Institute of Technology
Masaryk University
Brno, Czech Republic

Toshio Hakoshima
Structural Biology Laboratory
Nara Institute of Science and Technology
Ikoma, Nara, Japan

ISBN 978-3-319-91351-3 ISBN 978-3-319-91352-0 (eBook)
https://doi.org/10.1007/978-3-319-91352-0

Library of Congress Control Number: 2018947373

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

In terms of the intensity with which information can be interpreted, there is nothing better than a picture, and this helps structural biology stand at the forefront of interpretation for biochemistry. Plant structural biology is strong and is growing in importance even though there is a distinct lack of plant protein structures in the databases. With the number of structures arising from human proteins currently close to 1500, the highest ranked plant (*Spinacea oleracea*) has 50, but the long list of other species separating these two reduces the total contribution of plant protein structures to just a few percent of structures solved (based on RCSB PDB entries 2018). At a time when plant genomes are being sequenced at a fantastic rate, our understanding of what gene products do and the annotation necessary to help make sense of the genomic data are lagging behind. However, here lies opportunity, and the history of plant protein structural biology can be proud of its successes. This volume captures some of these successes in the field of hormone signalling.

Some readers may be entering the world of protein structures for the first time and others may be seeking to immerse themselves in the details of molecular signalling mechanisms, and both will find this volume rewarding. There are chapters introducing the various methodologies for solving protein structures, from X-ray crystallography as probably the most familiar to the increasingly popular cryo-electron microscopy. However, the bulk of the contributions are overviews of the structures of hormone receptors. In this, we are fortunate. Structures are introduced and explained for all the most important receptors and some of their interactors and, importantly, these structures are interpreted in terms of, e.g. the residues, loops and features which contribute mechanistically to function. As noted above, structural biology is visually strong and each article displays beautiful images to help explain how hormone perception works.

Most of the proteins described have been solved using crystallography. Fortunately, many plant hormone receptors are soluble proteins which have made their structural biology more amenable, but a subset are membrane-bound such as the receptors for ethylene and cytokinins. No structural biology project should be undertaken lightly given the immense efforts required to purify sufficient protein,

but difficulties for those interested in membrane proteins is amplified by the need to maintain protein integrity during extraction. It is with intrinsic membrane proteins that cryo-electron microscopy techniques are starting to play a vital role as its resolution gets better and better.

Collectively, the articles in this volume provide a welcome and exciting prelude to what I hope will be a wave of biochemical illuminations.

School of Life Sciences, University of Warwick, Coventry, UK

Richard Napier

Contents

Part I
Introduction

Chapter 1
Overview of Proteins in Plant Hormone Signaling

Toshio Hakoshima

1.1 Plant Hormones

Plant hormones (or phytohormones) play central roles in the integration of diverse environmental cues with signaling networks and the genetic programs of plants. In animals, hormones comprise any member of classes of signaling molecules produced by glands in multicellular organisms that are transported by the circulatory system to distant target organs so as to regulate physiology and behavior. Although plants do not possess glands that secrete hormones, they do possess several tissues that produce and also respond to hormones. Plant hormones comprise a set of structurally unrelated small organic compounds (Fig. 1.1), including auxin, jasmonic acid (JA), gibberellin (or gibberellic acid, GA), abscisic acid (ABA), brassinosteroid (BR), cytokinin, ethylene, strigolactone (SL), and salicylic acid (SA). Although karrikin (KAR) is not produced by plants, it is included as a plant hormone given its function as an "endocrine disruptor" or "environmental hormone".

In terms of chemical structure, some of the plant hormones could find animal counterparts produced from common precursors (Table 1.1) (Chow and McCourt 2006). Animals produce several steroid hormones that display similarity to BR. ABA and SL are produced from carotenoids, which are precursors of retinoic acid in animals. Plant JA and animal prostaglandins are both produced from fatty acids. Auxin is an indole acetic acid derived from indole, and its animal counterpart could be melatonin, which is produced in the pineal gland and regulates sleep and wakefulness. All of these animal counterparts seem not to possess similar functions as with the plant hormones. Generally, plant hormones possess a much broader range of functions compared to animal hormones, and a single plant cell can respond to more than one hormone, and a single hormone can affect different tissues in different

T. Hakoshima (✉)
Structural Biology Laboratory, Nara Institute of Science and Technology, Ikoma, Nara, Japan
e-mail: hakosima@bs.naist.jp

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_1

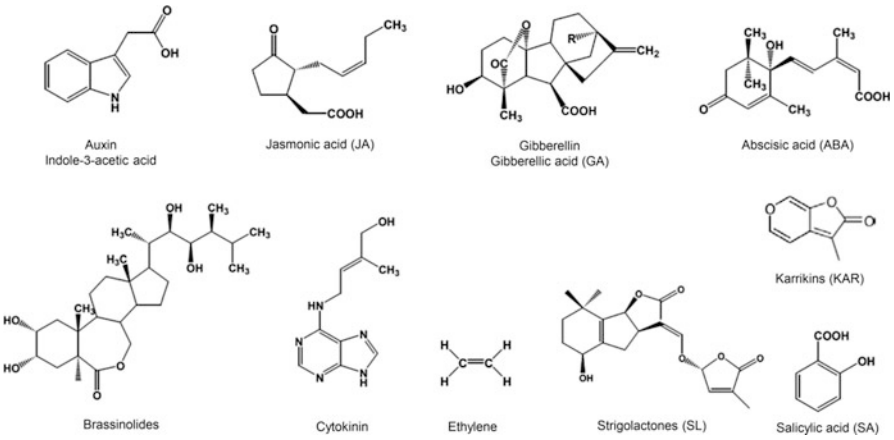


Fig. 1.1 Chemical structures of plant hormones

Table 1.1 Biosynthetic precursors of plant and animal hormones

Precursor	Plant	Animal
Sterol	Brassinosteroid (BR)	Progesterone
Fatty acid	Jasmonic acid (JA)	Prostaglandin
Indole	Auxin	Melatonin
Carotenoid	Abscisic acid (ABA)	Retinoic acid
	Strigolactone (SL)	
Geranylgeranyl-PP	Gibberellin (GA)	–
Purine	Cytokinin (CK)	–
Methionine	Ethylene	–
Phenylalanine	Salicylic acid (SA)	–

ways. GA is produced from geranylgeranyl-phosphates, cytokinin from purines, and ethylene from methionine, and these have no counterparts in animals. Animals do not produce SA, although SA exhibits remarkable pharmacological activity.

The number of recognized plant hormones may increase in the future, particularly if we are reminded that only 9 years have passed since SL was accepted as a plant hormone (Umehara et al. 2008; Gomez-Roldan et al. 2008). In addition to plant hormones comprised of organic compounds, we now know that signaling peptides also play prominent roles in regulating plant physiology by short-range intercellular communication mechanisms, and have established the concept of peptide hormones just as in the case of animals. The Arabidopsis genome encodes more than 1000 putative small signaling molecules, yet only a few polypeptides have been functionally characterized (Grienenberger and Fletcher 2015). Elucidation of the molecular functions and structures of these small peptides and their receptors is a frontier of current plant and structural biology.

1.2 Plant Hormone Nuclear Receptors

Animals possess a conserved nuclear receptor superfamily of transcription factors that perceive a variety of hormones. Humans have 48 nuclear receptors that contain orphan receptors that perceive unknown compounds and receptors that perceive compounds other than hormones, such as certain lipids (Evans and Mangelsdorf 2014). The genomic data, however, suggest that plants have no such nuclear receptor homologs. Instead, plants co-opt proteins from different protein families to perceive plant hormones (Table 1.2). One prominent characteristic found in plant hormone signaling is the central role played by the ubiquitin-proteasome system (UPS). In particular, cullin-RING ubiquitin ligase (CRL) complexes play a key role in the

Table 1.2 Plant hormone receptors

Hormone	Receptor	Receptor type
Nuclear receptors		
Auxin	TIR1 (TRANSPORT INHIBITOR RESPONSE 1)	LRR-type F-box protein: Ub-substrate[a] receptor
Jasmonic acid (JA)	COI1 (CORONATINE INSENSITIVE 1)	LRR-type F-box protein: Ub-substrate receptor
Gibberellin (GA)	GID1 (GIBERELLIN INSENSITIVE DWARF 1)	F-box protein adaptor (α/β hydrolase-superfamily)
		Ub-substrate receptor
Strigolactone (SL)	D14 (DWARF 14)	Hormone hydrolase (α/β hydrolase-superfamily)
		F-box protein adaptor: Ub-substrate receptor
Karrikin (KAR)[a]	KAI2 (KARRIKIN INSENSITIVE 2)	Hormone hydrolase (α/β hydrolase-superfamily)
	/D14L (D14-like)	F-box protein adaptor: Ub-substrate receptor
Salicylic acid (SA)	NPR3,4 (NONEXPRESSOR OF PR GENES 3, 4)	CUL3 adaptor (BTB domain protein)
		Ub-substrate receptor
Abscisic acid (ABA)	PYR/PYLs (PYRABACTIN RESISTANCE 1)	Phosphatase inhibitor (START-superfamily)
		Ub-substrate receptor
Receptor kinases		
Brassinosteroid (BR)	BRI1 (BRASSINOSTEROID INSENSITIVE 1)	LRR-type receptor kinase (Ser/Thr protein kinase domain)
Cytokinin (CK)	AHK2,3,4 (ARABIDOPSIS HISTIDINE KINASE 2,3,4)	Histidine kinase (single TM helix[b])
Ethylene	ETR1 (ETHYLENE RESPONSE1), ERS1 (ETHYLENE RESPONSE SENSOR 1),	Histidine kinase (3–4 TM helices)
		ETR2, EIN4, ERS2

[a]Ubiquitylation substrate (Ub-substrate)
[b]TM helix, transmembrane helix

signaling pathways of major plant hormones including auxin, JA, GA, SL, KAR, SA, ABA, and ethylene (details are reviewed in Chap. 2). A significant proportion of plant genomes is devoted to the encoding of UPS components. For example, inspection of the Arabidopsis genome suggests the presence of 500–600 F-box proteins, which is a large number when compared to the 78 F-box proteins in humans (Cardozo and Pagano 2004). The auxin and JA receptors, in fact, are the F-box proteins TIR1 and COI1, respectively. These F-box proteins possess leucine-rich repeat (LRR) domains, which recognize substrate proteins for ubiquitylation in a hormone-dependent manner (Chaps. 4, and 5). The GA, SL, and KAR receptors are members of the α/β hydrolase superfamily and act as F-box protein-bound adaptor proteins that recognize substrate proteins for ubiquitylation in a hormone-dependent manner. The GA receptor GID1 has no catalytic activity due to replacement of the catalytic His residue of the Ser-His-Asp catalytic triad with a Val/Ile residue (Chap. 6). In sharp contrast to GID1, the SL and KAR receptors possess a conserved catalytic triad system. These catalytic residues are essential for the action of SL and KAR receptors, and the SL receptor obviously exhibits catalytic activity with respect to SL hydrolysis, which is essential for SL function (Chap. 7).

Identification of the SA receptor remained elusive for quite some time. Recently, NPR (nonexpressor of PR genes) proteins have been reported to function as SA receptors in developing pathogen-induced systemic acquired resistance (SAR). One of these proteins, NPR1, was reported to function as a transcriptional co-activator (Wu et al. 2012), whereas two other proteins, NPR3 and NPR4, function as ankyrin-repeat-containing CRL3 E3 substrate receptors for ubiquitylation of NPR1 followed by degradation (Fu et al. 2012). NPR1, NPR3, and NPR4 contain a conserved BTB domain, share sequence homology, and were reported to bind SA. The precise manner by which SA binding differentiates the molecular functions of these three NPRs and regulates the interplay with each other and other transcription factors responsible for SAR-induced gene expression remains unclear, and will require structural studies of these proteins in an effort to delineate the mechanisms involved.

Unlike other plant hormone receptors that contain only a single or few members, the ABA receptor family contains multiple members (14 in Arabidopsis), and implies functional differentiation of receptor members. The ABA receptors PYR/PYLs belong to a protein family containing a steroidogenic acute regulatory protein-related lipid transfer (START) domain, and also belong to the Bet v 1 superfamily containing the birch pollen allergen Bet v 1a. The ABA receptors act as protein phosphatase inhibitors against type 2 Ser/Thr phosphatases (PP2Cs), ABI1, 2, HAB1, 2, HAI1, 2 and 3, and AHG1 (Chap. 8). This inhibition elevates the protein kinase activity of Subfamily III of Snf1-related kinases (SnRKs), which belong to the AMP-dependent protein kinase (AMPK) superfamily. The ABA signaling mediated by some ABA receptors is down-regulated by CRL4-mediated ubiquitylation and degradation (Chap. 2).

1.3 Plant Hormone Receptor Kinases

The other three plant hormones, BR, cytokinin, and ethylene, are perceived by receptor kinases located at plasma membranes. Among these, the BR receptor BRI1 is a receptor kinase possessing an LRR ectodomain, a single transmembrane (TM) helix, and a cytoplasmic domain containing an Ser/Thr protein kinase domain. The auxin and JA receptors also utilize LRR domains incorporated into the F-box proteins for hormone perception as mentioned above. LRR domains appear frequently in receptors such as peptide hormone receptors or Toll-like receptors (TLRs) in innate immune systems for the recognition of a variety of small compounds. In plants, receptor kinases form the largest family of membrane receptors and ca. 600 putative receptor kinases possessing Ser/Thr protein kinase domains have been identified in the Arabidopsis genome, although most of these receptors remain uncharacterized to date. Of these, the number of receptors that act as peptide hormone receptors is unknown at present. Considering that ca. 1000 putative small signaling molecules are encoded in the Arabidopsis genome, we expect a plethora of peptide hormone receptors to be identified (details are reviewed in Chap. 3).

The cytokinin and ethylene receptors are histidine kinases, which comprise the upstream module in the His-Asp phosphorelay (or two-component) signaling system. The downstream module of the signaling system comprises response regulators, which possess a conserved receiver domain containing an Asp residue for accepting a phosphate group. Thus, phosphoryl transfer occurs from ATP to the His residue of the histidine kinase and from the His residue to the Asp residue of the response regulator. Cytokinin and ethylene receptors are hybrid kinases that possess both histidine kinase and receiver domains. The cytokinin receptors AHK2, 3, and 4 comprise a CHASE ectodomain for cytokinin binding, a single TM helix, and a cytoplasmic hybrid kinase domain. The downstream elements of the receptor are AHP proteins with a conserved HPt domain for phosphoryl transfer (Kato et al. 1997). The ethylene receptors are divided into subfamily 1 (ETR1, ESR1) acting as His kinases with all hallmark residues conserved, and 2 (ETR2, ESR2 EI N4), which lack one or more hallmark residues and adopted Ser/Thr kinase activity. Ethylene perception is accomplished by the N-terminal copper ion-containing transmembrane domain comprised of three (subfamily 1) or four (subfamily 2) TM helices. It remains unknown how the gaseous nonpolar molecule is perceived by the protein. The copper ion likely participates in ethylene binding by mediating characteristic intermolecular interactions such as charge transfer between ion charges and π electrons of ethylene. However, no structural or biophysical information is currently available. Structural studies of the ethylene receptor remain to be the most challenging subject in structural plant biology.

1.4 Plant Hormone Perception

Hormone receptors possess two major molecular functions. One is hormone perception and the other is transfer of the hormone signal to downstream molecules. Plant hormones are bipartite molecules possessing both nonpolar hydrocarbon/aromatic

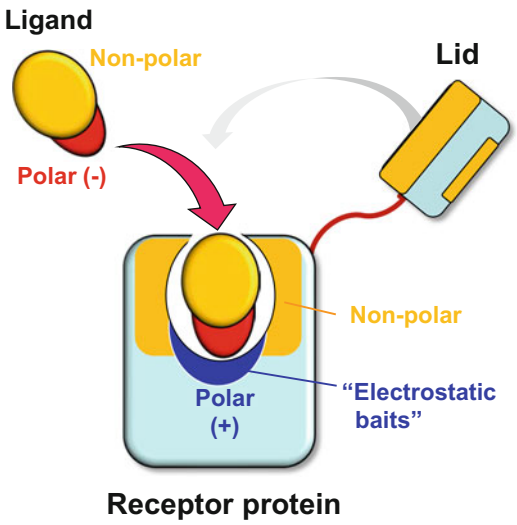


Fig. 1.2 Electrostatic guidance found in GA and other acid hormone binding

moieties and polar groups, but are mostly hydrophobic compounds exhibiting low solubility. Therefore, the hormone-binding sites of the receptors comprise primarily hydrophobic pockets with polar patches to facilitate specific interactions with the polar groups of hormones. Therefore, hormone perception inside the pockets is mediated by both nonpolar and polar intermolecular interactions, which contain hydrophobic contacts, salt bridges, and direct and water-mediated hydrogen bonds. A prominent and common polar interaction is found in the hormone-binding pockets for auxin, JA, GA, and ABA. These hormones are acids that possess one negatively charged carboxyl group. In the receptor-bound state, the carboxylic acid group is anchored to the positively charged Arg/Lys residues located at the bottom of the deep binding pockets of the receptors, and the negative charge of the hormone molecule is neutralized by the formation of salt bridges (Fig. 1.2). The anchoring of the carboxylic acid groups of plant hormone molecules by electrostatic interaction is reminiscent of the electrostatic guidance mechanism by which animal hormones enter the binding pocket of their nuclear receptors (Renaud et al. 1995). It is most likely that the carboxyl group enters the pocket first and is drawn down through the hydrophobic pocket to the anchoring site displaying an electrostatic bait. For the polar interaction, the GA receptor utilizes the region corresponding to the active site of the α/β hydrolase fold containing the oxyanion hole. SL/KAR receptors also comprise one α/β hydrolase domain with all active residues of the Ser-His-Asp triad, and the hormone molecules are hydrolyzed (Chap. 7). Compared to the deep hormone-binding pockets found in the auxin, JA, GA, ABA, and SL/KAR receptors that occlude the hormone molecules, the BR and cytokinin receptors provide surface pockets for hormone binding and receptor-bound BR and cytokinin molecules are still accessible from the solvent region.

1.5 Signal Transfer by Plant Hormone Receptors

When a hormone molecule binds the pocket of the receptor, perception of the hormone induces local and/or global structural changes in the receptor molecule. These changes switch on the interactions with the downstream effector molecule. The GA receptor GID1 displays a drastic conformational change on GA binding, referred to as the "close-the-lid" mechanism (Murase et al. 2008). In the first step of binding, the GA molecule still has an exposed hydrocarbon moiety to the solvent region since the carboxyl group of the GA molecule is placed first inside the deep binding pocket by the "electrostatic guidance" mechanism. This binding state induces a conformational change in the N-terminal switch region, which is conformationally flexible and forms no stable structure in the GA-free state. On GA binding, the N-terminal switch region is folded into three α-helices to form the lid of the binding pocket so as to contact the hydrocarbon moiety of GA and completely occlude the GA molecule (Fig. 1.3). The GA-induced lid also provides an active molecular surface containing nonpolar residues to recognize the effector proteins, DELLA proteins, which are the ubiquitylation substrates. ABA receptors possess a mechanism referred to as the "gate-latch-lock" mechanism with ABA as an allosteric inducer mechanism (Melcher et al. 2009). In this case, two loops, a gate and a latch, located at the entrance of the binding pocket, undergo marked conformational changes on ABA binding to form an interaction platform to trap the effector proteins, PP2C phosphatases. The auxin receptor TIR1 shows another mechanism by which the receptor, an F-box protein, recognizes its effector proteins, substrate proteins for ubiquitylation by the "molecular glue" mechanism in a hormone-

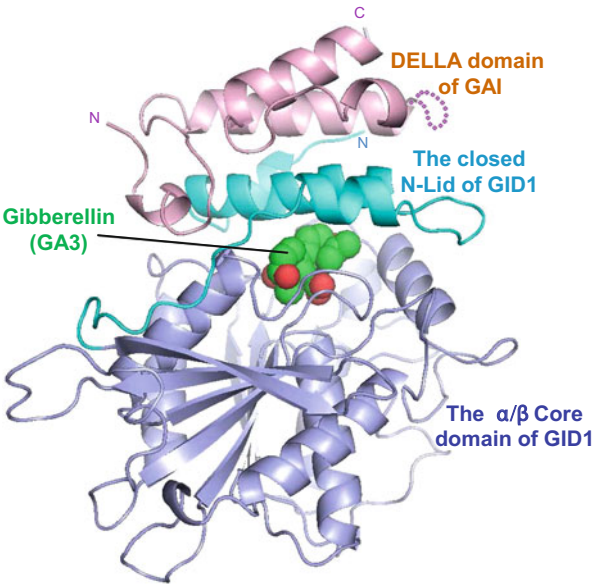


Fig. 1.3 The "closed-the-lid" mechanism found in the GA3-GID1-DELLA complex. The lid also provides binding site for DELLA protein

dependent manner (Chap. 4) (Tan et al. 2007). In this mechanism, the receptor-bound auxin molecule interacts directly with the ubiquitylation substrate protein bound to the auxin receptor. This binding mode shows that the bound substrate protein plays a role in the lid of the auxin-binding pocket. A similar mechanism is also seen in the JA receptors. Intriguingly, the BR receptor kinase and certain peptide receptor kinases adopt the "molecular glue" mechanism for co-receptor binding. BR possibly also cytokinin binding to receptor kinases defines the configuration of the receptor kinase dimer so as to activate the kinase domain located at the cytoplasm via transmembrane helices. As with animal receptor kinases, the nature of the conformational changes that take place and are essential for kinase activation remains obscure, and structural studies of receptor kinases as membrane proteins represent challenging projects in the future.

References

Cardozo T, Pagano MM (2004) The SCF Ubiquitin ligase: insights into a molecular machine. Nat Rev Mol Cell Biol 5:739–751

Chow B, McCourt P (2006) Plant hormone receptors: perception is everything. Genes Dev 20:1998–2008

Evans RM, Mangelsdorf DJ (2014) Nuclear receptors, RXR, and the Big Bang. Cell 157:255–266

Fu ZQ, Yan S, Saleh A, Wang W, Ruble J, Oka N, Mohan R, Spoel SH, Tada Y, Zheng N, Dong X (2012) NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. Nature 486:228–232

Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pages V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC, Bouwmeester H, Bécard G, Beveridge CA, Rameau C, Rochange SF (2008) Strigolactone inhibition of shoot branching. Nature 455:189–194

Grienenberger E, Fletcher JC (2015) Polypeptide signaling molecules in plant development. Curr Opin Plant Biol 23:8–14

Kato M, Mizuno T, Shimizu T, Hakoshima T (1997) Insights into multistep phosphorelay from the crystal structure of the C-terminal HPt domain of ArcB. Cell 88:717–723

Melcher K, Ng LM, Zhou XE, Soon FF, Xu Y, Suino-Powell KM, Park SY, Weiner JJ, Fujii H, Chinnusamy V, Kovach A, Li J, Wang Y, Li J, Peterson FC, Jensen DR, Yong EL, Volkman BF, Cutler SR, Zhu JK, Xu HE (2009) A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. Nature 462:602–608

Murase K, Hirano Y, Sun TP, Hakoshima T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. Nature 456:459–463

Renaud JP, Rochel N, Ruff M, Vivat V, Chambon P, Gronemeyer H, Moras D (1995) Crystal structure of the RAR-γ ligand-binding domain bound to all-trans retinoic acid. Nature 378:681–689

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446:640–645

Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, Magome H, Kamiya Y, Shirasu K, Yoneyama K, Kyozuka J, Yamaguchi S (2008) Inhibition of shoot branching by new terpenoid plant hormones. Nature 455:195–200

Wu Y, Zhang D, Chu JY, Boyle P, Wang Y, Brindle ID, De Luca V, Després C (2012) The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. Cell Rep 1:639–647

Chapter 2
Overview of Protein Degradation in Plant Hormone Signaling

Domnita V. Rusnac and Ning Zheng

2.1 Introduction to the Ubiquitin-Proteasome System

Protein degradation is a proteolytic process, which counteracts protein synthesis and determines the half-lives of all proteins in the cell. Although some proteins can be extremely long lived, the majority of cellular proteins has a measurable half-life, ranging from minutes to days (Toyama and Hetzer 2013; Hershko and Ciechanover 1998). Early studies of protein breakdown in animals and plants emphasized on its roles in protein quality control and amino acid reutilization, which help eukaryotic cells to cope with cellular and environmental stress as well as nutrient starvation. Recent advances, however, have unraveled an unexpected regulatory function of protein degradation in actively controlling the abundance of a variety of intracellular proteins, thereby, modulating their activities (Hershko and Ciechanover 1998).

The ubiquitin-proteasome system (UPS) is the central pathway for intracellular protein degradation and is evolutionarily conserved in all eukaryotes, including plants (Vierstra 2009; Callis 2014). In an ATP-dependent manner, the UPS is programmed to respond to diverse cellular cues and selectively label target proteins for rapid breakdown. Thanks to the groundbreaking work by Avram Hershko, Aaron Ciechanover, Irwin Rose, Alfred Goldberg, Alexander Varshavsky, and many other pioneers in the field, most of the key components of the UPS have now been identified and biochemically characterized in great details (Wilkinson 2005). Our mechanistic understanding of the UPS function has also benefited tremendously from the extensive structural studies in the past two decades. This chapter offers a brief overview of the UPS and its major constituents in eukaryotes and highlights its unique involvements in various hormonal signaling pathways in plants.

D. V. Rusnac · N. Zheng (✉)
Department of Pharmacology and Howard Hughes Medical Institute, University of Washington, Box 357280, Seattle, WA, USA
e-mail: nzheng@uw.edu

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_2

2.1.1 Ubiquitin and Ubiquitin-Like Proteins

Ubiquitin is a 76-amino acid protein universally found in all eukaryotic species and broadly expressed in different tissues of animals and plants. It has a highly conserved polypeptide sequence, which differs by three amino acids between the yeast and human orthologues. Ubiquitin is characterized by a compact β-grasp fold and a flexible C-terminal tail terminated by a di-Gly motif after maturation (Fig. 2.1a). In the UPS, ubiquitin serves as a protein posttranslational modifier, whose C-terminal carboxyl group is covalently conjugated to the ε-amino group of a substrate lysine residue via an isopeptide bond. As ubiquitin itself also has seven lysine residues, polyubiquitin chains can be formed when the carboxyl terminus of one ubiquitin molecule is linked to a lysine residue of another copy (Fig. 2.1a, b). Depending on which ubiquitin lysine residue is involved in chain elongation, polyubiquitin chains can be built with different linkages either in a homogeneous or branched fashion (Komander and Rape 2012; Meyer and Rape 2014). Among different types of ubiquitin chains, the Lys-48-linked tetraubiquitin chain has long been established as the minimal signal for proteasome targeting (Fig. 2.1b) (Thrower et al. 2000).

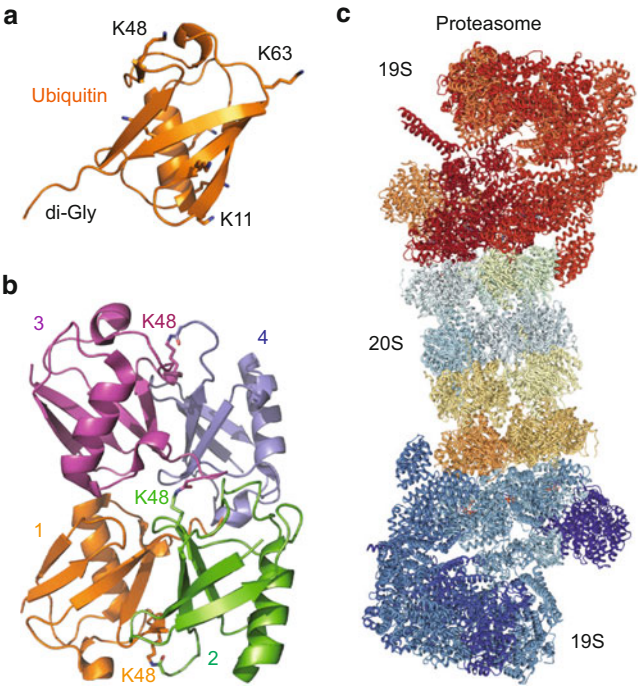


Fig. 2.1 (**a**) Ubiquitin with seven lysine residues (sticks) and a C-terminal di-Gly motif. Three commonly modified ubiquitin lysine residues are labeled. (**b**) Lys-48 linked tetraubiquitin chain (PDB:2O6V). (**c**) 26S proteasome with the 20S core particle and the 19S regulatory particle (PDB:5GJR)

In most, if not all, eukaryotic organisms, several proteins have been found to share sequence homology with ubiquitin and adopt the same ubiquitin fold, but do not serve as proteasome-targeting signal. These ubiquitin-like proteins (UBLs), such as NEDD8 and SUMO, also feature a C-terminal di-Gly motif after precursor processing and function as protein modifiers in regulating diverse cellular processes, including the UPS (van der Veen and Ploegh 2012). In most cases, these UBLs modify the substrate proteins in a monomeric form and elicit their effects by altering the structural topology, protein network, or cellular localization of the targets.

2.1.2 Proteasome as a Protein Degradation Machinery

The 26S proteasome is an intracellular multi-subunit proteolytic machinery, which is localized in both cytosolic and nuclear compartments and acts as the most downstream component of the UPS (Coux et al. 1996). Due to its protein destruction nature, the 26S proteasome has been evolved to safeguard its proteolytic activity at both architectural and functional levels (Tomko and Hochstrasser 2013). To achieve tight regulation of its protease function, the 26S proteasome is composed of two parts, the 20S core particle, which carries the catalytic activities, and the 19S regulatory particle, which controls the access of the active sites hidden inside the enzymatic core (Fig. 2.1c). Crystal structures of the 20S core particle revealed a cylindrical architecture, which consists of four stacked rings sequestering a central pore (Kish-Trier and Hill 2013). The inner two rings are each constructed by seven β-subunits, harboring three peptidase activities with the catalytic sites buried in the interior cavity, whereas the outer two rings are each formed by seven α-subunits, whose N-terminal regions converge at the center and together close up the proteolytic chamber of the core particle.

By docking to the outer rings of the 20S particle, the 19S regulatory particle of the proteasome is engaged with the proteasome core on its two ends and feed the degradation machinery with only polyubiquitinated protein substrates. Distinct from the 20S particle, the 19S particle has a highly asymmetric structure, which has historically been divided into two sub-complexes, the lid and the base (Lander et al. 2012). The base of the 19S particle contains six different ATPase subunits, which are assembled into a trimer-of-dimers ring-shaped structure. In addition, it also features three non-ATPase subunits, which have ubiquitin receptor functions. Together, these 19S base subunits are responsible for recognizing polyubiquitinated substrate, opening the gate of the 20S core, unfolding the folded substrate protein, and translocating the linearized polypeptide into the proteolytic chamber. The 19S lid complex, which consists of ten subunits, partially covers the base ATPases and makes direct contacts with the 20S core. Besides contributing to ubiquitin recognition, one important function of the 19S lid is to catalyze the removal of ubiquitin from the substrate before it is fed to the protease core. Recent advances in cryoelectron microscopy (cryo-EM) have not only allowed near atomic resolution structural determination of the entire proteasome, including the 19S regulatory

particle, but also helped reveal the protein degradation machinery in different functional states with substrate and/or nucleotides bound (Bhattacharyya et al. 2014).

2.1.3 The E1-E2-E3 Enzyme Cascade

Ubiquitin conjugation to a protein substrate, a process referred to as ubiquitination (or ubiquitylation), is the hallmark of ubiquitin-dependent protein degradation. Protein ubiquitination is catalyzed by the sequential actions of three enzymes, the E1 ubiquitin-activating enzyme, the E2 ubiquitin-conjugating enzyme, and the E3 ubiquitin-protein ligases (Pickart 2001) (Fig. 2.2a). Free ubiquitin is first activated by the E1 enzyme, which uses ATP-Mg^{2+} to catalyze the acyl adenylation of ubiquitin's C-terminal carboxyl group and then captures the activated ubiquitin tail with its catalytic cysteine via a thiolester bond. Upon binding to a ubiquitin-conjugating enzyme, the ubiquitin-activating E1 enzyme subsequently transfers ubiquitin to the active site cysteine residue on E2 through a trans-thiolesterification reaction. As a highly active enzyme, E1 is responsible for constitutively charging E2 enzymes with ubiquitin in the cell. Vertebrates have two E1 genes, whose protein products, known as UBE1 and UBA6, have been found to preferentially charge different E2s (Jin et al. 2007). In *Arabidopsis thaliana*, two ubiquitin E1 enzymes, UBA1 and UBA2, have also been identified with nonredundant functions (Goritschnig et al. 2007).

In contrast to the small number of E1 enzymes, the ubiquitin-conjugating E2 enzymes are numbered in 30–40 in higher eukaryotes and often act in different cellular pathways (Wenzel et al. 2011b). All E2 enzymes share a conserved ~150-amino acid catalytic core domain, which adopts a classic UBC fold with the active site cysteine tucked in a cleft between two loops. Certain E2s feature additional N-terminal or C-terminal extension sequences, whereas a specific subgroup of E2s contains an internal acid loop close to the active site cysteine. Although E2s were once thought to be simple ubiquitin "carriers," recent studies have shown that they display distinct intrinsic reactivity and often play a critical role in dictating the linkage specificity of a polyubiquitin chain (Stewart et al. 2016). Because many ubiquitin-charged E2s (Ub~E2s) selectively interact and function with specific types of ubiquitin E3 ligases (see below), their active sites can have characteristic reactivity toward different attacking groups, such as the ε-amino group of the lysine side chain and the thiol group of a cysteine residue. Furthermore, with the help of extra sequence elements or binding partners, some E2s can recognize a specific lysine residue on the receiver (proximal) ubiquitin, which accepts the C-terminus of the incoming donor (distal) ubiquitin during chain extension. Interestingly, some E2 variants, which lack the active site cysteine, have been shown to interact with a canonical E2 to confer linkage-specific polyubiquitin chain activities.

Although the thiolester bond in the Ub~E2 conjugate is less stable than the isopeptide bond linking ubiquitin and substrate, transfer of ubiquitin from an E2 to

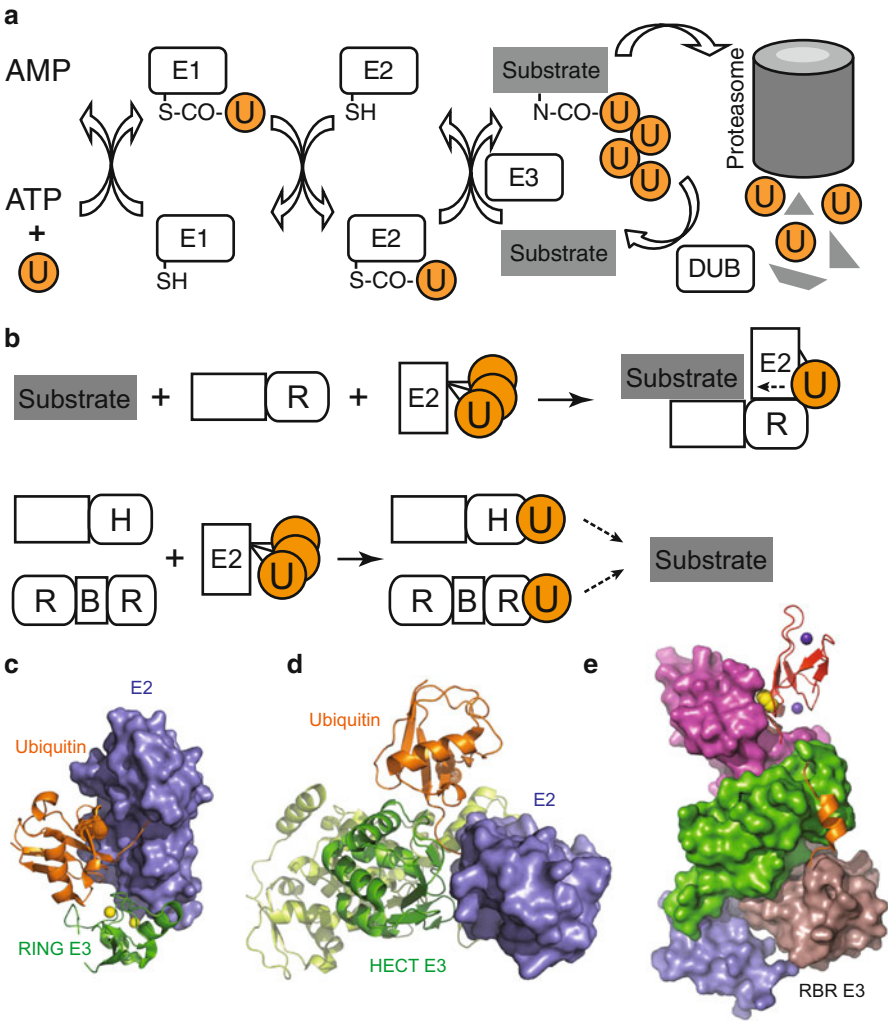


Fig. 2.2 (**a**) Ubiquitin-proteasome system with the E1-E2-E3 enzyme cascade acting upstream of the proteasome and the counteracting deubiquitinases (DUBs). (**b**) The three types of ubiquitin E3 ligases (R, RING; H, HECT; and RBR) and their different ubiquitin transfer mechanisms. (**c**) A RING E3-Ub~E2 complex (PDB:4AP4). (**d**) A HECT E3-Ub~E2 complex (PDB:3JW0). (**e**) An RBR E3 (PDB:4 K95)

a substrate does not occur efficiently until an E3 ubiquitin ligase is present (Pickart 2001). In the three-enzyme cascade, the E3 enzyme performs two critical functions to facilitate substrate ubiquitination. First, E3s stimulate the reactivity of a ubiquitin-charged E2 to accelerate ubiquitin discharge. Second, E3s provide a platform, onto which a specific protein substrate and the ubiquitin-charged E2 are recruited and brought together in close proximity. Ubiquitin E3 ligases, therefore, represent an

ideal class of enzymes favored by evolution for adopting novel functions that can couple protein ubiquitination and degradation with various upstream signals in diverse cellular pathways.

2.1.4 Three Types of Ubiquitin E3 Ligases

The functional importance and versatility of ubiquitin ligases in the UPS are best manifested by the different types of E3s and their sheer number in the eukaryotic genomes in comparison to other UPS enzymes. In *Arabidopsis thaliana*, more than one thousand genes have been identified to encode putative ubiquitin ligases (Vierstra 2009). Although this number varies among other plant species, the prevalence of E3s and their roles in regulating plant physiology are obvious. Intriguingly, plant pathogens are known to produce effector proteins that either mimic or hijack E3 ligases to take advantage of the host UPS and benefit their infection and life cycle (Banfield 2015). Such cross-kingdom functions further highlight the central roles played by ubiquitin ligases in the cell.

In all eukaryotes, three types of E3s have been identified, which are grouped based on their different signature sequence motifs and distinct catalytic mechanisms (Fig. 2.2b). The Really Interesting New Gene (RING) domain defines the largest family of ubiquitin ligases, known as RING-type E3s, which share a common protein fold consisting of two zinc-binding fingers with eight zinc-coordinating cysteine and histidine residues (Deshaies and Joazeiro 2009) (Fig. 2.2c). Besides the RING domain, these E3 ligases either contain a substrate-binding domain in the same polypeptide or belong to a multi-subunit ubiquitin ligase complex, which uses another subunit for recruiting substrate. The RING-type E3s are distinguished from other E3s by catalyzing the direct transfer of ubiquitin from an E2 to the subunit. Recent structural studies have shown that, upon binding to a ubiquitin-charged E2 enzyme, the RING domain makes contacts with both the E2 and the donor ubiquitin and stabilizes the Ub~E2 conjugate in a "closed" conformation (Fig. 2.2c) (Plechanovová et al. 2011; Dou et al. 2012; Pruneda et al. 2012). In doing so, a RING E3 activates the ubiquitin-charged E2 for ubiquitin transfer by presumably optimizing the geometry of the E2 active site for the nucleophilic attack by the side chain of a lysine residue in either a substrate or a receiver ubiquitin molecule.

The homology to E6AP carboxyl terminus (HECT) type of E3s represents a second family of ubiquitin ligases, which are characterized by their common C-terminal catalytic domain, known as HECT domain (Rotin and Kumar 2009) (Fig. 2.2d). With a bilobal structure, the HECT domain harbors an active site cysteine, which forms an obligate thiolester intermediate with ubiquitin to promote substrate ubiquitination (Fig. 2.2b) (Huang et al. 1999; Metzger et al. 2012). The first step of ubiquitin transfer mediated by the HECT E3s involves a trans-thiolesterification reaction, in which ubiquitin is passed from the active site cysteine of the E2 to that of the E3. Due to the nature of this specific reaction, HECT E3s only function with a small subset of E2 enzymes. Akin to single polypeptide RING E3s,

most known HECT E3s recognize their specific substrate through regions outside their catalytic domain. Although the human genome encodes nearly 30 HECT E3s, this family of ubiquitin ligases remains relatively small in plants (Marín 2013).

Remarkably, recent studies have unveiled a third family of E3s, which is named RING-in-Between-rings-RING (RBR) E3s (Spratt et al. 2014; Wenzel et al. 2011a). Despite the presence of several zinc finger-containing RING-like domains, RBR E3s are mechanistically closer to the HECT E3s than the RING E3s. While the RING1 domain of RBR E3s is responsible for recruiting a ubiquitin-charged E2 enzyme, ligation of ubiquitin to the substrate involves the formation of a ubiquitin~E3 intermediate, which is anchored at a strictly conserved catalytic cysteine found in the RING2 domain of the E3s (Fig. 2.2b, e). Similar to the HECT E3s, the RBR E3s relay ubiquitin to the substrate and display strong E2 preferences. Recent structural analyses of several RBR E3s have revealed that these multi-domain ubiquitin ligases almost exclusively adopt an auto-inhibited conformation in isolation (Trempe et al. 2013; Wauer and Komander 2013; Stieglitz et al. 2013; Lechtenberg et al. 2016) (Fig. 2.2e). Activation of these enzymes might be achieved by posttranslational modifications of the E3s or upon interactions with their binding partners, which presumably recruits specific substrates. So far, RBR E3s have been poorly studied in plants (Marín 2010). However, the potential functional connections of a RBR subfamily, Ariadne/HHARI, with the superfamily of cullin-RING E3s, as suggested by recent studies, might implicate a prominent role of the RBR E3 in plant hormone signaling (see below) (Scott et al. 2016).

2.1.5 Deubiquitinases

In the same way as most protein posttranslational modifications, protein ubiquitination is reversible, and the activities of ubiquitin ligases can be counterbalanced by enzymes capable of cleaving ubiquitin-linked isopeptide bonds (Fig. 2.2a). These isopeptidases, also known as deubiquitinases (DUBs), can either trim various ubiquitin chains with specific linkages or catalyze the removal of ubiquitin from substrate (Komander et al. 2009). Their activities not only enable ubiquitin recycling prior to substrate degradation by the proteasome but also provide a mechanism for regulating protein ubiquitination in a dynamic manner. In animals, DUBs are classified into six different subfamilies (USPs, UCHs, OTUs, MJDs, JAMM, and MINDYs) based on their sequence homology. The same six DUB families are also found in plants with a total of ~60 different family members in the *Arabidopsis* genome. Although little is known about their functions, it is expected that their deubiquitinase activities might be involved in fine-tuning the ubiquitination and degradation of many substrate polypeptides, including those implicated in hormone signaling.

2.2 Cullin-RING Ubiquitin Ligases

The cullin-RING ubiquitin ligase (CRL) complexes represent the largest family of multi-subunit E3s in all eukaryotic species (Deshaies 1999; Zhao et al. 2003; Zimmerman et al. 2010). In animals, these E3 machineries regulate diverse cellular functions, such as signal transduction, cell cycle progression, metabolic processes, DNA repair and replication, circadian clock, and stress responses. In plants, they not only participate in many of these functions that are conserved in all eukaryotic cells but also perform numerous plant-specific tasks, particularly, in perceiving and transducing phytohormone signals (Hua and Vierstra 2011). A rapid growing number of studies have now revealed the central roles CRLs in the signaling pathways of the vast majority of known plant hormones, including auxin, jasmonate (JA), gibberellin (GA), strigolactone, salicylic acid (SA), abscisic acid (ABA), and ethylene.

The CRL E3 machineries are built in a modular fashion, in which a common catalytic platform is used by numerous interchangeable substrate receptor subunits for ubiquitinating specific substrates (Li et al. 2003) (Fig. 2.3a). The CRL catalytic platform is formed between a ~80 kDa cullin scaffold protein and a ~20 kDa RING domain protein, RBX1. The cullin scaffold adopts an elongated overall structure with a more globular C-terminal domain (CTD) interacting with Rbx1 via an intermolecular β-sheet (Fig. 2.3b). The CRL substrate receptor subunits assemble with the catalytic platform by either directly interacting with the N-terminal

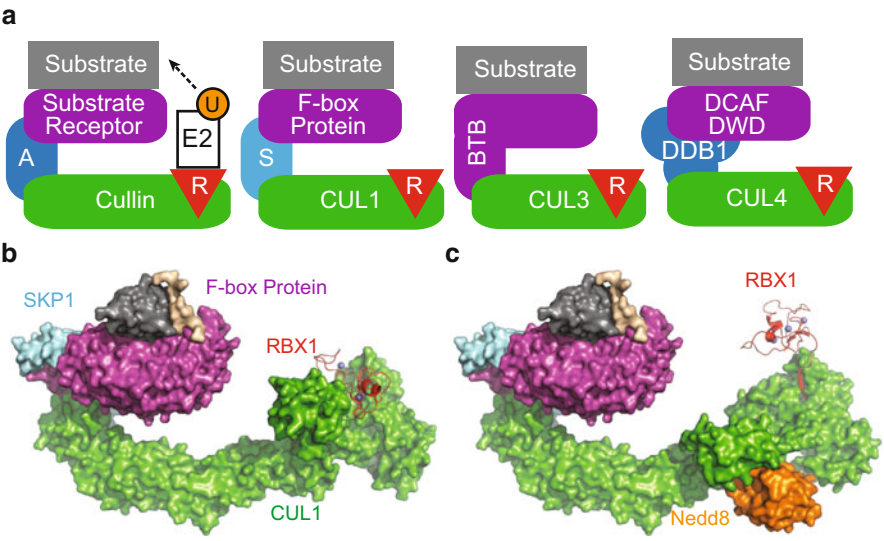


Fig. 2.3 (**a**) The assembly logic of cullin-RING E3 complexes and the three types of CRLs in plants, SCF/CRL1 (S, SKP1; R, RBX1), CRL3, and CRL4. (**b**) Structural model of a substrate-loaded SCF/CRL1 complex. (**c**) Structural model of a substrate-loaded NEDD8-modified SCF/CRL1 complex

domain (NTD) of the cullin scaffold or indirectly through an adaptor protein. When a substrate is presented by the substrate receptor subunit, CRLs promote sequential addition of ubiquitin to the substrate by successively engaging and activating a ubiquitin-charged E2 enzyme. Humans have six closely related cullin proteins (CUL1, 2, 3, 4A, 4B, and 5), which organize five different types of CRLs. By contrast, plants have only three cullin proteins (CUL1, CUL3, and CUL4), which give rise to three major CRL E3s, CRL1, CRL3, and CRL4 (Fig. 2.3a).

2.2.1 An Introduction to SCF/CRL1 E3s

The SCF (SKP1-CUL1-F-box proteins)/CRL1 E3 is the prototype of all CRLs and is composed of the CUL1-RBX1 catalytic core, the cullin adaptor SKP1, and members of the F-box protein family, which serve as the interchangeable substrate receptors (Deshaies 1999). Through a conserved ~40-amino acid F-box motif, F-box proteins interact with the relatively abundant SKP1 adaptor to form a battery of stable SCF substrate-binding modules. In addition to the F-box motif, the F-box proteins contain various protein-protein interaction domains, such as leucine-rich repeats (LRRs) domain or WD40-repeat domain, for binding specific substrates. Early studies of the SCF/CRL1 E3s in fungi and animals showed that F-box proteins often recognize their cognate substrates through a short linear sequence motif, known as degron (Zhao et al. 2003; Skaar et al. 2013). Their high affinity interaction frequently requires phosphorylation of one or two amino acids in the substrate degron by an upstream kinase, a prerequisite effectively linking protein ubiquitination to kinase signaling. Recent studies have shown that certain SCF substrates can not only bypass this requirement but also employ its entire protein fold to interface with the F-box protein (Xing et al. 2013).

In the past decade, research in plant hormone signaling has unraveled a stunning new paradigm of SCF/CRL1 functions, which places the E3 complexes as the central components in the signaling pathways of several key phytohormones (Fig. 2.4). These studies, together with high-resolution structural analyses, help establish novel mechanisms by which the SCF/CRL1 E3s bridge hormone sensing and transcriptional reprogramming via protein ubiquitination and degradation. In the auxin and JA signaling pathways, the F-box proteins, TIR1 and COI1, have been identified as the long-sought hormone receptors, which directly perceive the two phytohormones with a ligand-binding pocket constructed by their LRR domains (Dharmasiri et al. 2005; Kepinski and Leyser 2005; Xie et al. 1998; Tan et al. 2007; Sheard et al. 2010). In a hormone-dependent manner, the two F-box proteins interact with the degron sequences of AUX/IAAs and JAZs, respectively, and promote their polyubiquitination and proteasome-mediated degradation (Fig. 2.4a, b). AUX/IAA and JAZ proteins are transcription repressors in the auxin and JA signaling pathways. Their rapid hormone-triggered degradation alleviates their inhibitory effects and activates the gene expression governed by IAA and MYC2 transcription factors. Interestingly, crystallographic studies of the two systems have revealed a “molecular

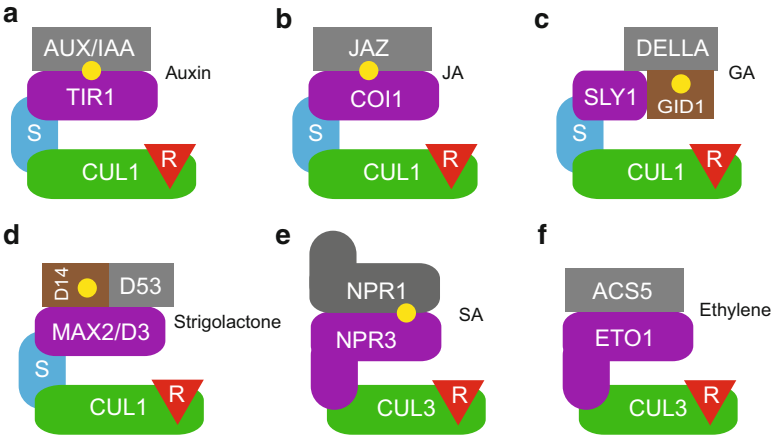


Fig. 2.4 (**a**) SCF^{TIR1} in complex with auxin and AUX/IAA. (**b**) SCF^{COI1} in complex with JA and JAZ. (**c**) $SCF^{SLY1\text{-}GID2}$ in complex with GA and DELLA. (**d**) $SCF^{MAX2\text{-}D14}$ in complex with strigolactone. (**e**) $CRL3^{NPH4}$ in complex with SA and NPR1. (**f**) $CRL3^{ETO1}$ in complex with ACS5. Yellow dots indicate hormone molecules

glue" mechanism by which each of the two hormones enhances the interactions between the F-box protein and its substrates by filling up a gap at their protein-protein interaction interface. The structural studies have also revealed the binding of specific inositol polyphosphate molecules to the two F-box proteins underneath their hormone-binding pockets. These soluble inositol polyphosphates could serve as a second signal for phosphate availability that is recognized by the F-box proteins. The details of the two systems are elaborated in the following chapters of this book.

Remarkably, the plant SCF/CRL1 E3s also play a major role in GA and strigolactone signaling (Shabek and Zheng 2014). These two phytohormones regulate a myriad of plant developmental processes, such as stem elongation (GA), flowering (GA), shoot branching (strigolactone), and symbiotic interactions with arbuscular mycorrhizal fungi (strigolactone). Similar to auxin and JA, GA and strigolactone are perceived by SCF complexes, which feature SLY1/GID2 and MAX2/D3 as the F-box proteins (Fig. 2.4c, d). Distinct from TIR1 and COI1, SLY1/GID2 and MAX2/D3 do not physically recognize the hormone molecules. Instead, these two F-box proteins interact with a cognate α/β hydrolase, GID1 and D14, respectively, which harbors a deep pocket for sensing the hormonal signal. With a nonproductive catalytic triad, GID1 loses its catalytic activity as a hydrolase. GA binding induces major conformational remodeling of the enzyme and enables it to interact with the DELLA transcription regulators. By promoting the ubiquitination and degradation of DELLA proteins, the GA-bound $SCF^{SLY1/GID2\text{-}GID1}$ E3 complex relays the hormonal signal to alter downstream gene expression (Murase et al. 2008; Shimada et al. 2008). Intriguingly, the strigolactone sensor protein D14 retains its hydrolase activity and is capable of slowly hydrolyzing the hormone as a substrate (Nakamura et al. 2013; Hamiaux et al. 2012). Upon binding to the F-box protein

MAX2/D3, D14 undergoes profound conformational changes and blocks the release of the hydrolysis product (Yao et al. 2016). Through its interaction with MAX2/D3, D14 is thought to recruit a class I Clp ATPase protein, D53, and promotes its degradation to inhibit axillary bud outgrowth and branching (Jiang et al. 2013; Zhou et al. 2013).

Besides acting as the sensors of plant hormones, SCF E3s also participate in hormonal signaling by regulating the abundance of key components of the pathways. In ethylene signaling, which dictates fruit ripening and several plant developmental and stress response processes, the F-box proteins EBF1/2 have been shown to destabilize EIN3, which is the transcription factor controlling ethylene responses (Guo and Ecker 2003; Potuschak et al. 2003). Interestingly, EIN3-EBF1/2 interactions can be negatively regulated by ethylene and positively promoted by red light through the action of photoreceptor phyB (Shi et al. 2016; Qiao et al. 2009). Furthermore, the F-box proteins ETP1/2 have been recently identified to promote the proteasomal degradation of EIN2, which is a transmembrane protein acting downstream of the ethylene receptors and upstream of EIN3. Additional examples of plant F-box proteins involved in hormonal signaling include the KISS ME DEADLY (KMD) proteins, which ubiquitinate type-B *Arabidopsis* response regulator (ARR) 2 in the cytokinin signaling pathway, and RCAR3 INTERACTING PROTEIN 1 (RIFP1), which negatively regulates the abscisic acid (ABA) receptor RCAR3 (Kim et al. 2013; Li et al. 2016).

2.2.2 An Introduction to CRL3-BTB E3s

The CRL3 E3s constitute a second large family of plant CRLs (Hua and Vierstra 2011). They differ from SCF/CRL1 complexes by having their cullin adaptor and substrate receptor functions combined into a single polypeptide, named Bric-a-brac, Tramtrack, Broad-complex (BTB)-domain protein. Members of this protein family share a conserved BTB domain that adopts a SKP1-like fold and directly binds CUL3. BTB-domain proteins have also evolved several other substrate-binding domains with characteristic protein folds and sequences such as ankyrin repeats, KELCH, MATH, and ZnF domain. The NPR (nonexpressor of PR genes) proteins constitute one small cohort of ankyrin-repeat-containing CRL3 E3 substrate receptors, which are particularly relevant to hormonal signaling. These proteins have been reported to function as the missing receptors for salicylic acid (SA) in developing pathogen-induced systemic acquired resistance (SAR) in plants (Fu et al. 2012). When a host is under the attack of a microbial pathogen, the cells around the local infection site produce SA, which will dissipate throughout the organism inducing the expression of a wide range of pathogenesis-related proteins (Fu and Dong 2013). NPR3 and NPR4 have recently been shown to directly bind SA and function as CUL3 adaptors to mediate NPR1 degradation (Fig. 2.4e). With different binding affinities toward SA, NPR3 and NPR4 are thought to enable differential responses to the variable cellular SA concentrations in the infected and neighboring cells (Fu et al.

2012). As a master regulator of SAR, NPR1 itself is a BTB-domain protein, which shares sequence homology with NPR3 and NPR4. It interacts with several transcription regulators. Although NPR1 has also been documented to bind SA, how it interplays with NPR3/4 and other transcription factors responsible for SAR-induced gene expression remains unclear (Wu et al. 2012).

Besides NPRs, several other BTB-domain proteins have been implicated in regulating plant hormone functions. Through its interaction with ACS5, a 1-aminocyclopropane-1-carboxylic acid synthase responsible for synthesizing ethylene, the BTB-domain protein ETO1 has been reported to block ethylene biosynthesis by both inhibiting the enzymatic activity of ACS5 and promoting its proteasome-dependent degradation (Fig. 2.4f) (Wang et al. 2004). NAKED PINS IN YUC MUTANTS (NPY) is another BTB-domain protein, which forms a complex with the PINOID kinase and regulates auxin-mediated organogenesis downstream of the auxin efflux carriers, PINs (Strader and Zhao 2016). Last but not the least, a BTB-domain protein, BT2, has been suggested to antagonize ABA signals and enhance certain auxin responses (Mandadi et al. 2009). The BTB protein family has at least 80 members in the *Arabidopsis* genome, and the number is nearly doubled in rice due to a major subfamily expansion (Gingerich et al. 2007). Given the importance of phytohormone in regulating plant physiology, it is highly likely that more BTB-domain proteins are involved in mediating hormonal signaling.

2.2.3 An Introduction to CRL4-DDB1-DCAF/DWD E3s

The CLR4 E3 complexes represent a unique family of CRLs specialized for ubiquitinating substrates mostly in the nucleus (Li et al. 2003). Distinct from other cullins, CUL4 interacts with a large adaptor protein, DDB1, which consists of three β-propeller domains and assembles with a family of WD40-repeat-containing proteins, DCAF/DWDs, as the substrate receptors (Angers et al. 2006; Li et al. 2006; Jin et al. 2006; He et al. 2006). To date, four DCAF/DWD proteins, DWA1, DWA2, DWA3, and ABD1, have been functionally characterized in hormone signaling (Lee et al. 2010; Lee et al. 2011; Seo et al. 2014). All of them negatively regulate ABA signal transduction by recruiting the ABA-responsive transcription factor, ABI5, to the CRL4 E3 for ubiquitination and proteasomal degradation. Interestingly, DDA1, a small noncanonical DCAF/DWD protein lacking a WD40-repeat domain, has also been documented to downregulate ABA signaling. DDA1 interacts with the ABA receptors, PYL8, PYL4, and PYL9, and induces their CRL4-mediated ubiquitination (Irigoyen et al. 2014). With ~80 family members in *Arabidopsis thaliana* and rice, the CRL4 E3s most likely play additional roles in regulating hormone responses beyond the ABA pathway (Hua and Vierstra 2011).

2.2.4 Regulation of CRL E3s by Nedd8 and COP9

AXR1 is one of the first few genes identified in the genetic screens of auxin-resistant mutants (Leyser et al. 1993). With high sequence similarity to the E1 ubiquitin-activating enzyme, AXR1 was later characterized as the E1 enzyme for the UBL protein, NEDD8. Analogous to the activation of kinases by phosphorylation, all CRL E3s are activated by the covalent modification of cullin scaffolds by NEDD8 (also known as RUB1) at a specific lysine residue next to the RBX1-binding site (Deshaies and Joazeiro 2009). Crystal structures of neddylated cullins have revealed an open topology of the catalytic platform, in which the E2-binding RING domain of RBX1 is released from a cullin CTD cleft, potentially allowing the ubiquitin-charged E2 to approach the substrate recruited by the substrate receptor subunit anchored at the cullin NTD (Duda et al. 2008) (Fig. 2.3c).

In addition to enhancing the ligase activities, NEDD8 modification has been suggested to play a role in facilitating the exchange of substrate-binding modules on the shared catalytic platform of CRL E3s (Lydeard et al. 2013). In eukaryotic cells, a 120 kDa HEAT-repeat protein, CAND1, has been identified to bind the unmodified cullin-RBX1 complexes and block the assembly of substrate-binding modules, such as the SKP1-F-box protein complexes and BTB-domain proteins (Goldenberg et al. 2004). Because NEDD8 modification of cullins inhibits CAND1 association, dynamic neddylation and deneddylation are thought to mediate the redistribution and recycling of the CRL catalytic platform among different substrate-recruiting modules (Pierce et al. 2013).

To cleave NEDD8 from cullins, eukaryotic cells have evolved a conserved eight-subunit protein complex, the COP9 signalosome or CSN, which can interact with all cullin-RBX1 complexes (Wei and Deng 2003; Lyapina et al. 2001). Among the eight CSN subunits, CSN5 is the JAMM-type isopeptidase responsible for catalyzing the cullin deneddylation reaction (Cope et al. 2002). Interestingly, each of the CSN subunits shares sequence homology to one of the subunits of the 19S proteasome lid, suggesting that the two complexes have evolutionary, and possibly functional, connections (Wei et al. 1998). Just like the NEDD8 E1 enzyme, CSN was first identified in plants (Chamovitz et al. 1996). Mutations of each CSN subunit led to the same constitutive photomorphogenesis phenotype. Interestingly, the same phenotype is shared by mutants of three other genes—DET1, a DCAF/DWD protein; COP10, a DET1-DDB1-interacting E2 variant; and COP1, which is a RING E3 that also interacts with DET1 (Lau and Deng 2012; Chen et al. 2006; Wertz et al. 2004). Together, these proteins and CSN might act in the same pathway to promote the ubiquitination and degradation of HY5, a key transcription factor regulating photomorphogenic development. Besides mediating light signaling, CSN also participates in hormonal signal transduction by controlling the assembly of several aforementioned CRL complexes.

2.3 Other E3s Involved in Hormone Signaling

Despite the apparent predominance of CRL E3s in sensing and transducing phytohormone signals, an increasing number of single polypeptide RING-type E3s have emerged from recent studies with functional roles in balancing the complex pathways of hormonal signaling and mediating hormone cross talks. A multi-domain RING E3, KEEP ON GOING (KEG), for example, has been shown to not only ubiquitinate the b-ZIP transcription factors, ABI5, ABF1, and ABF3(Chen et al. 2013), in ABA responses but also interact with and positively stabilize JAZ12 in the JA pathway (Liu and Stone 2013; Pauwels et al. 2015). A group of RING-domain ligases, including ABA insensitive RING proteins AIRP1–4, have also been identified to regulate ABA signal transduction (Yu et al. 2016). As variants of the RING-type E3s, the U-box proteins share a similar domain structure with the RING domain but lack the zinc-coordinating residues (Yee and Goring 2009). These E3 ligases are also prevalent in plants with 64 family members in *Arabidopsis thaliana* and several members implicated in hormone responses.

2.4 Summary

Plants utilize the ubiquitin-proteasome system to regulate almost every aspect of their physiology. A significant percentage of their genomes were devoted to encode the UPS components. Because of the special roles played by the E3 enzymes in dictating substrate specificity, ubiquitin ligases constitute the majority of the plant UPS and have been greatly diversified to allow seamless integration of protein ubiquitination and degradation with different cellular functions. To mediate hormone signaling, plants have evolved numerous cullin-RING ubiquitin ligase complexes to perceive and translate the hormonal signals into proteasomal degradation of transcriptional regulators. These E3 ligases combine multiple functions in a single enzymatic assembly and efficiently convert diffusible hormonal signals into gene expression regulation in a surprisingly simple pathway. As elaborated in the subsequent chapters of this book, structural biology has made major contributions to plant hormone studies, particularly in establishing the novel mechanisms by which E3 ligases perceive hormonal signals. With a large number of plant ubiquitin ligases and the rest of the plant UPS still poorly characterized, new paradigms of hormone signal transduction through protein degradation might be revealed in future studies.

Acknowledgments D. V. R. and N. Z. are supported by Howard Hughes Medical Institute. Published studies from the Zheng laboratory related to the topics discussed in this chapter have been sponsored by Howard Hughes Medical Institute, National Institutes of Health, National Science Foundation, The Pew Charitable Trusts, Burroughs Wellcome Fund, and the University of Washington, Seattle, WA, USA.

References

Angers S, Li T, Yi X, MacCoss MJ, Moon RT, Zheng N (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. Nature 443(7111):590–593. doi: nature05175 [pii]. https://doi.org/10.1038/nature05175

Banfield MJ (2015) Perturbation of host ubiquitin systems by plant pathogen/pest effector proteins. Cell Microbiol 17(1):18–25. https://doi.org/10.1111/cmi.12385

Bhattacharyya S, Yu H, Mim C, Matouschek A (2014) Regulated protein turnover: snapshots of the proteasome in action. Nat Rev Mol Cell Biol 15(2):122–133. https://doi.org/10.1038/nrm3741

Callis J (2014) The ubiquitination machinery of the ubiquitin system. Arabidopsis Book 12:e0174. https://doi.org/10.1199/tab.0174

Chamovitz DA, Wei N, Osterlund MT, von Arnim AG, Staub JM, Matsui M, Deng XW (1996) The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch. Cell 86(1):115–121

Chen H, Shen Y, Tang X, Yu L, Wang J, Guo L, Zhang Y, Zhang H, Feng S, Strickland E, Zheng N, Deng XW (2006) Arabidopsis CULLIN4 forms an E3 ubiquitin ligase with RBX1 and the CDD complex in mediating light control of development. Plant Cell 18(8):1991–2004. https://doi.org/10.1105/tpc.106.043224

Chen YT, Liu H, Stone S, Callis J (2013) ABA and the ubiquitin E3 ligase KEEP ON GOING affect proteolysis of the Arabidopsis thaliana transcription factors ABF1 and ABF3. Plant J 75 (6):965–976. https://doi.org/10.1111/tpj.12259

Cope GA, Suh GS, Aravind L, Schwarz SE, Zipursky SL, Koonin EV, Deshaies RJ (2002) Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1. Science 298 (5593):608–611. https://doi.org/10.1126/science.1075901

Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem 65:801–847. https://doi.org/10.1146/annurev.bi.65.070196.004101

Deshaies RJ (1999) SCF and Cullin/Ring H2-based ubiquitin ligases. Annu Rev Cell Dev Biol 15:435–467. https://doi.org/10.1146/annurev.cellbio.15.1.435

Deshaies RJ, Joazeiro CA (2009) RING domain E3 ubiquitin ligases. Annu Rev Biochem 78:399–434. https://doi.org/10.1146/annurev.biochem.78.101807.093809

Dharmasiri N, Dharmasiri S, Estelle M (2005) The F-box protein TIR1 is an auxin receptor. Nature 435(7041):441–445. doi:nature03543 [pii]. https://doi.org/10.1038/nature03543

Dou H, Buetow L, Sibbet GJ, Cameron K, Huang DT (2012) BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. Nat Struct Mol Biol 19(9):876–883. https://doi.org/10.1038/nsmb.2379

Duda DM, Borg LA, Scott DC, Hunt HW, Hammel M, Schulman BA (2008) Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. Cell 134 (6):995–1006. https://doi.org/10.1016/j.cell.2008.07.022

Fu ZQ, Dong X (2013) Systemic acquired resistance: turning local infection into global defense. Annu Rev Plant Biol 64:839–863. https://doi.org/10.1146/annurev-arplant-042811-105606

Fu ZQ, Yan S, Saleh A, Wang W, Ruble J, Oka N, Mohan R, Spoel SH, Tada Y, Zheng N, Dong X (2012) NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. Nature 486 (7402):228–232. https://doi.org/10.1038/nature11162

Gingerich DJ, Hanada K, Shiu SH, Vierstra RD (2007) Large-scale, lineage-specific expansion of a bric-a-brac/tramtrack/broad complex ubiquitin-ligase gene family in rice. Plant Cell 19 (8):2329–2348. https://doi.org/10.1105/tpc.107.051300

Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, Zheng N (2004) Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. Cell 119(4):517–528. doi: S0092867404009535 [pii]. https://doi.org/10.1016/j.cell.2004.10.019

Goritschnig S, Zhang Y, Li X (2007) The ubiquitin pathway is required for innate immunity in Arabidopsis. Plant J 49(3):540–551. https://doi.org/10.1111/j.1365-313X.2006.02978.x

Guo H, Ecker JR (2003) Plant responses to ethylene gas are mediated by SCF(EBF1/EBF2)-dependent proteolysis of EIN3 transcription factor. Cell 115(6):667–677

Hamiaux C, Drummond RS, Janssen BJ, Ledger SE, Cooney JM, Newcomb RD, Snowden KC (2012) DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. Curr Biol 22(21):2032–2036. https://doi.org/10.1016/j.cub.2012.08.007

He YJ, McCall CM, Hu J, Zeng Y, Xiong Y (2006) DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. Genes Dev 20(21):2949–2954. https://doi.org/10.1101/gad.1483206

Hershko A, Ciechanover A (1998) The ubiquitin system. Annu Rev Biochem 67:425–479. https://doi.org/10.1146/annurev.biochem.67.1.425

Hua Z, Vierstra RD (2011) The cullin-RING ubiquitin-protein ligases. Annu Rev Plant Biol 62:299–334. https://doi.org/10.1146/annurev-arplant-042809-112256

Huang L, Kinnucan E, Wang G, Beaudenon S, Howley PM, Huibregtse JM, Pavletich NP (1999) Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. Science 286(5443):1321–1326. doi:7996 [pii]

Irigoyen ML, Iniesto E, Rodriguez L, Puga MI, Yanagawa Y, Pick E, Strickland E, Paz-Ares J, Wei N, De Jaeger G, Rodriguez PL, Deng XW, Rubio V (2014) Targeted degradation of abscisic acid receptors is mediated by the ubiquitin ligase substrate adaptor DDA1 in Arabidopsis. Plant Cell 26(2):712–728. https://doi.org/10.1105/tpc.113.122234

Jiang L, Liu X, Xiong G, Liu H, Chen F, Wang L, Meng X, Liu G, Yu H, Yuan Y, Yi W, Zhao L, Ma H, He Y, Wu Z, Melcher K, Qian Q, Xu HE, Wang Y, Li J (2013) DWARF 53 acts as a repressor of strigolactone signalling in rice. Nature 504(7480):401–405. https://doi.org/10.1038/nature12870

Jin J, Arias EE, Chen J, Harper JW, Walter JC (2006) A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. Mol Cell 23(5):709–721. https://doi.org/10.1016/j.molcel.2006.08.010

Jin J, Li X, Gygi SP, Harper JW (2007) Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. Nature 447(7148):1135–1138. https://doi.org/10.1038/nature05902

Kepinski S, Leyser O (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. Nature 435(7041):446–451. doi:nature03542 [pii]. https://doi.org/10.1038/nature03542

Kim HJ, Chiang YH, Kieber JJ, Schaller GE (2013) SCF(KMD) controls cytokinin signaling by regulating the degradation of type-B response regulators. Proc Natl Acad Sci U S A 110(24):10028–10033. https://doi.org/10.1073/pnas.1300403110

Kish-Trier E, Hill CP (2013) Structural biology of the proteasome. Annu Rev Biophys 42:29–49. https://doi.org/10.1146/annurev-biophys-083012-130417

Komander D, Rape M (2012) The ubiquitin code. Annu Rev Biochem 81:203–229. https://doi.org/10.1146/annurev-biochem-060310-170328

Komander D, Clague MJ, Urbé S (2009) Breaking the chains: structure and function of the deubiquitinases. Nat Rev Mol Cell Biol 10(8):550–563. https://doi.org/10.1038/nrm2731

Lander GC, Estrin E, Matyskiela ME, Bashore C, Nogales E, Martin A (2012) Complete subunit architecture of the proteasome regulatory particle. Nature 482(7384):186–191. https://doi.org/10.1038/nature10774

Lau OS, Deng XW (2012) The photomorphogenic repressors COP1 and DET1: 20 years later. Trends Plant Sci 17(10):584–593. https://doi.org/10.1016/j.tplants.2012.05.004

Lechtenberg BC, Rajput A, Sanishvili R, Dobaczewska MK, Ware CF, Mace PD, Riedl SJ (2016) Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. Nature 529(7587):546–550. https://doi.org/10.1038/nature16511

Lee JH, Yoon HJ, Terzaghi W, Martinez C, Dai M, Li J, Byun MO, Deng XW (2010) DWA1 and DWA2, two Arabidopsis DWD protein components of CUL4-based E3 ligases, act together as negative regulators in ABA signal transduction. Plant Cell 22(6):1716–1732. https://doi.org/10.1105/tpc.109.073783

Lee JH, Terzaghi W, Deng XW (2011) DWA3, an Arabidopsis DWD protein, acts as a negative regulator in ABA signal transduction. Plant Sci 180(2):352–357. https://doi.org/10.1016/j.plantsci.2010.10.008

Leyser HM, Lincoln CA, Timpte C, Lammer D, Turner J, Estelle M (1993) Arabidopsis auxin-resistance gene AXR1 encodes a protein related to ubiquitin-activating enzyme E1. Nature 364 (6433):161–164. https://doi.org/10.1038/364161a0

Li C, Liu G, Xu C, Lee GI, Bauer P, Ling HQ, Ganal MW, Howe GA (2003) The tomato suppressor of prosystemin-mediated responses2 gene encodes a fatty acid desaturase required for the biosynthesis of jasmonic acid and the production of a systemic wound signal for defense gene expression. Plant Cell 15(7):1646–1661

Li T, Chen X, Garbutt KC, Zhou P, Zheng N (2006) Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. Cell 124 (1):105–117. doi:S0092-8674(05)01233-X [pii]. https://doi.org/10.1016/j.cell.2005.10.033

Li Y, Zhang L, Li D, Liu Z, Wang J, Li X, Yang Y (2016) The Arabidopsis F-box E3 ligase RIFP1 plays a negative role in abscisic acid signalling by facilitating ABA receptor RCAR3 degradation. Plant Cell Environ 39(3):571–582. https://doi.org/10.1111/pce.12639

Liu H, Stone SL (2013) Cytoplasmic degradation of the Arabidopsis transcription factor abscisic acid insensitive 5 is mediated by the RING-type E3 ligase KEEP ON GOING. J Biol Chem 288 (28):20267–20279. https://doi.org/10.1074/jbc.M113.465369

Lyapina S, Cope G, Shevchenko A, Serino G, Tsuge T, Zhou C, Wolf DA, Wei N, Deshaies RJ (2001) Promotion of NEDD-CUL1 conjugate cleavage by COP9 signalosome. Science 292 (5520):1382–1385. https://doi.org/10.1126/science.1059780

Lydeard JR, Schulman BA, Harper JW (2013) Building and remodelling Cullin-RING E3 ubiquitin ligases. EMBO Rep 14(12):1050–1061. https://doi.org/10.1038/embor.2013.173

Mandadi KK, Misra A, Ren S, McKnight TD (2009) BT2, a BTB protein, mediates multiple responses to nutrients, stresses, and hormones in Arabidopsis. Plant Physiol 150(4):1930–1939. https://doi.org/10.1104/pp.109.139220

Marín I (2010) Diversification and specialization of plant RBR ubiquitin ligases. PLoS One 5(7): e11579. https://doi.org/10.1371/journal.pone.0011579

Marín I (2013) Evolution of plant HECT ubiquitin ligases. PLoS One 8(7):e68536. https://doi.org/10.1371/journal.pone.0068536

Metzger MB, Hristova VA, Weissman AM (2012) HECT and RING finger families of E3 ubiquitin ligases at a glance. J Cell Sci 125(Pt 3):531–537. https://doi.org/10.1242/jcs.091777

Meyer HJ, Rape M (2014) Enhanced protein degradation by branched ubiquitin chains. Cell 157 (4):910–921. https://doi.org/10.1016/j.cell.2014.03.037

Murase K, Hirano Y, Sun TP, Hakoshima T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. Nature 456(7221):459–463. doi:nature07519 [pii]. https://doi.org/10.1038/nature07519

Nakamura H, Xue YL, Miyakawa T, Hou F, Qin HM, Fukui K, Shi X, Ito E, Ito S, Park SH, Miyauchi Y, Asano A, Totsuka N, Ueda T, Tanokura M, Asami T (2013) Molecular mechanism of strigolactone perception by DWARF14. Nat Commun 4:2613. https://doi.org/10.1038/ncomms3613

Pauwels L, Ritter A, Goossens J, Durand AN, Liu H, Gu Y, Geerinck J, Boter M, Vanden Bossche R, De Clercq R, Van Leene J, Gevaert K, De Jaeger G, Solano R, Stone S, Innes RW, Callis J, Goossens A (2015) The RING E3 ligase KEEP ON GOING modulates JASMONATE ZIM-DOMAIN12 stability. Plant Physiol 169(2):1405–1417. https://doi.org/10.1104/pp.15.00479

Pickart CM (2001) Mechanisms underlying ubiquitination. Annu Rev Biochem 70:503–533. https://doi.org/10.1146/annurev.biochem.70.1.503

Pierce NW, Lee JE, Liu X, Sweredoski MJ, Graham RL, Larimore EA, Rome M, Zheng N, Clurman BE, Hess S, Shan SO, Deshaies RJ (2013) Cand1 promotes assembly of new SCF complexes through dynamic exchange of F box proteins. Cell 153(1):206–215. https://doi.org/10.1016/j.cell.2013.02.024

Plechanovová A, Jaffray EG, McMahon SA, Johnson KA, Navrátilová I, Naismith JH, Hay RT (2011) Mechanism of ubiquitylation by dimeric RING ligase RNF4. Nat Struct Mol Biol 18 (9):1052–1059. https://doi.org/10.1038/nsmb.2108

Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, Genschik P (2003) EIN3-dependent regulation of plant ethylene hormone signaling by two arabidopsis F box proteins: EBF1 and EBF2. Cell 115(6):679–689

Pruneda JN, Littlefield PJ, Soss SE, Nordquist KA, Chazin WJ, Brzovic PS, Klevit RE (2012) Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. Mol Cell 47(6):933–942. https://doi.org/10.1016/j.molcel.2012.07.001

Qiao H, Chang KN, Yazaki J, Ecker JR (2009) Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis. Genes Dev 23 (4):512–521. https://doi.org/10.1101/gad.1765709

Rotin D, Kumar S (2009) Physiological functions of the HECT family of ubiquitin ligases. Nat Rev Mol Cell Biol 10(6):398–409. https://doi.org/10.1038/nrm2690

Scott DC, Rhee DY, Duda DM, Kelsall IR, Olszewski JL, Paulo JA, de Jong A, Ovaa H, Alpi AF, Harper JW, Schulman BA (2016) Two distinct types of E3 ligases work in unison to regulate substrate ubiquitylation. Cell 166(5):1198–1214.e1124. https://doi.org/10.1016/j.cell.2016.07.027

Seo KI, Lee JH, Nezames CD, Zhong S, Song E, Byun MO, Deng XW (2014) ABD1 is an Arabidopsis DCAF substrate receptor for CUL4-DDB1-based E3 ligases that acts as a negative regulator of abscisic acid signaling. Plant Cell 26(2):695–711. https://doi.org/10.1105/tpc.113.119974

Shabek N, Zheng N (2014) Plant ubiquitin ligases as signaling hubs. Nat Struct Mol Biol 21 (4):293–296. https://doi.org/10.1038/nsmb.2804

Sheard LB, Tan X, Mao H, Withers J, Ben-Nissan G, Hinds TR, Kobayashi Y, Hsu FF, Sharon M, Browse J, He SY, Rizo J, Howe GA, Zheng N (2010) Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. Nature 468(7322):400–405. doi:nature09430 [pii]. https://doi.org/10.1038/nature09430

Shi H, Shen X, Liu R, Xue C, Wei N, Deng XW, Zhong S (2016) The red light receptor phytochrome B directly enhances substrate-E3 ligase interactions to attenuate ethylene responses. Dev Cell 39(5):597–610. https://doi.org/10.1016/j.devcel.2016.10.020

Shimada A, Ueguchi-Tanaka M, Nakatsu T, Nakajima M, Naoe Y, Ohmiya H, Kato H, Matsuoka M (2008) Structural basis for gibberellin recognition by its receptor GID1. Nature 456 (7221):520–523. https://doi.org/10.1038/nature07546

Skaar JR, Pagan JK, Pagano M (2013) Mechanisms and function of substrate recruitment by F-box proteins. Nat Rev Mol Cell Biol 14(6):369–381. https://doi.org/10.1038/nrm3582

Spratt DE, Walden H, Shaw GS (2014) RBR E3 ubiquitin ligases: new structures, new insights, new questions. Biochem J 458(3):421–437. https://doi.org/10.1042/BJ20140006

Stewart MD, Ritterhoff T, Klevit RE, Brzovic PS (2016) E2 enzymes: more than just middle men. Cell Res 26(4):423–440. https://doi.org/10.1038/cr.2016.35

Stieglitz B, Rana RR, Koliopoulos MG, Morris-Davies AC, Schaeffer V, Christodoulou E, Howell S, Brown NR, Dikic I, Rittinger K (2013) Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. Nature 503(7476):422–426. https://doi.org/10.1038/nature12638

Strader LC, Zhao Y (2016) Auxin perception and downstream events. Curr Opin Plant Biol 33:8–14. https://doi.org/10.1016/j.pbi.2016.04.004

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446(7136):640–645. doi:nature05731 [pii]. https://doi.org/10.1038/nature05731

Thrower JS, Hoffman L, Rechsteiner M, Pickart CM (2000) Recognition of the polyubiquitin proteolytic signal. EMBO J 19(1):94–102. https://doi.org/10.1093/emboj/19.1.94

Tomko RJ, Hochstrasser M (2013) Molecular architecture and assembly of the eukaryotic proteasome. Annu Rev Biochem 82:415–445. https://doi.org/10.1146/annurev-biochem-060410-150257

Toyama BH, Hetzer MW (2013) Protein homeostasis: live long, won't prosper. Nat Rev Mol Cell Biol 14(1):55–61. https://doi.org/10.1038/nrm3496

Trempe JF, Sauvé V, Grenier K, Seirafi M, Tang MY, Ménade M, Al-Abdul-Wahid S, Krett J, Wong K, Kozlov G, Nagar B, Fon EA, Gehring K (2013) Structure of parkin reveals mechanisms for ubiquitin ligase activation. Science 340(6139):1451–1455. https://doi.org/10.1126/science.1237908

van der Veen AG, Ploegh HL (2012) Ubiquitin-like proteins. Annu Rev Biochem 81:323–357. https://doi.org/10.1146/annurev-biochem-093010-153308

Vierstra RD (2009) The ubiquitin-26S proteasome system at the nexus of plant biology. Nat Rev Mol Cell Biol 10(6):385–397. https://doi.org/10.1038/nrm2688

Wang KL, Yoshida H, Lurin C, Ecker JR (2004) Regulation of ethylene gas biosynthesis by the Arabidopsis ETO1 protein. Nature 428(6986):945–950. https://doi.org/10.1038/nature02516

Wauer T, Komander D (2013) Structure of the human Parkin ligase domain in an autoinhibited state. EMBO J 32(15):2099–2112. https://doi.org/10.1038/emboj.2013.125

Wei N, Deng XW (2003) The COP9 signalosome. Annu Rev Cell Dev Biol 19:261–286. https://doi.org/10.1146/annurev.cellbio.19.111301.112449

Wei N, Tsuge T, Serino G, Dohmae N, Takio K, Matsui M, Deng XW (1998) The COP9 complex is conserved between plants and mammals and is related to the 26S proteasome regulatory complex. Curr Biol 8(16):919–922

Wenzel DM, Lissounov A, Brzovic PS, Klevit RE (2011a) UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. Nature 474(7349):105–108. https://doi.org/10.1038/nature09966

Wenzel DM, Stoll KE, Klevit RE (2011b) E2s: structurally economical and functionally replete. Biochem J 433(1):31–42. https://doi.org/10.1042/BJ20100985

Wertz IE, O'Rourke KM, Zhang Z, Dornan D, Arnott D, Deshaies RJ, Dixit VM (2004) Human De-etiolated-1 regulates c-Jun by assembling a CUL4A ubiquitin ligase. Science 303(5662):1371–1374. https://doi.org/10.1126/science.1093549

Wilkinson KD (2005) The discovery of ubiquitin-dependent proteolysis. Proc Natl Acad Sci U S A 102(43):15280–15282. https://doi.org/10.1073/pnas.0504842102

Wu Y, Zhang D, Chu JY, Boyle P, Wang Y, Brindle ID, De Luca V, Després C (2012) The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. Cell Rep 1(6):639–647. https://doi.org/10.1016/j.celrep.2012.05.008

Xie DX, Feys BF, James S, Nieto-Rostro M, Turner JG (1998) COI1: an Arabidopsis gene required for jasmonate-regulated defense and fertility. Science 280(5366):1091–1094

Xing W, Busino L, Hinds TR, Marionni ST, Saifee NH, Bush MF, Pagano M, Zheng N (2013) SCF(FBXL3) ubiquitin ligase targets cryptochromes at their cofactor pocket. Nature 496(7443):64–68. https://doi.org/10.1038/nature11964

Yao R, Ming Z, Yan L, Li S, Wang F, Ma S, Yu C, Yang M, Chen L, Li Y, Yan C, Miao D, Sun Z, Yan J, Sun Y, Wang L, Chu J, Fan S, He W, Deng H, Nan F, Li J, Rao Z, Lou Z, Xie D (2016) DWARF14 is a non-canonical hormone receptor for strigolactone. Nature 536(7617):469–473. https://doi.org/10.1038/nature19073

Yee D, Goring DR (2009) The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. J Exp Bot 60(4):1109–1121. https://doi.org/10.1093/jxb/ern369

Yu F, Wu Y, Xie Q (2016) Ubiquitin-proteasome system in ABA signaling: from perception to action. Mol Plant 9(1):21–33. https://doi.org/10.1016/j.molp.2015.09.015

Zhao Y, Thilmony R, Bender CL, Schaller A, He SY, Howe GA (2003) Virulence systems of Pseudomonas syringae pv. Tomato promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway. Plant J 36(4):485–499. doi:1895 [pii]

Zhou F, Lin Q, Zhu L, Ren Y, Zhou K, Shabek N, Wu F, Mao H, Dong W, Gan L, Ma W, Gao H, Chen J, Yang C, Wang D, Tan J, Zhang X, Guo X, Wang J, Jiang L, Liu X, Chen W, Chu J, Yan C, Ueno K, Ito S, Asami T, Cheng Z, Lei C, Zhai H, Wu C, Wang H, Zheng N, Wan J (2013) D14-SCF(D3)-dependent degradation of D53 regulates strigolactone signalling. Nature 504(7480):406–410. https://doi.org/10.1038/nature12878

Zimmerman ES, Schulman BA, Zheng N (2010) Structural assembly of cullin-RING ubiquitin ligase complexes. Curr Opin Struct Biol 20(6):714–721. doi:S0959-440X(10)00135-1 [pii]. https://doi.org/10.1016/j.sbi.2010.08.010

Chapter 3
Structural Insight into Recognition of Plant Peptide Hormones by Plant Receptor Kinases

Jizong Wang, Guangzhong Lin, Rui Ma, Zhifu Han, and Jijie Chai

3.1 Introduction to Plant Peptide Hormones and Receptor Kinases (RKs)

Cell signaling transduction in multicellular organisms is mediated by signals such as peptides, steroids, and other small molecular compounds (Hunter 2000). In higher plants, intercellular communications involved in plant growth and development largely depend on the seven small molecular compounds: auxin, cytokinin, gibberellin, abscisic acid, ethylene, brassinosteroids, and jasmonic acid (Santner et al. 2009; Santner and Estelle 2009). These small molecules are collectively called conventional plant hormones. Over the past two decades, however, increasingly more studies have indicated that signaling peptides also have crucial roles in regulating plant physiology, including cell proliferation, growth, differentiation, innate immunity, and senescence (Grienenberger and Fletcher 2015; Marshall et al. 2011; Matsubayashi 2014; Murphy et al. 2012). Therefore, plant signaling peptides are starting to be accepted as peptide hormones, similar to those in animals (Tager and Steiner 1974). Since the first plant peptide hormone systemin was identified in tomato (Pearce et al. 1991), a dozen peptide hormones have been characterized and over 1000 are predicted to exist in the *Arabidopsis* genome (Lease and Walker 2006; Silverstein et al. 2007).

Plant peptide hormones can be classified into two groups: the small posttranslationally modified signaling peptides (called small peptide hormones hereafter) and the cysteine-rich peptides (CRPs) with ~5–20 and ~50 amino acids in size in their mature functional forms, respectively (Marshall et al. 2011; Matsubayashi

Jizong Wang and Guangzhong Lin are contributed equally to this work.

J. Wang · G. Lin · R. Ma · Z. Han · J. Chai (✉)
Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China
e-mail: chaijj@tsinghua.edu.cn

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_3

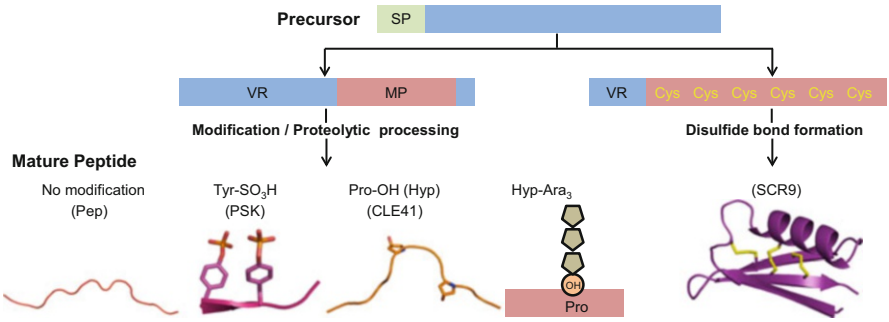


Fig. 3.1 Schematic representation of maturation of different plant peptide hormones. Tyr-SO_3H, Pro-OH (Hyp), and Hyp-Ara_3 indicate the modifications of tyrosine sulfation, proline hydroxylation, and hydroxyproline arabinosylation, respectively. An example (in parenthesis) is given for each of these modifications. Disulfide bonds are shown in yellow and stick. SP: N-terminal signal peptide; VR: central variable region; MP: C-terminal mature peptide domain. Pep: plant elicitor peptide; PSK: phytosulfokine; CLE41: CLAVATA3/Endosperm surrounding region- related (CLE) 41; SCR9: S-locus CRP

2014). Both groups are synthesized as precursors (Fig. 3.1) featuring an N-terminal signal peptide (SP), a central variable region (VR), and a C-terminal mature peptide domain (MP). Maturation of small peptide precursors requires posttranslational modifications such as tyrosine sulfation, proline hydroxylation, hydroxyproline arabinosylation, and extensive proteolytic processing (Matsubayashi 2014; Murphy et al. 2012). In contrast, activation of CRPs involves formation of intramolecular disulfide bonds and likely proteolytic processing. Besides the tomato systemin, other well-studied small signaling peptides include CLAVATA3/Endosperm surrounding region-related (CLE), inflorescence deficient in abscission (IDA), C-terminally encoded peptide (CEP), root meristem growth factor (RGF), and phytosulfokine (PSK) (Matsubayashi 2014). Typical examples of CRPs are epidermal patterning factor (EPF), LURE1/2, tapetum determinent 1 (TPD1), S-locus CRP (SCR), and rapid alkalinization factor (RALF) (Marshall et al. 2011).

In animals, both receptor tyrosine kinases (RTKs) and G protein-coupled receptors (GPCRs) function as the receptors of peptide hormones (Catt and Dufau 1977). By contrast, nearly all the receptors of plant peptide hormones identified thus far are RKs. RKs are a large family of membrane-anchored receptors with more than 600 encoded in *Arabidopsis* genome (Shiu and Bleecker 2001b). RKs have a conserved tripartite domain structure with a varied extracellular domain, a single transmembrane domain, and a conserved cytoplasmic serine/threonine kinase domain (Shiu and Bleecker 2001a). Based on their variable extracellular domains, plant RKs can be classified into over ten families. RKs with leucine-rich repeat (LRR-RK) ectodomain have more than 240 members in *Arabidopsis* and constitute the largest subfamily of RKs. LRR-RKs can be further divided into 13 subfamilies (Gou et al. 2010; Shiu and Bleecker 2001b; Torii 2004). Most of the peptide hormone receptors identified thus far belong to the LRR-RK family. These include the small signaling peptide receptor CLV1 (a receptor of CLAVATA3; (Hazak and

Hardtke 2016), PXY (a receptor of CLE41; (Fisher and Turner 2007), HAESA (a receptor of IDA; (Jinn et al. 2000), CEPRs (receptors of CEPs; (Tabata et al. 2014), and RGFRs (receptors of RGFs; (Ou et al. 2016; Shinohara et al. 2016; Song et al. 2016b). It is of interest to note that all these receptors belong to the LRR-XI subfamily of RKs.

Recent studies have elucidated the structures of several plant peptide hormones in complex with their RK receptors. These structural data provide significant insight into the recognition mechanisms of peptide hormones, in particular small peptide hormones, by their receptors. Here we summarize these structural data and suggest some directions for further structural study of plant peptide hormones.

3.2 Perception of Small Signaling Peptides by LRR-RKs Without an ID

The structural study of the plant elicitor peptide 1 in *Arabidopsis* (AtPep1) in complex with its LRR-RK receptor PEPR1 provided the first view of small signaling peptide recognition by RKs (Tang et al. 2015). Shortly after, several crystal structures of small peptide hormones bound by their cognate LRR-RK receptors including CLE41-PXY, IDA-HAESA and RGF1-RGFR were elucidated (Santiago et al. 2016; Song et al. 2016b; Zhang et al. 2016a). Structural comparison showed that these four LRR-RKs exhibit a superhelix structure (Fig. 3.2a), similar to those of other LRR-RKs (Hothorn et al. 2011; She et al. 2011) that recognize non-peptide hormones, further supporting the idea that the LRR domains of RKs have a conserved structure. The four receptor-bound peptides are similarly oriented and bind the inner surface of the LRR structures of these LRR-RKs (Fig. 3.2a). These structural features resemble those of the bacteria-derived flg22 binding to its receptor FLS2 (Sun et al. 2013b). The former four peptide hormones adopt different conformations when recognized by their respective receptors. AtPep1 and RGF1 are fully extended, whereas CLE41 and IDA have an "Ω"-like kink in the middle (Fig. 3.2a). No striking conformational changes occur to the receptors following binding of the small signaling peptides (Santiago et al. 2016; Zhang et al. 2016a).

Despite the different overall conformations of the four receptor-bound small peptide hormones, their C-terminal sides, in particular the last residues, are highly conserved when recognized by receptors (Fig. 3.2b). Remarkably, two critical arginine residues (called RxR motif hereafter) from the four receptors responsible for recognition of the last amino acids of the peptides are also conserved. Interestingly, the RxR motif is not only conserved among the four LRR-RKs but also largely invariable among the other members of LRR-RK XI family (Fig. 3.3a). Given the fact that the last amino acids of the four peptides, either histidine or asparagine, are also present in many other known small signaling peptides (Fig. 3.3b), these structural observations raise the possibility that LRR-RK XI family can recognize the small peptide hormones ending with the histidine or asparagine residues as

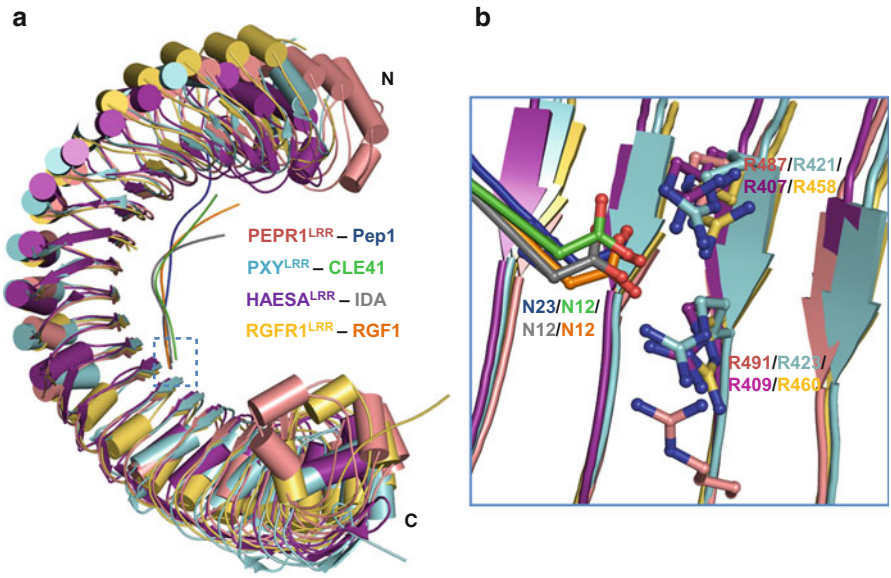


Fig. 3.2 Structural comparison of small signaling peptides in complex with LRR-RKs without an ID. (**a**) Structure alignment of the PEPR1^{LRR}-Pep1, PXY^{LRR}-CLE41, HAESA^{LRR}-IDA, and RGFR1^{LRR}-RGF1 complexes. Color codes are indicated. "N": N-terminus; "C": C-terminus. (**b**) A detailed view of the C-termini of AtPep1, CLE41, IDA, and RGF1 interaction with their respective receptors, which is boxed in blue dashed line in "**a**)". The residues involved in interactions are labeled with the same color codes shown in "**a**)"

ligands. As a matter of fact, a biochemical assay based on this hypothesis led to the identification of RGFRs (Song et al. 2016b), which were further confirmed by two other research groups using genetic screens (Ou et al. 2016; Shinohara et al. 2016). Supporting an important role of the RxR motif in ligand recognition, removal of the last asparagine residue from AtPep1 nearly abolished AtPep1-PEPR1 interaction in vitro and led to ~400-fold loss of immunity-inducing activity in suspension cells (Pearce et al. 2008; Tang et al. 2015).

A question raised by the likely conserved small peptide hormone recognition mechanism of LRR-RK XI subfamily is how their ligand specificities are achieved. While the C-termini of different subfamilies of peptide hormones are conserved, their N-terminal sides vary greatly in sequence (Fig. 3.3b). The nonconserved residues of peptide hormones are presumably important for their specific recognition by this family of LRR-RKs. In line with this possibility, different subfamilies of peptides hormones adopt distinct conformations when recognized by their receptors (Fig. 3.2a). Even within the same group of peptide hormones, variations in sequences can also confer their receptor specificity. For example, the nine RGF members in *Arabidopsis* are highly conserved at their N- and C-terminal sides. In contrast, their central regions are comparatively less conserved (Matsuzaki et al. 2010). Our structural study showed that it is these regions that dictate specific

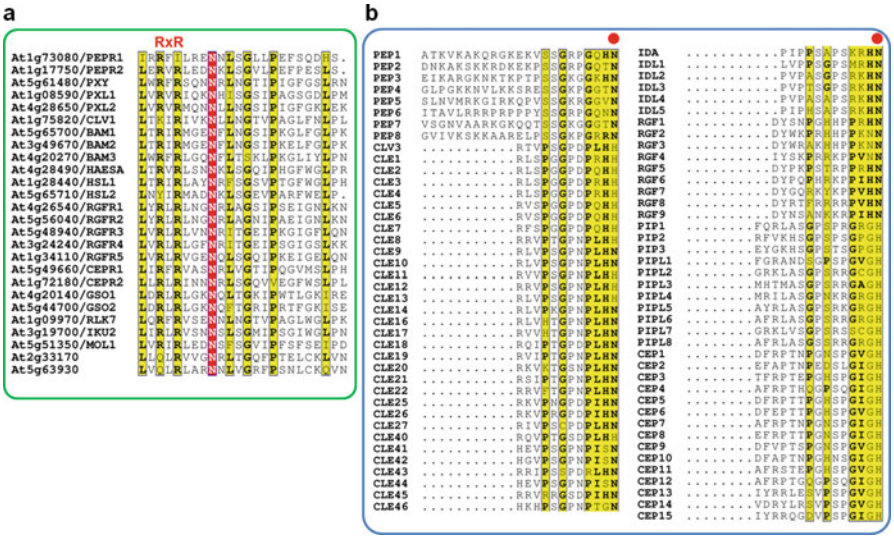


Fig. 3.3 Sequence alignment among the members of LRR-RK XI subfamily and small peptide hormones. (**a**) Sequence alignment of LRR-RKs XI subfamily members in *Arabidopsis* surrounding "RxR" motif. The conserved "RxR" motif is labeled. (**b**) Sequence alignment of mature small signaling peptides. The conserved C-termini are indicated by red solid circle

recognition of RGFs by different RGFRs (Song et al. 2016b), though cellular localizations may also have a role in this respect. Some subfamilies of peptide hormones such as CLEs (Fig. 3.3b) are highly conserved in their mature forms. In some cases, the same peptide hormone has been reported to have different receptors (Depuydt et al. 2013; Endo et al. 2013; Kang and Hardtke 2016). In addition to distinct cellular localizations, as will be discussed subsequently in detail, posttranslational modifications in small signaling peptides may also contribute to their specific recognition by receptors.

3.3 Perception of Small Signaling Peptides by LRR-RKs with an ID

There are non-LRR subdomains present in the ectodomain of some LRR-RKs such as the members of LRR-RK X subfamily and many RLPs (Fritz-Laylin et al. 2005). The non-LRR regions are called "island domain (ID)" as they are generally embedded between two consecutive LRR motifs and protrude out of the inner surface of the LRR structures. The available structural data support the idea that a major function of the IDs is dedicated to ligand binding (Hothorn et al. 2011; She et al. 2011; Wang et al. 2015). However, paucity of structural information renders it difficult to generalize how IDs are involved in ligand recognition. Furthermore, the numbers,

lengths, and positions of the IDs vary greatly among different members of LRR-RKs (Amano et al. 2007; Hothorn et al. 2011; She et al. 2011; Song et al. 2014; Wang et al. 2015; Zhao et al. 2002). Therefore, it is not surprising that this type of LRR-RKs have ligands with highly diversified chemical structures.

Three structures of ID-containing LRR-RKs were elucidated with two of them in ligand-binding forms (Hothorn et al. 2011; She et al. 2011; Song et al. 2014; Wang et al. 2015). One of the two ligand-bound LRR-RKs is PSKR, a member of LRR-RK X subfamily that recognizes the small signaling peptide phytosulfokine (PSK) (Matsubayashi et al. 2002). PSK is a pentapeptide with tyrosine sulfation in the first and third tyrosine residues (Matsubayashi and Sakagami 1996) and functions to regulate plant growth, development, and innate immunity (Sauter 2015). In *Arabidopsis*, five paralogous PSK precursor genes have been identified. These precursor proteins have about 80 amino acids with a secreted signal peptide at the N-terminus and a five amino acids PSK sequence close to the C-terminus (Matsubayashi et al. 2006).

Similar to the structures of other plant LRR-RKs (Hothorn et al. 2011; She et al. 2011; Sun et al. 2013b), $PSKR^{LRR}$ exhibits a highly curved superhelical structure (Wang et al. 2015) (Fig. 3.4a). $PSKR^{LRR}$ has a 36-amino acid island domain ($PSKR^{ID}$, residues 502–537). However, in contrast with that of BRI1 (Hothorn et al. 2011; She et al. 2011), folding of the $PSKR^{ID}$ is PSK-induced, as it is completely disordered in the structure of the PSK-free $PSKR^{LRR}$. Sandwiched between the $PSKR^{ID}$ and the inner side of $PSKR^{LRR}$, PSK mainly adopts a short β-strand forming an anti-parallel β-sheet with the β-strand from the $PSKR^{ID}$ (Fig. 3.4b). These structural observations indicate that PSK binding induces the insertion domain well-structured to complete the $PSKR^{ID}$. Interestingly, structural comparison showed that the $PSKR^{ID}$ together with PSK is similarly positioned to $BRI1^{ID}$ (Fig. 3.4c).

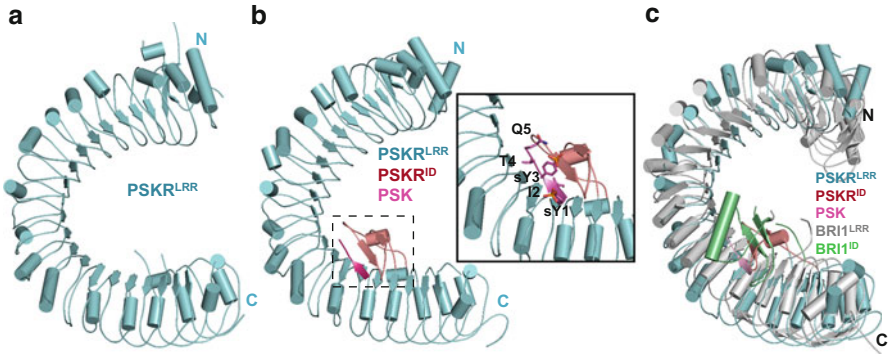


Fig. 3.4 Structures of PSK in complex with its receptor PSKR. (**a**) Overall structure of the free $PSKR^{LRR}$ with its island domain disordered. (**b**) Shown on the left is the overall structure of $PSKR^{LRR}$-PSK recognition complex. Color codes are indicated. Right: a close-up view of PSK interaction with the island domain and the inner surface of LRR boxed in black dashed line in the left panel. (**c**) The $PSKR^{ID}$ together with PSK is similarly positioned to $BRI1^{ID}$. Shown is the structural comparison of $PSKR^{LRR}$-PSK complex with the free $BRI1^{LRR}$

EXCESS MICROSPOROCYTES 1 (EMS1) and PSY1 receptor (PSYR) also belong to the LRR-RK X family, which control somatic and reproductive cell fates in the *Arabidopsis* anther (Zhao et al. 2002) and regulate plant growth and immunity (Amano et al. 2007), respectively. Both EMS1 and PSYR have an ID and the one from PSYR is similar to that of PSKR. Like PSKR, PSYR also recognizes a small signaling peptide, PSY (plant peptide containing sulfated tyrosine). PSY contains 18-amino acids with one tyrosine residue sulfated and one hydroxyproline arabinosylated, both of which are important for the full activity of PSY (Amano et al. 2007). In contrast with PSKR and PSYR, EMS1 likely perceives the CRP, TPD1 (TAPETUM DETERMINANT 1), as a ligand for cell specialization in the *Arabidopsis* anther (Yang et al. 2003). However, whether and how IDs from these LRR-RK and LRR-RLPs contribute to ligand recognition remains unknown, structural studies are needed to address these questions.

3.4 Perception of CRPs by RKs

The genome of *Arabidopsis* encodes ~1000 CRPs that play crucial roles in growth, development, reproduction, and immune defense (Marshall et al. 2011). Structurally, CRPs are similar to the peptide hormones such as growth hormone, insulin, and epidermal growth factor (EGF) in animals (Tager and Steiner 1974). Given the large number of CRPs, it is not surprising that they can be recognized by diverse subfamilies of RKs including LRR-RKs (Takeuchi and Higashiyama 2016; Torii 2012) and S-domain (SD)-RKs (Ivanov et al. 2010). This contrasts with the small peptide hormones, nearly all of which are recognized by LRR-RKs (Endo et al. 2014). Despite the large number and significance of CRPs in plants, thus far only one structure of CRP bound by its receptor has been solved. Therefore, many structural efforts directed to elucidate the molecular mechanisms of CRP recognition by their receptors are needed.

The known structure of CRP in complex with its receptor is about self-incompatibility (SI), which is a widespread phenomenon in flowering plants for prevention of self-fertilization and inbreeding (Ma et al. 2016). Genetic studies showed that a single polymorphic locus, termed *S* locus, controls the SI response in many species (Kitashiba and Nasrallah 2014; Takayama and Isogai 2005). The *S*-locus usually contains two tightly linked genes encoding the male and female *S* determinants, which have coevolved to generate a variety of different *S* haplotypes (Iwano and Takayama 2012). In *Brassica*, the male and female *S* determinants are the pollen expressed *S*-locus cysteine-rich protein (SCR, or *S*-locus protein 11) and the stigma expressed *S*-locus receptor kinase (SRK) (Schopfer and Nasrallah 2000; Takasaki et al. 2000; Takayama et al. 2000). The SI response in *Brassica* is mediated by specific recognition of an SCR by SRK variants encoded in the same *S*-locus haplotype (Kachroo et al. 2001; Takayama et al. 2001). SCR is a prototypic member of CRP hormones, whereas SRK belongs to RKs with an

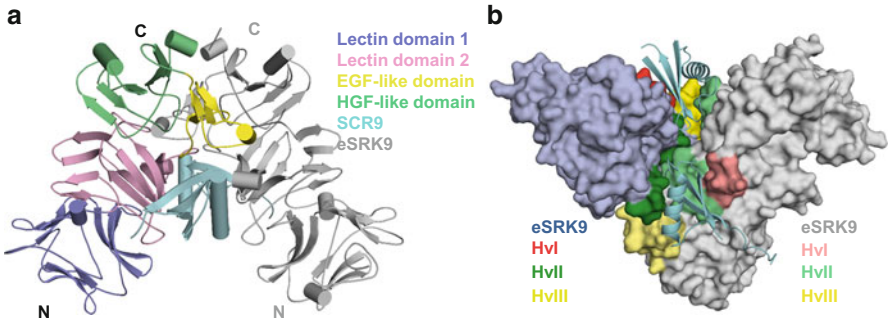


Fig. 3.5 Structures of CRP recognition by SD-RK. (**a**) Architecture of the SCR9-eSRK9 heterotetrameric complex. The overall crystal structure of the eSRK9-SCR9 complex is shown. The lectin domain 1, lectin domain 2, EGF-like domain, and HGF-like domain from one eSRK9 are shown in blue, pink, yellow, and green, respectively. For clarity, the other eSRK9 molecule is colored gray. (**b**) SCR specifically binds three Hv regions of eSRK9. The two eSRK9 molecules (blue and gray) are shown in surface representation and the two SCR9 molecules (cyan) are exhibited by ribbon diagrams. The three Hv regions (HvI, HvII, and HvIII) of the two eSRK9 molecules are colored red, green, yellow and pink, light green, light yellow respectively

extracellular S domain. The S domain consists of two N-terminal lectin domains, an EGF-like domain, and a C-terminal HGF-like domain (Ma et al. 2016).

Structural and biochemical studies showed that binding of SCR9 induced eSRK9 homodimerization by interacting with the second lectin domain and the EGF-like domain of eSRK9, forming a 2:2 SCR9:eSRK9 heterotetramer (Fig. 3.5a). These data provide further evidence for the dimerization model of RK activation (Han et al. 2014; Liu et al. 2012). The specific recognition of SCR9 is through three hypervariable regions (HvI, HvII, and HvIII) of eSRK9 (Fig. 3.5a), supporting the hypothesis that these regions of different SRK variants dictate specific recognition of their cognate SCRs (Boggs et al. 2009; Ivanov et al. 2010; Kemp and Doughty 2007; Kusaba et al. 1997; Sato et al. 2002). SCR9 interacts with the HvI, HvIII, the C-terminal half of HvII from one eSRK9 monomer, and the N-terminal half of HvII from the other eSRK9 monomer in chorus (Fig. 3.5b).

In the structure of the eSRK9-SCR9 complex, eSRK dimerization is both ligand- and receptor-mediated. Interestingly, interactions between the two eSRK9 molecules in the complex are mainly mediated by the HvII and HvIII regions (Fig. 3.5b), suggesting that the dimerization is SRK9-specific. Furthermore, simultaneous interaction of eSCR9 with HvI, HvII, and HvIII from two eSRK molecules suggests that a homodimer of SRK9 encodes its specific recognition of SCR9. Thus, the preformed homodimers of SRKs observed in planta (Giranton et al. 2000; Shimosato et al. 2007), likely transient in nature, may as a whole function as the receptors of their cognate SCRs. This affords an explanation for the observations that non-cognate SRK-SCR binding was detected but no SI response was induced in planta (Chookajorn et al. 2004; Kemp and Doughty 2007). It seems that, in addition to the important role in SRK activation, homodimerization of an SRK may also act as a fail-safe mechanism to ensure specific SI response. Such a mechanism would be

particularly important for specific SI in the heterozygous plants that sometimes encode highly conserved SRK variants on stigma (Naithani et al. 2007).

3.5 Posttranslational Modifications in Recognition of Peptide Hormones by Their Receptors

Posttranslational modifications including proteolytic processing, tyrosine sulfation, proline hydroxylation, and hydroxyproline arabinosylation are common among peptide hormones, in particular for the small peptide hormones. For this reason, small peptide hormones are also called posttranslationally modified peptides (Matsubayashi 2014; Murphy et al. 2012). In many cases, modifications are important for the activities of peptide homones (Amano et al. 2007; Ni et al. 2011; Ohyama et al. 2009; Tamaki et al. 2013). In this section, we discuss how these modifications contribute to peptide hormone recognition by their receptors.

3.5.1 Proline Hydroxylation

Proline hydroxylation is the most common form of posttranslational modification in peptide hormones and many of them such as CLV3, CLV9, CLE41, and IDA have been found to undergo this type of modification in vivo (Butenko et al. 2014; Ito et al. 2006; Kondo et al. 2006; Shinohara et al. 2012). A recent structural study showed that proline hydroxylation was required for recognition of the small peptide hormone IDA by its receptor HAESA (Santiago et al. 2016). In the structure, the hydroxyl group from the hydroxylated proline (Hyp7) of IDA establishes a hydrogen bond with Glu266 of HAESA. Mutagenesis study supported an important role of this single hydrogen bond in strong IDA interaction with the ecto-LRR domain of HAESA. However, in contrast to that of IDA, proline hydroxylation of CLE41 was dispensable for CLE41 interaction with PXY in vitro as supported by structural and biochemical data (Zhang et al. 2016a). It is still unknown whether and how proline hydroxylation contributes to the activity of CLE41. Roles other than ligand binding including trafficking, stability, and resistance to proteases have been proposed for proline hydroxylation in the activity of peptide hormones (Matsubayashi 2014). Alternatively or additionally, it is also possible that proline hydroxylation may alter the receptor specificity of a small peptide hormone. This would substantially increase the diversities of small signaling peptides and thus endow them with more physiological functions. Consistently, CLE45 was shown to be recognized by the two different LRR-RKs, BAM3 (Depuydt et al. 2013; Kang and Hardtke 2016) and SKM1 (Endo et al. 2013), though more studies are needed to reconcile these results.

3.5.2 Proteolytic Processing

Proteolytic cleavage at the N-terminal sides of precursor peptides is a critical step for maturation of small signaling peptides (Schardon et al. 2016; Srivastava et al. 2008). But some functional forms of small peptide hormones also contain flanking sequences at the C-terminal sides of their precursors. This suggests that C-terminal proteolytic processing may be required for the maturation of these precursors. This agrees with the study showing that the Zn^{2+} carboxypeptidase SOL1 catalyzed C-terminal processing of the CLE19 proprotein to produce the functional CLE19 peptide (Tamaki et al. 2013), though the mechanism underlying remains unclear. Consistent with these biochemical data, recently structural studies showed that, at least for some small peptide hormones, C-terminal proteolytic processing is essential for receptor binding (Santiago et al. 2016; Song et al. 2016b; Tang et al. 2015; Zhang et al. 2016a). The free carboxyl group of the last residue of CLE41, AtPep1, RGF1, or IDA with the RxR motif is critical for their receptor binding. For CLE41 and AtPep1, addition of an arginine residue at the C-termini of their mature forms greatly reduced their activities of interaction with PXY and PEPR1 in vitro, respectively. These two structures showed that C-terminal extensions will neutralize the negative charges at their C-termini of CLE41 and AtPep1 required for interaction with the RxR motif.

3.5.3 Tyrosine Sulfation

Thus far, the modification of sulfation in peptide hormones including PSK, PSY, and RGF has been shown to occur only for the tyrosine residue (Amano et al. 2007; Matsubayashi and Sakagami 1996; Matsuzaki et al. 2010). Recent structural data demonstrated a critical role of tyrosine sulfation in the recognition of RGFs and PSKs by RGFRs (Song et al. 2016b) and PSKRs (Wang et al. 2015), respectively. The sulfate groups of the sulfated PSK and RGF1 contribute to PSK-PSKR1^{LRR} and RGF1-RGFR1^{LRR} interactions via both polar interactions and van der Waals packing. The results from MST (Microscale Thermophoresis) indicated that PSK displayed about 30-fold higher binding affinity with PSKR than with the desulfated PSK (dPSK). More striking effect of sulfation was observed for binding of RGF1 to its receptor RGFR1, with enhancing about 200-fold binding affinity.

3.6 Co-Receptors with Peptide Hormone-Recognizing Receptors

Increasingly more data support the idea that ligand-induced RK homodimerization or heterodimerization with a co-receptor is an essential step for RK activation, although the mechanisms underlying vary among different RKs (Han et al. 2014;

Song et al. 2016a). In many cases, ligands act as a molecular glue to link an RK with its co-receptor for heterodimerization of two different RKs (Santiago et al. 2016, 2013; Sun et al. 2013a, b; Zhang et al. 2016b). In contrast with these gluing ligands, PSK allosterically induces heterodimerization of its receptor PSKR with SERK member (Wang et al. 2015). Ligand-induced homodimerization of an RK was only demonstrated for SRK9 at both structural and biochemical levels (Ma et al. 2016). SERKs are the co-receptors identified thus far, but recent genetic data suggest that glycosylphosphatidylinositol (GPI)-anchored proteins can function as co-receptors with FERONIA receptor kinase signaling in *Arabidopsis* (Li et al. 2015; Liu et al. 2016), though how GPI is involved in FERONIA-mediated signaling remains unclear.

Several small peptide hormones are known to induce heterodimerization of their receptors with a SOMATIC EMBRYOGENESIS RECEPTOR KINASE (SERK) member (Meng et al. 2016; Song et al. 2016b; Tang et al. 2015; Wang et al. 2015; Zhang et al. 2016b). The structures of the small peptide hormones CLE41 and IDA bound by their receptors together with SERK have been solved recently (Santiago et al. 2016; Zhang et al. 2016b). In the structures, the two peptides act as a molecular glue via their C-terminal sides to mediate interaction of their receptors with a SERK member. Structural comparison showed that the C-terminal sides of the two receptor-bound peptides assume a highly conserved conformation (Fig. 3.6a). The conserved conformation of the two peptides is less likely caused by interaction with a SERK member, as similar conformations are also observed for the C-terminal sides of the receptor-bound RGF and AtPep1 as well (Fig. 3.6a). This indicates that the conserved conformation of these peptide hormones is determined by interaction with their receptors. Interactions of CLE41 or IDA with a SERK are mediated by the penultimate residue that forms two main-chain hydrogen bonds with a conserved N-terminal loop of SERKs (Fig. 3.6b). Collectively, these structural observations suggest that RGFs and AtPeps can also have SERK members as co-receptors with their receptors in signaling complexes. In strong support of this hypothesis, SERKs were shown to form ligand-induced heterodimers with RGFRs (Song et al. 2016b) and PEPR1 (Tang et al. 2015). In vivo studies provided further evidence for a co-receptor role played by SERKs in RGF-induced signaling (Ou et al. 2016). Given the conserved RxR motif among XI LRR-RKs and C-termini of many small peptide hormones, these results raise the possibility that SERKs may also act as co-receptors with these RxR motif containing LRR-RKs. However, future studies are needed to verify or disprove this model.

3.7 Conclusion and Outlook

Recent structural studies have provided significant insight into the mechanisms of peptide hormone recognition by their receptors. The structure-based identification of the conserved RxR motif in a subfamily of LRR-RKs is expected to facilitate finding of their ligands. The structural data also further support the dimerization model of RK activation, which hypothesizes that two kinases from a heterodimeric RKs or a

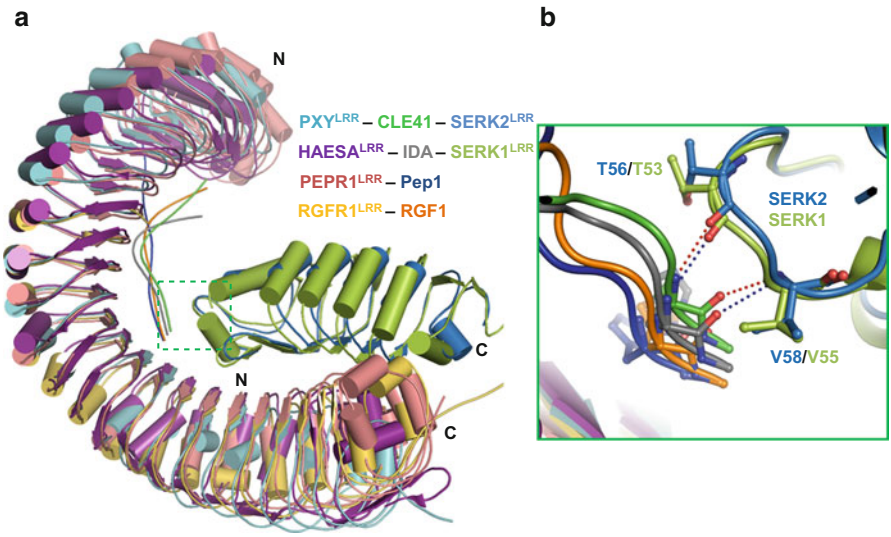


Fig. 3.6 Structural comparison of small peptide hormone-induced LRR-RK/SERK complexes. (**a**) Structure alignment of the PXY^{LRR}-CLE41- SERK2^{LRR}, HAESA^{LRR}-IDA-SERK1^{LRR}, PEPR1^{LRR}-AtPep1, and RGFR1^{LRR}-RGF1 complexes. Color codes are indicated. (**b**) A detailed view of the C-termini of CLE41 and IDA interaction with SERK2^{LRR} and SERK1^{LRR} respectively, which is boxed in green dashed line in "**a**)". The residues involved in interactions are labeled with the same color codes shown in "**a**)"

homodimeric RK are necessary and probably sufficient for their activation. But it remains possible that clustering of the hetero- or homodimeric complexes are required for their full activation as suggested by cell-based study of BRI1-SERK signaling (Bucherl et al. 2013; Wang et al. 2005) and shown for receptor tyrosine kinases (Huang et al. 2016). Despite the progresses made, many questions are still facing structural biologists regarding recognition of peptide hormones. For example, hydroxyproline arabinosylation has been shown important for the activity of the small peptide hormone CLV3 (Shinohara and Matsubayashi 2013; Xu et al. 2015). But how the modification contributes to CLV3-induced signaling still remains elusive. Nearly all the receptors of small peptide hormones verified thus far are LRR-RKs. It is still unknown whether and how the other types of RKs recognize small peptide hormones. Many more CRPs are encoded in the genome of *Arabidopsis*, and their receptors and functions are highly diversified. However, only one structure of the receptor-bound CRP is available. Due to the paucity of structural information, our understanding of the mechanisms underlying CRP recognition is still limited. More efforts directed at structural elucidation of CRPs complexed with their receptors are required to elucidate the mechanisms. These mechanisms in turn are expected to aid in matching of CRPs with their receptors.

References

Amano Y, Tsubouchi H, Shinohara H, Ogawa M, Matsubayashi Y (2007) Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in Arabidopsis. Proc Natl Acad Sci U S A 104:18333–18338

Boggs NA, Dwyer KG, Nasrallah ME, Nasrallah JB (2009) In vivo detection of residues required for ligand-selective activation of the S-locus receptor in *Arabidopsis*. Curr Biol 19:786–791

Bucherl CA, van Esse GW, Kruis A, Luchtenberg J, Westphal AH, Aker J, van Hoek A, Albrecht C, Borst JW, de Vries SC (2013) Visualization of BRI1 and BAK1(SERK3) membrane receptor heterooligomers during brassinosteroid signaling. Plant Physiol 162:1911–1925

Butenko MA, Wildhagen M, Albert M, Jehle A, Kalbacher H, Aalen RB, Felix G (2014) Tools and strategies to match peptide-ligand receptor pairs. Plant Cell 26:1838–1847

Catt KJ, Dufau ML (1977) Peptide hormone receptors. Annu Rev Physiol 39:529–557

Chookajorn T, Kachroo A, Ripoll DR, Clark AG, Nasrallah JB (2004) Specificity determinants and diversification of the *Brassica* self-incompatibility pollen ligand. Proc Natl Acad Sci U S A 101:911–917

Depuydt S, Rodriguez-Villalon A, Santuari L, Wyser-Rmili C, Ragni L, Hardtke CS (2013) Suppression of Arabidopsis protophloem differentiation and root meristem growth by CLE45 requires the receptor-like kinase BAM3. Proc Natl Acad Sci U S A 110:7074–7079

Endo S, Shinohara H, Matsubayashi Y, Fukuda H (2013) A novel pollen-pistil interaction conferring high-temperature tolerance during reproduction via CLE45 signaling. Curr Biol 23:1670–1676

Endo S, Betsuyaku S, Fukuda H (2014) Endogenous peptide ligand-receptor systems for diverse signaling networks in plants. Curr Opin Plant Biol 21:140–146

Fisher K, Turner S (2007) PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. Curr Biol 17:1061–1066

Fritz-Laylin LK, Krishnamurthy N, Tor M, Sjolander KV, Jones JD (2005) Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis. Plant Physiol 138:611–623

Giranton JL, Dumas C, Cock JM, Gaude T (2000) The integral membrane S-locus receptor kinase of Brassica has serine/threonine kinase activity in a membranous environment and spontaneously forms oligomers in planta. Proc Natl Acad Sci U S A 97:3759–3764

Gou X, He K, Yang H, Yuan T, Lin H, Clouse SD, Li J (2010) Genome-wide cloning and sequence analysis of leucine-rich repeat receptor-like protein kinase genes in Arabidopsis thaliana. BMC Genomics 11:19

Grienenberger E, Fletcher JC (2015) Polypeptide signaling molecules in plant development. Curr Opin Plant Biol 23:8–14

Han Z, Sun Y, Chai J (2014) Structural insight into the activation of plant receptor kinases. Curr Opin Plant Biol 20:55–63

Hazak O, Hardtke CS (2016) CLAVATA 1-type receptors in plant development. J Exp Bot 67:4827–4833

Hothorn M, Belkhadir Y, Dreux M, Dabi T, Noel JP, Wilson IA, Chory J (2011) Structural basis of steroid hormone perception by the receptor kinase BRI1. Nature 474:467–471

Huang Y, Bharill S, Karandur D, Peterson SM, Marita M, Shi X, Kaliszewski MJ, Smith AW, Isacoff EY, Kuriyan J (2016) Molecular basis for multimerization in the activation of the epidermal growth factor receptor. elife 5:e14107

Hunter T (2000) Signaling--2000 and beyond. Cell 100:113–127

Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H (2006) Dodeca-CLE peptides as suppressors of plant stem cell differentiation. Science 313:842–845

Ivanov R, Fobis-Loisy I, Gaude T (2010) When no means no: guide to Brassicaceae self-incompatibility. Trends Plant Sci 15:387–394

Iwano M, Takayama S (2012) Self/non-self discrimination in angiosperm self-incompatibility. Curr Opin Plant Biol 15:78–83

Jinn TL, Stone JM, Walker JC (2000) HAESA, an Arabidopsis leucine-rich repeat receptor kinase, controls floral organ abscission. Genes Dev 14:108–117

Kachroo A, Schopfer CR, Nasrallah ME, Nasrallah JB (2001) Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. Science 293:1824–1826

Kang YH, Hardtke CS (2016) Arabidopsis MAKR5 is a positive effector of BAM3-dependent CLE45 signaling. EMBO Rep 17:1145–1154

Kemp BP, Doughty J (2007) *S* cysteine-rich (SCR) binding domain analysis of the *Brassica* self-incompatibility *S*-locus receptor kinase. New Phytol 175:619–629

Kitashiba H, Nasrallah JB (2014) Self-incompatibility in Brassicaceae crops: lessons for interspecific incompatibility. Breed Sci 64:23–37

Kondo T, Sawa S, Kinoshita A, Mizuno S, Kakimoto T, Fukuda H, Sakagami Y (2006) A plant peptide encoded by CLV3 identified by in situ MALDI-TOF MS analysis. Science 313:845–848

Kusaba M, Nishio T, Satta Y, Hinata K, Ockendon D (1997) Striking sequence similarity in inter- and intra-specific comparisons of class I SLG alleles from *Brassica oleracea* and *Brassica campestris*: implications for the evolution and recognition mechanism. Proc Natl Acad Sci U S A 94:7673–7678

Lease KA, Walker JC (2006) The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. Plant Physiol 142:831–838

Li C, Yeh FL, Cheung AY, Duan Q, Kita D, Liu MC, Maman J, Luu EJ, Wu BW, Gates L et al (2015) Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for FERONIA receptor kinase signaling in Arabidopsis. elife 4:e06587

Liu T, Liu Z, Song C, Hu Y, Han Z, She J, Fan F, Wang J, Jin C, Chang J et al (2012) Chitin-induced dimerization activates a plant immune receptor. Science 336:1160–1164

Liu X, Castro C, Wang Y, Noble J, Ponvert N, Bundy M, Hoel C, Shpak E, Palanivelu R (2016) The role of LORELEI in pollen tube reception at the interface of the synergid cell and pollen tube requires the modified eight-cysteine motif and the receptor-like kinase FERONIA. Plant Cell 28:1035–1052

Ma R, Han Z, Hu Z, Lin G, Gong X, Zhang H, Nasrallah JB, Chai J (2016) Structural basis for specific self-incompatibility response in Brassica. Cell Res 26:1320–1329

Marshall E, Costa LM, Gutierrez-Marcos J (2011) Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. J Exp Bot 62:1677–1686

Matsubayashi Y (2014) Posttranslationally modified small-peptide signals in plants. Annu Rev Plant Biol 65:385–413

Matsubayashi Y, Sakagami Y (1996) Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of Asparagus officinalis L. Proc Natl Acad Sci U S A 93:7623–7627

Matsubayashi Y, Ogawa M, Morita A, Sakagami Y (2002) An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. Science 296:1470–1472

Matsubayashi Y, Ogawa M, Kihara H, Niwa M, Sakagami Y (2006) Disruption and overexpression of Arabidopsis phytosulfokine receptor gene affects cellular longevity and potential for growth. Plant Physiol 142:45–53

Matsuzaki Y, Ogawa-Ohnishi M, Mori A, Matsubayashi Y (2010) Secreted peptide signals required for maintenance of root stem cell niche in Arabidopsis. Science 329:1065–1067

Meng X, Zhou J, Tang J, Li B, de Oliveira MV, Chai J, He P, Shan L (2016) Ligand-induced receptor-like kinase complex regulates floral organ abscission in Arabidopsis. Cell Rep 14:1330–1338

Murphy E, Smith S, De Smet I (2012) Small signaling peptides in Arabidopsis development: how cells communicate over a short distance. Plant Cell 24:3198–3217

Naithani S, Chookajorn T, Ripoll DR, Nasrallah JB (2007) Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain. Proc Natl Acad Sci U S A 104:12211–12216

Ni J, Guo Y, Jin H, Hartsell J, Clark SE (2011) Characterization of a CLE processing activity. Plant Mol Biol 75:67–75

Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y (2009) A glycopeptide regulating stem cell fate in Arabidopsis thaliana. Nat Chem Biol 5:578–580

Ou Y, Lu X, Zi Q, Xun Q, Zhang J, Wu Y, Shi H, Wei Z, Zhao B, Zhang X et al (2016) RGF1 INSENSITIVE 1 to 5, a group of LRR receptor-like kinases, are essential for the perception of root meristem growth factor 1 in Arabidopsis thaliana. Cell Res 26:686–698

Pearce G, Strydom D, Johnson S, Ryan CA (1991) A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science 253:895–897

Pearce G, Yamaguchi Y, Munske G, Ryan CA (2008) Structure-activity studies of AtPep1, a plant peptide signal involved in the innate immune response. Peptides 29:2083–2089

Santiago J, Henzler C, Hothorn M (2013) Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. Science 341:889–892

Santiago J, Brandt B, Wildhagen M, Hohmann U, Hothorn LA, Butenko MA, Hothorn M (2016) Mechanistic insight into a peptide hormone signaling complex mediating floral organ abscission. elife 5:e15075

Santner A, Estelle M (2009) Recent advances and emerging trends in plant hormone signalling. Nature 459:1071–1078

Santner A, Calderon-Villalobos LI, Estelle M (2009) Plant hormones are versatile chemical regulators of plant growth. Nat Chem Biol 5:301–307

Sato K, Nishio T, Kimura R, Kusaba M, Suzuki T, Hatakeyama K, Ockendon DJ, Satta Y (2002) Coevolution of the *S*-locus genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa*. Genetics 162:931–940

Sauter M (2015) Phytosulfokine peptide signalling. J Exp Bot 66:5161–5169

Schardon K, Hohl M, Graff L, Pfannstiel J, Schulze W, Stintzi A, Schaller A (2016) Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. Science 354:1594–1597

Schopfer CR, Nasrallah JB (2000) Self-incompatibility. Prospects for a novel putative peptide-signaling molecule. Plant Physiol 124:935–940

She J, Han Z, Kim TW, Wang J, Cheng W, Chang J, Shi S, Wang J, Yang M, Wang ZY et al (2011) Structural insight into brassinosteroid perception by BRI1. Nature 474:472–476

Shimosato H, Yokota N, Shiba H, Iwano M, Entani T, Che FS, Watanabe M, Isogai A, Takayama S (2007) Characterization of the SP11/SCR high-affinity binding site involved in self/nonself recognition in brassica self-incompatibility. Plant Cell 19:107–117

Shinohara H, Matsubayashi Y (2013) Chemical synthesis of Arabidopsis CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. Plant Cell Physiol 54:369–374

Shinohara H, Moriyama Y, Ohyama K, Matsubayashi Y (2012) Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1 reveals diversified ligand recognition mechanisms of plant LRR-RKs. Plant J 70:845–854

Shinohara H, Mori A, Yasue N, Sumida K, Matsubayashi Y (2016) Identification of three LRR-RKs involved in perception of root meristem growth factor in Arabidopsis. Proc Natl Acad Sci U S A 113:3897–3902

Shiu SH, Bleecker AB (2001a) Plant receptor-like kinase gene family: diversity, function, and signaling. Sci STKE 2001:re22

Shiu SH, Bleecker AB (2001b) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci U S A 98:10763–10768

Silverstein KA, Moskal WA Jr, Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 51:262–280

Song W, Han Z, Sun Y, Chai J (2014) Crystal structure of a plant leucine rich repeat protein with two island domains. Sci China Life Sci 57:137–144

Song W, Han Z, Wang J, Lin G, Chai J (2016a) Structural insights into ligand recognition and activation of plant receptor kinases. Curr Opin Struct Biol 43:18–27

Song W, Liu L, Wang J, Wu Z, Zhang H, Tang J, Lin G, Wang Y, Wen X, Li W et al (2016b) Signature motif-guided identification of receptors for peptide hormones essential for root meristem growth. Cell Res 26:674–685

Srivastava R, Liu JX, Howell SH (2008) Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in Arabidopsis. Plant J 56:219–227

Sun Y, Han Z, Tang J, Hu Z, Chai C, Zhou B, Chai J (2013a) Structure reveals that BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. Cell Res 23:1326–1329

Sun Y, Li L, Macho AP, Han Z, Hu Z, Zipfel C, Zhou JM, Chai J (2013b) Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science 342:624–628

Tabata R, Sumida K, Yoshii T, Ohyama K, Shinohara H, Matsubayashi Y (2014) Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. Science 346:343–346

Tager HS, Steiner DF (1974) Peptide hormones. Annu Rev Biochem 43:509–538

Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K (2000) The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. Nature 403:913–916

Takayama S, Isogai A (2005) Self-incompatibility in plants. Annu Rev Plant Biol 56:467–489

Takayama S, Shiba H, Iwano M, Shimosato H, Che FS, Kai N, Watanabe M, Suzuki G, Hinata K, Isogai A (2000) The pollen determinant of self-incompatibility in Brassica campestris. Proc Natl Acad Sci U S A 97:1920–1925

Takayama S, Shimosato H, Shiba H, Funato M, Che FS, Watanabe M, Iwano M, Isogai A (2001) Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. Nature 413:534–538

Takeuchi H, Higashiyama T (2016) Tip-localized receptors control pollen tube growth and LURE sensing in Arabidopsis. Nature 531:245–248

Tamaki T, Betsuyaku S, Fujiwara M, Fukao Y, Fukuda H, Sawa S (2013) SUPPRESSOR OF LLP1 1-mediated C-terminal processing is critical for CLE19 peptide activity. Plant J 76:970–981

Tang J, Han Z, Sun Y, Zhang H, Gong X, Chai J (2015) Structural basis for recognition of an endogenous peptide by the plant receptor kinase PEPR1. Cell Res 25:110–120

Torii KU (2004) Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. Int Rev Cytol 234:1–46

Torii KU (2012) Mix-and-match: ligand-receptor pairs in stomatal development and beyond. Trends Plant Sci 17:711–719

Wang X, Li X, Meisenhelder J, Hunter T, Yoshida S, Asami T, Chory J (2005) Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. Dev Cell 8:855–865

Wang J, Li H, Han Z, Zhang H, Wang T, Lin G, Chang J, Yang W, Chai J (2015) Allosteric receptor activation by the plant peptide hormone phytosulfokine. Nature 525:265–268

Xu C, Liberatore KL, MacAlister CA, Huang Z, Chu YH, Jiang K, Brooks C, Ogawa-Ohnishi M, Xiong G, Pauly M et al (2015) A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat Genet 47:784–792

Yang SL, Xie LF, Mao HZ, Puah CS, Yang WC, Jiang L, Sundaresan V, Ye D (2003) Tapetum determinant1 is required for cell specialization in the Arabidopsis anther. Plant Cell 15:2792–2804

Zhang H, Lin X, Han Z, Qu LJ, Chai J (2016a) Crystal structure of PXY-TDIF complex reveals a conserved recognition mechanism among CLE peptide-receptor pairs. Cell Res 26:543–555

Zhang H, Lin X, Han Z, Wang J, Qu LJ, Chai J (2016b) SERK family receptor-like kinases function as co-receptors with PXY for plant vascular development. Mol Plant 9:1406–1414

Zhao DZ, Wang GF, Speal B, Ma H (2002) The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. Genes Dev 16:2021–2031

Part II
Nuclear Receptors

Chapter 4
Structural Biology of Auxin Signal Transduction

Hongwei Jing and Lucia C. Strader

4.1 Overview of Auxin Signal Transduction

The plant hormone auxin plays a central role in nearly all aspects of plant growth and development, including embryogenesis, organogenesis, hypocotyl elongation, organ polarity establishment, and tropic responses (reviewed in Sauer et al. 2013; Enders and Strader 2015). Extensive genetic and biochemical studies have revealed that auxin-responsive gene expression is mediated by the TRANSPORT INHIBITOR RESPONSE1/AUXIN SIGNALING F-BOX PROTEIN (TIR1/AFB) pathway (reviewed in Salehin et al. 2015). Recent structural data for multiple components of this pathway have illuminated the molecular basis for regulation of auxin-responsive gene expression and suggest new regulatory mechanisms to be explored.

The TIR1/AFBs auxin signal transduction pathway involves three major components: the $SCF^{TIR1/AFB}$ receptors complex, Auxin/INDOLE-3-ACETIC ACID (Aux/IAA) repressor proteins, and AUXIN RESPONSE FACTOR (ARF) transcription factors. The TIR1/AFBs are F-box proteins that form a SKP1–CUL1–F-box (SCF)-type E3 ubiquitin ligase when complexed with the proteins S PHASE KINASE-ASSOCIATED PROTEIN1 (SKP1) (or ASK1 in plants), CULLIN1 (CUL1), and RING BOX 1 (RBX1), called $SCF^{TIR1/AFB}$ (Pickart 2001; Gagne et al. 2002). Auxin-responsive gene transcription is directly regulated by the ARF proteins, which directly bind to auxin response elements (AuxREs) (Guilfoyle and Hagen 2007). When auxin levels are low, Aux/IAA proteins directly repress the ARF transcriptional activity. When auxin levels increase, auxin acts as a "molecular glue" to promote the formation of a complex consisting of a receptor TIR1/AFB and Aux/IAA (Fig. 4.1) (Kepinski and Leyser 2005; Blakeslee et al. 2007). The formation of the SCF^{TIR1}-Aux/IAA co-receptor causes ubiquitination of Aux/IAAs, which

H. Jing · L. C. Strader (✉)
Department of Biology, Washington University in St. Louis, St. Louis, MO, USA
e-mail: strader@wustl.edu

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_4

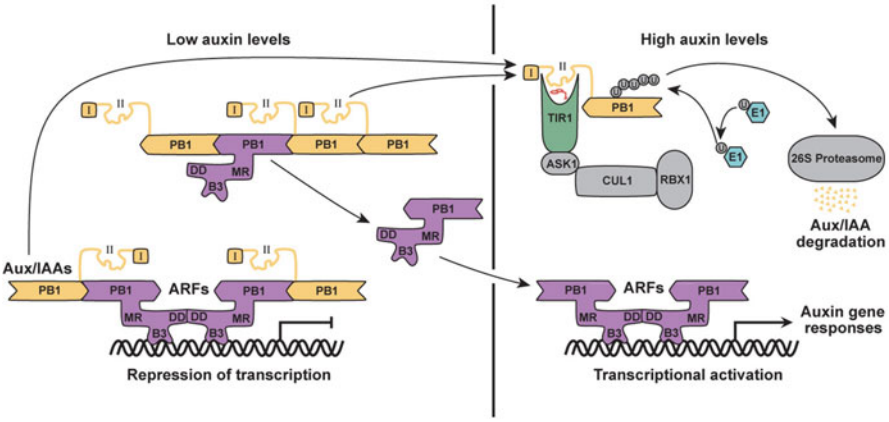


Fig. 4.1 Model for TIR1/AFB-mediated auxin signaling pathway. TIR1 is an F-box protein that directly binds auxin and targets Aux/IAA proteins for degradation. Under low auxin levels, Aux/IAA proteins and ARFs interact directionally, using the basic and acidic interfaces of the PB1 domain (previously called Domain III/IV), thereby repressing the ARF-dependent transcription of auxin response genes. When auxin levels are high, auxin acts as a "molecular glue" to promote ubiquitination and degradation of Aux/IAA proteins through the activity of the 26S proteasome, releasing ARFs from repression and allowing transcriptional activation of target genes

are subsequently subjected to degradation through the 26S proteasome (Gray et al. 2001; Tiwari et al. 2001; Leyser 2006; Mockaitis and Estelle 2008). The degradation of Aux/IAA proteins relieves ARFs from the repressive complex, allowing the ARFs to directly activate or repress the transcription of the downstream auxin response gene (Fig. 4.1) (Gray et al. 2001; Reed 2001; Tiwari et al. 2001; Liscum and Reed 2002).

4.2 TIR1-Aux/IAA Interactions Depend on the Presence of Auxin

Many components of the TIR1/AFB auxin signal transduction pathway were identified through genetic screens for mutants with altered auxin response in Arabidopsis (reviewed in Leyser 2006; Calderon-Villalobos et al. 2010; Salehin et al. 2015). One of these mutants, *tir1*, was isolated by screening for resistance to auxin transport inhibitors *N*-1-naphthylphthalamic acid (NPA), and was deficient in a variety of auxin-regulated growth processes including hypocotyl elongation and lateral root formation (Ruegger et al. 1997). The *TIR1* gene encodes a leucine-rich-repeat (LRR)-containing F-box protein that acts as a subunit of an E3-type ubiquitin–protein ligase and is a key player in auxin responses (Ruegger et al. 1998). Some years later, TIR1 was unequivocally demonstrated to be an auxin receptor by two independent labs (Dharmasiri et al. 2005a; Kepinski and Leyser 2005).

In Arabidopsis, the TIR1/AFB auxin receptor family is comprised of six members, including TIR1 and AUXIN SIGNALING F-BOX 1-5 (AFB1-5) proteins. AFB1 is most closely related to TIR1 and shares 70% identity, whereas AFB2 and AFB3 are 60% identical to TIR1, and AFB4 and AFB5 are approximately 50% identical to TIR1. Functional redundancy exists among the *TIR1/AFB* genes in Arabidopsis (Dharmasiri et al. 2005b; Parry et al. 2009). Genetic analysis of different combinations of *tir1*, *afb1*, *afb2*, and *afb3* mutants revealed that TIR1 and AFB2 play key roles during early seedling development (Parry et al. 2009), resulting in these components becoming the best characterized amongst the TIR1/AFB family.

Aux/IAAs were first identified as genes that were rapidly upregulated after auxin treatment in etiolated soybean (*Glycine max*) and pea (*Pisum sativum*) tissues (Walker and Key 1982; Theologis et al. 1985). Additional *Aux/IAA* genes were subsequently found based on sequence similarity to known *Aux/IAA* genes or in yeast two-hybrid assays with Aux/IAA proteins from various plant species, including tomato (*Lycopersicon esculentum*), tobacco (*Nicotiana tabacum*), rice (*Oryza sativa*), maize (*Zea mays*), and *Arabidopsis thaliana*. The Arabidopsis genome encodes 29 Aux/IAA proteins (Tiwari et al. 2001; Remington et al. 2004; Overvoorde et al. 2005). Expression and phenotypic analyses revealed that many members of Aux/IAA gene family have redundant function (Abel and Theologis 1996; Reed 2001; Remington et al. 2004; Overvoorde et al. 2005). Sequence analysis revealed that Aux/IAA proteins contain four highly conserved amino acid sequence motifs, named domains I, II, III, and IV (Reed 2001). Each domain contributes to the functional properties of the protein. The N-terminal domain I is a repressor domain and acts to recruit the transcriptional corepressor TOPLESS (TPL) (Long et al. 2006; Szemenyei et al. 2008). Domain II is the degron motif and a key determinant of Aux/IAA auxin-dependent degradation because it mediates interaction with the TIR1/AFBs (Tan et al. 2007). Domains III and IV are interaction domains that display high homology with ARF protein domains, also called III and IV, which mediate homo- and heterodimerization between Aux/IAA repressors and ARF transcription factors (Kim et al. 1997; Ulmasov et al. 1997b). Domains III and IV fold into a single globular protein structure and are now called the PB1 domain (see below).

Genetic studies provided important insight into the function of the *Aux/IAA* genes. Screens for Arabidopsis mutants with altered auxin responses or morphological phenotypes have identified many gain-of-function mutations in multiple different *Aux/IAA* genes: *IAA1/AXR5* (*axr5/auxin-resistant 5*) (Yang et al. 2004), *IAA3/SHY2* (*shy2/suppressor of hy2-2*) (Kim et al. 1996; Tian and Reed 1999), *IAA6/SHY1* (*shy1/suppressor of hy2-1*) (Kim et al. 1996; Liscum and Reed 2002), *IAA7/AXR2* (*axr2/auxin-resistant 2*) (Wilson et al. 1990; Nagpal et al. 2000), *IAA12/BDL* (*bdl/bodenlos*) (Hamann et al. 1999; Hamann et al. 2002), *IAA14/SLR* (*slr/solitary-root*) (Fukaki et al. 2002), *IAA16* (*iaa16-1*) (Rinaldi et al. 2012), *IAA17/AXR3* (*axr3/auxin-resistant 3*) (Leyser et al. 1996; Rouse et al. 1998), *IAA18/CRANE* (*iaa18/crane*) (Uehara et al. 2008), *IAA19/MSG* (*msg2/massugu2*) (Tatematsu et al. 2004), and *IAA28* (*iaa28-1*) (Rogg et al. 2001). In each case, the IAA domain II variant in these mutants results in increased Aux/IAA protein stability, suggesting that they act

as a negative regulators of auxin signaling and that their degradation is essential for normal auxin response.

Although the *tir1* mutant was identified as early as 1997, the connection between SCF^{TIR1}, Aux/IAA degradation, and auxin perception was not understood for several years. In 2001, genetic and biochemical studies showed that auxin regulates SCF^{TIR1}-dependent degradation of Aux/IAA proteins (Gray et al. 2001; Zenser et al. 2001). The interaction between the SCF^{TIR1} and Aux/IAA was demonstrated in pull-down assays (Gray et al. 2001); these interactions were dependent on the conserved domain II of Aux/IAA proteins (Gray et al. 2001). Further, experiments with Aux/IAA-luciferase or -GUS fusion proteins demonstrated that domain II is sufficient for auxin-dependent degradation (Zenser et al. 2001). These results suggested that auxin treatment stimulated the interaction between the SCF^{TIR1} and Aux/IAA proteins to promote their degradation (Gray et al. 2001; Zenser et al. 2001); however, the auxin-dependence of this mechanism was still unclear. Later studies clearly demonstrated that SCF^{TIR1} directly interacts with Aux/IAA proteins in an auxin-dependent manner (Kepinski and Leyser 2004; Dharmasiri et al. 2005a; Dharmasiri et al. 2005b; Kepinski and Leyser 2005) to form an auxin co-receptor complex (Tan et al. 2007).

4.3 TIR1 Structural Data Uncovered a New Mechanism for Ligand Perception

Structural studies revealed the unexpected, yet elegant, auxin perception mechanism (Tan et al. 2007). The Arabidopsis TIR1-ASK1 complex was crystallized in a complex with a 17-amino-acid peptide from IAA7 domain II in the presence of various auxins (Tan et al. 2007).

The TIR1-ASK1 crystal structure revealed a mushroom-shaped overall structure with the 18 TIR1 leucine-rich-repeat (LRR) domains being the "cap" and the TIR1 F-box motif and ASK1 forming the "stem" (Fig. 4.2a, b). The F-box domain is a nearly 40-residue motif at the TIR1 amino terminus that forms a three-helix bundle, extensively interacting with the carboxy-terminal four helices of ASK1. Immediately following the F-box domain, the 18 LRRs domain of TIR1 fold into a twisted horseshoe-shaped solenoid. Strikingly, an unexpected inositol hexakisphosphate ($InsP_6$) molecule is found near the center of the TIR1-LRR fold in close vicinity of the auxin-binding site. The top surface of the TIR1-LRR domain has a single surface pocket that binds both auxin and the domain II peptide. In the crystal structure, auxin docks to the bottom of the TIR1 pocket, whereas the IAA7 peptide sits on the top of auxin and covers up the pocket (Fig. 4.2b). Overall, TIR1-LRR and the conserved Aux/IAA degron peptide sandwich auxin in the middle, which nucleates a hydrophobic core among these three molecules (reviewed in Calderon-Villalobos et al. 2010). This structure, combined with the requirement for auxin in mediating the TIR1/AFB-Aux/IAA interaction, has led to auxin being called the "molecular glue" of the TIR1-auxin-Aux/IAA interaction.

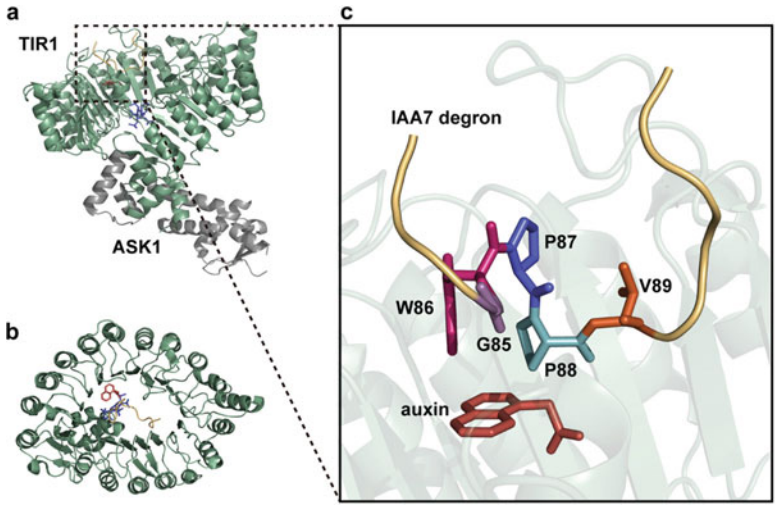


Fig. 4.2 Structure of the TIR1-ASK1 complex with IAA and the IAA7 degron peptide. (**a**) Overview of the complex of TIR1 (green), ASK1 (gray), IAA7 degron peptide (yellow), IAA (red) and InsP6 (blue). (**b**) Top view of the complex of TIR1 (green), IAA7 degron peptide (yellow), IAA (red) and InsP6 (blue). (**c**) Detailed view of the auxin-binding pocket, showing the conserved GWPPV motif of the IAA7 degron peptide and IAA, with TIR1 in the background

The TIR1-LRR domain consists of 18 LRRs and a carboxy-terminal cap sequence. Each TIR1-LRR contains a β-strand followed by an α-helix, and assembles into the expected solenoid structure with an overall horseshoe-like shape. The intra-repeat loops associated with LRRs 2,12, and 14 are unusually long and located at the top surface of TIR1-LRR solenoid. Of the three loops, the LRR-2 loop plays a very important role in constructing the auxin- and substrate-binding surface pocket by interacting with the nearby concave surface of the TIR1-LRR solenoid (Tan et al. 2007).

As revealed in the TIR1-IAA-Aux/IAA degron peptide, TIR1 recognizes auxin through the bottom portion of the surface pocket formed between the LRR-2 loop and the solenoid inner surface. IAA binds to the TIR1 pocket via two very important functional moieties, the side-chain carboxyl group and the indole ring. The carboxyl group of IAA anchors the plant hormone to the bottom of the TIR1 pocket by forming a salt bridge and two hydrogen bonds with two residues (Arg403 and Ser438) from the pocket floor. Meanwhile, the indole ring of IAA stacks on top of the pocket floor with its edge packing against the surrounding walls through hydrophobic interactions and van der Waals contacts. On the loop-2 side of the TIR1 pocket, the benzene region of the IAA indole ring interacts with two TIR1 phenylalanine residues (Phe79 and Phe82). On the concave surface side, the rest of the IAA indole ring is partially sandwiched between two parallel layers of TIR1 residues and is therefore mainly in contact with the TIR1 polypeptide backbone (Tan et al. 2007).

Upon binding to TIR1, the Aux/IAA degron peptide is docked above the auxin molecule, completely enclosing the TIR1 pocket. The IAA7 degron consists of

predominantly hydrophobic residues and binds to the TIR1-auxin pocket through extensive hydrophobic interactions. The IAA7 peptide adopts a highly coiled conformation so that the highly conserved central hydrophobic consensus motif Gly-Trp-Pro-Pro-Val (GWPPV) is positioned to cover the entire auxin molecule (Tan et al. 2007). Genetic screens have identified many mutations in this GWPPV motif that result in Aux/IAA proteins stabilization (Salehin et al. 2015). In the crystal, the tryptophan and the second proline in this GWPPV motif, interact with the surrounding hydrophobic wall of the TIR1 pocket and stack against the auxin molecule lying underneath, packing against the indole ring and the side chain of IAA (Tan et al. 2007). The positions of these two conserved residues are partially maintained by the first proline in the middle, which itself also forms hydrophobic interactions with surrounding TIR1 residues. The first glycine residue is also located at a critical position, where flexibility of the peptide is required for the N-terminal region of the substrate peptide to take a sharp turn and continue interacting with TIR1. In the structure, the valine residue at the end of the GWPPV motif shows a conserved hydrophobic feature, which also plays a crucial role in its interactions with the nearby hydrophobic residues of TIR1 (Fig. 4.2c) (Tan et al. 2007). Thus, auxin nucleates a hydrophobic core together with the GWPPV motif and TIR1 pocket, which provides the energy basis for enabling the high affinity interaction between TIR1 and Aux/IAA (reviewed in Calderon-Villalobos et al. 2010).

Crystallographic analysis of the TIR1-ASK1 complex revealed an island of unexpected electron density, which belongs to an Inositol hexakisphosphate ($InsP_6$) molecule co-purified with the TIR1-ASK1 protein complex from insect cells. In the crystal, the molecule $InsP_6$ is tightly surrounded by more than ten conserved positively charged residues at the concave surface of the TIR1-LRR domain. $InsP_6$ interacts with the auxin-binding pocket of TIR1 from underneath and is in direct contact with the basic residue binding to the carboxyl group of the auxin. The high affinity and the binding mode of $InsP_6$ at the core of the auxin receptor strongly suggest that it is a functional cofactor of TIR1 (Tan et al. 2007). However, the role of $InsP_6$ in the auxin signaling pathway is still a challenge for future study. The "molecule glue" model provided by the TIR1-auxin-Aux/IAA structure has been expanded to explain many other receptor–ligand interactions.

4.4 A New Model for F-Box Interaction with Ligands and Targets

In plants, protein degradation mediated by the ubiquitin–proteasome system is involved in various phytohormone signaling pathways, including auxin, jasmonic acid (JA), gibberellin acid (GA), strigolactone (SL), abscisic acid (ABA), and ethylene. For some of these pathways, the activation of signal transduction uses a relief-of-repression mechanism that is strikingly similar to that of auxin. In presence of auxin, $SCF^{TIR1/AFB}$ E3 ligases directly induce the degradation of Aux/IAA transcriptional repressors, releasing ARF transcription factors and thus allowing

auxin-mediated gene transcription (Fig. 4.1). In the presence of JA-Ile, SCF^{COI1} directs the degradation of JAZ transcriptional repressors, releasing transcription factors and thus inducing JA-mediated transcription (discussed in Chap. 5). Growth-repressive DELLA proteins are targeted for degradation by $SCF^{SLY1/SNZ}$ E3 ligases in response to GA perception (discussed in Chap. 6). Strigolactone and karriken signaling are similarly mediated by the F-box protein MAX2 (Arabidopsis) or DWARF3 (rice), which promote proteolysis of SUPPRESSOR OF MAX2-1 (SMAX1) or DWARF53 proteins to modulate gene expression (discussed in Chap. 7).

The ubiquitin–proteasome system is also important for human health and has been implicated in numerous human diseases, including cancer, cardiovascular disease, diabetes, immunological disorders, and neurological disorders (reviewed in Hussain et al. 2016). Indeed, E3 ubiquitin ligases have emerged as valid drug targets for the development of novel anticancer therapeutics. Bortezomib (also known as Velcade or PS-341) is the first general proteasome inhibitor approved by the Food and Drug Administration for treatment of relapsed/refractory multiple myeloma and mantle cell lymphoma (Colson et al. 2004; Orlowski and Kuhn 2008). Further, the cullin neddylation inhibitor MLN4924 is a cancer treatment (Soucy et al. 2009) that inactivates cullin-RING ligases (CRLs) to cause the accumulation of CRL substrates, ultimately triggering cell cycle arrest, senescence, and/or apoptosis to suppress the cancer cell growth (Luo et al. 2012; Zhao et al. 2014). Thus far, MLN4924 has shown promising anticancer activity in a broad spectrum of in vitro and in vivo preclinical tumor models, as well as in phase I clinical trials (Zhao et al. 2014; Wu and Yu 2016). Recently, Petzold et al. (2016) reported the structural basis of lenalidomide-induced casein kinase 1α (CK1α) degradation by the CUL4-RBX1-DDB1-CRBN ($CRL4^{CRBN}$) E3 ubiquitin ligase. Structural results show that CK1α binding to $CRL4^{CRBN}$ is strictly dependent on the presence of a lenalidomide, providing a mechanistic insight into how small molecules exploit an ubiquitin ligase for the destabilization of therapeutic targets (Petzold et al. 2016), using a mechanism similar to the auxin signaling pathway. The molecular glue model of a ligand promoting interaction of an E3 ligase with a substrate may allow future discovery of specific drugs that target a particular subset of SCF-controlled pathways to eliminate disease-causing proteins with minimal cell toxicity.

4.5 Regulation of $SCF^{TIR1/AFB}$-Aux/IAA Co-receptor Formation

In Arabidopsis, six TIR1/AFBs and 29 Aux/IAAs may participate in auxin co-receptor formation. Different TIR1/AFB-Aux/IAA combinations display distinct affinities for one another and for different natural and synthetic auxins (Calderón Villalobos et al. 2012; Lee et al. 2014; Shimizu-Mitao and Kakimoto 2014). Both

in vitro pull-down and yeast two-hybrid protein interaction experiments revealed that different TIR1/AFBs displayed varied affinity for the same Aux/IAA (Parry et al. 2009; Calderón Villalobos et al. 2012). In addition, using an auxin signaling system recapitulated in yeast, Havens et al. (2012) found that Aux/IAA protein degradation rate differed whether TIR1 and AFB2 participated in the co-receptor, suggesting that TIR1 and AFB2 have very different activities even with the same Aux/IAA protein (Havens et al. 2012). In addition, Aux/IAA protein degradation assays indicate that different Aux/IAAs display distinct degradation rates (Calderón Villalobos et al. 2012). Furthermore, different combinations of TIR1 and Aux/IAA proteins form co-receptor complexes with a wide range of auxin-binding affinities (Calderón Villalobos et al. 2012). Taken together, various combinations of TIR1/AFB and Aux/IAA proteins with a variety of auxin affinities contribute to the ability of auxin to regulate a dynamic range of plant processes.

$SCF^{TIR1/AFB}$-Aux/IAA co-receptor formation can be regulated by posttranslational modifications. In 2012, Terrile et al. reported that nitric oxide (NO) influences auxin signal transduction through *S*-nitrosylation of the auxin receptor TIR1 in Arabidopsis; NO donors increase auxin-dependent gene expression and enhance TIR1-Aux/IAA interaction (Terrile et al. 2012). In addition, Jing et al. (2015) reported that isomerization of a conserved proline residue in the Aux/IAA degron affects Aux/IAA recruitment to the $SCF^{TIR1/AFB}$-Aux/IAA complex, illustrating another layer of complexity to auxin signaling (Jing et al. 2015). Mutations in the rice *LATERAL ROOTLESS2* (*LRT2*), which encodes a cyclophilin-type peptidyl-prolyl *cis/trans* isomerase (PPIase) that catalyzes the isomerization of peptide bonds at proline residues, resulted in auxin resistance and lateral root defects. Biochemical and genetic studies showed that LRT2 catalyzes the *cis/trans* isomerization of rice OsIAA11 at the Trp^{104}-Pro^{105} peptide bond, accelerating the Pro^{105}-*cis* conformer formation, and thereby facilitating their binding to the auxin receptor OsTIR1 for forming an $SCF^{TIR1/AFB}$-Aux/IAA complex (Jing et al. 2015). This new discovery of cyclophilin isomerase enzyme LRT2 adds a new twist to regulation of auxin signal transduction (Su et al. 2015). Furthermore, increased temperature promotes rapid accumulation of the auxin receptor TIR1, dependent on the molecular chaperone HSP90 (Wang et al. 2016). Protein interaction data reveal that HSP90 and the co-chaperone SGT1 each interact with TIR1, suggesting that TIR1 is in a complex with HSP90 and SGT1. Inhibition of HSP90 activity results in TIR1 degradation and defects in auxin-regulated processes (Wang et al. 2016), suggesting that HSP90 regulates temperature-dependent seedling growth by stabilizing the auxin receptor TIR1. Furthermore, TIR1 oligomeric state may be of importance for its activity. TIR1 is capable of oligomerization *in planta*, and mutations that abolish TIR1 oligomerization impair the degradation of SCF^{TIR1} substrates and fail to complement the *tir1* mutant (Dezfulian et al. 2016), suggesting that TIR1 homo-oligomerization may play an important role in regulation of SCF^{TIR1} function and auxin signaling. In summary, posttranslational modifications of auxin signaling pathway components and oligomerization of these components may fine-tune the auxin response.

4.6 ARF DNA-Binding Domain Structure Uncovers a Dimerization Domain

ARF proteins play critical roles in the auxin response pathway. ARF1 was first discovered in a yeast one-hybrid system with an Arabidopsis cDNA expression library to screen for transcription factors that recognize the Auxin Response Element (AuxRE) TGTCTC (Ulmasov et al. 1997a). After subsequent genetic, genomic, and molecular studies, 22 *ARF* genes and one *ARF* pseudogene have been identified from Arabidopsis (Liscum and Reed 2002; Guilfoyle and Hagen 2007). Most ARF proteins consist of an N-terminal B3-type DNA-binding domain (DBD), a variable middle region that functions as an activation domain (AD) or repression domain (RD), and a carboxy-terminal interaction domain (PB1 domain, formerly called domains III and IV) (reviewed in Chandler 2016; Li et al. 2016). The ARF DBD is a plant-specific B3-type protein domain that efficiently binds to AuxREs (Ulmasov et al. 1999b; Guilfoyle and Hagen 2001) as a dimer (Boer et al. 2014). The ARF protein variable AD and RD regions are located carboxy-terminal to the DBDs and contain biased amino acid sequences that confer either transcriptional activation or repression activity. ADs are enriched in glutamine (Q), serine (S), and leucine (L), whereas RDs are enriched in serine (S) and in some cases proline (P), leucine (L), and/or glycine (G) residues (Ulmasov et al. 1999b). These middle-region amino acids play critical roles in determining ARF function, with S-rich ARFs acting as transcriptional repressors and Q-rich ARFs acting transcriptional activators (Tiwari et al. 2003; Guilfoyle and Hagen 2007). In Arabidopsis, five ARF proteins (ARF5, ARF6, ARF7, ARF8, and ARF19) are considered transcriptional activators, whereas ARF2, ARF3, ARF4, and ARF9 have been shown experimentally to function as transcriptional repressors (Ulmasov et al. 1999a; Tiwari et al. 2003), and the remaining Arabidopsis ARF proteins are hypothesized to also act as transcriptional repressors, based on the similarity of their variable region to the characterized repressors.

In 2014, Boer et al. determined high-resolution crystal structures of the DNA-binding domains of the repressing ARF1 and activating ARF5, as well as complexes between DNA and the ARF1 and ARF5 DBDs (Fig. 4.3) (Boer et al. 2014). These structures revealed that the ARF DBDs are composed of three distinct structural domains. In addition to the predicted B3 domain (residues 120-226 in ARF1-DBD and 154-260 in ARF5-DBD), which folds in a seven-stranded open β barrel structure, the regions flanking the B3 domain form a second domain that facilitates ARF DBD dimerization, called the dimerization domain (DD). Thus, the B3 domain is flanked by the DD. In addition, the DBD C-terminal residues folds in a small five-stranded β-barrel-like structure to form an ancillary domain (AD) that tightly interacts with the DD (Boer et al. 2014). Crystallographic data suggested that the ARF-DBDs might homodimerize through their DD domain and small-angle X-ray scattering (SAXS) experiments verified that ARF DBD dimerization occurs in solution (Boer et al. 2014). Mutational analysis studies suggest that DD interface

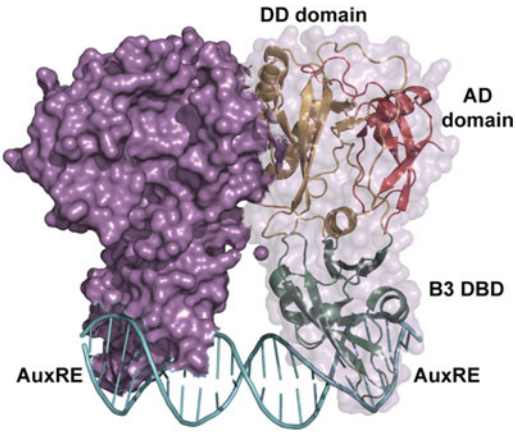


Fig. 4.3 Structure of the ARF1 dimerized DNA-binding domain with ER7 complex. ARF1 dimerized DNA-binding domain (DBD) bound to the ER7 everted repeat containing two AuxREs is shown. The left ARF1 DBD monomer is a surface rendering whereas the right monomer is shown as a ribbon diagram. Each ARF1 DBD monomer is composed of three distinct structural domains, named as B3 domain (green), the dimerization domain (DD, yellow), and the ancillary domain (AD, red)

residues are required for ARF5 transcriptional activity *in planta*, consistent with the possibility that ARF DBD dimerization is required for ARF activity.

To determine the structural basis for DNA binding by ARF proteins, the authors co-crystallized ARF1-DBD and a double-stranded oligonucleotide containing ER7, an everted repeat of the AuxRE element spacing by seven nucleotides (Ulmasov et al. 1997a), revealing that the ARF1 DNA-binding interface is located at the tips of the U-shaped dimer created by the DD interaction (Boer et al. 2014). The two B3 domains bind to the inverted AuxRE TGTCTC elements located at either extreme of the oligonucleotide, and the connecting DNA sequence bridges the gap between the B3 domains. Binding of two AuxRE sites by an ARF dimer generates cooperative DNA-binding behavior (Boer et al. 2014). Further studies demonstrated that the ARF1 and ARF5 B3 domains do not have qualitatively distinct DNA-binding specificity but rather bind to the same spectrum of motifs with quantitatively different efficiencies. Thus, ARF proteins DBD dimerization contributes to DNA-binding affinity, but not to the specificity of DNA motif recognition. In addition to quantitative differences at the level of binding sites, ARF1 and ARF5 differ markedly in their ability to bind complex motifs depending on the spacing of the two binding sites (Boer et al. 2014). In summary, structure–function analysis indicates the DNA-binding domain as an ARF dimerization domain, suggests that ARF dimers bind complex sites as molecular calipers with ARF-specific spacing preference, and provides an atomic-scale explanation for DNA-binding specificity in auxin response.

4.7 ARF and Aux/IAA PB1 Domain Structures Suggest Multimerization Amongst These Proteins

ARF and Aux/IAA proteins interaction are central to auxin signaling and occur through a conserved C-terminal motif (Kim et al. 1997; Guilfoyle and Hagen 2012). Bioinformatic analysis of ARF and Aux/IAA proteins suggested this C-terminal region may form a type I/II Phox and Bem1p (PB1) protein–protein interaction domain (Guilfoyle and Hagen 2012). Recently, crystallographic evidence of the C-terminal regions of ARF5 and ARF7 confirms that domain III/IV adopts a PB1 fold to mediate interaction with ARF and Aux/IAA proteins in Arabidopsis (Fig. 4.4a, b) (Korasick et al. 2014; Nanao et al. 2014). Further, NMR structures of the PB1 domains of Arabidopsis IAA17 and *Pisum sativum* IAA4 conformed to a similar architecture in these repressor proteins (Han et al. 2014; Dinesh et al. 2015).

The ARF7 C-terminal region adopts a canonical PB1 fold with slight modifications (Korasick et al. 2014). The N-terminal domain III consists of an antiparallel β-sheet (β1–β2) and α1, and the conserved lysine (K1042) is on the surface-exposed face of β1. The C-terminal domain IV contains a second antiparallel β-sheet (β3–β5) and two α-helices (α2 and α3), and the cluster of acidic residues forming the OPCA-like motif is located on the loops flanking β4. This PB1 scaffold places the lysine (K1042) and OPCA motif residues (D1092, E1094, D1096, and D1102) on the opposite faces of the ARF7PB1 structure (Fig. 4.4c, d) (Korasick et al. 2014). Previous studies show that PB1 domains adopt a β-grasp fold and may display an acidic surface (type I), a basic surface (type II), or both surfaces (type I/II) on opposite faces of the domain structure to allow for front-to-back orientation of multiple PB1 domains (Noda et al. 2003; Sumimoto et al. 2007). Crystal structure data confirmed this analysis, and the results show that the ARF7PB1 electrostatic surface potential reveals positive and negative interaction interfaces, containing the invariant lysine (K1042) and OPCA motif, respectively (Fig. 4.4c, d) (Korasick et al. 2014).

The crystal structure of the domain III/IV of ARF5 also shows strong structural similarity with the PB1 domain (Nanao et al. 2014). In the crystal, domain III/IV of ARF5 form oligomers in which monomers interact with each other in a head-to-tail manner, with each monomer interaction occurring through two charged interfaces: one generally positive and another generally negative (Nanao et al. 2014). ARF5 K797 from the positive side of one monomer interacts with D847, E849, and D851 of the OPCA motif of another monomer as similarly described for the ARF7PB1 (Korasick et al. 2014; Nanao et al. 2014). In addition, the Arabidopsis IAA17 domain III/IV structure exhibits a compact β-grasp fold with a highly dynamic insert helix (Han et al. 2014). IAA17 and ARF5 associate to form homo- or hetero-oligomers using a common scaffold and binding interfaces (Han et al. 2014). Furthermore, the PB1 domain of *Pisum sativum* PsIAA4 features two distinct surface patches of oppositely charged amino acid residues, mediating front-to-back multimerization via electrostatic interactions (Dinesh et al. 2015). Mutations of conserved basic or acidic residues on either face suppressed PsIAA4 PB1

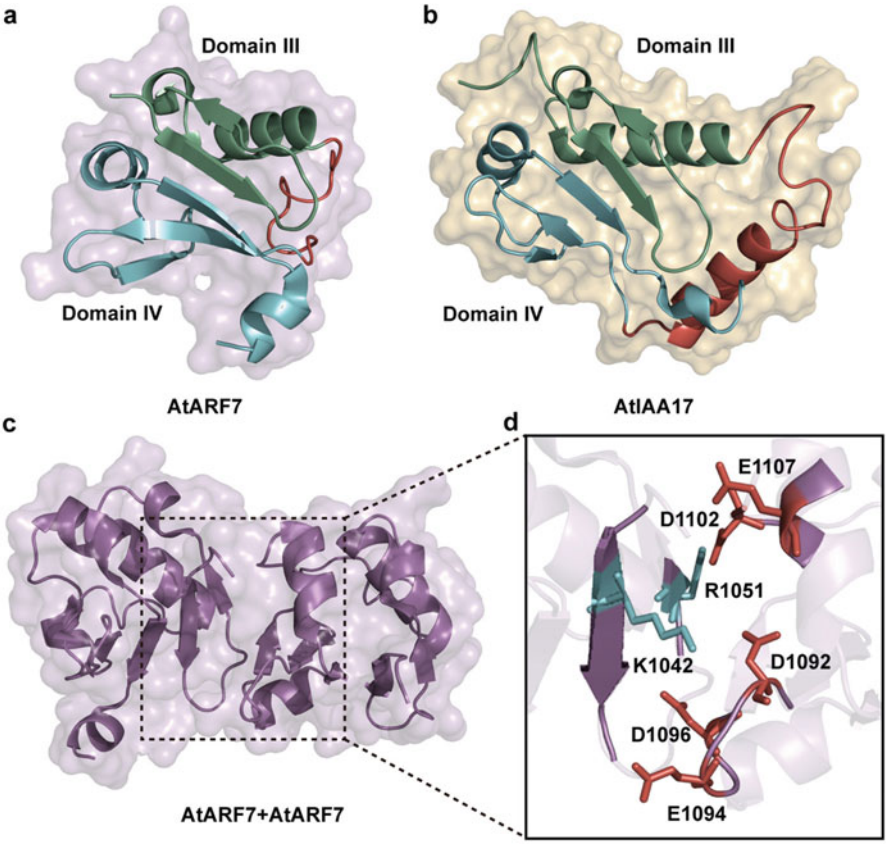


Fig. 4.4 Structures of the ARF7 and IAA17 PB1 domains and the protein–protein interaction. (**a**, **b**) The structures of the PB1 domain of ARF7 and Aux/IAA17, which were before called domains III (green), and IV (blue), show a common fold. (**c**) ARF7 PB1 domain interactions are driven by the conserved residues on opposing positive (+) and negative (−) interaction surfaces. (**d**) Detail of the ARF7 dimer interaction surface shows the conserved positive K1042 (red) and the negative D1092, E1094, D1096, and D1102 (green) participate in charge–charge interactions

homo-oligomerization in vitro, confirming directional interaction of full-length PsIAA4 in vivo (Dinesh et al. 2015). Altogether, PB1 domain interactions are driven by electrostatic differences on either face of these domains to mediate interactions among ARF and Aux/IAA proteins.

The ARF7 PB1 domain crystallized with 16 molecules in the asymmetric unit (Korasick et al. 2014). In this arrangement, the invariant ARF7PB1 lysine orients toward the OPCA motif of the next PB1 domain to form a curved helix topology. Several chains within the ARF7 PB1 domain crystal structure packed together to form a pentameric arrangement of ARF7PB1 molecules with the negative charges toward the outer side of the curve and positive charges along the inner face and places the PB1 domain N-terminal side staggered along the curved multimeric outer

face (Korasick et al. 2014). In addition, heterologously expressed ARF5 (Nanao et al. 2014), ARF7 (Korasick et al. 2014), IAA17 (Han et al. 2014), and PsIAA4 (Dinesh et al. 2015) behave as higher-order oligomers in solution. Thus, ARF and Aux/IAA proteins may form PB1 domain-mediated oligomers; however, functional roles for this oligomerization have not yet been fully elucidated. Overexpressing the stabilized iaa16-1 protein results in restricted growth in Arabidopsis, whereas overexpressing the stabilized PB1 domain variants *iaa16-1*^{K122A} or *iaa16-1*^{opca} results in wild-type phenotypes (Korasick et al. 2014), consistent with the possibility that IAA16 multimerization with target ARF proteins is necessary for repressor activity. Further, overexpressing IAA17 in Arabidopsis protoplasts strongly inhibits auxin responsiveness, whereas overexpression of IAA17 variants capable of interacting with targets with a single PB1 domain electrostatic face results in decreased repression (Nanao et al. 2014). Conversely, overexpression of IAA19 PB1 single-face variants in Arabidopsis protoplasts (Nanao et al. 2014) or expression of iaa14-1 PB1 single-face variants (Pierre-Jerome et al. 2016) had little effect on repressive function, suggesting that IAA multimerization is not necessary for the activity for all Aux/IAA proteins. Why multimerization appears to be necessary for function of some, but not all, IAA proteins remains an open question.

In plants, TOPLESS (TPL) and TOPLESS-related (TPR) proteins predominantly function as corepressors through interaction with transcription factors to regulate plant development, stress responses, and hormone signal transduction (Causier et al. 2012). TPL/TPR corepressors interact with small ethylene response factor associated amphiphilic repression (EAR) motifs (amino acid sequence LxLxL) in diverse transcriptional repressors (Kagale et al. 2010; Kagale and Rozwadowski 2011). In the year 2015, Ke et al. reported the structural and biochemical studies of interactions between TPL proteins and EAR motifs (Ke et al. 2015). The OsTPR2 TOPLESS Domain (TPD) forms an extended tetrameric structure. The TPD monomers form a fold composed of nine α helices (α1 to α9), two short 3_{10} helices, and connecting loops (Ke et al. 2015). Structure results show that the tetramer of TPD is a dimer of dimers. The first dimerization interface is mediated by helices α1, α2, and α9 from each monomer, and the second interface is formed by a new dimerization motif, in which the antiparallel helices α6 and α7 from each monomer pack perpendicularly against helices α6 and α7 from the neighboring monomer (Ke et al. 2015).

To better understand the atomic detail of the interaction between TPL/TPD and LxLxL EAR motifs, the author crystallized OsTPR2 TPD in complex with the NINJIA EAR motif (Ke et al. 2015). Structures show that each EAR motif peptide binds to a groove linked by hydrophobic and positively charged residues contained within each OsTPR2 TPD monomer; thus, four NINJA EAR motif peptides are bound by one OsTPR2 tetramer (Ke et al. 2015). To test whether the EAR-binding mechanism is conserved, the author also determined the structures of OsTPR2 TPD complexed with EAR motifs from IAA1 and IAA10 (Ke et al. 2015), revealing that the TPD conformations and overall binding modes are similar amongst these complexes. Thus, LxLxL-type EAR motifs likely share a conserved mode of TPD interaction mediated by hydrophobic interactions between the conserved leucine

residues of the EAR motif and the highly conserved hydrophobic and positively charged cleft residues of the TPD. Taken together, sequence alignments and structure-based mutagenesis data indicate that the TPL/TPD corepressor-binding mode is highly conserved in many transcriptional repressors (Ke et al. 2015), thus providing a general mechanism for gene repression mediated by the TPL corepressors to regulate a multiple aspects of plant development, including many plant hormone signaling pathways.

4.8 Summary

Auxin is a key regulator of plant growth and development. Over the past 20 years, extensive molecular and genetic techniques identified the major components in auxin signal transduction pathway. Recently, structural and biophysical studies have clarified and refined our understanding of the auxin signal transduction pathway. The first structure of TIR1-auxin-Aux/IAA complex revealed auxin as a "molecular glue" to mediate TIR1 and Aux/IAA interaction. Afterward, ARF dimerization through DNA-binding domains provided an atomic-scale mechanistic model for DNA-binding specificity. ARF and Aux/IAA PB1 domain structures provided insight into how protein interactions may attenuate auxin signal transduction. In addition, TPL/TPD and EAR motifs define the molecular basis of how the TPL family corepressors interact with and are recruited by diverse repressors to regulate plant development and hormone signaling pathways. In summary, these structure studies provide new insights into molecular mechanism of the auxin response pathway. However, the complexity of combining different auxin signaling components, multimerization of transcription factors, protein posttranslational modifications, and interaction of auxin signaling components with other signaling pathways may all contribute to the diversity, complexity, and specificity of auxin response. In the future, combining structural research, genetic, genomics, and biochemical studies on complexes involved in auxin signaling will help to gain new insights into molecular mechanisms of auxin signaling pathway.

References

Abel S, Theologis A (1996) Early genes and auxin action. Plant Physiol 111:9–17

Blakeslee JJ, Bandyopadhyay A, Lee OR, Mravec J, Titapiwatanakun B, Sauer M, Makam SN, Cheng Y, Bouchard R, Adamec J et al (2007) Interactions among PIN-FORMED and P-glycoprotein auxin transporters in *Arabidopsis*. Plant Cell 19:131–147

Boer DR, Freire-Rios A, van den Berg WAM, Saaki T, Manfield IW, Kepinski S, Lopez-Vidrieo I, Franco-Zorrilla JM, de Vries SC, Solano R et al (2014) Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. Cell 156:577–589

Calderón Villalobos LIA, Lee S, De Oliveira C, Ivetac A, Brandt W, Armitage L, Sheard LB, Tan X, Parry G, Mao HB et al (2012) A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. Nat Chem Biol 8:477–485

Calderon-Villalobos LI, Tan X, Zheng N, Estelle M (2010) Auxin perception-structural insights. Cold Spring Harb Perspect Biol 2:a005546

Causier B, Lloyd J, Stevens L, Davies B (2012) TOPLESS co-repressor interactions and their evolutionary conservation in plants. Plant Signal Behav 7:325–328

Chandler JW (2016) Auxin response factors. Plant Cell Environ 39:1014–1028

Colson K, Doss DS, Swift R, Tariman J, Thomas TE (2004) Bortezomib, a newly approved proteasome inhibitor for the treatment of multiple myeloma: nursing implications. Clin J Oncol Nurs 8:473–480

Dezfulian MH, Jalili E, Roberto DK, Moss BL, Khoo K, Nemhauser JL, Crosby WL (2016) Oligomerization of SCF^{TIR1} is essential for Aux/IAA degradation and auxin signaling in Arabidopsis. PLoS Genet 12:e1006301

Dharmasiri N, Dharmasiri S, Estelle M (2005a) The F-box protein TIR1 is an auxin receptor. Nature 435:441–445

Dharmasiri N, Dharmasiri S, Weijers D, Lechner E, Yamada M, Hobbie L, Ehrismann JS, Jürgens G, Estelle M (2005b) Plant development is regulated by a family of auxin receptor F box proteins. Dev Cell 9:109–119

Dinesh DC, Kovermann M, Gopalswamy M, Hellmuth A, Calderon Villalobos LI, Lilie H, Balbach J, Abel S (2015) Solution structure of the PsIAA4 oligomerization domain reveals interaction modes for transcription factors in early auxin response. Proc Natl Acad Sci U S A 112:6230–6235

Enders TA, Strader LC (2015) Auxin activity: past, present, and future. Am J Bot 102:180–196

Fukaki H, Tameda S, Masuda H, Tasaka M (2002) Lateral root formation is blocked by a gain-of-function mutation in the *SOLITARY-ROOT/IAA14* gene of *Arabidopsis*. Plant J 29:153–168

Gagne JM, Downes BP, Shiu SH, Durski AM, Vierstra RD (2002) The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. Proc Natl Acad Sci U S A 99:11519–11524

Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M (2001) Auxin regulates SCF^{TIR1}-dependent degradation of AUX/IAA proteins. Nature 414:271–276

Guilfoyle TJ, Hagen G (2001) Auxin response factors. J Plant Growth Regul 20:281–291

Guilfoyle TJ, Hagen G (2007) Auxin response factors. Curr Opin Plant Biol 10:453–460

Guilfoyle TJ, Hagen G (2012) Getting a grasp on domain III/IV responsible for auxin response factor-IAA protein interactions. Plant Sci 190:82–88

Hamann T, Mayer U, Jürgens G (1999) The auxin-insensitive *bodenlos* mutation affects primary root formation and apical-basal patterning in the *Arabidopsis* embryo. Development 126:1387–1395

Hamann T, Benkova E, Bäurle I, Kientz M, Jürgens G (2002) The *Arabidopsis BODENLOS* gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. Genes Dev 16:1610–1615

Han M, Park Y, Kim I, Kim EH, Yu TK, Rhee S, Suh JY (2014) Structural basis for the auxin-induced transcriptional regulation by Aux/IAA17. Proc Natl Acad Sci U S A 111:18613–18618

Havens KA, Guseman JM, Jang SS, Pierre-Jerome E, Bolten N, Klavins E, Nemhauser JL (2012) A synthetic approach reveals extensive tunability of auxin signaling. Plant Physiol 160:135–142

Hussain M, Lu Y, Liu YQ, Su K, Zhang J, Liu J, Zhou GB (2016) Skp1: implications in cancer and SCF-oriented anti-cancer drug discovery. Pharmacol Res 111:34–42

Jing H, Yang X, Zhang J, Liu X, Zheng H, Dong G, Nian J, Feng J, Xia B, Qian Q et al (2015) Peptidyl-prolyl isomerization targets rice Aux/IAAs for proteasomal degradation during auxin signalling. Nat Commun 6:7395

Kagale S, Rozwadowski K (2011) EAR motif-mediated transcriptional repression in plants an underlying mechanism for epigenetic regulation of gene expression. Epigenetics-Us 6:141–146

Kagale S, Links MG, Rozwadowski K (2010) Genome-wide analysis of ethylene-responsive element binding factor-associated amphiphilic repression motif-containing transcriptional regulators in Arabidopsis. Plant Physiol 152:1109–1134

Ke J, Ma H, Gu X, Thelen A, Brunzelle JS, Li J, Xu HE, Melcher K (2015) Structural basis for recognition of diverse transcriptional repressors by the TOPLESS family of corepressors. Sci Adv 1:e1500107

Kepinski S, Leyser O (2004) Auxin-induced SCF^{TIR1}-Aux/IAA interaction involves stable modification of the SCF^{TIR1} complex. Proc Natl Acad Sci U S A 101:12381–12386

Kepinski S, Leyser O (2005) The *Arabidopsis* F-box protein TIR1 is an auxin receptor. Nature 435:446–451

Kim BC, Soh MS, Kang BJ, Furuya M, Nam HG (1996) Two dominant photomorphogenic mutations of *Arabidopsis thaliana* identified as suppressor mutations of *hy2*. Plant J 9:441–456

Kim J, Harter K, Theologis A (1997) Protein-protein interactions among the Aux/IAA proteins. Proc Natl Acad Sci U S A 94:11786–11791

Korasick DA, Westfall CS, Lee SG, Nanao MH, Dumas R, Hagen G, Guilfoyle TJ, Jez JM, Strader LC (2014) Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression. Proc Natl Acad Sci U S A 111:5427–5432

Lee S, Sundaram S, Armitage L, Evans JP, Hawkes T, Kepinski S, Ferro N, Napier RM (2014) Defining binding efficiency and specificity of auxins for $SCF^{TIR1/AFB}$-Aux/IAA co-receptor complex formation. ACS Chem Biol 9:673–682

Leyser O (2006) Dynamic integration of auxin transport and signalling. Curr Biol 16:R424–R433

Leyser HMO, Pickett FB, Dharmasiri S, Estelle M (1996) Mutations in the *AXR3* gene of *Arabidopsis* result in altered auxin response including ectopic expression from the *SAUR-AC1* promoter. Plant J 10:403–413

Li SB, Xie ZZ, Hu CG, Zhang JZ (2016) A review of auxin response factors (ARFs) in plants. Front Plant Sci 7:47

Liscum E, Reed JW (2002) Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol Biol 49:387–400

Long JA, Ohno C, Smith ZR, Meyerowitz EM (2006) TOPLESS regulates apical embryonic fate in *Arabidopsis*. Science 312:1520–1523

Luo Z, Pan Y, Jeong LS, Liu J, Jia L (2012) Inactivation of the Cullin (CUL)-RING E3 ligase by the NEDD8-activating enzyme inhibitor MLN4924 triggers protective autophagy in cancer cells. Autophagy 8:1677–1679

Mockaitis K, Estelle M (2008) Auxin receptors and plant development: a new signaling paradigm. Annu Rev Cell Dev Biol 24:55–80

Nagpal P, Walker LM, Young JC, Sonawala A, Timpte C, Estelle M, Reed JW (2000) *AXR2* encodes a member of the Aux/IAA protein family. Plant Physiol 123:563–573

Nanao MH, Vinos-Poyo T, Brunoud G, Thevenon E, Mazzoleni M, Mast D, Laine S, Wang SC, Hagen G, Li HB et al (2014) Structural basis for oligomerization of auxin transcriptional regulators. Nat Commun 5:3617

Noda Y, Kohjima M, Izaki T, Ota K, Yoshinaga S, Inagaki F, Ito T, Sumimoto H (2003) Molecular recognition in dimerization between PB1 domains. J Biol Chem 278:43516–43524

Orlowski RZ, Kuhn DJ (2008) Proteasome inhibitors in cancer therapy: lessons from the first decade. Clin Cancer Res 14:1649–1657

Overvoorde PJ, Okushima Y, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Liu A, Onodera C, Quach H et al (2005) Functional genomic analysis of the *AUXIN/INDOLE-3-ACETIC ACID* gene family members in *Arabidopsis thaliana*. Plant Cell 17:3282–3300

Parry G, Calderon-Villalobos LI, Prigge M, Peret B, Dharmasiri S, Itoh H, Lechner E, Gray WM, Bennett M, Estelle M (2009) Complex regulation of the TIR1/AFB family of auxin receptors. Proc Natl Acad Sci U S A 106:22540–22545

Petzold G, Fischer ES, Thoma NH (2016) Structural basis of lenalidomide-induced CK1 alpha degradation by the CRL4(CRBN) ubiquitin ligase. Nature 532:127–130

Pickart CM (2001) Mechanisms underlying ubiquitination. Annu Rev Biochem 70:503–533

Pierre-Jerome E, Moss BL, Lanctot A, Hagemena A, Nemhauser JL (2016) Functional analysis of molecular interactions in synthetic auxin response circuits. Proc Natl Acad Sci 113:11354–11359

Reed JW (2001) Roles and activities of Aux/IAA proteins in *Arabidopsis*. Trends Plant Sci 6:420–425

Remington DL, Vision TJ, Guilfoyle TJ, Reed JW (2004) Contrasting modes of diversification in the *Aux/IAA* and *ARF* gene families. Plant Physiol 135:1738–1752

Rinaldi MA, Liu J, Enders TA, Bartel B, Strader LC (2012) A gain-of-function mutation in *IAA16* confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. Plant Mol Biol 79:359–373

Rogg LE, Lasswell J, Bartel B (2001) A gain-of-function mutation in *IAA28* suppresses lateral root development. Plant Cell 13:465–480

Rouse D, Mackay P, Stirnberg P, Estelle M, Leyser O (1998) Changes in auxin response from mutations in an *AUX/IAA* gene. Science 279:1371–1373

Ruegger M, Dewey E, Hobbie L, Brown D, Bernasconi P, Turner J, Muday G, Estelle M (1997) Reduced naphthylphthalamic acid binding in the *tir3* mutant of Arabidopsis is associated with a reduction in polar auxin transport and diverse morphological defects. Plant Cell 9:745–757

Ruegger M, Dewey E, Gray WM, Hobbie L, Turner J, Estelle M (1998) The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast Grr1p. Genes Dev 12:198–207

Salehin M, Bagchi R, Estelle M (2015) $SCF^{TIR1/AFB}$-based auxin perception: mechanism and role in plant growth and development. Plant Cell 27:9–19

Sauer M, Robert S, Kleine-Vehn J (2013) Auxin: simply complicated. J Exp Bot 64:2565–2577

Shimizu-Mitao Y, Kakimoto T (2014) Auxin sensitivities of all Arabidopsis Aux/IAAs for degradation in the presence of every TIR1/AFB. Plant Cell Physiol 55:1450–1459

Soucy TA, Smith PG, Milhollen MA, Berger AJ, Gavin JM, Adhikari S, Brownell JE, Burke KE, Cardin DP, Critchley S et al (2009) An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. Nature 458:732–736

Su SH, Gray WM, Masson PH (2015) Auxin: shape matters. Nat Plants 1:15097

Sumimoto H, Kamakura S, Ito T (2007) Structure and function of the PB1 domain, a protein interaction module conserved in animals, fungi, amoebas, and plants. Sci STKE 2007:re6

Szemenyei H, Hannon M, Long JA (2008) TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. Science 319:1384–1386

Tan X, Calderon-Villalobos LIA, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446:640–645

Tatematsu K, Kumagai S, Muto H, Sato A, Watahiki MK, Harper RM, Liscum E, Yamamoto KT (2004) *MASSUGU2* encodes Aux/IAA19, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in *Arabidopsis thaliana*. Plant Cell 16:379–393

Terrile MC, Paris R, Calderon-Villalobos LIA, Iglesias MJ, Lamattina L, Estelle M, Casalongue CA (2012) Nitric oxide influences auxin signaling through S-nitrosylation of the Arabidopsis TRANSPORT INHIBITOR RESPONSE 1 auxin receptor. Plant J 70:492–500

Theologis A, Huynh TV, Davis RW (1985) Rapid induction of specific mRNAs by auxin in pea epicotyl tissue. J Mol Biol 183:53–68

Tian Q, Reed JW (1999) Control of auxin-regulated root development by the *Arabidopsis thaliana SHY2/IAA3* gene. Development 126:711–721

Tiwari SB, Wang X-J, Hagen G, Guilfoyle TJ (2001) Aux/IAA proteins are active repressors, and their stability and activity are modulated by auxin. Plant Cell 13:2809–2822

Tiwari SB, Hagen G, Guilfoyle T (2003) The roles of auxin response factor domains in auxin-responsive transcription. Plant Cell 15:533–543

Uehara T, Okushima Y, Mimura T, Tasaka M, Fukaki H (2008) Domain II mutations in CRANE/IAA18 suppress lateral root formation and affect shoot development in *Arabidopsis thaliana*. Plant Cell Physiol 49:1025–1038

Ulmasov T, Hagen G, Guilfoyle TJ (1997a) ARF1, a transcription factor that binds to auxin response elements. Science 276:1865–1868

Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ (1997b) Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. Plant Cell 9:1963–1971

Ulmasov T, Hagen G, Guilfoyle TJ (1999a) Activation and repression of transcription by auxin-response factors. Proc Natl Acad Sci U S A 96:5844–5849

Ulmasov T, Hagen G, Guilfoyle TJ (1999b) Dimerization and DNA binding of auxin response factors. Plant J 19:309–319

Walker JC, Key JL (1982) Isolation of cloned cDNAs to auxin-responsive poly(A)RNAs of elongating soybean hypocotyl. Proc Natl Acad Sci U S A 79:7185–7189

Wang RH, Zhang Y, Kieffer M, Yu H, Kepinski S, Estelle M (2016) HSP90 regulates temperature-dependent seedling growth in Arabidopsis by stabilizing the auxin co-receptor F-box protein TIR1. Nat Commun 7:10269

Wilson AK, Pickett FB, Turner JC, Estelle M (1990) A dominant mutation in *Arabidopsis* confers resistance to auxin, ethylene and abscisic acid. Mol Gen Genet 222:377–383

Wu S, Yu L (2016) Targeting cullin-RING ligases for cancer treatment: rationales, advances and therapeutic implications. Cytotechnology 68:1–8

Yang XQ, Lee S, So JH, Dharmasiri S, Dharmasiri N, Ge L, Jensen C, Hangarter R, Hobbie L, Estelle M (2004) The IAA1 protein is encoded by *AXR5* and is a substrate of SCF^{TIR1}. Plant J 40:772–782

Zenser N, Ellsmore A, Leasure C, Callis J (2001) Auxin modulates the degradation rate of Aux/IAA proteins. Proc Natl Acad Sci U S A 98:11795–11800

Zhao Y, Morgan MA, Sun Y (2014) Targeting Neddylation pathways to inactivate Cullin-RING ligases for anticancer therapy. Antioxid Redox Signal 21:2383–2400

Chapter 5
Structural Biology of Jasmonic Acid Metabolism and Responses in Plants

Cynthia K. Holland and Joseph M. Jez

5.1 Introduction to Jasmonates

To coordinate growth and development during times of stress and predation, plants evolved the capability of synthesizing and responding to an array of complex modified oxylipins, or oxygenated fatty acid derivates, known as jasmonates (Kombrink 2012). Methyl jasmonate (MeJA) was first isolated from the oil of jasmine (*Jasminum grandiflorum*), although its effect on plant physiology remained elusive for nearly two decades (Demole et al. 1962; Ueda and Kato 1980; Dathe 1981). Since this discovery, jasmonates have been recognized as critical phytohormones involved in plant defense and fertility.

Chemically, the jasmonates encompass jasmonic acid (JA) and its derivatives, including MeJA, 3*R*, 7*S*-jasmonyl-L-isoleucine (JA-Ile), cis-jasmone, and jasmonyl-1-aminocyclopropane-1-carboxylic acid (JA-ACC) (Fig. 5.1a) (Browse 2009). Jasmonate biosynthesis proceeds through the lipooxygenase-catalyzed peroxidation of linoleic acid to form 13-hydroperoxide, which is modified and cyclized to form 12-oxo-phytodienoic acid (OPDA). OPDA is reduced and then undergoes three rounds of β-oxidation to form JA. Further biochemical modifications of JA, such as glycosylation, decarboxylation, reduction of either the C_6 carbonyl or $C_{9,10}$ double bond, or hydroxylation of either C_{11} or C_{12}, can convert the core JA scaffold into over 30 different jasmonate compounds with diverse physiological roles.

Investigations of the role of jasmonates in plant physiology reveal their role in stamen and trichome development, vegetative growth, cell cycle regulation, senescence, regulation of anthocyanin pigment biosynthesis, and responses to various biotic and abiotic stresses (Creelman and Mullet 1997; Wasternack 2007; Howe and Jander 2008; Browse 2009; Avanci et al. 2010; Pauwels and Goossens 2011).

C. K. Holland · J. M. Jez (✉)
Department of Biology, Washington University in St. Louis, St. Louis, MO, USA
e-mail: jjez@wustl.edu

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_5

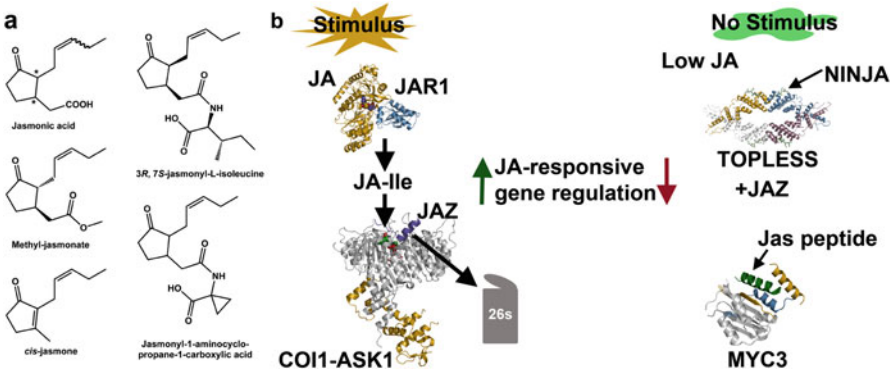


Fig. 5.1 Overview of jasmonate molecules and jasmonate-linked responses. (**a**) Chemical structures of representative jasmonates. The * denotes a chiral center that can be either in the *R* or *S* configuration. The carbon–carbon double-bond drawn as a squiggle can be either cis or trans. (**b**) Overview of jasmonate signaling and gene regulation pathways. Stimulus that leads to production of JA-Ile by the enzyme JAR1 results in perception of the bioactive hormone by the COI1-ASK1 receptor. Ubiquitinylation of JAZ proteins by the complex activates JA-responsive gene expression. Under low JA conditions, JAZ proteins interact with different transcriptional factors, such as TOPLESSS and MYC3, to repress expression

Jasmonates are also key players in induced resistance of plants (Kunkel and Brooks 2002), including rhizobacteria-mediated induced systemic resistance (Pieterse et al. 1998).

As one of the major plant defense compounds, jasmonates induce defense mechanisms to ward off insects, herbivores, and a broad spectrum of fungal and bacterial pathogens (Howe and Jander 2008; Browse 2009). Wounding of the plant tissue leads to elevated levels of jasmonates at the site of attack, which in turn induces a systemic increase of JAs through the regulation of JA-responsive genes that allow for production of diverse compounds as the attack continues (Reymond et al. 2004; Glauser et al. 2008). These compounds include glucosinolates, camalexins, alkaloids, artemisinin, and volatile organic compounds, such as terpenes, which provide a chemical arsenal to combat predators (Bolter 1993; Paré and Tumlinson 1999; Engelberth et al. 2004; De Geyter et al. 2012).

When stressed or otherwise stimulated, various jasmonates accumulate and are converted to JA-Ile, the only known bioactive jasmonate hormone (Fig. 5.1b) (Wasternack and Hause 2013; Fonseca et al. 2009a, b). JA-Ile forms a co-receptor complex with the F-box protein CORONATINE INSENSITIVE 1(COI1) and a JASMONATE-ZIM DOMAIN PROTEIN (JAZ) (Katsir et al. 2008). This complex allows for ubiquitinylation of the JAZ protein, which results in its degradation by the 26S proteasome (Fonseca et al. 2009a; Thines et al. 2007; Chini et al. 2007). JAZ proteins are nuclear repressors of transcription factors and repress JA-responsive genes through recruiting the co-repressor TOPLESS (TPL) by the adaptor protein NOVEL INTERACTOR OF JAZ (NINJA) and also competing with MEDIATOR25 (MED25) for interaction with members of the MYC transcription factor family

(Pauwels et al. 2010; Zhang et al. 2015). Degradation of JAZ proteins thereby alters transcriptional patterns in response to JA-mediated signals.

Structural efforts to characterize the JA biosynthetic enzymes, starting from chloroplast membrane lipids, have allowed for a better understanding of the generation of 3*R*, 7*S*-jasmonate needed to produce JA-Ile to elicit a physiological response. Although the jasmonate biosynthesis and JA-related signaling pathways are well-studied, atomic-level studies aimed at deciphering the mechanisms and molecular interactions responsible for large-scale physiological responses are just beginning. Recent structural investigations reveal the biochemistry and protein–protein interaction details of enzymes and proteins involved in bioactive jasmonate biosynthesis, metabolism, perception, and transcriptional responses.

5.2 Production of JA-Ile, the Bioactive Jasmonate Phytohormone

Plant responses to developmental cues and the environment are primarily modulated by the activity and concentration of phytohormones (Westfall et al. 2010). Fluctuation in concentrations of plant growth regulators and modifications of the chemical structure of a hormone affects receptor interactions to alter developmental responses. Although jasmonates can be decorated with methyl groups, sugars, amino acids, and other chemical moieties (Staswick and Tiryaki 2004), the functions of all of the various modified jasmonates remain unclear. Presumably, plants have developed mechanisms, such as conjugation with amino acids, to regulate hormone levels. For example, conjugation of isoleucine to JA by the acyl acid amido synthetase JAR1 gives rise to JA-Ile, which is the active form of the hormone (Fig. 5.2a) (Staswick et al. 2002).

In plants, the GH3 (Gretchen Hagen 3) acyl acid amido synthetases contribute to maintaining hormone levels in plants by conjugating a hormone, including auxins, jasmonates, and benzoates, to an amino acid as either a storage or activation mechanism (Hagen and Guilfoyle 1985; Westfall et al. 2013). The biological function of this class of enzymes was first established when the *jasmonate resistant 1* (*jar1–1*) mutant was identified in *Arabidopsis thaliana* and shown to exhibit decreased sensitivity to exogenous MeJA and reduced male fertility (Staswick et al. 1992, 2002). The *A. thaliana* JAR1 protein (also known as AtGH3.11) functions as a JA-amido synthetase and conjugates Ile to JA (Staswick and Tiryaki 2004). AtGH3.11/JAR1 functions as a monomer to catalyze the adenylation of JA using ATP in the first half-reaction of a ping-pong kinetic mechanism with subsequent AMP release and JA-amino acid conjugate formation in the second half-reaction (Chen et al. 2010). JA-Ile can then regulate JA-responsive genes through binding the F-box receptor protein COI1 (Feys et al. 1994; Xie et al. 1998).

The X-ray crystal structure of AtGH3.11/JAR1 was determined in complex with JA-Ile to reveal its catalytic mechanism (Westfall et al. 2012). The overall structure

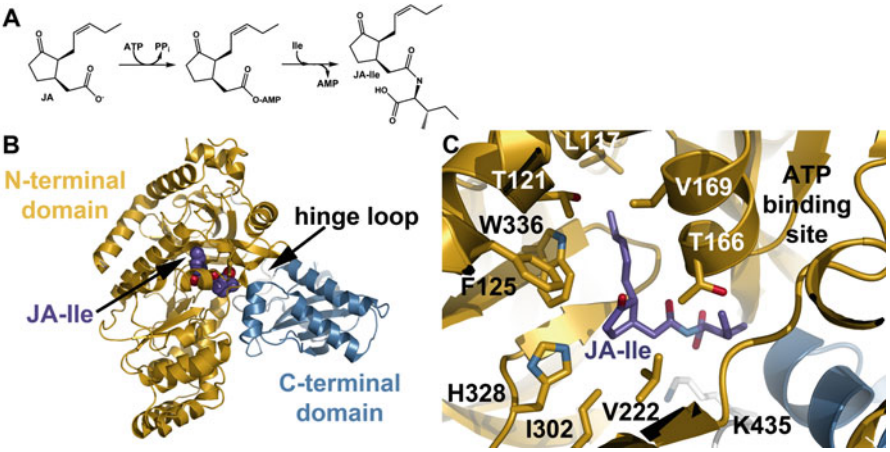


Fig. 5.2 Conversion of JA to the bioactive jasmonate JA-Ile. (**a**) Overall reaction catalyzed by the JAR1 acyl acid amido synthetase. Each half-reaction is shown. (**b**) Three-dimensional structure of JAR1 from *Arabidopsis thaliana*. The ribbon diagram of JAR1 in complex with JA-Ile (space-filling model; purple) is shown. The N- and C-terminal domains of JAR1 are colored gold and blue, respectively. The 'hinge-loop' that alters in conformation, along with the C-terminal domain, is indicated. (**c**) Active site of JAR1. This view shows a close-up of the JAR1 active site with JA-Ile bound. Key residues are noted. The ATP binding site, based on other GH3 protein structures, is also indicated

of AtGH3.11/JAR1 consists of a large N-terminal domain containing α-helices that flank a β-barrel and two β-sheets; the C-terminal domain is related to the adenylating firefly luciferase (ANL) enzyme superfamily (Fig. 5.2b) (Gulick 2009). The active site is located at the intersection of the two domains with a hinge loop between Leu ^{427}-Arg^{439} providing the conformational flexibility necessary for pivoting the C-terminal domain at each stage of catalysis.

The structure of AtGH3.11/JAR1 in complex with JA-Ile provides information on how the enzyme binds the oxylipin molecule (Fig. 5.2c). At the interface of the two domains, JA-lle is situated in a large apolar binding site. Multiple residues, including Thr166, Val222, Phe223, and Ile304 encompass the cyclopentane ring of JA-Ile with a water-mediated hydrogen bond formed with His328. The pentenyl tail of JA-Ile binds in a pocket formed by Leu117, Thr121, Phe125, Thr166, Val169, and Trp336. The isoleucine moiety is oriented with its carboxylate toward residues of the nucleotide binding P-loop. Although no nucleotide was bound in the structure, the highly conserved adenosine binding pocket architecture can be inferred based on the structure of another *A. thaliana* GH3 protein (i.e., AtGH3.12/PBS3) (Westfall et al. 2012). The nucleotide-binding site consists of a canonical phosphate-binding loop motif, a β-turn-β-structure motif, and a ribose interaction motif. When the C-terminal domain is oriented in the "open" active site conformation, the active site allows for binding of ATP, Mg^{2+}, and JA for the adenylation reaction to occur. Upon adenylation, pyrophosphate is released, and rotation of the C-terminal domain is

triggered, trapping the adenylated intermediate for the second half-reaction. This rotation also creates a tunnel that allows access of isoleucine into the active site for nucleophilic attack on the intermediate. This releases AMP and results in formation of the conjugated product JA-Ile.

5.3 Perception of JA-Ile by the Hormone Receptor F-Box Protein COI1

The hormone receptor for JA-Ile is the F-box protein COI1, which functions in recruiting substrates to the Skp1-Cullin1-F-box protein (SCF) ubiquitin E3 ligase complex (Feys et al. 1994; Xie et al. 1998). In the presence of JA-Ile, JAZ transcriptional repressors bind SCF^{COI1} (Chini et al. 2007; Yan et al. 2007). The active jasmonate hormone mediates protein–protein interaction, which leads to the poly-ubiquitylation and degradation of JAZ proteins by the 26S proteasome. This relieves repression of the MYC2 transcription factor, or other transcription factors, and allows for transcriptional changes in jasmonate-responsive genes (Lorenzo et al. 2004; Chini et al. 2007).

Structural studies of the *A. thaliana* COI1-ASK1 complex bound to either JA-Ile or the jasmonate mimic coronatine and a peptide of a bipartite JAZ1 degron reveal how the hormone acts as molecular glue in this macromolecular assembly (Fig. 5.3a) (Sheard et al. 2010). The COI1-JAZ co-receptor forms through interactions between the C-terminal Jas motif of the JAZ1 protein (Glu^{200}-Val^{220}). The overall structure of COI1 resembles the F-box complex of the auxin receptor, TIR1 (Tan et al. 2007).

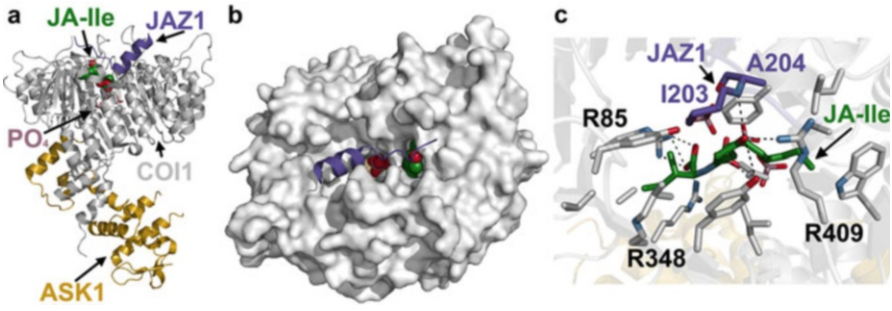


Fig. 5.3 Perception of JA-Ile by the jasmonate receptor COI1. (**a**) Overall structure of the COI1-ASK1 jasmonate receptor complex. The overall structure is shown as a ribbon diagram with JA-Ile (green), phosphate (PO_4; rose) bound in the inositol pentakisphosphate-binding site, and the JAZ1 peptide (purple) shown as space-filling models. (**b**) JA-Ile as the molecular glue between COI1 and JAZ proteins. Top view of the COI1-ASK1 complex shown as a molecular surface model with JA-Ile (green) and the JAZ1 peptide (purple) represented as a space-filling model and a ribbon, respectively. (**c**) Interactions in the JA-Ile binding site. A close-up view of the JA-Ile (green) binding site and polar interactions (dotted lines) made within the active site pocket with both COI1 (white) and JAZ1 (purple) is shown

COI1 consists of a tri-helical N-terminal F-box motif that binds ASK1 and a horseshoe-shaped C-terminal solenoid of 18 tandem leucine-rich repeats (LRRs). The top of the LRR domain binds the hormone and the C-terminal Jas motif of JAZ1 through three intra-repeat loops (Loop 2, 12, and 14), and a fourth loop (Loop C) partially covers Loop 2. Large, polar residues provide multiple contacts for formation of the functional complex (Fig. 5.3b). Any mutations of residues in the site disrupt COI1-JAZ1 interactions. The keto group of the ligand, either JA-Ile or coronatine, points upward and interacts with COI1 residues at the entrance of the pocket and is solvent exposed in the absence of the JAZ degron. Three arginine residues (Arg85, Arg348, and Arg409) bind the amide and carboxyl groups of the ligand forming the floor of the pocket. Additional density in the structure led to the conclusion that inositol pentakisphosphate ($InsP_5$) binds below the hormone-binding pocket. Although the specific binding interactions and the mechanism remain to be explored, $InsP_5$ was identified as a critical component of the co-receptor for sensitive jasmonate perception. Therefore, the jasmonate receptor complex is a three-molecule complex, requiring COI1, the JAZ degron, and $InsP_5$ for hormone binding.

The complete structure of a JAZ protein remains to be determined. JAZ proteins appear to be intrinsically disordered in their native form, which allows for repressor plasticity in its binding interactions (Pazos et al. 2013; Pietrosemoli et al. 2013). *A. thaliana* contains 12 JAZ proteins that are thought to be redundant in signaling pathways as repressors (Chini et al. 2007; Yan et al. 2007; Browse 2009). In the COI1-ASK1 structure, the degron peptide of JAZ1 forms a bipartite structure. Five N-terminal amino acids interact with COI1 and completely cover the opening where JA-Ile binds, which traps the ligand in the pocket. These co-receptor interactions explain the stereospecific preferences for (*3R, 7S*)-JA-Ile verus (*3R, 7R*)-JA-Ile, as the stereochemistry may place the aliphatic chain unfavorably close to nearby JAZ1 and COI1 residues. The A-helix of the C-terminus of the JAZ1 degron binds the surface of the COI1 LRR domain, thus strengthening the interactions that form the co-receptor complex. There is considerable variability on the amino acid sequence level of the N-terminal clamp region of the JAZ degron among JAZ proteins. Further investigation is needed to tease apart these interactions between COI1 and the variable JAZ proteins and the responses of these co-receptors.

5.4 Transcriptional Regulation of Jasmonate-Responsive Genes

In the absence of JA-Ile, JAZ repressors bind either helix-loop-helix transcription factors or adaptor proteins to recruit co-repressors needed to down-regulate jasmonate-targeted gene expression (Chini et al. 2007; Thines et al. 2007; Pauwels et al. 2010; Ke et al. 2015). The general co-repressor TOPLESS (TPL) can be recruited by an array of repressors, transcription factors, and transcriptional adaptor proteins (Causier et al. 2012).

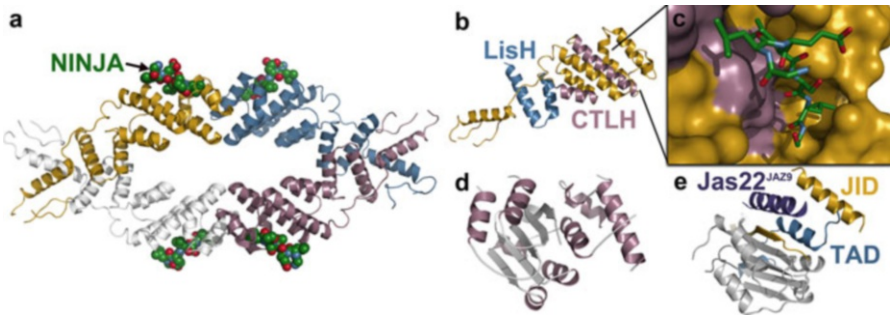


Fig. 5.4 Transcriptional control of jasmonate responses. (**a**) Structural insight on TOPLESS (TPD)-NINJA interaction. A ribbon diagram of the tetrameric N-terminal domain of TOPLESS (TPD) co-complexed with NINJA peptides (green space-filling molecule) is shown. Each monomer of the TPD is colored differently. (**b**) Interaction between the LisH (blue) and CTLH (rose) domains of the TPD (gold; domains colored separately). (**c**) NINJA binding site. The NINJA peptide (green) binds TPD near the CTLH domain (rose). (**d**) Structure of the MYC3 transcription factor. The ribbon diagram shows the overall structure of MYC3 and is colored by secondary structure (α-helices, rose; β-strand, white). (**e**) JAZ protein interaction promotes transcriptional activation. Binding of the Jas motiff of a JAZ protein (purple ribbon) between Jas-interacting domain (JID) and the transcription activation domain (TAD) of MYC3 alters interaction with the MED25 subunit of the Mediator complex

TPL regulates gene expression by interacting with the ethylene response factor associated amphiphillic repression (EAR) motifs of certain transcriptional repressors (Causier et al. 2012). The N-terminal domain of TPL, or the TOPLESS domain (TPD), interacts with EAR motifs of NINJA proteins, and also JAZ repressors themselves, when JA-Ile levels are low. The N-terminal domain contains a lissencephaly homologous (LisH) dimerization motif and a C-terminal to LisH (CTLH) motif, followed by a proline- and glutamine-rich linker that connects the N-terminal domain to the two WD40 domains, a 40 amino acid motif that terminates with a tryptophan-aspartate dipeptide, of the C-terminus (Fig. 5.4a). X-ray crystal structures of the TPD from *A. thaliana* and rice (*Oryza sativa*) revealed a dimer of dimers, with each monomeric unit containing two dimerization interfaces with a CTLH motif sandwiched in between (Fig. 5.4b) (Ke et al. 2015). Each TPD monomer contains nine α-helices and two small 3_{10} helices. Co-crystallization of a TPD with peptides of the NINJA EAR motif showed that four NINJA peptides could bind one TPD tetramer in a shallow groove composed of the CTLH motif (Fig. 5.4c). Homologous WD40 domains in animals and yeast co-repressors, notably Groucho and Tup1, have been found to interact with repressors, but repressors that interact with plant TPL WD40 domains remain to be identified, which would give further insight into the overall roles of TPL in plant growth and development (Goldstein et al. 1999; Jimenez et al. 1997; Muhr et al. 2001; Paroush et al. 1994; Zhang et al. 2001; Komachi et al. 1994).

While the Jas domain of JAZ proteins is an integral part of the COI1 F-box receptor complex, this domain also directly regulates the transcription factor MYC3

(Zhang et al. 2015). *A. thaliana* MYC proteins contain a Jas-interacting domain and a transcription activation domain. In the three-dimensional structure, MYC3 forms a helix-sheet-helix sandwich fold of five antiparallel β-sheets surrounded by eight α-helices (Fig. 5.4d). This structure was the first example of a non-complexed well-resolved MYC transcription activation domain, as they are generally unstructured in their unbound state. Additionally, MYC3 has been co-crystallized with a Jas peptide (i.e., Ser218-Met239) of JAZ9 (Fig. 5.4e) (Zhang et al. 2015). The Jas peptide was nestled between the Jas-interacting domain (JID) and the transcription activation domain (TAD). While the first five amino acids of the Jas domain peptide are necessary for COI1-JAZ binding interactions, they do not form critical interactions with MYC3. The Jas peptide was a continuous helix, in contrast to its bipartite loop and short C-terminal helix in the COI1 co-receptor complex structure. This suggests that JAZ proteins employ a switch mechanism to form distinct confirmations in the MYC-JAZ transcriptional repression resting state versus the COI1-JAZ activated complex. Such a mechanism implies that MYC and COI1 potentially compete for binding the Jas domain of JAZ proteins. Elucidation of how MYC and Jas domains interact provided the first molecular insight on how transcriptional repression is achieved in a plant hormone response pathway.

To promote transcription of jasmonate-responsive genes, RNA polymerase II is recruited to promoters by the action of the MED25 subunit of the Mediator complex (Kazan and Manners 2013). MYC3 directly binds MED25, although JAZ repressors can outcompete MED25 for MYC3. While MED25 is a critical component of gene regulation in response to jasmonates, biophysical protein–protein interaction and structural data are absent and would provide insight into the flexibility of MYC3 binding both MED25 and JAZ repressors (Chen et al. 2012; Cevik et al. 2012; Kidd et al. 2009). For further information on the MYC family of transcriptional activators, basic helix-loop-helix repressors, and other JAZ targets readers are referred to a recent review (Chini et al. 2016).

5.5 Biosynthesis and Chemical Modifications of Jasmonates

5.5.1 Jasmonic Acid Biosynthesis

To gain an understanding of JA-signaling regulation, structural efforts have also focused on the biosynthesis of jasmonates and their stereo-selectivity in generating 3*R*, 7*S*-jasmonate, the precursor to JA-Ile. Jasmonates, while they are a diverse class of compounds, all stem from the same polyunsaturated fatty acid (PUFA) intermediates, linoleic acid (18:2) and hexadecatrienoid acid (16:3), from the chloroplast membrane (Schaller and Stintzi 2009). These PUFAs can be oxidized to form fatty acid regio- and stereo-specific hydroperoxides using lipooxygenases within the chloroplast. The oxygenated fatty acid hydroperoxide products are dehydrated to form unstable allylic epoxides, or allene oxides, which are cyclized by allene oxide

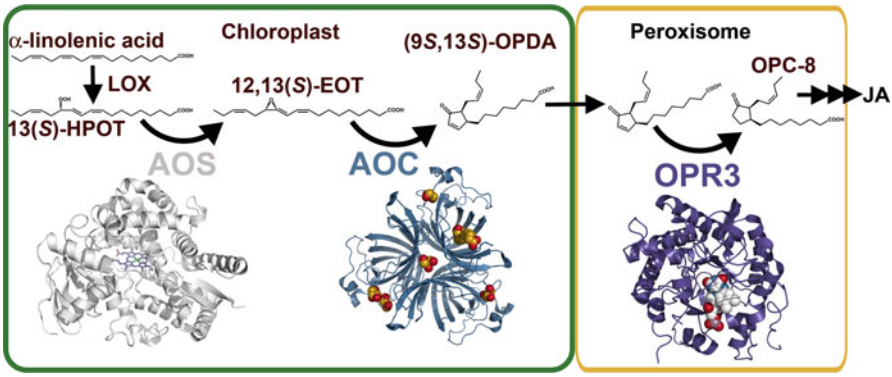


Fig. 5.5 Overview of JA biosynthesis and structures of key enzymes. Biosynthesis of JA begins in the chloroplast with the conversion of α-linolenic acid to (9*S*, 13*S*)-OPDA. X-ray crystal structures of AOS and AOC, which catalyze the second and final steps in the chloroplast pathway, provide chemical insight on these biosynthetic conversions. OPDA is transported from the chloroplast to the peroxisome, where the enzyme OPR3 forms OPC-8. Subsequent β-oxidation of OPC-8 leads to formation of JA

cyclases to form optically pure (9*S*,13*S*)-oxophytodienoic acid (OPDA). OPDA is then transported out of the chloroplast by an unidentified transporter.

X-ray crystal structures of the CYP74 allene oxide synthase (AOS) from *A. thaliana* and guayule (*Parthenium argentatum*) (Lee et al. 2008; Li et al. 2008) reveal that plant CYP74s share a common fold with other P450 enzymes with a heme-binding loop that is eight amino acids longer than similar P450s in other organisms (Fig. 5.5). The allene oxide cyclase 2 (AOC2) crystal structure gave insight into the reaction chemistry, specificity, and evolution that gives rise to the first cyclic, optically pure bioactive molecule in the JA biosynthetic pathway, OPDA (Hoffman et al. 2006). The AOC2 trimer contains a lipocalin fold comprised of an eight-stranded antiparallel β-barrel followed by a partially helical C-terminus (Fig. 5.5). To identify the active site, the enzyme was co-crystallized with the inhibitor vernolic acid, which was bound in the hydrophobic barrel cavity. Modeling of the 12,13(*S*)-epoxy-9(Z),11,15(Z)-octadecatrienoic acid (12,13-EOT) substrate and the OPDA product into the active site cavity led to the prediction of the mechanism of cyclization, which includes the opening of the epoxide to form an anionic intermediate followed by a *trans* to *cis* conformational change between C10 and C11 to allow for the pericyclic ring closure with absolute stereo-selectivity.

To complete jasmonate biosynthesis, the peroxisomal ATP-binding cassette (ABC) transporter COMATOSE (CTS) (Footitt et al. 2002; Theodoulou et al. 2005) transports OPDA and other β-oxidation substrates into the peroxisome. Once inside, OPDA is reduced to 3-oxo-2-(2′(*Z*)-pentenyl)-cyclopentane-1-octanoic acid (OPC-8:0). The crystal structure of tomato 12-OPDA reductase 3 has revealed the active site architecture responsible for the substrate stereo-selectivity and its regulation through self-inhibited dimerization (Breithaupt et al. 2006). A Co-A ester is added by ATP-dependent acyl-activating enzymes to prepare OPC-8:0 for β-oxidation. OPC-8:CoA ligase 1 (OPCL1), is the only enzyme in the *A. thaliana*

acyl-activating enzyme family that is JA-specific (Koo et al. 2006). Two or three rounds of β-oxidation result in the shortened JA-CoA, and a thioesterase cleaves the CoA moiety to release free JA.

Overall, the crystallographic investigations of enzymes in jasmonate biosynthesis have revealed the mechanisms underlying the variety of oxylipins that can be generated while maintaining strict specificity; however, the roles of JA biosynthesis pathway intermediates in signaling remain to be understood.

5.5.2 Methyl-Jasmonate

Plants use volatile emissions to attract pollinators and seed dispersers and repel unwanted herbivores (Seo et al. 2001). MeJA is one such volatile involved in communicating defense responses within the organism and amongst plants (Seo et al. 2001; Karban et al. 2000; Kessler et al. 2006; Baldwin et al. 2006). Plant exposure to MeJA triggers JA-mediated defense responses, and exogenous MeJA application can lead to herbivore resistance in many plant species (McConn et al. 1997; Baldwin 1998; Li et al. 2002; Wu et al. 2008).

To generate MeJA, the enzyme JA carboxyl methyltransferase (JMT) uses S-adenosyl-L-methionine (SAM) to catalyze the methylation of JA (Seo et al. 2001). Although a JMT structure remains to be determined, the X-ray crystal structure of a related enzyme, the *Clarkia breweri* salicylic acid carboxyl methyltransferase (SAMT) (Zubieta et al. 2003), was solved. Using this structure as a guide, the active site architecture can be inferred by sequence alignment, which predicts a highly conserved SAM/SAH binding site. *Clarkia* SAMT and *A. thaliana* JMT vary by six residues in the active site (Tyr147 to Ser, Met150 to His, Ile225 Gln, Met308 to Ile, Phe347 to Tyr, and Asn349 to Ile), and mutagenesis of these residues was able to introduce methylation of JA into the SAMT but not a complete switch in specificity. A structure of a plant JMT would allow for a more thorough investigation of key residues involved in JMT specificity.

As noted above, the exogenous application of MeJA elicits a physiological response, which results from metabolism of MeJA to JA (Wu et al. 2008). Although an MeJA-esterase catalyzes the hydrolysis of MeJA (Stuhlfelder et al. 2002, 2004), the three-dimensional structure of this enzyme remains to be determined and would contribute to our understanding of the molecular basis of plant JA perception.

5.5.3 Homeostasis of JA-Ile: Hydroxylation, Carboxylation, and Turnover

Mechanisms governing the homeostasis of JA-Ile are essential to our understanding of plant defenses (Koo et al. 2011). Cytochrome P450 oxygenases (CYPs) constitute a vast family of heme-containing enzymes that catalyze both monooxygenations and

hydroxylations (Schuler and Werck-Reichhart 2003). CYPs can also regulate hormonal pathways by modifying the chemical structures of phytohormones (Mizutani and Ohta 2010). The JA metabolite 12-hydroxy-JA-Ile forms through hydroxylation of JA-Ile by both CYP94B3 and CYP94B1 (Koo et al. 2011, 2014). CYP94B3 is specific for JA-Ile and is unable to hydroxylate JA to form 12-hydroxy JA or 12-hydroxy-JA-Ile to form 12-carboxy-JA-Ile. In contrast, CYP94B1 can convert JA-Ile to 12-carboxy-JA-Ile, although to a lesser extent (Koo et al. 2011, 2014; Kitaoka et al. 2011; Heitz et al. 2012). Together, these two enzymes account for more than 95% of 12-hydroxylation of JA-Ile and are localized to the endoplasmic reticulum (Koo et al. 2014). The carboxy-derivative of JA is preferentially formed from 12-hydroxy-JA-Ile by a third CYP, CYP94C1 (Heitz et al. 2012). Because these CYP enzymes inactivate JA-Ile, structural studies are necessary to provide insight into the evolution of CYP substrate specificity and to understand the mechanisms underlying JA modifications as a means of decreasing the pool of physiologically active JA.

While JA-Ile is the active form of the jasmonate hormone, the amino acid must be removed for it to be deactivated (Widemann et al. 2013). Two amidohydrolases, IAR3 (indole-3-acetic acid (IAA)-Ala resistant 3) and ILL6 (IAA-Leu resistant1-like 6), have been characterized in the catalysis of JA-Ile turnover. Not only does IAR3 cleave JA-Ile, but it also hydrolyzes the 12-hydroxy-JA-Ile conjugate. Although there is an X-ray crystallographic structure for the IAA-amino acid hydrolyzing ILL2, the structure of a JA-Ile specific amidohydrolase remains to be determined (Bitto et al. 2009).

5.6 Summary

In the past decade, structural studies have deepened our understanding of the complexity regulating jasmonate biosynthesis and signaling. Active jasmonate is synthesized by the conjugation of the amino acid Ile to JA by JAR1, which is perceived by the COI1-ASK1 complex. In the presence of JA-Ile, the JAZ repressor binds the JA-Ile-bound COI1-ASK1 complex, which leads to the degradation of JAZ by the 26S proteasome. When JA-Ile levels decline, JAZ recruits the co-repressor TOPLESS through interactions with the adaptor protein NINJA to bind MYC repressors. While structural work on this signaling pathway has revealed the general scheme, efforts to understand the JAZ-binding interactors and the mechanisms governing selectivity in this large family of repressors are necessary to gain a full appreciation of the complexity of JA-responsive gene regulation. The adapter protein NINJA has only recently been discovered, and the C-terminal domain should be investigated for its role in WD40 domain interactions in signaling and for, potentially, the discovery of cross talk between the JA signaling pathway and other hormone pathways. Jasmonates encompass a broad family of compounds, and the physiological role of these decorated jasmonates continues to escape our understanding.

Future investigations are needed to elucidate the physiological and potential signaling role of the diversity of jasmonate compounds, as well as the role of the modifying enzymes in regulating the JA pool, and thus regulating JA responses. Overall, structural, biochemical, and biophysical studies have brought to light the intricacies of JA signaling, but a great deal remains to be discovered about the untapped potential of other JAs in influencing physiological responses.

References

Avanci NC, Luche DD, Goldman GH, Goldman MH (2010) Jasmonates are phytohormones with multiple functions, including plant defense and reproduction. Genet Mol Res 9:484–505

Baldwin IT (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. Proc Natl Acad Sci U S A 95:8113–8118

Baldwin IT, Halitschke R, Paschold A, von Dahl CC, Preston CA (2006) Volatile signaling in plant-plant interactions: "talking trees" in the genomics era. Science 311:812–815

Bitto E, Bingman CA, Bittova L, Houston NL, Boston RS, Fox BG, Phillips GN Jr (2009) X-ray structure of ILL2, an auxin-conjugate amidohydrolase from *Arabidopsis thaliana*. Proteins 74:61–71

Bolter CJ (1993) Methyl jasmonate induces papain inhibitor(s) in tomato leaves. Plant Physiol 103:1347–1353

Breithaupt C, Kurzbauer R, Lilie H, Schaller A, Strassner J, Huber R, Macheroux P, Clausen T (2006) Crystal structure of 12-oxophytodienoate reductase 3 from tomato: self-inhibition by dimerization. Proc Natl Acad Sci U S A 103:14337–14342

Browse J (2009) Jasmonate passes muster: a receptor and targets for the defense hormone. Annu Rev Plant Biol 60:183–205

Causier B, Ashworth M, Guo W, Davies B (2012) The TOPLESS interactome: a framework for gene repression in Arabidopsis. Plant Physiol 158:423–438

Cevik V, Kidd BN, Zhang P, Hill C, Kiddle S, Denby KJ, Holub EB, Cahill DM, Manners JM, Schenk PM, Beynon J, Kazan K (2012) MEDIATOR25 acts as an integrative hub for the regulation of jasmonate-responsive gene expression in Arabidopsis. Plant Physiol 160:541–555

Chen Q, Westfall CS, Hicks LM, Wang S, Jez JM (2010) Kinetic basis for the conjugation of auxin by a GH3 family indole acetic acid-amido synthetase. J Biol Chem 285:29780–29786

Chen R, Jiang H, Li L, Zhai Q, Qi L, Zhou W, Liu X, Li H, Zheng W, Sun J, Li C (2012) The Arabidopsis Mediator subunit MED25 differentially regulates jasmonate and abscisic acid signaling through interacting with the MYC2 and ABI5 transcription factors. Plant Cell 24:2898–2916

Chini A, Fonseca S, Fernandez G, Adie B, Chico JM, Lorenzo O, Garcia-Casado G, Lopez Vidriero I, Lozano FM, Ponce MR, Micol JL, Solano R (2007) The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448:666–671

Chini A, Gimenez-Ibanez S, Goossens A, Solano R (2016) Redundancy and specificity in jasmonate signalling. Curr Opin Plant Biol 33:147–156

Creelman RA, Mullet JE (1997) Biosynthesis and action of jasmonates in plants. Annu Rev Plant Physiol Plant Mol Biol 48:355–381

Dathe W (1981) Endogenous plant hormones of the broad bean, *Vicia faba L.*(−)-jasmonic acid, a plant growth inhibitor in pericarp. Planta 153:530–535

De Geyter N, Gholami A, Goormachtig S, Goossens A (2012) Transcriptional machineries in jasmonate-elicited plant secondary metabolism. Trends Plant Sci 17:349–359

Demole E, Lederer E, Mercier D (1962) Isolement et détermination de la structure du jasmonate de méthyle, constituant odorant charactéristique de l'essence de jasmin. Helv Chim Acta 45:675–685

Engelberth J, Alborn HT, Schmelz EA, Tumlinson JH (2004) Airborne signals prime plants against insect herbivore attack. Proc Natl Acad Sci U S A 101:1781–1785

Feys BJF, Benedetti CE, Penfold CN, Turner JG (1994) Arabidopsis mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. Plant Cell 6:751–759

Fonseca S, Chini A, Hamberg M, Adie B, Porzel A, Kramell R, Miersch O, Wasternack C, Solano R (2009a) (+)-7-Iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. Nat Chem Biol 5:344–350

Fonseca S, Chico JM, Solano R (2009b) The jasmonate pathway: the ligand, the receptor and the core signalling module. Curr Opin Plant Biol 12:539–547

Footitt S, Slocombe SP, Larner V, Kurup S, Wu Y, Larson T, Graham I, Baker A, Holdsworth M (2002) Control of germination and lipid mobilization by COMATOSE, the Arabidopsis homologue of human ALDP. EMBO J 21:2912–2922

Glauser G, Grata E, Dubugnon L, Rudaz S, Farmer EE, Wolfender JL (2008) Spatial and temporal dynamics of jasmonate synthesis and accumulation in Arabidopsis in response to wounding. J Biol Chem 283:16400–16407

Goldstein RE, Jimenez G, Cook O, Gur D, Paroush Z (1999) Huckebein repressor activity in *Drosophila* terminal patterning is mediated by Groucho. Development 126:3747–3755

Gulick AM (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. ACS Chem Biol 4:811–827

Hagen G, Guilfoyle TJ (1985) Rapid induction of selective transcription by auxins. Lol Cell Biol 5:1197–1203

Heitz T, Widemann E, Lugan R, Miesch L, Ullmann P, Désaubry L, Holder E, Grausem B, Kandel S, Miesch M, Werck-Reichhart D, Pinot F (2012) Cytochromes P450 CYP94C1 and CYP94B3 catalyze two successive oxidation steps of plant hormone jasmonoyl-isoleucine for catabolic turnover. J Biol Chem 287:6296–6306

Hoffman E, Zerbe P, Schaller F (2006) The crystal structure of *Arabidopsis thaliana* allene oxide cyclase: insights into the oxylipin cyclization reaction. Plant Cell 18:3201–3217

Howe GA, Jander G (2008) Plant immunity to insect herbivores. Annu Rev Plant Biol 59:41–66

Jimenez G, Paroush Z, Ish-Horowicz D (1997) Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. Genes Dev 11:3072–3082

Karban R, Baldwin IT, Baxter KJ, Laue G, Felton GW (2000) Communication between plants: induced resistance in wild tobacco plants following clipping of neighboring sagebrush. Oecologia 125:66–71

Katsir L, Schilmiller AL, Staswick PE, He SY, Howe GA (2008) COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. Proc Natl Acad Sci U S A 105:7100–7105

Kazan K, Manners JM (2013) MYC2: the master in action. Mol Plant 6:686–703

Ke J, Ma H, Gu X et al (2015) Structural basis for recognition of diverse transcriptional repressors by the TOPLESS family of corepressors. Sci Adv 1:e1500107

Kessler A, Halitschke R, Diezel C, Baldwin IT (2006) Priming of plant defense responses in nature by airborne signaling between *Artemisia tridentata* and *Nicotiana attenuata*. Oecologia 148:280–292

Kidd BN, Edgar CI, Kumar KK, Aitken EA, Schenk PM, Manners JM, Kazan K (2009) The mediator complex subunit PFT1 is a key regulator of jasmonate-dependent defense in Arabidopsis. Plant Cell 21:2237–2252

Kitaoka N, Masubara T, Sato M, Takahashi K, Wakuta S, Kawaide H, Matsui H, Nabeta K, Matsuura H (2011) Arabidopsis CYP94B3 encodes jasmonyl-L-isoleucine 12-hydroxylase, a key enzyme in the oxidative catabolism of jasmonate. Plant Cell Physiol 52:1757–1765

Komachi K, Redd MJ, Johnson AD (1994) The WD repeats of Tup1 interact with the homeo domain protein α2. Genes Dev 8:2857–2867

Kombrink E (2012) Chemical and genetic exploration of jasmonate biosynthesis and signaling in plants. Planta 236:1351–1366

Koo AJ, Chung HS, Kobayashi Y, Howe GA (2006) Identification of a peroxisomal acyl activating enzyme involved in the biosynthesis of jasmonic acid in Arabidopsis. J Biol Chem 281:33511–33520

Koo AJ, Cooke TF, Howe GA (2011) Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. Proc Natl Acad Sci U S A 108:9298–9303

Koo AJ, Thireault C, Zemelis S, Poudel AN, Zhang T, Kitaoka N, Brandizzi F, Matsuura H, Howe GA (2014) Endoplasmic reticulum-associated inactivation of the hormone jasmonyl-L-isoleucine by multiple members of the cytochrome P450 94 family in Arabidopsis. J Biol Chem 298:29728–29738

Kunkel BN, Brooks DM (2002) Cross talk between signaling pathways in pathogen defense. Curr Opin Plant Biol 5:325–331

Lee D, Nioche P, Hamberg M, Raman CS (2008) Structural insights into the evolutionary paths of oxylipin biosynthetic enzymes. Nature 455:363–368

Li C, Williams MM, Loh YT, Lee GI, Howe GA (2002) Resistance of cultivated tomato to cell content-feeding herbivores is regulated by the octadecanoid-signaling pathway. Plant Physiol 130:494–503

Li L, Chang Z, Pan Z, Fu Z, Wang X (2008) Modes of heme binding and substrate access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide synthase. Proc Natl Acad Sci U S A 105:13883–13888

Lorenzo O, Chico JM, Sanchez-Serrano JJ, Solano R (2004) JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. Plant Cell 16:1938–1950

McConn M, Creelman RA, Bell E, Mullet JE, Browse J (1997) Jasmonate is essential for insect defense in Arabidopsis. Proc Natl Acad Sci U S A 94:5473–5477

Mizutani M, Ohta D (2010) Diversification of P450 genes during land plant evolution. Annu Rev Plant Biol 61:291–315

Muhr J, Andersson E, Persson M, Jessell TM, Ericson J (2001) Groucho-mediated transcriptional repression establishes progenitor cell pattern and neuronal fate in the ventral neural tube. Cell 104:861–873

Paré PW, Tumlinson JH (1999) Plant volatiles as a defense against insect herbivores. Plant Physiol 121:325–331

Paroush Z, Finley RL Jr, Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D (1994) Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. Cell 79:805–815

Pauwels L, Goossens A (2011) The JAZ proteins: a crucial interface in the jasmonate signaling cascade. Plant Cell 23:3089–3100

Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, Perez AC, Chico JM, Bossche RV, Sewell J, Gil E, Garcia-Casado G, Witters E, Inze D, Long JA, De Jaeger G, Solano R, Goossens A (2010) NINJA connects the co-repressor TOPLESS to jasmonate signalling. Nature 464:788–791

Pazos F, Pietrosemoli N, Garcia-Martin JA, Solano R (2013) Protein intrinsic disorder in plants. Front Plant Sci 4:363

Pieterse CM, van Wees SC, van Pelt JA, Knoester M, Laan R, Gerrits H, Weisbeek PJ, van Loon LC (1998) A novel signaling pathway controlling induced systemic resistance in Arabidopsis. Plant Cell 10:1571–1580

Pietrosemoli N, Garcia-Martin JA, Solano R, Pazos F (2013) Genome-wide analysis of protein disorder in *Arabidopsis thaliana*: implications for plant environmental adaptation. PLoS One 8: e55524

Reymond P, Bodenhausen N, Van Poecke RM, Krishnamurthy V, Dicke M, Farmer EE (2004) A conserved transcript pattern in response to a specialist and a generalist herbivore. Plant Cell 16:3132–3147

Schaller A, Stintzi A (2009) Enzymes in jasmonate biosynthesis- structure, function, regulation. Phytochemistry 70:1532–1153

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annu Rev Plant Biol 54:629–667

Seo HS, Song JT, Cheong JJ, Lee YH, Lee YW, Hwang I, Lee JS, Choi YD (2001) Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. Proc Natl Acad Sci U S A 98:4788–4793

Sheard LB, Tan X, Mao H et al (2010) Jasmonate perception by inositol-phosphate potentiated COI1-JAZ co-receptor. Nature 465:400–405

Staswick PE, Tiryaki I (2004) The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in Arabidopsis. Plant Cell 16:2117–2127

Staswick PE, Su W, Howell SH (1992) Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an *Arabidopsis thaliana* mutant. Proc Natl Acad Sci U S A 89:6837–6840

Staswick PE, Tiryaki I, Rowe ML (2002) Jasmonate response locus JAR1 and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. Plant Cell 14:1405–1415

Stuhlfelder C, Lottspeich F, Mueller MJ (2002) Purification and partial amino acid sequences of an esterase from tomato. Phytochemistry 60:233–240

Stuhlfelder C, Mueller MJ, Warzecha H (2004) Cloning and expression of a tomato cDNA encoding a methyl jasmonate cleaving esterase. Eur J Biochem 271:2976–2983

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446:640–645

Theodoulou FL, Job K, Slocombe SP, Footitt S, Holdsworth M, Baker A, Larson TR, Graham IA (2005) Jasmonic acid levels are reduced in COMATOSE ATP-binding cassette transporter mutants: implications for transport of jasmonate precursors into peroxisomes. Plant Physiol 137:835–840

Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, Browse J (2007) JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. Nature 448:661–665

Ueda J, Kato J (1980) Isolation and identification of a senescence-promoting substance from wormwood (*Artemisia absinthium L.*). Plant Physiol 66:246–249

Wasternack C (2007) Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. Ann Bot 100:681–697

Wasternack C, Hause B (2013) Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in annals of botany. Ann Bot 111:1021–1058

Westfall CS, Herrmann J, Chen Q, Wang S, Jez JM (2010) Modulating plant hormones by enzyme action. Plant Signal Behav 5:1607–1612

Westfall CS, Zubieta C, Herrmann J, Kapp U, Nanao MH, Jez JM (2012) Structural basis for pre-receptor modulation of plant hormones by GH3 family proteins. Science 336:1708–1711

Westfall CS, Muehler AM, Jez JM (2013) Enzyme action in the regulation of plant hormone responses. J Biol Chem 288:19304–19311

Widemann E, Miesch L, Lugan R, Holder E, Heinrich C, Aubert Y, Miesch M, Pinot F, Heitz T (2013) The amindohydrolases IAR3 and ILL6 contribute to jasmonyl-isoleucine horomone turnover and generate 12-hydroxyjasmonic acid upon wounding in Arabidopsis leaves. J Biol Chem 288:31701–31714

Wu J, Wang L, Baldwin IT (2008) Methyl jasmonate-elicited herbivore resistance: does MeHA function as a signal without being hydrolyzed to JA? Planta 227:1161–1168

Xie D, Feys BF, James S, Nieto-Rostro M, Turner JG (1998) COI1: an Arabidopsis gene required for jasmonate-regulated defense and fertility. Science 280:1091–1094

Yan Y, Stolz S, Chetelat A, Reymond P, Pagni M, Dubugnon L, Farmer EE (2007) A downstream mediator in the growth repression limb of the jasmonate pathway. Plant Cell 19:2470–2483

Zhang H, Levine M, Ashe HL (2001) Brinker is a sequence-specific transcriptional repressor in the *Drosophila* embryo. Genes Dev 15:261–266

Zhang F, Yao J, Ke J, Zhang L, Lam VQ, Xin XF, Zhou XE, Chen J, Brunzelle J, Griffin PR, Zhou M, Xu HE, Melcher K, He SY (2015) Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. Nature 525:269–273

Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP (2003) Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. Plant Cell 15:1704–1716

Chapter 6
Gibberellin

Sayaka Takehara and Miyako Ueguchi-Tanaka

6.1 Overview of GA Signal Transduction

Gibberellins (GAs) are well-known phytohormones that are essential for many developmental processes in plants, including seed germination, stem elongation, leaf expansion, pollen maturation, and the development of flowers, fruit, and seeds (Olszewski et al. 2002; Itoh et al. 2008). GA was first identified in the pathogenic fungus *Gibberella fujikuroi*, the causal agent of the "foolish-seedling" disease of rice, resulting in the excessive elongation of the infected rice plants (Yabuta and Sumiki 1938). In the 1950s, GAs were further recognized as endogenous hormones in higher plants (Radley 1956). Since these discoveries, over 130 GAs have been identified from various plants, fungi, and bacteria (MacMillan 2002). Among the GAs produced in plants, only a few GAs act as bioactive GAs, with the others being precursors or catabolites of bioactive GAs (Yamaguchi 2008).

Over the past few decades, most of the components involved in GA signaling have been identified through the genetic screening of rice (*Oryza sativa*) and *Arabidopsis thaliana*. One of the key components, the GA receptor GIBBERELLIN INSENSITIVE DWARF1 (GID1) was first identified from rice as a crucial gene in the *GA i*nsensitive *d*warf mutant, *gid1* (Ueguchi-Tanaka et al. 2005). The rice *gid1* mutant affects various responses, such as GA's promotion of leaf sheath elongation, α-amylase's induction in seeds and the feedback of GA biosynthetic enzymes (Ueguchi-Tanaka et al. 2005). GID1 is a soluble protein that has structural similarity to hormone-sensitive lipases (HSLs) involved in lipid metabolism in animals (Yeaman 2004). GID1-green fluorescent protein has been detected in the nucleus and had a faint signal in the cytosol (Ueguchi-Tanaka et al. 2005). Glutathione S-transferase-tagged GID1 interacted directly with bioactive GAs but not with inactive

S. Takehara · M. Ueguchi-Tanaka (✉)
Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, Japan
e-mail: mueguchi@nuagr1.agr.nagoya-u.ac.jp

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_6

GAs, and the affinities for the bioactive GAs, and the rapid interaction/dissociation of the GAs, are consistent with in vivo GA responses. Thus, it was concluded that GID1 is a soluble GA receptor. Another important property of GID1 is that it interacts with a rice DELLA protein (Ueguchi-Tanaka et al. 2005, see below). Since the discovery of the GID1 receptor in rice, the GID1 soluble receptor has been identified from various plants, such as Arabidopsis (Griffiths et al. 2006; Nakajima et al. 2006; Willige et al. 2007), populus (Sterky et al. 2004; Israelsson et al. 2005), grapevine (Velasco et al. 2007), cotton (Aleman et al. 2008), and lycophytes (Hirano et al. 2007; Yasumura et al. 2007). There is one GA receptor gene in rice, *OsGID1*, whereas there are three GA receptor genes in Arabidopsis, *GID1a*, *GID1b*, and *GID1c,* which have overlapping but also distinct functions in regulating different developmental processes (Griffiths et al. 2006; Nakajima et al. 2006). The Arabidopsis *gid1* triple mutant results in severe GA-insensitive phenotypes, including failure of seed to germinate, late flowering, severe dwarfism, and complete infertility (Griffiths et al. 2006; Iuchi et al. 2007; Willige et al. 2007; Suzuki et al. 2009).

In 2008, two groups reported the structure of GID1 in rice and Arabidopsis using X-ray crystallography. The structural analyses led to the redefining of "bioactive GAs" as ligands for the GID1 receptor (Murase et al. 2008; Shimada et al. 2008). Bioactive GAs, which include GA_1, GA_3, GA_4, and GA_7, are tetracyclic diterpenoid carboxylic acids possessing the *ent*-GA skeleton, and they contain a C6 carboxyl group, a γ-lactone ring between C4 and C10, a hydroxyl group at C3, and no hydroxyl group at C2 (Yamaguchi 2008) (Fig. 6.1).

Other important components of the GA-signaling pathway are the DELLA proteins (DELLAs), which function as negative regulators of GA responses. They belong to the GRAS superfamily of putative transcription factors. DELLA proteins contain DELLA and TVHYNP motifs at their N-termini and a GRAS domain at their C-termini, which defines them as members of the GRAS family of putative transcriptional regulators (Bolle 2004). Rice has only one DELLA protein, SLENDER RICE1 (SLR1); by contrast, Arabidopsis has five DELLAs, GA-INSENSITIVE (GAI), REPRESSOR OF GA1–3 (RGA), and three RGA-like proteins, RGL-1, RGL-2, and RGL-3, which play distinct and also overlapping functions in repressing GA responses (Peng et al. 1997; Ikeda et al. 2001; Silverstone et al. 2001; Lee et al. 2002; Wen and Chang 2002; Itoh et al. 2003; Tyler et al. 2004). GA is perceived by a GA-receptor, GID1, and then elicits the degradation of DELLAs, (Ueguchi-Tanaka et al. 2005; Griffiths et al. 2006) resulting in a de-repressed GA-signaling state. Moreover, an F-box protein, referred to as GIBBERELLIN-INSENSITIVE DWARF2 (GID2) was identified in rice, and SLEEPY1 (SLY1) and SNEEZY were also identified in Arabidopsis (McGinnis et al. 2003; Sasaki et al. 2003). The F-box protein binds to DELLA once it has perceived the GA signal (Ueguchi-Tanaka et al. 2005, 2007a; Achard and Genschik 2009). This GA-dependent interaction confers the specificity of SKP1–CULLIN–F box protein (SCF) E3 ubiquitin ligase complex to DELLAs, which promotes the addition of a polyubiquitin chain, thereby targeting DELLAs for their subsequent degradation by the 26S proteasome complex in response to GAs (Ueguchi-Tanaka et al. 2007b).

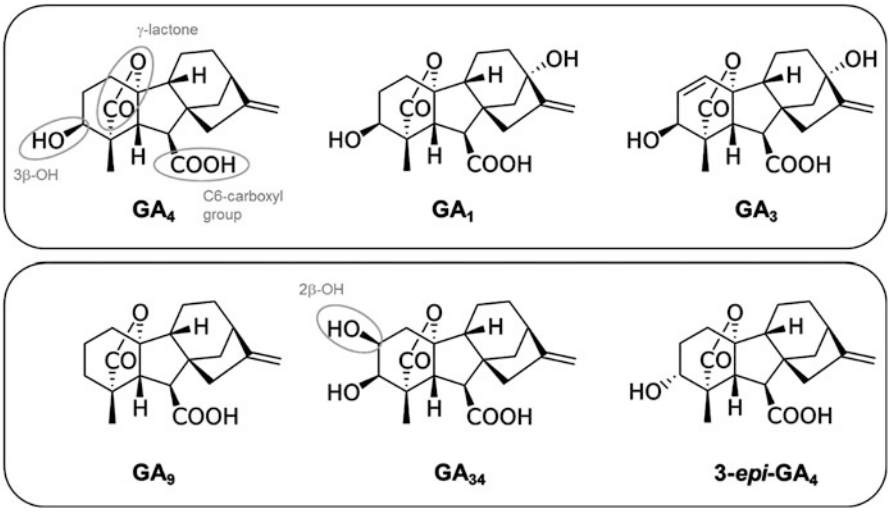


Fig. 6.1 Biologically active and inactive GAs in flowering plants. Bioactive GAs in flowering plants contain a γ-lactone structure in the A-ring, 3β-hydroxylation (3β-OH) of the A-ring, and a C-6 carboxyl group in the B-ring (light gray ellipses in the structure of GA_4). In the bottom panel, 2β-hydroxylation (2β-OH) of the A-ring, such as in GA_{34}, leads to inactivation. The GA preference of SmGID1b from *Selaginella moellendorffii* does not fit the rule of flowering plants. Thus, SmGID1b interacts with inactive GAs, GA_9 and the synthetic GA 3-*epi*-GA_4, but hardly interacts with bioactive GAs, GA_1 or GA_3 (Hirano et al. 2007)

The discovery of GID1, DELLA, and GID2/SLY1 in GA signaling produced an overview of GA signaling. When GA is absent, DELLA represses the GA response, and when GA is present, GID1 binds to GA, which induces the formation of the GID1–GA–DELLA complex, followed by the degradation of the DELLA protein and consequent occurrence of GA responses.

6.2 Structure of GID1

6.2.1 GID1 Resembles α/β-Hydrolase (ABH)

The structure of OsGID1 from rice bound with GA_4 and GA_3 was resolved at a 1.9 Å resolution using X-ray crystallography (Shimada et al. 2008). The crystal structure of the GA–AtGID1–DELLA (partial DELLA peptide) complex was also resolved at a 1.8 Å resolution from Arabidopsis (Murase et al. 2008). As was hypothesized based on the sequence similarity, the overall structure of GID1 shows an ABH fold similar to that of the HSLs (Fig. 6.2). GID1 has two major domains, the lid and the

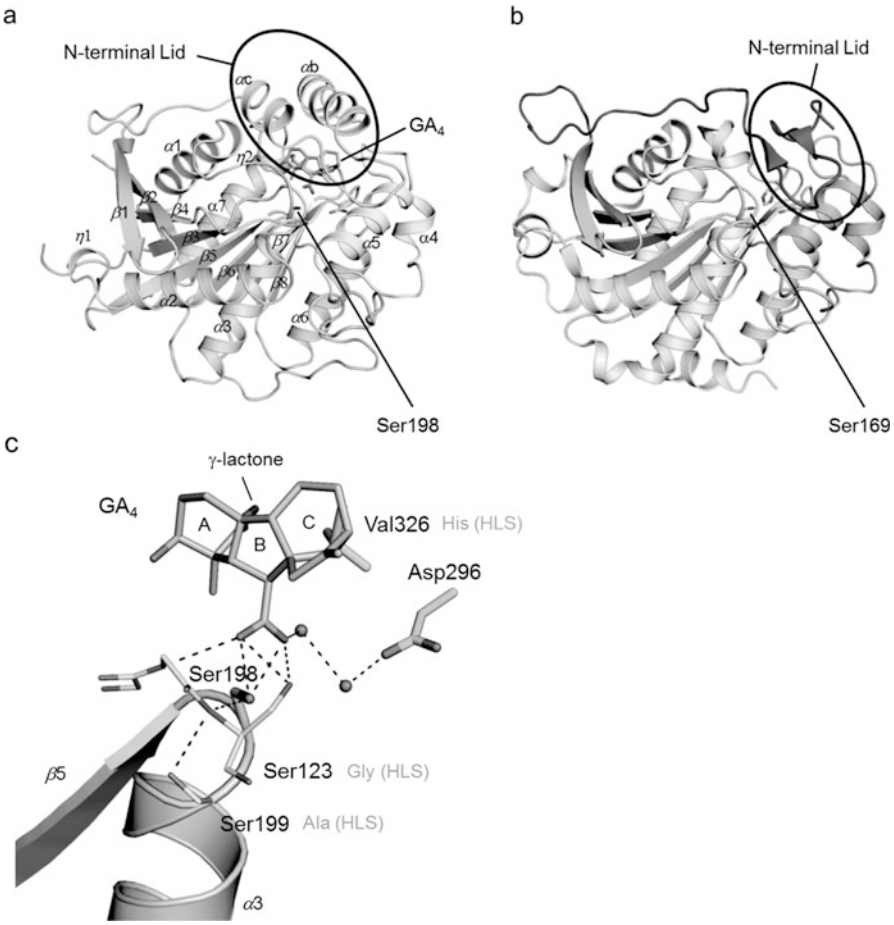


Fig. 6.2 (**a**) The GA_4–GID1 complex structure is shown as a ribbon diagram. (**b**) A ribbon diagram of plant carboxylesterase CXE1 from *Actinidia eriantha* (PDB code: 2O7R). (**c**) The GA-binding site structure of GID1, including corresponding residues for the HGGG (HGGS in GID1) motif and the catalytic triad of the HSLs (Ser198, Asp296, and Val326). The residues of GID1 are indicated, and the corresponding residues of AeCXE1 are also presented (HSL). The rings of the *ent*-gibberellane structure are labeled as A, B, and C, for GA_4, and water molecules are shown as spheres. (Figure is based on that of Shimada et al. 2008)

ABH fold (Osterlund 2001; Ueguchi-Tanaka et al. 2005, 2007a; Nakajima et al. 2006; Hirano et al. 2008; Murase et al. 2008). This ABH fold domain is composed of an eight-stranded β-sheet with α-helices packing the sides. This core domain contains the conserved HGG and GXSXG motifs characteristic of hormone-sensitive lipases and other carboxylesterase.

6.2.2 GA-Binding Pocket

GID1 proteins lack hydrolase activities because the His residue within the Ser-His-Asp catalytic triad of the ABH family replaced other amino acid residues, such as Val in rice (Nakajima et al. 2006). Instead, this pocket forms the binding core of bioactive GAs (Shimada et al. 2008).

The characteristics of the GA-binding site within GID1 are the hydrophilic network, which recognizes the C6 carboxyl group and the C3 hydroxyl group of GA_4, and the region formed by covering the GID1 N-terminal lid of the core hydrophobic region (Fig. 6.3). The C6 carboxyl group of GA_4 forms hydrogen bonds with Gly122, Ser198 and Ser123. Ser123 is stabilized by a network of hydrogen bonds formed with Arg251 and Asp 250 in α5. Moreover, the C6 carboxyl

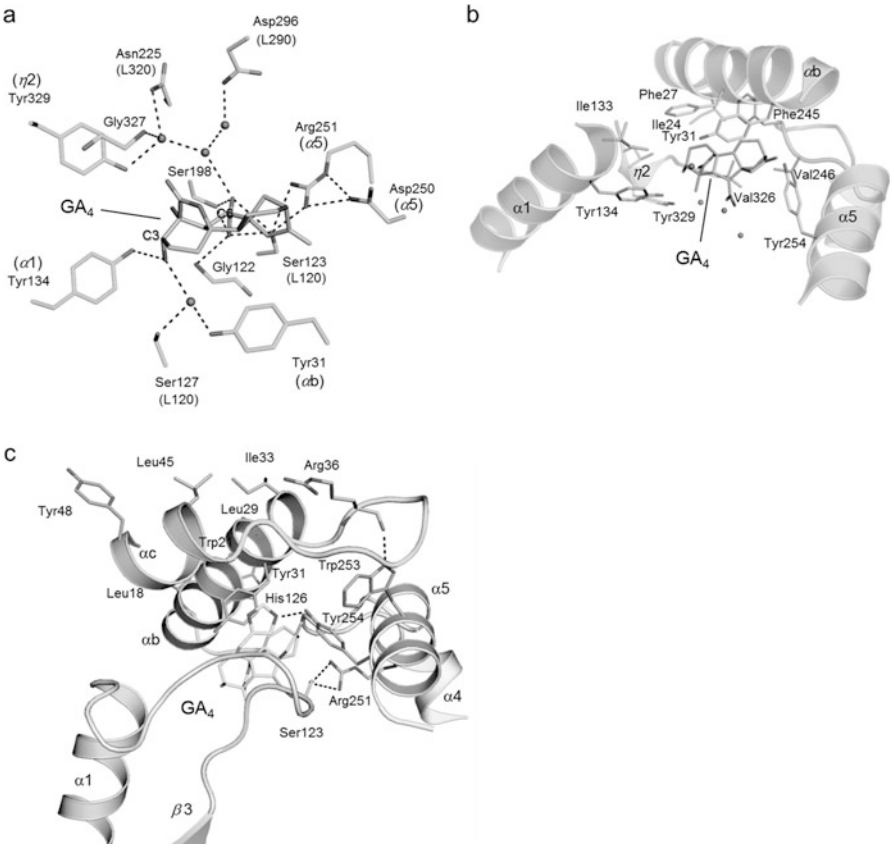


Fig. 6.3 **(a)** Polar residues that interact with GA_4. Water molecules are shown as spheres. The positions in the secondary structural element are shown in parentheses. **(b)** Residues with van der Waals or hydrophobic interactions with GA_4. **(c)** Functional residues in the lid involved in closuring of GID1. (Figure based on that of Shimada et al. 2008)

group is linked to Tyr329, Asn225, and Asp 296 by means of two water molecules. The C3 hydroxyl group of GA_4 forms hydrogen bonds with Tyr 134, Ser127, and Tyr31 indirectly through a water molecule. In the case of GA_3, there is an additional interaction by means of a hydrogen bond between the C13 hydroxyl group and Asp 250.

Phe27, and Tyr31 and Tyr254 in the N-terminal lid contact the A-ring and C-ring of GA_4, respectively. Thus, these residues contribute to keeping GA_4 held firmly by hydrophobic interactions, indicating that GA_4 likely acts to close the N-terminal lid over the binding pocket.

6.2.3 N-Terminal Lid (DELLA Protein Interaction)

The structural analysis of the GID1–GA_4 complex revealed that several amino acid side chains in the N-terminal lid, such as Leu18, Trp21, Leu29, and Ile 33 at αb, and Leu45 and Tyr48 at αc, extrude from the molecule's surface (Shimada et al. 2008). The replacement of these six residues markedly diminished the interaction of OsGID1 with SLR1 in yeast cells, but was not accompanied by a decline in the GA-binding activity (Shimada et al. 2008). This indicated that a number of hydrophobic residues protruding from the outer surface of the N-terminal lid in OsGID1 are required for protein–protein interactions with DELLA SLR1 (Shimada et al. 2008). Moreover, the crystal structure of the Arabidopsis GID1a–GA–DELLA peptide complex showed that GID1a binds the DELLA GAI directly, and the hydrophobic residues of GID1 are the major sites of interactions between the closed lid (outer surface) of GID1a and DELLA GAI (Murase et al. 2008). This GID1a N-terminal lid consists of a loop and three α-helices (αa, αb, and αc), covers GA in a deep binding pocket and then binds with the DELLA domain, containing DELLA, VHYNP, and LExLE motifs. In fact, previous analyses using yeast two hybrid assays and in vitro interaction experiments also suggested that GA-dependent GID1 interactions with SLR1 and RGA require the N-terminal DELLA and TVHYNP motifs (Griffiths et al. 2006; Ueguchi-Tanaka et al. 2007b), supporting the above result. Based on this research, the current model of GA perception proposes that the hydrophobic residues of GA interact with the GID1 N-terminal lid to induce a conformational change upon GA binding. Hydrophobic interactions between the GA and GID1 N-terminal lid are likely involved in pulling the lid closed. This folded N-terminal lid creates a binding surface to induce DELLA proteins on the residues extruding from the molecular surface of GID1. Finally, DELLA proteins are polyubiquitinated by the $SCF^{GID2/SLY1}$ ubiquitin E3 ligase complex and degraded through the 26S proteasome pathway (Fig. 6.4).

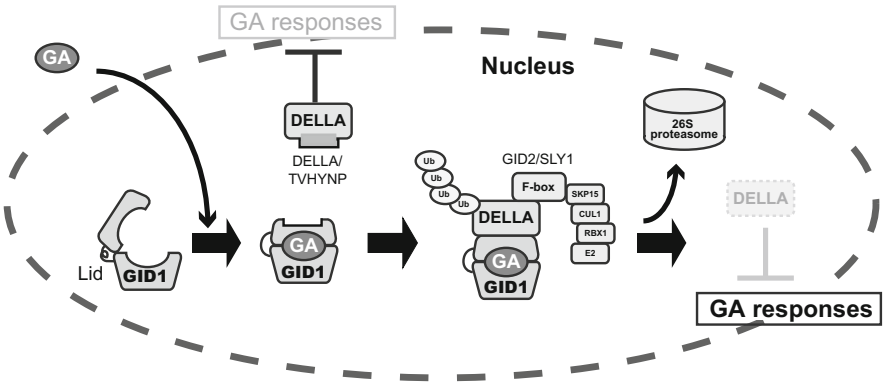


Fig. 6.4 Model for GA action through DELLA protein degradation. In the absence of GA, GA activity is repressed by the DELLA protein. When GA is present, the GID1 receptor binds GA. The GID1–GA complex then interacts with the DELLA and TVHYNP motifs of the DELLA protein, resulting in the recognition of the DELLA protein by the $SCF^{GID2/SLY1}$ complex. After the DELLA protein is polyubiquitinated by the $SCF^{GID2/SLY1}$ complex, it is degraded through the 26S proteasome pathway, and as a consequence GA is activated. This consecutive reaction is predicted to occur in the nucleus

6.2.4 Evolution of the GID1 Receptor

Two studies have explored when plants evolved the ability to perceive GA through the GID1/DELLA pathway and examined these GA-signaling components in the lycophyte *Selaginella moellendorffii* and the bryophyte *Physcomitrella patens* (Hirano et al. 2007; Yasumura et al. 2007). Homologs of DELLA, GID1, and GID2 in *S. moellendorffii* were able to functionally substitute for SLR1, GID1, and GID2 in rice (Hirano et al. 2007), respectively, and the *Selaginella kraussiana* DELLA repressed growth in Arabidopsis (Yasumura et al. 2007). In contrast with *S. moellendorffii*, no functional homologs were found in the moss *P. patens*, although there are some genes encoding proteins homologous to rice GID1, SLR1, and GID2 (Hirano et al. 2007). The GID1 homologs in *P. patens*, PpGID1L1 and PpGID1L2, did not exhibit in vitro GA-binding activities in either a thermodynamic analysis using isothermal titration calorimetry or a subtraction assay of the nonspecific binding activity, nor did they interact with any DELLA proteins in yeast cells in the presence of various kinds of GAs. Transgenic rice expressing the gene encoding PpDELLAL1, a protein that contains conserved GRAS domains similar to those in seed plant DELLA proteins but does not contain typical DELLA/TVHYNP domains, did not show any GA-insensitive phenotypes. Moreover, the overexpression of PpGID2L1 in rice *gid2-1* did not rescue its dwarf phenotype (Hirano et al. 2007). Based on these results, it was hypothesized that *P. patens* does not contain a GA-perception mechanism mediated by the GID1/DELLA system. Thus, the GA receptor/response system appeared after the

divergence of vascular plants from the moss lineage (~430 million years ago). The presence of the GID1–GA DELLA system in *S. moellendorffii,* which diverged early from the lineage of ferns (~400 million years ago) and seed plants (~300 million years ago), strongly suggests that the last common ancestor of vascular plants had a GID1–GA DELLA system.

However, the GA specificity and sensitivity of the lycophyte *S. moellendorffii* GID1s (SmGID1s) are lower than those in flowering plants (Hirano et al. 2007). This indicates that some important amino acid residues were gradually refined to obtain the high affinity and specificity to the higher plant GAs, such as GA_4 and GA_1, which are replaced in SmGID1s (Shimada et al. 2008).

6.2.5 ABH Family in Plants

Nuclear receptors are also presents in animals, but the ABH family has only been reported in plants, in which it regulates many physiological functions. Among the ABH family, the strigolactone and karrikin-smoke receptors, DAD2/D14 and KAI2, respectively, which are involved in shoot branching, are known to function as ligand receptors, as well as GA receptors (Hamiaux et al. 2012; Guo et al. 2013; Kagiyama et al. 2013; Zhao et al. 2014; Shimada et al. 2008). Moreover, the Arabidopsis genome contains hundreds of uncharacterized ABH-like genes (Lamesch et al. 2012; Mitchell et al. 2015; Mindrebo et al. 2016). Thus, there may be unknown diverse functions of ABH superfamily members in plants.

6.2.6 Structure of the GRAS Family Proteins

There are no reports of structural analyses of DELLA proteins containing C-terminal GRAS domains at present; however, recent studies on the other GRAS family proteins revealed their structures, including those of SCARECROW-LIKE 7 in rice (OsSCL7) and SHORT-ROOT (SHR)–SCARECROW (SCR) proteins. While OsSCL7 forms a homodimer in a side-by-side configuration (Li et al. 2016), SHR–SCR forms a heterodimer in a head-to-head configuration (Hirano et al. 2017). Moreover, SHR–SCR plays a role as a transcription cofactor, binding to target genes indirectly by interacting with the N-terminus of a conserved zinc finger, ZF4 α-helix, of the INDETERMINATE DOMAIN transcription factors (Hirano et al. 2017), while OsSCL7 directly interacts with DNA (Li et al. 2016). Thus, there is a possibility that there are subtypes of GRAS proteins with different properties. Further structural studies on DELLA proteins are required.

6.3 Recent Discoveries of DELLA Functions

6.3.1 Trapping Function Against Other Transcription Factors

DELLAs are predicted to have dual functions as transcriptional regulators because DELLA is considered to lack a DNA-binding domain (Davière and Achard 2013). It is possible that DELLA interacts with DNA-binding domains of other transcription factors and suppresses the expression of genes downstream of GA signaling in various developmental processes. In the absence of GA, the transcription of many GA-responsive genes is turned off. This is because DELLA interacts with the transcription factor, binding cis-sequences existing upstream of those genes, preventing the transcription factor from binding to its target genes' promoters and regulating gene expression. Conversely, in the presence of GA, the DELLA protein's ubiquitination and proteasome-mediated degradation are triggered, releasing the transcription factor from the negative effects of the DELLA protein. DELLA is reported to have a trapping function against several transcription factors.

For example, Arabidopsis DELLA interacts with two basic helix-loop-helix (bHLH)-type transcription factors, PHYTOCHROME INTERACTING FACTOR 3 and 4, and blocks their binding to the promoters of their target genes, resulting in short hypocotyls in light-grown plants (Feng et al. 2008; Lucas et al. 2008). In addition, DELLAs also interact with ALCATRAZ, another bHLH-type transcription factor, resulting in impaired seed dispersal (Arnaud et al. 2010). Moreover, recent studies have revealed that cross talk between GAs and jasmonates (JAs) is involved in the control of plant growth and defense in response to environmental and endogenous signals (Hou et al. 2013). In particular, the DELLA protein RGA binds to JA ZIM-domain proteins to compete with MYC2, which is involved in the JA-signaling pathway (Boter et al. 2004), and thus promotes MYC2-dependent JA signaling (Hou et al. 2010).

6.3.2 Transactivation

Additionally, DELLA interacts with other transcription factors as transcriptional co-activators to promote the expression of downstream genes (Zentella et al. 2007; Gallego-Bartolomé et al. 2011). In this case, the downstream genes encode the positive regulators in GA signaling, such as the GA biosynthetic enzymes GA20ox and GA3ox, GA receptor GID1, and transcriptional regulators, such as SCL3, resulting in GA-feedback regulation. In 2014, two groups revealed this control mechanism using yeast hybrid screenings and transient assays (Yoshida et al. 2014; Fukazawa et al. 2014). According to these studies, a co-activator/co-repressor-exchange regulatory system exists in which DELLA and SCL3 are used as a transcriptional co-activator and a co-repressor, respectively, and the

INDETERMINATE DOMAIN protein family is used as a transcriptional scaffold for DNA binding to regulate the expression of SCL3 and other genes involved in GA signaling. This system may be involved in the transcriptional regulation of other signaling pathways.

6.4 Conclusions

Structural analyses of OsGID1–GA and GID1a–GA–DELLA revealed that GID1's overall structure resembles that of the HSLs. GID1 lacks hydrolase activity because the His residue within the Ser-His-Asp catalytic triad of the ABH family has been replaced with other amino acid residues, such as Val in rice. Plant GID1 may have evolved uniquely from a receptor family derived from lipases because the ABH family is plant-specific. Moreover, various biochemical studies and the detailed crystal structure showed important residues for GA binding and provided a clear understanding of the GA perception mechanism. In addition, recent studies of DELLA functions and crystal structures in the GRAS family have added valuable new information. Future investigations will increase the understanding of these regulators, which are involved in plant-specific signaling networks.

Acknowledgments This work was supported by a Grants-in-Aid for Scientific Research on Innovative Areas (No. 3806) and a MEXT KAKENHI Grant (16H04907). The authors declare that they do not have any competing financial interests.

References

Achard P, Genschik P (2009) Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. J Exp Bot 60(4):1085–1092

Aleman L, Kitamura J, Abdel-Mageed H, Lee J, Sun Y, Nakajima M, Allen RD (2008) Functional analysis of cotton orthologs of GA signal transduction factors GID1 and SLR1. Plant Mol Biol 68(1–2):1–16

Arnaud N, Girin T, Sorefan K, Fuentes S, Wood TA, Lawrenson T, Østergaard L (2010) Gibberellins control fruit patterning in Arabidopsis thaliana. Genes Dev 24(19):2127–2132

Bolle C (2004) The role of GRAS proteins in plant signal transduction and development. Planta 218 (5):683–692

Boter M, Ruíz-Rivero O, Abdeen A, Prat S (2004) Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes Dev 18(13):1577–1591

Davière JM, Achard P (2013) Gibberellin signaling in plants. Development 140:1147–1151

Feng S, Martinez C, Gusmaroli G, Wang Y, Zhou J, Wang F, Deng XW (2008) Coordinated regulation of *Arabidopsis thaliana* development by light and gibberellins. Nature 451 (7177):475–479

Fukazawa J, Teramura H, Murakoshi S, Nasuno K, Nishida N, Ito T, Takahashi Y (2014) DELLAs function as coactivators of GAI-ASSOCIATED FACTOR1 in regulation of gibberellin homeostasis and signaling in Arabidopsis. Plant Cell 26(7):2920–2938

Gallego-Bartolomé J, Alabadí D, Blázquez MA (2011) DELLA-induced early transcriptional changes during etiolated development in *Arabidopsis thaliana*. PLoS One 6(8):e23918

Griffiths J, Murase K, Rieu I, Zentella R, Zhang Z-L, Powers SJ, Thomas SG (2006) Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. Plant Cell 18(12):3399–3414

Guo S, Xu Y, Liu H, Mao Z, Zhang C, Ma Y, Zhang Q, Meng Z, Chong K (2013) The interaction between OsMADS57 and OsTB1 modulates rice tillering via DWARF14. Nat Commun 4:1566

Hamiaux C, Drummond RS, Janssen BJ, Ledger SE, Cooney JM, Newcomb RD, Snowden KC (2012) DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. Curr Biol 22(21):2032–2036

Hirano K, Nakajima M, Asano K, Nishiyama T, Sakakibara H, Kojima M, Matsuoka M (2007) The GID1-mediated gibberellin perception mechanism is conserved in the Lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. Plant Cell 19(10):3058–3079

Hirano K, Ueguchi-Tanaka M, Matsuoka M (2008) GID1-mediated gibberellin signaling in plants. Trends Plant Sci 13(4):192–199

Hirano Y, Nakagawa M, Suyama T, Murase K, Shirakawa M, Takayama S, Hakoshima T (2017) Structure of the SHR–SCR heterodimer bound to the BIRD/IDD transcriptional factor JKD. Nat Plants 3(17):17010

Hou X, Lee LYC, Xia K, Yan Y, Yu H (2010) DELLAs modulate Jasmonate signaling via competitive binding to JAZs. Dev Cell 19(6):884–894

Hou X, Ding L, Yu H (2013) Crosstalk between GA and JA signaling mediates plant growth and defense. Plant Cell Rep 32(7):1067–1074

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, Yamaguchi J (2001) Slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. Plant Cell 13 (5):999–1010

Israelsson M, Sundberg B, Moritz T (2005) Tissue-specific localization of gibberellins and expression of gibberellin-biosynthetic and signaling genes in wood-forming tissues in aspen. Plant J 44 (3):494–504

Itoh H, Matsuoka M, Steber CM (2003) A role for the ubiquitin-26S-proteasome pathway in gibberellin signaling. Trends Plant Sci 8(10):492–497

Itoh H, Ueguchi-Tanaka M, Matsuoka M (2008) Molecular biology of gibberellins signaling in higher plants. Int Rev Cell Mol Biol 268:191–221

Iuchi S, Suzuki H, Kim YC, Iuchi A, Kuromori T, Ueguchi-Tanaka M, Nakajima M (2007) Multiple loss-of-function of Arabidopsis gibberellin receptor AtGID1s completely shuts down a gibberellin signal. Plant J 50(6):958–966

Kagiyama M, Hirano Y, Mori T, Kim SY, Kyozuka J, Seto Y, Yamaguchi S, Hakoshima T (2013) Structures of D14 and D14L in the strigolactone and karrikin signaling pathways. Genes Cells 18(2):147–160

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012) The Arabidopsis information resource (TAIR): improved gene annotation and new tools

Lee S, Cheng H, King KE, Wang W, He Y, Hussain A, Peng J (2002) Gibberellin regulates Arabidopsis seed germination via RGL2, a GAI/RGA-like gene whose expression is up-regulated following imbibition. Genes Dev 16:646–658

Li S, Zhao Y, Zhao Z, Wu X, Sun L, Liu Q, Wu Y (2016) Crystal structure of the GRAS domain of SCARECROW-LIKE 7 in *Oryza sativa*. Plant Cell 28(5):1025–1034

Lucas M, Davière J-M, Rodríguez-Falcón M, Pontin M, Iglesias-Pedraz JM, Lorrain S, Prat S (2008) A molecular framework for light and gibberellin control of cell elongation. Nature 451 (7177):480–484

MacMillan J (2002) Occurrence of gibberellins in vascular plants, fungi and bacteria. J Plant Growth Regul 20:387–442

McGinnis KM, Thomas SG, Soule JD, Strader LC, Zale JM, Sun T, Steber CM (2003) The Arabidopsis SLEEPY1 gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. Plant Cell 15(5):1120–1130

Mindrebo JT, Nartey CM, Seto Y, Burkart MD, Noel JP (2016) Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. Curr Opin Struct Biol 41:233–246

Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJ, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD (2015) The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res 43:D213–D221

Murase K, Hirano Y, Sun T, Hakoshima T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. Nature 456(7221):459–463

Nakajima M, Shimada A, Takashi Y, Kim YC, Park SH, Ueguchi-Tanaka M, Yamaguchi I (2006) Identification and characterization of Arabidopsis gibberellin receptors. Plant J 46(5):880–889

Olszewski N, Sun T-P, Gubler F (2002) Gibberellin signaling: biosynthesis, catabolism, and response pathways. Plant Cell 14:S61–S80

Osterlund T (2001) Structure-function relationships of hormone-sensitive lipase. Eur J Biochem 268(7):1899–1907

Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP (1997) The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev 11(23):3194–3205

Radley M (1956) Occurrences of substances similar to gibberellic acid in higher plants. Nature 178:1070–1071

Sasaki A, Itoh H, Gomi K, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Matsuoka M (2003) Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. Science 299(5614):1896–1898

Shimada A, Ueguchi-Tanaka M, Nakatsu T, Nakajima M, Naoe Y, Ohmiya H, Matsuoka M (2008) Structural basis for gibberellin recognition by its receptor GID1. Nature 456(7221):520–523

Silverstone AL, Jung HS, Dill A, Kawaide H, Kamiya Y, Sun TP (2001) Repressing a repressor: gibberellin-induced rapid reduction of the RGA protein in Arabidopsis. Plant Cell 13 (7):1555–1566

Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Jansson S (2004) A Populus EST resource for plant functional genomics. Proc Natl Acad Sci U S A 101 (38):13951–13956

Suzuki H, Park SH, Okubo K, Kitamura J, Ueguchi-Tanaka M, Iuchi S, Nakajima M (2009) Differential expression and affinities of Arabidopsis gibberellin receptors can explain variation in phenotypes of multiple knock-out mutants. Plant J 60(1):48–55

Tyler L, Thomas SG, Hu J, Dill A, Alonso JM, Ecker JR, Sun T-P (2004) Della proteins and gibberellin-regulated seed germination and floral development in Arabidopsis. Plant Physiol 135(2):1008–1019

Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Matsuoka M (2005) GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature 437(7059):693–698

Ueguchi-Tanaka M, Nakajima M, Katoh E, Ohmiya H, Asano K, Saji S, Matsuoka M (2007a) Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. Plant Cell 19(7):2140–2155

Ueguchi-Tanaka M, Nakajima M, Motoyuki A, Matsuoka M (2007b) Gibberellin receptor and its role in gibberellin signaling in plants. Annu Rev Cell Dev Biol 58:183–198

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2(12):e1326

Wen C-K, Chang C (2002) Arabidopsis RGL1 encodes a negative regulator of gibberellin responses. Plant Cell 14(1):87–100

Willige BC, Ghosh S, Nill C, Zourelidou M, Dohmann EMN, Maier A, Schwechheimer C (2007) The DELLA domain of GA INSENSITIVE mediates the interaction with the GA INSENSITIVE DWARF1A gibberellin receptor of Arabidopsis. Plant Cell 19(4):1209–1220

Yabuta T, Sumiki Y (1938) On the crystal of gibberellin, a substance to promote plant growth. J Agric Chem Soc Jpn 14:1526

Yamaguchi S (2008) Gibberellin metabolism and its regulation. Annu Rev Plant Biol 59 (1):225–251

Yasumura Y, Crumpton-Taylor M, Fuentes S, Harberd NP (2007) Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land-plant evolution. Curr Biol 17 (14):1225–1230

Yeaman SJ (2004) Hormone-sensitive lipase--new roles for an old enzyme. Biochem J 379:11–22

Yoshida H, Hirano K, Sato T, Mitsuda N, Nomoto M, Maeo K, Ueguchi-Tanaka M (2014) DELLA protein functions as a transcriptional activator through the DNA binding of the indeterminate domain family proteins. Proc Natl Acad Sci U S A 111(21):7861–7866

Zentella R, Zhang Z-L, Park M, Thomas SG, Endo A, Murase K, Sun T (2007) Global analysis of della direct targets in early gibberellin signaling in Arabidopsis. Plant Cell 19(10):3037–3057

Zhao J, Wang T, Wang M, Liu Y, Yuan S, Gao Y, Yin L, Sun W, Peng L, Zhang W, Wan J, Li X (2014) DWARF3 participates in an SCF complex and associates with DWARF14 to suppress rice shoot branching. Plant Cell Physiol 55(6):1096–1109

Chapter 7
Strigolactone and Karrikin Signaling Proteins

Toshio Hakoshima

7.1 Introduction

7.1.1 Strigolactone (SL)

Strigolactone (SL) is a terpenoid lactone product derived from the metabolic pathway of carotenoids (Matusova et al. 2005) and consists of a tricyclic lactone (ABC ring) linked to a butenolide moiety (D ring) by an enol–ether bond (Fig. 7.1a). The butenolide ring is conserved in carlactone-derived compounds showing SL-like activity but display diverse chemical structures. At least 20 different naturally occurring SL compounds have been characterized (Tokunaga et al. 2015). Generally, SLs are chemically unstable and the concentration of naturally occurring SLs in plants is relatively low. These obstacles that impeded investigations of SLs were overcome by the synthesis of stable SL analogs such as the GR series of compounds, named after their inventor Gerald Rosebery (Zwanenburg and Pospíšil 2013). Of these analogs, GR24 is the most active and widely used in SL research (Besserer et al. 2008) (Fig. 7.1b). The SL biosynthesis pathway is mediated by cytochrome P450-type oxygenases containing CAROTENOID CLEAVAGE DIOXYGENASE 7 (CCD7) and CCD8, which are encoded by genes such as *RAMOSUS 5* (*RMS*5)/*MORE AXILLARY GROWTH* 3 (*MAX3*) and *RMS1*/*MAX4*. Mutations of these genes display an excess branching phenotype (Sorefan et al. 2003; Booker et al. 2004), suggesting that SLs function as branch-inhibiting hormones (Umehara et al. 2008; Gomez-Roldan et al. 2008). SLs were first discovered in root exudates as bioactive substances that stimulate germination of the parasitic weed *Striga lutea* (Cook et al.

T. Hakoshima (✉)
Structural Biology Laboratory, Nara Institute of Science and Technology, Takayama, Japan
e-mail: hakosima@bs.naist.jp

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_7

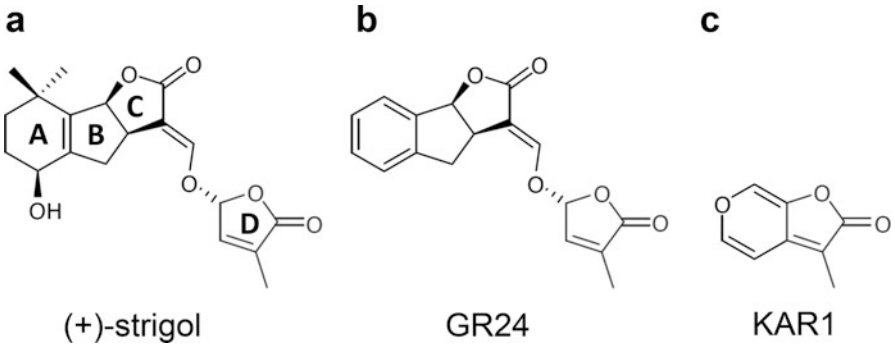


Fig. 7.1 Chemical structures of SL and KAR. (**a**) Strigol is the first discovered SL, (**b**) GR24, (**c**) KAR1

1966) and other parasitic weeds such as *Orobanche* and *Alectra* (Yoneyama et al. 2010, for review), with the designation of SLs being based on their functional similarity with the first SL, strigol, and the lactone ring-containing chemical structure.

7.1.2 Karrikin (KAR)

Karrikin (KAR) is not a product of the plant itself, but is formed by heating or combustion of carbohydrates such as cellulose (Flematti et al. 2011). In fact, the first KAR was isolated from smoke that promotes seed germination and enhances seedling responses to light (Flematti et al. 2004). The chemical structure of KAR apparently shares a substituted butenolide moiety with SLs and the pyran ring is important for the induction of seed germination (Flematti et al. 2007, 2010) (Fig. 7.1c). Smoke elicits a significant ecological impact on post-fire environments and stimulates the germination of over a thousand plant species. It is significant that designation of karrikin is based on the Australian aboriginal term "karrik" for smoke. So far, six KARs (KAR1 to KAR6), which differ in their methyl substitutions, have been identified in plant-derived smoke and KAR1 is the main seed germination stimulant (Flematti et al. 2009). The major biological function of KARs is seed germination stimulation, which makes KARs indistinguishable from real plant hormones. This situation is reminiscent of endocrine disruptors or environmental pollutants acting on signaling pathways in humans. These disruptors/pollutants are collectively designated as "environmental hormones" in Japanese.

7.1.3 The SL and KAR Signaling Pathways

SLs and KARs share highly similar perception and signaling components and pathways. In addition to metabolic enzymes such as MAX3 and MAX4, two gene

products are known to be essential in the SL signaling pathway (Yamaguchi and Kyozuka 2010; Beveridge and Kyozuka 2010). One encodes D3/MAX2 (rice/Arabidopsis gene nomenclature), which is a leucine-rich repeat (LRR)-type F-box protein similar to TIR1 and COI1, which are auxin and jasmonic acid (JA) receptors, respectively. D3/MAX2 forms a SKP1-CULLIN-F-box (SCF), an E3 ubiquitin ligase complex that disrupts downstream target proteins that act as negative regulators of branching through ubiquitin-mediated protein degradation (Stirnberg et al. 2002; Ishikawa et al. 2005). The other is the *D14* gene product, which belongs to the α/β-fold hydrolase superfamily (Arite et al. 2009; Ishikawa et al. 2005). D14 was naturally expected to assume the role of SL receptor that directly perceives SL since a member of the α/β-fold hydrolase superfamily, GID1, is known to act as the gibberellin (GA) receptor (Murase et al. 2008; Ueguchi-Tanaka et al. 2005). Importantly, GR24 promotes physical interactions between D14 and D3/MAX2 (Hamiaux et al. 2012; Zhao et al. 2014, 2015), suggesting SL-dependent formation of the SCF E3 ligase complex containing D14. This is analogous to formation of the SCF E3 ligase complex with GID1 in GA signaling, where GID1 functions as an adaptor protein of GA-dependent recognition of substrate DELLA proteins. Genetic and biochemical studies have revealed that the substrate of the $SCF^{D3/MAX2\text{-}D14}$ E3 ligase complex is D53/SUPPRESSOR OF MAX2 1 (SMAX1)-LIKE (SMXL) (Jiang et al. 2013; Zhou et al. 2013). D53/SMXL, which displays weak homology to an ATPase protein of the ClpB/HEAT SHOCK PROTEIN100 (HSP100) class, has been shown to function as a signaling repressor that is degraded upon treatment with GR24, although the mechanisms involved in the molecular function of these proteins remain unknown (Jiang et al. 2013; Stanga et al. 2013; Zhou et al. 2013). Among the eight members comprising the SMXL family (Stanga et al. 2013), SMXL6, SMXL7, and SMXL8 (SMXL678) are D53 co-orthologs and mediate SL signaling (Soundappan et al. 2015) (Fig. 7.2a).

The D14 homolog D14-LIKE (D14L) in rice and D14 paralog KARRIKIN INSENSITIVE 2 (KAI2) have been characterized as KAR receptors in rice and Arabidopsis, respectively (Arite et al. 2009; Hamiaux et al. 2012). D14 and D14L/KAI2 possess relatively high amino acid sequence identity (~50%) (Arite et al. 2009). In the KAR signaling pathway, D3/MAX2 is common but the substrate proteins are SMAX1 and SMAX2 (Stanga et al. 2013, 2016; Soundappan et al. 2015) (Fig. 7.2b).

In sharp contrast to the GA receptor GID1, which lacks the His residue of the Ser-His-Asp catalytic triad, the Ser-His-Asp catalytic triad is strictly conserved in D14, which is essential for both SL hydrolysis and signaling functions (Hamiaux et al. 2012; Nakamura et al. 2013; Waters et al. 2015a). The rate of SL hydrolysis by D14, however, was found to be too slow (50% loss in 3 h in vitro) to function as a key enzyme in the production of a bioactive SL-derived compound that may be perceived by another protein (Hamiaux et al. 2012; Zhao et al. 2013). In fact, neither of the final products of SL hydrolysis was found to act as shoot branching suppression signals (Hamiaux et al. 2012; Nakamura et al. 2013; Zhao et al. 2013). As with D14, the Ser-His-Asp catalytic triad is conserved in D14L/KAI2, and is also required for KAI2 function (Waters et al. 2015b). KAR hydrolysis, however, is not expected given its chemical properties (Zhao et al. 2013, 2015).

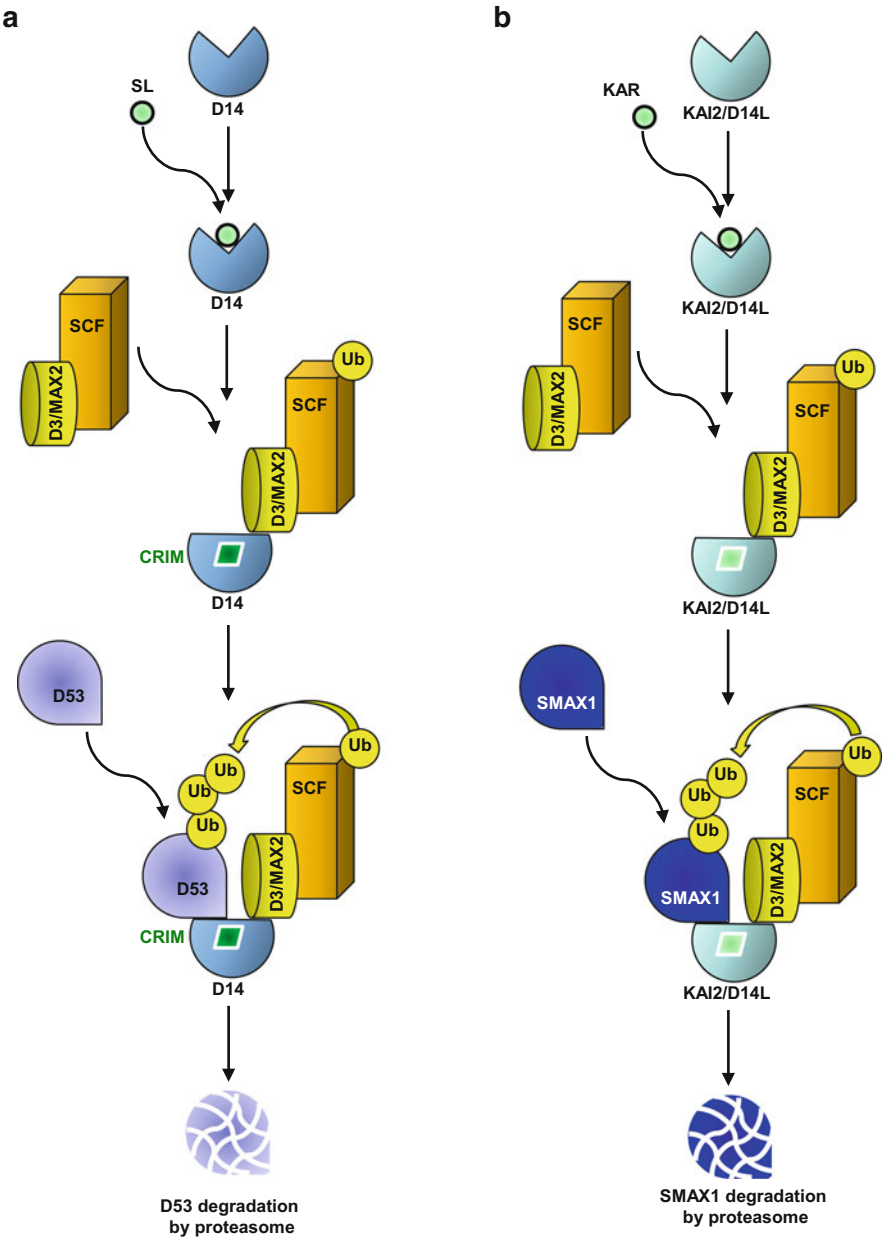


Fig. 7.2 SL and KAR signaling pathways. (**a**) The SL pathway. (**b**) The KAR pathway

7.2 Physical Data for Direct Binding

D14 was initially assumed to be the SL receptor given the available indirect evidence at the time, which comprised observations of SL-dependent DAD2 binding to PhMAX2A, identification of Petunia D3/MAX2, and the observed destabilization of D14 by addition of SL using differential scanning fluorimetry (DSF) (Hamiaux et al. 2012). These observations, however, did not necessarily mean that D14 is the SL receptor. The first direct evidence indicating the direct binding of SL to D14 was provided by isothermal titration calorimetry (ITC) (Kagiyama et al. 2013). Titration of D14 with GR24 and of AtD14L/KAI2 with KAR1 each gave rise to sigmoidal titration curves, strongly suggesting direct specific binding. The dissociation constant K_D is 300 nM for GR24 binding to AtD14. Interestingly, the K_D for KAR1 binding to AtD14L/KAI2 is 4.6 μM, so that the binding affinity of KAR1 to KAI2 is one order of magnitude weaker than that of D14. The thermodynamic parameters indicate that both bindings are entropy-driven rather than enthalpy-driven. The hormone-binding site in the free state is probably occupied with water molecules, and hormone-binding to this site is accompanied by the release of bound water molecules, thereby resulting in an increase in entropy. The major difference between GR24-D14 and KAR1-KAI2 binding is the smaller enthalpy change (ΔH) associated with KAR1-KAI2 binding. This is reasonable since the molecular size of KAR1 is smaller than that of GR24, suggesting loss of protein–ligand contact area in KAR1-KAI2 binding. It should be noted that no heat was observed in the ITC experiment involving KAI2 being titrated with GR24, indicating the absence of GR24 binding to KAI2, while the absence of a saturation titration curve for the titration of D14 with KAR1 suggests nonspecific binding of KAR1 to D14. The difference could be attributable to the difference in size of the active site pockets: the pocket of KAI2 being smaller than that of D14 (see below).

7.3 Receptor Structure

7.3.1 Strigolactone Receptor

Crystal structures of SL receptors have been reported for petunia (*Petunia hybrida*) D14 ortholog DAD2, rice (*Oryza sativa*) D14, and Arabidopsis (*Arabidopsis thaliana*) AtD14 (Hamiaux et al. 2012; Kagiyama et al. 2013; Bythell-Douglas et al. 2013; Zhao et al. 2013). Like many other rice proteins, rice D14 has an N-terminal nonconserved region of 54 residues containing repeated Gly or Ser residues. These residues were absent in the structural study. The C-terminal α/β-hydrolase domain is well conserved with 74% amino acid sequence identity with Arabidopsis, and the domain structure was determined at a high resolution (1.45 Å) (Fig. 7.3a, b). The α/β-hydrolase domains of all SL receptors from rice, Arabidopsis, and petunia display essentially the same structure, which is composed

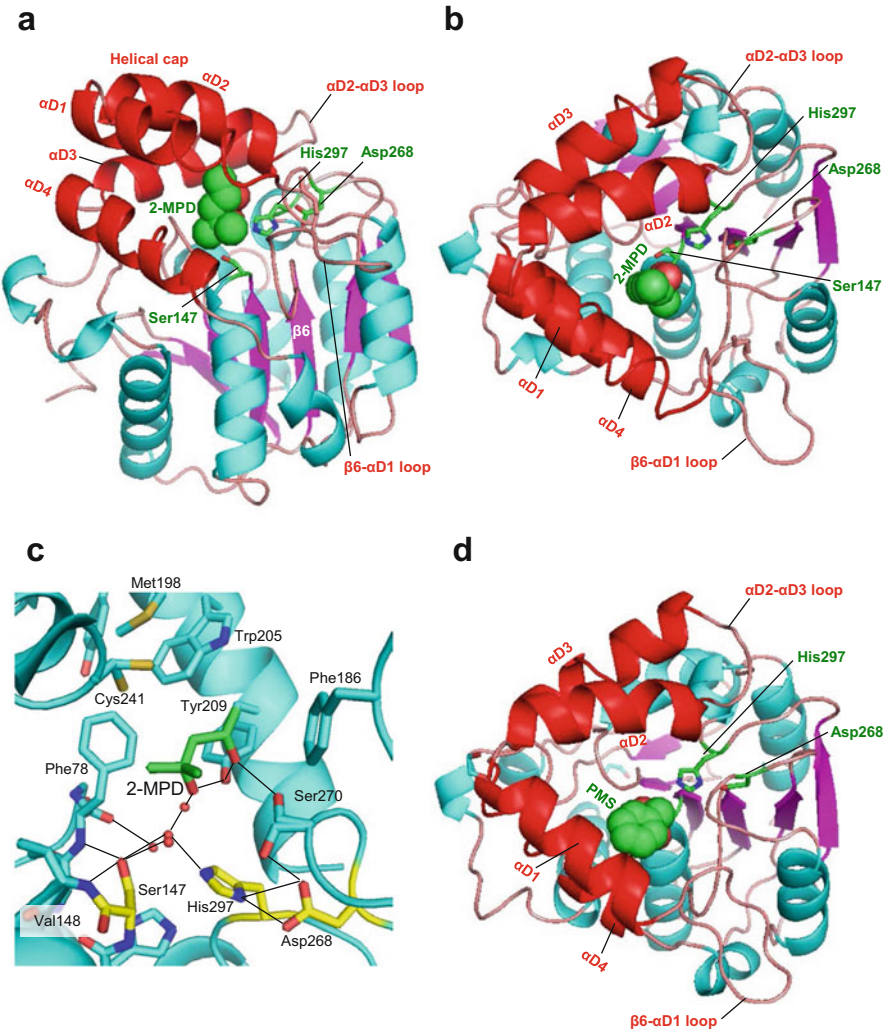


Fig. 7.3 D14 structures. (**a**) The free form (a side view). (**b**) The free (a top view). (**c**) The active site contains 2MPD and water molecules. (**d**) The PMS-bound form. The free form contains 2-MPD. Both 2-MPD and PMS-linked D14 structures (PDB ID 3W04 and 3W05 by Kagiyama et al. 2013, respectively) are designated as the "open state"

of seven β-strands forming the central β-sheet and ten α-helices. Compared with the canonical α/β hydrolase domain (Nardini and Dijkstra 1999; Heikinheimo et al. 1999), D14 and other SL receptors have an additional insertion of four α-helices (αD1-αD4) between β6 and αD (Fig. 7.4). The inserted four helices collectively form a "helical cap," as with members of the α/β hydrolase family (Nardini and Dijkstra 1999), which covers the active site pocket, while a small hole remains open to the solvent region (Fig. 7.4b). Similar structures were obtained from DAD2 and AtD14,

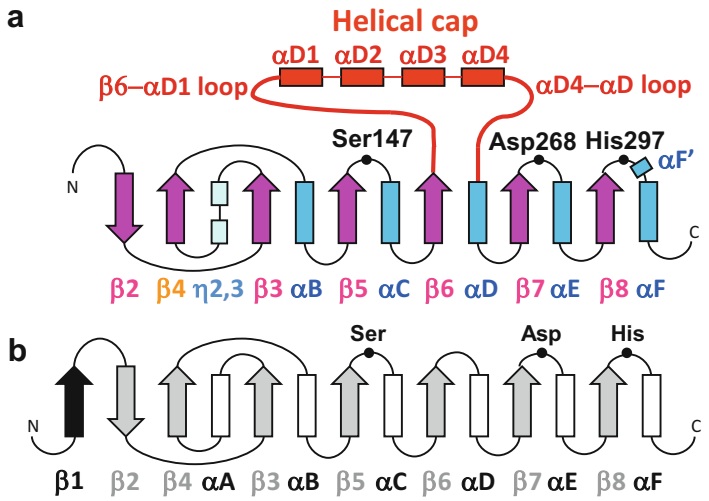


Fig. 7.4 Topology of D14. D14 displays an α/β hydrolase fold with insertion of four helices forming the helical cap (**a**) D14, (**b**) The core domain of the canonical α/β-hydrolase

and are designated as the "open state" forms. At the active site pocket, we found 2-methyl-pentane-diol (2-MPD), a precipitant from the crystallization solution, bound to a cluster of four water molecules hydrogen bonded to Ser147 and His297 of the catalytic triad. The 2-MPD is located at the nonpolar site close to the entrance of the pocket (Fig. 7.3c). The helical cap is distinct from the "helical lid" found in GID1: the helical lid is formed by three α-helices and is induced in the N-terminal switch region by GA binding (Murase et al. 2008). The Ser-His-Asp catalytic triad (Ser147, His297 and Asp268 in rice D14) is located in the catalytic pocket in a strictly conserved orientation and the conserved oxyanion hole formed with two main chain amide groups (Val148 and Phe78), suggesting its intrinsic catalytic activity (Fig. 7.3c). The oxyanion hole is well known to be essential for stabilization of the transition intermediate by neutralizing the localized charges on the intermediate. This activity is well demonstrated by formation of Ser-modified D14 following mixing of D14 with phenyl-methane-sulfonyl fluoride (PMSF), a typical inhibitor of serine proteases possessing the Ser-His-Asp catalytic triad. The structure of the modified D14 shows that the PMS-bound Ser residue at the active center and the phenyl group of PMS are trapped in the nonpolar site of the pocket (Fig. 7.3d). This nonpolar site corresponds to the 2-MPD binding site and also the KAR binding site of KAI2 (see below).

7.3.2 Karrikin Receptor

The KAR receptors display an overall structure similar to that of D14 with a small root-mean-square (rms) deviation (1.2 Å). As with D14, the α/β-hydrolase domain is

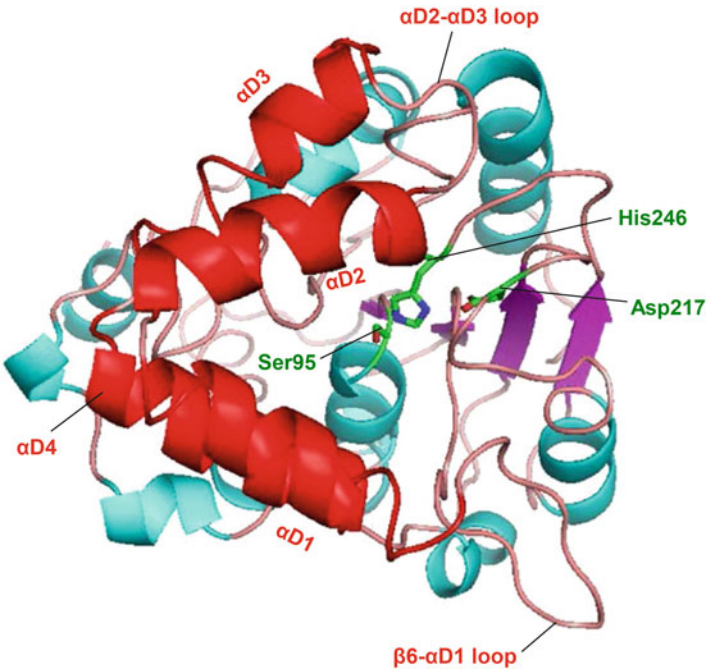


Fig. 7.5 AtD14L/KAI2 structure. (PDB ID 3W06 by Kagiyama et al. 2013)

covered by a helical cap formed by four α-helices with the conserved configuration of the catalytic triad residues (Ser95, His246 and Asp217) (Fig. 7.5). Two prominent local deviations are found at loop regions β6-αD1 loop and αD2-αD3 loop. The shift in the αD2-αD3 loop, which is part of the helical cap, is interesting since it seems to be induced by a one-residue (Asp167) insertion in AtD14L/KAI2. This site is far from the active site and the effect of this D14L/KAI2-specific insertion on activity is unclear. Interestingly, αD2 and αD3 helices of AtD14 undergo a marked conformational change when bound to D3/MAX2, and this insertion may have some effect on the open-to-closed state transition (Yao et al. 2016). Importantly, β6-αD1 loop is facing toward the active site as part of the cap and makes contacts with αD4 helix. Through these contacts, the shift in this loop is accompanied with a shift in the C-terminal end of αD4 helix, and results in narrowing of the pocket entrance and also the active site space around the catalytic triad (Fig. 7.5). More importantly, nonpolar residues forming the binding site are replaced with more bulky residues, for example Val is replaced with Ile residues, in AtD14L/KAI2. These replacements result in a smaller pocket, which could account for the lack of binding of GR24 to D14L/KAI2, as shown by our ITC experiment.

7.3.3 Hormone-Bound Forms of SL and KAR Receptors

Attempts to obtain cocrystals of the receptor and hormone failed, probably due to the catalytic action of the receptors. In an effort to overcome that obstacle, one approach was to engineer and use an enzyme with no activity, for example by mutating one of the catalytic residues. However, these engineered D14 or KAI2 proteins were found to exhibit worse physical properties such as instability or poor solubility. The first cocrystal structure elucidating the bound hormone molecule appeared for the KAI2-KAR1 complex at a high (1.3 Å) resolution (Guo et al. 2013). In this crystallization, a relatively high concentration of KAR1 (5 mM) was employed at pH 6.5. The KAR1 ligand was found to sit in the opening to the active site distal from the catalytic triad (Fig. 7.6). This binding site is comprised of aromatic nonpolar residues, which encapsulate the KAR1 ring. It is noteworthy that the binding site includes aromatic residues (Phe134 and Phe194) from αD4 helix and β6-αD1 loop, and differs conformationally from D14 as described above. Nevertheless, the KAR1-binding site corresponds to the nonpolar binding site for 2-MPD found in the D14 crystal (Kagiyama et al. 2013) (Fig. 7.3a, b) and also utilized for the binding site of CRIM (see below). Structural comparison of KAR1-bound KAI2 with apo KAI2 reveals that the two structures are similar, with small conformational shifts being found in the side chains. This result was somewhat unexpected, but is also true for

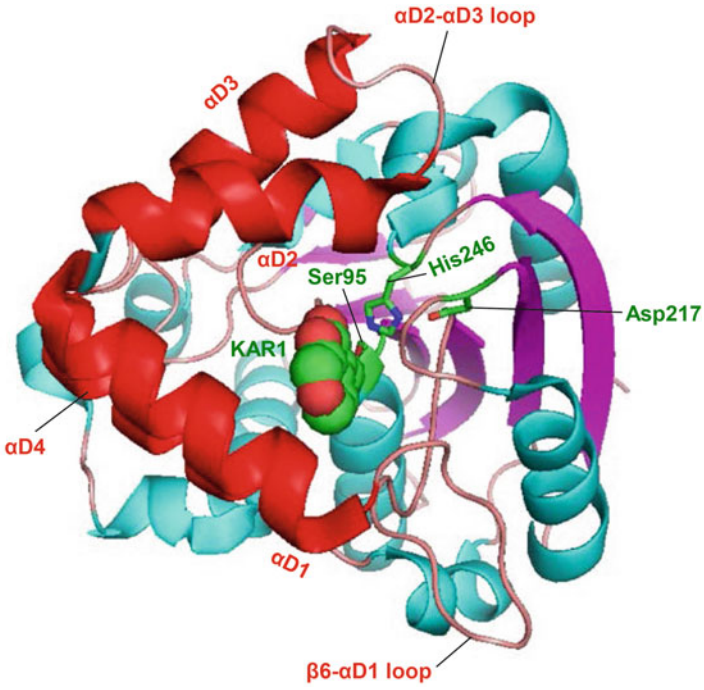


Fig. 7.6 A KAR1-bound KAI2 structure. (PDB ID 4JYM by Guo et al. 2013)

D14. An attempt to crystallize the D14-GR24 complex resulted in a structure of AtD14 with bound GR24 without significant global conformational change from apo AtD14, although the electron density for the bound GR24 is poor and details of the configuration of the bound GR24 remain obscure (Zhao et al. 2015). Moreover, no global conformational change was observed in the crystal structure of D14 covalently bound to a GR24 hydrolysis intermediate (Zhao et al. 2013) or in a crystal structure of D14 bound to a D-ring hydrolysis product (Nakamura et al. 2013). It is an important conclusion from our structural biology studies that no global conformational change occurs on GR24/KAR1 binding to D14/KAI2, which implies that the conformational switch that should occur on these receptors on hormone binding requires another protein/factor that may associate with the receptor. It is obvious that our long-awaited structures to be determined are structures of higher-level complexes such as that of D14-D3 or D14-D3-D53 (see also below).

7.4 Signaling Protein Complex Structure

7.4.1 CLIM Found in the AtD14-D3-ASK1 Complex

Since D14 and KAI2 are folded into an α/β hydrolase with the Ser-His-Asp catalytic triad and exhibit catalytic activity, it might be possible to isolate the intermediate complex for crystallization. D14 is found to catalyze GR24 hydrolysis via intermediate 2,4,4,-trihydroxy-3-methyl-3-butenal (TMB), with the final products being hydroxymethyl butenolide (D-OH) and tricyclic lactone (ABC-OH) (Hamiaux et al. 2012; Nakamura et al. 2013; Zao et al. 2013). Notably, the presence of D3 inhibits AtD14-mediated GR24 hydrolysis and suppresses the release of one of the final products, hydroxymethyl butenolide (D-OH). This observation suggests that D3-bound AtD14 could retain a hydrolytic D-ring-derived intermediate inside the closed active site pocket of AtD14 for SL signaling. In an effort to investigate the mechanism, the cocrystal structure of AtD14 with rice F-box protein D3 and Arabidopsis SKP1 homolog ASK1, hereafter referred to as the AtD14-D3-ASK1 complex, was determined at 3.3 Å resolution (Fig. 7.7) (Yao et al. 2016). Indeed, electron density was found between two active residues, Ser97 and His247, suggesting that a small molecule links the two residues, and was interpreted to comprise a $C_5H_5O_2$ modification as suggested by tandem mass spectrometry (MS/MS) analysis. Importantly, the same $C_5H_5O_2$ modification was also detected *in planta* when plants were treated with GR24. The $C_5H_5O_2$ modification is referred to as a covalently linked intermediate molecule (CLIM), which is derived from the D-ring as a result of SL hydrolysis (Fig. 7.7).

A study of RAMOSUS3 (RMS3), a pea (*Pisum sativum*) ortholog of D14, showed formation of a covalent RMS3-D-ring complex, in which the D ring was

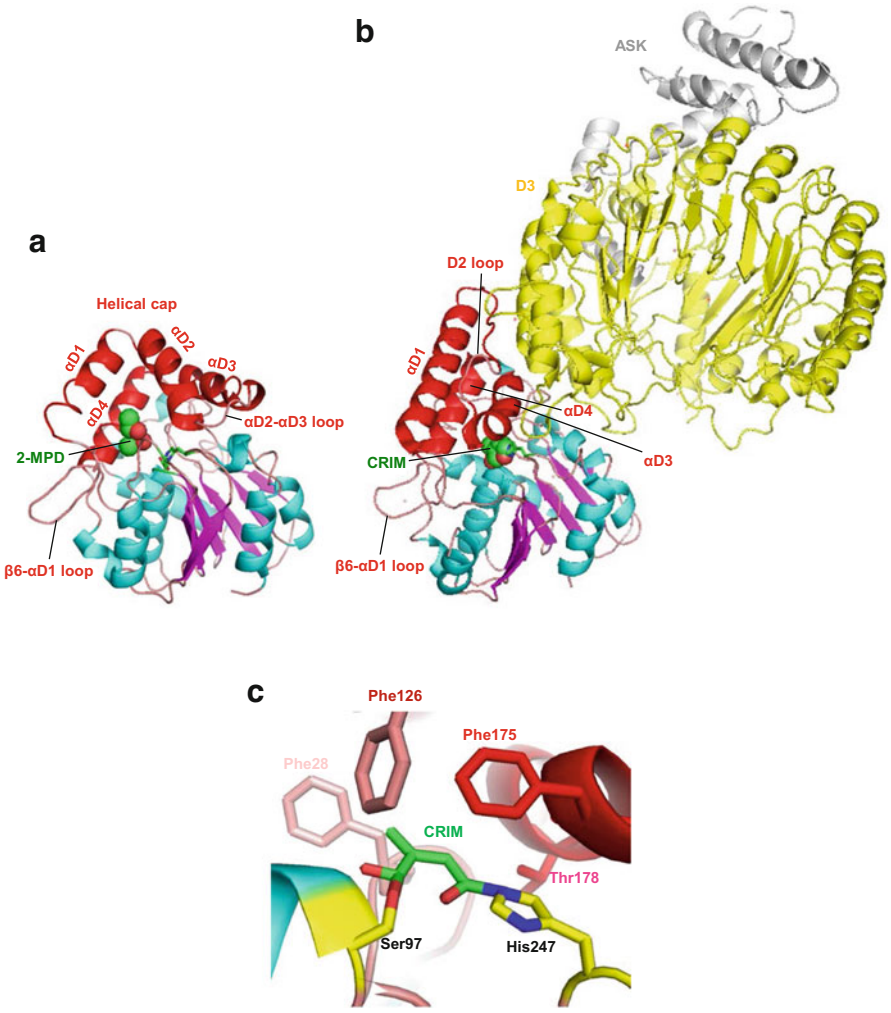


Fig. 7.7 The At14-D3-ASK complex. (**a**) The free form of D14. (**b**) The overall structure of the At14-D3-ASK complex. (PDB ID 4JYM by Guo et al. 2013). (**c**) A close-up view of CRIM

attached to His247 of the catalytic triad, and that the formation is essential for bioactivity (de Saint Germain et al. 2016). Isolation of the RMS3-D-ring complex as well as the AtD14-CLIM complex suggests that the receptors perform an irreversible enzymatic reaction to generate its own ligand and are subjected to degradation in a MAX2-dependent and MAX2-independent manner, respectively (Chevalier et al. 2014; Waters et al. 2015a).

7.4.2 The Open-to-Closed Conformational Transition

The structure of the At14-D3-ASK complex revealed dramatic conformational changes in AtD14 upon binding to D3 in the presence of GR24. Compared to the apo AtD14 structure in the "open state" (Fig. 7.7a), the structure of AtD14 bound to D3 undergoes significant conformational changes to form a new structure in the closed state (Fig. 7.7b). This open-to-closed transition is achieved by collapse of the helical cap formed by four α-helices in the open state. In the closed state, the helical cap is formed by three α-helices with the helix-to-coil transition of the second α-helix (the αD2 helix in the D3-free state, Fig. 7.7a) coupled by the coil-to-helix transition of the loop between the first and the second α-helices of the helical cap. The conformational transition is obviously induced by binding to D3, which forms a Leucine-rich repeat (LRR) domain and provides the C-terminal part of the LRR domain for direct binding to the helical cap of AtD14. As a result of the transition, the closed state of AtD14 contains a small closed active site pocket (80 $Å^3$), which is in contrast to a large open active site pocket (420 $Å^3$). The small pocket seems to be suitable for embedding small molecules (such as CLIM). Notably, the catalytic residues S97 and H247 remain almost in the original positions of the catalytic triad, and the Ser-CLIM-His segment at the catalytic center is sealed in a hydrophobic cavity together with a subset of nonpolar residues (including Phe28, Phe126, Phe175, Leu179, and Val194), the side chains of which are shifted by the induced fit to contact CLIM. These relatively small conformational shifts trigger the open-to-closed transition. As described above, most of these nonpolar residues comprise the nonpolar 2-MPD/PMS-binding sites of D14 and KAR1-binding site of KAI2, both in the closed forms. Unlike the catalytic residues S97 and H247, the other catalytic residue D218 moves away from the catalytic center by disordering of the loop containing D218. This destabilized loop may be involved in interaction with other proteins such as the repressors D53/SMXLs in an SL-dependent manner. However, there is no experimental evidence for the D53/SMXL-binding site at present. Further structural and biophysical studies are needed to provide experimental evidence for the binding site.

7.4.3 Proteins Assisting Formation of Transition-State Complexes

The SL receptor is an enzyme with relatively poor activity in the isolated state. However, the catalytic activity is essential for SL signaling. Like the SL receptor, some enzymes are known to exhibit only poor activity. Usually, these enzymes interact with assisting proteins and the poor catalytic activity is accelerated by binding to the assisting protein, which accelerates formation of a transition-state complex. A typical example of such assisting proteins is the GTPase activating protein (GAP). Formation of a GTPase-GAP complex stabilizes the transition state

of the hydrolysis of GTP to yield GDP and inorganic phosphate. The GTPase-GAP complex is a transition complex that is intrinsically unstable and is swiftly transferred to the next state, in which product molecules are delivered. In an effort to obtain a stable transition-state complex between GTPase RhoA and RhoGAP, crystallization was performed using an artificial transition-state analog compound, GDP-AlF_4^-, which mimics the intermediate formed by attack of a water molecule on the γ-phosphate of GTP. The obtained crystal structure reveals how GAP accelerates the catalytic reaction in the three-dimensional molecular complex, where a critical Arg residue, referred to as the arginine finger, is inserted into the active site of the GTPase and stabilizes the transition-state molecules by neutralizing the localized electron charges on the γ-phosphate (Rittinger et al. 1997a). Using a non-hydrolyzable GTP analog, GMPPNP, GAP, and GTPase form a stable ground-state complex that can be crystallized. The obtained structure of the ground-state complex was distinct from that of the transition-state complex and provided no useful information concerning the mechanism by which GAP action stabilizes the transition state (Rittinger et al. 1997b).

D3 may play a role as an assisting protein that accelerates formation of the transition-state complex, but may also function as an inhibitory protein by suppressing the release of one of the final products. These dual functions enable formation of the AtD14-D3 complex, a stable transition-state complex containing an intermediate molecule. I speculate that D53 is another assisting protein and that formation of the D14-D3-D53 complex is the genuine transition-state complex. The observed inhibitory function of D3 is induced by the absence of assisting protein D53. Since D3 is a substrate protein of the SCF E3 complex containing D14 and D3, formation of the transition-state complex enables ubiquitylation of D53, in addition to swift transfer to the state releasing the product D-OH. Thus, the catalysis of SL hydrolysis by D14 is coupled with recognition of the D53 substrate followed by ubiquitylation of D53.

7.5 Concluding Remarks

The last decade has yielded incredible insights pertaining to the SL and KAR signaling pathways. The notion that SL is a plant hormone was recognized in 2008 (Umehara et al. 2008: Gomez-Roldan et al. 2008). Since then, extensive studies have been performed not only at the levels of plant biology and phytochemistry, but at the molecular and structural levels. The main axis of the signaling pathway is SL-D14-D3/MAX2-D53/SMXL678 and KAR-KAI2-D3/MAX2-SMAX1,2. In these pathways, the main switch molecules are the hormone receptors D14 and KAI2, which act as the input device, and where the output device is the ubiquitylation substrate, being D53/SMXL678 for SL and SMAX1,2 for KAR signaling. Unfortunately, the molecular architecture and function of the substrates remains to be determined. Since D14 and KAI2 are not canonical hormone receptors but are also enzymes, an understanding of these proteins requires an investigation of

higher-order complex structures such as the D14-D3-D53 complex. The structure of the CLIM-D14-D3 complex (Yao et al. 2016) shed a strong light on the mechanisms by which the chemical conversion of SL is capable of switching on the signaling pathway. However, several fundamental points remain to be addressed in terms of molecular structures and the chemistry involved. First, our understanding of hormone perception by these closely related receptors remains poor. Additionally, the highly similar structures of D14 and KAI2 and their enzymatic action make it even more difficult for us to provide answers. It could be that perception is executed in a metastable state such as a substate of the transition state of the chemical reaction. Elucidation of SL and KAR hormone perception may provide a clue from the perspective of cross talking or interference between these hormones. The second issue that needs to be addressed is substrate specificity. With this issue I am currently optimistic since solving the structure of the D14-D3-D53 complex is expected to provide an almost complete answer to this question.

References

Arite T, Umehara M, Ishikawa S, Hanada A, Maekawa M, Yamaguchi S, Kyozuka J (2009) d14, a strigolactone insensitive mutant of rice, shows an accelerated outgrowth of tillers. Plant Cell Physiol 50:1416–1424

Besserer A, Bécard G, Jauneau A, Roux C, Séjalon-Delmas N (2008) GR24, a synthetic analog of strigolactones stimulates the mitosis and growth of the arbuscular mycorrhizal fungus Gigaspora rosea by boosting its energy metabolism. Plant Physiol 148:402–413

Beveridge CA, Kyozuka J (2010) New genes in the strigolactone-related shoot branching pathway. Curr Opin Plant Biol 13:34–39

Booker J, Auldridge M, Wills S, McCarty D, Klee H, Leyser O (2004) MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. Curr Biol 14:1232–1238

Bythell-Douglas R, Waters MT, Scaffidi A, Flematti GR, Smith SM, Bond CS (2013) The structure of the karrikin-insensitive protein (KAI2) in Arabidopsis thaliana. PLoS One 8(1):e54758

Chevalier F, Nieminen K, Sánchez-Ferrero JC, Rodríguez ML, Chagoyen M, Hardtke CS, Cubas P (2014) Strigolactone promotes degradation of DWARF14, an α/β hydrolase essential for strigolactone signaling in Arabidopsis. Plant Cell 26:1134–1150

Cook CE, Whichard LP, Turner B, Wall ME, Egley GH (1966) Germination of witchweed (Striga lutea Lour.): isolation and properties of a potent stimulant. Science 154:1189–1190

de Saint Germain A, Clavé G, Badet-Denisot M-A, Pillot J-P, Cornu D, Le Caer J-P, Burger M, Pelissier F, Retailleau P, Turnbull C, Bonhomme S, Chory J, Rameau C, Boyer FD (2016) An histidine covalent receptor and butenolide complex mediates strigolactone perception. Nat Chem Biol 12:787–794

Flematti GR, Ghisalberti EL, Dixon KW, Trengove RD (2004) A compound from smoke that promotes seed germination. Science 305:977–977

Flematti GR, Goddard-Borger ED, Merritt DJ, Ghisalberti EL, Dixon KW, Trengove RD (2007) Preparation of 2H-furo[2.3-c]pyran-2-one derivatives and evaluation of their germination-promoting activity. J Agric Food Chem 55:2189–2194

Flematti GR, Ghisalberti EL, Dixon KW, Trengove RD (2009) Identification of alkyl substituted 2H-furo[2.3-c]pyran-2-ones as germination stimulants present in smoke. J Agric Food Chem 57:9475–9480

Flematti GR, Scaffidi A, Goddard-Borger ED, Heath CH, Nelson DC, Commander LE, Stick RV, Dixon KW, Smith SM, Ghisalberti EL (2010) Structure-activity relationship of karrikin germination stimulants. J Agric Food Chem 58:8612–8617

Flematti GR, Scaffidi A, Dixon KW, Smith SM, Ghisalbert EL (2011) Production of the seed germination stimulant karrikinolide from combustion of simple carbohydrates. J Agric Food Chem 59:1195–1198

Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pages V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC, Bouwmeester H, Bécard G, Beveridge CA, Rameau C, Rochange SF (2008) Strigolactone inhibition of shoot branching. Nature 455:189–194

Guo Y, Zheng Z, La Clair JJ, Chory J, Noel JP (2013) Smoke-derived karrikin perception by the alpha/beta-hydrolase KAI2 from Arabidopsis. Proc Natl Acad Sci U S A 110:8284–8289

Hamiaux C, Drummond RS, Janssen BJ, Ledger SE, Cooney JM, Newcomb RD, Snowden KC (2012) DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. Curr Biol 22:2032–2036

Heikinheimo P, Goldman A, Jeffries C, Ollis DL (1999) Of barn owls and bankers: a lush variety of α/β hydrolases. Struct Fold Des 7:R141–R146

Ishikawa S, Maekawa M, Arite T, Onishi K, Takamure I, Kyozuka J (2005) Suppression of tiller bud activity in tillering dwarf mutants of rice. Plant Cell Physiol 46:79–86

Jiang L, Liu X, Xiong G, Liu H, Chen F, Wang L, Meng X, Liu G, Yu H, Yuan Y, Yi W, Zhao L, Ma H, He Y, Wu Z, Melcher K, Qian Q, Xu HE, Wang Y, Li J (2013) DWARF53 acts as a repressor of strigolactone signalling in rice. Nature 504:401–405

Kagiyama M, Hirano Y, Mor T, Kim SY, Kyozuka J, Seto Y, Yamaguchi S, Hakoshima T (2013) Structures of D14 and D14L in the strigolactone and karrikin signaling pathways. Genes Cells 18:147–160

Matusova R, Rani K, Verstappen FWA, Franssen MCR, Beale MH, Bouwmeester HJ (2005) The strigolactone germination stimulants of the plant-parasitic *striga* and *orobanche* spp. are derived from the carotenoid pathway. Plant Physiol 139:920–934

Murase K, Hirano Y, Sun TP, Hakoshima T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. Nature 456:459–463

Nakamura H, Xue YL, Miyakawa T, Hou F, Qin HM, Fukui K, Shi X, Ito E, Ito S, Park SH, Miyauchi Y, Asano A, Totsuka N, Ueda T, Tanokura M, Asami T (2013) Molecular mechanism of strigolactone perception by DWARF14. Nat Commun 4:2613

Nardini M, Dijkstra BW (1999) α/β hydrolase fold enzymes: the family keeps growing. Curr Opin Struct Biol 9:732–737

Rittinger K, Walker PA, Eccleston JF, Smerdon SJ, Gamblin SJ (1997a) Structure at 1.65 a of RhoA and its GTPase-activating protein in complex with a transition-state analogue. Nature 389:758–762

Rittinger K, Walker PA, Eccleston JF, Nurmahomed K, Owen D, Laue E, Gamblin SJ, Smerdon SJ (1997b) Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP. Nature 388:693–697

Sorefan K, Booker J, Haurogne K, Goussot M, Bainbridge K, Foo E, Chatfield S, Ward S, Beveridge C, Rameau C, Leyser O (2003) MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in Arabidopsis and pea. Genes Dev 17:1469–1474

Soundappan I, Bennett T, Morffy N, Liang Y, Stanga JP, Abbas A, Leyser O, Nelson DC (2015) SMAX1-LIKE/D53 family members enable distinct MAX2-dependent responses to strigolactones and karrikins in Arabidopsis. Plant Cell 27:3143–3159

Stanga JP, Smith SM, Briggs WR, Nelson DC (2013) SUPPRESSOR OF MORE AXILLARY GROWTH2 1 controls seed germination and seedling development in Arabidopsis. Plant Physiol 163:318–330

Stanga JP, Morffy N, Nelson DC (2016) Functional redundancy in the control of seedling growth by the karrikin signaling pathway. Planta 243:1397–1406

Stirnberg P, van de Sande K, Leyser HMO (2002) MAX1 and MAX2 control shoot lateral branching in Arabidopsis. Development 129:1131–1141

Tokunaga T, Hayashi H, Akiyama K (2015) Medicaol, a strigolactone identified as a putative didehydro-orobanchol isomer from Medicago truncatula. Phytochemistry 111:91–97

Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow TY, Hsing YI, Kitano H, Yamaguchi I, Matsuoka M (2005) *GIBBERELLIN INSENSITIVE DWARF1* encodes a soluble receptor for gibberellin. Nature 437:693–698

Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, Magome H, Kamiya Y, Shirasu K, Yoneyama K, Kyozuka J, Yamaguchi S (2008) Inhibition of shoot branching by new terpenoid plant hormones. Nature 455:195–200

Waters MT, Scaffidi A, Flematti G, Smith SM (2015a) Substrate-induced degradation of the α/β-fold hydrolase KARRIKIN INSENSITIVE2 requires a functional catalytic triad but is independent of MAX2. Mol Plant 8:814–817

Waters MT, Scaffidi A, Moulin SL, Sun YK, Flematti GR, Smith SM (2015b) A Selaginella moellendorffii ortholog of KARRIKIN INSENSITIVE2 functions in Arabidopsis development but cannot mediate responses to karrikins or strigolactones. Plant Cell 27:1925–1944

Yamaguchi S, Kyozuka J (2010) Branching hormone is busy both underground and overground. Plant Cell Physiol 51:1091–1094

Yao R, Ming Z, Yan L, Li S, Wang F, Ma S, Yu C, Yang M, Chen L, Chen L, Li Y, Yan C, Miao D, Sun Z, Yan J, Sun Y, Wang L, Chu J, Fan S, He W, Deng H, Nan F, Li J, Rao Z, Lou Z, Xie D (2016) DWARF14 is a non-canonical hormone receptor for strigolactone. Nature 536:469–473

Yoneyama K, Awad AA, Xie X, Yoneyama K, Takeuchi Y (2010) Strigolactones as germination stimulants for root parasitic plants. Plant Cell Physiol 51:1095–1103

Zhao LH, Zhou XE, Wu ZS, Yi W, Xu Y, Li S, Xu TH, Liu Y, Chen RZ, Kovach A, Kang Y, Hou L, He Y, Xie C, Song W, Zhong D, Xu Y, Wang Y, Li J, Zhang C, Melcher K, Xu HE (2013) Crystal structures of two phytohormone signal transducing α/β hydrolases: karrikin-signaling KAI2 and strigolactone-signaling DWARF14. Cell Res 23:436–439

Zhao J, Wang T, Wang M, Liu Y, Yuan S, Gao Y, Yin L, Sun W, Peng L, Zhang W, Wan J, Li X (2014) DWARF3 participates in an SCF complex and associates with DWARF14 to suppress rice shoot branching. Plant Cell Physiol 55:1096–1109

Zhao LH, Zhou XE, Yi W, Wu Z, Liu Y, et al, Kang Y, Hou L, de Waal PW, Li S, Jiang Y, Scaffidi A, Flematti GR, Smith SM, Lam VQ, Griffin PR, Wang Y, Li J, Melcher K, Xu HE (2015) Destabilization of strigolactone receptor DWARF14 by binding of ligand and E3-ligase signaling effector DWARF3. Cell Res 25:1219–1236

Zhou F, Lin Q, Zhu L, Ren Y, Zhou K, Shabek N, Wu F, Mao H, Dong W, Gan L, Ma W, Gao H, Chen J, Yang C, Wang D, Tan J, Zhang X, Guo X, Wang J, Jiang L, Liu X, Chen W, Chu J, Yan C, Ueno K, Ito S, Asami T, Cheng Z, Wang J, Lei C, Zhai H, Wu C, Wang H, Zheng N, Wan J (2013) D14-SCFD3-dependent degradation of D53 regulates strigolactone signalling. Nature 504:406–410

Zwanenburg B, Pospíšil T (2013) Structure and activity of strigolactones: new plant hormones with a rich future. Mol Plant 6:38–62

Chapter 8
Abscisic Acid Signaling and Biosynthesis: Protein Structures and Molecular Probes

Jonathan D. M. Helander and Sean R. Cutler

8.1 Introduction

CO_2 uptake is essential for plant growth, however its entry through open stomata allows water loss by transpiration, a trade-off that necessitates coordination between stomatal conductance and available water. To navigate this balancing act, plants regulate the levels of the phytohormone abscisic acid (ABA; Fig. 8.1), which rises in response to water deficit to elicit stomatal closure, which in turn reduces carbon fixation and growth. ABA acts through a signaling module that emerged with the transition of plants to land and is present in all land plant genomes sequenced (Cutler et al. 2010). Many years of physiological and genetic analyses have demonstrated that ABA plays a central role in regulating transpiration (Melotto et al. 2006; McAinsh et al. 1990). In addition, it functions in the induction of seed dormancy and development (Hilhorst and Karssen 1992), shoot growth (Watts et al. 1981; Creelman et al. 1990; Saab et al. 1990), leaf senescence (Liang et al. 2014; Mayak and Halevy 1972; Yang et al. 2014; Gao et al. 2016), root system architecture (De Smet et al. 2003, 2006; Zhao et al. 2014), cold tolerance (Gusta et al. 2005), sugar sensing (Rolland et al. 2006), and numerous abiotic and biotic responses (Anderson et al. 2004; Chiwocha et al. 2005; Fedoroff 2002; Duan et al. 2013; Ton et al. 2009).

ABA has been identified in all kingdoms of life aside from archaeans (Hartung 2010). In the moss *Physcomitrella patens*, ABA acts through the core ABA response module (described below) to control desiccation tolerance, which was likely a key ancestral function of ABA that developed as plants moved onto land (Khandelwal et al. 2010). Recent analyses of a SnRK2 loss-of-function mutant in the fern

J. D. M. Helander · S. R. Cutler (✉)
Institute for Integrative Genome Biology, Center for Plant Cell Biology, and Department of Botany and Plant Sciences, University of California, Riverside, CA, USA
e-mail: cutler@ucr.edu

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_8

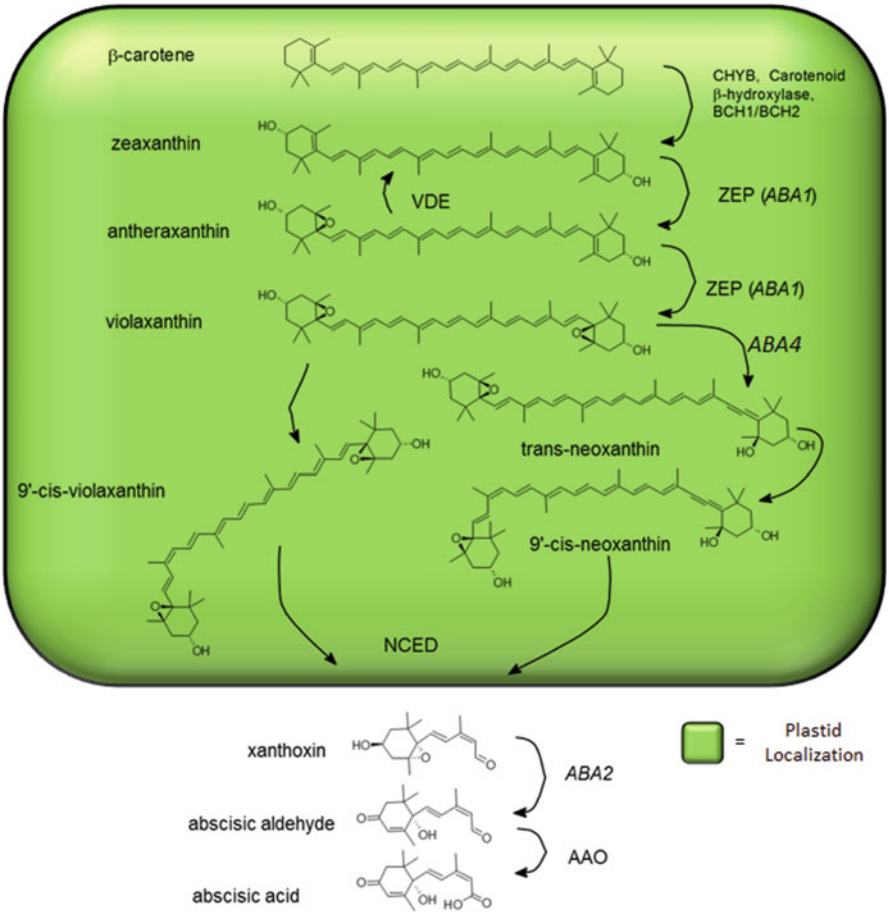


Fig. 8.1 Biosynthesis of ABA starting from β-carotene. Green box indicates reactions taking place within the plastid organelles

Ceratopteris richardii suggest that ABA's ancestral function may have been to regulate spore dormancy and reproductive tissue sex ratios, which has been interpreted to imply that this reflects an ancestral role of ABA, which was later co-opted to control the transpiration and seed dormancy in seed plants (McAdam et al. 2016). However, stomata are present in all land plant lineages except liverworts, and stomata in both Physcomitrella and the lycophyte *Selaginella uncinata* respond to high exogenous ABA concentrations, which is suggestive of an early origin of ABA-mediated control of stomatal aperture (Sussmilch et al. 2017; Lind et al. 2015), although this point is still debated (McAdam et al. 2016). Nonetheless, ABA has numerous biological activities across land plants and acts through an ancient core signaling pathway to control different cellular outputs.

The exogenous application of ABA to diverse angiosperms is sufficient to mimic many physiological responses observed during water deficit. For example, in *Arabidopsis* and maize, ABA application inhibits shoot growth and at high concentrations also inhibits primary root growth (De Smet et al. 2003; Duan et al. 2013; Fujii and Zhu 2009). At low ABA concentrations, primary root growth can be stimulated, mimicking effects of mild drought stress (Watts et al. 1981; Creelman et al. 1990; Geng et al. 2013; Sharp et al. 2004). In *Arabidopsis*, ABA shapes root architecture because the low ABA concentrations that stimulate primary root growth also inhibit lateral root growth; together, these ABA effects create deeper root systems (De Smet et al. 2003; Duan et al. 2013; Fujii and Zhu 2009; Deak and Malamy 2005). These and other potentially beneficial ABA effects have stimulated efforts to leverage the ABA pathway to improve crop yields during drought (Helander et al. 2016). Early examples toward this goal exploited anti-sense transgenes that confer ABA hypersensitivity by reducing farnesyltransferase activity (Wang et al. 2005; Cutler et al. 1996). The direct application of ABA to crop plants has been shown to improve yield in wheat and soybean (Reinoso et al. 2011; Zhang et al. 2016; Travaglia et al. 2010), however, ABA is photolabile, rapidly catabolized in vivo, and moderately complex in structure, which together make its utilization in row crops currently impractical (Todoroki et al. 2001; Wenjian et al. 2013; Abrams et al. 1997; Kushiro et al. 2004). Nonetheless, ABA's effects on yield during conditions of modest drought have motivated interest in designing inexpensive synthetic ABA mimics that can be used as agrochemicals (Helander et al. 2016), a point we return to later in this chapter.

Here we review ABA's biosynthesis, perception, and its core signaling network, focusing on the wealth of X-ray crystallographic data for the receptors, phosphatases, and kinases that form the core ABA response module. This understanding is enabling the design of new genetic and chemical tools to modulate water use and stress tolerance. We additionally discuss structural features of synthetic ligands that modulate both ABA biosynthesis and receptor function. We note that there are a number of important topics that we cannot discuss due to space constraints including ABA catabolism by the CYP707A family (Kushiro et al. 2004), ABA transport by multiple ABC-class transporters (Seo 2014), and ABA conjugation by glucosyltransferases (Xu et al. 2002).

8.2 ABA Biosynthesis

ABA is a C_{15} apocarotenoid that, in plants, is derived from β-carotene synthesized through the "indirect" chloroplastic MEP/IPP pathway, which is distinguished from the direct cytosolic pathway used by fungi from a mevalonic acid precursor. A simplified version of the major steps of the plant pathway is shown in Fig. 8.1; more detailed descriptions can be found in many excellent reviews (Schwartz and Zeevaart 2010; Endo et al. 2014; Nambara and Marion-Poll 2005). In this pathway, β-carotene, a C_{40} hydrocarbon, is oxidized to all-*trans*-zeaxanthin by the enzyme

CHYB (*β-carotene hydroxylase*); all-*trans*-zeaxanthin is next oxidized to antheroxanthin and then to all-*trans*-violaxanthin by the enzyme *zeaxanthin epoxidase* (ZEP), which is encoded in *Arabidopsis* by the locus *ABA1* and catalyzes both reactions. All-*trans*-violaxanthin can then follow two routes toward either 9′-*cis*-neoxanthin or 9-*cis*-violaxanthin in steps that involve an unidentified isomerase and *neoxanthin synthase* (whose identity is still being clarified). At this point, the 9-*cis*-xanthophyll products are cleaved at their 11,12 double bonds by a family of enzymes called *9-Cis-epoxycarotenoid dioxygenases* (NCEDs) to yield the C_{15} compound xanthoxin and different C_{25} coproducts, depending on the particular substrate cleaved. The synthesis of xanthoxin marks the first committed step to ABA from the plastidial MEP pathway and the NCEDs play a critical role in stress-induced ABA biosynthesis (discussed below). Although both the 9′-*cis*-neoxanthin and 9-*cis*-violaxanthin NCED substrates are present *in planta*, it is believed that 9′-*cis*-neoxanthin is the predominant precursor to ABA (Schwartz and Zeevaart 2010). Once formed, xanthoxin is exported to the cytosol, where it is oxidized by *xanthoxin dehydrogenase* (encoded in *Arabidopsis* by *ABA2*) to yield abscisyl alcohol and finally oxidized to 2-*cis*,4-*trans*-*S*-(+)-ABA by a molybdenum cofactor-dependent oxidase encoded in *Arabidopsis* by *abscisyl aldehyde oxidase* (*AAO3*).

The NCEDs are central regulators of ABA levels and have been the focus of extensive investigation. The NCED gene family was first identified in maize through the characterization of a family of viviparous mutants in which seeds germinate precociously on the ear instead of acquiring ABA-imposed seed dormancy. The molecular cloning of a *Viviparous 14* (*Vp14)* transposon-induced lesion (Schwartz et al. 1997) revealed that *Vp14* possesses sequence similarity to a bacterial *Lignostilbene-□-β-dioxygenase* involved in lignin degradation. Together with feeding studies and metabolite profiling, this similarity pointed toward *Vp14's* role in synthesizing xanthoxin, which was demonstrated directly in vitro using recombinant VP14 (Schwartz et al. 1997). This pioneering work established *Vp14* as the founding member of a large family of *carotenoid cleavage dioxygenases* (CCDs) that are widely distributed in prokaryotes and eukaryotes and are, for example, responsible for retinal synthesis from β-carotene in vertebrates (Harrison and Bugg 2014). In *Arabidopsis*, there are nine CCDs of which five are NCEDs that participate in ABA biosynthesis. CCD7/MAX3 is involved in strigolactone synthesis and the remaining CCDs participate in the synthesis of diverse apocarotenoids (Auldridge et al. 2006). Specific NCED mRNAs increase in response to drought and extensive evidence has defined these enzymes as central regulators of drought-stress-induced ABA (Qin and Zeevaart 1999).

The crystal structures of multiple CCDs have revealed a conserved active site architecture that involves a catalytic iron coordinated by 4 highly conserved histidine residues flanked by an outer shell of three conserved glutamic acid residues (Harrison and Bugg 2014). $^{18}O_2$ labelling studies using CCD1 demonstrated that CCDs are indeed dioxygenases, as both of CCD1's β-carotene cleavage products incorporate ^{18}O (Schmidt et al. 2006), although some earlier studies suggested that they are monooxygenases (see (Harrison and Bugg 2014) for discussion). The proposed

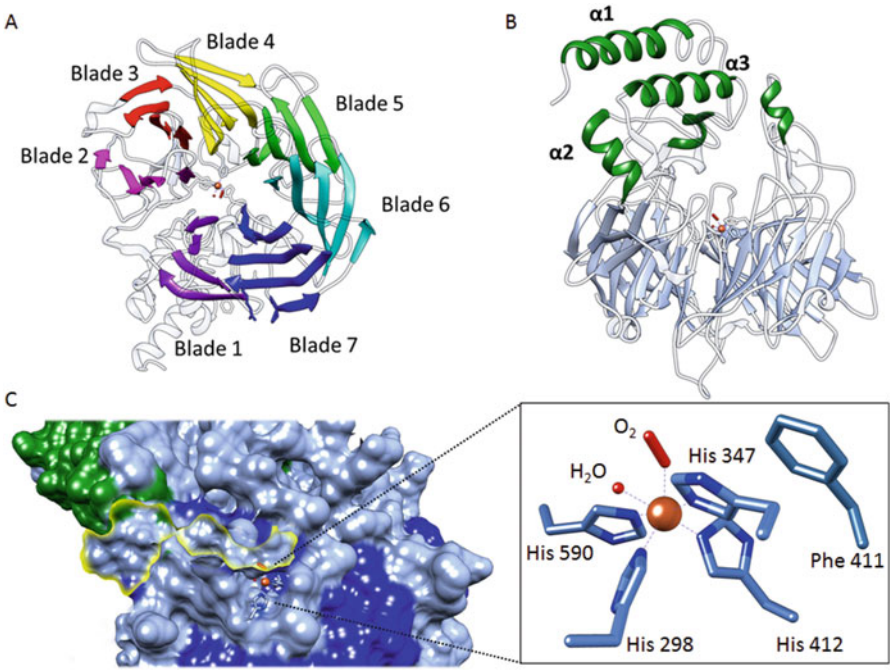


Fig. 8.2 (**a**) "Top-down" view of the seven-blade propeller structure of VP14 with each blade indicated and colored. (**b**) "Side-view" indicating the alpha-helices that form a "cap" over the propeller. (**c**) Surface view with highlighted solvent-exposed volume that likely is responsible for encapsulating the substrate, with an enlarged magnification of the enzymatic site and important residues centered around the Fe^{2+}

CCD reaction mechanism involves single electron transfer from Fe^{2+} to O_2, generating a reactive radical that attacks the carbon–carbon double bond to form a carbocation intermediate. Two paths from this intermediate to final products have been proposed, one which proceeds through a transient peroxide/dioxetane that spontaneously converts into two aldehyde products (Messing et al. 2010), or a path that proceeds through a water-assisted cleavage (Harrison and Bugg 2014; Messing et al. 2010).

There is currently one plant NCED crystal structure, for VP14 (PDB ID:3NPE) (Messing et al. 2010). VP14 possesses a 7-bladed β-propeller structure (Fig. 8.2). An NCED-specific domain is formed from four α-helical regions inserted between blades of the propeller architecture that assemble into a cap domain above the catalytic active site (Fig. 8.2a, b) (Messing et al. 2010). The β-propeller architecture creates a central solvent-exposed tunnel that displays the histidine-coordinated catalytic Fe^{2+}, which is proximal to a water molecule and an O_2 (Fig. 8.2c). The water and histidines (590, 412, and 347) occupy a plane coordinated to the catalytic iron. The O_2 and His 298 are situated above and below this plane, respectively.

Aside from the four coordinating His groups, Phe-411 is the only other residue within 5 Å of the Fe^{2+}.

Based on modeling studies, it has been proposed that 9-*cis*-xanthophylls bind to hydrophobic residues in the propeller tunnel such that the double bond between C11 and C12 is positioned immediately below the activated O_2. VP14's α-helical cap is lined by a large cluster of solvent-exposed hydrophobic residues that have been proposed to anchor VP14 in the thylakoid membrane proximal to its xanthophyll substrates, which are photosynthetic accessory pigments (Messing et al. 2010).

The *Arabidopsis* NCEDs display different substrate binding, subchloroplastic localization, and mRNA expression patterns (Tan et al. 2003). In in vitro chloroplast import assays, VP14 was associated with both stromal and thylakoid membrane fractions (Tan et al. 2003). Three *Arabidopsis* NCEDs (Melotto et al. 2006; McAinsh et al. 1990; Creelman et al. 1990) are stromal localized while NCED5 is exclusively thylakoid membrane associated and NCED9 is present in both fractions (Tan et al. 2003). In *Arabidopsis*, NCED3 is the key drought-induced family member and loss-of-function *nced3* mutants are ABA deficient and wilty, although NCED9 is also mildly induced by water deficit (Tan et al. 2003; Iuchi et al. 2001). Thus, at least some of the NCEDs are transcriptionally regulated to directly regulate ABA levels. The identity of the primary sensor(s) of water deficit that regulate NCED expression remains unknown, but the rapidity with which osmotic signals can be transduced from osmotically stressed roots to alter shoot ABA-biosynthesis suggests that changes in turgor pressure likely drive long distance responses to drought experience by roots (Christmann et al. 2007). In this context, the recently identified *OSCA* (*REDUCED HYPEROSMOLALITY-INDUCED* Ca^{2+} *INCREASE*) family of osmotic stress sensors (Yuan et al. 2014) may provide new leads for connecting osmotic stress sensing to the regulation of ABA biosynthesis.

The NCEDs have attracted interest as targets for the rational design of ABA-biosynthetic inhibitors. Prior to the identification of NCEDs, it was established using carbon monoxide studies that the enzymatic conversion of 9-*cis*-xanthophylls to xanthoxin is not heme-mediated (Schwartz and Zeevaart 2010). This led to the hypothesis that the reaction was catalyzed by lipoxygenases (LOXs), which are non-heme dioxygenases that cleave *cis*-double bonds in polyunsaturated fatty acids to produce diverse oxylipins. Although we now know that NCEDs are not homologous to LOXs, this early speculation inspired tests of known LOX inhibitors for their effects on drought-induced ABA accumulation in Xanthium leaves (Creelman et al. 1992). These tests established that the LOX inhibitor nordihydroguaiaretic acid (NDGA; Fig. 8.3) blocks drought-induced ABA biosynthesis (Creelman et al. 1992).

Mechanistically, NDGA is known to change the oxidation state of a LOX catalytic iron (Louis-Flamberg et al. 1988; Shimizu et al. 1984), which suggested that 9-*cis*-xanthophyll cleavage might require a similar catalytic site. Additionally, it was observed that alkylation of NDGA's phenols abolished LOX antagonism without disrupting 9-*cis*-xanthophyll cleavage, which in turn suggested a strategy for increasing the NCED-selectivity of this scaffold (Whitman et al. 2002). Based on these observations, NDGA was used as a scaffold for modification, additionally incorporating lessons learned from structure activity relationships of inhibitors of the

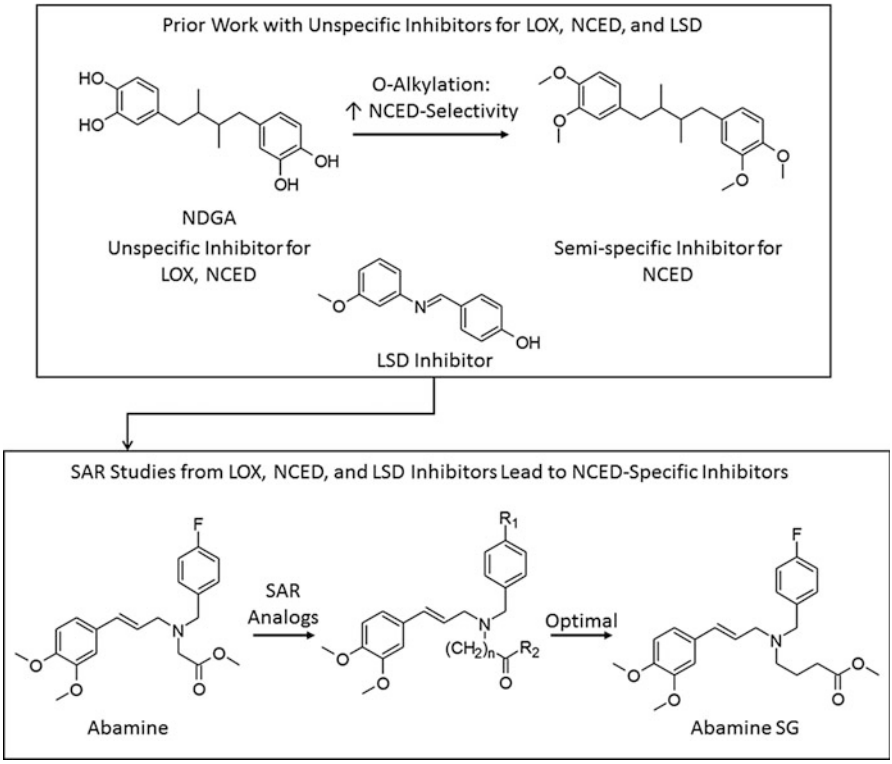


Fig. 8.3 Prior work on LOX, LSD, and NCED inhibitors (top) were used as starting points for the design of selective NCED inhibitors. Specifically, the scaffold of NDGA was observed to lose activity for LOX with *O*-alkylation, hypothesizing an increase in NCED selectivity. Further, the double bond of the imino LSD inhibitor was included in SAR studies on NCED inhibition. Eventually, Abamine was developed but further optimized to Abamine SG by varying three regions to remove off-target effects

bacterial CCD *lignostilbene-α,β-dioxygenase* (LSD) (Han et al. 2002, 2003). These efforts eventually led to the synthesis of abamine (Fig. 8.3) (Han et al. 2004), which contains a central tertiary amine, positioned such that it may coordinate with the catalytic iron or facilitate better binding by mimicking the presumptive carbocation NCED reaction intermediate (Harrison and Bugg 2014). Abamine treatments reduce mannitol-induced ABA accumulation in both spinach (*Spinacia oleracea*) and *Arabidopsis* by a maximum of ~40% but also inhibit seedling growth by a mechanism unrelated to ABA biosynthesis-inhibition, leaving ample room for improvement. To address this limitation, structure-activity relationship (SAR) studies were performed to vary the R2 ester functionality, the length of the linker between the tertiary amine and ester (*n*), and finally the substitutions, R1 on the benzyl appendage (Fig. 8.3), resulting in the discovery of abamineSG, which reduces mannitol-induced ABA accumulation by up to 77% in *Arabidopsis* without inhibiting seedling growth (Kitahata et al. 2006). Although abamineSG is a useful research tool, one

limitation is that its effects on ABA biosynthesis appear selective for eudicots (M. Otani personal communication and our own observations); thus, next-generation NCED inhibitors with broader spectrum activity and improved potency would be very beneficial as probe molecules. From the perspective of agrochemical development, improved inhibitors could find use in improving germination rates to improve synchronization of stand establishment and possibly as agents to break bud dormancy in tree crops. Such efforts would benefit from higher resolution NCED structures (3.2 Å for VP14) and co-crystal structures with existing inhibitors and substrates.

8.3 The Core ABA Response Pathway

ABA responses are controlled by a 3-component negative regulatory signaling module that is composed of soluble *pyrabactin resistance 1/PYR1-like/regulatory component of ABA receptor* (PYR/PYL/RCAR; PYL for short) ABA receptors, clade A *type 2C protein phosphatases* (PP2Cs), and group III *subfamily 2 Snf1-related kinases* (SnRK2s; Fig. 8.4). The SnRK2s phosphorylate and regulate the activity of several downstream effectors, such as bZIP transcription factors, NADPH oxidases, and ion channels that are required for guard cell closure (Kulik et al. 2011). In vitro, the SnRK2s autoactivate by cis- and trans-autophosphorylation on a critical activation loop residue, but their activity is suppressed by the PP2Cs, which dephosphorylate the activation loops and prevent phosphorylation by docking in the SnRK2's catalytic pockets (Soon et al. 2012; Ng et al. 2011). When ABA binds to

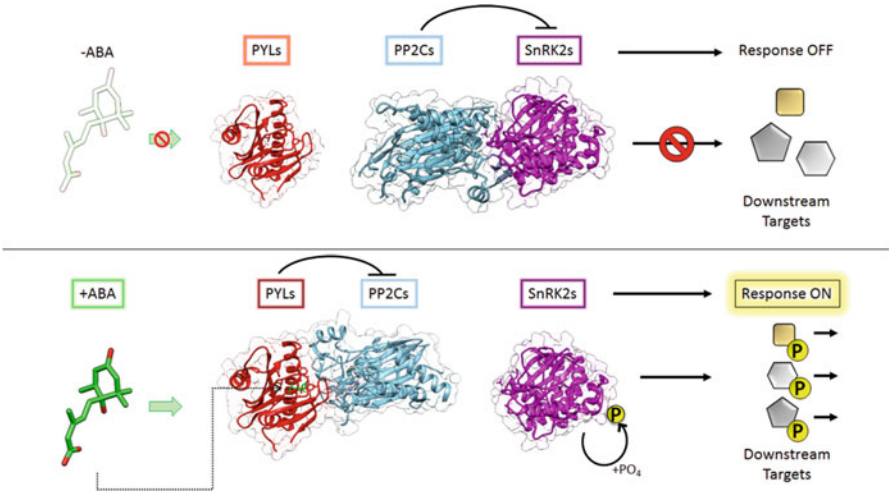


Fig. 8.4 Panel A – illustration of the core pathway ABA-signaling pathway. Top represents unstressed, low ABA (-ABA) conditions, and bottom represents when ABA is present in high concentrations (+ABA). Yellow-encircled P indicates phosphorylation

soluble PYL ABA receptors, a conformational change enables the ligand-bound receptors to form a complex with the PP2Cs that inhibits phosphatase activity; this in turn enables accumulation of activated SnRK2 kinases and activation of downstream responses including transcriptional responses that regulate a large protein–protein interaction network (Lumba et al. 2014). Thus, ABA controls cellular outputs by regulating SnRK2 kinase activity via ABA-induced formation of al PYL/PP2C complex; this is one of the many plant hormone-stabilized protein–protein interactions that has emerged as a common theme in plant signaling (Provart et al. 2015).

The core pathway is highly conserved in land plants and appears to have emerged with the evolutionary transition of plants from water to land (Hauser et al. 2011). For example, the genomes of *Physcomitrella patens, Selaginella moellendorffii*, and numerous seed plants possess multiple PYLs, PP2Cs, and SnRK2 (Hauser et al. 2011; Umezawa et al. 2010a). Moreover, downstream targets of the SnRK2s including bZIP transcription factors homologous to the *Arabidopsis* ABF/AREB/ABI5 have been described in some moss genomes, but not in genomes of aquatic charophytes (Hauser et al. 2011; Umezawa et al. 2010a). Together, these findings indicate that the core pathway co-evolved with colonization of land by plants. Extensive genetic data supports the core model. For example, multi-locus PYL receptor mutants display differing degrees of ABA insensitivity (Gonzalez-Guzman et al. 2012), multi-locus PP2C mutants are ABA hypersensitive or constitutively activated (Rubio et al. 2009), and multi-locus SnRK2 mutants are strongly ABA insensitive (Fujii and Zhu 2009; Fujita et al. 2009) (discussed below).

8.3.1 PYL Receptors

The PYL ABA receptors were identified by parallel strategies in two separate labs in 2009. To address the challenges posed by genetic redundancy typical of plant genomes, Park et al. used forward chemical genetics to identify pyrabactin, a synthetic selective ABA agonist (Fig. 8.5) that inhibits seed germination (Park et al. 2009). Pyrabactin inhibits seeds but is only weakly active in vegetative tissues; this selectivity suggested that it might be a useful tool for identifying receptors, which had not emerged from extensive genetic screens (Park et al. 2009). A forward genetic screen for pyrabactin-resistant mutants was conducted, which defined *Pyr1* as necessary for pyrabactin's seed germination inhibition activity (Park et al. 2009). Map-based cloning defined PYR1 as a small (22 KDa), soluble, predicted ligand-binding protein in the START domain superfamily (Iyer et al. 2001). HSQC ligand-binding analyses using ^{15}N-labelled PYR1 revealed that PYR1 could bind to both ABA and pyrabactin and multiple lines of evidence showed that ABA agonists stimulate the formation of an inhibitory complex between PYR1 and PP2Cs. Loss-of-function *pyr1* mutants do not display strong ABA insensitivity due to the action of other PYLs; in contrast, *pyr1* mutants are insensitive to pyrabactin because the agonist selectively activates PYR1, which is highly expressed in seeds (Fig. 8.5). In parallel, Ma et al. searched for interactors of ABI2 using a yeast-two hybrid screen

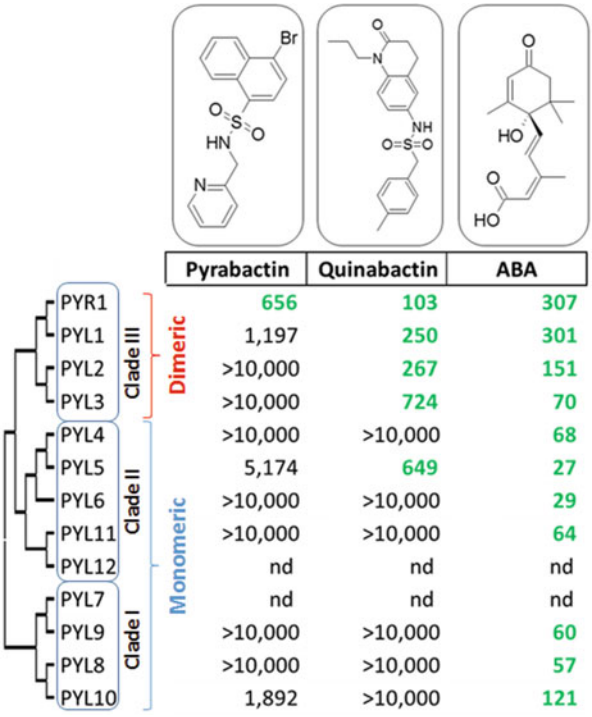


Fig. 8.5 Structures of pyrabactin, quinabactin, and natural *S*-(+)-ABA, and their IC_{50}'s mapped to the PYL receptor family, which is divided into the three clades. Values are in nM

(Ma et al. 2009a). Their screen identified RCAR1 (PYL9) (Ma et al. 2009a). Biochemical characterization of RCAR1 demonstrated that it binds to ABA and inhibits PP2C activity. Importantly, Ma et al. demonstrated by ITC that inclusion of ABI2 in binding titrations with PYL9 increased apparent ABA affinity, the PP2Cs can be therefore considered co-receptors (Ma et al. 2009a), although subsequent structural studies showed that PP2Cs make only a single indirect water-mediated contact to ABA's ketone and that ABA is enclosed within PYL receptors through direct and indirect interactions with 21 ligand binding pocket residues (Melcher et al. 2009; Yin et al. 2009; Miyazono et al. 2009). Thus, these parallel efforts independently identified different members of the *Arabidopsis* 14-member PYR/PYL/RCAR gene family. Moreover, the identification of pyrabactin defined ABA receptors as druggable targets that could be leveraged for the development of new agrochemicals (Helander et al. 2016).

The PYL receptor gene family is unusually large for a single hormone and forms three phylogenetically distinct subfamilies that can be found in all angiosperm plant genomes sequenced (Hauser et al. 2011; Umezawa et al. 2010b; Raghavendra et al. 2010; Gonzalez-Guzman et al. 2014). mRNA expression profiling experiments have shown that the expression patterns of the receptors show substantial overlap and are frequently co-expressed in guard cells and other cell types. While the specific biological roles of each subtype are still being elucidated, it is clear from work using synthetic ABA agonists (described below) that activating the PYR1-subfamily

(subfamily III) is sufficient to cause guard cell closure, reduce transpiration, and induce a genome-wide ABA-like transcriptional response in diverse species (Park et al. 2015). These data demonstrate the specific importance of subfamily III receptors as targets for developing anti-transpiration compounds. These lower-affinity dimeric receptors appear tuned to respond to the high concentrations of ABA induced by abiotic stress.

Biochemical characterization of *Arabidopsis* PYLs has shown that the receptor subfamilies possess distinct properties with respect to the oligomeric state, ABA affinity, and PP2C interactions. Subfamily I and II receptors are monomeric and, in general, show a higher ABA-affinity, while subfamily III receptors are lower affinity and mostly dimeric (with the exception of PYL3 which is in monomer-dimer exchange) (Nishimura et al. 2009; Dupeux et al. 2011a; Hao et al. 2011). As noted above, the ABA-affinities measured in vitro are influenced by whether the measurements are conducted in the presence or absence of a PP2C. For example, ITC measurements have shown that PYL9 (RCAR1) has a K_d for ABA of ~660 nM in isolation, which drops to ~64 nM in the presence of ABI2 (Ma et al. 2009b). Similarly, PYL5 has a K_d of ~1.1 μM in isolation and ~38 nM in the presence of HAB1 (Santiago et al. 2009a). The structures of ternary ABA-PYL-PP2C complexes, described below, provide a simple rationalization for this; PP2Cs bind to the closed PYL conformer enclosing the ligand-binding pocket, which should increase affinity by lowering K_{off}.

The dimeric receptors possess lower intrinsic ABA affinity than monomeric receptors due to differences between the ligand-binding pockets of monomeric and dimeric type receptors; for example, X-ray crystallographic studies have shown that in comparison to PYR1, the high affinity monomeric receptor PYL9 forms an additional hydrogen bond to ABA's COOH due to a difference in residues lining its ligand-binding pocket (Nakagawa et al. 2014). In addition, in order for dimeric receptors to interact with PP2Cs, they must overcome a thermodynamic barrier imposed by homodimer dissociation, which is required for PP2C binding because the two binding interfaces are overlapping (Dupeux et al. 2011a). Upon ABA binding, dimers are destabilized, as indicated by a shift in the proportion of monomeric receptor observed in gel filtration experiments (Dupeux et al. 2011a) as well as higher self-association K_ds measured in analytical ultracentrifugation experiments (Zhang et al. 2012). For example, the dimeric low-affinity receptor PYL2 shifts from a homodimerization K_d of ~7 μM in its apo form to ~56 μM in the presence of saturating ABA concentrations (Zhang et al. 2012). Thus, dimeric receptors are auto-inhibited from interacting with PP2Cs in the absence of ABA. The apo-forms of monomeric receptors, on the other hand, can interact with PP2Cs (Park et al. 2009; Ma et al. 2009a) and often display higher basal activity in PP2C-inhibition assays (Melcher et al. 2009; Hao et al. 2011), although these in vitro effects may be experimental artifacts stemming from the inclusion of BSA in assays (Li et al. 2015).

Several X-ray crystallographic experiments have revealed the structures of multiple PYLs (Table 8.1). The PYLs possess a conserved helix-grip fold (Iyer et al. 2001) in which a C-terminal alpha-helix is surrounded by a seven-stranded β-sheet (Fig. 8.6) (Nishimura et al. 2009). The PYLs possess two highly conserved surface-

Table 8.1 Available crystal structures of ABA receptors and PP2Cs, *Arabidopsis thaliana*

PDB ID	Receptor	Species	Ligand	Chains	Resolution (Å)	Publication
3K3K	PYR1	*A.t.*	ABA	Homodimer	1.7	Nishimura et al. (2009)
3WG8	PYR1	*A.t.*	AS6	Monomer	2.3	Takeuchi et al. (2014)
3KDJ	PYL1	*A.t.*	ABA	ABI1	1.88	Yin et al. (2009)
3NMN	PYL1	*A.t.*	Pyrabactin	ABI1	2.15	Melcher et al. (2010)
3KL1	PYL2	*A.t.*	ABA	Homodimer	1.55	Zhang et al. (2012)
3NS2	PYL2	*A.t.*	Pyrabactin	Homodimer	1.63	Yuan et al. (2010)
4LA7	PYL2	*A.t.*	Quinabactin	HAB1	1.98	Okamoto et al. (2013)
4DSC	PYL3	*A.t.*	ABA	Homodimer	1.95	Zhang et al. (2012)
4JDL	PYL5	*A.t.*	Apo	Monomer	2.65	Zhang et al. (2013)
3W9R	PYL9	*A.t.*	ABA	Monomer	1.9	Nakagawa et al. (2014)
3RT2	PYL10	*A.t.*	Apo	Monomer	1.5	Hao et al. (2011)
3R6P	PYL10	*A.t.*	ABA	HAB1	2.7	
4N0G	PYL13	*A.t.*	Apo	PP2CA	2.38	Li et al. (2013)
4OIC	Soluble ABA receptor	*Oryza sativa*	ABA	Putative PP2C	2.0	He et al. (2014)

loops that are critical for ligand- and PP2C-binding, termed the "gate" and "latch," which connect β3-β4 and β5-β6, respectively. The flexible gate loop is the most mobile structural element and can adopt open and closed forms; its closed state is stabilized by interactions with ABA and the latch loop. When a ternary complex forms with a PP2Cs, a conserved Trp in a clade A–specific recognition loop inserts between the gate and latch loops and "locks" the closed receptor conformer (see next section) (Melcher et al. 2009; Yin et al. 2009; Miyazono et al. 2009) (Fig. 8.7).

Ligands bind PYLs in a central cavity created by α2, α3, and the seven-stranded β-sheet domain (Fig. 8.6). In their apo forms, the binding pocket is solvent exposed and provides direct ligand access. The gate loop contains a 5-amino acid sequence SGLPA- that is nearly invariant amongst receptors; the latch loop contains a 3-amino acid -HRL- sequence. Closure of the gate loop is associated with conformational change that alters the orientation of the serine in the *S*GLPA gate loop so that it becomes more solvent exposed (Fig. 8.6) and positioned to interact with the PP2C by inserting into its catalytic active site, where it acts as a product mimic. The receptors are thus ABA-regulated competitive PP2C inhibitors, which has been demonstrated

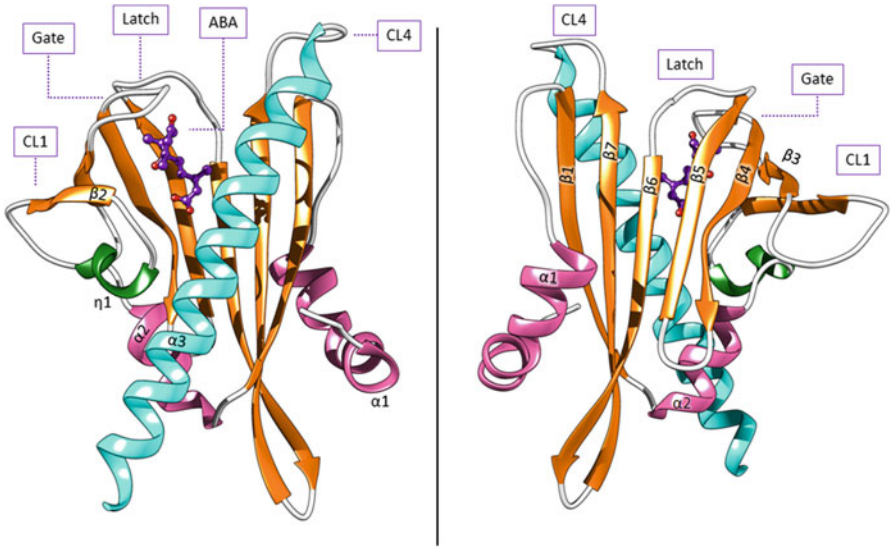


Fig. 8.6 ABA-bound crystal structure of PYR1 (PDB ID: 3K3K). The helix-grip fold structure produced by the seven-stranded β-sheet (orange) wrapped concavely over α2 (pink) and α3 (cyan). Helix α1 arches over the opposite face of the β-sheet. The flexible gate-loop and latch are indicated and ABA is depicted in purple as a ball-and-stick model

in vitro using competition experiments (Melcher et al. 2009). The closed gate conformer is stabilized by a water-mediated interaction between the -SGL*P*A- proline and the backbone amide of the arginine from the -H*R*L- latch (Melcher et al. 2009); the arginine sidechain forms a water-mediated contact to ABA's ketone (Santiago et al. 2009b) as does the -SGLP*A*- alanine backbone amide; thus polar interactions between the gate and latch loops and ABA are critical for stabilizing the activated PYL conformer. Closure of the gate around the ligand is the primary allosteric mechanism for receptor activation (Weiner et al. 2010). The importance of the gate and latch regions for receptor function is supported by mutational studies in PYR1; the gate mutation P88S (targeting -SGL*P*A-) or the latch mutation H115A (targeting *H*RL) abolish ABA-mediated PP2C inhibition without altering ABA binding (Melcher et al. 2009). Within the ligand-binding pocket, ABA is stabilized by a network of water-mediated hydrogen bonds to its carboxylate and C1'-OH, as well as a salt bridge between a conserved lysine (K59 in PYR1) and ABA's COOH. In PYR1, the highly conserved E141 makes two water-mediated contacts to ABA's COOH and the mutation E141K, which was isolated as a *pyr1* allele, disrupts PYR1 function in vivo (Park et al. 2009).

The wealth of structural, biochemical, and genetic data available for PYR1 have made it and its close relatives excellent scaffolds for protein engineering. Although beyond the scope of this review, we note that PYR1's ligand-binding specificity has been reprogrammed through mutagenesis of its ligand-binding pocket to create orthogonal receptor-ligand pairs that can be expressed in transgenic plants to enable

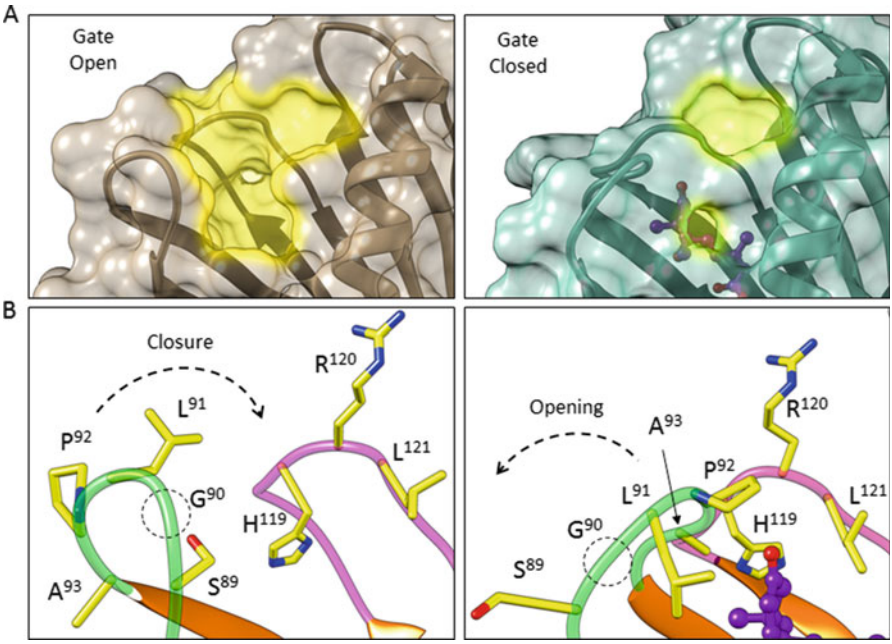


Fig. 8.7 Binding conformation changes of the gate and latch loop. (**a**) The binding pocket of PYL2 highlighted in yellow with the open state (left, PDB: 3KDH) and the closed, ABA-bound conformation (right, PDB: 3KDI). (**b**) The conserved -SGLPA- gate loop (green) and -HRL- latch loop (pink) conformations in open (left) and closed (right). ABA is depicted in purple ball-and-stick, and Gly90 is circled, as it does not have a side chain

agrochemical control of drought tolerance (Park et al. 2015). In addition, the PYL-PP2C module has been repurposed to develop FRET-based ABA sensors that have opened new avenues for investigating ABA dynamics in vivo (Waadt et al. 2014; Jones et al. 2014). Furthermore, saturation mutagenesis of PYR1's PP2C-binding interface defined a set of mutations clustering around the gate and C-terminal helix that stabilize PYL-PP2C interactions. When combined, these mutations create constitutively active receptor variants that bind and inhibit PP2C activity in the absence of ABA at 1:1 stoichiometry (Mosquna et al. 2011).

8.3.2 PP2Cs

The PP2Cs are type 2 protein Ser/Thr phosphatases (Luan 2003) and represent the largest protein phosphatase family in *Arabidopsis*, comprising 76 members in 10 subfamilies (clades A through J) (Kerk et al. 2002; Schweighofer et al. 2004). PP2Cs require bivalent cations (Mg^{2+} or Mn^{2+}) and can be differentiated from other protein phosphatases based on inhibitor insensitivity to PP2A and PP1 inhibitors

such as okadaic acid (Cohen 1989). One important role of the PP2Cs across eukaryotes is to reverse stress-induced mitogen-activated protein kinase (MAPK) and receptor-like kinase (RLK) signaling (Schweighofer et al. 2004; Ofek et al. 2003; Menges et al. 2002).

Plant clade A PP2Cs contain negative regulators of ABA signaling that in *Arabidopsis* include: *ABA insensitive 1 and 2* (ABI1 & 2), *ABA hypersensitive germination 1 and 3* (Yoshida et al. 2006a; Bhaskara et al. 2012), HAB1 and HAB2 (*Homolog of ABI1 1 and 2*) (Rodriguez et al. 1998a; Saez et al. 2004), HAI1, HAI2, and HAI3 (*highly ABA-induced 1–3*), and AHG1 (*ABA hypersensitive germination 1*) (Nishimura et al. 2007). The family's role in ABA signaling was first defined by the identification of the dominant *abi1–1* and *abi2–1* alleles, which confer reduced sensitivity to ABA-mediated seed germination inhibition (Koornneef et al. 1984). Map-based cloning established that both genes encode closely related PP2Cs (Leung et al. 1994; Leung et al. 1997; Meyer et al. 1994; Rodriguez et al. 1998b). Both dominant mutations are caused by a Gly → Asp mutation in a homologous residue proximal to a catalytic metal ion (G180 in ABI1, G162 in ABI2). In vitro, the dominant alleles greatly reduce (~5-fold), but do not abolish, PP2C activity (Robert et al. 2006; Leube et al. 1998), which led some (Rodriguez et al. 1998b) to conclude the mutant alleles encode dominant negative proteins that interfere with a normally positive role in signaling. Numerous subsequent studies characterizing loss-of-function alleles established that the PP2Cs negatively regulate ABA responses (Nishimura et al. 2007; Gosti et al. 1999) and analyses of multi-locus PP2C mutants demonstrated that they display ABA-hypersensitivity and constitutive ABA responses (Rubio et al. 2009). The mutant proteins encoded by *abi1–1* and *abi2–1* are dominant hypermorphs that mutate a key PP2C residue that contacts the PYL gate loop and are thus resistant to PYL-mediated inhibition (Park et al. 2009) and presumably retain sufficient enzymatic activity to dephosphorylate and inactivate SnRK2s kinases.

The catalytic region of the clade A PP2C HAB1 can be occupied by three Mn^{2+} cations that are coordinated in place primarily by four highly conserved Asp residues (Fig. 8.8b, c). The catalytic region is located within a channel toward the center of the protein, flanked by a clade A–specific recognition loop (Weiner et al. 2010) that harbors a highly conserved Trp residue (Trp385 in HAB1) that is critical for PYL interactions. The recognition loop projects outward and is centrally located. Trp indole-NH makes a hydrogen bond to the conserved gate-latch water, which is also bound to ABA's ketone and gate and latch residues; thus, HAB1's Trp385 forms a water-mediated hydrogen bond network between ABA's ketone and the PYLs (Fig. 8.8d, e). This Trp–H_2O–ABA interaction is conserved with all the PP2Cs except AHG1, which lacks a Trp in this position and is immune to receptor-mediated inhibition (Yin et al. 2009; Miyazono et al. 2009). Mutating this conserved Trp abolishes PYL-mediated PP2C inhibition without reducing intrinsic phosphatase activity (Dupeux et al. 2011b); when over-expressed in transgenic *Arabidopsis*, $HAB1^{W385A}$ induces ABA insensitivity and is thus an engineered hypermorphic allele (Dupeux et al. 2011b). Given the large number of PP2Cs present in plants, the clade A recognition loop plays an important role in providing specificity to PYL–

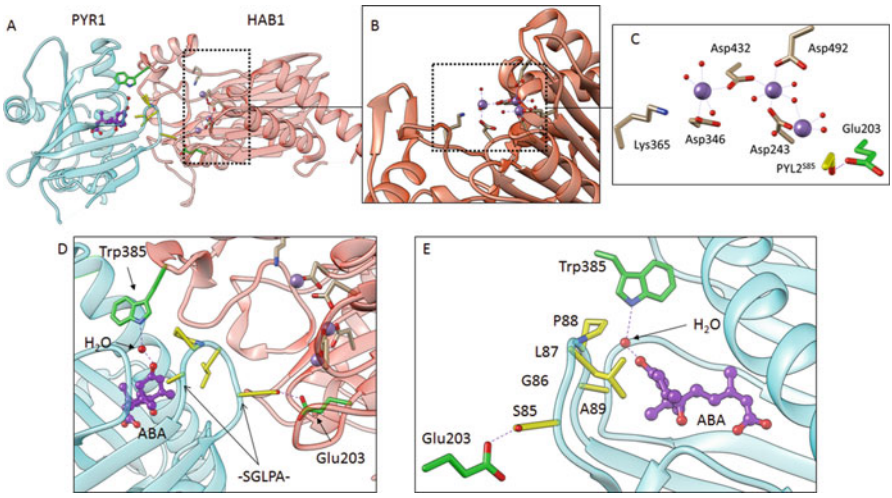


Fig. 8.8 Depiction of PYR1-HAB1 interaction. (**a**) Overall binding pose of PYR1 (teal), and HAB1 (light red) with ABA in purple and important residues shown. (**b**) Closeup of the catalytic site of HAB1. (**c**) The highly conserved residues of the clade A PP2C catalytic site as depicted with HAB1. Purple spheres are coordinated Mg^{2+} cations. Residues facilitating PYL-PP2C interaction, $HAB1^{Glu203}$ (green) and $PYL2^{S85}$ (yellow) depicted. (**d**) The gate (depicted by yellow residues in -SGLPA- sequence) closure–HAB1 interaction area. The highly conserved PP2C tryptophan residue can be seen (green) in proximity to the gate-latch water that facilitates a Trp–H_2O–ABA interaction. Glu203 is shown in green in proximity to PYR1 Ser85. (**e**) Additional view of the conserved Glu203 and Trp385 from HAB1 in proximity to the gate-latch residues and the ABA-binding pocket. (**f**) Close-up focused on the conserved water with four hydrogen bonds, to the backbones of $PYR1^{P88}$, $PYR1^{R116}$, the sidechain of $HAB1^{W385}$, and to the keto-oxygen of ABA

PP2C interactions. As discussed below, the recognition loop also provides specificity to SnRK2–PP2C interactions, which mirror many of the PYL–PP2C interactions (Soon et al. 2012). Once bound, the gate of the PYL proteins is stabilized in a closed position, locking ABA within the pocket. Ser85, from the conserved gate sequence -*S*GLPA-, directly hydrogen bonds with HAB1's Glu203 (Fig. 8.8d, e) (Yin et al. 2009), further stabilizing this interaction. These interactions inhibit the PP2C phosphatase function by preventing the access of substrates.

8.3.3 SnRK2s

In *Arabidopsis*, there are 38 Snf1-related kinases that fall into three subgroups (SnRK1–3) (Hrabak et al. 2003). All SnRKs belong to the large AMP-dependent protein kinase (AMPK) superfamily whose founding member participates in signaling cellular energy status (Wang et al. 2013; Hardie 2007). The 3 SnRK1s are most closely related to AMPKs and function in signaling energy status (Kulik et al. 2011; Robaglia et al. 2012). There are 10 SnRK2s that can be subdivided into three

subgroups; all except SnRK2.9 appear to be activated by osmotic stress, which involves both ABA-dependent and ABA-independent mechanisms (Kulik et al. 2011). Subfamily III SnRK2 kinase activities are strongly controlled by ABA and are the direct substrates of clade A PP2Cs.

Plant SnRK2s were first identified with the isolation of the wheat gene PKABA, whose mRNA is ABA- and drought-responsive (Anderberg and Walker-Simmons 1992; Hedbacker and Carlson 2008). Although not classified as a SnRK2 at the time, subsequent analyses led to the identification of *ABA-activated protein kinase* (AAPK) from fava bean guard cells (Li and Assmann 1996; Li et al. 2000), which was similar to PKABA and later recognized as a member of the SnRK2 family. The first compelling genetic evidence for the importance of SnRK2s in vivo came from the characterization of the *Arabidopsis* mutant *open stomata 1* (OST1/SnRK2.6), which was identified by virtue of its reduced leaf temperature caused by increased transpiration (Mustilli et al. 2002). Map-based cloning of *Ost1* demonstrated that it was closely related to AAPK, PKABA, and other stress-induced kinases (Mustilli et al. 2002). Promoter-GUS studies showed that *Ost1* is expressed in guard cells and both root and shoot vasculature, however microarray data show that *Ost1* is also expressed at lower levels in mesophyll and other cells (Yang et al. 2008; Penfield et al. 2006). The *snrk2.2/snrk2.3/ost1* triple mutant, which genetically removes all group III SnRK2s, displays an extremely ABA-insensitive phenotype and can germinate on 50 μM of ABA, a concentration 50 times higher than that which can inhibit WT *Arabidopsis thaliana* (Fujii and Zhu 2009; Fujita et al. 2009; Nakashima et al. 2009); these and other observations have demonstrated that SnRK2s are essential for eliciting ABA responses (Fujita et al. 2009; Nakashima et al. 2009; Yoshida et al. 2006b), however, phosphoproteomic studies have shown that a small number of ABA-dependent phosphorylation events occur in *snrk2.2/snrk2.3/ost1* plants, indicating that other minor players (possibly subfamily II SnRK2s) participate in ABA-mediated phosphorylation (Wang et al. 2013). SnRK2.2 and SnRK2.3 are redundant, and a double mutant is required to observe ABA-insensitivity and enhanced water loss.

Prior to defining the "core pathway", yeast two hybrid studies had suggested that the PP2C ABI1 can bind to OST1, which provided an important clue to the subsequent definition of the core response module (Yoshida et al. 2006c). SnRK2 kinase activity is reduced in a *pyr1/pyl1/pyl2/pyl4* quadruple mutant, which further supported the core model (Fig. 8.4) (Park et al. 2009). This pathway can be reconstituted in vitro using members of each core component and a kinase substrate (PYR1, ABI1, OST1, and an ABF) (Fujii et al. 2009). Downstream substrates of ABA-dependent SnRK2 phosphorylation include the bZIP *A*BA-*r*esponsive *e*lement-*b*inding (AREB) transcription factors, the *sl*ow *a*nion *c*hannel-associated 1 (SLAC1) ion channel proteins, the K^+ channel in *Arabidopsis thaliana* (KAT1), the *r*espiratory *b*urst *o*xidase *h*omolog *F* (RBOHF), and more recently *Sn*RK2-*s*ubstrate 1 (SNS1) (Wang et al. 2013) as well as other proteins.

Subfamily III SnRK2s display basal kinase activity in vitro due to autoactivation by both cis- and trans-autophosphorylation of their activation loops. Mutational analyses indicated that phosphorylation of the SnRK2 activation loop (S175 in

SnRK2.6) is required for full activity, as the S175A mutant reduces ABF phosphorylation by SnRK2.6 ~98% in vitro (Ng et al. 2011). OST1/SnRK2.6 possesses higher (5–10x) autoactivation in vitro than SnRK2.2/2.3. Disrupting OST1 *cis*-phosphorylation by introducing a catalytic mutation does not prevent activation loop phosphorylation by wild-type protein, which indicates that transphosphorylation occurs (Ng et al. 2011).

High-resolution crystal structures of SnRK2.3 and 2.6 were determined using molecular replacement of X-ray data collected from crystals of proteins containing surface mutants predicted to lower entropy (Soon et al. 2012). As anticipated, these revealed that SnRKs possess a canonical kinase fold and are most similar to both AMPK and Snf1. Like these kinases, the SnRK2s contain a bilobal structure with a small (N-lobe) and large lobe (C-lobe) that sandwiches an ATP molecule in a catalytic cleft that forms between the two domains connected by a flexible hinge domain that enables the kinases to cycle between open (inactive) and closed (active) conformations (Fig. 8.9a) (Pearce et al. 2010; Littler et al. 2010). These two conformers were resolved because SnRK2.6 adopted the open conformer, whereas

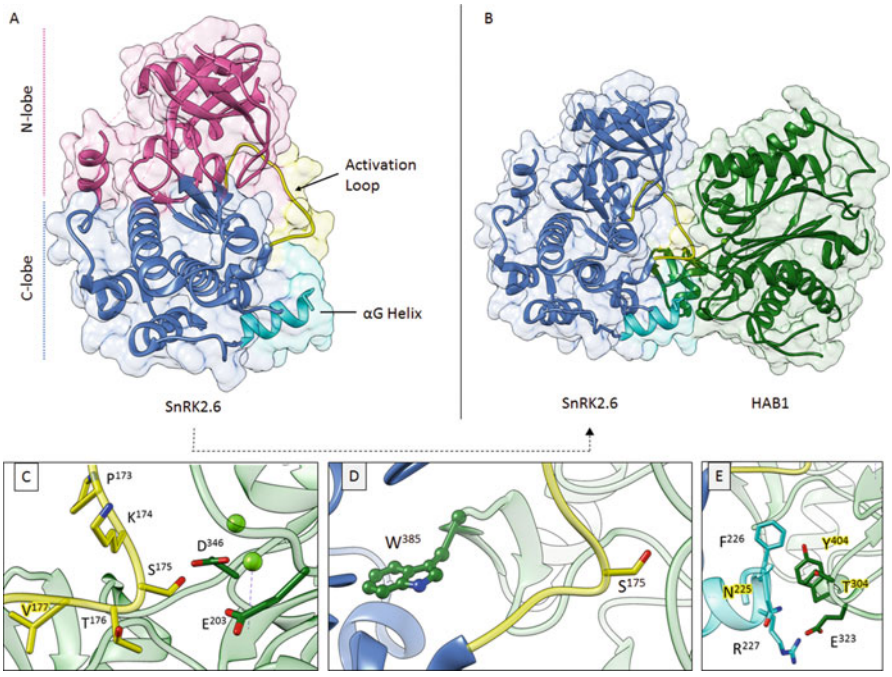


Fig. 8.9 SnRK2.6–HAB1 interactions and important residues. (**a**) SnRK2.6 bilobal structure with the N-lobe (pink) and C-lobe (blue), and the important interactor regions of the activation loop (yellow) and αG helix (cyan). (**b**) Surface rendition of SnRK2.6 (left) and HAB1 (dark green, right). (**c**) The activation loop (yellow) residues, including SnRK2.6^{S175} that inserts into the catalytic region, proximal to HAB1^{D346} and HAB1^{E203}. (**d**) Depiction of the conserved tryptophan and serine that mimics PYL–PP2C structural interaction mechanisms. (**e**) Residues of the αG helix (cyan) and proximal residues of HAB1. Yellow text-highlighting is to increase contrast

SnRK2.3 crystallized in the closed conformer. The active (closed) conformer positions the mobile αC helix and other active site ATP- and Mg^{++}-binding residues in a catalytically favorable orientation and orients the activation loop serine adjacent to the catalytic sites, poised for phosphorylation. By analogy to other kinases, the transition from open to closed conformers is likely stabilized by activation loop phosphorylation. The activation loops in the SnRK2.3 and 2.6 structures were not phosphorylated, which indicates that the active conformation can, under some circumstances, be adopted without activation loop phosphorylation; transitions between the two conformers has been proposed to explain the high basal autoactivation of SnRK2s in vitro. All plant SnRK2s contain a C-terminal SnRK2-specific domain called the SnRK2-box (also called domain I in (Kulik et al. 2011)), which forms a single α-helix (αG) that packs against the N-terminal lobe and αC helix. It has been proposed that the αC helix-SnRk2 box interaction stabilizes the closed conformer and underlies the relatively high basal activity of plant SnRK2s, in particular SnRK2.6, which has higher basal activity than SnRK2.2 and 2.3.

Obtaining diffraction quality crystals for an SnRK2.6-HAB1 complex was challenging and required the use of a synthetic fusion protein that linked the two proteins via an artificial linker, as well as the introduction of surface mutations into SnRK2.6 (Soon et al. 2012). The crystal structure of the complex obtained demonstrated that an extensive protein–protein interface forms through mutual packing of the kinase and phosphatase active sites that remarkably mirrors PYL–PP2C interactions. These interactions involve multiple domains on SnRK2.6: its activation loop, the catalytic cleft, and the αG helix that forms the SnRK2-box (Fig. 8.9a, e) (Soon et al. 2012). The activation loop serine ($SnRK2.6^{S175}$) packs tightly into the HAB1 catalytic pocket mimicking the positioning of the PYL -SGLPA- loop. The PP2C lock tryptophan ($HAB1^{W385}$) inserts into the SnRK2 catalytic site and physically blocks substrate access (Fig. 8.9d). The SnRK2-box αG helix interacts with PP2C residues proximal to W385, further stabilizing the complex formed (Fig. 8.9e). Thus, two key themes of molecular recognition that provide specificity to PYL–PP2C interactions are exploited in PP2C–SnRK2 interactions, suggesting that regulation of the entire signaling module likely co-evolved.

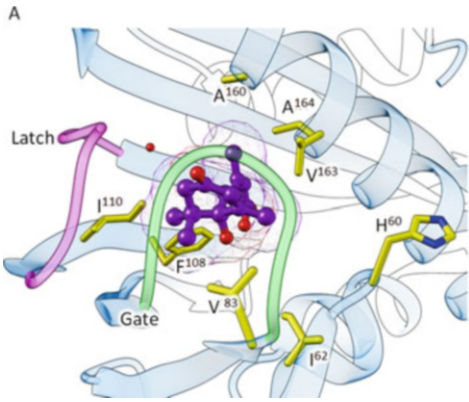
Subfamily III SnRK2s contain a conserved acidic "ABA-box" at their C-terminus that enhances PP2C–SnRK2 interactions, however this region was not resolved in the crystal structure (Belin et al. 2006). PP2C–SnRK2.6 interactions have ~2–8 μM K_ds while the ABA-box in isolation binds to HAB1 with ~10-fold lower affinity, but can disrupt SnRK2–PP2C interactions when supplied as a competitor. The ABA-box interacts with a poly-basic surface region on HAB1 and stabilizes the formation of an inhibitory PP2C-SnRK2 complex. The net effects of the SnRK2–PP2C interactions is that PP2Cs inactivate SnRK2s both catalytically, by dephosphorylating pS175, and stoichiometrically, by forming a stable PP2C:SnRK2 complex that prevents activation loop phosphorylation (Soon et al. 2012). Mutational studies have demonstrated that both of these mechanisms work in parallel to elicit full PP2C-mediated SnRK2 inhibition in vitro; mutations that disrupt either PP2C catalytic activity or SnRK2-PP2C binding via the ABA-box result in only partial disruption of function;

however, simultaneously eliminating both PP2C catalytic activity and PP2C-SnRK2 binding yields a HAB1 mutant that cannot inactivate SnRk2.6 kinase activity in vitro.

8.3.4 *Structural Differences Between* Arabidopsis *ABA Receptors*

Sequence differences between ABA receptor ligand-binding pockets influence the selectivity of ABA agonists. There are eight distinct variable positions within the proximal ($\leq$ 5 Å from ABA) binding pocket of the 14 receptors in *Arabidopsis thaliana* that are positioned to influence agonist affinity and selectivity; in PYR1 these correspond to H60, I62, V83, F108, I110, A160, V163, V164 (Figs. 8.8b, c and 8.10) [31]. Understanding the molecular determinants of agonist selectivity will help guide the design of new selective probes of receptor function. Characterization of diverse agonists has shown that a few key residues influence the affinity and selectivity of multiple agonist–receptor interactions.

The first molecular understanding of agonist selectivity came from analyzing the synthetic agonist pyrabactin (Fig. 8.5), which differentially affects PYR1 and PYL2 (IC_{50} 307 nM and > 10,000 nM, respectively) (Park et al. 2009; Okamoto et al. 2013). A genetic selection of mutagenized PYR1 was conducted to identify point mutations that eliminate pyrabactin responsiveness but retain ABA responsiveness, using a simple yeast-two-hybrid (Y2H)-based selection (Peterson et al. 2010). This defined two mutations in PYR1 (I62V and I110V) that recapitulate natural sequence differences between PYL2 and PYR1. Residues homologous to I110 in PYR1 were also defined as selectivity determinants by structural biological approaches (Melcher



B

Clade III	PYR1	H60	I62	V83	F108	I110	A160	V163	V164
	PYL1	H87	I89	V110	F135	I137	A190	V193	I194
	PYL2	H65	V67	V87	F112	V114	V166	V169	V170
	PYL3	H80	I82	V107	F132	V134	V189	V192	V193
Clade II	PYL4	H82	L84	V105	F130	V132	V179	I182	V183
	PYL5	N88	I90	V111	F136	V138	V185	I188	V189
	PYL6	H91	V93	V114	F139	V141	A193	I196	V197
	PYL11	Q40	V42	V62	I87	I89	A138	I141	V142
	PYL12	H40	V42	V62	I87	I89	A137	I140	V141
	PYL13	R39	V41	L67	V92	I94	V144	I147	I148
Clade I	PYL7	P66	I68	V87	I112	I114	V164	L167	I168
	PYL9	P64	V66	V85	I110	I112	V162	L165	I166
	PYL8	P62	I64	V83	I108	I110	V160	L163	I164
	PYL10	P57	I59	L79	I104	I106	V156	L159	I160

Fig. 8.10 (**a**) Structural depiction of variable residues within the PYR1 binding pocket, colored in yellow. The latch loop is highlighted in purple, and the gate in green. ABA is depicted in ball-and-stick form in purple. (**b**) Variable residues by homologous position between the 14 *Arabidopsis* receptors

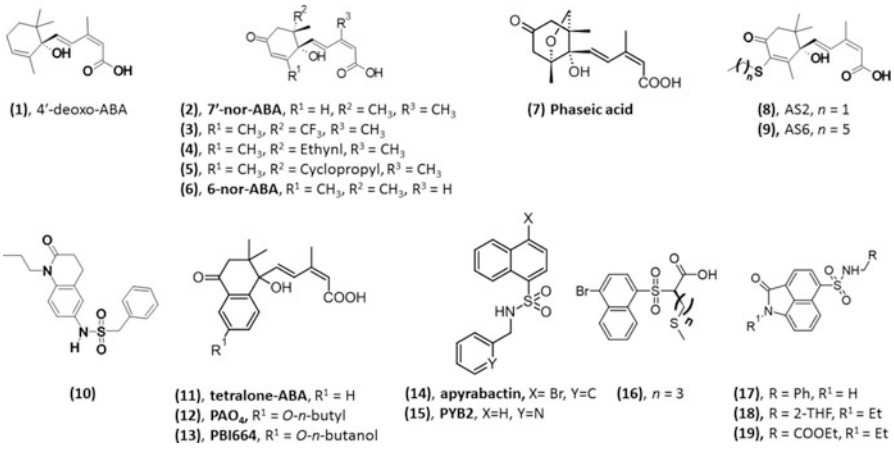


Fig. 8.11 Structures of various analogs from the ABA or pyrabactin scaffold that were useful in structure–activity relationship studies

et al. 2010; Yuan et al. 2010); I110V affects pyrabactin response more strongly and a $PYR1^{I62V,I110V}$ double almost fully eliminates pyrabactin responsiveness (Peterson et al. 2010). Thus, converting PYR1's ligand-binding pocket to be more PYL2-like recapitulates PYL2's insensitivity to pyrabactin. Conversely, the $PYL2^{V67I,V114I}$ mutation imbues pyrabactin-responsiveness upon PYL2 (Peterson et al. 2010).

Unexpectedly, both PYR1 and PYL2 were found to bind pyrabactin (Peterson et al. 2010). The difference in pyrabactin sensitivity between PYR1 and PYL2 is due to differences in pyrabactin orientation; the larger PYL2 pocket enables pyrabactin to rotate ~180° and adopt an orientation that blocks gate closure. In PYR1, I110 makes hydrophobic contacts to the pyridyl ring and blocks the rotation that is possible in PYL2, which has a smaller valine residue at the homologous position. The non-productive binding orientation is able to antagonize ABA action on PYL2, although pyrabactin is a weak antagonist in comparison to the rationally designed antagonist AS6 (see below).

Pyrabactin was also informative for defining the importance of contacts made with the conserved lysine residue (homologous to PYR1 K59) of the receptors. Apyrabactin (**14**, Fig. 8.11), an analog of pyrabactin that lacks the pyridyl nitrogen, is biologically inactive because it cannot form a hydrogen bond to K59 (Melcher et al. 2010). The homologous K64A mutation in PYL2 abolishes ABA binding, demonstrating the importance of this residue to ABA recognition (Yin et al. 2009). Among *Arabidopsis* receptors, PYL13 is the only receptor lacking this lysine and instead contains a Gln in the homologous position. PYL13 was initially thought to be ABA-insensitive because it does not inhibit ABI1, HAB1, or HAB2 in response to ABA in vitro (Li et al. 2013). However, it was subsequently shown that PYL13 selectively regulates ABI1, ABI2, and PP2CA in an ABA-dependent manner (Fuchs

et al. 2014). Thus, a lysine in the K59 position is not always essential for ABA binding.

The hydrophobic gate-latch interface is important for interactions with ABA's cyclohexenone ring and analogs that increase hydrophobic contact with the gate, such as 8′-trifluoro ABA (**3**, Fig. 8.11) (Todoroki et al. 1995), 8′-acetylene-ABA (**4**, Fig. 8.11) (Benson et al. 2015), or 8′-cyclopropyl ABA (**5**, Fig. 8.11) (Benson et al. 2015) retain ABA-like potency in vitro. The analog 7′-nor-ABA, (**2**, Fig. 8.11) showed reduced activity, consistent with the importance of hydrophobic interactions near the gate-latch to binding. Similarly, pyrabactin's naphthyl ring makes several contacts in this region and the analog PYB2, which replaces pyrabactin's bromine with a hydrogen atom (**15**, Fig. 8.11), dramatically reduces activity (Hao et al. 2010). The available PYL1-pyrabactin structures suggest that the bromine substituent improves activity by stabilizing the gate-latch closure through interactions with V110 and L114 (Hao et al. 2010).

Results from a virtual screening effort that sought to identify pyrabactin-analogs produced multiple active agonists that closely resemble pyrabactin (Fig. 8.11) (Melcher et al. 2010). The methionine derivative of pyrabactin, where the pyridyl ring is swapped for a methionine sidechain (**16**, Fig. 8.11) showed stronger in vitro activity than pyrabactin. From the same screen, other pyrabactin analogs containing a tricyclic naphtholactam moiety (**17–19**, Fig. 8.11) gave moderate activity in a phosphatase inhibition assay.

Another synthetic sulfonamide ABA-receptor agonist, quinabactin (Fig. 8.5), was identified in two independent small molecule screens, one using a yeast two-hybrid system and the other a direct PYR1-HAB1 biochemical assay (Cao et al. 2013). Quinabactin activates a larger subset of the ABA-receptors compared to pyrabactin and has greater potency, displaying nanomolar IC_{50} values on PYR1, PYL1, PYL2, PYL3, and PYL5 in receptor-mediated phosphatase inhibition assays (Fig. 8.5). Unlike pyrabactin, quinabactin elicits vegetative ABA responses due to the expanded activity on the receptor family, indicating that activation of the dimeric receptors is sufficient for this response. Overlaying the 3D conformations of PYL-bound quinabactin and pyrabactin reveals a similar binding orientation. Both synthetic agonists display a hydrophobe-hydrophile-hydrophobe general structure (Fig. 8.12). However, quinabactin, unlike pyrabactin, makes a hydrogen bond to the conserved gate-latch water H-bond network, mimicking ABA's cyclohexenone-oxygen. Removal of the cyclohexenone-oxygen from ABA (4′-deoxo-ABA, **1**, Fig. 8.11) greatly diminishes bioactivity, supporting the importance of interactions with the gate-latch to binding (Takahashi et al. 1986). ABA's C6-methyl is positioned in a small pocket (the C6-cleft) that can also be occupied by quinabactin's 4-methylbenzyl moiety (Fig. 8.12b). Quinabactin-analogs lacking the 4-methyl group (replaced by a H, **10**, Fig. 8.11) lose activity (Helander et al. 2016; Cao et al. 2013), which is also true of ABA analogs lacking the C6-methyl (**6**, Fig. 8.11) (Takeuchi et al. 2015a) suggesting that occupancy of the C6-cleft is important for agonist activity. Moreover, sequence variability in the C6-cleft (corresponding to A160 in PYR1; Fig. 8.10) influences ABA sensitivity (Nakagawa et al. 2014). Thus,

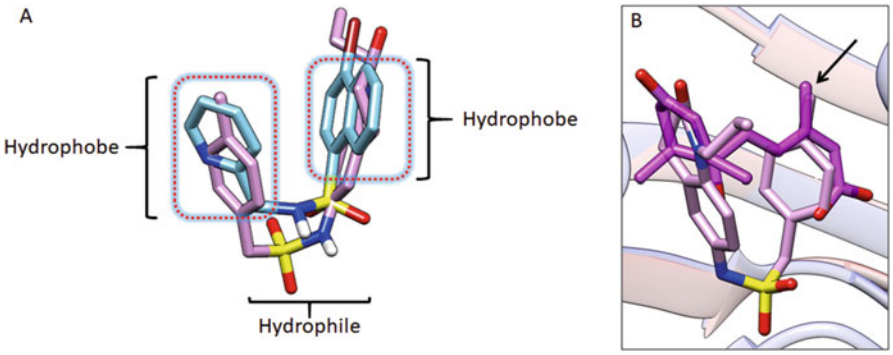


Fig. 8.12 (**a**) The hydrophobe-hydrophile-hydrophobe structure of overlaid pyrabactin and quinabactin. (**b**) Overlay of ABA (pink, transparent, PDB ID: 3K3K) and quinabactin (plum, PDB ID: 4LA7) with arrow depicting the C6-methyl of ABA superimposed over the toluyl methyl of quinabactin

the C6-cleft is important to agonist activity and influences natural differences in ABA-sensitivity between receptors.

Characterizing unnatural ABA analogs, such as the *R*-(−)-ABA isomer, has also been useful for probing molecular determinants of agonist selectivity. *R*-(−)-ABA can bind some of the receptors, with reduced affinity, by rotation of its cyclohexenone ring 180° (Nishimura et al. 2009), which leads to its 8′- and 9′-methyl groups interacting with residues that would normally interact (+)-ABA's 7′-methyl (Nishimura et al. 2009; Zhang et al. 2013; Milborrow 1974). As noted above, the ABA receptor family possesses variability in multiple ABA-proximal residues including residues near ABA's C8'-methyl (Fig. 8.10). PYL9 possesses a relatively low affinity for (−)-ABA, due in part to variability in a conserved hydrophobic residue near (+)-ABA's C8'-methyl (V66 in PYL9, see Fig. 8.10). The mutation $PYL9^{V66I}$ confers partial responsive to *R*-(−)-ABA (Zhang et al. 2013). These observations mirror agonist selectivity differences between PYR1 and PYL2 where variation at the homologous position (V66 in PYR1) influences pyrabactin sensitivity.

Given the variability of ABA receptor pocket architecture on agonist selectivity, it is likely that there are underlying functional differences that impact physiology. We previously speculated that the variation may determine selectivity to different endogenous ligands. This has been validated by elegant studies that have shown that the ABA catabolite phaseic acid (PA) (**7**, Fig. 8.11) is a selective PYL agonist. It is most active on PYL3, PYL5, and PYL6, with IC_{50} values of 264, 567, and 637 nM, respectively (Weng et al. 2016). Weng et al. produced PYL-PA crystal structures for both PYL2 and PYL3 which depicted for the lesser affinity in PYL2 (4504 nM) that there was a perceived ~60° tilt of the cyclohexenone head group which simultaneously disrupted the water-mediated hydrogen-bond network normally seen in the PYL2-ABA complex. The dramatic difference between PYL3 (264 nM) and PYR1 (9556 nM) affinities for PA were partially explained by a

residue difference at PYR1 homologous position of I110, which would result in a steric clash if placed into PYL3's respective V134. This underlying sequence basis for PA selectivity is similar to that described for pyrabactin, where I110/V114 in PYR1 and PYL2 function as the major selectivity determinant (Peterson et al. 2010).

8.3.5 *Structure-Aided Design of PYL Agonists and Antagonists*

The availability of multiple crystal structures is enabling for virtual screening and rational design. Virtual screening has yielded some pyrabactin analogs with improved properties (Melcher et al. 2010) and rational design has produced a new family of ABA-receptor antagonists (Takeuchi et al. 2014). One interesting difference between quinabactin and pyrabactin is the presence of quinabactin's hydrophobic n-propyl chain, which interacts with a small solvent-exposed pore situated above ABA's C3' that has been named the "3′-tunnel" (Takeuchi et al. 2014). This pore is formed in PYR1 by the residues P88, F159, V163, L87, and F61. With the exception of V163, these are invariant across ABA receptor subfamilies. Quinabactin analogs with shortened *N*-alkyl chains are less potent agonists (Cao et al. 2013), pointing to this region as exploitable for tuning agonist activity. The 3′-tunnel contacts PP2Cs in ternary complexes, which suggested that it could be exploited to design antagonists. Pursuing this idea, Todoroki and co-workers created a series 3′-alkylsulfanyl ABA analogs. The two-carbon chain on 3′-ethylsulfanyl-ABA (AS2, compound 8 in Fig. 8.11) is positioned to occupy the 3′-tunnel without disrupting PP2C interactions; indeed it is a potent ABA agonist in vitro and in vivo (Takeuchi et al. 2014) (Fig. 8.13). By lengthening the AS chain to 5-carbons and beyond (AS5-AS12, AS6, **9**, Fig. 8.11), it was demonstrated that there was dramatically reduced PP2C-inhibition by PYR1 with these compounds in vitro. This mechanism of antagonism was also evident in vivo where seeds that were co-incubated with ABA and the 3′-alkylsulfanyl antagonists showed increased germination in comparison to ABA-treated seeds (Takeuchi et al. 2014).

Additional studies showed that the 3′-tunnel can be exploited with the bicyclic tetralone-ABA scaffold (**11**, Fig. 8.11) (Nyangulu et al. 2006). By extending an *alkoxy* linker at the 11′-position (analogous to the 3′-position of ABA, **12**, Fig. 8.11) on the tetralone-ABA derivative, the alkyl chain was predicted to occupy the 3-′-tunnel. Short chains of $n = 1$–3 carbon lengths showed improved agonist activity, while chains of $n \geq 4$ produced antagonist activity (Takeuchi et al. 2015b; Rajagopalan et al. 2016). Interestingly, the addition of a propanol (PBI664; **13**, Fig. 8.11) created a pan-antagonist while the PBI911 (identical to PAO4, **12**, Fig. 8.11) was a selective antagonist (Rajagopalan et al. 2016).

Rational design efforts will be enhanced by the discovery of new scaffolds that broaden opportunities for chemical synthesis. To date, virtual screening has yielded sulfonamides relatively similar to pyrabactin (Melcher et al. 2010). It is likely that

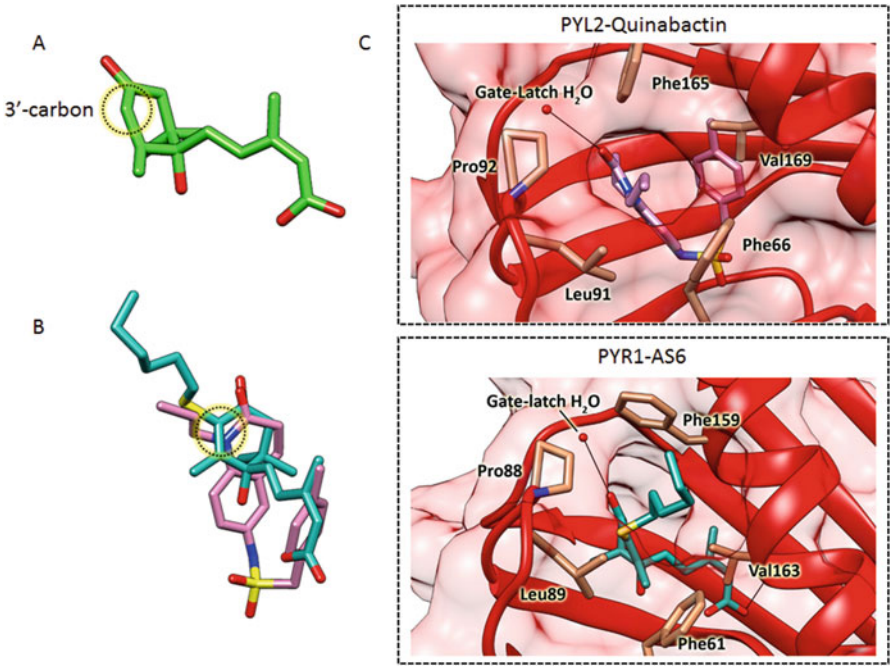


Fig. 8.13 Design of PYL-antagonist compound AS6. (**a**) 3D conformation of ABA, with targeted 3′-carbon indicated. (**b**) Overlay of quinabactin's (pink) *N*-alkyl chain, which extends into the 3′-tunnel, with AS6 (teal) with an extended alkyl chain designed to protrude from the 3′-tunnel. (**c**) Viewpoint looking down the 3′-tunnel, formed around the *N*-propyl chain with quinabactin binding to PYL2 (top, PDB ID: 4LA7) and similar viewpoint looking into the 3′-tunnel of AS6 binding to PYR1 (bottom, PDB ID: 3 K90)

the sulfonamide linker can be manipulated to other hydrophilic moieties and phosphonamide pyrabactin agonists have been described (Van Overtveldt et al. 2015). New antagonist scaffolds discovered have also been discovered by high throughput screening employing chemical arrays of natural products immobilized on glass slide arrays, however the compounds identified lack useful activity in seed germination assays due to instability in vivo (Ito et al. 2015).

8.4 Conclusion

Our current understanding of ABA biosynthesis and signaling has evolved to a state where rational design can be employed to produce selective probes for key targets such as ABA receptors and NCEDs. The large, evolutionarily conserved, size of the ABA receptor gene family raises the question as to why plants require so many receptors for a single hormone, especially when activating a small subset of dimeric receptors is sufficient to induce many classic ABA responses. It is clear that dimeric

(subfamily III) ABA receptors are validated targets for manipulating transpiration, but it remains to be determined whether other receptor subtypes may prove similarly exploitable for agricultural uses. The development of new agonists selective for monomeric ABA receptors could help address this issue, as will the phenotypic characterization of mutants lacking each subfamily of receptors. In this context, the availability of substantial structural receptor data could prove useful for designing selective probes. It will additionally be beneficial to design next-generation NCED inhibitors with broad spectrum activity against monocot and dicot NCEDs. Lastly, the identification of PA (Weng et al. 2016) as an endogenous selective ABA agonist could point to the existence of other plant ligands used to control receptor subfamily-specific outputs. Numerous tools for discovering such ligands have been created including yeast-two hybrid reporter strains and simple and robust receptor-mediated PP2C inhibition assays.

The rapid rise of plant genome editing (Lozano-Juste and Cutler 2014) may afford the possibility to rationally exploit existing structural data to manipulate plant phenotypes in unprecedented ways. For example, we recently described engineered ABA receptors that contain 6 mutations and respond to the agrichemical mandipropamid, rather than ABA, with nM sensitivity. Once mature, targeted genome editing may allow the creation of non-transgenic plant varieties that activate ABA signaling in response to mandipropamid, which would allow an existing agrochemical to control transpiration in a non-transgenic plant. A detailed understanding of structure–function relationships for important plant target proteins, such as ABA receptors, will help guide the rational modification of plant phenotypes to benefit agricultural applications and expand opportunities for engineering novel chemically controlled phenotypes.

References

Abrams SR, Rose PA, Cutler AJ, Balsevich JJ, Lei B, Walker-Simmons MK (1997) 8[prime]-methylene Abscisic acid (an effective and persistent analog of Abscisic acid). Plant Physiol 114:89–97

Anderberg RJ, Walker-Simmons MK (1992) Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinases. Proc Natl Acad Sci U S A 89:10183–10187

Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, Maclean DJ, Ebert PR, Kazan K (2004) Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. Plant Cell 16:3460–3479

Auldridge ME, McCarty DR, Klee HJ (2006) Plant carotenoid cleavage oxygenases and their apocarotenoid products. Curr Opin Plant Biol 9:315–321

Belin C, de Franco P-O, Bourbousse C, Chaignepain S, Schmitter J-M, Vavasseur A, Giraudat J, Barbier-Brygoo H, Thomine S (2006) Identification of features regulating OST1 kinase activity and OST1 function in guard cells. Plant Physiol 141:1316–1327

Benson CL, Kepka M, Wunschel C, Rajagopalan N, Nelson KM, Christmann A, Abrams SR, Grill E, Loewen MC (2015) Abscisic acid analogs as chemical probes for dissection of abscisic acid responses in Arabidopsis thaliana. Phytochemistry 113:96–107

Bhaskara GB, Nguyen TT, Verslues PE (2012) Unique drought resistance functions of the highly ABA-induced clade a protein phosphatase 2Cs. Plant Physiol 160:379–395

Cao M, Liu X, Zhang Y, Xue X, Zhou XE, Melcher K, Gao P, Wang F, Zeng L, Zhao Y, Zhao Y, Deng P, Zhong D, Zhu J-K, Xu HE, Xu Y (2013) An ABA-mimicking ligand that reduces water loss and promotes drought resistance in plants. Cell Res 23:1043–1054

Chiwocha SDS, Cutler AJ, Abrams SR, Ambrose SJ, Yang J, Ross ARS, Kermode AR (2005) The etr1-2 mutation in Arabidopsis thaliana affects the abscisic acid, auxin, cytokinin and gibberellin metabolic pathways during maintenance of seed dormancy, moist-chilling and germination. Plant J 42:35–48

Christmann A, Weiler EW, Steudle E, Grill E (2007) A hydraulic signal in root-to-shoot signalling of water shortage. Plant J 52:167–174

Cohen P (1989) The structure and regulation of protein phosphatases. Annu Rev Biochem 58:453–508

Creelman RA, Mason HS, Bensen RJ, Boyer JS (1990) Water deficit and abscisic acid cause differential inhibition of shoot versus root growth in soybean seedlings analysis of growth, sugar accumulation, and gene expression. *Plant Physiology 92*(1):205–214

Creelman RA, Bell E, Mullet JE (1992) Involvement of a lipoxygenase-like enzyme in abscisic acid biosynthesis. Plant Physiol 99:1258–1260

Cutler S, Ghassemian M, Bonetta D, Cooney S, McCourt P (1996) A protein farnesyl transferase involved in abscisic acid signal transduction in Arabidopsis. Science 273:1239–1241

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: emergence of a core signaling network. Annu Rev Plant Biol 61:651–679

De Smet I, Signora L, Beeckman T, Inzé D (2003) An abscisic acid-sensitive checkpoint in lateral root development of Arabidopsis. The Plant

De Smet I, Zhang H, Inzé D, Beeckman T (2006) A novel role for abscisic acid emerges from underground. Trends Plant Sci 11:434–439

Deak KI, Malamy J (2005) Osmotic regulation of root system architecture. Plant J 43:17–28

Duan L, Dietrich D, Ng CH, Chan PMY, Bhalerao R, Bennett MJ, Dinneny JR (2013) Endodermal ABA signaling promotes lateral root quiescence during salt stress in Arabidopsis seedlings. Plant Cell 25:324–341

Dupeux F, Santiago J, Betz K, Twycross J, Park S-Y, Rodriguez L, Gonzalez-Guzman M, Jensen MR, Krasnogor N, Blackledge M, Holdsworth M, Cutler SR, Rodriguez PL, Márquez JA (2011a) A thermodynamic switch modulates abscisic acid receptor sensitivity. EMBO J 30:4171–4184

Dupeux F, Antoni R, Betz K, Santiago J, Gonzalez-Guzman M, Rodriguez L, Rubio S, Park S-Y, Cutler SR, Rodriguez PL, Márquez JA (2011b) Modulation of abscisic acid signaling in vivo by an engineered receptor-insensitive protein phosphatase type 2C allele. Plant Physiol 156:106–116

Endo A, Okamoto M, Koshiba T (2014) ABA biosynthetic and catabolic pathways. In: Zhang D-P (ed) Abscisic acid: metabolism, transport and signaling. Springer, Dordrecht, pp 21–45

Fedoroff NV (2002) Cross-talk in abscisic acid signaling. Sci STKE 2002:re10

Fuchs S, Tischer SV, Wunschel C, Christmann A, Grill E (2014) Abscisic acid sensor RCAR7/PYL13, specific regulator of protein phosphatase coreceptors. Proc Natl Acad Sci U S A 111:5741–5746

Fujii H, Zhu J-K (2009) Arabidopsis mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. Proc Natl Acad Sci U S A 106:8380–8385

Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park SY, Cutler SR, Sheen J, Rodriguez PL, Zhu JK (2009) In vitro reconstitution of an abscisic acid signalling pathway. Nature 462:660–664

Fujita Y, Nakashima K, Yoshida T, Katagiri T, Kidokoro S, Kanamori N, Umezawa T, Fujita M, Maruyama K, Ishiyama K, Kobayashi M, Nakasone S, Yamada K, Ito T, Shinozaki K, Yamaguchi-Shinozaki K (2009) Three SnRK2 protein kinases are the main positive regulators

of abscisic acid signaling in response to water stress in Arabidopsis. Plant Cell Physiol 50:2123–2132
Gao S, Gao J, Zhu X, Song Y, Li Z, Ren G, Zhou X, Kuai B (2016) ABF2, ABF3, and ABF4 promote ABA-mediated chlorophyll degradation and leaf senescence by transcriptional activation of chlorophyll catabolic genes and senescence-associated genes in Arabidopsis. Mol Plant 9:1272–1285
Geng Y, Wu R, Wee CW, Xie F, Wei X, Chan PMY, Tham C, Duan L, Dinneny JR (2013) A spatio-temporal understanding of growth regulation during the salt stress response in Arabidopsis. Plant Cell 25:2132–2154
Gonzalez-Guzman M, Pizzio GA, Antoni R, Vera-Sirera F, Merilo E, Bassel GW, Fernández MA, Holdsworth MJ, Perez-Amador MA, Kollist H, Rodriguez PL (2012) Arabidopsis PYR/PYL/RCAR receptors play a major role in quantitative regulation of stomatal aperture and transcriptional response to abscisic acid. https://doi.org/10.1105/tpc.112.098574
Gonzalez-Guzman M, Rodriguez L, Lorenzo-Orts L, Pons C, Sarrion-Perdigones A, Fernandez MA, Peirats-Llobet M, Forment J, Moreno-Alvero M, Cutler SR, Albert A, Granell A, Rodriguez PL (2014) Tomato PYR/PYL/RCAR abscisic acid receptors show high expression in root, differential sensitivity to the abscisic acid agonist quinabactin, and the capability to enhance plant drought resistance. J Exp Bot 65(15):4451–4464
Gosti F, Beaudoin N, Serizet C, Webb AA, Vartanian N, Giraudat J (1999) ABI1 protein phosphatase 2C is a negative regulator of abscisic acid signaling. Plant Cell 11:1897–1910
Gusta LV, Trischuk R, Weiser CJ (2005) Plant cold acclimation: the role of Abscisic acid. J Plant Growth Regul 24:308–318
Han S-Y, Inoue H, Terada T, Kamoda S, Saburi Y, Sekimata K, Saito T, Kobayashi M, Shinozaki K, Yoshida S, Asami T (2002) Design and synthesis of lignostilbene-α,-β-dioxygenase inhibitors. Bioorg Med Chem Lett 12:1139–1142
Han S-Y, Inoue H, Terada T, Kamoda S, Saburi Y, Sekimata K, Saito T, Kobayashi M, Shinozaki K, Yoshida S, Asami T (2003) N -Benzylideneaniline and N -Benzylaniline are Potent Inhibitors of Lignostilbene-α,β-dioxygenase, a Key Enzyme in Oxidative Cleavage of the Central Double Bond of Lignostilbene. J Enzyme Inhib Med Chem 18:279–283
Han S-Y, Kitahata N, Sekimata K, Saito T, Kobayashi M, Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K, Yoshida S, Asami T (2004) A novel inhibitor of 9-cis-epoxycarotenoid dioxygenase in abscisic acid biosynthesis in higher plants. Plant Physiol 135:1574–1582
Hao Q, Yin P, Yan C, Yuan X, Li W, Zhang Z, Liu L, Wang J, Yan N (2010) Functional mechanism of the abscisic acid agonist pyrabactin. J Biol Chem 285:28946–28952
Hao Q, Yin P, Li W, Wang L, Yan C, Lin Z, Wu JZ, Wang J, Yan SF, Yan N (2011) The molecular basis of ABA-independent inhibition of PP2Cs by a subclass of PYL proteins. Mol Cell 42:662–672
Hardie DG (2007) AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. Nat Rev Mol Cell Biol 8:774–785
Harrison PJ, Bugg TDH (2014) Enzymology of the carotenoid cleavage dioxygenases: reaction mechanisms, inhibition and biochemical roles. Arch Biochem Biophys 544:105–111
Hartung W (2010) The evolution of abscisic acid (ABA) and ABA function in lower plants, fungi and lichen. Funct Plant Biol 37:806–812
Hauser F, Waadt R, Schroeder JI (2011) Evolution of abscisic acid synthesis and signaling mechanisms. Curr Biol 21:R346–R355
He Y, Hao Q, Li W, Yan C, Yan N, Yin P (2014) Identification and characterization of ABA receptors in Oryza sativa. PLoS One 9:e95246
Hedbacker K, Carlson M (2008) SNF1/AMPK pathways in yeast. Front Biosci 13:2408–2420
Helander JDM, Vaidya AS, Cutler SR (2016) Chemical manipulation of plant water use. Bioorg Med Chem 24:493–500
Hilhorst HWM, Karssen CM (1992) Seed dormancy and germination: the role of abscisic acid and gibberellins and the importance of hormone mutants. Plant Growth Regul 11:225–238

Hrabak EM, Chan CWM, Gribskov M, Harper JF, Choi JH, Halford N, Kudla J, Luan S, Nimmo HG, Sussman MR, Thomas M, Walker-Simmons K, Zhu J-K, Harmon AC (2003) The Arabidopsis CDPK-SnRK superfamily of protein kinases. Plant Physiol 132:666–680

Ito T, Kondoh Y, Yoshida K, Umezawa T, Shimizu T, Shinozaki K, Osada H (2015) Novel Abscisic acid antagonists identified with chemical Array screening. Chembiochem 16:2471–2478

Iuchi S, Kobayashi M, Taji T, Naramoto M, Seki M, Kato T, Tabata S, Kakubari Y, Yamaguchi-Shinozaki K, Shinozaki K (2001) Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in Arabidopsis. Plant J 27:325–333

Iyer LM, Koonin EV, Aravind L (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43:134–144

Jones AM, Danielson JA, Manojkumar SN, Lanquar V, Grossmann G, Frommer WB (2014) Abscisic acid dynamics in roots detected with genetically encoded FRET sensors. elife 3:e01741

Kerk D, Bulgrien J, Smith DW, Barsam B, Veretnik S, Gribskov M (2002) The complement of protein phosphatase catalytic subunits encoded in the genome of Arabidopsis. Plant Physiol 129:908–925

Khandelwal A, Cho SH, Marella H, Sakata Y, Perroud P-F, Pan A, Quatrano RS (2010) Role of ABA and ABI3 in desiccation tolerance. Science 327:546

Kitahata N, Han S-Y, Noji N, Saito T, Kobayashi M, Nakano T, Kuchitsu K, Shinozaki K, Yoshida S, Matsumoto S, Tsujimoto M, Asami T (2006) A 9-cis-epoxycarotenoid dioxygenase inhibitor for use in the elucidation of abscisic acid action mechanisms. Bioorg Med Chem 14:5555–5561

Koornneef M, Reuling G, Karssen CM (1984) The isolation and characterization of abscisic acid-insensitive mutants of Arabidopsis thaliana. Physiol Plant 61:377–383

Kulik A, Wawer I, Krzywińska E, Bucholc M, Dobrowolska G (2011) SnRK2 protein kinases—key regulators of plant response to abiotic stresses. OMICS 15:859–872

Kushiro T, Okamoto M, Nakabayashi K, Yamagishi K, Kitamura S, Asami T, Hirai N, Koshiba T, Kamiya Y, Nambara E (2004) The Arabidopsis cytochrome P450 CYP707A encodes ABA 8′-hydroxylases: key enzymes in ABA catabolism. EMBO J 23:1647–1656

Leube MP, Grill E, Amrhein N (1998) ABI1 of Arabidopsis is a protein serine/threonine phosphatase highly regulated by the proton and magnesium ion concentration. FEBS Lett 424:100–104

Leung J, Bouvier-Durand M, Morris PC, Guerrier D, Chefdor F, Giraudat J (1994) Arabidopsis ABA response gene ABI1: features of a calcium-modulated protein phosphatase. Science 264:1448–1452

Leung J, Merlot S, Giraudat J (1997) The Arabidopsis ABSCISIC ACID-INSENSITIVE2 (ABI2) and ABI1 genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. Plant Cell 9:759–771

Li J, Assmann SM (1996) An Abscisic acid-activated and calcium-independent protein kinase from guard cells of fava bean. Plant Cell 8:2359–2368

Li J, Wang XQ, Watson MB, Assmann SM (2000) Regulation of abscisic acid-induced stomatal closure and anion channels by guard cell AAPK kinase. Science 287:300–303

Li W, Wang L, Sheng X, Yan C, Zhou R, Hang J, Yin P, Yan N (2013) Molecular basis for the selective and ABA-independent inhibition of PP2CA by PYL13. Cell Res 23:1369–1379

Li J, Shi C, Sun D, He Y, Lai C, Lv P, Xiong Y, Zhang L, Wu F, Tian C (2015) The HAB1 PP2C is inhibited by ABA-dependent PYL10 interaction. Sci Rep 5:10890

Liang C, Wang Y, Zhu Y, Tang J, Hu B, Liu L, Ou S, Wu H, Sun X, Chu J, Chu C (2014) OsNAP connects abscisic acid and leaf senescence by fine-tuning abscisic acid biosynthesis and directly targeting senescence-associated genes in rice. Proc Natl Acad Sci U S A 111:10013–10018

Lind C, Dreyer I, López-Sanjurjo EJ, von Meyer K, Ishizaki K, Kohchi T, Lang D, Zhao Y, Kreuzer I, Al-Rasheid KAS, Ronne H, Reski R, Zhu J-K, Geiger D, Hedrich R (2015) Stomatal guard cells co-opted an ancient ABA-dependent desiccation survival system to regulate stomatal closure. Curr Biol. https://doi.org/10.1016/j.cub.2015.01.067

Littler DR, Walker JR, Davis T, Wybenga-Groot LE, Finerty PJ Jr, Newman E, Mackenzie F, Dhe-Paganon S (2010) A conserved mechanism of autoinhibition for the AMPK kinase domain: ATP-binding site and catalytic loop refolding as a means of regulation. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:143–151

Louis-Flamberg P, Krupinski-Olsen R, Shorter AL, Kemal C (1988) Reductive inhibition of soybean Lipoxygenase-1 by NDGA a possible mechanism for regulation of Lipoxygenase Activitya. Ann N Y Acad Sci 524:382–384

Lozano-Juste J, Cutler SR (2014) Plant genome engineering in full bloom. Trends Plant Sci 19:284–287

Luan S (2003) Protein phosphatases in plants. Annu Rev Plant Biol 54:63–92

Lumba S, Toh S, Handfield L-F, Swan M, Liu R, Youn J-Y, Cutler SR, Subramaniam R, Provart N, Moses A, Desveaux D, McCourt P (2014) A mesoscale abscisic acid hormone interactome reveals a dynamic signaling landscape in Arabidopsis. Dev Cell 29:360–372

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E (2009a) Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science 324:1064

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E (2009b) Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science 324:1064

Mayak S, Halevy AH (1972) Interrelationships of ethylene and abscisic acid in the control of rose petal senescence. Plant Physiol 50:341–346

McAdam SAM, Brodribb TJ, Banks JA, Hedrich R, Atallah NM, Cai C, Geringer MA, Lind C, Nichols DS, Stachowski K, Geiger D, Sussmilch FC (2016) Abscisic acid controlled sex before transpiration in vascular plants. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.1606614113

McAinsh MR, Brownlee C, Hetherington AM et al (1990) Abscisic acid-induced elevation of guard cell cytosolic Ca2+ precedes stomatal closure. Nature 343:186–188

Melcher K, Ng L-M, Zhou XE, Soon F-F, Xu Y, Suino-Powell KM, Park S-Y, Weiner JJ, Fujii H, Chinnusamy V, Kovach A, Li J, Wang Y, Li J, Peterson FC, Jensen DR, Yong E-L, Volkman BF, Cutler SR, Zhu J-K, Xu HE (2009) A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. Nature 462:602–608

Melcher K, Xu Y, Ng L-M, Zhou XE, Soon F-F, Chinnusamy V, Suino-Powell KM, Kovach A, Tham FS, Cutler SR, Li J, Yong E-L, Zhu J-K, Xu HE (2010) Identification and mechanism of ABA receptor antagonism. Nat Struct Mol Biol 17:1102–1108

Melotto M, Underwood W, Koczan J, Nomura K, He SY (2006) Plant stomata function in innate immunity against bacterial invasion. Cell 126:969–980

Menges M, Hennig L, Gruissem W, Murray JAH (2002) Cell cycle-regulated gene expression in Arabidopsis. J Biol Chem 277:41987–42002

Messing SAJ, Gabelli SB, Echeverria I, Vogel JT, Guan JC, Tan BC, Klee HJ, McCarty DR, Mario Amzel L (2010) Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. https://doi.org/10.1105/tpc.110.074815

Meyer K, Leube MP, Grill E (1994) A protein phosphatase 2C involved in ABA signal transduction in Arabidopsis thaliana. Science 264:1452–1455

Milborrow BV (1974) The chemistry and physiology of Abscisic acid. Annu Rev Plant Physiol 25:259–307

Miyazono K-I, Miyakawa T, Sawano Y, Kubota K, Kang H-J, Asano A, Miyauchi Y, Takahashi M, Zhi Y, Fujita Y, Yoshida T, Kodaira K-S, Yamaguchi-Shinozaki K, Tanokura M (2009) Structural basis of abscisic acid signalling. Nature 462:609–614

Mosquna A, Peterson FC, Park S-Y, Lozano-Juste J, Volkman BF, Cutler SR (2011) Potent and selective activation of abscisic acid receptors in vivo by mutational stabilization of their agonist-bound conformation. Proc Natl Acad Sci U S A 108:20838–20843

Mustilli A-C, Merlot S, Vavasseur A, Fenzi F, Giraudat J (2002) Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. Plant Cell 14:3089–3099

Nakagawa M, Kagiyama M, Shibata N, Hirano Y, Hakoshima T (2014) Mechanism of high-affinity abscisic acid binding to PYL9/RCAR1. Genes Cells 19:386–404

Nakashima K, Fujita Y, Kanamori N, Katagiri T, Umezawa T, Kidokoro S, Maruyama K, Yoshida T, Ishiyama K, Kobayashi M, Shinozaki K, Yamaguchi-Shinozaki K (2009) Three Arabidopsis SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. Plant Cell Physiol 50:1345–1363

Nambara E, Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol 56:165–185

Ng L-M, Soon F-F, Zhou XE, West GM, Kovach A, Suino-Powell KM, Chalmers MJ, Li J, Yong E-L, Zhu J-K, Griffin PR, Melcher K, Xu HE (2011) Structural basis for basal activity and autoactivation of abscisic acid (ABA) signaling SnRK2 kinases. Proc Natl Acad Sci U S A 108:21259–21264

Nishimura N, Yoshida T, Kitahata N, Asami T, Shinozaki K, Hirayama T (2007) ABA-hypersensitive Germination1 encodes a protein phosphatase 2C, an essential component of abscisic acid signaling in Arabidopsis seed. Plant J 50:935–949

Nishimura N, Hitomi K, Arvai AS, Rambo RP, Hitomi C, Cutler SR, Schroeder JI, Getzoff ED (2009) Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science 326:1373–1379

Nyangulu JM, Nelson KM, Rose PA, Gai Y, Loewen M, Lougheed B, Wilson Quail J, Cutler AJ, Abrams SR (2006) Synthesis and biological activity of tetralone abscisic acid analogues. Org Biomol Chem 4:1400–1412

Ofek P, Ben-Meir D, Kariv-Inbal Z, Oren M, Lavi S (2003) Cell cycle regulation and p53 activation by protein phosphatase 2Cα. J Biol Chem 278:14299–14305

Okamoto M, Peterson FC, Defries A, Park S-Y, Endo A, Nambara E, Volkman BF, Cutler SR (2013) Activation of dimeric ABA receptors elicits guard cell closure, ABA-regulated gene expression, and drought tolerance. Proc Natl Acad Sci U S A 110:12132–12137

Park S-Y, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow T-FF, Alfred SE, Bonetta D, Finkelstein R, Provart NJ, Desveaux D, Rodriguez PL, McCourt P, Zhu J-K, Schroeder JI, Volkman BF, Cutler SR (2009) Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science 324:1068–1071

Park S-Y, Peterson FC, Mosquna A, Yao J, Volkman BF, Cutler SR (2015) Agrochemical control of plant water use using engineered abscisic acid receptors. Nature 520:545–548

Pearce LR, Komander D, Alessi DR (2010) The nuts and bolts of AGC protein kinases. Nat Rev Mol Cell Biol 11:9–22

Penfield S, Gilday AD, Halliday KJ, Graham IA (2006) DELLA-mediated cotyledon expansion breaks coat-imposed seed dormancy. Curr Biol 16:2366–2370

Peterson FC, Burgie ES, Park S-Y, Jensen DR, Weiner JJ, Bingman CA, Chang C-EA, Cutler SR, Phillips GN Jr, Volkman BF (2010) Structural basis for selective activation of ABA receptors. Nat Struct Mol Biol 17:1109–1113

Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C, Colot V, Cutler S, Dangl J, Ehrhardt D, Friesner JD, Frommer WB, Grotewold E, Meyerowitz E, Nemhauser J, Nordborg M, Pikaard C, Shanklin J, Somerville C, Stitt M, Torii KU, Waese J, Wagner D, McCourt P (2015) 50 years of Arabidopsis research: highlights and future directions. New Phytol. https://doi.org/10.1111/nph.13687

Qin X, Zeevaart JA (1999) The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stressed bean. Proc Natl Acad Sci U S A 96:15354–15361

Raghavendra AS, Gonugunta VK, Christmann A, Grill E (2010) ABA perception and signalling. Trends Plant Sci 15:395–401

Rajagopalan N, Nelson KM, Douglas AF, Jheengut V, Alarcon IQ, McKenna SA, Surpin M, Loewen MC, Abrams SR (2016) Abscisic acid analogues that act as universal or selective antagonists of Phytohormone receptors. Biochemistry 55:5155–5164

Reinoso H, Travaglia C, Bottini R (2011) ABA Increased Soybean Yield by Enhancing Production of Carbohydrates and Their Allocation in Seed. Soybean–Biochemistry, Chemistry and Physiology InTech, Rijeka, pp 577–598

Robaglia C, Thomas M, Meyer C (2012) Sensing nutrient and energy status by SnRK1 and TOR kinases. Curr Opin Plant Biol 15:301–307

Robert N, Merlot S, N'Guyen V, Boisson-Dernier A, Schroeder JI (2006) A hypermorphic mutation in the protein phosphatase 2C HAB1 strongly affects ABA signaling in Arabidopsis. FEBS Lett 580:4691–4696

Rodriguez PL, Leube MP, Grill E (1998a) Molecular cloning in Arabidopsis thaliana of a new protein phosphatase 2C (PP2C) with homology to ABI1 and ABI2. Plant Mol Biol 38:879–883

Rodriguez PL, Benning G, Grill E (1998b) ABI2, a second protein phosphatase 2C involved in abscisic acid signal transduction in Arabidopsis. FEBS Lett 421:185–190

Rolland F, Baena-Gonzalez E, Sheen J (2006) Sugar sensing and signaling in plants: conserved and novel mechanisms. Annu Rev Plant Biol 57:675–709

Rubio S, Rodrigues A, Saez A, Dizon MB, Galle A, Kim T-H, Santiago J, Flexas J, Schroeder JI, Rodriguez PL (2009) Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. Plant Physiol 150:1345–1355

Saab IN, Sharp RE, Pritchard J, Voetberg GS (1990) Increased endogenous abscisic acid maintains primary root growth and inhibits shoot growth of maize seedlings at low water potentials. Plant Physiol 93:1329–1336

Saez A, Apostolova N, Gonzalez-Guzman M, Gonzalez-Garcia MP, Nicolas C, Lorenzo O, Rodriguez PL (2004) Gain-of-function and loss-of-function phenotypes of the protein phosphatase 2C HAB1 reveal its role as a negative regulator of abscisic acid signalling. Plant J 37:354–369

Santiago J, Rodrigues A, Saez A, Rubio S, Antoni R, Dupeux F, Park S-Y, Márquez JA, Cutler SR, Rodriguez PL (2009a) Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade a PP2Cs. Plant J 60:575–588

Santiago J, Dupeux F, Round A, Antoni R, Park S-Y, Jamin M, Cutler SR, Rodriguez PL, Márquez JA (2009b) The abscisic acid receptor PYR1 in complex with abscisic acid. Nature 462:665–668

Schmidt H, Kurtzer R, Eisenreich W, Schwab W (2006) The Carotenase AtCCD1 from Arabidopsis thaliana is a Dioxygenase. J Biol Chem 281:9845–9851

Schwartz SH, Zeevaart JAD (2010) Abscisic acid biosynthesis and metabolism. In: Davies PJ (ed) Plant hormones. Springer, Dordrecht, pp 137–155

Schwartz SH, Tan BC, Gage DA, Zeevaart JAD, McCarty DR (1997) Specific oxidative cleavage of carotenoids by VP14 of maize. Science 276:1872–1874

Schweighofer A, Hirt H, Meskiene I (2004) Plant PP2C phosphatases: emerging functions in stress signaling. Trends Plant Sci 9:236–243

Seo M (2014) ABA transmembrane transport and transporters. In: Zhang D-P (ed) Abscisic acid: metabolism, transport and signaling. Springer, Dordrecht, pp 47–59

Sharp RE, Poroyko V, Hejlek LG, Spollen WG, Springer GK, Bohnert HJ, Nguyen HT (2004) Root growth maintenance during water deficits: physiology to functional genomics. J Exp Bot 55:2343–2351

Shimizu T, Rådmark O, Samuelsson B (1984) Enzyme with dual lipoxygenase activities catalyzes leukotriene A4 synthesis from arachidonic acid. Proc Natl Acad Sci U S A 81:689–693

Soon F-F, Ng L-M, Zhou XE, West GM, Kovach A, Tan MHE, Suino-Powell KM, He Y, Xu Y, Chalmers MJ, Brunzelle JS, Zhang H, Yang H, Jiang H, Li J, Yong E-L, Cutler S, Zhu J-K, Griffin PR, Melcher K, Xu HE (2012) Molecular mimicry regulates ABA signaling by SnRK2 kinases and PP2C phosphatases. Science 335:85–88

Sussmilch FC, Brodribb TJ, McAdam SAM (2017) What are the evolutionary origins of stomatal responses to abscisic acid in land plants? J Integr Plant Biol 59:240–260

Takahashi S, Oritani T, Yamashita K (1986) Synthesis and biological activities of (±)-Deoxy-abscisic acid isomers. Agric Biol Chem 50:3205–3206

Takeuchi J, Okamoto M, Akiyama T, Muto T, Yajima S, Sue M, Seo M, Kanno Y, Kamo T, Endo A, Nambara E, Hirai N, Ohnishi T, Cutler SR, Todoroki Y (2014) Designed abscisic acid analogs as antagonists of PYL-PP2C receptor interactions. Nat Chem Biol 10:477–482

Takeuchi J, Ohnishi T, Okamoto M, Todoroki Y (2015a) The selectivity of 6-nor-ABA and 7′-nor-ABA for abscisic acid receptor subtypes. Bioorg Med Chem Lett 25:3507–3510

Takeuchi J, Ohnishi T, Okamoto M, Todoroki Y (2015b) Conformationally restricted 3′-modified ABA analogs for controlling ABA receptors. Org Biomol Chem 13:4278–4288

Tan B-C, Joseph LM, Deng W-T, Liu L, Li Q-B, Cline K, McCarty DR (2003) Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. Plant J 35:44–56

Todoroki Y, Hirai N, Koshimizu K (1995) 8′,8′-Difluoro- and 8′,8′,8′-trifluoroabscisic acids as highly potent, long-lasting analogues of abscisic acid. Phytochemistry 38:561–568

Todoroki Y, Tanaka T, Kisamori M, Hirai N (2001) 3′-Azidoabscisic acid as a photoaffinity reagent for abscisic acid binding proteins. Bioorg Med Chem Lett 11:2381–2384

Ton J, Flors V, Mauch-Mani B (2009) The multifaceted role of ABA in disease resistance. Trends Plant Sci 14:310–317

Travaglia C, Reinoso H, Cohen A, Luna C, Tommasino E, Castillo C, Bottini R (2010) Exogenous ABA increases yield in field-grown wheat with moderate water restriction. J Plant Growth Regul 29:366–374

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K (2010a) Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. Plant Cell Physiol 51:1821–1839

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K (2010b) Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. Plant Cell Physiol 51:1821–1839

Van Overtveldt M, TSA H, Verstraeten I, Geelen D, Stevens CV (2015) Phosphonamide pyrabactin analogues as abscisic acid agonists. Org Biomol Chem 13:5260–5264

Waadt R, Hitomi K, Nishimura N, Hitomi C, Adams SR, Getzoff ED, Schroeder JI (2014) FRET-based reporters for the direct visualization of abscisic acid concentration changes and distribution in Arabidopsis. elife 3:e01739

Wang Y, Ying J, Kuzma M, Chalifoux M, Sample A, McArthur C, Uchacz T, Sarvas C, Wan J, Dennis DT, McCourt P, Huang Y (2005) Molecular tailoring of farnesylation for plant drought tolerance and yield protection. Plant J 43:413–424

Wang P, Xue L, Batelli G, Lee S, Hou Y-J, Van Oosten MJ, Zhang H, Tao WA, Zhu J-K (2013) Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. Proc Natl Acad Sci U S A 110:11205–11210

Watts S, Rodriguez JL, Evans SE, Davies WJ (1981) Root and shoot growth of plants treated with Abscisic acid. Ann Bot 47:595–602

Weiner JJ, Peterson FC, Volkman BF, Cutler SR (2010) Structural and functional insights into core ABA signaling. Curr Opin Plant Biol 13:495–502

Weng J-K, Ye M, Li B, Noel JP (2016) Co-evolution of hormone metabolism and signaling networks expands plant adaptive plasticity. Cell 166:881–893

Wenjian L, Xiaoqiang H, Yumei X, Jinlong F, Yuanzhi Z, Huizhe L, Mingan W, Zhaohai Q (2013) Synthesis, photostability and bioactivity of 2,3-cyclopropanated abscisic acid. Phytochemistry 96:72–80

Whitman S, Gezginci M, Timmermann BN, Holman TR (2002) Structure−activity relationship studies of Nordihydroguaiaretic acid inhibitors toward soybean, 12-human, and 15-human Lipoxygenase. J Med Chem 45:2659–2661

Xu Z-J, Nakajima M, Suzuki Y, Yamaguchi I (2002) Cloning and characterization of the abscisic acid-specific glucosyltransferase gene from adzuki bean seedlings. Plant Physiol 129:1285–1295

Yang Y, Costa A, Leonhardt N, Siegel RS, Schroeder JI (2008) Isolation of a strong Arabidopsis guard cell promoter and its potential as a research tool. Plant Methods 4:6

Yang J, Worley E, Udvardi M (2014) A NAP-AAO3 regulatory module promotes chlorophyll degradation via ABA biosynthesis in Arabidopsis leaves. Plant Cell. https://doi.org/10.1105/tpc.114.133769

Yin P, Fan H, Hao Q, Yuan X, Wu D, Pang Y, Yan C, Li W, Wang J, Yan N (2009) Structural insights into the mechanism of abscisic acid signaling by PYL proteins. Nat Struct Mol Biol 16:1230–1236

Yoshida T, Nishimura N, Kitahata N, Kuromori T, Ito T, Asami T, Shinozaki K, Hirayama T (2006a) ABA-hypersensitive germination3 encodes a protein phosphatase 2C (AtPP2CA) that strongly regulates abscisic acid signaling during germination among Arabidopsis protein phosphatase 2Cs. Plant Physiol 140:115–126

Yoshida R, Umezawa T, Mizoguchi T, Takahashi S, Takahashi F, Shinozaki K (2006b) The regulatory domain of SRK2E/OST1/SnRK2.6 interacts with ABI1 and integrates abscisic acid (ABA) and osmotic stress signals controlling stomatal closure in Arabidopsis. J Biol Chem 281:5310–5318

Yoshida R, Umezawa T, Mizoguchi T, Takahashi S, Takahashi F, Shinozaki K (2006c) The regulatory domain of SRK2E/OST1/SnRK2. 6 interacts with ABI1 and integrates abscisic acid (ABA) and osmotic stress signals controlling stomatal closure in Arabidopsis. J Biol Chem 281:5310–5318

Yuan X, Yin P, Hao Q, Yan C, Wang J, Yan N (2010) Single amino acid alteration between valine and isoleucine determines the distinct pyrabactin selectivity by PYL1 and PYL2. J Biol Chem 285:28953–28958

Yuan F, Yang H, Xue Y, Kong D, Ye R, Li C, Zhang J, Theprungsirikul L, Shrift T, Krichilsky B, Johnson DM, Swift GB, He Y, Siedow JN, Pei Z-M (2014) OSCA1 mediates osmotic-stress-evoked Ca2+ increases vital for osmosensing in Arabidopsis. Nature 514:367–371

Zhang X, Zhang Q, Xin Q, Yu L, Wang Z, Wu W, Jiang L, Wang G, Tian W, Deng Z, Wang Y, Liu Z, Long J, Gong Z, Chen Z (2012) Complex structures of the abscisic acid receptor PYL3/RCAR13 reveal a unique regulatory mechanism. Structure 20:780–790

Zhang X, Jiang L, Wang G, Yu L, Zhang Q, Xin Q, Wu W, Gong Z, Chen Z (2013) Structural insights into the abscisic acid stereospecificity by the ABA receptors PYR/PYL/RCAR. PLoS One 8:e67477

Zhang X, Zhang X, Liu X, Shao L, Sun H, Chen S (2016) Improving winter wheat performance by foliar spray of ABA and FA under water deficit conditions. J Plant Growth Regul 35:83–96

Zhao Y, Xing L, Wang X, Hou Y-J, Gao J, Wang P, Duan C-G, Zhu X, Zhu J-K (2014) The ABA receptor PYL8 promotes lateral root growth by enhancing MYB77-dependent transcription of auxin-responsive genes. Sci Signal 7:ra53

Part III
Transmembrane Receptors

Chapter 9
Brassinosteroid Sensing and Signaling in Plants

Ulrich Hohmann and Michael Hothorn

9.1 Brassinosteroids and the Core Brassinosteroid Signaling Pathway

In 1979, a growth-promoting hormone was isolated from rape pollen (Grove et al. 1979). Its crystal structure revealed a hydroxyprolinated steroid featuring a B-ring lactone with structural similarities to the insect molting hormone ecdysone (Fig. 9.1a, b). It was named brassinolide and since its discovery dozens of related brassinosteroids, which can promote cell elongation, division, and differentiation, have been reported from different plant species (Bajguz 2007). Using synthetic brassinolide, brassinosteroid-deficient and -insensitive mutants were identified in forward genetic screens in the model plant *Arabidopsis*. In this way, the core biosynthetic pathway for the hormone and many components of the brassinosteroid signal transduction cascade were defined. One of the first genes to be cloned was BRI1 (Clouse et al. 1996; Li and Chory 1997), a plasma-membrane-localized receptor kinase with an extracellular leucine-rich repeat (LRR) domain, a single membrane spanning helix and a cytoplasmic kinase domain (Li and Chory 1997) (Fig. 9.2). BRI1 is one of ~200 LRR receptor kinases found in the *Arabidopsis* genome (Shiu and Bleecker 2001). BRI1 acts as receptor for brassinosteroids and binds the steroid hormone using its LRR-domain (Wang et al. 2001; Kinoshita et al. 2005).

BRI1 can interact with the BRI1 ASSOCIATED KINASE 1 (BAK1), which is a positive regulator of brassinosteroid signaling (Li et al. 2002; Nam and Li 2002) (Fig. 9.2). BAK1 is part of the SERK family of co-receptor kinases (Brandt and

U. Hohmann · M. Hothorn (✉)
Structural Plant Biology Laboratory, Department of Botany and Plant Biology, University of Geneva, Geneva, Switzerland
e-mail: michael.hothorn@unige.ch

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_9

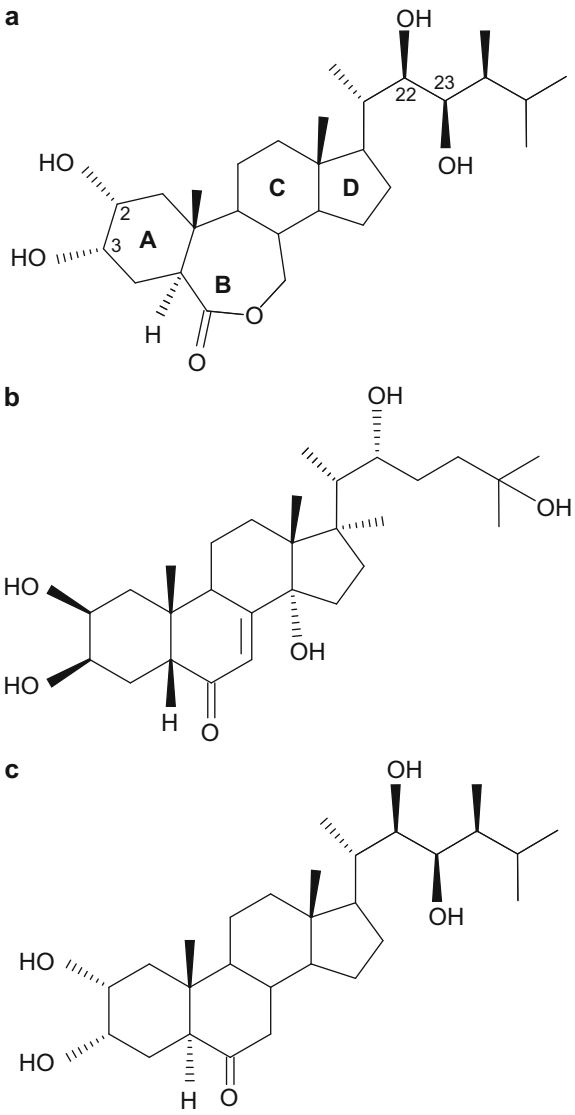


Fig. 9.1 Chemical structure of different steroid molecules. (**a**) Brassinolide, (**b**) ecdysone, and (**c**) castasterone

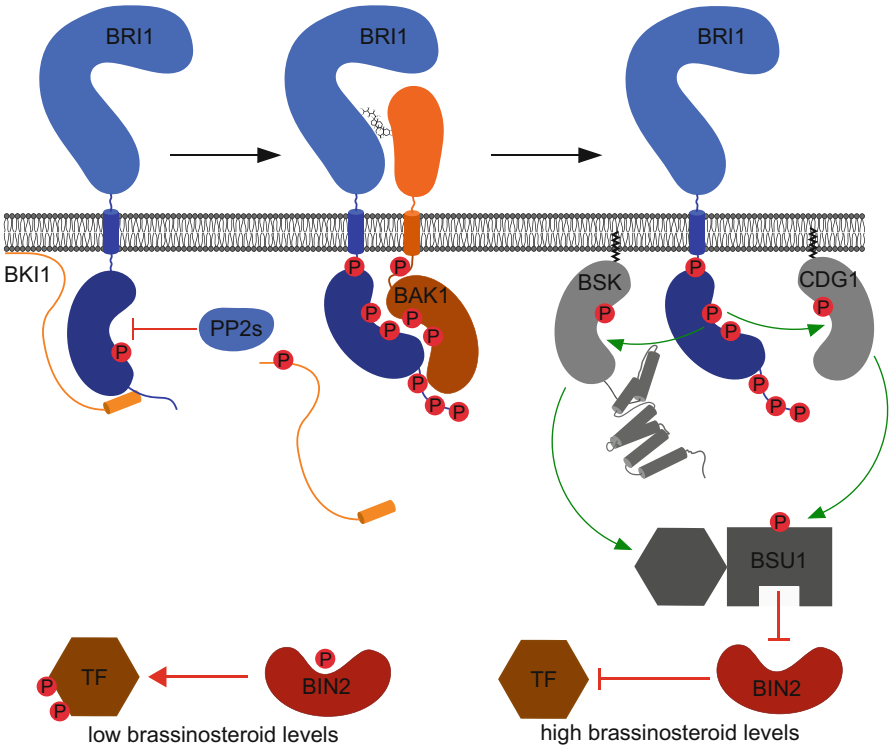


Fig. 9.2 Schematic view of the brassinosteroid signaling cascade. In the absence of brassinosteroids, BRI1 is kept inactive by the inhibitor protein BKI1 and by PP2 family phosphatases. Upon ligand binding, the co-receptor BAK1 is recruited, inhibitory BKI1 is released, and the cytoplasmic kinase domains of BRI1 and BAK1 can auto- and trans-phosphorylate each other. This enables BRI1 to trans-phosphorylate membrane-attached cytoplasmic kinases (BSKs and CDG1), which activates the BSU1 phosphatase. BSU then dephosphorylates the GSK3 kinase BIN2. Thus, when brassinosteroid signaling is active, BIN2 cannot phosphorylate the transcription factors (TFs) BES1 and BZR1, thereby activating them. Green arrows indicate activating phosphorylations and red arrows inhibitory phosphorylation events

Hothorn 2016). Active BRI1 can trans-phosphorylate different cytoplasmic kinases, which are attached to the plasma-membrane via covalent lipid anchors: the BRI1 SIGNALING KINASES (BSKs) (Tang et al. 2008) and CONSTITUTIVE DIFFERENTIAL GROWTH (CDG1) (Kim et al. 2011) (Fig. 9.2). These cytoplasmic kinases can activate the down-stream phosphatase BRASSINOSTEROID INSENSITIVE SUPRESSOR 1 (BSU1) (Mora-García et al. 2004; Kim et al. 2011), which in turn dephosphorylates and inactivates the GSK3-type kinase BRASSINOSTEROID INSENSITIVE 2 (BIN2) (Li et al. 2001; Li and Nam 2002; Mora-García et al. 2004; Kim et al. 2011) (Fig. 9.2). When brassinosteroid levels are low and BRI1 is inactive, BIN2 efficiently phosphorylates the transcription factors BRASSINAZOLE RESISTANT 1 (BZR1) (Wang et al. 2002) and BRI1-EMS-SUPPRESSOR 1 (BES1) (Yin et al. 2002), keeping them in an inactive state

bound to 14-3-3 proteins (Gampala et al. 2007) (Fig. 9.2). When brassinosteroids are sensed by BRI1, BIN2 is inactivated, leading to unphosphorylated BZR1/BES1, which accumulate in the nucleus and bind to their target promoters (Yin et al. 2002; Zhao et al. 2002; Vert and Chory 2006). Over the past few years, a combination of structural biology, quantitative biochemistry, and plant genetics has provided detailed mechanistic insights into the early steps of brassinosteroid signal transduction, namely, how the hormone is specifically sensed by its receptor BRI1 and how ligand binding at the cell surface leads to activation of the cytoplasmic signaling cascade.

9.2 The Structure of the BRI1 Extracellular Domain

Different brassinosteroid-insensitive mutant alleles map to the *bri1* locus (Clouse et al. 1996; Li and Chory 1997). Protein engineering (He et al. 2000) and ligand-binding assays using radiolabeled brassinosteroids (Wang et al. 2001; Kinoshita et al. 2005) revealed that BRI1 acts as a receptor for brassinosteroids. The hormone is directly sensed by the extracellular LRR domain of BRI1, which is exposed at the cell surface and to the lumen of endosomes (Friedrichsen et al. 2000; Geldner et al. 2007).

Individual leucine-rich repeats are 20–30 amino-acid long sequence stretches rich in leucine residues and forming an α/β hairpin structure (Kajava 1998) (Fig. 9.3a). Several repeats stack together to form a solenoid, which in the case of bacterial and human LRR proteins, such as ribonuclease inhibitor and Toll-like receptor 3, is reminiscent of a horseshoe (Kobe and Deisenhofer 1993; Choe et al. 2005)

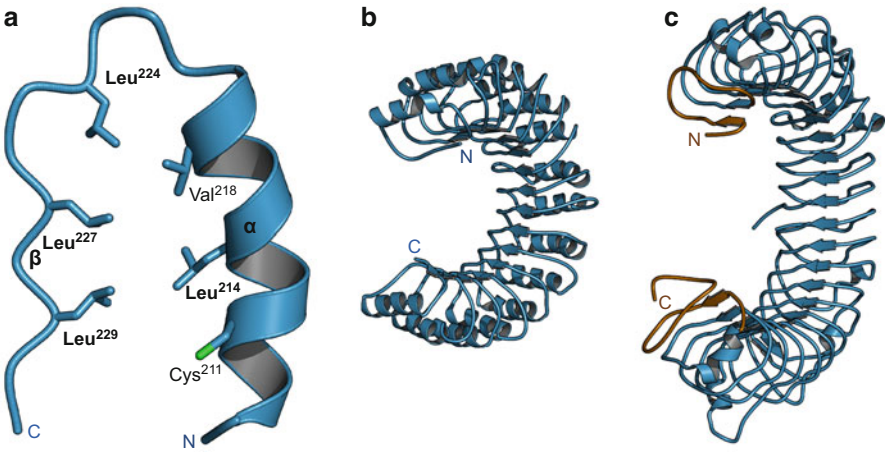


Fig. 9.3 Structural features of animal and bacterial LRR domains. (**a**) Ribbon diagram of a canonical LRR (ribonuclease inhibitor residues 207–231, PDB-ID 1DFJ) featuring an α-helix, a β-sheet, and conserved leucine residues forming a hydrophobic core (in bonds representation), (**b**) ribbon diagram of the horseshoe-shaped bacterial ribonuclease inhibitor (PDB-ID 1DFJ), and (**c**) human Toll-like receptor 3 (PDB-ID 1ZIW) LRR domains. The N- and C-terminal capping domains are shown in brown

(Fig. 9.3b, c). The inner surface of the horseshoe is formed by a parallel β-sheet, the outer surface is composed of an array of helices, and the hydrophobic core is formed by the conserved leucine residues (Fig. 9.3b, c). At their N- and C-termini, the hydrophobic LRRs are shielded by hydrophilic capping domains (Fig. 9.3c).

The crystal structure of the *Arabidopsis* BRI1 LRR domain unexpectedly revealed a right-handed superhelix composed of 25 individual LRRs with an inner diameter of ~30 Å and an out diameter of ~60 Å, and not the canonical horseshoe structure (Hothorn et al. 2011; She et al. 2011) (Fig. 9.4a). The helix completes one entire turn, with a rise of about ~70 Å. Crystal structures of different plant LRR domains now reveal that many if not all plant LRRs form helical or highly twisted assemblies (See Chap. 3). The twist is generated by the presence of a non-canonical, second β-sheet, which is oriented perpendicular to the central β-sheet lining the inner side of the LRR solenoid (Fig. 9.4a). This second β-sheet in plant LRR proteins is caused by the Lt/sGxIP consensus sequence of the plant-specific LRR subfamily (Kajava 1998). The LRRs of BRI1 are flanked by N- and C-terminal capping domains, which are stabilized by disulfide bonds (Fig. 9.4b). In the weak *bri1-5* mutant allele, Cys69 is mutated to a tyrosine residue, disrupting a disulfide bond in the N-terminal capping domain of BRI1 (Noguchi et al. 1999). This structural alteration of the N-terminal cap de-stabilizes the receptor and affects its secretion to the plasma membrane (Hong et al. 2008). Five additional disulfide bonds map to the LRR core itself, covalently linking consecutive LRR segments (Fig. 9.4b) (Hothorn et al. 2011). In addition, the LRR ectodomains of BRI1 and many but not all plant receptor kinases carry N-glycosylation sites. In the BRI1 structure, large surface areas of the LRR domain are thus masked by carbohydrate (Fig. 9.4c). These carbohydrate structures may be involved in the structural stabilization of the LRR domain but may also control protein–protein interactions (Hothorn et al. 2011).

9.3 The BRI1 Island Domain and the Steroid-Binding Site

While most of the 25 individual LRR motifs in BRI1 are connected via short loops, a large insertion is present between LRRs 21 and 22 (Hothorn et al. 2011; She et al. 2011) (Figs. 9.4a and 9.5a). This 70 amino-acid stretch forms a small "island" domain that folds back onto the LRR core, making extensive polar and apolar interactions with LRRs 13–25 (Fig. 9.5a, b). The island domain contains a central anti-parallel β-sheet sandwiched between the BRI1 LRR domain and a small helix and stabilized by a disulfide bond (Fig. 9.5a). Structures of the BRI1 ectodomain crystallized in the presence of brassinolide, a potent brassinosteroid in *Arabidopsis*, reveal one steroid molecule located in a binding pocket formed by both the LRR core (LRRs 21-25) and by the island domain (Hothorn et al. 2011; She et al. 2011) (Fig. 9.5b, c). Comparing the steroid-bound structure with the structures of "apo" BRI1 indicates that the island domain is rather mobile and flexible in the absence of ligand. Steroid binding induces a conformational rearrangement and fixing of the island domain, making BRI1 competent to engage in protein–protein interactions (Hothorn et al. 2011).

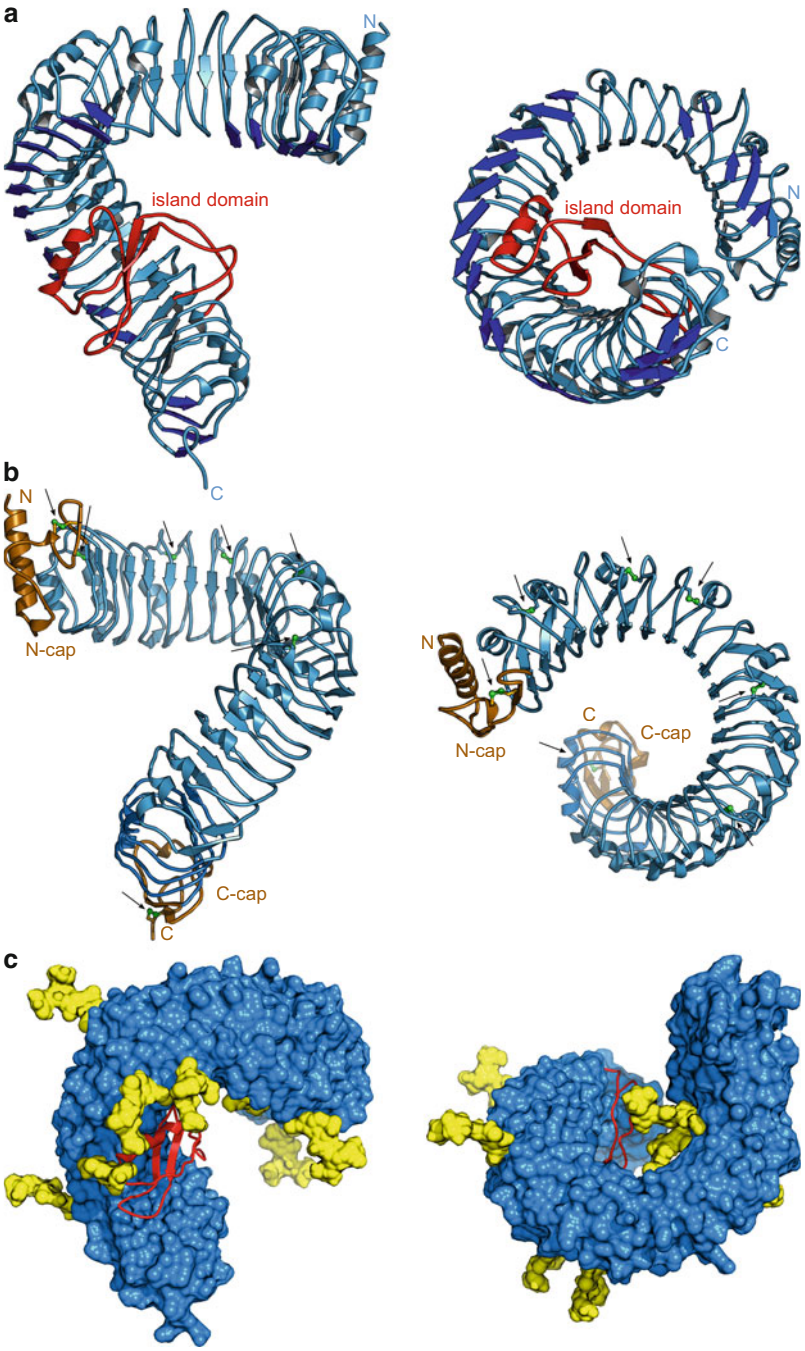


Fig. 9.4 The BRI1 extracellular domain folds into a spiral-shaped solenoid. (**a**) A plant-specific 2nd β-sheet forces the BRI1 LRR domain (PDB-ID 3RIZ, in light-blue) into a spiral shape. The *island domain* (in red) inserts between LRR 21 and 22 and maps to the inside of the helix. The plant specific β-sheet is highlighted in dark blue. Side view on the left, top view on the right. (**b**) The

The steroid-binding site provides a ~550 Å^2 hormone-receptor interface formed by BRI1 LRRs 23–25. This hydrophobic surface is in contact with the A–D rings of brassinolide (Fig. 9.5b) (Hothorn et al. 2011; She et al. 2011). The alkyl chain of the hormone is bound in a small pocket that is formed by residues originating from the LRRs 21 and 22 and from the island domain (Fig. 9.5c). Polar main-chain and side-chain interactions with the 22, 23-diol moiety of brassinolide originate from the island domain (Fig. 9.5c). There are few specific interactions between BRI1 and the B-ring lactone of brassinolide, rationalizing why brassinosteroids carrying modifications of their B-ring, such as castasterone (Fig. 9.1), can be sensed by the receptor (Wang et al. 2001; Back and Pharis 2003; Hothorn et al. 2011) (Fig. 9.5b, c). However, large parts of the steroid hormone, including the crucial 2α,3α-diol moiety are not in contact with BRI1 (Back and Pharis 2003; Hothorn et al. 2011; She et al. 2011). A similar mode of brassinolide binding has been reported for the BRI1 homologue BRL1 (She et al. 2013).

9.4 Receptor Activation of BRI1 Requires Shape-Complementary Co-receptor Kinases

Based on the sequence similarities between the LRR ectodomains of BRI1 and animal Toll-like receptors, ligand-induced homo-dimerization had been proposed to mediate BRI1 receptor activation (Wang et al. 2005b). Indeed, homo-oligomers of BRI1 have been observed in planta, however they appear to be constitutive rather than induced by brassinosteroid binding (Wang et al. 2005b). The purified BRI1 ectodomain behaves as a monomer in solution, and in contrast to Toll-like receptors (Leonard et al. 2008; Liu et al. 2008) shows no tendency to oligomerize in the presence of the steroid ligand (Hothorn et al. 2011; She et al. 2011; Bojar et al. 2014). This finding, together with the observation that the steroid-binding site maps to the inner face of the BRI1 helix, suggested that BRI1 activation requires interaction with a shape-complementary helper protein (Hothorn et al. 2011). A BRI1 co-receptor candidate, BAK1, had already been identified using forward genetics and protein–protein interaction screens (Li et al. 2002; Nam and Li 2002). BAK1 loss-of-function mutants display mild brassinosteroid-insensitive phenotypes (Li and Nam 2002; Li et al. 2002). It was then noted that BAK1 acts redundantly with other SERK family LRR receptor kinases (Nam and Li 2002; Karlova et al. 2006; He et al. 2007). Importantly, *serk1 serk3* (*bak1*) *serk4* triple knock-out mutants phenocopy *bri1* null mutants, suggesting that BRI1 and SERKs are both essential to sense brassinosteroids at the plasma membrane (Gou et al. 2012).

Fig. 9.4 (continued) BRI1 LRR domain is shielded by N- and C-terminal capping domains (in brown) and stabilized by disulfide bridges (in green, highlighted by arrows). The island domain has been omitted for clarity. (**c**) Multiple glycosylations mask large areas of the receptor surface (LRR domain in blue and carbohydrates in yellow, in surface representation, island domain in red and as ribbon diagram)

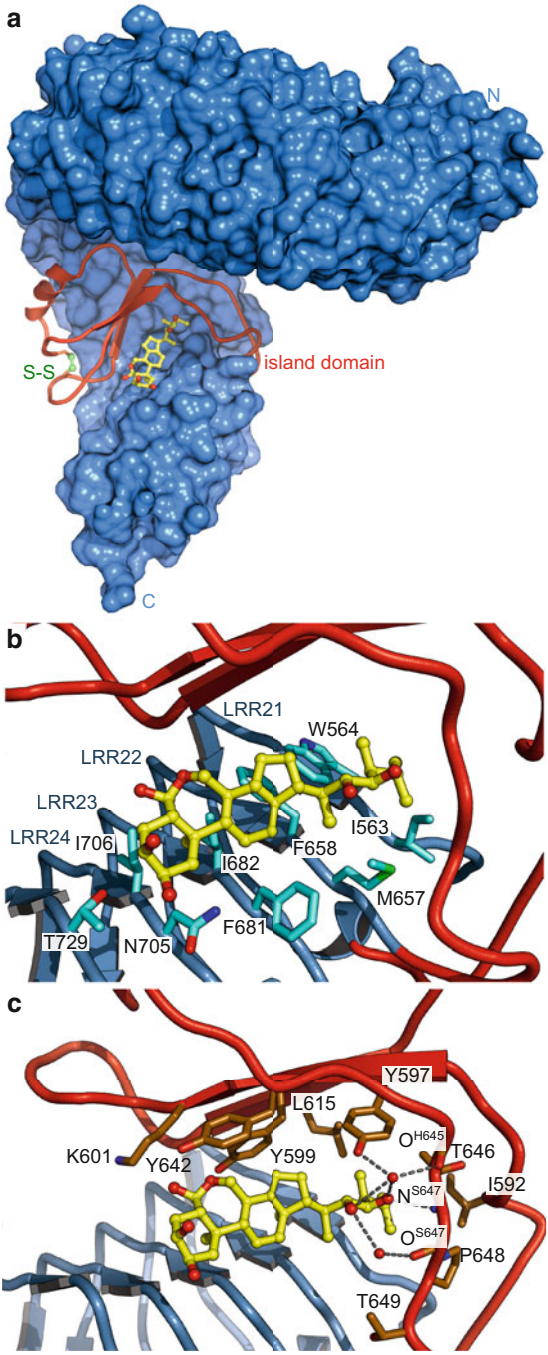


Fig. 9.5 Brassinosteroids bind to LRR domain of BRI1. (**a**) Binding of a brassinolide molecule to the LRR (in blue, surface representation) and *island domain* (in red, ribbon diagram) of BRI1 (PDB-ID 3RIZ). Brassinolide is shown in yellow (in bonds representation). (**b**) A hydrophobic surface formed by LRRs 23–25 interacts with the A-D rings of the steroid (colors as in **a**, BRI1 as

SERK proteins share their overall architecture with BRI1. The crystal structure of SERK1 revealed a short LRR ectodomain with 5 repeats (Santiago et al. 2013). In vitro, the LRR domains of SERK1 or BAK1 form tight heterodimers with the BRI1 ectodomain, but only in the presence of brassinosteroids (Santiago et al. 2013; Sun et al. 2013; Bojar et al. 2014). Crystal structures of BRI1-brassinolide-SERK1 (Santiago et al. 2013) and BRI1-brassinolide-BAK1 (Sun et al. 2013) complexes revealed that SERK co-receptors directly bind to the BRI1 LRR domain, to the island domain, and importantly to the brassinolide itself (Fig. 9.6a). The SERK1/BAK1 N-terminal capping domain completes the steroid-binding site, with the hormone acting as a "molecular glue," which promotes the association between receptor and co-receptor (Fig. 9.6a) (Santiago et al. 2013; Sun et al. 2013).

The effects of several genetic loss- and gain-of-function alleles in BRI1 and in BAK1 can be rationalized in light of the complex structures: Mutation of Gly644 in BRI1 into aspartate causes the loss-of-function phenotype of *bri1-6* plants (Fig. 9.6b) (Noguchi et al. 1999). The affected glycine residue is located in the center of the BRI1 island domain and may be important for both brassinosteroid binding and for proper recruitment of the co-receptor (Hothorn et al. 2011; Santiago et al. 2013; Hohmann et al. 2018). The neighboring Gly643 in the island domain is a glutamate residue in the BRI1 gain-of-function mutant *sud1* (Belkhadir et al. 2012). The BRI1 *sud1* mutant protein has been crystallized, and its structure revealed that Glu643 stabilizes the island domain. This possibly promotes the binding of steroid ligands as well as recruitment of the co-receptor (Fig. 9.6c) (Santiago et al. 2013). The semi-dwarf mutant *093AR* in barley maps to the *bri1* locus, replacing Thr573 with a lysine residue (Gruszka et al. 2011). The corresponding Thr649 in *Arabidopsis* BRI1 again maps to the island domain of the receptor, highlighting its important role in steroid sensing and co-receptor recruitment (Fig. 9.6b) (Hothorn et al. 2011; Santiago et al. 2013). Thr750, which is located in LRR 25 outside the steroid-binding pocket, is found mutated into an isoleucine in the strong loss-of-function mutant *bri1-102* (Fig. 9.6b) (Friedrichsen et al. 2000). Consistently, this mutant does not interfere with steroid sensing itself, but rather inhibits binding of the SERK co-receptor (Wang et al. 2001; Santiago et al. 2013). The *elg* (elongated) mutant of BAK1 causes a gain-of-function brassinosteroid signaling phenotype (Jaillais et al. 2011a). Asn122 in the LRR ectodomain of BAK1 is replaced by an asparagine residue in *elg* mutant plants, again supporting an essential role for SERK co-receptors in brassinosteroid sensing and signaling (Fig. 9.6b) (Jaillais et al. 2011a). The mutation has been recapitulated in rice SERK2 and the mutant protein has been crystallized (McAndrew et al. 2014). In this structure, Asn122 disrupts the formation of a nearby salt-bridge, but it remains to be understood how this structural

Fig. 9.5 (continued) ribbon diagram, relevant residues as well as brassinolide in bonds representation). (**c**) A pocket formed by the island domain and LRRs 21 and 22 binds the steroid's alkyl chain. The island domain interacts strongly with the 22,23-diol moiety of brassinolide (colors as in **b**, important island domain residues in bonds representation, dotted lines represent interactions, red spheres depict important water molecules)

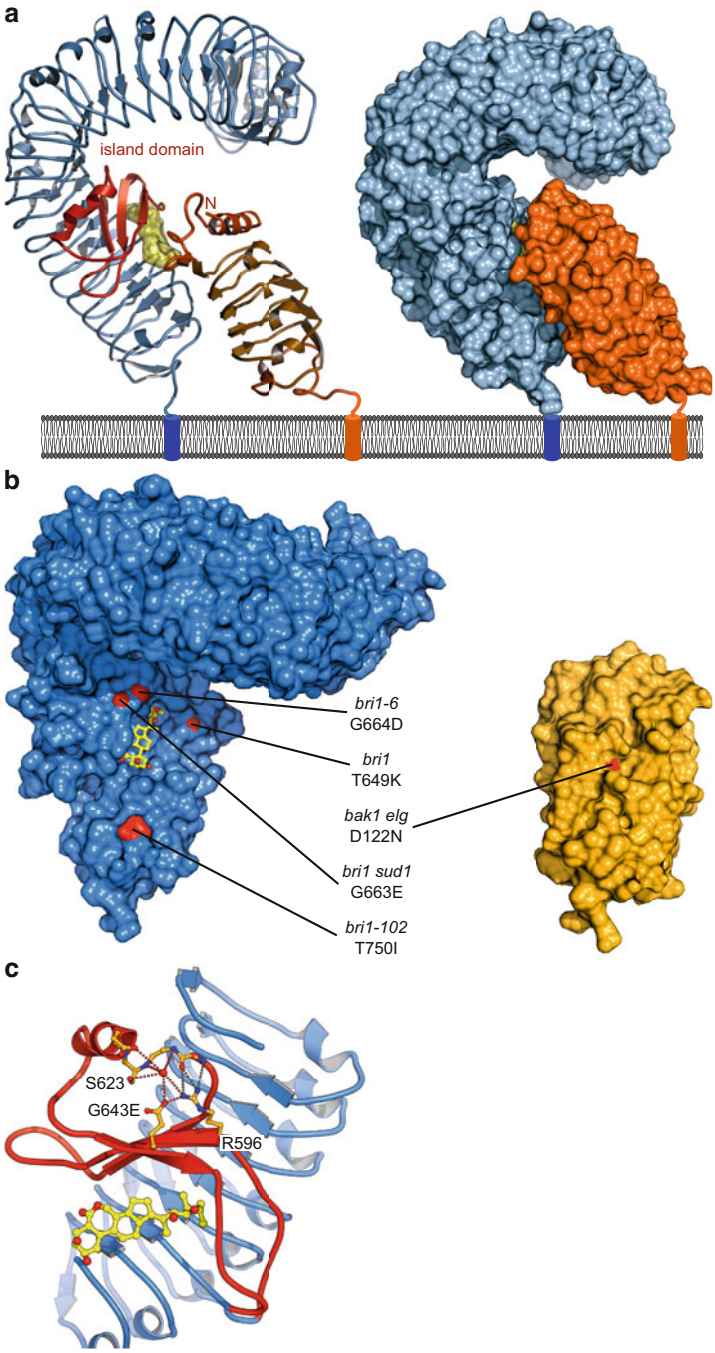


Fig. 9.6 Receptor activation of BRI1 involves a shape-complementary co-receptor. (**a**) Overview of the ternary BRI1-brassinolide-SERK1 complex (PDB-ID 4LSX, left ribbon diagrams, right panel surface views). The BRI1 LRR domain is shown in blue, the *island domain* in red, the SERK1 LRR domain in orange, and brassinolide highlighted in yellow. (**b**) Known gain- and loss-of-function

change affects brassinosteroid signaling (Jaillais et al. 2011a; McAndrew et al. 2014). Taken together, several genetic alleles in BRI1 and in SERKs highlight the importance of their LRR ectodomains for steroid hormone sensing and brassinosteroid receptor activation.

9.5 The Kinase Domains of BRI1 and SERKs Can Activate Each Other in the Cytosol

Formation of a BRI1-brassinosteroid-SERK signaling complex at the cell surface is not sufficient to trigger the cytoplasmic side of the brassinosteroid signaling pathway (Fig. 9.2). It was noted early on that the kinase domains of receptor and co-receptor are being phosphorylated in response to brassinosteroid sensing (Wang et al. 2005b; Wang et al. 2008) and that the isolated cytoplasmic domains of BRI1 and BAK1 can interact in vitro (Li et al. 2002; Nam and Li 2002).

The cytoplasmic domains of BRI1 and BAK1 are each composed of a catalytic kinase core, an N-terminal juxta-membrane region, and a C-terminal tail (Fig. 9.2). Current working models suggest that in the absence of brassinosteroids, the kinase domains of BRI1 and of SERKs are kept in a basal state and are under negative regulation by protein phosphatases (Wang et al. 2016). In its basal state, the kinase domain of BRI1 may also bind the BRI1 inhibitor protein BKI1 (Wang and Chory 2006; Jaillais et al. 2011b). Upon brassinosteroid sensing, the extracellular LRR domains of BRI1 and an SERK interact at the cell surface (Jaillais et al. 2011a; Santiago et al. 2013; Sun et al. 2013). The resulting, ligand-induced heterodimeric signaling complex brings the trans-membrane helices of the receptor and co-receptor in close proximity (Fig. 9.6a) (Santiago et al. 2013; Sun et al. 2013). This, in turn, allows the kinase domains of BRI1 and SERKs to interact (Wang et al. 2008). BRI1 and SERKs then trans-phosphorylate and activate each other, rendering the BRI1 kinase domain competent to phosphorylate its substrates, including BSKs, CDG1, and BKI1 (Fig. 9.2) (Wang and Chory 2006; Tang et al. 2008; Jaillais et al. 2011b; Kim et al. 2011).

While the mechanistic details of BRI1 receptor activation remain to be resolved, crystal structures of the isolated BRI1 and BAK1 kinase domains have been reported (Yan et al. 2012; Bojar et al. 2014). These structures revealed that plant receptor kinases share significant homology with the animal IRAK family of kinases (Fig. 9.7a, b). Both plant receptor kinases (including BRI1 and BAK1) and animal IRAK kinases are dual-specificity kinases, able to auto- and trans-phosphorylate on

Fig. 9.6 (continued) alleles map to the receptor – co-receptor interaction surface (BRI1 LRR domain in blue, in surface representation, island domain omitted, brassinolide in yellow and in bonds representation, SERK1 on the right in yellow). (**c**) The glutamate residue 643 in BRI1^{sud1} stabilizes the island domain (PDB-ID 4LSA, ribbon diagram, colors as in **a**, interacting residues in orange and in bonds representation, red sphere depicts a water molecule)

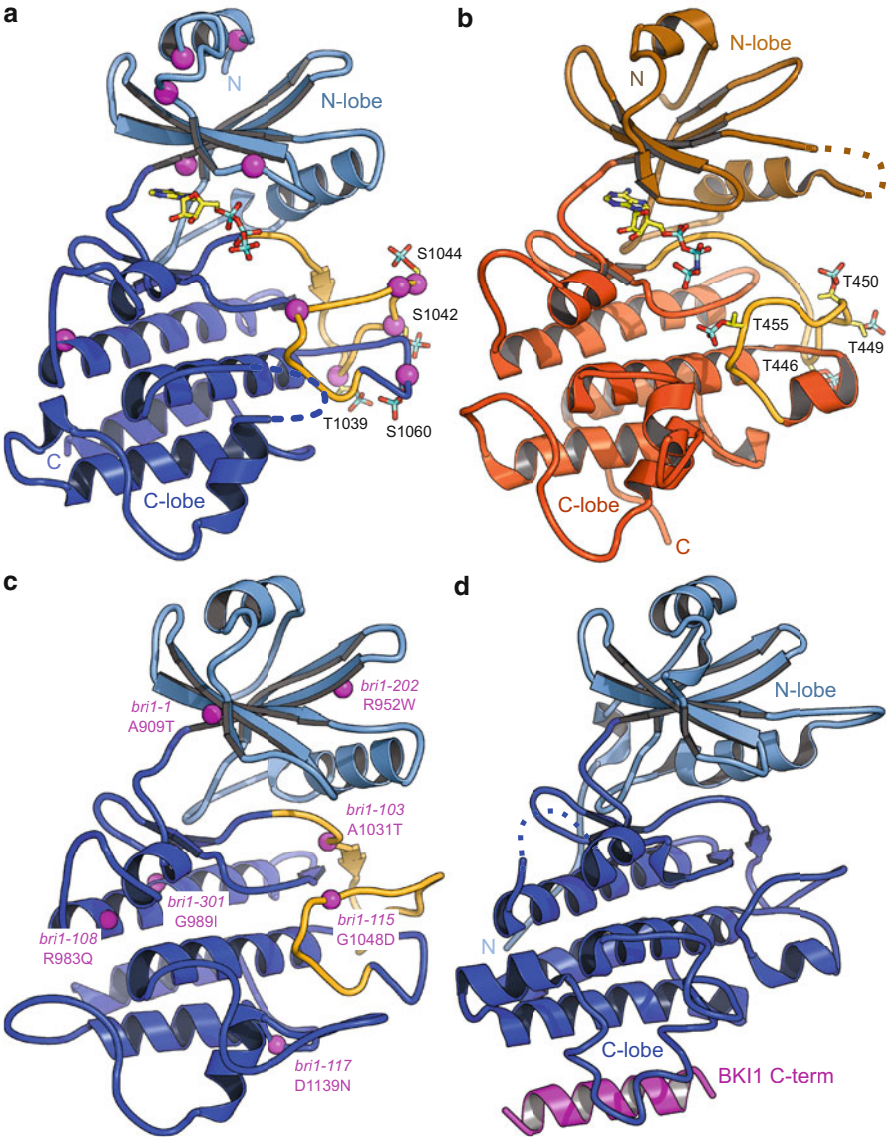


Fig. 9.7 Structures of the BRI1 and BAK1 kinase domains. (**a**) A ribbon diagram of the BRI1 kinase domain reveals a canonical kinase fold with *N-lobe* (light blue), *C-lobe* (dark blue), and activation loop (yellow). ATP is bound to the active site (in bonds representation), phosphorylation sites are highlighted with magenta spheres (PDB-ID 5LPY). (**b**) Crystal structure of the BAK1 kinase domain (features as in **a**, but *N-lobe* in brown and *C-lobe* in orange, PDB-ID 3UIM). (**c**) Known loss-of-function alleles (highlighted with spheres in magenta) map to the BRI1 kinase domain (structure as in **a**, nucleotide omitted). (**d**) Binding of the C-terminal helix of BKI1 to the C-lobe of the BRI1 kinase domain competes with BAK1 interaction or substrate binding (BKI1 in magenta as ribbon diagram, BRI1 kinase as in **a**, PDB-ID 4OH4)

Ser/Thr and on Tyr residues (Wang et al. 2006; Oh et al. 2009; Jaillais et al. 2011b). The BRI1 and BAK1 kinase domain structures feature the canonical N- and C-terminal kinase lobes, with parts of the juxtamembrane segments and the C-terminal tails being disordered (Fig. 9.7a, b) (Yan et al. 2012; Bojar et al. 2014). The BRI1 activation loop, which determines which substrates can bind, contains structural motifs normally found in either Ser/Thr or tyrosine kinases (Bojar et al. 2014). How BRI1 can both phosphorylate Ser/Thr and tyrosine substrates is poorly understood. Both BRI1 and BAK1 carry a complex pattern of phosphorylation sites, some of which are recapitulated in their crystal structures (Fig. 9.7a, b) (Oh et al. 2000; Wang et al. 2005a, 2005b, 2008; Oh et al. 2009; Bajwa et al. 2013). A large number of weak and strong *bri1* loss-of-function mutations map to the BRI1 kinase domain, likely interfering with the activity of the enzyme (Fig. 9.7c) (Clouse et al. 1996; Li and Chory 1997; Friedrichsen et al. 2000; Xu et al. 2008).

The interaction surface between the BRI1 and BAK1 kinase domains remains to be identified, but it is interesting to note that the inhibitor protein BKI1 disrupts the cytosolic interaction between the receptor and its SERK co-receptor (Jaillais et al. 2011b). BKI1 contains a conserved helical motif at its very C-terminus, which binds to the C-lobe of the BRI1 kinase (Fig. 9.7d) (Jaillais et al. 2011b; Wang et al. 2014). This suggests, that, in the absence of the inhibitor BKI1, BRI1 may form an asymmetric hetero-dimer with BAK1 or other SERKs, bringing the two kinase domains in a configuration that allows them to trans-phosphorylate each other (Bojar et al. 2014). BRI1 itself is able to phosphorylate BKI1 on a Tyr residue, triggering the release of the inhibitor protein from the plasma membrane into the cytosol (Fig. 9.2) (Jaillais et al. 2011b).

Taken together, crystallographic, biochemical, and genetic evidence supports a model, in which the plant steroid receptor BRI1 and a co-receptor kinase together sense brassinosteroids. The ligand induced receptor – co-receptor complex formation at the cell surface-brings the cytoplasmic kinase domains of BRI1 and SERK proteins in close proximity, allowing for trans-phosphorylation and subsequent activation of the cytosolic brassinosteroid signaling pathway. The use of spiral-shaped LRR ligand-binding domains and shape-complementary co-receptors represents a plant-unique signaling paradigm.

Acknowledgments Financial support by the Swiss National Science Foundation (grant number 156920) is gratefully acknowledged.

Bibliography

Back TG, Pharis RP (2003) Structure-activity studies of brassinosteroids and the search for novel analogues and mimetics with improved bioactivity. J Plant Growth Regul 22:350–361. https://doi.org/10.1007/s00344-003-0057-0

Bajguz A (2007) Metabolism of brassinosteroids in plants. Plant Physiol Biochem PPB 45:95–107. https://doi.org/10.1016/j.plaphy.2007.01.002

Bajwa VS, Wang X, Blackburn RK et al (2013) Identification and functional analysis of tomato BRI1 and BAK1 receptor kinase phosphorylation sites. Plant Physiol 163:30–42. https://doi.org/10.1104/pp.113.221465

Belkhadir Y, Jaillais Y, Epple P et al (2012) Brassinosteroids modulate the efficiency of plant immune responses to microbe-associated molecular patterns. Proc Natl Acad Sci U S A 109:297–302. https://doi.org/10.1073/pnas.1112840108

Bojar D, Martinez J, Santiago J et al (2014) Crystal structures of the phosphorylated BRI1 kinase domain and implications for brassinosteroid signal initiation. Plant J 78:31–43. https://doi.org/10.1111/tpj.12445

Brandt B, Hothorn M (2016) SERK co-receptor kinases. Curr Biol CB 26:R225–R226. https://doi.org/10.1016/j.cub.2015.12.014

Choe J, Kelker MS, Wilson IA (2005) Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. Science 309:581–585. https://doi.org/10.1126/science.1115253

Clouse SD, Langford M, McMorris TC (1996) A brassinosteroid-insensitive mutant in Arabidopsis thaliana exhibits multiple defects in growth and development. Plant Physiol 111:671–678. https://doi.org/10.1104/pp.111.3.671

Friedrichsen DM, Joazeiro CAP, Li J et al (2000) Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. Plant Physiol 123:1247–1256. https://doi.org/10.1104/pp.123.4.1247

Gampala SS, Kim T-W, He J-X et al (2007) An essential role for 14-3-3 proteins in brassinosteroid signal transduction in Arabidopsis. Dev Cell 13:177–189. https://doi.org/10.1016/j.devcel.2007.06.009

Geldner N, Hyman DL, Wang X et al (2007) Endosomal signaling of plant steroid receptor kinase BRI1. Genes Dev 21:1598–1602. https://doi.org/10.1101/gad.1561307

Gou X, Yin H, He K et al (2012) Genetic evidence for an indispensable role of somatic embryogenesis receptor kinases in brassinosteroid signaling. PLoS Genet 8:e1002452. https://doi.org/10.1371/journal.pgen.1002452

Grove MD, Spencer GF, Rohwedder WK et al (1979) Brassinolide, a plant growth-promoting steroid isolated from Brassica napus pollen. Nature 281:216–217. https://doi.org/10.1038/281216a0

Gruszka D, Szarejko I, Maluszynski M (2011) New allele of HvBRI1 gene encoding brassinosteroid receptor in barley. J Appl Genet. 52(3):257–68. https://doi.org/10.1007/s13353-011-0031-7

He Z, Wang ZY, Li J et al (2000) Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. Science 288:2360–2363

He K, Gou X, Yuan T et al (2007) BAK1 and BKK1 regulate brassinosteroid-dependent growth and brassinosteroid-independent cell-death pathways. Curr Biol CB 17:1109–1115. https://doi.org/10.1016/j.cub.2007.05.036

Hohmann U, Santiago J, Nicolet J, Olsson V, Spiga FM, Ludwig Hothorn A, Butenko MA, Hothorn M (2018) Mechanistic basis for the activation of plant membrane receptor kinases by SERK-family coreceptors. Proc Natl Acad Sci 115(13):3488–3493

Hong Z, Jin H, Tzfira T, Li J (2008) Multiple mechanism-mediated retention of a defective brassinosteroid receptor in the endoplasmic reticulum of Arabidopsis. Plant Cell 20:3418–3429. https://doi.org/10.1105/tpc.108.061879

Hothorn M, Belkhadir Y, Dreux M et al (2011) Structural basis of steroid hormone perception by the receptor kinase BRI1. Nature 474:467–471. https://doi.org/10.1038/nature10153

Jaillais Y, Belkhadir Y, Balsemão-Pires E et al (2011a) Extracellular leucine-rich repeats as a platform for receptor/coreceptor complex formation. Proc Natl Acad Sci U S A 108:8503–8507. https://doi.org/10.1073/pnas.1103556108

Jaillais Y, Hothorn M, Belkhadir Y et al (2011b) Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor. Genes Dev 25:232–237. https://doi.org/10.1101/gad.2001911

Kajava AV (1998) Structural diversity of leucine-rich repeat proteins. J Mol Biol 277:519–527. https://doi.org/10.1006/jmbi.1998.1643

Karlova R, Boeren S, Russinova E et al (2006) The Arabidopsis somatic embryogenesis receptor-like kinase1 protein complex includes brassinosteroid-insensitive1. Plant Cell 18:626–638. https://doi.org/10.1105/tpc.105.039412

Kim T-W, Guan S, Burlingame AL, Wang Z-Y (2011) The CDG1 kinase mediates brassinosteroid signal transduction from BRI1 receptor kinase to BSU1 phosphatase and GSK3-like kinase BIN2. Mol Cell 43:561–571. https://doi.org/10.1016/j.molcel.2011.05.037

Kinoshita T, Caño-Delgado A, Seto H et al (2005) Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. Nature 433:167–171. https://doi.org/10.1038/nature03227

Kobe B, Deisenhofer J (1993) Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature 366:751–756. https://doi.org/10.1038/366751a0

Leonard JN, Ghirlando R, Askins J et al (2008) The TLR3 signaling complex forms by cooperative receptor dimerization. Proc Natl Acad Sci U S A 105:258–263. https://doi.org/10.1073/pnas.0710779105

Li J, Chory J (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell 90:929–938

Li J, Nam KH (2002) Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. Science 295:1299–1301. https://doi.org/10.1126/science.1065769

Li J, Nam KH, Vafeados D, Chory J (2001) BIN2, a new brassinosteroid-insensitive locus in Arabidopsis. Plant Physiol 127:14–22

Li J, Wen J, Lease KA et al (2002) BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell 110:213–222

Liu L, Botos I, Wang Y et al (2008) Structural basis of toll-like receptor 3 signaling with double-stranded RNA. Science 320:379–381. https://doi.org/10.1126/science.1155406

McAndrew R, Pruitt RN, Kamita SG et al (2014) Structure of the OsSERK2 leucine-rich repeat extracellular domain. Acta Crystallogr D Biol Crystallogr 70:3080–3086. https://doi.org/10.1107/S1399004714021178

Mora-García S, Vert G, Yin Y et al (2004) Nuclear protein phosphatases with Kelch-repeat domains modulate the response to brassinosteroids in Arabidopsis. Genes Dev 18:448–460. https://doi.org/10.1101/gad.1174204

Nam KH, Li J (2002) BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell 110:203–212

Noguchi T, Fujioka S, Choe S et al (1999) Brassinosteroid-insensitive dwarf mutants of Arabidopsis accumulate brassinosteroids. Plant Physiol 121:743–752

Oh MH, Ray WK, Huber SC et al (2000) Recombinant brassinosteroid insensitive 1 receptor-like kinase autophosphorylates on serine and threonine residues and phosphorylates a conserved peptide motif in vitro. Plant Physiol 124:751–766

Oh M-H, Wang X, Kota U et al (2009) Tyrosine phosphorylation of the BRI1 receptor kinase emerges as a component of brassinosteroid signaling in Arabidopsis. Proc Natl Acad Sci U S A 106:658–663. https://doi.org/10.1073/pnas.0810249106

Santiago J, Henzler C, Hothorn M (2013) Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. Science 341:889–892. https://doi.org/10.1126/science.1242468

She J, Han Z, Kim T-W et al (2011) Structural insight into brassinosteroid perception by BRI1. Nature 474:472–476. https://doi.org/10.1038/nature10178

She J, Han Z, Zhou B, Chai J (2013) Structural basis for differential recognition of brassinolide by its receptors. Protein Cell 4:475–482. https://doi.org/10.1007/s13238-013-3027-8

Shiu SH, Bleecker AB (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci U S A 98:10763–10768. https://doi.org/10.1073/pnas.181141598

Sun Y, Han Z, Tang J et al (2013) Structure reveals that BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. Cell Res 23:1326–1329. https://doi.org/10.1038/cr.2013.131

Tang W, Kim T-W, Oses-Prieto JA et al (2008) BSKs mediate signal transduction from the receptor kinase BRI1 in Arabidopsis. Science 321:557–560. https://doi.org/10.1126/science.1156973

Vert G, Chory J (2006) Downstream nuclear events in brassinosteroid signalling. Nature 441:96–100. https://doi.org/10.1038/nature04681

Wang X, Chory J (2006) Brassinosteroids regulate dissociation of BKI1, a negative regulator of BRI1 signaling, from the plasma membrane. Science 313:1118–1122. https://doi.org/10.1126/science.1127593

Wang ZY, Seto H, Fujioka S et al (2001) BRI1 is a critical component of a plasma-membrane receptor for plant steroids. Nature 410:380–383. https://doi.org/10.1038/35066597

Wang ZY, Nakano T, Gendron J et al (2002) Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. Dev Cell 2:505–513

Wang X, Goshe MB, Soderblom EJ et al (2005a) Identification and functional analysis of in vivo phosphorylation sites of the Arabidopsis BRASSINOSTEROID-INSENSITIVE1 receptor kinase. Plant Cell 17:1685–1703. https://doi.org/10.1105/tpc.105.031393

Wang X, Li X, Meisenhelder J et al (2005b) Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. Dev Cell 8:855–865. https://doi.org/10.1016/j.devcel.2005.05.001

Wang Z, Liu J, Sudom A et al (2006) Crystal structures of IRAK-4 kinase in complex with inhibitors: a serine/threonine kinase with tyrosine as a gatekeeper. Struct Lond Engl 1993 14:1835–1844. https://doi.org/10.1016/j.str.2006.11.001

Wang X, Kota U, He K et al (2008) Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. Dev Cell 15:220–235. https://doi.org/10.1016/j.devcel.2008.06.011

Wang J, Jiang J, Wang J et al (2014) Structural insights into the negative regulation of BRI1 signaling by BRI1-interacting protein BKI1. Cell Res 24:1328–1341. https://doi.org/10.1038/cr.2014.132

Wang R, Liu M, Yuan M et al (2016) The brassinosteroid-activated BRI1 receptor kinase is switched off by dephosphorylation mediated by cytoplasm-localized PP2A B' subunits. Mol Plant 9:148–157. https://doi.org/10.1016/j.molp.2015.10.007

Xu W, Huang J, Li B et al (2008) Is kinase activity essential for biological functions of BRI1? Cell Res 18:472–478. https://doi.org/10.1038/cr.2008.36

Yan L, Ma Y, Liu D et al (2012) Structural basis for the impact of phosphorylation on the activation of plant receptor-like kinase BAK1. Cell Res 22:1304–1308. https://doi.org/10.1038/cr.2012.74

Yin Y, Wang ZY, Mora-Garcia S et al (2002) BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. Cell 109:181–191

Zhao J, Peng P, Schmitz RJ et al (2002) Two putative BIN2 substrates are nuclear components of brassinosteroid signaling. Plant Physiol 130:1221–1229. https://doi.org/10.1104/pp.102.010918

Chapter 10
Cytokinin and Ethylene Signaling

Blanka Pekarova, Agnieszka Szmitkowska, Josef Houser, Michaela Wimmerova, and Jan Hejátko

10.1 Introduction to Cytokinins and Ethylene

Cytokinins and ethylene directly regulate and/or participate in complex hormonal interactions controlling many cellular and developmental processes in plants. Cytokinins control cell division and differentiation, gametophyte formation, vascular tissue development, induce the formation of shoots, regulate axillary bud growth and apical dominance, delay ageing in leaves, and support chloroplast development. Cytokinins also regulate the uptake of nutrients and play a role in clock-related responses, interactions with pathogens, light and abiotic stresses (Kieber and Schaller 2014). Gaseous hormone ethylene, one of the first plant hormones discovered, regulates seed germination, root hair initiation, gravitropic growth, leaf and petal abscission, flower fading, and fruit ripening and acts as a stress hormone during most biotic and abiotic stress conditions (Abeles et al. 1992; Bleecker and Kende 2000).

Since a discovery of the first cytokinin, kinetin, by Miller and Skoog (Miller et al. 1955), many natural and synthetic compounds have been classified as cytokinins. Chemically, natural cytokinins are adenine derivatives with a short aliphatic (isoprenoid) or aromatic side chain at the N^6-position [Fig. 10.1; (Mok and Mok 2001)]. N^6-(2-isopentenyl)adenine (iP), *cis*-zeatin (cZ), *trans*-zeatin (tZ), and dihydrozeatin (DHZ) are the most abundant isoprenoid cytokinin types found in higher plants. tZ is an active cytokinin in all plant species. In spite of cZ occurring ubiquitously in the plant kingdom, sometimes even in the higher abundance than tZ (Gajdosova et al. 2011), it shows only limited cytokinin activity (Heyl et al. 2012; Romanov et al. 2006; Spichal et al. 2004). One of the most highly active natural aromatic cytokinin is N^6-benzyladenine with its hydroxy- (topolins) and methoxy-derivatives. All naturally

B. Pekarova · A. Szmitkowska · J. Houser · M. Wimmerova · J. Hejátko (✉)
Central European Institute of Technology, Masaryk University, Brno, Czech Republic
e-mail: hejatko@sci.muni.cz

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_10

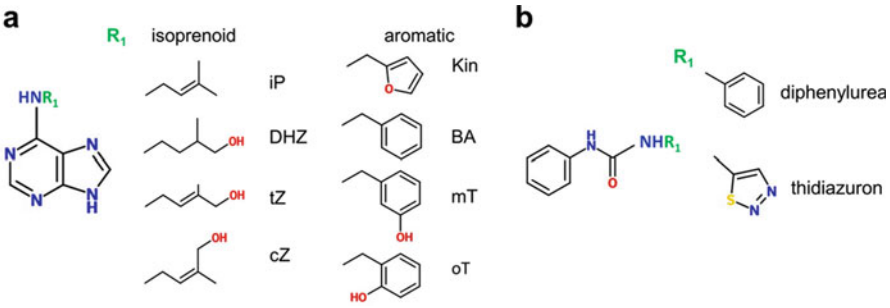


Fig. 10.1 Chemical structure of cytokinins. (**a**) Adenine-type cytokinins. R_1, side chain; iP, N^6-(2-isopentenyl)adenine; DHZ, dihydrozeatin; tZ, *trans*-zeatin; cZ, *cis*-zeatin; Kin, kinetin; BA, N^6-benzyladenine; mT, *meta*-topolin; oT, *ortho*-topolin, (**b**) phenylurea-type cytokinins

occurring cytokinins are present as free bases and corresponding conjugates such as cytokinin nucleosides, in which the cytokinin base is attached via its N^9 to the ribose moiety or cytokinin nucleotides (nucleosides monophosphates). Cytokinins can also be conjugated with sugars (O- and N-glycosides), most commonly with glucose or xylose, or amino acid residues such as alanine. Nevertheless, all these conjugates seem to be inactive cytokinin forms (Lomin et al. 2015) and a plethora of cytokinin activating and inactivating enzymes was described (Sakakibara 2006). Other compounds with cytokinin activity are derivatives of phenylurea, for example, diphenylurea and thidiazuron. For additional information on cytokinin biosynthesis, modification, and signaling see several recent reviews (Spichal 2012; Kieber and Schaller 2014; Zurcher and Muller 2016).

During the past two decades, substantial progress in elucidating the molecular and biochemical mechanism of cytokinin and ethylene action in plants was achieved. With the completion of the *Arabidopsis thaliana* genome sequence (Arabidopsis Genome 2000), the entire spectrum of genes coding for receptors, downstream intermediate signaling molecules, and transcription factors was identified and functionally characterized. In the following sections, we highlight the recent knowledge of structural characteristics of cytokinin and ethylene signaling proteins. On selected examples, we demonstrate how the knowledge of structure of signal transduction proteins at atomic resolution facilitates our mechanistic understanding of signaling specificity and multiple signal integration, signal transduction, and translation of the signal to specific response.

10.2 Cytokinin Signaling

10.2.1 Overview of Cytokinin Signaling Pathway

The cytokinin signal in plants is perceived and transmitted via what is called histidyl-to-aspartyl (His-to-Asp) multistep phosphorelay (MSP), which appears to have evolved from the bacterial two-component system (Schaller et al. 2011).

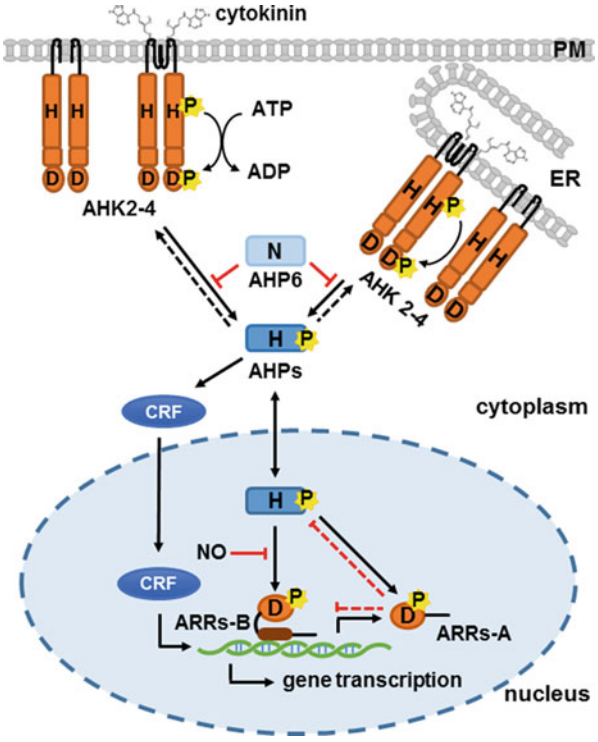


Fig. 10.2 Model of cytokinin signaling in *Arabidopsis thaliana*. Cytokinin signal is perceived by three cytokinin receptors, AHK2, AHK3, and AHK4 localized (dominantly) in the endoplasmic reticulum (ER), but also in the plasma membrane (PM). Upon cytokinin binding, the signal is transduced via multistep phosphorelay through conserved histidine (H) residue of the histidine kinase domain (orange rectangle) and aspartate (D) residue of the receiver domain (orange oval) of cytokinin receptors to small cytosolic AHPs. AHK4 functions as a kinase and a phosphatase in the presence or absence of cytokinins, respectively. AHPs transfer the signal to the nuclear type-B response regulators ARRs-B. Activated ARRs-B induce the expression of effector genes directly or via transcriptional upregulation of CRFs, which translocate to the nucleus in the AHP-dependent way. ARRs-B also upregulate ARRs-A, the cytokinin primary response genes and mediators of the negative feedback regulation (the mechanism of which is still unknown). AHP6 and nitric oxide (NO) are inhibitors of the phosphorelay. Phosphate is shown as a yellow star. Dashed lines indicate regulation by an unclear mechanism or proposed regulatory interactions

MSP-mediated cytokinin signaling (Fig. 10.2) initiates by the recognition of cytokinin molecule via input domain of the sensor histidine kinase (HK). That triggers the kinase activity, allowing autophosphorylation of the conserved His residue in the intracellular HK domain. The phosphate moiety is subsequently carried over to the conserved Asp residue of the C-terminal receiver domain (RD). The first intramolecular phosphotransfer is followed by intermolecular transphosphorylation to the conserved His residue of a small cytoplasmic His-containing phosphotransfer (HPt) protein. Three HKs, known as

ARABIDOPSIS HISTIDINE KINASE (AHK)2, AHK3, and AHK4 (also called CYTOKININ RESISTANT (CRE)1 or WOODENLEG (WOL)), were identified as cytokinin receptors in *Arabidopsis* (Higuchi et al. 2004; Inoue et al. 2001; Nishimura et al. 2004; Ueguchi et al. 2001; Mahonen et al. 2000). Among them, only AHK4 was proven to have a dual function, that is, kinase in the presence while phosphatase in the absence of cytokinins (Mahonen et al. 2006b). That dual regulatory role provides flexibility and rapid control over the pathway. *Arabidopsis* and maize cytokinin receptors are dominantly localized in the endoplasmic reticulum (ER) membrane (Caesar et al. 2011; Lomin et al. 2011; Wulfetange et al. 2011). However, the originally assumed plasma membrane (PM) localization (Kim et al. 2006) also cannot be excluded and recently published evidence suggest functional importance of PM localization of AHK3 (Zurcher et al. 2016).

Arabidopsis HPt family consists of eight proteins (AHP1–8). Analysis of multiple insertion mutants indicated that AHP1–3 and AHP5 function as redundant positive regulators of cytokinin signaling, whereas AHP4 is thought to have dual function, acting in some cases also as a negative regulator (Hutchison and Kieber 2007; Hutchison et al. 2006). AHP6, also called pseudo-HPt, has the phospho-accepting His residue replaced by Asp and attenuates cytokinin signaling via yet unknown mechanisms (Mahonen et al. 2006a). AHP7 and AHP8 were newly identified via detecting evolutionarily conserved elementary protein modules based on network analysis (DECMNA) approach (Dabravolski et al., unpublished) and their role in the *Arabidopsis* development remains to be identified. According to the current concept, AHPs undergo constant bidirectional nucleocytoplasmic partitioning independent of their phosphorylation status (Punwani et al. 2010).

In the nucleus, phosphorylated AHPs activate type-B ARABIDOPSIS RESPONSE REGULATORS (ARRs-B). ARRs-B act as partially redundant, cytokinin-regulated transcription factors that activate the transcription of numerous cytokinin primary response genes via binding to *cis* element in their promoters (Hwang and Sheen 2001). These include also type-A ARRs [*ARRs-A*; (D'Agostino et al. 2000)] or *CYTOKININ RESPONSE FACTORS* [*CRFs*; (Rashotte et al. 2006)]. In turn, ARRs-A function as partially redundant negative regulators of MSP signaling, mediating thus negative feed-back loop (To et al. 2004; To and Kieber 2008). However, besides that, ARRs-A seem to play also a direct regulatory role in the control of cytokinin-regulated plant development (Zhang et al. 2011; Salome et al. 2006). In addition, AHPs also interact with a set of CRFs (Cutcliffe et al. 2011), representing a parallel branch of cytokinin signaling to that mediated by ARRs-B. Numerous excellent reviews in the recent years have focused on individual aspects of the cytokinin perception and signaling, and we refer the reader to these articles for more details (Zurcher and Muller 2016; Gruhn and Heyl 2013; Kieber and Schaller 2014).

10.2.2 Cytokinin Signal Recognition

The ligand-binding properties of *Arabidopsis* cytokinin receptors have been studied intensively using heterologous bacterial and yeast model systems expressing individual receptors (Suzuki et al. 2001a; Yamada et al. 2001; Ueguchi et al. 2001; Inoue et al. 2001). *Escherichia coli*–based direct receptor binding assay (Romanov et al. 2005) demonstrated that AHKs bind cytokinins in nanomolar range, but individual sensor histidine kinases differ in their preference for individual cytokinin types (Spichal et al. 2004; Romanov et al. 2005, 2006; Stolz et al. 2011). Both AHK2 and AHK4 exhibit high affinity to tZ and iP and a considerably lower affinity to DHZ, whereas AHK3 exhibits a higher affinity to DHZ and tZ than to iP (Romanov et al. 2006; Spichal et al. 2004; Stolz et al. 2011). Cytokinin recognition preference was also demonstrated for cytokinin receptors from other plants, for example, *Zea mays* (Lomin et al. 2011) and *Brassica napus* (Kuderova et al. 2015). However, in contrast to the aforementioned *E. coli*–based binding assay, the recently introduced in planta assay relaying on *AHKs* expression in tobacco suggests the ability of cytokinin receptors to bind with high affinity only free bases, but not the cytokinin nucleosides (Lomin et al. 2015).

Cytokinin receptors sense the cytokinin signal via the N-terminal ligand-binding domain so-called *c*yclases/*h*istidine kinases *a*ssociated *s*ensory *e*xtracellular (CHASE), which is common among prokaryotes and lower eukaryotes (Anantharaman and Aravind 2001; Heyl et al. 2007; Mougel and Zhulin 2001; Ueguchi et al. 2001). As cytokinin receptors are predominantly located at the ER membrane, the term "extracellular" is supposed to be excluded from the full receptor designation (Steklov et al. 2013). The presence of CHASE sensory domain distinguishes cytokinin receptors from other sensor HKs in *Arabidopsis*.

The receptor-ligand binding mechanism was elucidated by crystallization of the CHASE domain from the AHK4 receptor in complexes with various cytokinins, such as iP, tZ, tZ riboside, DHZ, N^6-benzyladenine, kinetin, and thidiazuron (Hothorn et al. 2011). Interestingly, the structural differences found among individual complexes were negligible. The overall structure consists of a homodimer with helices α1 and α2 forming the dimerization interface (Fig. 10.3). The N-terminus of the monomeric CHASE domain forms a long helical stalk followed by two Per-Arnt-Sin (PAS)-like domains connected by a helical linker. The C-terminal β-strand of the membrane-proximal PAS-like domain is covalently linked to the N-terminus of the stalk helix via a disulfide bridge, which makes the domain more rigid and compact. Similar tertiary structures of sensory domains were previously identified in bacterial signal transduction proteins, including HKs and chemotaxis receptors (Zhou et al. 2008).

Cytokinin binding site is located on the membrane-distal PAS-like domain. Structural plasticity in the PAS-like domain allows binding of cytokinins with either aliphatic or aromatic side chains without the need of major structural rearrangements (Fig. 10.4a). The lower half of the cytokinin binding pocket consists of the central β-sheet of the membrane-distal PAS-like domain lined by small hydrophobic amino

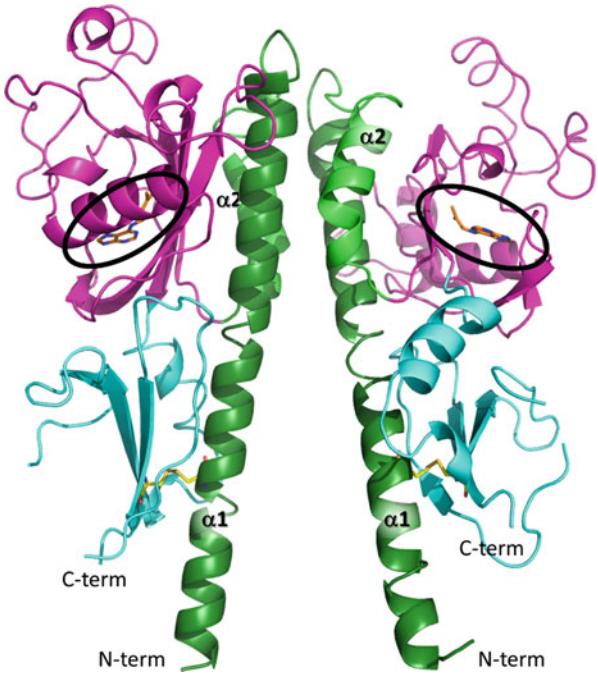


Fig. 10.3 AHK4 CHASE domain dimer (3T4J). Color code: dimerization interface helices α1 (dark green) and α2 (light green), membrane-proximal PAS-like domain (magenta) and membrane-distal PAS-like domain (cyan). Stabilizing disulfide bridges shown as yellow sticks, coordinated ligand N^6-(2-isopentenyl)adenine (iP) shown as orange sticks, cytokinin binding site highlighted by black oval

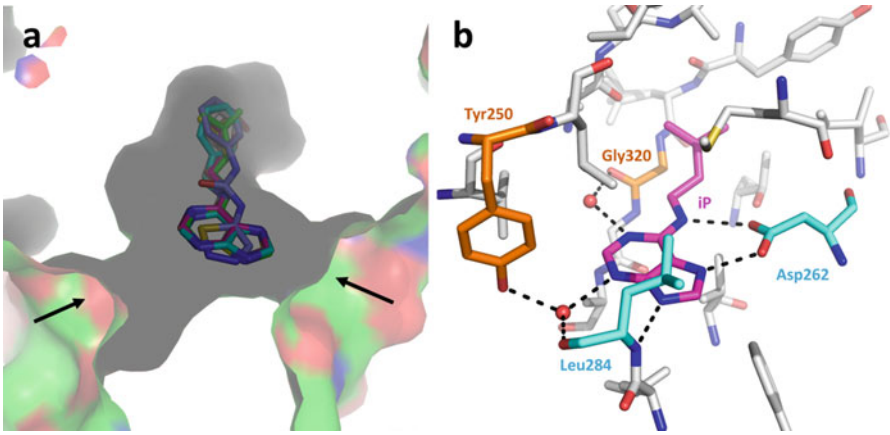


Fig. 10.4 Detail of cytokinin-binding site of AHK4 CHASE domain. (**a**) Cytokinin-recognizing site of AHK4 CHASE domain with individual cytokinins shown as sticks. Possible binding site accessing tunnels marked by arrows. N^6-(2-isopentenyl)adenine (iP) – green, N-benzyl adenine – cyan, *trans*-zeatin (tZ) – magenta, dihydrozeatin (DHZ) – yellow, kinetin – gray, thidiazuron – blue. (**b**) iP in complex with AHK4 CHASE domain. Hydrogen bridging residues shown in cyan, water-mediated only bridging residues shown in orange with water molecules as red spheres

acid residues. Substitutions of these small residues with larger amino acids, that is, the substitution of Thr278 with isoleucine as observed in cytokinin-insensitive *wol* mutation (Mahonen et al. 2000), blocks cytokinin binding, presumably via steric interference. The hydrophobic upper half of the binding pocket is formed by two short β-strands of the distal PAS-like domain. The adenine moiety of bound cytokinin is properly oriented and stabilized in the binding pocket via hydrogen and water-mediated bonds (Fig. 10.4b). Two hydrogen bonds between carboxy group of Asp262 and the adenine ring seem to be critical for receptor function. The total number of amino acids forming the ligand-binding pocket is approximately 20. The remaining residues are involved in hydrophobic interactions with both the adenine ring and, in particular, the side chain of cytokinin (Hothorn et al. 2011). Interestingly, cZ and tZ differ in the way they are recognized by the CHASE domain of AHK4. While tZ forms hydrogen bond with Thr294 residue via the hydroxyl group of the side chain, such bond is missing in cZ, thus explaining the higher affinity of AHK4 for tZ (Romanov et al. 2006). The AHK4 structure-based homology models of AHK3 receptor in complex with tZ and the *Zea mays* cytokinin receptor ZmHK1 in complex with tZ or thidiazuron (Lomin et al. 2015) suggest that the structure of the cytokinin-binding site is generally conserved among cytokinin receptors and that certain differences in amino acid composition in the binding pockets may be responsible for observed cytokinin-type preference. The small size of the binding cavity, which completely embraces the ligand (Fig. 10.4a), additionally explains the inactivity of the glycosylated cytokinins, which are too large to fit into the site (Spichal et al. 2004; Hothorn et al. 2011). Altogether, these complexes provide a physical basis for the design of novel compounds with cytokinin activity, which must feature a planar ring structure occupying the adenine-binding pocket, followed by a linker that can form hydrogen bonds with highly conserved Asp (Asp262 in AHK4) and a planar aliphatic or aromatic side chain (Hothorn et al. 2011).

10.2.3 Cytokinin Signal Transduction

10.2.3.1 Transducing the Cytokinin Signal Across Plasma Membrane

The CHASE domain is flanked on both sides by hydrophobic transmembrane α-helices. The number of α-helices adjacent to the N terminus of CHASE domain varies among cytokinin receptors from 1 to 4, whereas single α-helix spanning the membrane between CHASE domain and the C-terminal cytosolic HK domain is typical for all cytokinin-responsive HKs (Steklov et al. 2013). Point mutations G435C, F436S, and M447T in the latter α-helix of AHK4 [according to bioinformatic analysis of (Steklov et al. 2013) M447 is located beyond transmembrane α-helix] render the receptor constitutively active regardless of its ability to bind cytokinins (Miwa et al. 2007). This is suggesting the importance of the helix in the transduction of the cytokinin binding to the CHASE domain across the PM. Studies describing analogous amino acid substitutions in AHK2 and AHK3 further

strengthen the conclusion that mutations in this α-helix cause conformational change resembling those caused by cytokinin binding, thus stimulating HK activity even in the absence of the hormone (Miwa et al. 2007). Short linker (~10 amino acids) between the C terminus of CHASE domain and the following transmembrane α-helix is highly conserved (Steklov et al. 2013). The N-terminal part of the linker belongs to the membrane-proximal PAS-like domain of the CHASE domain, whose function remains unknown. The C-terminal part of the linker is proposed to participate together with the adjacent transmembrane α-helix in the interdomain signal transmission (Steklov et al. 2013). However, the detailed mechanisms of how the cytokinin signal is transmitted from the sensory domain to the intracellular HK domain is recently unknown.

10.2.3.2 Activation of Intracellular Domains of Receptors

The intracellular portion of cytokinin receptors is composed of two domains, HK domain and C-terminal RD (HK_{RD}). The HK domain comprises two subdomains, an N-terminal dimerization and histidine phosphotransfer domain, and the C-terminal catalytic and ATP binding domain. No structural data of any cytosolic domain are yet available for cytokinin receptors. Experimental high-resolution crystal structures exist for homologous *Arabidopsis* ethylene receptors and HKs CYTOKININ INDEPENDENT (CKI)1 and CKI2/AHK5 (see section "Receiver Domain"). Thanks to high structural conservancy, these structures may serve as an input body for the modeling of cytosolic domains of cytokinin receptors (Fig. 10.5). However, for a detailed view on cytokinin signal transduction, experimental structures of cytosolic domains of at least one cytokinin receptor would be of high importance.

Recently published crystal structures and conformational dynamics in solution of three RD forms (free, Mg^{2+} bound, and beryllofluoridated) of CKI1 ($CKI1_{RD}$) (Otrusinova et al. 2017) provide the first insight into the phosphorylation-mediated activation of HK_{RD} in the eukaryotic system. In comparison to bacterial RR_{RD}s exhibiting structural changes upon phosphorylation and divalent cation binding, all three $CKI1_{RD}$ structures are identical and similar to the phosphorylated state of bacterial RDs. Nevertheless, the three $CKI1_{RD}$ forms exhibit different conformational dynamics in solution for the β3-α3 loop located close to the phosphorylation site. This conformational freedom can be viewed as a mechanism allowing the unphosphorylated β3-α3 loop to be involved in interactions with different domains of CKI1 during signal transduction. In its non-phosphorylated "inactive" conformation, the β3-α3 loop of $CKI1_{RD}$ seems to mediate interaction with ATP lid of catalytic and ATP-binding domain, thus allowing phosphorylation of $CKI1_{RD}$ by the HK domain. The phosphorylation shifts the equilibrium of preexisting conformational states of the β3-α3 loop to the "active" ones, favoring the interaction with its downstream signaling partners, the HPt proteins (Otrusinova et al. 2017). The functional importance of the observed conformational flexibility was confirmed in D1057E, G1058N mutant of $CKI1_{RD}$. The mutated residues of $CKI1_{RD}$ were replaced by amino acids present in corresponding positions in RD of ethylene

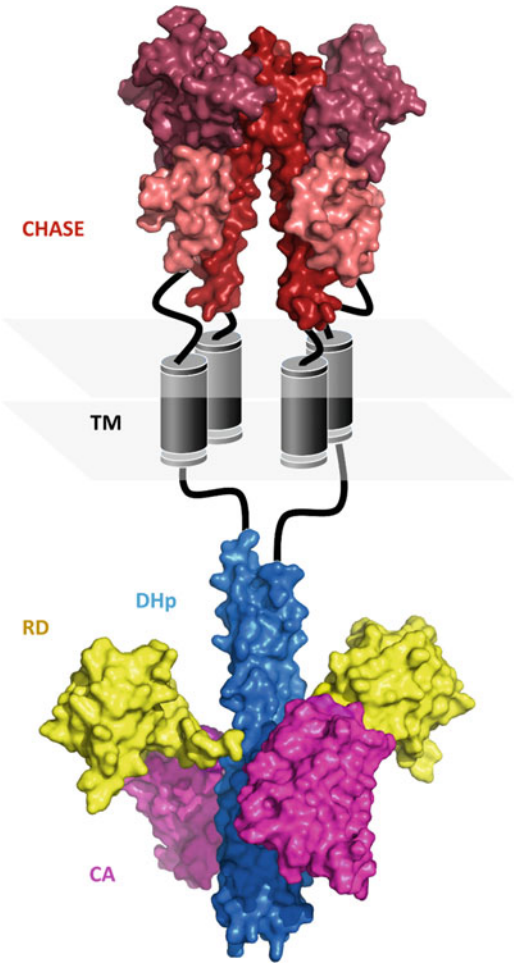


Fig. 10.5 Schematic representation of complete AHK4 protein. The model is based on the analogy with ETR1, see Fig. 10.13. Individual domains labeled accordingly: CHASE – red, transmembrane (TM) – gray, dimerization and histidine phosphotransfer (DHp) – blue, catalytic and ATP binding (CA) – magenta, receiver (RD) – yellow

receptor ETR1, which in contrast to $CKI1_{RD}$ reveals absence of conformational dynamics in its β3-α3 loop and was proposed as atypical RD acting in the phosphorylation-independent way (Hung et al. 2016); see also section "Receiver Domain". In line with that, the D1057E, G1058N $CKI1_{RD}$ version was unable to be phosphorylated in vitro and was inactive in the in vivo signaling assay (Otrusinova et al. 2017).

10.2.3.3 Histidine-Containing Phosphotransfer Proteins

HPt proteins, which function as mobile elements transferring cytokinin signal from the membrane-localized sensor HKs to the nuclear-localized RRs, are small cytosolic single-domain proteins. To date, three-dimensional X-ray structures of five HPt

proteins in free forms and one in a complex with HK_{RD} from several higher plant species have been determined (Bae et al. 2010; Ruszkowski et al. 2013; Bauer et al. 2013; Degtjarik et al. 2013; Sugawara et al. 2005). All structures are highly conserved and retain an entirely α-helical fold consisting of six α-helices connected by loops. Two pairs of C-terminal antiparallel helices, longer α3 and α6 and shorter α4 and α5, form the four-helix bundle typical of all plant, yeast, or bacterial monomeric HPts [Fig. 10.6; (Bae et al. 2010; Bauer et al. 2013; Ruszkowski et al. 2013; Kato et al. 1997; Sugawara et al. 2005)]. The two N-terminal helices, α1 and α2, are not directly included in the four-helix bundle architecture and are connected with α6 and α3 by hydrophobic interactions (Bae et al. 2010).

The four-helix bundle is stabilized by hydrophobic contacts on the inner side of the helices (Bae et al. 2010; Sugawara et al. 2005). In addition, intramolecular sulfur-

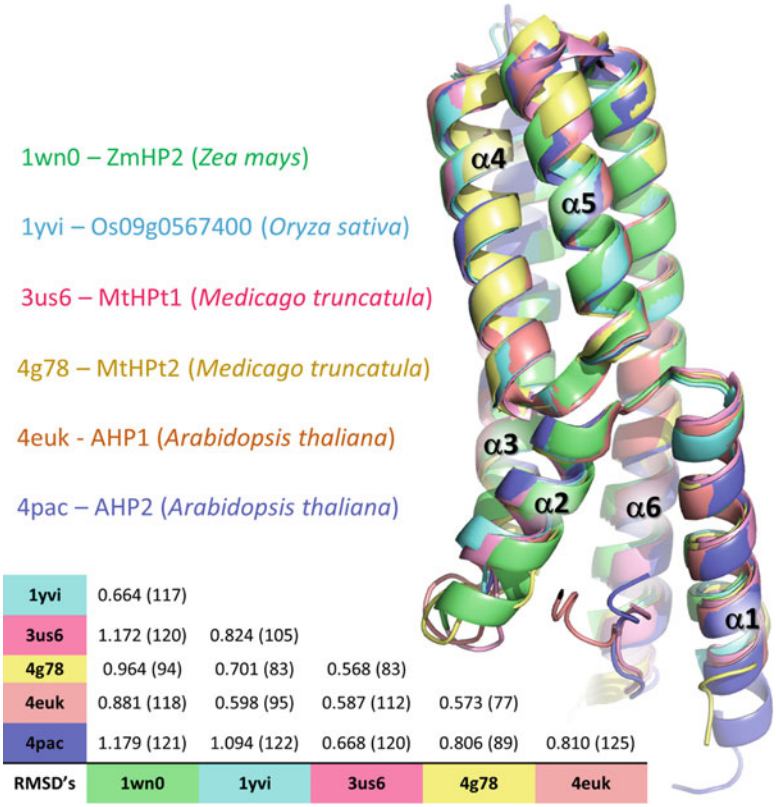


Fig. 10.6 Structural alignment of histidine-containing phosphotransfer proteins (HPts) from plants. RMSDs of individual domains' backbones are listed in the table with number of aligned residues in brackets. Color code: ZmHP2 from *Zea mays* (1WN0) – green, Os09g0567400 from *Oryza sativa* (1YVI) – cyan, MtHPt1 from *Medicago truncatula* (3US6) – pink, MtHPt2 from *M. truncatula* (4G78) – yellow, AHP1 from *Arabidopsis thaliana* (4EUK) – brown, AHP2 from *Arabidopsis thaliana* (4PAC) – blue

π interactions with aromatic residues (Ruszkowski et al. 2013) seem to play a key role in stabilizing the four-helical bundle structure, providing it with high rigidity. Conserved phospho-accepting His residue is positioned in the middle of helix α4 and is solvent exposed thanks to surrounding amino acid residues with shorter side chains (Sugawara et al. 2005). As shown for bacterial HPt protein ArcB and yeast Ypd1, the exchange of some of these residues by others with long side chains significantly reduce phosphorylation efficiency (Janiak-Spens and West 2000; Matsushika and Mizuno 1998).

10.2.3.4 Response Regulators

Based on the protein structure, 21 RRs, which are the final elements of MSP signaling in *Arabidopsis*, are divided into two major families, type-A and type-B [ARRs-A and ARRs-B, respectively; (Imamura et al. 1999; Kieber and Schaller 2014)]. ARRs-A (10 members) are relatively small and possess N-terminal RD, followed by a short C-terminal fragment of unknown function. ARRs-B (11 members) also contain RD followed by a large variable C-terminal extension with a highly conserved, approximately 60-amino-acid-long myeloblastosis (Myb)-related DNA-binding Golgi-associated retrograde protein (GARP) motif, also called B-motif (Hosoda et al. 2002; Imamura et al. 1999). The *Arabidopsis* genome encodes two additional ARRs, structurally similar to ARRs-A, and classified as type-C (ARRs-C). In contrast to ARRs-A, the expression of ARRs-C is not influenced by cytokinins. However, their ectopic expression affects cytokinin signaling output (Kiba et al. 2004; Gattolin et al. 2006), therefore ARRs-C might modulate cytokinin signaling under yet unknown conditions.

While the bacterial RRs with their N-terminal RDs are intensively structurally studied proteins (Gao and Stock 2009), the structure of none of the plant RDs has been solved. Despite low sequence similarity, RDs associated with either HK or RR (HK_{RD} and RR_{RD}, respectively) are generally structurally highly conserved among all kingdoms and exhibit a typical $(\alpha/\beta)_5$ fold (see further in the text and Fig. 10.12). RR_{RD} functions as a final phosphate acceptor in the MSP signaling cascade. Its structural rearrangement upon phosphorylation leads to the activation of RR, allowing thus cytokinin-dependent activation of transcription. However, only little is known about plant Myb-related domains, including the B-motif. To get more mechanistical insight, the structure of a 64-amino-acid-long DNA-binding motif of ARR10 (ARR10-B) was solved employing nuclear magnetic resonance (NMR) spectroscopy (Hosoda et al. 2002). The ARR10-B structure in solution consists of three α-helices and an N-terminal flexible arm (Fig. 10.7). The helices α1 and α2 as well as α2 and α3 are connected by a type I and type II β-turn, respectively, and a hydrophobic core stabilizes the whole structure. Helices α2, α3, and the type II β-turn form the helix-turn-helix (HTH) motif responsible for DNA binding. HTH motif represents a common DNA recognition unit in numerous DNA-binding proteins, such as c-Myb repeats and homeodomains (Riechmann et al. 2000). ARR10-B recognizes the core sequence AGATT located in the major groove of the target DNA

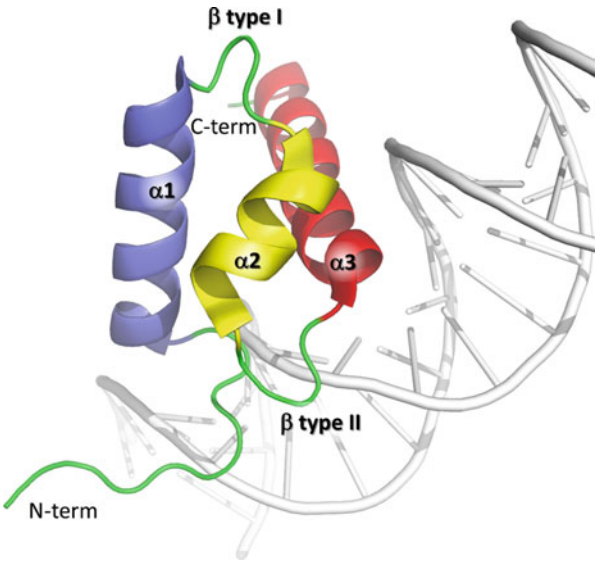


Fig. 10.7 *A. thaliana* B-type response regulator ARR10 (1IRZ). Individual helices and loops labeled. DNA structure (white) placed according to the structure of homologous PhoP-DNA complex of *Mycobacterium tuberculosis* (5ED4)

via residues located in the helix α3 and contacts the adjacent minor groove of DNA with the N-terminal flexible arm (Fig. 10.7). The surface around the HTH motif is positively charged. These positively charged amino acids are a part of the nuclear localization signal, as shown by the nuclear localization of GFP-ARR10 fusion protein in onion epidermal cells (Hosoda et al. 2002).

10.2.4 Molecular Basis of the Interaction Between HPt and RD

The only resolved crystal structure of C-terminal RD of AHK5 ($AHK5_{RD}$) in complex with one of its cognate HPts, AHP1 (Bauer et al. 2013), provides the first mechanistic insight into the interaction mode between plant HK_{RD} and HPt acting in MSP. In the complex, α1 of $AHK5_{RD}$ protrudes into a groove formed by three AHP1 helices (the non-bundle helix α2 and helices α3 and α4 of the four-helix bundle) and α2-α3 adjacent loop (Fig. 10.8a). One half of the resulting interaction interface involves hydrophobic bonds while the other one consists mostly of polar residues of AHP1 forming hydrogen bonds with $AHK5_{RD}$. Three hydrogen bonds and two hydrophobic residues of the non-bundle helix α2 also contribute to the $AHK5_{RD}$-AHP1 complex interface. Hence, the complex structure indicates that the contact area consists of two major features: the hydrogen bond-mediated docking ridge and the hydrophobic patch (Bauer et al. 2013). The recently published in silico derived homology model of AHP1 complexed with 16–175 amino acid region of ARR4 (Verma et al. 2015) mimics the interaction mechanism described in the $AHK5_{RD}$-AHP1 complex, indicating the structural conservancy within MSP proteins and

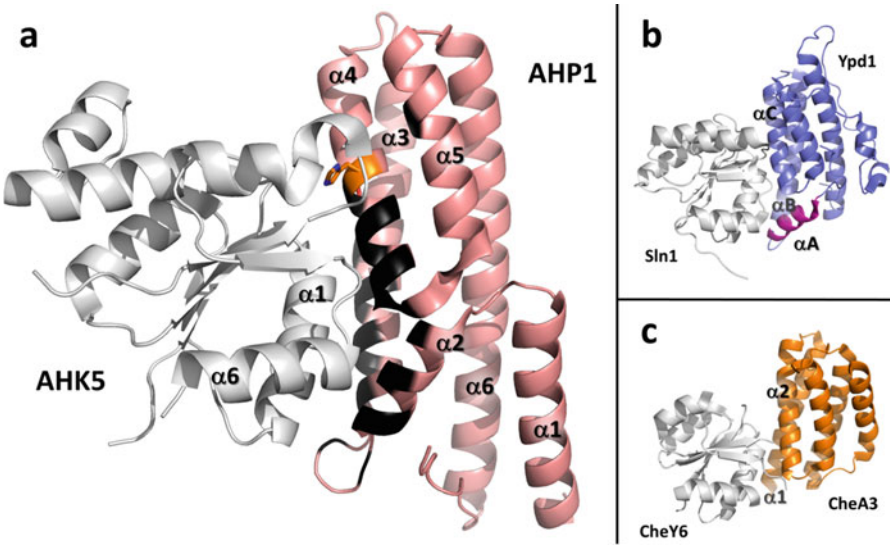


Fig. 10.8 Complex of the receiver domain of histidine kinase (HK_{RD}) and histidine-containing phosphotransfer (HPt) protein mediating the first intermolecular transfer in the MSP signaling. (**a**) $AHK5_{RD}$ (white)-AHP1 (pink) complex from *A. thaliana* (4EUK). AHP1 interacting part of helices 2, 3, and 4 colored in black. Conserved catalytic histidine residue shown in orange. (**b**) HK_{RD} Sln1 (white) in complex with HPt protein Ypd1 (dark blue) (1OXB) from *Saccharomyces cerevisiae*. αA helix corresponding to α2 helix of AHP1 shown in purple. (**c**) HK_{RD} CheY6 (white) complexed by HPt protein CheA3 (orange) (3KYJ) from *Rhodopseudomonas sphaeroides*. Note that the non-bundle helix of HPt protein (corresponding to α2 of AHP1 or αA of Ypd1) is missing in the bacterial system

similar interaction mechanism between AHPs and upstream and downstream RDs in *Arabidopsis*. However, solid experimental data are necessary to confirm that.

The Sln1-Ypd1 co-crystal structure from *Saccharomyces cerevisiae* (Xu et al. 2003) partially resembles $AHK5_{RD}$-AHP1 complex in *Arabidopsis* (Fig. 10.8b), thus suggesting the existence of similar interaction mechanism in yeast. On the other hand, the bacterial HPt proteins are mostly missing the N-terminal non-bundle helices present in eukaryotic HPts. As a consequence, in bacterial complexes the α1 of RR_{RD} dominantly interacts with helices α1 and α2 of cognate HPt (corresponding to α3 and α4 of AHP1, respectively) [Fig. 10.8c; (Bell et al. 2010; Kato et al. 1999; Capra et al. 2010; Skerker et al. 2008; Casino et al. 2009)]. Thus, involvement of the N-terminal non-bundle helix (represented by α2 of AHP1 or the similarly oriented αA of YPD1; Fig. 10.8a, b, respectively) in the interaction interface in eukaryotic HK/RD cognate pairs implies certain structural uniqueness in molecular recognition of eukaryotic vs. bacterial MSP.

Bimolecular fluorescence complementation experiments in transiently transformed tobacco epidermal leaf cells have revealed that $AHK5_{RD}$ interacts with all six AHP proteins except for AHP4. Further, surface plasmon resonance

experiments showed that $AHK5_{RD}$ interacts with AHP1, AHP2, and AHP3 with comparable affinities (Bauer et al. 2013), suggesting that $AHK5_{RD}$ does not prefer any of the AHPs. Accordingly, the residues found to be involved in the interaction between $AHK5_{RD}$ and AHP1 are highly conserved among all six AHP proteins, which together with the aforementioned binding specificity suggests similar interaction mode for all $AHK5_{RD}$-AHP complexes. On the other hand, CKI1 and other *Arabidopsis* HKs show certain level of interaction specificity (Horak et al. 2011; Pekarova et al. 2011). CKI1 interacts with only a subset of AHPs with different affinities (Pekarova et al. 2011). Molecular dynamics modelling based on the structures of free AHP2 (Degtjarik et al. 2013); Degtjarik et al. unpublished) and $CKI1_{RD}$ (Pekarova et al. 2011) confirmed the structural conservancy of molecular docking between *Arabidopsis* HK_{RD}s and AHP proteins, as described for the $AHK5_{RD}$-AHP1 complex. Nevertheless, important structural differences between $AHK5_{RD}$-AHP1 and $CKI1_{RD}$-AHP2 complexes were also observed (Degtjarik et al. unpublished). Several AHP residues were found to determine the position of both interacting partners (Degtjarik et al. unpublished). This is suggesting that not only the interaction interface but also the mutual positioning of both interaction partners might be of key importance in determining the specificity of the RD/HPt interaction. In support of that, swapping the position-determining residues between AHP1 and AHP2 was sufficient to rewire AHP1 and to make it interact with its non-cognate interaction partner $CKI1_{RD}$ (Degtjarik et al. unpublished).

Collectively, these findings rather impugn the previously proposed role of AHP proteins as a redundant signaling hub in MSP (Urao et al. 2000; Suzuki et al. 2001b; Imamura et al. 1999; Dortay et al. 2006) and substantially contribute to our understanding of MSP signaling. The aforementioned discoveries highlight the importance of molecular recognition, acting potentially as a molecular mechanism allowing integration of diverse signals into a single MSP pathway while maintaining the information about the origin of the signal.

10.3 Ethylene Signaling

10.3.1 Overview of the Ethylene Signaling Pathway

In *Arabidopsis,* ethylene is perceived via a five-member receptor family, consisting of ETHYLENE RESPONSE (ETR)1, ETR2, ETHYLENE RESPONSE SENSOR (ERS)1, ERS2, and ETHYLENE INSENSITIVE (EIN)4 (Bleecker et al. 1988; Chang et al. 1993; Hua et al. 1998; Sakai et al. 1998; Hua and Meyerowitz 1998). Ethylene receptors are located predominantly in the ER membrane (Grefen et al. 2008; Bisson et al. 2009; Chen et al. 2002). The ethylene sensors are further divided into two subfamilies based on their phylogenetic and functional features. In *Arabidopsis*, subfamily 1 includes ETR1 and ERS1, both of which possess HK activity. More diverged subfamily 2 includes serine/threonine (Ser/Thr) kinases

ETR2, ERS2, and EIN4. ERS1 is the only receptor displaying both His and Ser/Thr kinase activity in vitro (Moussatche and Klee 2004). Genetic data indicate that HK activity of ethylene sensors is not required for ethylene signal transmission, but might play a modulating role, for example, in the regulation of interaction strength between receptors and downstream components of the signaling pathway (Hall et al. 2012; Wang et al. 2003). *Arabidopsis* subfamily 1 plays a dominant role in ethylene responses, but the signaling mechanism is still largely unclear (Qu et al. 2007). Ethylene receptors have both overlapping (Hua et al. 1995, 1998; Hua and Meyerowitz 1998; Hall and Bleecker 2003; Binder et al. 2004), but also distinct functions (Binder et al. 2012; Shakeel et al. 2012; Wang et al. 2003; Hall and Bleecker 2003). The unique role of subfamily 1 is clearly demonstrated by the inability of receptors of subfamily 2 to functionally substitute for receptors of subfamily 1 (Wang et al. 2003).

In contrast to cytokinin signaling, where cytokinin binding activates the kinase activity of cytokinin sensors, in the air (absence of ethylene), the ethylene receptors are active, functioning as negative regulators of the ethylene signaling [Fig. 10.9a; (Wang et al. 2006; Hua and Meyerowitz 1998)]. Active ethylene receptors associate with and activate Ser/Thr protein kinase CONSTITUTIVE TRIPLE RESPONSE (CTR)1 (Kieber et al. 1993; Clark et al. 1998; Huang et al. 2003), the Raf-like kinase, and another negative regulator of the pathway. Active CTR1 forms a homodimer (Mayerhofer et al. 2012) and phosphorylates metal transporter-like protein EIN2 (Ju et al. 2012). Phosphorylated EIN2 is subsequently degraded by 26S proteasome under the action of two F-box proteins, EIN2 TARGETING F-BOX PROTEIN (ETP)1 and ETP2 (Qiao et al. 2009). Meanwhile, in the nucleus, EIN3-BINDING F-BOX PROTEIN (EBF)1 and EBF2 target the key transcription factors EIN3 and its homolog EIL1 (EIN3-like 1) for the 26S proteasomal degradation pathway, resulting in suppression of downstream ethylene responses (Guo and Ecker 2003, 2004; Potuschak et al. 2003; Gagne et al. 2004; Binder et al. 2007).

Binding of ethylene inactivates ethylene receptors that (via yet unknown mechanism) attenuates the kinase activity of CTR1, resulting in the inability of CTR1 to phosphorylate EIN2 (Fig. 10.9b). Dephosphorylated EIN2 is proteolytically cleaved and its C-terminal portion (EIN2C) translocates to the nucleus (Ju et al. 2012; Wen et al. 2012; Qiao et al. 2012). That allows EIN2-dependent 26S proteasomal degradation of negative regulators EBF1 and EBF2, resulting into concomitant accumulation and stabilization of the transcription factors EIN3 and EIL1 in the nucleus (Guo and Ecker 2003; An et al. 2010). EIN3 and EIL1 seem to function as homodimers (Solano et al. 1998; Li et al. 2012) and activate the downstream transcriptional cascade controlling the ethylene-responsive gene expression (Alonso et al. 2003; Konishi and Yanagisawa 2008). Although the nuclear localization of EIN2C in ethylene signaling is pivotal, two recent reports (Li et al. 2015; Merchante et al. 2015) reveal a novel role of EIN2 in the control of translation. On top of the role of EIN2 in 26S proteasome-mediated degradation of EBF1 and EBF2, the authors found that EIN2C is required to inhibit translation of both negative regulators. EIN2C binds to the 3′ UTR of EBF1/2 and targets EBF1/2 mRNA to cytoplasmic

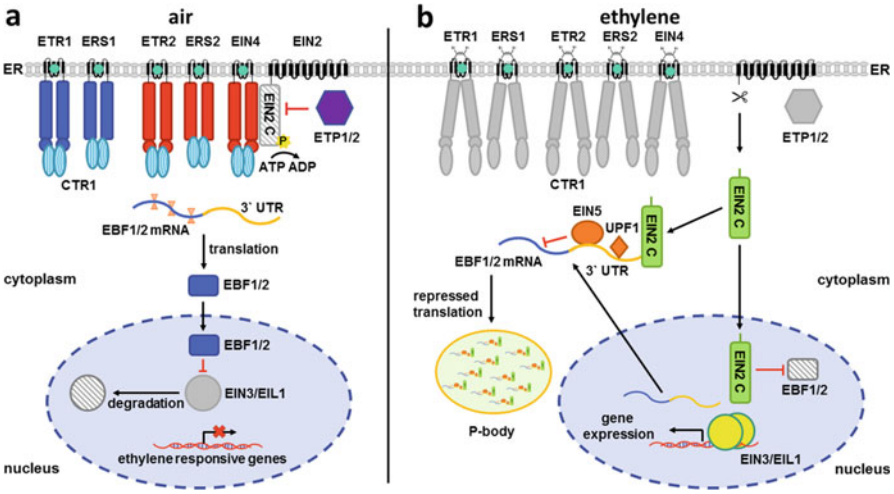


Fig. 10.9 Model of the ethylene signal transduction. (**a**) In the air (absence of ethylene) the ethylene receptors ETR1, ERS1 (subfamily 1, dark blue) and ETR2, ERS2, and EIN4 (subfamily 2, red) are negative regulators of the ethylene signaling. Copper ion (green circles) is an ethylene-binding cofactor. The receptors activate CTR1 (blue dotted ovals) that dimerizes when active and suppresses the ethylene response. CTR1 inactivates EIN2 (gray rectangle) by phosphorylation (yellow star). EIN2 can directly interact with the kinase domain of the receptors. The level of EIN2 is negatively regulated by the F-box proteins ETP1/2 (violet hexagon). The mRNA of two other F-box proteins EBF1/2 (dark blue curly line) is translated and the proteins are transported to the nucleus. There, EBF1/2 (dark blue square) mediates proteasomal degradation of transcription factors EIN3/EIL1 (gray circle) that switches off the transcription of the ethylene response genes. (**b**) In the presence of ethylene, the receptors (gray) bind the hormone and become inactive, which in turn switches off the CTR1 (dotted gray ovals). This inactivation prevents the phosphorylation of EIN2 (green). The C-terminal end of non-phosphorylated EIN2 is cleaved off by an unknown mechanism and attaches to 3′-UTR of *EBF1/2* (yellow line). Together with EIN5 (orange oval) and UPF1 (orange diamond), EIN2 blocks the translation of *EBF1/2* mRNAs and redirects them for P-body-mediated turnover. The transcription factors EIN3/EIL1 (yellow circles) dimerize and activate the expression of ethylene target genes, including the F-box genes *EBF1/2*. Inactive molecules are colored gray while stripped molecules depict proteins targeted to proteasome-mediated degradation. Phosphate is shown as a yellow star and ribosomes as orange hourglasses

processing body (P-body) via interaction with several P-body factors, including 5′-3′ exoribonuclease EIN5 (also known as EXORIBONUCLEASE 4) and nonsense-mediated decay proteins (UPFs). For more detailed information on ethylene signaling, we refer the reader to recent reviews (Ju and Chang 2015; Cho and Yoo 2015).

10.3.2 Ethylene Signal Recognition

The basic functional unit of the receptors is a disulfide-linked dimer (Schaller et al. 1995; Hall et al. 1999) binding one copper ion (Cu^{+}) and one molecule of ethylene (Rodriguez et al. 1999). The copper ion is supplied by the intracellular copper

transporter P-type ATPase RESPONSIVE-TO-ANTAGONIST (RAN)1 and appears to be required for proper conformation and activity of ethylene receptors (Hirayama et al. 1999). Both ethylene and copper ion are bound by the highly conserved N-terminal binding domain. Within the ethylene-binding domain of ETR1, mutation analysis identified seven residues to be important for ethylene recognition (Wang et al. 2006) and two residues to be potential binding sites for copper ion (Rodriguez et al. 1999). For further information on ethylene-copper complex chemistry, we refer the reader to a recent review (Light et al. 2016). Although the crystal structure of the binding domain is not available, the current model suggests that the domain consists of three membrane-spanning α-helices (Rodriguez et al. 1999). Receptors of subfamily 2 contain an additional N-terminal transmembrane domain that may function as a membrane targeting signal sequence (Sakai et al. 1998; Hua et al. 1998).

10.3.3 Ethylene Signal Transduction

10.3.3.1 Intracellular Domains of Ethylene Receptors

Intracellular cytosolic portion of ethylene receptors is composed of a c*G*MP-specific phosphodiesterase, *a*denylyl cyclase and *F*lhA binding (GAF) domain, and a HK domain. Three of five *Arabidopsis* receptors (ETR1, ETR2, and EIN4) possess an additional C-terminal RD. While the GAF domain mediates interaction among different types of receptor dimers to form non-covalent higher-order clusters (Gao et al. 2008; Grefen et al. 2008), HK domain and RD are involved in the interaction with CTR1 and EIN2, the downstream components of the canonical ethylene pathway (Clark et al. 1998; Bisson et al. 2009; Bisson and Groth 2010; Chen et al. 2010). Excepting GAF domain, high-resolution crystal structures of cytosolic domains (single representative for each) are currently available.

Dimerization and Histidine Phosphotransfer (DHp) Domain

The three-dimensional crystal structure of DHp domain, participating in the dimerization of the receptor, was solved for ERS1 ($ERS1_{DHp}$) (Mayerhofer et al. 2015). The $ERS1_{DHp}$ structure reveals an asymmetric homodimer, with each monomer consisting of two antiparallel (hairpin) α-helices. Approximately 30 of the N-terminal residues of the hairpin helices compose an antiparallel left-handed coiled-coil structure, while the remaining C-terminal portions assemble a structurally conserved four-helix bundle (Fig. 10.10a). The coiled-coil motif interconnects the ethylene-binding transmembrane and GAF domains and it is likely to be included in a transduction of the conformational changes upon ethylene binding, potentially leading to the change in the activity of the intracellular portion of the sensor.

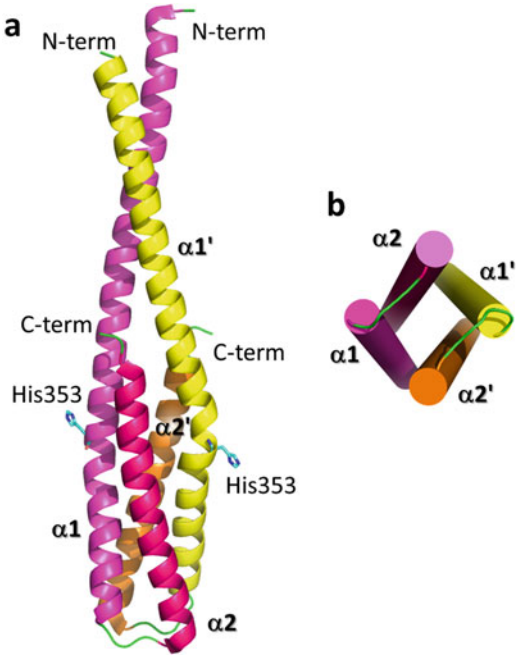


Fig. 10.10 Dimer of ERS1 dimerization and histidine phosphotransfer (DHp) domain (4MT8). (**a**) Individual monomers shown in pink/magenta and yellow/orange, respectively. Catalytic histidine shown in cyan. (**b**) Bottom view of the dimer highlighting the four-helix bundle organization determining the self-phosphorylation of the domain in cis- or trans-position; the latter mechanism is predicted for ERS1/ETR1 as based on the bundle architecture

The His353, corresponding to the phospho-accepting His in other sensor HKs, is located on helix α1 of the four-helix bundle and it is oriented outward for autophosphorylation. Two residues prior to His353, there is a kink common to all known DHp structures. The bending angle in one monomer is ca. 18°, whereas in the other it is ca. 31°. The partial asymmetry of $ERS1_{DHp}$ dimer presumably influences the orientation of both preceding GAF and succeeding catalytic and ATP-binding domains. Molecular dynamics calculations indicate that the kinks are mutually replaceable, indicating some flexibility in the structure, which might be important for hormone signal transduction (Mayerhofer et al. 2015). The topology of loop-connecting hairpins (Fig. 10.10b) suggests phosphorylation in the trans position, that is, in the dimer each monomer phosphorylates its interaction partner.

Catalytic and ATP Binding (CA) Domain

Phosphorylation of the conserved His residue in the DHp domain is mediated by CA domain. CA binds ATP, which then donates its γ-phosphate to the conserved His residue of the DHp domain. The three-dimensional crystal structure of ETR1 CA domain ($ETR1_{CA}$) in complex with ADP and $CdSO_4$ was determined (Mayerhofer et al. 2015). The structure reveals detailed interactions between the $ETR1_{CA}$ and ADP, providing us with structural insight into the $ETR1_{CA}$ in a state ready to donate γ-phosphate. Noteworthy, the crystallization of the $ETR1_{CA}$ was not successful in the presence of biologically active cations such as magnesium (Mg^{2+}) or manganese

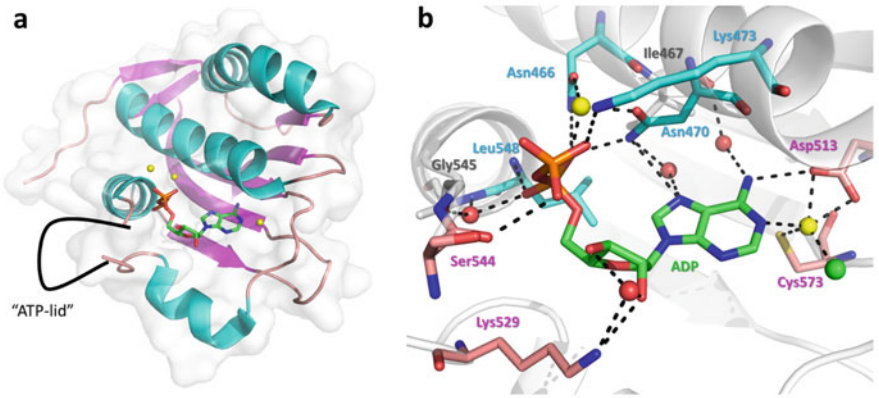


Fig. 10.11 ETR1 catalytic and ATP binding domain with hydrolyzed ATP (ADP) bound (4PL9). (**a**) Overall view of the domain. Cd^{2+} ions in the vicinity of ADP (highlighting possible positions of Mg^{2+} ions in vivo) shown as yellow spheres. Approximate position of disordered "ATP-lid" shown as black curve. (**b**) Detail of ADP positioning in the binding site. Amino acids involved in ADP binding shown as sticks in colors corresponding to panel "a" when intermediating hydrogen bond or white when binding through water bridge only. Water molecules involved in ligand binding shown as red spheres, Cd^{2+} ions as yellow spheres, chloride as green sphere. Hydrogen bonds shown as black dashed lines

(Mn^{2+}), the latter being a necessary cofactor for ETR1 HK activity (Moussatche and Klee 2004). Instead, $ETR1_{CA}$ was crystallized in the presence of Cd^{2+} [see above; (Mayerhofer et al. 2015)]. The structure has an expected αβ sandwich fold containing two layers: the first consisting of five β-strands and the second one formed by three α-helices and a pair of short antiparallel β-strands (Fig. 10.11a). Sequences of the CA domains from both eukaryotic and prokaryotic sensors have well conserved N, G1, F, G2, and G3 box sequence motifs. Both conserved and partially conserved residues from the individual boxes participate in the formation of the hydrophobic ATP-binding pocket. In the structure, the conserved Asp513 of the G1 box forms a direct hydrogen bond with the amino group of the adenine moiety and a Cd^{2+}-mediated hydrogen bond with N1 of the adenine ring. The Cd^{2+} ion is further coordinated with Cys573 from the G3 box and the Cl^{-} ion. Ribose moiety is stabilized by Lys529 side chain through both direct and water-mediated interactions. The α- and β-phosphate groups of ADP are in contact with Asn466, Asn470, and Lys473 side chains, the main chain amide of Leu548 and backbone oxygen of Ser544. Oxygens of the phosphate groups are ligated together via side chains of Asn466 and Asn470 through another Cd^{2+} ion (Fig. 10.11b). Interestingly, the Cys573 residue required for metal (Mg^{2+} or Mn^{2+}) binding is present only in subfamily 1 of ethylene receptors. Larger residues at this position presumably prevent binding of metal ion by members of the subfamily 2. Whether the absence of the metal-binding site in ethylene receptors of subfamily 2 is responsible for the lack of their HK activity, however, remains to be experimentally verified. The highly flexible loop rich in glycine residues and covering the ATP-binding pocket, called

the "ATP-lid," is disordered in ETR1. Importantly, the flexibility of the loop is necessary for binding the ATP as well as for interaction with the His-containing phosphorylation site in structurally characterized DHp in bacterial systems (Wang 2012; Marina et al. 2001; Casino et al. 2014).

Receiver Domain

$ETR1_{RD}$ was the first *Arabidopsis* HK_{RD} with determined structure (Muller-Dieckmann et al. 1999). To date, high-resolution crystal structures of two previously mentioned HK_{RD}s, both from *Arabidopsis*, are available, namely, $CKI1_{RD}$ (Pekarova et al. 2011) and $AHK5_{RD}$ in complex with AHP1 (Bauer et al. 2013). All three domains show a high level of conformational similarity to both prokaryotic and eukaryotic HK_{RD}s and RR_{RD}s of the CheY-like superfamily. The HK_{RD} structure resembles a typical $(\alpha/\beta)_5$ fold consisting of a five-stranded β-sheet core (composed of mainly hydrophobic amino acid residues) sandwiched by two α-helices on one side and three on the other. The phosphorylation and Mg^{2+} binding sites are located in an acidic pocket near the C-termini of the three central β-strands (Fig. 10.12). Conserved phospho-accepting Asp residue is located at the C-terminus of β3-strand (Bourret 2010; Volz 1993; Wilson et al. 2009). A root mean-square

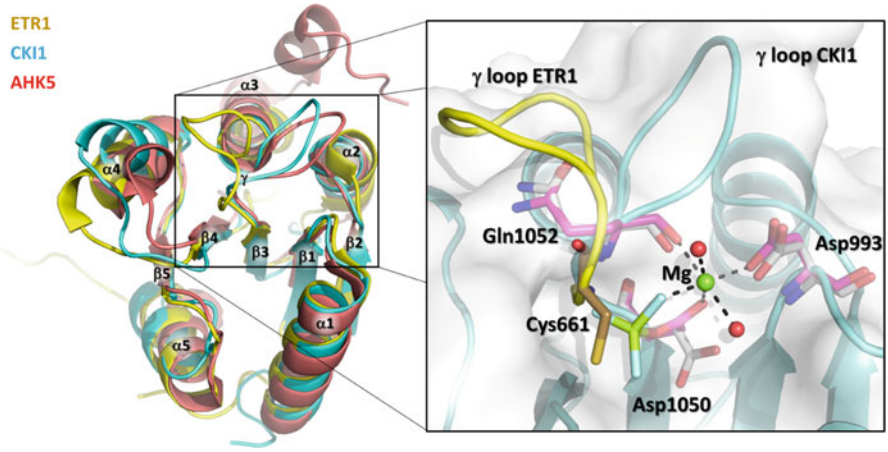


Fig. 10.12 Structural alignment of C-terminal receiver domains of *Arabidopsis* histidine kinases. ETR1 (yellow, 1DCF), CKI1 (cyan, 3MM4), and AHK5 (pink, 4EUK) aligned with secondary structure elements and catalytic γ loop labeled. Note the opposite localization of the loop in $ETR1_{RD}$ when compared to other RDs. Active site of $CKI1_{RD}$ enlarged. Magnesium (Mg^{2+}, green sphere) binding residues highlighted as sticks in magenta (Mg^{2+}-bound form) and white (Mg^{2+}-free form). ETR1 Cys661 corresponding to CKI1 Gln1052 is shown in brown. Water molecules represented as red spheres. The position of ETR1 γ loop shown in yellow. RMSD [Å] of individual structure comparisons are as follows, in parenthesis is the number of aligned amino acid residues: ETR1 (1DCF):CKI1 (3MM4) 0.953 (92); ETR1 (1DCF):AHK5 (4EUK) 1.097 (81); CKI1 (3MM4):AHK5 (4EUK) 1.061 (84)

deviation (RMSD) between α-carbons of $ETR1_{RD}$ and $CKI1_{RD}$, $ETR1_{RD}$ and $AHK5_{RD}$, and $CKI1_{RD}$ and $AHK5_{RD}$ is 0.953 Å for 92 aligned residues, 1.097 Å for 81 residues, and 1.061 Å for 84 residues (PyMol), respectively, thus indicating a high level of structural similarity. Two major features distinguish $ETR1_{RD}$ from $AHK5_{RD}$ and $CKI1_{RD}$: (i) position of the loop neighboring the phosphorylation site and connecting strand β3 to the helix α3 (also called γ loop) and (ii) conformational dynamics of the loop.

In $CKI1_{RD}$, $AHK5_{RD}$, and other (bacterial and yeast) structurally determined RDs, the γ loop is in a close proximity to the helix α2 (Pekarova et al. 2011; Bauer et al. 2013; Volz 1993). In contrast, the corresponding loop in $ETR1_{RD}$ is flipped to the opposite side, interacting with the helix α4 (Muller-Dieckmann et al. 1999). Due to this atypical orientation, the active site in $ETR1_{RD}$ is incomplete. While other residues involved in the Mg^{2+} coordination are conserved, the flipped orientation of the ETR1 γ loop displaces the oxygen of the backbone carbonyl of Cys661 (corresponding to Gln1052 in CKI1) outward from the active site, preventing thus coordination of Mg^{2+} ion (Fig. 10.12). In addition to those data from X-ray structure, a recent NMR-based study showed that $ETR1_{RD}$ is unable to bind Mg^{2+} as well as the phosphate analog beryllofluoride (BeF_3^-) in solution (Hung et al. 2016), indicating the inability of $ETR1_{RD}$ to be phosphorylated.

The incorporation of divalent cation and phosphorylation stabilize the otherwise flexible tertiary structure of typical bacterial RR_{RD}s (Feher et al. 1997; Ocasio et al. 2015; Knaggs et al. 2007). We have shown that γ loop in $CKI1_{RD}$ is also highly flexible in solution and the binding of Mg^{2+} leads to its stabilization (Pekarova et al. 2011). In contrast to $CKI1_{RD}$, the structure of $ETR1_{RD}$ is rather rigid in solution even in the absence of Mg^{2+} (Hung et al. 2016; Otrusinova et al. 2017). Limited structural flexibility and inability to bind both Mg^{2+} and BeF_3^- make $ETR1_{RD}$ unfavorable for phosphorylation. Thus, $ETR1_{RD}$ might be classified as an atypical RD, functioning presumably via binding with the other components of the pathway independently on phosphorylation.

10.3.3.2 Model of the Intracellular Portion of Sensor HK

The knowledge of the aforementioned individual cytosolic domain structures ($ERS1_{DHp}$ and $ETR1_{RD}$) combined with the GAF domain from *Deinococcus radiodurans* bacteriophytochrome (Wagner et al. 2007) allowed their assembly into the molecular envelope of the cytosolic part of ETR1 determined via small angle X-ray scattering (SAXS) (Mayerhofer et al. 2015). In spite of a relatively high level of uncertainty, the resulting first model of plant sensory HK provides valuable structural insight into the inter-domain organization of the catalytic core of not only ethylene receptors but possibly all plant HKs. The model indicates that the cytosolic portion of ETR1 is a dimer of a dumbbell-like shape in solution. The GAF domain is located on the top (membrane-proximal) and CA and RD on the bottom (membrane-distal), expanding outward from the central DHp helical stalk (Fig. 10.13). The GAF and HK domains adopt defined positions relative to each other without any physical

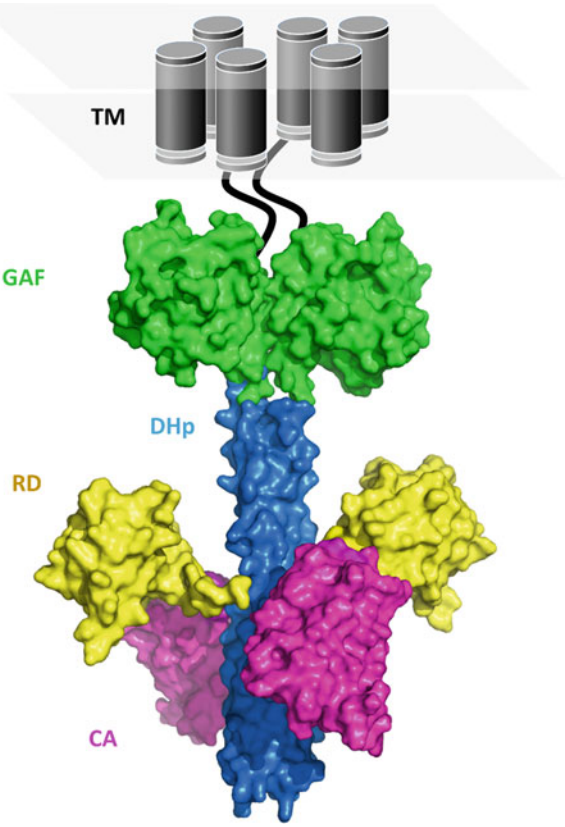


Fig. 10.13 Schematic representation of complete ETR1 protein. Individual domains designated as follows: transmembrane (TM) – gray, GAF – green, dimerization and histidine phosphotransfer (DHp) – blue, catalytic and ATP binding (CA) – magenta, receiver (RD) – yellow. (Based on Mayerhofer et al. (2015) and modified)

contact between them. In contrast, RDs are flexible and separated from each other. In none of the obtained model variants did the two RDs come close enough to homodimerize, thus confirming recent NMR findings that $ETR1_{RD}$ is monomer in solution (Hung et al. 2016). In the most frequent conformation of the dimer observed in an ensemble analysis, the RDs adopt a position suited for interaction with other receptors or proteins. This is consistent with the fact that interactions of CTR1 with the HK domain and RD of ETR1 and ERS1 are essential for CTR1 activation (vide infra) (Clark et al. 1998; Gao et al. 2003; Huang et al. 2003).

10.3.3.3 Downstream Signaling Proteins, CTR1 and EIN2

The cytosolic protein CTR1, active in the absence of ethylene but deactivated after ethylene perception, acts immediately downstream of *Arabidopsis* ethylene receptors. CTR1 consists of two domains: N-terminal regulatory domain and C-terminal protein kinase domain. CTR1 interacts with ethylene receptors via its N-terminal regulatory domain (Clark et al. 1998; Gao et al. 2003; Huang et al. 2003), but the

molecular mechanism of how ethylene receptors transfer the signal to CTR1 and control its activity is still unclear. The interaction leads to co-localization of CTR1 with ethylene receptors at the ER membrane and its upregulation, possibly via CTR1 stabilization (Gao et al. 2003). Upon deletion of the N-terminal domain, CTR1 becomes a constitutively active kinase (Yoo et al. 2008).

The C-terminal kinase domain of CTR1 ($CTR1_{KD}$) is similar to a catalytic domain of Raf-like mitogen-activated protein kinase kinase kinase (MAPKKK) (Kieber et al. 1993), but the existence of the mitogen-activated protein kinase cascade downstream of CTR1 has never been reliably identified (Ecker 2004). C-terminal $CTR1_{KD}$ was crystallized as an enzymatically active (triphosphorylated) and inactive (non-phosphorylated) form in a complex with kinase inhibitor staurosporine (Mayerhofer et al. 2012). Inactive (kinase-dead) form was prepared by replacing the strictly conserved Asp676 located in the catalytic loop with asparagine. The structure of $CTR1_{KD}$ adopts the characteristic protein kinase fold comprising two lobes connected by a short linker. The N-lobe (residues 540–626) consists of five-stranded antiparallel β-sheet, including a typical α-helix (residues 587–601; designated as helix C), whereas the larger C-lobe (residues 632–821) has a predominantly α-helical structure [Fig. 10.14; (Mayerhofer et al. 2012)].

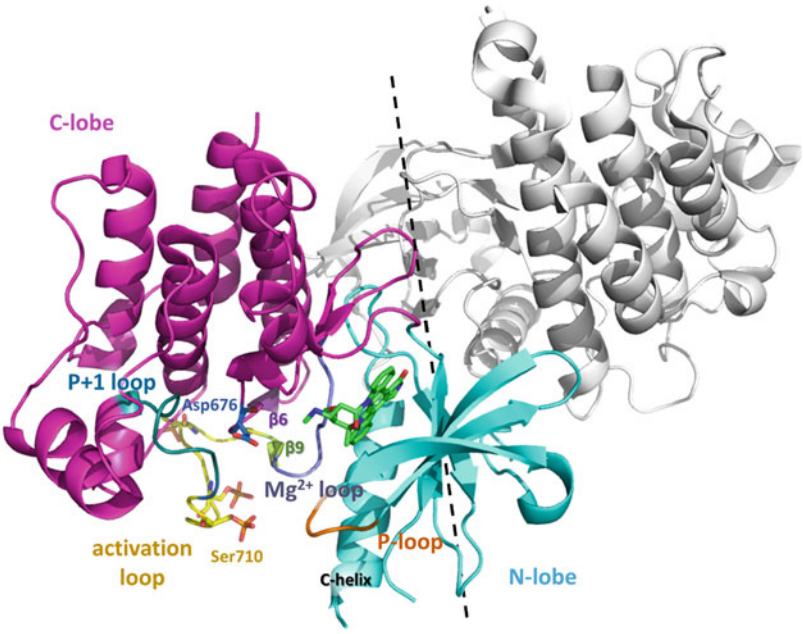


Fig. 10.14 Dimer of C-terminal kinase domain of CTR1 (3PPZ). Main structural elements on one monomer highlighted: N-lobe including C-helix – cyan, C-lobe – magenta, conserved P-loop – orange, Mg^{2+} loop – violet, activation loop with phosphorylated residues as sticks – yellow, P + 1 loop – teal, β6 – dark purple, β9 – lime, catalytic Asp676 – dark blue, inhibitor staurosporine – green. The dimer symmetry axes shown by dashed line

CTR1 belongs to the RD kinase family (RD kinases contain conserved arginine preceding conserved catalytic aspartate in the catalytic loop), being often activated by the phosphorylation of their activation segments. The activation segment of RD kinase consists of several secondary motifs, which include (from N to C terminus) Mg^{2+} binding loop, β9, the activation loop, and the P + 1 loop (Nolen et al. 2004). Three phosphorylation sites (Thr704, Ser707, and Ser710) are concentrated within the activation loop located in the C-lobe of $CTR1_{KD}$ (Fig. 10.14). The functionally important activation loop is unstructured in the inactive non-phosphorylated form of $CTR1_{KD}$. Phosphorylation of the Ser710 initiates folding of β9 (Arg699/Leu700/Lys701), which is disordered in the inactive $CTR1_{KD}$, via formation of hydrogen interactions between phosphorylated Ser710 and positively charged residues Arg675 and Arg699 in the surrounding basic pocket. That results in the stabilization of the activation loop via the formation of a short antiparallel β-sheet between β6 (Pro671/Ile672/Val673) preceding the catalytic loop and β9 within the activation loop (Mayerhofer et al. 2012).

Besides phosphorylation of the activation segment, there are also phosphorylation-independent mechanisms of RD kinase activation. Activation of RAF-B kinase was shown to be dimerization dependent (Rajakulendran et al. 2009). Importantly, kinase domains of both CTR1 and homologous RAF-B kinase form virtually identical dimers in the crystal, thus implying a similar mechanism of activation in case of CTR1 too (Mayerhofer et al. 2012). Accordingly, the wild-type $CTR1_{KD}$ forms a dimer also in solution, while the kinase-dead D676N mutant as well as several catalytically inactive activation loop mutants are monomeric (Mayerhofer et al. 2012). Thus, homodimerization seems to be the mechanism involved in CTR1 activation. However, active CTR1 forms a "back to back" dimer, placing the substrate binding sites (located close to the P + 1 loop motif and Asp676) at the opposite ends of the dimer (Fig. 10.14). Consequently, the activation mediated by phosphorylation cannot occur between the individual protomers of the dimer, but rather between "face-to-face" interacting protomers of neighboring dimers. In support of that, the activation interface formed between dimers in adjacent asymmetric units of $CTR1_{KD}$ seems to resemble the activation interface of the *Mycobacterium tuberculosis* PknB receptor Ser/Thr kinase (Mieczkowski et al. 2008).

Based on the aforementioned structural and functional characterization of CTR1, the mechanism for CTR1 activation by ethylene receptors was suggested. In the model (Fig. 10.15), each receptor dimer binds the dimer of CTR1 (2:2 stoichiometry). That allows the formation of receptor kinase units in which ethylene might activate CTR1 via phosphorylation through their back-to-back dimer interface. Once activated, the activation status could spread throughout receptor clusters [the formation of which was described (Gao et al. 2008)] via front-to-front enzyme–substrate interactions allowing transphosphorylation between neighboring dimers]. This model provides a plausible explanation for the broad range of ethylene sensitivity (0.2 nL/L–1000 μL/L) as well as the dominant nature of ethylene-insensitive mutants via the formation of non-covalent higher-order complexes between ethylene receptors (Mayerhofer et al. 2012).

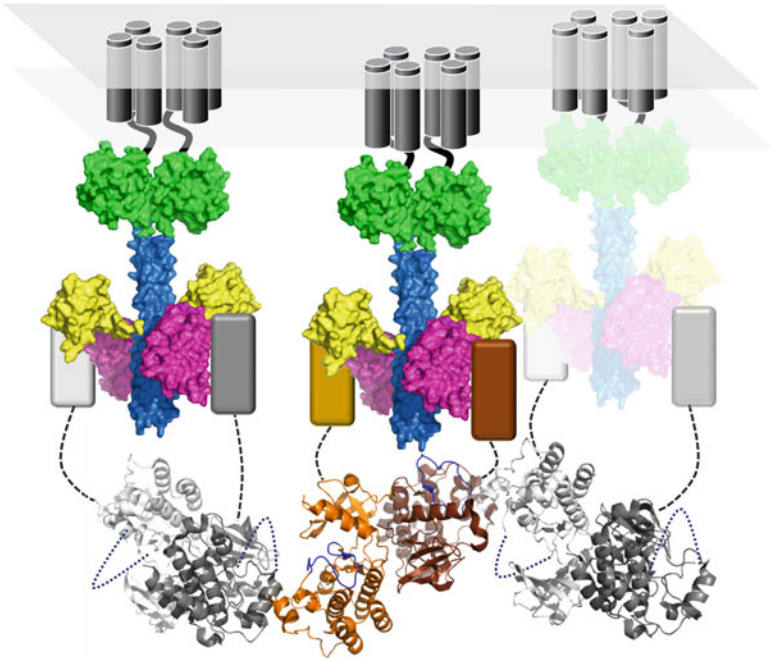


Fig. 10.15 Scheme of CTR1 activation and receptor oligomerization. Receptors shown as in Fig. 10.13, keeping the domain color code: transmembrane (TM) – gray, GAF – green, dimerization and histidine phosphotransfer (DHp) – blue, catalytic and ATP binding (CA) – magenta, receiver (RD) – yellow. CTR1 kinase domains shown in their chain-forming interactions including both back-to-back (dark-light) and front-to-front (gray-brown) inter-dimer interfaces as seen in crystal. The structured activation loop shown in blue for central (activated) dimer, for other (inactive) dimers, the position of the loop is shown as blue dashed line. The N-terminal regulatory domains shown as cartoons and colored according to kinase domains (activated kinase domain in brown, inactive in gray), linkers between CTR1 domains shown as black dashed lines

EIN2 is a key component of the ethylene signaling pathway connecting ethylene perception at the ER membrane and nuclear responses. EIN2 consists of an ER membrane-anchored N-terminal regulatory domain and cytosolic C-terminal domain (Alonso et al. 1999). The N-terminal regulatory domain is predicted to be composed of 12 transmembrane α-helices. Domain sequence comparison studies show similarity of EIN2 to Nramp family of metal-ion transporter proteins. However, no ion transport activity could be detected for EIN2. Hydrophilic C-terminal domain (EIN2C) directly interacts with ethylene receptors and CTR1. EIN2C with its intrinsic nuclear localization signal is sufficient to activate ethylene responses (Bisson and Groth 2010, 2011). Phosphorylation of EIN2 by CTR1 on at least two residues (Ser645 and Ser924) of EIN2C (Qiao et al. 2012) is a key regulatory mechanism of ethylene signal transmission. Preventing EIN2 phosphorylation on Ser645 and Ser924 by alanine substitution results in constitutive nuclear localization of EIN2C and constitutive activation of ethylene responses (Ju et al. 2012). However, the structural basis of neither EIN2 phosphorylation nor EIN2C cleavage has been defined.

10.3.3.4 Ethylene-Controlled Transcription Factors

EIN3 and EIL (EIL1–3) proteins are nucleus-localized transcription factors that bind to the promoter regions of the downstream genes and regulate their expression, thus controlling the vast majority of ethylene responses (Chang et al. 2013; Chao et al. 1997). The EIN3 and EIL proteins from different higher plant species are highly homologous, especially in their N-terminal half (~ 300 amino acid residues) (Yamasaki et al. 2005). This N-terminal half except for the first ~ 80 residues is rich in basic residues (Yamasaki et al. 2005) and was found to be essential for DNA-binding activity (Solano et al. 1998; Kosugi and Ohashi 2000), although showing no sequence similarity to other known DNA-binding proteins (Chao et al. 1997; Riechmann et al. 2000). Additionally, the DNA-binding part of EIN3 and EIL proteins contains a region rich in proline residues. Considering that transcription factors function as homodimers and that homodimerization of EIN3 requires a proline-rich region (Solano et al. 1998), the region is suggested to act as a dimerization interface. The function of the unstructured C-terminal part of proteins is currently unknown.

The three-dimensional structure of the *Arabidopsis* EIL3 DNA-binding region (Ser162-Gln288) determined in solution by NMR spectroscopy revealed a fold dissimilar to known DNA-binding domain structures. The structure consists of five α-helices packing against one another in a parallel or an anti-parallel manner, forming together a globular structure as a whole (Fig. 10.16). The packing of the α-helices is facilitated by interactions between hydrophobic side-chains mostly of aliphatic or aromatic residues (Yamasaki et al. 2005). All of these residues are highly conserved among the EIN3 and EIL proteins, indicating that the DNA-binding domains possess a similar structural architecture. As shown by NMR titration experiments, residues of helices α1 and α3, and, in part, adjacent loops connecting helices α1 and α2 and helices α2 and α3 were mostly influenced by (and therefore involved in) DNA-binding. However, to specify the DNA-EIN3/EIL protein

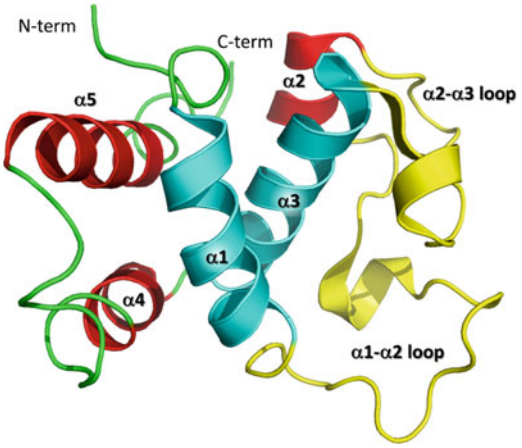


Fig. 10.16 EIL3 DNA-binding region (1WIJ). Secondary structure elements labeled. Probable DNA-interaction regions are shown in cyan (helices) and yellow (loops), respectively

recognition interface and to reveal the DNA-recognition mechanism, determination of the DNA-protein complex structure will be necessary.

10.4 Cytokinin/Ethylene Crosstalk

The canonical CTR1/EIN2/EIN3 signaling pathway is the dominant signaling pathway in ethylene perception in *Arabidopsis* and likely in most of the (higher) plants. Nevertheless, the facts that *ctr1* loss-of-function mutants display residual ethylene-insensitive (Larsen and Chang 2001) and that ethylene-insensitive dominant receptor mutants partially rescue the *ctr1* mutation (Hua and Meyerowitz 1998; Hua et al. 1998), strengthen the existence of an alternative ethylene responses pathways that bypasses CTR1-regulated pathway. Accumulating evidence [reviewed by (Zdarska et al. 2015)] indicates signaling crosstalk between ethylene receptors and MSP pathway. ETR1 seems to be one of the best candidates in mediating the crosstalk, based on its HK activity and the presence of all the domains necessary for mediating phosphorelay signaling. In vitro ETR1 in its functional dimeric state directly interacts with AHP1 and the ETR1-AHP1 complex formation seems to be dependent on the phosphorylation status of both proteins (Scharein and Groth 2011). Truncated ETR1 lacking the C-terminal RD does not interact with AHP1, suggesting the importance of RD for this interaction (Scharein et al. 2008). Yeast two-hybrid analysis showed that ETR1 also interacts with AHP1 and AHP3 (Urao et al. 2000). ARR2, member of the ARR-B family, was found to be phosphorylated in ETR1-dependent way and proposed to act downstream of ETR1 in the ethylene signaling (Hass et al. 2004). HK-deficient ETR1 receptor lines as well as mutations in another ARR-B ARR1 reduce sensitivity to ethylene in the ethylene-mediated inhibition of cell proliferation in the *Arabidopsis* root apical meristem (Street et al. 2015). In line with those findings, our recent unpublished data suggest an important role of HK activity of ETR1-mediated ethylene signaling in the control of MSP, integrating thus cytokinin and ethylene signaling into a single MSP pathway. Importantly, our data also clearly show spatiotemporal specificity of cytokinin- and ethylene-induced MSP activity, suggesting the existence of a mechanism allowing to distinguish the origin of both signal types (Zdarska et al., manuscript in preparation).

10.5 Conclusions and Perspectives

Our current insights into structural aspects of MSP-mediated cytokinin and ethylene signaling in plants are limited to the knowledge of several structures of individual MSP domains, being represented mostly by a single structure for each signaling module (except HPt proteins and HK_{RD}s). Although the overall structural fold seems to be highly conserved, even small differences can lead to important differences in

their ability to translate both (and possibly other) signals to specific responses. For instance, the ability to interact with and to be phosphorylated by individual sensor histidine kinases seems to be highly dependent on structural characteristics of individual HPt proteins in *Arabidopsis* (Degtjarik et al. unpublished). High-resolution crystal structures of individual signaling components and their complexes with upstream and downstream partners will provide us with essential information about the recognition mode in frame of individual signaling events and will shed light on the mechanism of maintaining specificity and mediating signal integration. However, the main gap in the molecular characterization of cytokinin and ethylene signaling pathways remains the structural characterization of full-length receptors in their non-activated and activated forms, and hence the mechanism of conformational changes involved in the transmission of signals across membranes and the activation of receptor cytosolic domains that further leads to activation of downstream signaling cascade. For example, in the case of canonical ethylene signaling, the key step, that is, ethylene-mediated inactivation of CTR1 or topology of ER-located cytokinin receptors and its functional importance in cytokinin signaling, still remains a complete black box. Finally yet importantly, to be able to understand MSP- and CTR1-mediated cytokinin and ethylene signaling in its specificity and complexity, it is essential to structurally characterize members of these pathways in different plant species. Development of novel approaches in the recent structural biology including the CryoEM tomography of receptors and their complexes in plant cells (see Appendix 3) seem to be very promising.

Acknowledgments This work was supported by the Czech Science Foundation, grant 13-25280S, and by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601) and CZ.02.1.01/0.0/0.0/16_026/0008446.

References

Abeles FB, Morgan PW, Saltveit ME (1992) Ethylene in plant biology, 2nd edn. Academic Press, San Diego

Alonso JM, Hirayama T, Roman G, Nourizadeh S, Ecker JR (1999) EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. Science 284(5423):2148–2152

Alonso JM, Stepanova AN, Solano R, Wisman E, Ferrari S, Ausubel FM, Ecker JR (2003) Five components of the ethylene-response pathway identified in a screen for weak ethylene-insensitive mutants in Arabidopsis. Proc Natl Acad Sci U S A 100(5):2992–2997

An F, Zhao Q, Ji Y, Li W, Jiang Z, Yu X, Zhang C, Han Y, He W, Liu Y, Zhang S, Ecker JR, Guo H (2010) Ethylene-induced stabilization of ETHYLENE INSENSITIVE3 and EIN3-LIKE1 is mediated by proteasomal degradation of EIN3 binding F-box 1 and 2 that requires EIN2 in Arabidopsis. Plant Cell 22(7):2384–2401

Anantharaman V, Aravind L (2001) The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem Sci 26 (10):579–582

Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408(6814):796–815

Bae E, Bitto E, Bingman CA, Allard ST, Wesenberg GE, Wrobel RL, Fox BG, Phillips GN (2010) Crystal structure of a putative histidine-containing phosphotransfer protein from Oryza saliva. J Korean Soc Appl Biol Chem 53(6):852–856

Bauer J, Reiss K, Veerabagu M, Heunemann M, Harter K, Stehle T (2013) Structure-function analysis of Arabidopsis thaliana histidine kinase AHK5 bound to its cognate phosphotransfer protein AHP1. Mol Plant 6(3):959–970

Bell CH, Porter SL, Strawson A, Stuart DI, Armitage JP (2010) Using structural information to change the phosphotransfer specificity of a two-component chemotaxis signalling complex. PLoS Biol 8(2):e1000306

Binder BM, O'Malley RC, Wang W, Moore JM, Parks BM, Spalding EP, Bleecker AB (2004) Arabidopsis seedling growth response and recovery to ethylene. A kinetic analysis. Plant Physiol 136(2):2913–2920

Binder BM, Walker JM, Gagne JM, Emborg TJ, Hemmann G, Bleecker AB, Vierstra RD (2007) The Arabidopsis EIN3 binding F-box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling. Plant Cell 19(2):509–523

Binder BM, Chang C, Schaller GE (2012) Perceptione of ethylene by plants: ethylene receptors. In: McManus MT (ed) Annual plant reviews, the plant hormone ethylene, vol 44. Wiley-Blackwell, Oxford, pp 117–145

Bisson MM, Groth G (2010) New insight in ethylene signaling: autokinase activity of ETR1 modulates the interaction of receptors and EIN2. Mol Plant 3(5):882–889

Bisson MM, Groth G (2011) New paradigm in ethylene signaling: EIN2, the central regulator of the signaling pathway, interacts directly with the upstream receptors. Plant Signal Behav 6 (1):164–166

Bisson MM, Bleckmann A, Allekotte S, Groth G (2009) EIN2, the central regulator of ethylene signalling, is localized at the ER membrane where it interacts with the ethylene receptor ETR1. Biochem J 424(1):1–6

Bleecker AB, Kende H (2000) Ethylene: a gaseous signal molecule in plants. Annu Rev Cell Dev Biol 16:1–18

Bleecker AB, Estelle MA, Somerville C, Kende H (1988) Insensitivity to ethylene conferred by a dominant mutation in Arabidopsis thaliana. Science 241(4869):1086–1089

Bourret RB (2010) Receiver domain structure and function in response regulator proteins. Curr Opin Microbiol 13(2):142–149

Caesar K, Thamm AM, Witthoft J, Elgass K, Huppenberger P, Grefen C, Horak J, Harter K (2011) Evidence for the localization of the Arabidopsis cytokinin receptors AHK3 and AHK4 in the endoplasmic reticulum. J Exp Bot 62(15):5571–5580

Capra EJ, Perchuk BS, Lubin EA, Ashenberg O, Skerker JM, Laub MT (2010) Systematic dissection and trajectory-scanning mutagenesis of the molecular Interface that ensures specificity of two-component signaling pathways. PLoS Genet 6(11):e1001220

Casino P, Rubio V, Marina A (2009) Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell 139(2):325–336

Casino P, Miguel-Romero L, Marina A (2014) Visualizing autophosphorylation in histidine kinases. Nat Commun 5:3258

Chang C, Kwok SF, Bleecker AB, Meyerowitz EM (1993) Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. Science 262(5133):539–544

Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, Huang SS, Schmitz RJ, Urich MA, Kuo D, Nery JR, Qiao H, Yang A, Jamali A, Chen H, Ideker T, Ren B, Bar-Joseph Z, Hughes TR, Ecker JR (2013) Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. elife 2:e00675

Chao QM, Rothenberg M, Solano R, Roman G, Terzaghi W, Ecker JR (1997) Activation of the ethylene gas response pathway in Arabidopsis by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. Cell 89(7):1133–1144

Chen YF, Randlett MD, Findell JL, Schaller GE (2002) Localization of the ethylene receptor ETR1 to the endoplasmic reticulum of Arabidopsis. J Biol Chem 277(22):19861–19866

Chen YF, Gao Z, Kerris RJ 3rd, Wang W, Binder BM, Schaller GE (2010) Ethylene receptors function as components of high-molecular-mass protein complexes in Arabidopsis. PLoS One 5 (1):e8640

Cho YH, Yoo SD (2015) Novel connections and gaps in ethylene signaling from the ER membrane to the nucleus. Front Plant Sci 5:733

Clark KL, Larsen PB, Wang X, Chang C (1998) Association of the Arabidopsis CTR1 Raf-like kinase with the ETR1 and ERS ethylene receptors. Proc Natl Acad Sci U S A 95(9):5401–5406

Cutcliffe JW, Hellmann E, Heyl A, Rashotte AM (2011) CRFs form protein-protein interactions with each other and with members of the cytokinin signalling pathway in Arabidopsis via the CRF domain. J Exp Bot 62(14):4995–5002

D'Agostino IB, Deruere J, Kieber JJ (2000) Characterization of the response of the Arabidopsis response regulator gene family to cytokinin. Plant Physiol 124(4):1706–1717

Degtjarik O, Dopitova R, Puehringer S, Nejedla E, Kuty M, Weiss MS, Hejatko J, Janda L, Kuta Smatanova I (2013) Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of AHP2, a signal transmitter protein from Arabidopsis thaliana. Acta Crystallogr Sect F Struct Biol Cryst Commun 69(Pt 2):158–161

Degtjarik O, Dopitova R, Reha D, Puehringer S, Otrusinova O, Pekarova B, Szmitkowska A, Hrdinova V, Valkova M, Kuty M, Jayasree A, Weiss MS, Janda L, Kuta-Smatanova I, Zidek L, Hejatko J. Structural insights into the specificity of sensory histidine kinase signaling in Eukaryotes (unpublished)

Dortay H, Mehnert N, Burkle L, Schmulling T, Heyl A (2006) Analysis of protein interactions within the cytokinin-signaling pathway of Arabidopsis thaliana. FEBS J 273(20):4631–4644

Ecker JR (2004) Reentry of the ethylene MPK6 module. Plant Cell 16(12):3169–3173

Feher VA, Zapf JW, Hoch JA, Whiteley JM, McIntosh LP, Rance M, Skelton NJ, Dahlquist FW, Cavanagh J (1997) High-resolution NMR structure and backbone dynamics of the Bacillus subtilis response regulator, Spo0F: implications for phosphorylation and molecular recognition. Biochemistry 36(33):10015–10025

Gagne JM, Smalle J, Gingerich DJ, Walker JM, Yoo SD, Yanagisawa S, Vierstra RD (2004) Arabidopsis EIN3-binding F-box 1 and 2 form ubiquitin-protein ligases that repress ethylene action and promote growth by directing EIN3 degradation. Proc Natl Acad Sci U S A 101 (17):6803–6808

Gajdosova S, Spichal L, Kaminek M, Hoyerova K, Novak O, Dobrev PI, Galuszka P, Klima P, Gaudinova A, Zizkova E, Hanus J, Dancak M, Travnicek B, Pesek B, Krupicka M, Vankova R, Strnad M, Motyka V (2011) Distribution, biological activities, metabolism, and the conceivable function of cis-zeatin-type cytokinins in plants. J Exp Bot 62(8):2827–2840

Gao R, Stock AM (2009) Biological insights from structures of two-component proteins. Annu Rev Microbiol 63:133–154

Gao Z, Chen YF, Randlett MD, Zhao XC, Findell JL, Kieber JJ, Schaller GE (2003) Localization of the Raf-like kinase CTR1 to the endoplasmic reticulum of Arabidopsis through participation in ethylene receptor signaling complexes. J Biol Chem 278(36):34725–34732

Gao Z, Wen CK, Binder BM, Chen YF, Chang J, Chiang YH, Kerris RJ 3rd, Chang C, Schaller GE (2008) Heteromeric interactions among ethylene receptors mediate signaling in Arabidopsis. J Biol Chem 283(35):23801–23810

Gattolin S, Alandete-Saez M, Elliott K, Gonzalez-Carranza Z, Naomab E, Powell C, Roberts JA (2006) Spatial and temporal expression of the response regulators ARR22 and ARR24 in Arabidopsis thaliana. J Exp Bot 57(15):4225–4233

Grefen C, Stadele K, Ruzicka K, Obrdlik P, Harter K, Horak J (2008) Subcellular localization and in vivo interactions of the Arabidopsis thaliana ethylene receptor family members. Mol Plant 1 (2):308–320

Gruhn N, Heyl A (2013) Updates on the model and the evolution of cytokinin signaling. Curr Opin Plant Biol 16(5):569–574

Guo H, Ecker JR (2003) Plant responses to ethylene gas are mediated by SCF(EBF1/EBF2)-dependent proteolysis of EIN3 transcription factor. Cell 115(6):667–677

Guo H, Ecker JR (2004) The ethylene signaling pathway: new insights. Curr Opin Plant Biol 7 (1):40–49

Hall AE, Bleecker AB (2003) Analysis of combinatorial loss-of-function mutants in the Arabidopsis ethylene receptors reveals that the ers1 etr1 double mutant has severe developmental defects that are EIN2 dependent. Plant Cell 15(9):2032–2041

Hall AE, Chen QG, Findell JL, Schaller GE, Bleecker AB (1999) The relationship between ethylene binding and dominant insensitivity conferred by mutant forms of the ETR1 ethylene receptor. Plant Physiol 121(1):291–300

Hall BP, Shakeel SN, Amir M, Ul Haq N, Qu X, Schaller GE (2012) Histidine kinase activity of the ethylene receptor ETR1 facilitates the ethylene response in Arabidopsis. Plant Physiol 159 (2):682–695

Hass C, Lohrmann J, Albrecht V, Sweere U, Hummel F, Yoo SD, Hwang I, Zhu T, Schafer E, Kudla J, Harter K (2004) The response regulator 2 mediates ethylene signalling and hormone signal integration in Arabidopsis. EMBO J 23(16):3290–3302

Heyl A, Wulfetange K, Pils B, Nielsen N, Romanov GA, Schmulling T (2007) Evolutionary proteomics identifies amino acids essential for ligand-binding of the cytokinin receptor CHASE domain. BMC Evol Biol 7:62

Heyl A, Riefler M, Romanov GA, Schmulling T (2012) Properties, functions and evolution of cytokinin receptors. Eur J Cell Biol 91(4):246–256

Higuchi M, Pischke MS, Mahonen AP, Miyawaki K, Hashimoto Y, Seki M, Kobayashi M, Shinozaki K, Kato T, Tabata S, Helariutta Y, Sussman MR, Kakimoto T (2004) In planta functions of the Arabidopsis cytokinin receptor family. Proc Natl Acad Sci U S A 101 (23):8821–8826

Hirayama T, Kieber JJ, Hirayama N, Kogan M, Guzman P, Nourizadeh S, Alonso JM, Dailey WP, Dancis A, Ecker JR (1999) RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in Arabidopsis. Cell 97(3):383–393

Horak J, Janda L, Pekarova B, Hejatko J (2011) Molecular mechanisms of Signalling specificity via Phosphorelay pathways in Arabidopsis. Curr Protein Pept Sci 12(2):126–136

Hosoda K, Imamura A, Katoh E, Hatta T, Tachiki M, Yamada H, Mizuno T, Yamazaki T (2002) Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators. Plant Cell 14(9):2015–2029

Hothorn M, Dabi T, Chory J (2011) Structural basis for cytokinin recognition by Arabidopsis thaliana histidine kinase 4. Nat Chem Biol 7(11):766–768

Hua J, Meyerowitz EM (1998) Ethylene responses are negatively regulated by a receptor gene family in Arabidopsis thaliana. Cell 94(2):261–271

Hua J, Chang C, Sun Q, Meyerowitz EM (1995) Ethylene insensitivity conferred by Arabidopsis ERS gene. Science 269(5231):1712–1714

Hua J, Sakai H, Nourizadeh S, Chen QG, Bleecker AB, Ecker JR, Meyerowitz EM (1998) EIN4 and ERS2 are members of the putative ethylene receptor gene family in Arabidopsis. Plant Cell 10 (8):1321–1332

Huang Y, Li H, Hutchison CE, Laskey J, Kieber JJ (2003) Biochemical and functional analysis of CTR1, a protein kinase that negatively regulates ethylene signaling in Arabidopsis. Plant J 33 (2):221–233

Hung YL, Jiang I, Lee YZ, Wen CK, Sue SC (2016) NMR study reveals the receiver domain of Arabidopsis ETHYLENE RESPONSE1 Ethylene receptor as an atypical type RESPONSE regulator. PLoS One 11(8):e0160598

Hutchison CE, Kieber JJ (2007) Signaling via histidine-containing phosphotransfer proteins in Arabidopsis. Plant Signal Behav 2(4):287–289

Hutchison CE, Li J, Argueso C, Gonzalez M, Lee E, Lewis MW, Maxwell BB, Perdue TD, Schaller GE, Alonso JM, Ecker JR, Kieber JJ (2006) The Arabidopsis histidine phosphotransfer proteins are redundant positive regulators of cytokinin signaling. Plant Cell 18(11):3073–3087

Hwang I, Sheen J (2001) Two-component circuitry in Arabidopsis cytokinin signal transduction. Nature 413(6854):383–389

Imamura A, Hanaki N, Nakamura A, Suzuki T, Taniguchi M, Kiba T, Ueguchi C, Sugiyama T, Mizuno T (1999) Compilation and characterization of Arabidopsis thaliana response regulators implicated in His-Asp phosphorelay signal transduction. Plant Cell Physiol 40(7):733–742

Inoue T, Higuchi M, Hashimoto Y, Seki M, Kobayashi M, Kato T, Tabata S, Shinozaki K, Kakimoto T (2001) Identification of CRE1 as a cytokinin receptor from Arabidopsis. Nature 409(6823):1060–1063

Janiak-Spens F, West AH (2000) Functional roles of conserved amino acid residues surrounding the phosphorylatable histidine of the yeast phosphorelay protein YPD1. Mol Microbiol 37 (1):136–144

Ju C, Chang C (2015) Mechanistic insights in ethylene perception and signal transduction. Plant Physiol 169(1):85–95

Ju C, Yoon GM, Shemansky JM, Lin DY, Ying ZI, Chang J, Garrett WM, Kessenbrock M, Groth G, Tucker ML, Cooper B, Kieber JJ, Chang C (2012) CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in Arabidopsis. Proc Natl Acad Sci U S A 109(47):19486–19491

Kato M, Mizuno T, Shimizu T, Hakoshima T (1997) Insights into multistep phosphorelay from the crystal structure of the C-terminal HPt domain of ArcB. Cell 88(5):717–723

Kato M, Shimizu T, Mizuno T, Hakoshima T (1999) Structure of the histidine-containing phosphotransfer (HPt) domain of the anaerobic sensor protein ArcB complexed with the chemotaxis response regulator CheY. Acta Crystallogr D Biol Crystallogr 55:1257–1263

Kiba T, Aoki K, Sakakibara H, Mizuno T (2004) Arabidopsis response regulator, ARR22, ectopic expression of which results in phenotypes similar to the wol cytokinin-receptor mutant. Plant Cell Physiol 45(8):1063–1077

Kieber JJ, Schaller GE (2014) Cytokinins. The Arabidopsis Book 12:e0168

Kieber JJ, Rothenberg M, Roman G, Feldmann KA, Ecker JR (1993) CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases. Cell 72(3):427–441

Kim HJ, Ryu H, Hong SH, Woo HR, Lim PO, Lee IC, Sheen J, Nam HG, Hwang I (2006) Cytokinin-mediated control of leaf longevity by AHK3 through phosphorylation of ARR2 in Arabidopsis. Proc Natl Acad Sci U S A 103(3):814–819

Knaggs MH, Salsbury FR Jr, Edgell MH, Fetrow JS (2007) Insights into correlated motions and long-range interactions in CheY derived from molecular dynamics simulations. Biophys J 92 (6):2062–2079

Konishi M, Yanagisawa S (2008) Ethylene signaling in Arabidopsis involves feedback regulation via the elaborate control of EBF2 expression by EIN3. Plant J 55(5):821–831

Kosugi S, Ohashi Y (2000) Cloning and DNA-binding properties of a tobacco Ethylene-Insensitive3 (EIN3) homolog. Nucleic Acids Res 28(4):960–967

Kuderova A, Gallova L, Kuricova K, Nejedla E, Curdova A, Micenkova L, Plihal O, Smajs D, Spichal L, Hejatko J (2015) Identification of AHK2- and AHK3-like cytokinin receptors in Brassica napus reveals two subfamilies of AHK2 orthologues. J Exp Bot 66(1):339–353

Larsen PB, Chang C (2001) The Arabidopsis eer1 mutant has enhanced ethylene responses in the hypocotyl and stem. Plant Physiol 125(2):1061–1073

Li J, Li Z, Tang L, Yang Y, Zouine M, Bouzayen M (2012) A conserved phosphorylation site regulates the transcriptional function of ETHYLENE-INSENSITIVE3-like1 in tomato. J Exp Bot 63(1):427–439

Li W, Ma M, Feng Y, Li H, Wang Y, Ma Y, Li M, An F, Guo H (2015) EIN2-directed translational regulation of ethylene signaling in Arabidopsis. Cell 163(3):670–683

Light KM, Wisniewski JA, Vinyard WA, Kieber-Emmons MT (2016) Perception of the plant hormone ethylene: known-knowns and known-unknowns. J Biol Inorg Chem 21(5–6):715–728

Lomin SN, Yonekura-Sakakibara K, Romanov GA, Sakakibara H (2011) Ligand-binding properties and subcellular localization of maize cytokinin receptors. J Exp Bot 62(14):5149–5159

Lomin SN, Krivosheev DM, Steklov MY, Arkhipov DV, Osolodkin DI, Schmulling T, Romanov GA (2015) Plant membrane assays with cytokinin receptors underpin the unique role of free cytokinin bases as biologically active ligands. J Exp Bot 66(7):1851–1863

Mahonen AP, Bonke M, Kauppinen L, Riikonen M, Benfey PN, Helariutta Y (2000) A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes Dev 14(23):2938–2943

Mahonen AP, Bishopp A, Higuchi M, Nieminen KM, Kinoshita K, Tormakangas K, Ikeda Y, Oka A, Kakimoto T, Helariutta Y (2006a) Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. Science 311(5757):94–98

Mahonen AP, Higuchi M, Tormakangas K, Miyawaki K, Pischke MS, Sussman MR, Helariutta Y, Kakimoto T (2006b) Cytokinins regulate a bidirectional phosphorelay network in Arabidopsis. Curr Biol 16(11):1116–1122

Marina A, Mott C, Auyzenberg A, Hendrickson WA, Waldburger CD (2001) Structural and mutational analysis of the PhoQ histidine kinase catalytic domain: INSIGHT INTO THE REACTION MECHANISM. J Biol Chem 276(44):41182–41190

Matsushika A, Mizuno T (1998) The structure and function of the histidine-containing phosphotransfer (HPt) signaling domain of the Escherichia coli ArcB sensor. J Biochem 124 (2):440–445

Mayerhofer H, Panneerselvam S, Mueller-Dieckmann J (2012) Protein kinase domain of CTR1 from Arabidopsis thaliana promotes ethylene receptor cross talk. J Mol Biol 415(4):768–779

Mayerhofer H, Panneerselvam S, Kaljunen H, Tuukkanen A, Mertens HD, Mueller-Dieckmann J (2015) Structural model of the cytosolic domain of the plant ethylene receptor 1 (ETR1). J Biol Chem 290(5):2644–2658

Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enriquez P, Binder BM, Heber S, Stepanova AN, Alonso JM (2015) Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. Cell 163(3):684–697

Mieczkowski C, Iavarone AT, Alber T (2008) Auto-activation mechanism of the Mycobacterium tuberculosis PknB receptor Ser/Thr kinase. EMBO J 27(23):3186–3197

Miller CO, Skoog F, Vonsaltza MH, Strong FM (1955) Kinetin, a cell division factor from deoxyribonucleic acid. J Am Chem Soc 77(5):1392–1392

Miwa K, Ishikawa K, Terada K, Yamada H, Suzuki T, Yamashino T, Mizuno T (2007) Identification of amino acid substitutions that render the Arabidopsis cytokinin receptor histidine kinase AHK4 constitutively active. Plant Cell Physiol 48(12):1809–1814

Mok DWS, Mok MC (2001) Cytokinin metabolism and action. Annu Rev Plant Physiol Plant Mol Biol 52:89–118

Mougel C, Zhulin IB (2001) CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem Sci 26(10):582–584

Moussatche P, Klee HJ (2004) Autophosphorylation activity of the Arabidopsis ethylene receptor multigene family. J Biol Chem 279(47):48734–48741

Muller-Dieckmann HJ, Grantz AA, Kim SH (1999) The structure of the signal receiver domain of the Arabidopsis thaliana ethylene receptor ETR1. Structure 7(12):1547–1556

Nishimura C, Ohashi Y, Sato S, Kato T, Tabata S, Ueguchi C (2004) Histidine kinase homologs that act as cytokinin receptors possess overlapping functions in the regulation of shoot and root growth in Arabidopsis. Plant Cell 16(6):1365–1377

Nolen B, Taylor S, Ghosh G (2004) Regulation of protein kinases; controlling activity through activation segment conformation. Mol Cell 15(5):661–675

Ocasio VJ, Correa F, Gardner KH (2015) Ligand-induced folding of a two-component signaling receiver domain. Biochemistry 54(6):1353–1363

Otrusinova O, Demo G, Padrta P, Jasenakova Z, Pekarova B, Gelova Z, Szmitkowska A, Kaderavek P, Jansen S, Zachrdla M, Klumpler T, Marek J, Hritz J, Janda L, Iwai H, Wimmerova M, Hejatko J, Zidek L (2017) Conformational dynamics as a key factor of signaling mediated by the receiver domain of sensor histidine kinase from Arabidopsis thaliana. J Biol Chem. https://doi.org/10.1074/jbc.M117.790212

Pekarova B, Klumpler T, Triskova O, Horak J, Jansen S, Dopitova R, Borkovcova P, Papouskova V, Nejedla E, Sklenar V, Marek J, Zidek L, Hejatko J, Janda L (2011) Structure and binding specificity of the receiver domain of sensor histidine kinase CKI1 from Arabidopsis thaliana. Plant J 67(5):827–839

Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, Genschik P (2003) EIN3-dependent regulation of plant ethylene hormone signaling by two arabidopsis F box proteins: EBF1 and EBF2. Cell 115(6):679–689

Punwani JA, Hutchison CE, Eric Schaller G, Kieber JJ (2010) The subcellular distribution of the Arabidopsis histidine phosphotransfer proteins is independent of cytokinin signaling. Plant J 62:473–482

Qiao H, Chang KN, Yazaki J, Ecker JR (2009) Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis. Genes Dev 23 (4):512–521

Qiao H, Shen Z, Huang SS, Schmitz RJ, Urich MA, Briggs SP, Ecker JR (2012) Processing and subcellular trafficking of ER-tethered EIN2 control response to ethylene gas. Science 338 (6105):390–393

Qu X, Hall BP, Gao Z, Schaller GE (2007) A strong constitutive ethylene-response phenotype conferred on Arabidopsis plants containing null mutations in the ethylene receptors ETR1 and ERS1. BMC Plant Biol 7:3

Rajakulendran T, Sahmi M, Lefrancois M, Sicheri F, Therrien M (2009) A dimerization-dependent mechanism drives RAF catalytic activation. Nature 461(7263):542–545

Rashotte AM, Mason MG, Hutchison CE, Ferreira FJ, Schaller GE, Kieber JJ (2006) A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. Proc Natl Acad Sci U S A 103(29):11081–11085

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290(5499):2105–2110

Rodriguez FI, Esch JJ, Hall AE, Binder BM, Schaller GE, Bleecker AB (1999) A copper cofactor for the ethylene receptor ETR1 from Arabidopsis. Science 283(5404):996–998

Romanov GA, Spichal L, Lomin SN, Strnad M, Schmulling T (2005) A live cell hormone-binding assay on transgenic bacteria expressing a eukaryotic receptor protein. Anal Biochem 347 (1):129–134

Romanov GA, Lomin SN, Schmulling T (2006) Biochemical characteristics and ligand-binding properties of Arabidopsis cytokinin receptor AHK3 compared to CRE1/AHK4 as revealed by a direct binding assay. J Exp Bot 57(15):4051–4058

Ruszkowski M, Brzezinski K, Jedrzejczak R, Dauter M, Dauter Z, Sikorski M, Jaskolski M (2013) Medicago truncatula histidine-containing phosphotransfer protein: structural and biochemical insights into the cytokinin transduction pathway in plants. FEBS J 280(15):3709–3720

Sakai H, Hua J, Chen QG, Chang C, Medrano LJ, Bleecker AB, Meyerowitz EM (1998) ETR2 is an ETR1-like gene involved in ethylene signaling in Arabidopsis. Proc Natl Acad Sci U S A 95 (10):5812–5817

Sakakibara H (2006) Cytokinins: activity, biosynthesis, and translocation. Annu Rev Plant Biol 57:431–449

Salome PA, To JP, Kieber JJ, McClung CR (2006) Arabidopsis response regulators ARR3 and ARR4 play cytokinin-independent roles in the control of circadian period. Plant Cell 18 (1):55–69

Schaller GE, Ladd AN, Lanahan MB, Spanbauer JM, Bleecker AB (1995) The ethylene response mediator ETR1 from Arabidopsis forms a disulfide-linked dimer. J Biol Chem 270 (21):12526–12530

Schaller GE, Shiu SH, Armitage JP (2011) Two-component systems and their co-option for eukaryotic signal transduction. Curr Biol 21(9):R320–R330

Scharein B, Groth G (2011) Phosphorylation alters the interaction of the Arabidopsis phosphotransfer protein AHP1 with its sensor kinase ETR1. PLoS One 6(9):e24173

Scharein B, Voet-van-Vormizeele J, Harter K, Groth G (2008) Ethylene signaling: identification of a putative ETR1-AHP1 phosphorelay complex by fluorescence spectroscopy. Anal Biochem 377(1):72–76

Shakeel SN, Wang X, Binder BM, Schaller GE (2012) Mechanisms of signal transduction by ethylene: overlapping and non-overlapping signalling roles in a receptor family. AoB Plants 5: plt010

Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT (2008) Rewiring the specificity of two-component signal transduction systems. Cell 133(6):1043–1054

Solano R, Stepanova A, Chao Q, Ecker JR (1998) Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. Genes Dev 12(23):3703–3714

Spichal L (2012) Cytokinins - recent news and views of evolutionally old molecules. Funct Plant Biol 39(4):267–284

Spichal L, Rakova NY, Riefler M, Mizuno T, Romanov GA, Strnad M, Schmulling T (2004) Two cytokinin receptors of Arabidopsis thaliana, CRE1/AHK4 and AHK3, differ in their ligand specificity in a bacterial assay. Plant Cell Physiol 45(9):1299–1305

Steklov MY, Lomin SN, Osolodkin DI, Romanov GA (2013) Structural basis for cytokinin receptor signaling: an evolutionary approach. Plant Cell Rep 32(6):781–793

Stolz A, Riefler M, Lomin SN, Achazi K, Romanov GA, Schmulling T (2011) The specificity of cytokinin signalling in Arabidopsis thaliana is mediated by differing ligand affinities and expression profiles of the receptors. Plant J 67(1):157–168

Street IH, Aman S, Zubo Y, Ramzan A, Wang X, Shakeel SN, Kieber JJ, Schaller GE (2015) Ethylene inhibits cell proliferation of the Arabidopsis root meristem. Plant Physiol 169 (1):338–350

Sugawara H, Kawano Y, Hatakeyama T, Yamaya T, Kamiya N, Sakakibara H (2005) Crystal structure of the histidine-containing phosphotransfer protein ZmHP2 from maize. Protein Sci 14 (1):202–208

Suzuki T, Miwa K, Ishikawa K, Yamada H, Aiba H, Mizuno T (2001a) The Arabidopsis sensor His-kinase, AHK4, can respond to cytokinins. Plant Cell Physiol 42(2):107–113

Suzuki T, Sakurai K, Ueguchi C, Mizuno T (2001b) Two types of putative nuclear factors that physically interact with histidine-containing phosphotransfer (Hpt) domains, signaling mediators in His-to-Asp phosphorelay, in Arabidopsis thaliana. Plant Cell Physiol 42(1):37–45

To JP, Kieber JJ (2008) Cytokinin signaling: two-components and more. Trends Plant Sci 13 (2):85–92

To JP, Haberer G, Ferreira FJ, Deruere J, Mason MG, Schaller GE, Alonso JM, Ecker JR, Kieber JJ (2004) Type-A Arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling. Plant Cell 16(3):658–671

Ueguchi C, Koizumi H, Suzuki T, Mizuno T (2001) Novel family of sensor histidine kinase genes in Arabidopsis thaliana. Plant Cell Physiol 42(2):231–235

Urao T, Miyata S, Yamaguchi-Shinozaki K, Shinozaki K (2000) Possible His to Asp phosphorelay signaling in an Arabidopsis two-component system. FEBS Lett 478(3):227–232

Verma V, Sivaraman J, Srivastava AK, Sadanandom A, Kumar PP (2015) Destabilization of interaction between cytokinin signaling intermediates AHP1 and ARR4 modulates Arabidopsis development. New Phytol 206(2):726–737

Volz K (1993) Structural conservation in the CheY superfamily. Biochemistry 32 (44):11741–11753

Wagner JR, Zhang J, Brunzelle JS, Vierstra RD, Forest KT (2007) High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. J Biol Chem 282(16):12298–12309

Wang S (2012) Bacterial two-component systems: structures and signaling mechanisms. In: Huang C (ed) Protein phosphorylation in human health. IntechOpen, London, United Kingdom, pp 339–464

Wang W, Hall AE, O'Malley R, Bleecker AB (2003) Canonical histidine kinase activity of the transmitter domain of the ETR1 ethylene receptor from Arabidopsis is not required for signal transmission. Proc Natl Acad Sci U S A 100(1):352–357

Wang W, Esch JJ, Shiu SH, Agula H, Binder BM, Chang C, Patterson SE, Bleecker AB (2006) Identification of important regions for ethylene binding and signaling in the transmembrane domain of the ETR1 ethylene receptor of Arabidopsis. Plant Cell 18(12):3429–3442

Wen X, Zhang C, Ji Y, Zhao Q, He W, An F, Jiang L, Guo H (2012) Activation of ethylene signaling is mediated by nuclear translocation of the cleaved EIN2 carboxyl terminus. Cell Res 22(11):1613–1616

Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, Chothia C, Gough J (2009) SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res 37(Database issue):D380–D386

Wulfetange K, Lomin SN, Romanov GA, Stolz A, Heyl A, Schmulling T (2011) The Cytokinin receptors of Arabidopsis thaliana are locating mainly to the endoplasmic reticulum. Plant Physiol 156(4):1808–1818

Xu Q, Porter SW, West AH (2003) The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems. Structure 11(12):1569–1581

Yamada H, Suzuki T, Terada K, Takei K, Ishikawa K, Miwa K, Yamashino T, Mizuno T (2001) The Arabidopsis AHK4 histidine kinase is a cytokinin-binding receptor that transduces cytokinin signals across the membrane. Plant Cell Physiol 42(9):1017–1023

Yamasaki K, Kigawa T, Inoue M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2005) Solution structure of the major DNA-binding domain of Arabidopsis thaliana ethylene-insensitive3-like3. J Mol Biol 348(2):253–264

Yoo SD, Cho YH, Tena G, Xiong Y, Sheen J (2008) Dual control of nuclear EIN3 by bifurcate MAPK cascades in C2H4 signalling. Nature 451(7180):789–795

Zdarska M, Dobisova T, Gelova Z, Pernisova M, Dabravolski S, Hejatko J (2015) Illuminating light, cytokinin, and ethylene signalling crosstalk in plant development. J Exp Bot 66:4913–4931

Zhang W, To JP, Cheng CY, Eric Schaller G, Kieber JJ (2011) Type-A response regulators are required for proper root apical meristem function through the post-transcriptional regulation of PIN auxin efflux carriers. Plant J 68:1–10

Zhou YF, Nan B, Nan J, Ma Q, Panjikar S, Liang YH, Wang Y, Su XD (2008) C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. J Mol Biol 383 (1):49–61

Zurcher E, Muller B (2016) Cytokinin synthesis, signaling, and function–advances and new insights. Int Rev Cell Mol Biol 324:1–38

Zurcher E, Liu J, di Donato M, Geisler M, Muller B (2016) Plant development regulated by cytokinin sinks. Science 353(6303):1027–1030

Part IV
Method Guidance

Chapter 11
X-Ray Crystallography

Oksana Degtjarik, Gabriel Demo, Michaela Wimmerova, and Ivana Kuta Smatanova

11.1 Introduction

X-ray crystallography is a powerful experimental method for studying the three-dimensional (3D) structure of proteins and their ensembles, providing the most complete information about the spatial location of every atom in the molecule. Proteins perform all essential functions in the living cell—signaling, regulatory, transport, protective, and structural functions. In higher plants proteins play a leading role in sensing and reaction to various abiotic (drought, salinity, extreme temperatures) and biotic (pathogen attack) stressors. Knowing the 3D protein structure enables us to understand how proteins perform their molecular functions in the

O. Degtjarik
Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

Institute of Chemistry and Biochemistry, University of South Bohemia, České Budějovice, Czech Republic

G. Demo
CEITEC -Central European Institute of Technology, Masaryk University, Brno, Czech Republic

RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA, USA

M. Wimmerova
CEITEC -Central European Institute of Technology, Masaryk University, Brno, Czech Republic

I. Kuta Smatanova (✉)
Institute of Chemistry and Biochemistry, University of South Bohemia, České Budějovice, Czech Republic

Center for Nanobiology and Structural Biology of the Institute of Microbiology, Academy of Sciences of the Czech Republic, Nove Hrady, Czech Republic
e-mail: ivanaks@seznam.cz

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_11

cell. This information, in turn, opens opportunities for the rational modification of plants in order to increase their resilience to various stressors and enhance their productivity. Despite the use of rapidly developing structural biology methods such as cryo-electron microscopy and cryo-electron tomography, X-ray crystallography still plays a leading role in determining the structures of macromolecules—around 90% of the structures of macromolecules deposited to the Protein Data Bank (www.rcsb.org) were resolved just by this method. Of all the structures in the PDB databank only 3.1% belong to the subkingdom of higher plants, a proportion that is rather low considering the importance of plants for human beings.

X-ray crystallography is a complex method consisting of several stages (Fig. 11.1). The basic model for X-ray crystallography is to mount the crystal between an X-ray source and an X-ray detector. The crystal is so placed in an intense beam of X-rays (monochromatic X-rays). The beam of X-rays can strike the crystal and is scattered into many different directions. The detector is able to detect these scattered directions as 'dark spots', called reflections, which reflect the regular pattern from the planes of atoms in a crystal. The position of a reflection can be used to obtain the direction in which the particular beam was diffracted by the crystal. Simultaneously, the intensity of the recorded reflection (spot) from the diffracted beam is measured on the detector. The two parameters for each diffracted beam, beam intensity and direction, are computationally taken together with the physical, chemical, and biological information of the studied system to reconstruct the model of molecules in the unit cell of the crystal.

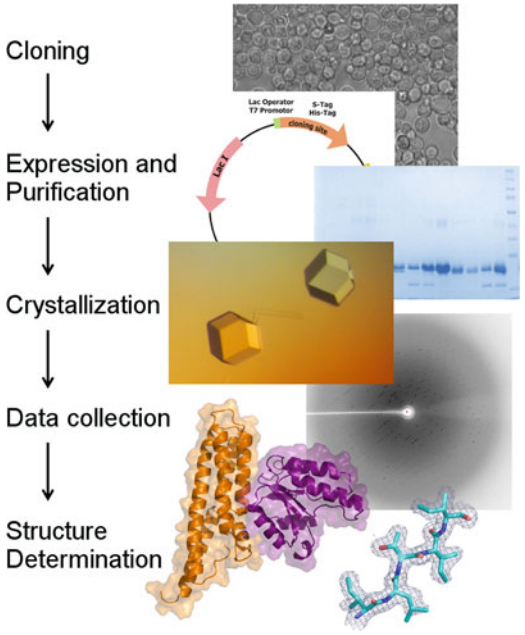


Fig. 11.1 Overview of protein structure determination stages. (Modified from Rupp 2010)

In this chapter we provide a short overview of the X-ray crystallography method used for the determination of the 3D structure of macromolecules, in order to encourage researchers to apply this method in current plant biology studies.

11.2 Crystal Geometry

A crystal is a 3D periodic arrangement of molecules organized in a crystal lattice. The molecules within a crystal are held together by weak non-covalent interactions. Protein crystals consist of repeated building blocks called *unit cells*, which contain all the structural and symmetry information (Fig. 11.2). The unit cell is defined by the length of the cell edges (a, b, and c) and the angles between them (α, β, and γ) (Rupp 2010).

The smallest unit of a crystal needed to reproduce the whole unit cell by applying crystallographic symmetry operations is called an *asymmetric unit* (Fig. 11.2). In protein crystallography, only rotation, translation, or screw axes (combination of rotation and translation) symmetry operations are allowed (Rhodes 2010).

The internal symmetry of the unit cell and its content is described by its *space group*. Space groups belong to seven crystal systems: triclinic, monoclinic, orthorhombic, trigonal, tetragonal, hexagonal, and cubic. In total, there are 230 possible space groups. However, biological macromolecules can crystallize in only 65 of them. Knowledge of the space group and dimensions of the unit cell is essential for the interpretation of diffraction data (Rhodes 2010).

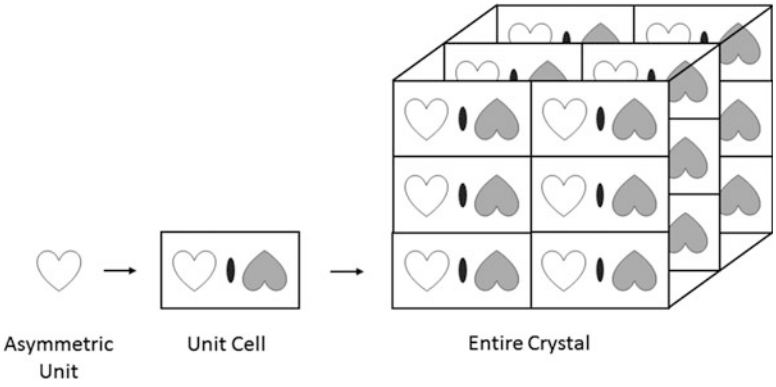


Fig. 11.2 Relationship between asymmetric unit, unit cell, and whole crystal. The asymmetric unit (white heart), by 180° rotation around the two-fold crystallographic symmetry axis (black oval), produces the second copy (grey heart). The two hearts together comprise the unit cell. A three-dimensional (3D) crystal is built by transitional repetitions of the unit cell in three directions

11.3 X-Ray Theory

In order for an object to diffract, the wavelength of the light is ideally required to be no larger than the studied object. X-rays are a form of electromagnetic radiation, of wavelengths 0.1–100 Å, which are able to produce an image of individual atoms in protein molecules, in which bonded atoms are only about 0.15 nm (1.5 Å) long.

X-rays can be obtained by bombarding a piece of metal (copper or molybdenum) with electrons produced by a heated cathode and accelerated by an electric field. Electrons with high energy can collide with electrons from lower energy orbitals and displace them from the orbitals in the metal atom. The vacant orbitals from a lower energy level are filled with electrons from higher energy orbitals, a process that results in the emission of energy as X-ray photons (conventional sources are X-ray tubes or rotating anodes). The energy of the emission has specific wavelengths (characteristic lines) defined by the characteristic energy-level spacing of the element. In the case of copper, the electrons dropping from the L shell to replace the displaced K electrons ($K\alpha$- transition) emit X-rays of wavelength 1.54 Å (Rhodes 2010).

The most powerful X-ray sources are represented by particle storage rings (synchrotrons), which produce synchrotron radiation in the X-ray region. In these rings, electrons or positrons circulate at velocities near the speed of light, driven by energy from radio frequency transmitters and maintained in circular motion by powerful magnets. A charged body such as an electron emits energy (synchrotron radiation) when forced into curved motion, and in accelerators, the energy is emitted as X-rays. Accessory devices called wigglers cause additional bending of the beam, thus increasing the intensity of radiation. Systems of focusing mirrors and monochromators tangential to the storage ring provide powerful monochromatic X-rays at selectable wavelengths. Whatever the source of the X-rays, the beam is directed through a collimator, a narrow metal tube that selects and reflects the X-rays into parallel paths, producing a narrow beam. After collimation, the beam diameter can be further reduced with systems of so-called focusing mirrors. The synchrotrons generally allow us to measure complete datasets of a small crystal (40 μm) in a short time—nowadays less than 5 min (Marek and Trávníček 2002).

The interaction of X-ray photons with a crystal lattice results in two possible outcomes: absorption of a photon by the atom in the lattice or photon scattering. Absorption causes the temperature of the atom to increase and can lead to a photoelectric effect or to fluorescence. In the case of scattering, the X-ray photons deflect (scatter) from their original route in many different directions (Blundell and Johnson 1976). The scattered waves can interfere with each other and when they meet in different phases they cancel each other out (destructive interference). When the scattered waves are in phase they interfere constructively and can produce a set of diffraction spots (reflections) on the detector if the so-called Bragg's condition is fulfilled:

$$n\lambda = 2d\sin\theta$$

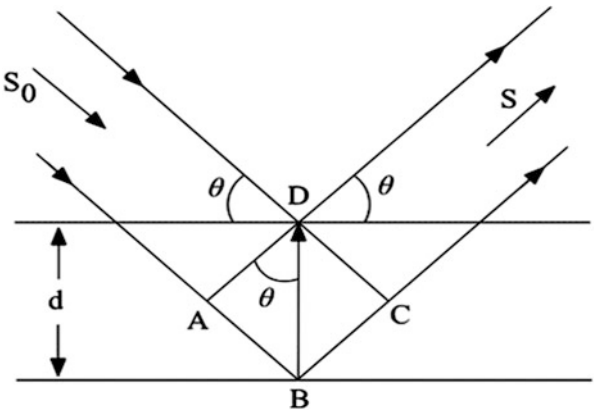


Fig. 11.3 Bragg's law geometrical description. A pair of incident waves (incident beam – $\vec{S_o}$) with the same wavelength approach two lattice planes separated by distance *d* and reflect (reflected beam – $\vec{S}$) with an angle θ. The difference in the tracks of the waves scattered in points *D* and *B* is equal to $AB + BC = 2d\sin\theta$

According to this condition, constructive interference occurs when the path length between the incident and reflected waves of each lattice plane differ by any multiple (n) of the incident wavelength (λ). The difference in the path length depends on the scattering angle (θ) of the incidence (and reflection) and distance (d) between the parallel planes in the crystal lattice (Fig. 11.3) (Giacovazzo 2002).

11.4 Protein Crystallization

Having a high-quality protein crystal is a necessary requirement for X-ray structure determination. Crystallization is the most critical step on the path to X-ray protein structure determination. To date, it has not been possible to predict the exact conditions that enable the formation of single crystals of a specific protein. Protein molecules by their nature are large, flexible, and irregularly shaped. Therefore, to get them assembled into a regular periodic crystal lattice is not an easy task. Despite the various methods developed during the past few years, protein crystallization still remains problematic and dependent on empirical approaches (McPherson 2001).

Protein crystals form a wide open network of macromolecules with a high solvent content, and this high solvent content is the major difference from the characteristics of inorganic and/or organometallic crystals (Bugg 1986; McPherson 1999). A protein crystal contains about 25–85% of solvent, represented mainly by water molecules. The protein-bound water forms the ordered hydration shell of the protein. The ordered water molecules play a role in the formation of hydrogen bonds that supplement and strengthen direct interactions between the protein molecules (Salemme et al. 1988). These interactions can be described as specific interactions, and occur together with electrostatic interactions (involving the amino acid side-chains and ions included in the solvent) (Salemme et al. 1988; Kierzek and Zielenkiewicz 2001). Van der Waals interactions and hydrophobic interactions are

considered to be nonspecific. All these mentioned interactions can contribute to protein structure stabilization, which can vary from case to case.

Crystallization is a complex process, involving multiple equilibria between different states of the crystallizing species. The main principle of protein crystallization is to induce supersaturation in the protein solution by manipulating physical and chemical parameters (protein concentration, pH, temperature, nature and concentration of the precipitant, and other features). The two main stages of crystallization are nucleation and crystal growth. During the first stage, in the supersaturated thermodynamically metastable state, protein molecules can overcome an energy barrier, separate from the solution, and self-assemble into a periodic 3D array. The second step, crystal growth, starts spontaneously after a nucleus of critical size has formed (McPherson 1999; Chayen and Saridakis 2008; Rupp 2010) by the adsorption of free molecules to the nucleus surface and their incorporation into its lattice (McPherson 1999). The process finishes in final crystal form when the solution reaches a sufficient concentration of protein molecules or the crystal becomes contaminated by impurities in the solution. The region suitable for protein crystallization is generally represented by the solubility curve on a phase diagram (Fig. 11.4).

The phase diagram shows the solubility of the protein as a function of precipitant concentration in an ideal crystallization experiment (McPherson 1999; Chayen 2004). The precipitant can be any chemical or physical variable that affects protein solubility. In an undersaturated region, the conditions are unfavorable for nucleation or crystal growth owing to very low protein and/or precipitant concentrations. The solubility curve is the border between undersaturated and supersaturated regions. It represents dynamic equilibrium in the crystallization experiment. The metastable

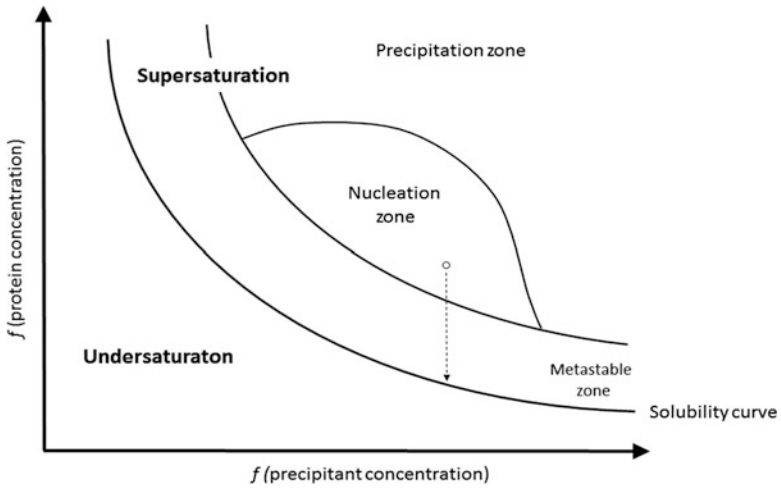


Fig. 11.4 Simplified phase diagram for protein crystallization. The protein concentration is represented as a function of the concentration of the precipitant agent. (Adapted from Snell et al. 2008)

zone supports crystal growth, but not nucleation. In the absence of a nucleant in this zone, the solution will remain clear. The region above the metastable zone is a zone that is favorable for spontaneous homogeneous nucleation. The closer the system is to the precipitation zone, the more nuclei will appear, resulting in the growth of numerous low-quality crystals. When nuclei have formed, the concentration of the protein in the solution will decrease, leading the system to the metastable zone. The process of crystal growth continues until the system reaches dynamic equilibrium (solubility curve). At a very high level of supersaturation, protein precipitation occurs. This region is unfavorable for crystallization, because the precipitates form faster than the crystals (Asherie 2004; Chayen 2005; Chayen and Saridakis 2008; Luft et al. 2014).

11.5 Crystallization Techniques

Numerous techniques have been developed for the crystallization of macromolecules: vapor diffusion (hanging-drop and sitting-drop), batch, microbatch under oil, microdialysis, free-interface diffusion, and counter-diffusion (Fig. 11.5).

11.5.1 Vapor Diffusion

Vapor diffusion (sitting-drop and hanging-drop) is a favorite technique of most protein crystallographers owing to its simplicity and ease of set up. The conditions in this technique (Hampel et al. 1968) are manipulated by diffusion through air. In the hanging-drop method (Fig. 11.5a), a small drop of the protein solution is mixed with a droplet of the precipitant solution and placed on a cover slide. The cover slide is flipped and sealed over the reservoir containing the precipitant solution (Fig. 11.5a). In the resulting isolated system the concentration of the precipitant in

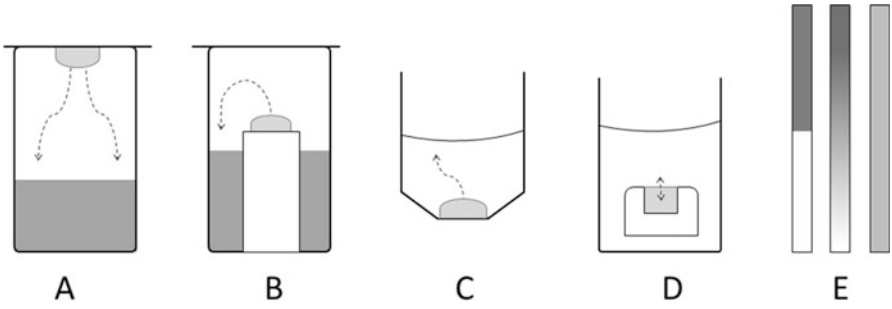


Fig. 11.5 Schematic representation of common crystallization techniques: hanging-drop vapor diffusion (**a**), sitting-drop vapor diffusion (**b**), microbatch under oil (**c**), microdialysis (**d**), free-interface diffusion (**e**). (Modified from Rupp 2010)

the reservoir is higher than that in the drop. Owing to the difference in the concentrations, the reservoir absorbs water from the drop through the vapor and drives the solution to the supersaturated state, thus enabling crystallization. The sitting-drop method (Fig. 11.5b) has the same principle as the hanging-drop, except that the drop is placed on a concave sitting-drop post. The system is isolated from the environment by sealing with a transparent tape or a cover slide. Both the sitting-drop and hanging-drop methods are suitable for initial crystallization screening, as well as for optimization of the discovered conditions. Sitting-drop and hanging-drop vapor diffusion methods have been optimized for automated robotic systems that allow minimization of the sample volume and provide high-throughput screening of crystallization conditions within several minutes (Bergfors 2009).

11.5.2 Batch Crystallization

The batch crystallization technique is the simplest and fastest method for protein crystallization. The principle is that the addition of the precipitant to the protein solution suddenly brings the drop to the supersaturated state. A modification of the batch technique is the microbatch under oil technique (Fig. 11.5c), where the drop containing mixed protein and precipitant solutions is covered by oil (Chayen et al. 1992). The oil does not interact with common precipitants and prevents water evaporation. The mixing of water-permeable silicon oil with paraffin oil promotes the evaporation of water from the drop, thus promoting supersaturation. The advantages of the batch method include minimization of the protein and precipitant volume, avoidance of condensation during temperature fluctuations, and the possibility of using automated systems for high-throughput crystallization screening (Rupp 2010).

11.5.3 Microdialysis

The basis of the microdialysis technique (Fig. 11.5d) is that the sample is separated from a large volume of the precipitant by a semipermeable membrane, which allows water and small molecules to pass, but prevents the passage of protein macromolecules. The protein can slowly reach its precipitation point by dialysis against a solution of a concentrated salt or an organic solvent. The technique allows salting in and salting out processes, as well as the change of pH of the crystallization solution. The benefit of this method is that the precipitant solution can be easily exchanged. The disadvantages include a tricky set up and a quite big sample consumption per trial (Rupp 2010).

11.5.4 Free-Interface Diffusion

In the free-interface diffusion technique (Fig. 11.5e), the protein and precipitant solutions are in contact with each other in a narrow glass capillary, without premixing. Over time, the two solutions diffuse into each other, creating a gradient of protein/precipitant concentrations along the capillary length (Ng et al. 2003). Crystallization of the protein may occur in the high sample/low precipitant or low sample/high precipitant regions of the capillary. Similar to the free-interface diffusion technique, there is also a counter-diffusion technique. An example of this technique is a gel acupuncture method, where microcapillaries containing a protein solution are inserted into a gel, which is covered by a layer of precipitant. The precipitant diffuses into the gel and forms a gradient in the microcapillary, promoting crystallization along its length (Garcia-Ruiz et al. 1998). The diffusion time of the precipitant through the gel to the capillary can be regulated by the penetration depth of the capillary in the gel.

11.6 Optimization of Crystallization

The initial screening of crystallization conditions does not always lead to high diffraction-quality crystals. Very often many crystal-like formations, but not well-shaped crystals, can be observed under a stereomicroscope during the inspection of crystallization plates (Fig. 11.6). In order to improve crystal size and quality, different optimization strategies can be applied:

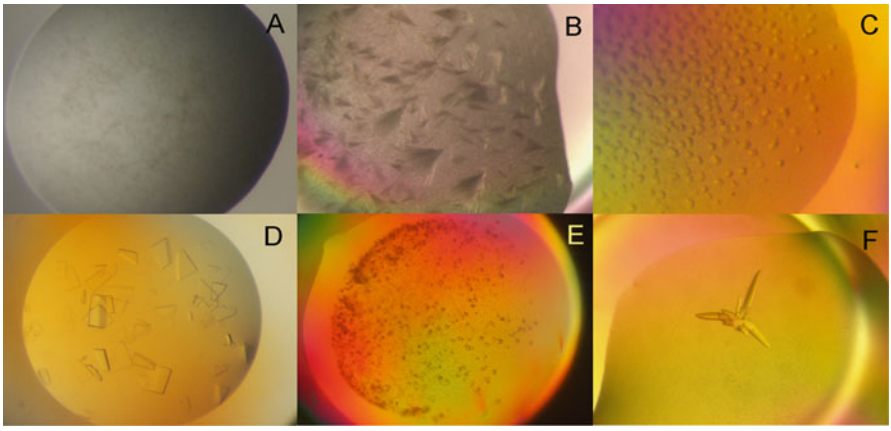


Fig. 11.6 Outcomes of different crystallization experiments (obtained in our laboratory). (**a**) Light precipitation; (**b**) needle clusters of LinB70; (**c**) spherulites of LinB81; (**d**) thin plates of the motor subunit of EcoR1241 restriction enzyme; (**e**) microcrystals of aldehyde dehydrogenase; (**f**) rod clusters of aldehyde dehydrogenase

11.6.1 Fine-Tuning of Initial Crystallization Conditions

A wide variety of physical, chemical, and biochemical parameters (the purity of the sample, sample handling, and the properties of the crystallization solution—composition, temperature, and pH) affect protein crystallization processes and thus the crystal size and quality (McPherson 1990). Many factors can influence protein solubility, which depends on the protein's surface charge. The protein's solubility is usually quite sensitive to pH and to temperature and precipitant composition (Green 1932). Protein precipitants are divided into four categories: salts, organic solvents, (long-chain) polymers, and non-volatile organic compounds. Salts influence the crystallization procedure based on the 'salting out' principle (the water molecules predominantly form bonds with ions, which ends in 'dehydration' of the protein). The order of salting out effectiveness of (an)ions follows the Hofmeister series (Collins and Washabaugh 1985)—$F^- \sim SO_4^{2-} > HPO_4^{2-} > CH_3COO^- >$ citrate^{3-} > tartrate^{2-} > $HCO_3^- > Cl^- > NO_3^- > ClO_3^-$—for negatively charged proteins, but is reversed for positively charged ones (cations are less effective) (Guilloteau et al. 1992; Carbonnaux et al. 1995; Baird et al. 2001). Further, various metal ions have been observed to induce the crystallization of proteins (essential for the protein's biological activity, stimulation of crystal growth, or stabilization of intermolecular contacts) (McPherson 1990; Trakhanov and Quiocho 1995; McPherson 1999). The last important aspect of protein crystallization is the concentration of the protein itself. The protein solution should be concentrated enough to enable crystal nucleation. The most common range is 5–20 mg/ml. The factors elaborated on above are the most well-known to influence protein crystallization.

11.6.2 Use of Additives

In the context of protein crystallization, an additive is a chemical compound that facilitates this process. Additives may already be present in the precipitant solution or they may be added to the macromolecule solution. According to their usage, additives are classified as metal cations, small alcohols, detergents, reducing agents, and others. The addition of metal cations can improve protein crystallization by the promotion of intermolecular interactions (Trakhanov and Quiocho 1995). Detergents act as stabilizing agents and increase the solubility of the protein (McPherson et al. 1986). Reducing agents, such as β-mercaptoethanol and dithiothreitol, stabilize the protein by preventing cysteine oxidation and are frequently added to the protein solution during protein preparation (McPherson 1990). Moreover, physiological ligands, cofactors, substrates, and inhibitors can also function as stabilizing agents and improve crystal quality compared with that of apoprotein.

11.6.3 Seeding

The main principle of the seeding technique is to introduce an existing low-quality crystal (seed) to the crystallization drop equilibrated at a low level of supersaturation, which is optimal for crystal growth, but adverse for spontaneous nucleation. Seeding can be applied for the initial screening, as well as for increasing the size and quality of the crystals (Bergfors 2003). Based on the size of the seeds, two main seeding methods are distinguished: macroseeding and microseeding. Microseeding implies the transfer of crystal fragments crushed by vortexing, seed beads, glass rods, or other tools. A modification of microseeding is streak seeding, where seeds are transferred by streaking with a whisker or a fiber through the new drop. During macroseeding, a single small crystal is washed and transferred to a new crystallization drop in order to increase the crystal size.

Generally, the seeds are prepared by crushing crystals. Nowadays, so-called nucleants are also used; these are created from porous material with different-sized pores. These pores can accommodate different large molecules, such as proteins. Pores with a size similar to that of the crystallized protein can act as a nucleation center and the protein can thus reach a higher supersaturated state to initialize crystal growth. A disadvantage of the use of nucleants is the presence of salts in the precipitant solution, which will then be forced to crystallize more rapidly than the required protein. To avoid this disadvantage, new polymer materials, called molecularly imprinted polymers (MIPS), are used. These polymer materials have pores designed directly for the sizes of the crystallized proteins (Saridakis et al. 2011).

11.7 X-Ray Data Collection

Obtaining data from a crystal starts by mounting the crystal from a crystallization drop onto a cryoloop, which is secured to a pin with a magnetic base. During the mounting, the crystal needs to remain in the mother liquor in order to prevent desiccation and further disintegration of the crystal. The data collection usually takes place under cryogenic temperatures. For this, after the mounting, the cryoloop with the crystal is flash-cooled in liquid nitrogen or directly attached to a goniometer head supplied with a Cryostream (Oxford Cryosystems, Oxford, United Kingdom). The method requires pre-soaking of the crystals in a cryoprotectant solution (glycerol, polyethylene glycol, glucose, xylitol) to avoid the formation of crystalline ice, which diminishes the diffraction quality. Cooling the crystals under liquid nitrogen temperature minimizes X-ray radiation damage and prolongs the lifetime of the crystals.

For the data collection, the cryoloop with the crystal is attached to a goniometer head, which allows the crystal to be oriented in an X-ray beam. The goniometer head is situated in the path of the X-ray beam between the radiation source and the detector. X-rays produced by the radiation source are directed onto the crystal and

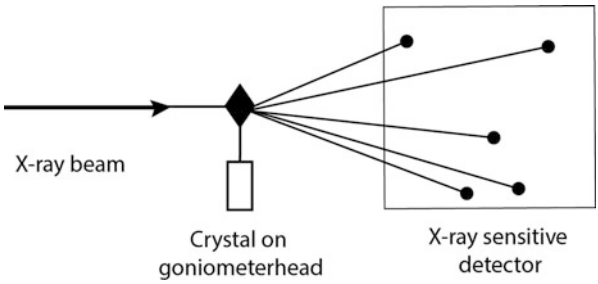


Fig. 11.7 Schematic representation of X-ray diffraction experiment

are subsequently scattered in many directions by the electrons of every atom in the crystal. The scattered X-rays interfere with each other when in phase to produce a diffraction pattern of regularly arranged spots (reflections) on the detector (Fig. 11.7). The measured intensities of these spots contain information about the arrangement of molecules within a protein crystal in atomic detail (Rhodes 2010).

To obtain complete data, it is necessary to obtain as many reflections as possible. This can be achieved by the rotation of the crystal around a single axis, at small increments (0.1–1.0°), while it is exposed to X-rays. The result of the data collection is a set of diffraction images with recorded intensities for each single reflection.

After data collection, the raw intensities must be processed to improve their consistency. In fact, data processing has to reconstitute the original, undistorted 3D lattice of reflections. In order to process the data, the dimensions of the unit cell need to be defined; this is done by indexing (determination of the symmetry—space group). Indexing is generally accomplished using an autoindexing routine and the data are integrated afterward (programs used are XDS, MOSFLM, or HKL3000) (Minor et al. 2006; Kabsch 2010; Battye et al. 2011). The integration procedure converts the hundreds of diffraction pattern images, collected on the detector containing thousands of reflections, into a single file, which consists of the Miller indices and intensity for each recorded reflection. After the data processing, the data must be merged and scaled (part of the CCP4 or PHENIX program package) (Adams et al. 2010; Winn et al. 2011). In the merging step, it is necessary to identify and merge the reflections that appear in two or more diffraction images and also to reject the outliers. The data scaling procedure is necessary because of the experimental differences of intensities in the recorded reflections. Therefore, all reflections must be located on a common scale; this means that the symmetry of the crystal needs to be reflected in the symmetry of the reflections; also, the reflections that are symmetry related must have the same intensity. The scaling of reflections must always be done on one dataset or multiple datasets, from two different crystals or from two different detector positions, because the intensity of the beam might change from frame to frame; also, the crystal might be larger in one direction than another, or different regions of the detector may have different sensitivities. The quality of the collected data can then be measured by a calculated score for merging or symmetry-related R-factor (R_{merge} or R_{sym}) (Giacovazzo 2002).

11.8 Solving of Protein Structure, and the Phase Problem

The reflections obtained after the data collection can be characterized by a structure factor *(F_{hkl})*, a complex number that represents the total scattering from all atoms of the unit cell, and that includes a structure factor amplitude ($|F_{hkl}|$) and a relative phase angle (Φ_{hkl}) between the diffracted waves.

The main objective of X-ray crystallography is to calculate an *electron density map* (distribution of electrons in space), which allows us to obtain the coordinates of atoms in the protein molecule. The relationship between the structure factor and the electron-density map can be mathematically expressed through *Fourier transformation*, where the structure factors containing information about amplitudes and phases are used as Fourier coefficients to generate the electron density map of the protein (Rhodes 2010):

$$p(xyz) = \frac{1}{V} \sum_{\substack{hkl \\ -\infty}}^{+\infty} \underbrace{|F(hkl)|}_{\text{Amplitudes}} . \, e^{-2\pi i[hx+ky+lz-\underbrace{\phi(hkl)}_{\text{Phases}}]}$$

Symbol ρ represents the density of electrons per unit volume at any point (x, y, z); V is the volume of the unit cell; and hkl are reflection indices. The structure factor amplitudes $|F_{hkl}|$ can be calculated as the square roots of the measured intensities (I_{hkl}). But no information about phases (Φ_{hkl}) can be obtained experimentally. This is the main problem of X-ray crystallography, called the *phase problem* (Rhodes 2010).

Several methods are used to solve the phase problem and determine the structure of a protein: molecular replacement, isomorphous replacement, anomalous dispersion, isomorphous replacement with anomalous scattering, and radiation damage-induced phasing. These methods are discussed below in brief.

11.8.1 Molecular Replacement (MR)

Molecular replacement is the most rapid and most frequently used method for solving protein structure. The phases are obtained from the known structure of a protein that is homologous to the unknown protein and has a similar fold and sufficient sequence identity (preferably at least 40%). The main challenge of MR is the correct transfer of the known molecule to the unit cell of the unknown molecule. This transfer can be achieved by two steps: rotation and translation. During the rotation step, the spatial orientation of the known and unknown molecules with respect to each other is determined. In the translation step, the movement requires the superimposition of one molecule onto the other, while the orientation found during the rotation step is calculated (Jan 1999).

11.8.2 Isomorphous Replacement (IR)

To determine protein structure by the IR method, it is necessary to obtain diffraction data not only from the native protein crystal, but also from the crystal of single (SIR) or multiple (MIR) heavy atom derivatives. The incorporation of heavy atoms into the protein crystal can be achieved by soaking the crystal or by co-crystallization with the heavy atoms. The introduced heavy atoms will change the scattered intensity of the crystal significantly (different atoms contribute to the scattered intensity in proportion to the square of the number of electrons). The native and derivative crystals should have the same unit cell dimensions; thus, they should be isomorphous. In the IR method differences in the structure factor amplitudes between native (FP) and derivative (FPH) protein crystals are used to identify the positions of the heavy atoms (FP-FPH). The resulting heavy atom structure (FH) can be used as a reference structure for phase determination. Phase determination with SIR leads to two possible solutions (phase ambiguity), while when MIR is used unambiguous phase determination can be achieved (Faust et al. 2008).

11.8.3 Single and Multiple Anomalous Dispersion (SAD and MAD)

Application of the SAD and MAD method requires the presence of strong anomalously scattering atoms, which can be incorporated into the protein (e. g., selenomethionine derivative or metal-containing proteins) or soaked into a crystal (heavy atoms, halogens, lanthanides, etc). The anomalous scattering properties of elements occur at specific wavelengths that are close to the absorption edge of the element. In the case of a MAD experiment, up to four data sets are collected at several wavelengths around the absorption edge of the element, where the anomalous scattering factors of the element are significantly different from each other. In the SAD method, just one dataset is collected at the wavelength corresponding to the absorption peak of the element. The phases are derived by using the differences in structure factor amplitudes that arise from anomalous scattering. These differences are generally smaller than in the case of IR; thus, accurate measurement of intensities is required during the data collection (Taylor 2003; Faust et al. 2010).

11.8.4 Single and Multiple Isomorphous Replacement with Anomalous Scattering (SIRAS and MIRAS)

The SIRAS and MIRAS methods are combinations of IR and anomalous dispersion. The use of anomalous scattering can prevent the phase ambiguity in SIR experiments.

11.8.5 Radiation Damage-Induced Phasing (RIP)

The RIP method is based on the specific effects of X-ray- or ultraviolet (UV)-induced radiation damage to protein crystals. With this method, two datasets are collected from a single crystal with the so-called burn dose, where the crystal is exposed to high doses of X-rays or UV. Radiation damage can cause significant differences in intensities between these two datasets. These differences can be used for the determination of phases by a SIR-type method. The RIP method works well for proteins that contain disulfide bridges (Nanao et al. 2005; Ravelli et al. 2005).

11.9 Interpretation of Electron Density Maps and Model Refinement

The quality of the phasing experiment may be influenced by phase errors and, therefore, also by errors in the electron density map. To improve phase quality, *density modification* techniques (programs—Solve/Resolve or PARROT) (Zhang et al. 1997; Terwilliger and Berendzen 1999) are used. Basically, these techniques work with some characteristics of a good electron density map as a template, which can recompute phases that will be more accurate than the original ones. In the *solvent flattening* technique, the density map is modified based on the solvent region (which may be the source of the phase error), which is flattened. The main idea is that the protein crystal is occupied by not only well-ordered protein molecules, but also by disordered solvent (mostly water molecules, which constitute 40–60% of the asymmetric cell volume). Based on the knowledge that the disordered solvent should have flat, featureless electron density, the boundaries between the disordered solvent and the ordered protein can be set and the entire electron density can be modified (Rhodes 2010). An *averaging* procedure can be used to improve the quality of the electron density map in a structure that contains two or more molecules of the same protein in the asymmetric unit. These protein molecules are related by non-crystallographic symmetry and usually show differences, since there are no crystallographic symmetry restraints to induce identity. Averaging the electron density values between the protein copies in the asymmetric unit can cancel out random errors, and the phases thus become more accurate. Another technique used, an algorithm called *histogram matching*, is based on the fact that the electron density distribution values in a map are basically the same for each structure at the same resolution. Altering the histogram of the electron density map values of the studied protein to resemble the histogram of a known high-quality protein can dramatically improve phases (Giacovazzo 2002).

Protein structure determination is followed by refinement of the structure. The refinement is carried out by performing iterative cycles of map calculation and model building in attempts to improve the agreement between the protein model and the observed data. The initial interpretation of the map is often not easy, and it depends

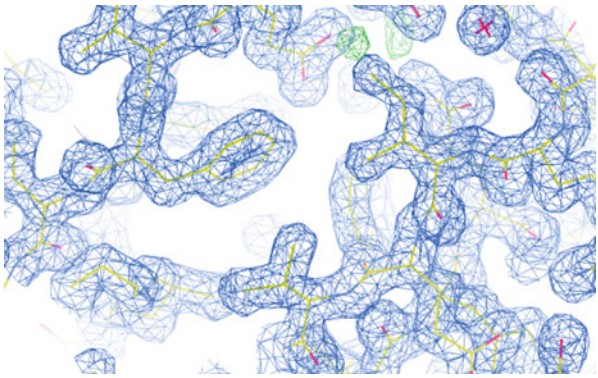


Fig. 11.8 Displayed electron density map with an atomic model in the program Coot. Coot allows model manipulations such as idealization, real space refinement, or manual rotation/translation of the protein residues and other molecules

on the resolution from which the phases are available. In a map with 3-Å resolution, only the contours of the polypeptide chain are visible. High-resolution data (1–2 Å) allow us to see the details of amino acid side chains and enable automatic chain tracking. The interpretation of the electron density map is performed on a graphic display (program Coot) (Emsley et al. 2010) (Fig. 11.8), where the piece of the protein chain can be fitted into the electron density map.

Molecular dynamics calculation is a widespread technique used for the refinement of protein structures. For crystallographic refinement, the X-ray information is used to restrain the energy of the system. The total potential energy of the system is categorized as empirical potential energy and experimental potential energy. The molecular dynamics simulation is provided by heating up the system from room temperature and then cooling the system down to ambient temperature (simulated annealing). The local minimum of the model thus reached can lead to an increase of convergence with respect to classical least squares.

11.10 Structure Validation

The main idea of structure validation is to check the consistency of the atomic model with the experimental data and also to check the consistency of the model with known physical and chemical properties. In small-molecule crystallography there is a very clear mathematical relationship between the data and the atomic model—the *R-factor*. This factor measures the agreement between the experimental data and the values recalculated from the atomic model. However, the *R-factor* alone is not sufficient in macromolecular crystallography, because the ratio between observations and parameters is limited (overrefinement). Therefore, another factor, which is virtually independent of refinement—R_{free}, is considered as an indicator of structure quality. In practice, a limited proportion of randomly chosen data, usually 5–10%, is left out of the refinement process and is used only to calculate the R value. R_{free} is generally higher than R but it reflects more closely the real information content of the

molecular model. At 2-Å resolution, the value of the *R* factor of the refined structure is usually less than 0.20 and the R_{free} value is between 0.20 and 0.30. The physical and chemical properties of the model are generally checked according to some specific indications. One of the most classical indicators is represented by the Ramachandran plot, which is widely used to inspect amino acid torsion angles. The plot checks amino acid conformations to determine whether they fall in favorable zones (allowed conformations) or whether they need to be corrected (or at least verified) because of their presence in a forbidden region of the plot (Rhodes 2010).

For structure validation (program Coot, PROCHECK in CCP4 package, or Molprobity) (Chen et al. 2010; Emsley et al. 2010; Winn et al. 2011), it is also useful to look at the static and dynamic disorder in the solved structure. The static disorder is basically determined by so-called *B-factors* (thermal vibrations of atoms). A very high *B-factor* value for some amino acids can indicate intrinsic disorder of part of the protein or may indicate misinterpretation of the electron density. The solvent content (water molecules) is closely related to the dynamic disorder of the protein, as the flexibility of the macromolecule is larger compared to tightly packed small molecules. Water molecules that are bound to the macromolecule can be identified during refinement and can be classified into three types: water molecules that interact with the side chains of amino acid residues to make hydrogen bonds (the first coordination shell of liquid water); water molecules that serve as a bridge between the main chain of the protein and other structural elements in the protein structure; and, finally, water molecules located in the internal cavities of the protein. All three of these possible water bindings in the protein structure may have a large impact on protein stabilization and therefore can improve the refinement and validation procedure (Giacovazzo 2002; Marek and Trávníček 2002).

References

Adams PD, Afonine PV et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66(Pt 2):213–221

Asherie N (2004) Protein crystallization and phase diagrams. Methods 34(3):266–272

Baird JK, Scott SC et al (2001) Theory of the effect of pH and ionic strength on the nucleation of protein crystals. J Cryst Growth 232(1):50–62

Battye TGG, Kontogiannis L et al (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr 67(4):271–281

Bergfors T (2003) Seeds to crystals. J Struct Biol 142(1):66–76

Bergfors TM (2009) Protein crystallization. International University Line, La Jolla

Blundell TL, Johnson LN (1976) Protein crystallography. Academic Press, New York

Bugg CE (1986) The future of protein crystal growth. J Cryst Growth 76(3):535–544

Carbonnaux C, Ries-Kautt M et al (1995) Relative effectiveness of various anions on the solubility of acidic Hypoderma lineatum collagenase at pH 7.2. Protein Sci 4(10):2123–2128

Chayen NE (2004) Turning protein crystallisation from an art into a science. Curr Opin Struct Biol 14(5):577–583

Chayen NE (2005) Methods for separating nucleation and growth in protein crystallisation. Prog iBiophys Mol Biol 88(3):329–337

Chayen NE, Saridakis E (2008) Protein crystallization: from purified protein to diffraction-quality crystal. Nat Methods 5(2):147–153
Chayen NE, Stewart PDS et al (1992) Microbatch crystallization under oil – a new technique allowing many small-volume crystallization trials. J Cryst Growth 122(1–4):176–180
Chen VB,, Arendall WB 3rd et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66(Pt 1):12–21
Collins KD, Washabaugh MW (1985) The Hofmeister effect and the behaviour of water at interfaces. Q Rev Biophys 18(04):323–422
Emsley P, Lohkamp B et al (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66(Pt 4):486–501
Faust A, Panjikar S et al (2008) A tutorial for learning and teaching macromolecular crystallography. J Appl Crystallogr 41:1161–1172
Faust A, Puehringer S et al (2010) Update on the tutorial for learning and teaching macromolecular crystallography. J Appl Crystallogr 43:1230–1237
Garcia-Ruiz JM, Moreno A et al (1998) Teaching protein crystallisation by the gel acupuncture method. J Chem Educ 75(4):442–446
Giacovazzo C (2002) Fundamentals of crystallography. Oxford University Press, Oxford
Green AA (1932) Studies in the physical chemistry of the proteins: X. The solubility of hemoglobin in solutions of chlorides and sulfates of varying concentration. J Biol Chem 95(1):47–66
Guilloteau J-P, Ries-Kautt MM et al (1992) Variation of lysozyme solubility as a function of temperature in the presence of organic and inorganic salts. J Cryst Growth 122(1):223–230
Hampel A, Labanauskas M et al (1968) Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNA's. Science 162(3860):1384–1387
Jan D (1999) Principles of protein X-ray crystallography. Springer, New York
Kabsch W (2010) Xds. Acta Crystallographica Section D-Biological Crystallography 66:125–132
Kierzek AM, Zielenkiewicz P (2001) Models of protein crystal growth. Biophys Chem 91(1):1–20
Luft JR, Newman J et al (2014) Crystallization screening: the influence of history on current practice. Acta Crystallographica Section F-Structural Biology Communications 70:835–853
Marek J, Trávníček Z (2002) Monokrystalová rentgenová strukturní analýza. Vydavatelství Univerzity, Palackého
McPherson A (1990) Current approaches to macromolecular crystallization. Eur J Biochem 189 (1):1–23
McPherson A (1999) Crystallization of biological macromolecules. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
McPherson A (2001) A comparison of salts for the crystallization of macromolecules. Protein Sci 10(2):418–422
McPherson A, Koszelak S et al (1986) The effects of neutral detergents on the crystallization of soluble-proteins. J Cryst Growth 76(3):547–553
Minor W, Cymborowski M et al (2006) HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. Acta Crystallographica Section D-Biological Crystallography 62:859–866
Nanao MH, Sheldrick GM et al (2005) Improving radiation-damage substructures for RIP. Acta Crystallographica Section D-Biological Crystallography 61:1227–1237
Ng JD, Gavira JA et al (2003) Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. J Struct Biol 142(1):218–231
Ravelli RBG, Nanao MH et al (2005) Phasing in the presence of radiation damage. J Synchrotron Radiat 12:276–284
Rhodes G (2010) Crystallography made crystal clear: a guide for users of macromolecular models. Elsevier Science, Burlington
Rupp B (2010) Biomolecular crystallography: principles, practice, and application to structural biology. Garland Science, New York
Salemme FR, Genieser L et al (1988) Molecular factors stabilizing protein crystals. J Cryst Growth 90(1):273–282

Saridakis E, Khurshid S et al (2011) Protein crystallization facilitated by molecularly imprinted polymers. Proc Natl Acad Sci U S A 108(27):11081–11086
Snell EH, Nagel RM et al (2008) The application and use of chemical space mapping to interpret crystallization screening results. Acta Crystallographica Section D-Biological Crystallography 64:1240–1249
Taylor G (2003) The phase problem. Acta Crystallographica Section D-Biological Crystallography 59:1881–1890
Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 55(Pt 4):849–861
Trakhanov S, Quiocho FA (1995) Influence of divalent-cations in protein crystallization. Protein Sci 4(9):1914–1919
Winn MD, Ballard CC et al (2011) Overview of the CCP4 suite and current developments. Acta Crystallographica Section D-Biological Crystallography 67:235–242
Zhang KY, Cowtan K et al (1997) Combining constraints for electron-density modification. Methods Enzymol 277:53–64

Chapter 12
Nuclear Magnetic Resonance

Milan Zachrdla, Zuzana Jaseňáková, and Lukáš Žídek

12.1 Introduction

The purpose of the following pages is to introduce the method of *nuclear magnetic resonance* (NMR). The limited space obviously does not allow us to explain theoretical principles, describe instrumentation, and present the armory of NMR experiments at the level of a university course. Fortunately, there is no need for it. Excellent textbooks have been written giving a beginner chance to reach a high level of understanding. The land of England seems to be particularly fertile in this sense as the books written by Malcolm Levitt (2008) and James Keeler (2010) are real gems of chemical education. More textbooks covering all important aspects of protein NMR spectroscopy are also available (Cavanagh et al. 2007; Rule and Hitchens 2006). Therefore, we present here only a light appetizer and direct an interested reader to a rich menu of NMR-related literature. Considering the aim of this chapter, the references are limited to reviews and general texts, original papers, and specialized articles are not cited.

Although NMR is a very versatile technique, applicable to chemical compounds of various composition in solid, liquid, or gas phase, only the basics of NMR spectroscopy of proteins are discussed in this chapter. Experiments for studies of nucleic acids in solution, rapidly developing methods of solid-state NMR spectroscopy, metabolomic analyses, and various imaging techniques are based on the same physical principles but utilize specific technical tricks, which are beyond the scope of this short overview.

M. Zachrdla · Z. Jaseňáková · L. Žídek (✉)
Central European Institute of Technology, Masaryk University, Brno, Czech Republic
e-mail: lzidek@chemi.muni.cz

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_12

12.1.1 Basic Principles

Atomic nuclei of most isotopes interact with magnetic fields. This interaction is usually described in terms of two physical vectors: *magnetic induction* of the field and *intrinsic magnetic moment* of the nucleus. In a static homogeneous magnetic field, vectors of the magnetic moments μ *precess* about the direction of the vector of the magnetic induction B_0, in the same manner as the axis of a spinning top precesses about the direction of the gravitational field of the Earth. *Nuclear magnetic resonance spectroscopy* is a method of measurement of the frequency of the precession of nuclear magnetic moments in macroscopic samples (e.g., in a solution containing the molecule of interest, Fig. 12.1a). In such samples, the observed quantity is the bulk *magnetization* **M** (density of the magnetic moments), given by the distribution of orientations of magnetic moments of the observed nuclei (Fig. 12.1b) in the magnetic field. According to the Boltzmann's law, the distribution at the thermodynamic equilibrium is given by energies of individual orientations. As the magnetic moments parallel with the direction of the magnetic induction have the lowest energy (Fig. 12.1c), the parallel orientations are more probable, and the vector describing the bulk magnetization, **M**, has the same direction as the vector of the magnetic field induction $\mathbf{B}_0$ (Fig. 12.1d), which also represents the axis of the precession (Fig. 12.1e). Obviously, the vector of the magnetization does not precess because it is collinear with the axis of the precession. To measure the precession frequency, the distribution of the magnetic moments should not be polarized along the direction of the magnetic induction (as it is at the equilibrium) but in a different direction (ideally perpendicular). This is achieved by irradiating the sample with electromagnetic waves with a frequency resonating with the frequency of the precession of the observed magnetic moments. The frequency range at the current spectrometers overlaps with the very short radio waves used in the frequency-modulated (FM) broadcasting. The magnetic induction of the electromagnetic wave, $\mathbf{B}_1$, combines with $\mathbf{B}_0$, and rotates the bulk magnetization **M** from its equilibrium direction. The irradiation time is carefully calibrated to rotate **M** to a plane perpendicular to $\mathbf{B}_0$ (Fig. 12.1e). After the irradiation is switched off, **M** continues to rotate in this plane, representing a source of rotating magnetic field. If a detector loop is placed close to the sample, the magnetic flux through the loop oscillates with the frequency of the rotation of the magnetization and generates an oscillating electric current in the loop. The electric current is converted to a digital form, and its time dependence is transformed into a *frequency spectrum*. The mathematical procedure used in the last step is usually the *Fourier transformation*.

12.1.2 Influence of Structural Environment

The described principle of measuring precession frequencies of nuclear magnetic moments does not explain what NMR spectroscopy offers to chemists and biologists. The major applications of NMR in chemistry, biology, medicine, or material

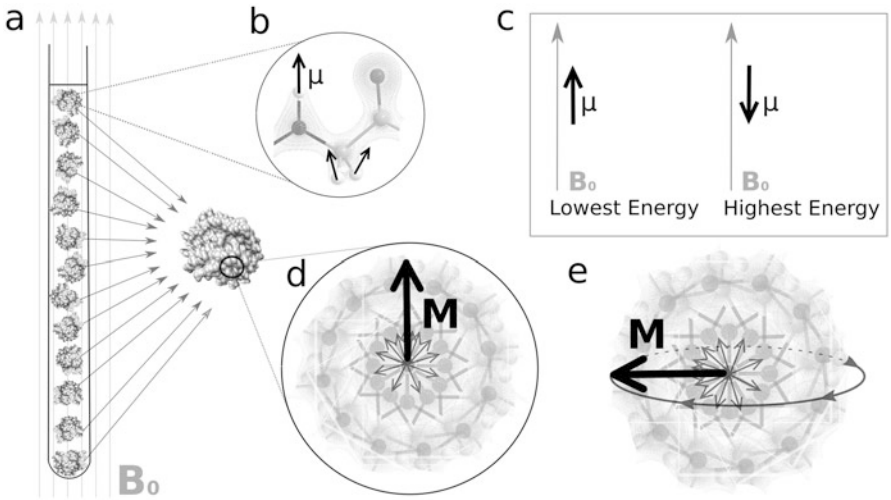


Fig. 12.1 Principle of protein NMR spectroscopy in solution. (**a**) Solution containing the studied molecules (receiver-domain sensor histidine kinase CKI1 from *Arabidopsis thaliana* is shown as an example) is placed in a homogeneous static magnetic field, described by the vector of magnetic induction $\mathbf{B}_0$. (**b**) Magnetic moments of observed nuclei (protons in a magnified glycine residue are shown as examples) interact with the external field but also with the microscopic fields of electrons and other nuclei. Proton nuclei and their magnetic moments are shown as white spheres and black arrows, respectively; electrons are visualized by drawing the electron density distribution. (**c**) The orientations of magnetic moments with the lowest and highest energies are parallel resp. antiparallel with the external magnetic fields. (**d**) NMR in a homogeneous field is influenced only by changes of orientation of the molecule, not by its position in the test tube. Therefore, models of all molecules of the receiver domain are overlaid with the superimposed position of the amide proton in the selected glycine residue (placed in the center). The magnified view shows that the molecules in isotropic liquids are assumed to adopt all orientations with the same probability (note the symmetry of the displayed glycine residues), but the distribution of magnetic moment orientations is biased by the magnetic fields, with the energetically more favorable orientations being more probable (only magnetic moments of amide protons are shown for the sake of clarity). As a result, the bulk magnetization **M**, depicted as a thick black arrow, is oriented along $\mathbf{B}_0$. (**e**) In a typical experiment, radio waves resonating with the precession frequency of magnetic moments are applied to change the distribution of magnetic moment orientations, and to rotate **M** into a plane perpendicular to $\mathbf{B}_0$. **M** then rotates about $\mathbf{B}_0$ as indicated

science, including structural studies of plant proteins, are based on slight perturbations of the aforementioned static magnetic fields by microscopic magnetic fields affecting the observed nucleus. Intrinsic magnetic moments of some nuclei and unpaired electrons (if they are present in the molecule, e.g., in certain metalloproteins) are important sources of the microscopic fields. Intrinsic moments of paired electrons cancel each other, but an orbital motion of such electrons also represents a source of a magnetic moment, known as *orbital magnetic moment*. Local fields of intrinsic and orbital moments are small and rapidly fluctuate as the molecule moves. The coherent and stochastic components of the microscopic fields have distinct effects.

Coherent modifications of $\mathbf{B}_0$ slightly shift the observed frequency. If the source of the modification is an interaction with orbital magnetic moments of electrons, the deviation of the precession frequency is called *chemical shift*. Coherent effects of direct (through-space) interactions with nuclear magnetic moments are observable only if the studied molecules are not oriented completely randomly (as in Fig. 12.1a), but some orientations of the molecule are preferred. This is the case if the molecule is present in crystal, liquid crystal, stretched gel, or other anisotropic environment. If the anisotropy is weak, the observed interaction is known as the *residual dipolar coupling* (RDC). Indirect interactions with magnetic moments of other nuclei, mediated by electrons in chemical bonds, are not only observable in the NMR spectra but also make more advanced methods of *correlated NMR spectroscopy* possible (vide infra). This type of interaction is called *J-coupling* or *scalar coupling*.

Effects of *stochastic* components of the microscopic fields are known as *relaxation phenomena*. Due to the relaxation, the measured frequencies are not defined precisely, and the lines in NMR spectra are not sharp as the lines in optical atomic spectra. Because the origin of fluctuations of the macroscopic magnetic fields is the molecular motions, broadening of the spectral lines depends on internal and Brownian motions of the studied molecules. In the case of well-structured proteins (and nucleic acids) in solution, the line broadening is proportional to the size (molecular mass) of the studied molecule, which makes NMR studies of large biopolymers challenging. On the other hand, NMR relaxation is a very useful phenomenon. Redistribution of magnetic moment orientations, the prerequisite of any NMR experiment, is solely due to the relaxation. Moreover, the most important NMR techniques of structure determination and monitoring molecular motions, discussed below, are based on relaxation phenomena.

Due to the interactions of the observed nuclei with the microscopic magnetic fields, NMR spectra carry useful information on structure and dynamics of the studied molecule. Nuclei of the same isotope (e.g., ^{1}H) in different chemical groups have different chemical shifts. Therefore, NMR spectra contain multiple peaks at frequencies, or chemical shifts, of nuclei in individual chemical groups. In addition, peaks of individual nuclei are further split due to interactions with other nuclei connected by chemical bonds (Fig. 12.2).

12.1.3 Correlated Spectroscopy

The basic experiment described above (consisting of irradiation by radio waves, acquisition of the signal, and a recovery period) often provides sufficient information to verify a structure of a small molecule. More advanced experiments are usually needed for structure determination of unknown compounds or for atomic-resolution studies of proteins, nucleic acids, and other biomolecules. The common denominator of the advanced NMR techniques is the concept of *correlated spectroscopy*. Sequences of several short pulses of radio waves are applied to the sample. The pulses are separated by delays during which the magnetic moments are allowed to

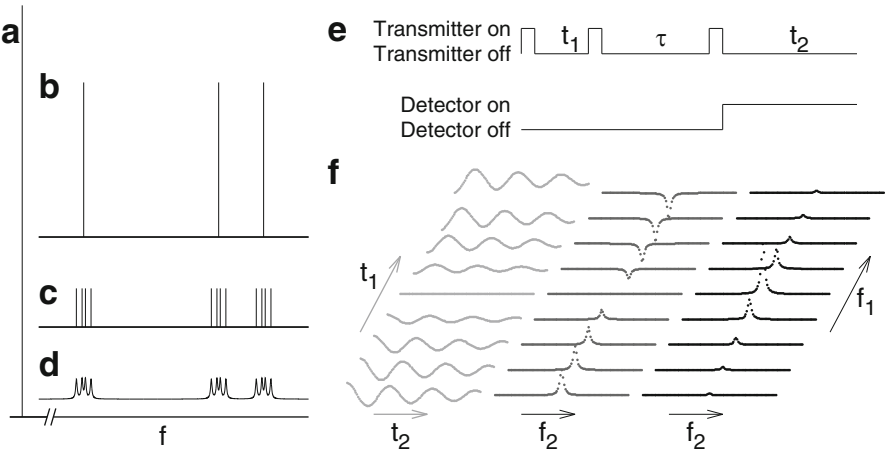


Fig. 12.2 Effect of microscopic magnetic fields on NMR spectrum (**a**–**d**) and principles of two-dimensional NMR spectroscopy (**e**, **f**). A hypothetical spectrum of a free proton would have a form of a sharp line at a negative frequency (**a**). Spectra of protons in molecules are influenced by the effects of microscopic magnetic fields. The coherent effect of interactions with the orbital magnetic fields of electrons shifts the proton frequencies to values that are less negative and distinct for hydrogen atoms in chemically different position (**b**). For example, three lines would appear in a spectrum of the glycine shown in Fig. 12.1b (neglecting other amino acids and solvent). The coherent effect of through-bond interactions with the other protons splits the peaks into multiplets (**c**). Finally, fluctuations of the microscopic fields due to molecular motions cause broadening of the lines (**d**). Two-dimensional proton spectrum with frequencies correlated via through-space interactions of the proton magnetic moments can be acquired by applying three pulses of radio waves separated by two delays and followed by the detection period t_2 (**e**). Length of the pulses is chosen so that they rotate magnetization by 90°. The delay τ must be sufficiently long, so that the interaction of magnetic moments changes distribution of their orientations. The experiment is repeated many times with the delay t_1 incremented. The obtained signal is modulated by varying the delay t_1 (**f**), with the modulating frequency being the frequency of the interacting nucleus. The raw signal acquired as a function of time is shown in light gray, the signal processed by applying Fourier transform in the t_2 dimension is shown in dark gray, and the signal converted to a 2D spectrum by applying Fourier transform in both time dimensions is shown in black. Correlated 2D spectra utilizing through-bond interactions differ in the sequence of radio-wave pulses (shown in panel **e**), but are processed in the same manner (shown in panel **f**)

interact mutually. In such a manner, effects of external (radio waves) and local (other nuclei) fields are combined to create magnetic states of the nuclei in which the distribution of magnetic moment orientations of the first nucleus is *correlated* with the distribution of magnetic moment orientations of the second nucleus (Fig. 12.2). As a consequence, only magnetizations of interacting nuclei are observed because only magnetic moment orientations of such nuclei can be correlated.

Usually, frequencies of the correlated magnetizations are recorded in *multidimensional spectra*. If the experiment is repeated several times, and one of the delays between radio-wave pulses is increased by a small increment in each repeat, the correlation is no longer fully efficient, but its efficiency is *modulated*. If the experiment is set up so that the efficiency depends on the frequency of the first

nucleus, but the frequency of the second nucleus is actually measured, the result of the measurement can be plotted as a two-dimensional spectrum with *cross peaks* at the frequencies of the interacting nuclei (Fig. 12.2). The *heteronuclear single-quantum coherence* (HSQC) spectroscopy represents an example of such a correlated experiment. If frequencies of hydrogen and nitrogen in a protein are correlated as described above, the recorded spectrum can be plotted as a two-dimensional map with the horizontal and vertical positions of peaks corresponding to hydrogen and nitrogen frequencies, respectively. If the delays between the radio-wave pulses are chosen so that only the directly bonded hydrogen and nitrogen nuclei are correlated, coordinates of each peak correspond to the frequencies of hydrogen and nitrogen in a single NH group. Frequencies of protons attached to carbon, oxygen, or sulfur are not observed, which simplifies the spectrum dramatically. Moreover, the spectrum contains a useful structural information (the correlation says which hydrogen is attached to which nitrogen) and usually represents a starting point in a study of a particular protein (vide infra).

The outlined principle of NMR experiments is straightforward and allows us to correlate precession frequencies in many different fashions. However, the physical reality restricts the possible applications, especially in the case of relatively large molecules such as proteins or nucleic acids:

- Energy difference between the most and least favored orientation of the magnetic moment (i.e., along and against the direction of $\mathbf{B}_0$, respectively, see Fig. 12.1c) is very small at the ambient temperature. Therefore, the most favorable orientation is only slightly preferred, and the resulting magnetization is relatively small. As a consequence, sensitivity of conventional NMR experiments is low.[1]
- Relaxation properties of nuclei with the spin quantum number higher than ½ make frequencies of these nuclei immeasurable in large molecules (with a notable exception of ^{2}H). For example, the aforementioned HSQC experiment can correlate proton frequency with the frequency of ^{15}N (spin number of ½), but not with the frequency of rapidly relaxing ^{14}N (spin number of 1).
- Line broadening is proportional to the molecular mass, as mentioned above. Conventional experiments employing the spin-½ nuclei are usually replaced by relaxation-optimized variants for proteins of mass larger than 30 kDa, and/or the number of spin-½ nuclei interacting with the observed nucleus is reduced to suppress relaxation (typically ^{1}H is replaced by ^{2}H).
- Nuclei with the spin number of ½ often represent minor isotopes of biogenic elements. Proton, ^{19}F, and ^{31}P are major isotopes of hydrogen, fluorine, and phosphorus, but the natural abundance of ^{13}C and ^{15}N is low (1.1% and 0.4%, respectively), and suitable isotopes of oxygen or sulfur are not available.

[1] A promising possibility of improving the sensitivity is *dynamic nuclear polarization*, taking advantage of a much larger difference between the energies of magnetic moments of unpaired electrons.

To improve the sensitivity, the isotopic composition of proteins larger than approx. 10 kDa is usually altered using isotope labeling approaches discussed in Sect. 12.2.1.

12.2 Sample Preparation

The great advantage of NMR spectroscopy is that it can be applied to diverse types of samples, including in vivo or in situ measurements. Nevertheless, it is highly recommended to optimize the sample composition, considering the inherently low sensitivity of the method in combination with generally high price and limited solubility of proteins. The basic recommendations have been summarized by Primrose in a very clear and detailed manner (Primrose 1993). Here, we only briefly review the most important issues.

12.2.1 Isotope Labeling

Isotope enrichment is one of the techniques that made NMR studies of biomacromolecules possible. The very first studies on isotopically labeled proteins were carried out in the 1960s, resulting in the production of isotopically labeled proteins extracted from bacteria and plants grown in media containing isotopically labeled nutrients (Crespi et al. 1968; Putter et al. 1969). Over the years, many isotope labeling strategies have been developed. Employing labeled metabolic precursors, cell-free expression, protein splicing, and other tricks, it is possible to selectively label chosen regions of proteins, amino acids of a chosen type (e.g., all methionines), atoms in specific positions in amino acids, etc. Here, we briefly describe only the most common protocol of *uniform isotope labeling* in a bacterial expression system.

Typically, a suitable *Escherichia coli* strain is grown in a minimal medium containing sole sources of nitrogen and/or carbon in a form of commercially available isotope-labeled substances (ammonium salts, glucose, glycerol). For example, the M9 medium is composed of 6 g/l Na_2HPO_4, 3 g/l KH_2PO_4, and 0.5 g/l NaCl, supplemented with 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 10 mM $FeSO_4$, 10 mg/l biotin, 10 mg/l thiamine, and antibiotics. When expressing uniformly [^{13}C, ^{15}N] labeled proteins, 2–4 g/l [$^{13}C_6$] D-glucose, and 0.5–1 g/l $^{15}NH_4Cl$ as the sole of carbon and nitrogen sources are added. Proteins larger than 25 kDa are often deuterated by using ^2H_2O (D_2O) in the growth medium instead of natural-abundance water (Zhang and van Ingen 2016).

12.2.2 Solubility

Protein sample should be fully dissolved in the aqueous solution that represents the physiological environment for the molecule. If the protein is only partially dissolved in a given solvent, the signal in a NMR spectrum will come out only from the soluble part (and the solvent).

12.2.3 Concentration and Volume

As NMR is relatively insensitive, the recorded signal is usually an average of several repeated measurements. Sensitivity of the NMR experiment is given by the signal-to-noise ratio, which is proportional to the square root of the number of times the signal is averaged. Therefore, doubling the sample concentration reduces the measurement time needed to obtain the same sensitivity by a factor of four. It documents that it is always good to use concentration as high as allowed by the circumstances (protein aggregation, sample availability, biological relevance of the conditions). It should be also emphasized that the optimal sensitivity is obtained only if the sensed volume of the given NMR probe is filled, which usually requires at least 0.3 ml of the protein solution.

12.2.4 Stability

The sample must be stable for the amount of time necessary to acquire all data at the temperature of the measurement. The lifetime of the sample can be prolonged by reducing microbial contamination, air oxidation, enzymatic, or chemical hydrolysis. Therefore, contamination by proteases or metal ions should be avoided, and antimicrobial agents such as sodium azide should be used routinely.

12.2.5 Buffer Composition

Proteins is solution are flexible molecules that undergo different inter- and intramolecular dynamical changes. The closer the sample composition to the physiological conditions is, the more biologically relevant data are obtained. In principle, NMR does not require any specific choice of buffer. However, any protons in the buffer molecule that do not exchange rapidly with water give rise to the signal in ^{1}H NMR spectrum. Therefore, the used buffer concentrations are often low (10–50 mM), but the buffering capacity should be also taken into account. To detect signals of protons that slowly exchange with water, most importantly of amide protons in the protein

backbone, experiments are performed in 90–95% H_2O (containing 5–10% D_2O for technical reasons). The presence of the huge excess of protons of the water molecules requires to apply special techniques suppressing the signal of water. Exchange of amide protons with water also makes measurements at low pH preferable.

Other additives may also influence the result of NMR measurement. High ionic strength decreases sensitivity of the NMR measurements. Salt concentrations up to 0.2 M can be tolerated, but concentrations higher than 0.5 M have a significant negative effect. Compounds increasing viscosity (glycerol, sucrose) broaden the spectral lines by slowing down tumbling of protein molecules and thus lower the sensitivity. As mentioned in Sect. 12.2.4, antimicrobial agents and metal chelators are often added to protect protein samples from degradation.

12.3 Assignment of NMR Spectra

The most basic and simple NMR experiments are nonselective. Irradiating the sample by a radio wave results in a signal of all nuclei with frequencies sufficiently close to the frequency of the applied wave (typically all protons or ^{13}C nuclei in all aliphatic groups). More advanced methods of multidimensional correlated spectroscopy greatly simplify spectra, improve the resolution, and encode the structural information in the correlations of the observed frequencies. The conventional strategy of converting the information hidden in the spectra to structural models consists of several steps. The first step represents assigning the frequencies observed in NMR spectra to particular atoms.

Although the principle of frequency assignment is general, the choice of the best-suited approach depends on the nature of the particular individual sample. With an increasing number of amino acids in the polypeptide chain, the spectrum becomes more crowded, and peaks of different nuclei are more likely to overlap. The conformational state of the protein plays an important role as well. A well-ordered protein with a molecular weight of 15 kDa behaves differently than an intrinsically disordered protein of the same molecular weight. Amino acids of the same type are surrounded by different micro-environments, and their nuclei are influenced by different microscopic magnetic fields. It is therefore likely that nuclei in the same position within the amino acid, but in different position in the protein molecule, e.g., alpha protons of valines in different regions of the protein, have distinct chemical shifts. On the contrary, the average environment of all valines in a disordered protein is very similar, and their chemical shifts are likely to be similar too. Therefore, we observe *high chemical shift dispersion* (broad distribution of frequencies of nuclei in the same chemical groups) for well-ordered proteins, but *low chemical shift dispersion* for intrinsically disordered proteins. On the other hand, molecular motions of relatively rigid well-structured proteins are slower and proportional to the molecular mass, whereas molecular motions of flexible disordered proteins are faster. Consequently, peaks of well-ordered proteins are broader than those of disordered proteins.

In conclusion, higher resolution is needed to assign the frequencies of a disordered protein, and higher sensitivity is needed to assign frequencies a well-ordered protein of the same size.

12.3.1 Assignment at Natural Isotope Abundance

Natural isotope abundance seriously limits NMR studies of large molecules. In principle, two-dimensional *heteronuclear* spectra (correlating frequencies of different isotopes, e.g., ^{1}H and ^{13}C, or ^{1}H and ^{15}N) of proteins can be acquired, but sample concentration and experimental time demands rarely make such approach practical. Therefore, NMR studies of proteins at the natural isotope assignment rely on 1D and 2D ^{1}H *homonuclear* spectra (correlating frequencies of protons in different positions in the molecule). Given the limited dimensionality, NMR studies of proteins exceeding approximately 10 kDa molecular mass represent a tedious or virtually impossible task. In the case of small proteins, the assignment of both backbone and side chains is done using 2D experiments known as *COrrelated SpectroscopY* (COSY), *Totally Correlated SpectroscopY* (TOCSY), and *Nuclear Overhauser Effect SpectroscopY* (NOESY). The 2D COSY and 2D TOCSY spectra provide through-bond correlations, i.e., each cross peak corresponds to the correlation between ^{1}H atoms separated by one or more chemical bonds. In contrast, 2D NOESY provides through-space correlations, i.e., each cross peak corresponds to the correlation between protons that are close in space (up to 5–6 Å), no matter whether the protons are in the same residue or are separated in the sequence by dozens of residues. The assignment protocols are based on search for correlation between frequencies of nearby protons. Considering the limited use of this approach, we do not discuss it here but refer to the detailed description by its pioneer Kurt Wüthrich (Wüthrich 1986).

12.3.2 Assignment of Isotope-Labeled Samples

Introduction of isotope labeling as a routine procedure of protein sample preparation started a new era of NMR spectroscopy based on heteronuclear correlations. Multidimensional NMR experiments involving various correlations were developed, making studies of proteins with molecular mass up to 30–40 kDa feasible. The experiments utilized in the conventional assignment protocols, described below, have been reviewed by Sattler et al. (Sattler et al. 1999).

The process of assignment usually starts by recording a 2D spectrum correlating ^{1}H and ^{15}N nuclei (the HSQC experiment, mentioned in Sect. 12.1.3, is used most often). This spectrum serves as a fingerprint of the studied protein (Fig. 12.3). Protons in the amide frequency region are correlated with the directly bonded nitrogens providing one peak for each amide group in the protein. Comparing the number of the observed signals with the number of expected signals and the

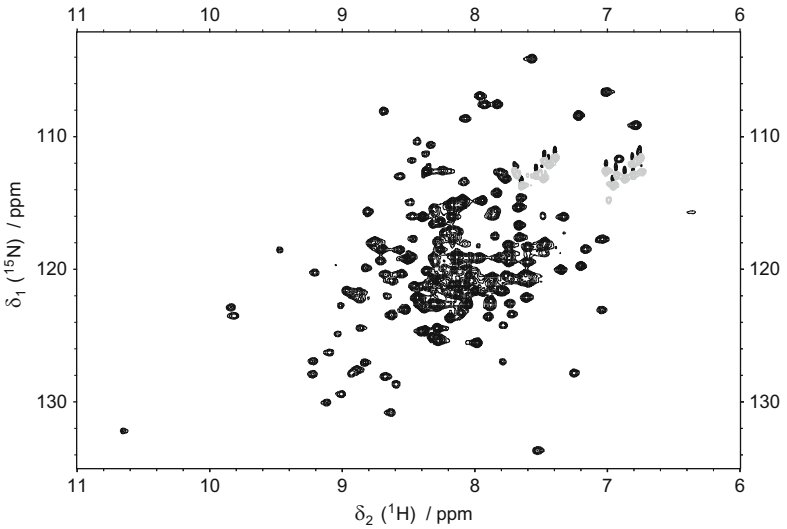


Fig. 12.3 2D HSQC spectrum correlating ^{1}H and ^{15}N nuclei of the receiver domain of the sensor histidine kinase CKI1 from *Arabidopsis thaliana*, with the active site phosphorylation mimicked by the beryllofluoride group (BeF_3^-). Each black peak in the spectrum corresponds to one NH group in the backbone of the protein. The gray peaks correspond to NH_2 groups of asparagine and glutamine side chains. Peaks of a 60-residue long disordered N-terminal region (histidine tag with a linker region preceding the sequence of the receiver domain) are clustered in the center of the spectrum, whereas peaks of the well-structured receiver domain are distributed in the whole range of frequencies plotted

chemical shift dispersion can be good indicators of whether the sensitivity is sufficient and whether the protein is folded.

The standard approach to frequency assignment of isotopically enriched proteins relies on recording a set of 3D spectra that correlate different atoms in the protein. There are numerous variants of NMR experiments using different building blocks (consisting of sequences of radio-wave pulses separated by carefully chosen delays) designed to achieve the desired correlation. Fortunately, the experiments are usually named in a self-explanatory manner, simply listing abbreviations of the atoms that are correlated. For example, the 3D HNCO experiment indicates that each signal obtained in a three-dimensional spectrum corresponds to the correlation of H^N, N, C' atoms in a protein backbone. Typically, the assignment is divided into two consecutive steps: backbone assignment and side-chain assignment.

Backbone assignment usually starts by recording a pair of 3D HNCACB and 3D CBCA(CO)NH experiments. Figure 12.4 illustrates how these experiments are used to identify sequential neighbors. The 3D CBCA(CO)NH experiment correlates C^{α}, C^{β}, and amide group first with ^{13}C of the carbonyl group. To keep dimensionality of the experiment equal to three, frequency of the carbonyl ^{13}C is not measured (highlighted by the parentheses in the name of the experiment). Instead, carbonyl ^{13}C is further correlated with amide ^{1}H and ^{15}N of the following amino acid. In this

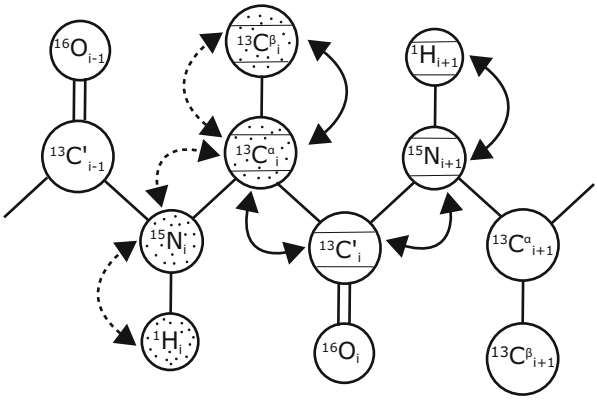


Fig. 12.4 Scheme of the backbone assignment procedure. A 2D ^{1}H-^{15}N experiment provides a spectrum where each peak corresponds to a single amide group in the protein. To assign the peaks to the corresponding amino acids, two 3D spectra were recorded. In the 3D HNCACB spectrum, the amide group is correlated (dotted arrows) with the carbons from the same amino acid (C^{α}_{i} and C^{β}_{i}). Two observed peaks, thus, correspond to a correlation of H^{N}_{i}, N_i with C^{α}_{i} or C^{β}_{i} atoms (circles filled with a dotted pattern). In the 3D CBCA(CO)NH spectrum, each amino acid is correlated (solid arrows) with carbon atoms of the amino acid preceding in the sequence (circles with horizontal strips). The amide group of the residue following in the sequence ($i + 1$) can be identified by finding a pair of peaks with matching C^{α} and C^{β} frequencies in both spectra (circles filled with both strips and dots). We can then continue to search for sequential neighbors in both directions in the same manner, and establish sequential connectivities for all amino acids (except for prolines)

manner, correlation of amide ^{1}H and ^{15}N of the given amino acid with the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ carbons from the preceding amino acid is obtained. The 3D HNCACB experiment correlates ^{1}H and ^{15}N of the amide group directly with the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ carbons of the same amino acid.[2] Comparison of the correlations observed in the described experiments (matching pairs of chemical shifts in both spectra) identifies *sequential connectivity* in the studied molecule. Once a stretch of sequentially connected amino acids is established, the obtained fragment is located in the sequence of the protein based on the comparison of the observed chemical shifts and average chemical shifts for a particular amino acid types. Although the 3D HNCACB and 3D CBCA(CO)NH principally contain all the information necessary for obtaining the complete backbone assignment, limited resolution and sensitivity of these experiments often lead to missing or ambiguous assignments. Therefore, additional spectra are usually recorded. A pair of 3D HNCA and 3D HN(CO)CA experiments correlating only $^{13}C^{\alpha}$ carbons and a pair of 3D HNCO and 3D HN(CA)CO experiments correlating carbonyl ^{13}C with amide groups offer higher sensitivity and resolution.

A drawback of all 3D experiments discussed so far is the need for a presence of an amide proton. An obvious obstacle, thus, arises with the presence of proline residues.

[2]In practice, the correlation with the carbons of the preceding amino acid is often observed too because the one-bond and two-bond ^{15}N-$^{13}C^{\alpha}$ interactions are of similar strength.

If the number of prolines is small, the sequential context is usually sufficient to assign the proline sequences. Hardware advancement together with the development of carbon-detected or nitrogen-detected experiments, which do not correlate amide proton, represent an alternative assignment strategy, applicable to proline-rich proteins.

When the backbone atoms are assigned, the procedure is completed by assigning the nuclei in the amino-acid side chains. The most direct and most sensitive experiment used to assign side-chain nuclei is 3D HCCH-TOCSY that correlates side-chain hydrogen nuclei with the side-chain carbons. Starting from the known C^{α} and C^{β} chemical shifts, obtained during the backbone assignment, the remaining chemical shifts for each amino acid can be assigned. However, the spectrum often suffers from severe signal overlaps. Acquisition of additional spectra may considerably facilitate the side-chain assignment in such a case. Examples of such auxiliary experiments are 3D HSQC-TOCSY, 3D H(CC)(CO)NH, and 3D (H)CC(CO)NH. 3D HSQC-TOCSY correlates the amide group ^{1}H and ^{15}N nuclei with the protons in the side chain of the same amino acid, while 3D H(CC)(CO)NH correlates the amide group ^{1}H and ^{15}N nuclei with the side-chain hydrogens from the preceding amino acid. The analogous 3D (H)CC(CO)NH experiment correlates amide group ^{1}H and ^{15}N nuclei with the aliphatic carbons from the preceding amino acid. It should be mentioned that the aforementioned experiments provide only assignment of the aliphatic side chains. The 3D HCCH-TOCSY experiment can be employed to correlate frequencies within aromatic side chains. Experiments connecting aliphatic and aromatic groups have been designed (Sattler et al. 1999), but the atoms of aromatic rings are often assigned based on the through-space interactions (vide infra).

12.3.3 Assignment and Studies of Large Proteins

Overall tumbling of proteins larger than 30 kDa is so slow that the resulting line broadening becomes a serious challenge. Moreover, the high number of residues of large proteins results in more frequent peak overlap. Several methods can be employed to improve the sensitivity and resolution. One possibility is to substitute the side-chain ^{1}H nuclei by ^{2}H. Due to the lower magnetogyric ratio, interactions of the ^{2}H nuclei are weaker, and the resulting relaxation effects are less severe. Sufficiently sharp lines can be obtained for 50 kDa proteins, but an obvious drawback is the absence of side-chain protons, which are essential, for example, for structure calculation (vide infra). For this reason, a partial deuteration is sometimes used, but the presence of the mixture of ^{1}H and ^{2}H isotopes in the CH_2 and CH_3 groups presents additional problems. More sophisticated labeling schemes, combined with site-directed mutagenesis, are used for assigning frequencies and structure determination of large proteins. The proteins are completely deuterated, and protons are re-introduced at few selected positions (Tugarinov and Kay 2003; Kainosho et al. 2006).

Another way to overcome the fast relaxation due to the high molecular mass is to modify the application of radio waves so that the undesired effects of relaxation are suppressed. *Transverse Relaxation Optimized SpectroscopY* (TROSY) may serve as an example (Pervushin et al. 1997). As mentioned in Sect. 12.1.2, interactions of the observed nucleus with other nuclei connected by chemical bonds split the observed peaks. For example, lines of directly bonded ^{1}H and ^{15}N nuclei are split to doublets. The widths of individual lines of the doublet are given by combinations of relaxation effects of interactions with electrons and with the other nucleus: one line is broadened by the sum of the mentioned interactions, the other one by their difference. The splitting is usually eliminated by application of radio waves to obtain single peaks and thus improve the sensitivity and resolution. However, the line widths in large proteins differ so much, that it is beneficial to preserve the splitting and observe only the narrower line. Because the interactions with electrons depend on $\mathbf{B}_0$, the actual line width depends on the magnetic field. For the amide NH group, the optimum is close to $\mathbf{B}_0$ of the strongest commercially available magnets (in 2018). The advantage of the described TROSY approach is that it can be implemented in the standard experiments. Therefore, similar assignment strategy can be used as described for smaller proteins. For even larger proteins (>100 kDa), it is beneficial to correlate frequencies using through-space interactions of magnetic moments, or to combine through-space and through-bond correlations (Riek et al. 1999). The latter approach is known as *cross relaxation-enhanced polarization transfer* (CRINEPT). Deuteration and TROSY/CRINEPT techniques work best when combined. These methods, as shown shortly after their invention (Fiaux et al. 2002), extended the size limit of protein NMR spectroscopy to at least 1 MDa. However, it should be stressed that such studies are far from routine.

The slow tumbling of large proteins can be also accelerated by changing environment of the protein. Decreasing viscosity by elevating temperature is not applicable to all proteins. An interesting possibility is to encapsulate the protein within the aqueous core of a reversed micelle dissolved in a low-viscosity solvent such as compressed pentane or ethane (Wand et al. 1998; Dodevski et al. 2014). Phospholipid bilayer nanodisks provide an opportunity to study membrane proteins by NMR (Hagn et al. 2013). Finally, investigation in the solid state may be advantageous in many cases.

The approaches discussed above were developed to overcome the problem of line broadening that seriously complicates NMR studies of large well-ordered proteins. As mentioned above, the line broadening is not so severe in the case of intrinsically disordered proteins. The major limitation of assignment strategies of disordered proteins is signal overlap. Sufficient resolution can be achieved by observing rotation of magnetization for a longer time, the recording high-dimensional (4D, 5D) spectra, and measuring frequencies of nuclei exhibiting better dispersion of chemical shifts than proton (Nováček et al. 2014). As mentioned in Sect. 12.1.3, multidimensional spectra are obtained by repeating the basic experiment with various lengths of specific delays between pulses of radio waves. To perform high-resolution multidimensional experiments in an acceptable time, the delays cannot be varied in a stepwise manner (a typical setup for conventional 2D and

3D experiments) but nonuniformly. Various schemes of setting up the delays and spectra processing programs such as NMRPipe (Delaglio et al. 1995), SMILE (Ying et al. 2016), SCRUB (Coggins et al. 2012), MDD (Orekhov and Jaravine 2011), NESTA (Sun et al. 2015), CAMERA (Worley 2016), hmsIST (Hyberts et al. 2012), MFT (Stanek and Koźmiński 2010), etc. have been developed, but further description is far beyond the reach of this chapter.

12.4 Structure Determination

12.4.1 Chemical Shifts

A brief look at a protein spectrum might evoke an impression that the chemical shift dispersion is purely random. However, chemical shifts, being results of interactions with surrounding electrons (Fig. 12.1), carry information about the distribution of electron density in the molecule. Detailed analysis of large protein sets revealed the relation between the chemical shifts and local conformation (Wishart et al. 1991; Yao et al. 1997). Chemical shifts represent a source of structural information obtained already during the frequency assignment. Once the backbone assignment is finished, chemical shifts might be used to obtain useful information about the protein. Although both α-helices and β-sheets represent more ordered state, the local environment is different, and as a consequence, the chemical shifts in these secondary structures tend to deviate from the random-coil chemical shifts differently. For example, difference between the C^{α} observed chemical shift and a random-coil chemical shift, so-called secondary chemical shift, in α-helix will be positive while being negative when located in β-sheet. Secondary chemical shifts can be used to calculate the *secondary structure propensities* (Marsh et al. 2006; Hafsa et al. 2015; Camilloni et al. 2012) describing the preference of the given residue to form helical or stretched structures.

In principle, the complete set of assigned chemical shifts should be sufficient to calculate three-dimensional structure of the studied protein. However, successful application of such an approach has been presented only for small proteins (Shen et al. 2008). In practice, chemical shifts are used to estimate the φ and ψ torsion angles based on the structure database search (Shen and Bax 2013; Cheung et al. 2010). Such an information is routinely used as a structural restraint in the structure calculation (vide infra).

12.4.2 Nuclear Overhauser Effect

Magnetic moment of the observed nucleus directly interacts with magnetic moments of nuclei close in space (Fig. 12.1). From the structural point of view, it is important that such an interaction does not require the nuclei to be bound by chemical bonds,

but it is mediated through space, and that is indirectly proportional to the *sixth power* of the *distance* between the two interacting spins. Correlation of two nuclei via the through-space interaction of their magnetic moments is known as the *nuclear Overhauser effect* (NOE). NOE is observable between various close-in-space nuclei. However, the ^{1}H-^{1}H NOE is most useful for structure determination due to a high occurrence of hydrogens in proteins and due to the strength of the ^{1}H-^{1}H interaction being largest among all stable nuclei. NOE has been applied to structural studies of small molecules shortly after its experimental verification, and the idea of using NOE to measure the interatomic distances in proteins tempted the scientists from early 1970s. Nevertheless, the first successful protein structure calculation based on NOE was reported by Wüthrich and coworkers (Williamson et al. 1985) in mid 1980's (and awarded by the Nobel Prize to Kurt Wüthrich in 2002). The first experiments involved selective irradiation of one nucleus and monitoring the signal intensity change of other nuclei in a 1D spectrum, after a delay (called *mixing time*) during which the orientations of the interacting magnetic moments were redistributed. 2D and 3D NOESY experiments, mentioned already in Sect. 12.3.1, are examples of correlated spectroscopy providing quantitative measurement of NOE, and nowadays represent the basis of protein structure calculation. The limiting factor of using NOE in protein structure calculation is the fact that NOE is usually negligible for distances greater than 5–6 Å. Moreover, the through-space correlation involves all magnetic moments close in space. The strength of NOE can be increased by prolonging the mixing time. However, the undesired interaction with more distant nuclei, called *spin diffusion*, becomes more pronounced for longer mixing times. Therefore, the choice of the optimal mixing time depends on whether sensitivity or accuracy of the NOE measurement is more important. Usually, longer mixing times are used if the main purpose of the NOESY experiment is the frequency assignment (Sect. 12.3.1).

The structure calculation of small proteins benefits from the fact that 2D NOESY spectrum is already recorded for the assignment. Once the assignment is completed, cross peaks corresponding to interactions between individual ^{1}H-^{1}H pairs can be identified. The distance information is encoded in the signal strength; therefore, a list of peak intensities, or peak volumes, is created and used as an input for the structure calculation. For larger proteins (>10 kDa), 2D spectra are too crowded to distinguish all ^{1}H-^{1}H pairs. A set of 3D ^{13}C-edited and ^{15}N-edited NOESY-HSQC experiments, combining through-space ^{1}H-^{1}H correlation with through-bond ^{1}H-^{13}C or ^{1}H-^{15}N correlation, is usually recorded.

The total number of NOE cross peaks for a protein with 100 amino acids can be up to several thousands. Manual assignment of such a vast number of peaks is tedious and complicated by ambiguities resulting from similar values of chemical shifts of side-chain protons in different residues. Several software tools have been developed to automatically select and/or assign the NOE cross peaks (Herrmann et al. 2002; Rieping et al. 2007). Although individual programs use different approaches, they share common features. Usually, the peaks are assigned iteratively in parallel with the structure calculation. The least ambiguous peaks are assigned

first, and the initial structural model is calculated. Then, some additional peaks are assigned to fit the structure, and the whole process is repeated.

In principle, simple triangulation can be used to build structural models from known interatomic distances (*distance geometry*). However, relatively large experimental errors of estimated distances make structures calculated in this manner highly imprecise. Therefore, the distances are usually used as *structural restraints* in *molecular dynamics simulations* to obtain structural models with a correct local geometry. Various programs for such structure calculations are available (Güntert and Buchner 2015; Brunger et al. 1998; Schwieters et al. 2003). The process of a structure calculation is commonly based on molecular dynamics using a simulated annealing (Kirkpatrick et al. 1983). The protein model is repeatedly heated and cooled down to reach a global energy minimum. The calculation can be performed either in the *Cartesian space*, where unique coordinates describe position of each atom in the 3D space, or in *torsion space* which relies only on the optimization of backbone torsion angles. Combination of both approaches is used as well. The distance information obtained from the peak intensity is always partially biased (due to spin diffusion, observing averaged NOEs, etc.). Rather than with exactly defined distances, molecular dynamics is performed with distances restrained to a certain range. Although the distance in the calculated structure is within the restrained range, none or minor energy penalty is applied, otherwise energy of the calculated structure is artificially increased. As already mentioned, the reach of a NOE contact is limited to 5–6 Å. Although a relatively small number of NOE restrains can be sufficient to get the overall fold, increasing number of distance restraints can provide more subtle structural information, such as the side chains orientation. In general, the most important restraints are distances between methyl groups and aromatic rings, typically buried inside the protein core and thus defining the fold most directly. The distance restraints are often combined with other types of restraints such as hydrogen bonds derived from identified secondary structure elements, torsion angles (Sect. 12.4.1), and residual dipolar couplings (Sect. 12.4.3). The structure calculation is typically repeated many times in vacuum, and a subset of calculated structures with the lowest energy and best agreement with the experimental restraints is further refined in an explicitly defined solvent.

12.4.3 Residual Dipolar Couplings

As described above, through-space interactions between magnetic moments provide important structural information (NOE), but their coherent effect is averaged to zero if the protein can freely tumble in a solution (Sect. 12.1.2). On the other hand, this type of interaction is very strong in solid samples and needs to be suppressed in solid-state NMR. In solution-state NMR, coherent effects of through-space interactions can be observed as the *residual dipolar couplings* in weakly aligned samples (Sect. 12.1.2).

The partial alignment can be achieved in several ways. The protein of interest can be dissolved in a liquid crystalline phase (made of bicelles, filamentous phages, etc.), placed in a stretched gel, labeled by a paramagnetic tag, and so on. The observed peak splitting (or other manifestation of the coupling) is usually due to the sum of RDC and scalar coupling. Therefore, a reference measurement in an isotropic solution is also needed. By subtraction of the splitting observed in an anisotropic and an isotropic medium, the value of RDC can be obtained. Most often, large RDC values of NH and CH groups are evaluated, but small couplings in other groups are also measurable. The importance of RDC for structure determination lies in a fact that the coupling value depends on the averaged orientation of the internuclear vector in a reference frame. The frame which is defined by the alignment and, in the case of rigid, well-structured proteins, the same for the whole molecule. RDC can therefore be used as another restraint, describing mutual orientation of structural fragments that can be distant in space. It is advantageous to acquire of a large set of RDC data in different media (Hus et al. 2008), but even RDC of one type determined in a single aligning environment represents a useful structural information. Similarly to NOE, RDC restraints are incorporated in the structure determination protocols by penalizing structural models that violate the experimental data.

12.4.4 Validation and Visualization of Calculated Structures

As in every method, the presence of errors in the process of structure determination is inevitable. As the number of structures deposited in the Protein Data Bank is growing constantly (more than 140,000 up to this date), the importance of validations tools increases. The approach to validation is influenced also by the method of structure determination. Various available programs for NMR structure validation (Bhattacharya et al. 2007; Laskowski et al. 1996; Vriend 1990) have been integrated into one suite named CING (Doreleijers et al. 2012). Local geometry (bond lengths, bond angles, packing quality, side-chain planarity) is checked in the same manner as when validating structures determined by diffraction techniques. Particular attention is paid to the torsion angles defining conformation of the backbone (φ and ψ, correlated in so-called Ramachandran plot) and side chains ($\chi 1$ and $\chi 2$, correlated in so-called Janin plot) and to virtual dihedral angles measuring direction of the backbone. Outliers from the normal distribution of the checking parameters are highlighted to identify structural features that are not necessarily wrong but deserve careful inspection. Restraint validation is of course specific for NMR-derived structures. The agreement between the experimental data and the resulting structures is evaluated and found violations are reported. The restraints inconsistent with the structure should be checked and, if they are found to be erroneous, the structure calculation should be repeated with a refined set of input data.

In contrast to crystal structures, results of NMR structure calculation are not presented as single sets of atomics coordinates. Instead, ensembles of 10–20 structural models, representing a selection from hundreds of calculated structures, are

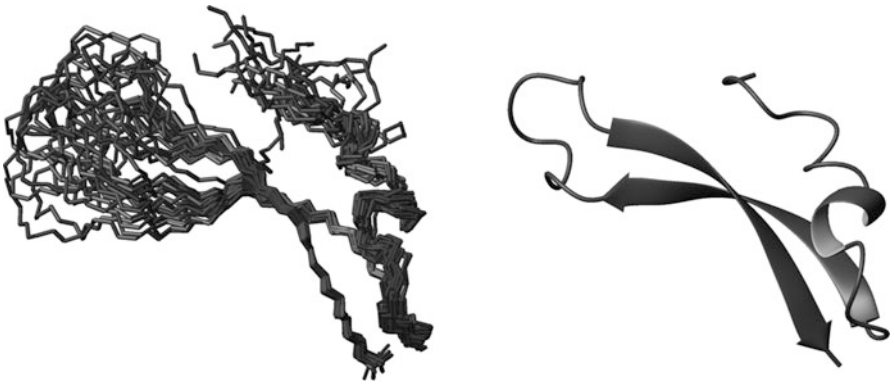


Fig. 12.5 Solution structure of plant peptide hormone regulating stomatal density from *Arabidopsis thaliana*. An ensemble of 20 calculated structures (left) and a representative structure with a ribbon representation of the secondary structure (right) are presented. (Structure solved by Ohki et al. 2011)

deposited. A typical ensemble of superimposed structures is shown in Fig. 12.5. It can be seen that some regions are highly similar in all structures, but some part have various conformations. The differences might be due to missing or incorrect restraints, or they may reflect internal dynamics of the studied molecule. Although low root-mean-square deviation between individual structural models is often interpreted as a sign of high precision of the structure calculation, a more loose representation may better reflect dynamics of the molecule. As an alternative to presenting bundles of structural models, representation of the result of the structure determination in a form of a single "regmean structure" has been proposed (Gottstein et al. 2012).

12.5 Intermolecular Interactions

Various NMR experiments have been designed to investigate complexes of proteins with other proteins (Anglister et al. 2016), nucleic acids (Yadav and Lukavsky 2016), or small ligands (Gossert and Jahnke 2016). Here, we mention only two straightforward approaches.

12.5.1 Chemical Shift Mapping

Once NMR frequencies of a protein are assigned, and its structure is known, interactions of the protein can be monitored very efficiently using *chemical shift mapping*. Protein fingerprint spectra (such as the one in Fig. 12.3) can be recorded in

minutes. In the presence of an interacting ligand, chemical shifts (frequencies) of residues in the vicinity of the interaction site are affected by the binding. Mapping the chemical shift changes of individual residues to the known 3D structure of the protein allows one to quickly identify the interaction interface without any additional structure calculation. In favorable cases, dissociation constant can be estimated by evaluating gradual changes of chemical shifts or peak volumes in a series of spectra recorded at various ligand concentrations. Fast and slow binding can be also distinguished: if the process is slow, peaks of bound and free proteins are detected separately in the spectra, if the interaction is faster than the difference of frequencies of the free and bound form, a single peak is observed at a frequency of the average of chemical shifts of the individual forms, weighted by their populations at the given ligand concentration. Methods described in Sect. 12.6.2 make quantitative description of the binding kinetics possible.

Studies of protein–protein (or protein–nucleic acid) interactions are facilitated by the possibility to selectively label any of the interacting partners with ^{13}C or ^{15}N isotopes. Relatively large complexes can be monitored because isotope labeling keeps the spectra simple (by observing only one component of the complex in each spectrum), and line broadening can be suppressed by TROSY or CRINEPT approaches, combined with deuteration if needed (Sect. 12.3.3).

12.5.2 Isotope-Filtered NOESY

NOE (Sect. 12.4.2) can be also observed between nuclei of different molecules if they are close in space in the complex. NOESY spectra can be recorded for complexes, but the intermolecular distance restraints must be identified among a large excess of intramolecular NOE data. Fortunately, isotope labeling allows one to detect the intermolecular contacts selectively. It is possible to selectively correlate frequency of proton attached to ^{13}C with frequency of ^{12}C-bound protons close in space, without observing through-space interactions of proton magnetic moments in ^{1}H-^{13}C, ^{1}H-^{13}C or ^{1}H-^{12}C, ^{1}H-^{12}C pairs (Anglister et al. 2016). Spectra of a complex consisting of one ^{13}C-labeled and one natural-abundance protein thus contain only peaks of protons of the binding interface. Such experiments provide distance restraints that can be used to solve structure of the complexes at atomic resolution.

12.6 Dynamics

The vast majority of proteins in the cell are not rigid structural elements, but mediators of dynamic processes. The static pictures of the determined protein structures are not sufficient to describe the dynamic functions of proteins, and it is highly desirable to extend molecular characterization of proteins to the description of

their motions. As the molecular motions represent the sole source of NMR relaxation phenomena, *relaxation measurements* offer a unique insight into the protein dynamics.

12.6.1 Fast Motions

The basic observables in the relaxation experiments are relaxation rates R_1 and R_2. R_1 measures the return of orientations of nuclear magnetic moments, perturbed in the NMR experiment, to their equilibrium distribution. R_2 combines rates of two processes, return to the equilibrium and loss of the coherence of the magnetic moment precession. If two nuclei interact, the return to the equilibrium distribution of the orientations of the observed magnetic moment depends also on the actual orientation of its interacting partner. Such a mutual dependence is the already described *nuclear Overhauser effect* (NOE). Rates of many other relaxation processes are measurable and provide useful information, but their discussion is beyond the scope of this text.

Special NMR experiments have been designed to measure R_1, R_2, and NOE quantitatively, separated from other effects (for review, see Korzhnev et al. 2001). This can been achieved most easily for the NH group in the peptide bond, well separated from other nitrogens in the protein molecule. As the NH group is present in almost all amino acids (except for proline), it is also an ideal probe of the backbone dynamics from the structural point of view. Therefore, relaxation of ^{15}N in the peptide bond is measured most often. However, numerous NMR experiments probing dynamics of other atoms in the protein backbone and side chains are also available (Engelke and Rüterjans 1999).

R_1, R_2, and NOE reflect molecular motions on the timescale of 10^{-15}–10^{-8} s. It should be emphasized that the timescale of the NMR experiment is much longer, starting in the millisecond range. As a consequence, the relaxation measurements provide only *statistical description* of the molecular motions. Because three relaxation parameters (R_1, R_2, and NOE) are typically measured for each observed nucleus, the molecular motions are usually described by three parameters too. If the protein has a well-defined structure, its overall tumbling (unrestricted in isotropic solution) and internal motions of the observed nucleus (restricted by the surrounding atoms) can be assumed to be independent. In such a case, the dynamics is typically described by three parameters: by two *correlation times*, specifying timescales of the tumbling of the whole protein and of the internal motion, and by one *order parameter*, defining restriction of the internal motion (value of 0 corresponds to unrestricted motion, value of 1 corresponds to the complete restriction, i.e., to the absence of motion). The correlation time of the overall tumbling represents a physical limit of monitoring internal motions, only faster internal changes can be observed. In practice, correlation times of internal motions in the range of 10^{-11}–10^{-10} s can be determined, whereas the order parameter can be measured for any motion faster than the correlation time of the overall tumbling. The outlined

procedure obviously represents a simplification of the real molecular dynamics, including multiple motional modes, but typically provides a very useful dynamic picture of the molecule with the atomic resolution. More advanced models and alternative approaches have been developed to treat cases when the assumption of two independent motions is not acceptable (Ishima and Nagayama 1995; Buevich and Baum 1999; Meirovitch et al. 2010).

12.6.2 Exchange Processes

Molecular processes slower than the overall tumbling of the protein can be also observed if they change distribution of electron density in the proximity of the observed nucleus. Such events are known as *conformational* or *chemical exchange*. The chemical exchange may include binding processes, (de)protonation, or other chemical reactions. Exchange occurring on the micro- to millisecond timescale leads to a loss of the coherence of the precession motion and, therefore, affects R_2. The relaxation experiments designed to measure R_2 (mentioned in Sect. 12.6.1) try to suppress the exchange effects by applying carefully programmed pulses of radio waves or by continuous irradiation. On the other hand, the exchange processes can be studied by deliberately decreasing efficiency of the suppressing radio waves and evaluating the resulting increase of the apparent R_2. Such experiments are known as *relaxation dispersion measurements*. Modulation of the power of a continuously applied radio wave is called the $R_{1\rho}$ *experiment*, and the approach based on variation of delays between radio-wave pulses is named after its inventors the *Carr–Purcel–Meiboom–Gill (CPMG)* experiment (Korzhnev et al. 2001). In an ideal case, the relaxation dispersion provides equilibrium and rate constants of the exchange, and the difference of precession frequencies of the individual exchanging forms. These parameters can be obtained even if the population of the minor form is too low to be observed directly. Therefore, relaxation dispersion is not only a method of monitoring the exchange rates but also a technique making structural studies of minor states of proteins possible. The timescale of exchange processes observable by the relaxation dispersion is approximately 10^{-4}–10^{-3} s for $R_{1\rho}$ and 10^{-3}–10^{-2} s for CPMG.

Chemical exchange saturation transfer (CEST), a technique based on irradiating the sample with low-power radio-waves selectively affecting the minor state, represents a noteworthy alternative to the relaxation dispersion, less sensitive to unwanted interference of nearby nuclei and better covering the slower processes (Vallurupalli et al. 2012).

In addition to the time windows described above, NMR can be used to monitor motions and events at other timescales. *Exchange spectroscopy*, an approach utilizing the fact that exchange processes create correlations observable in NOESY spectra (Ernst et al. 1987), is sensitive to exchange events on the timescale between 10^{-5} s and 10^{-1} s. *Real-time* measurements, with the time resolution shorter than 1 s, can capture slow processes. On the other hand, the gap between the nano- and microsecond timescales can be bridged by *RDC measurements*, able to provide order

parameters for processes on the timescales ranging from picosecond to millisecond (Briggman and Tolman 2003).

Acknowledgment Financial contribution made by the Ministry of Education, Youths, and Sports of the Czech Republic within special support paid from the National Programme for Sustainability II funds, project CEITEC 2020 (LQ1601), is gratefully acknowledged.

References

Anglister J, Srivastava G, Naider F (2016) Detection of intermolecular NOE interactions in large protein complexes. Prog Nucl Magn Reson Spectrosc 97:40–56

Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66(4):778–795

Briggman KB, Tolman JR (2003) De novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy. J Am Chem Soc 125(34):10164–10165

Brunger AT, Adams PD, Clore GM, Gros P, Grosse-Kunstleve RW, Jiang J, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system (CNS), a new software suite for macromolecular structure determination. Acta Cryst D54:905–921

Buevich AV, Baum J (1999) Dynamics of unfolded proteins: incorporation of distributions of correlation times in the model free analysis of NMR relaxation data. J Am Chem Soc 121:8671–8672

Camilloni C, De Simone A, Vranken WF, Vendruscolo M (2012) Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. Biochemistry 51(11):2224–2231

Cavanagh J, Skelton NJ, Fairbrother WJ, Rance M, Palmer AG (2007) Protein NMR spectroscopy: principles and practice, 2nd edn. Academic Press, Cambridge

Cheung M, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J Magn Reson 202(2):223–233

Coggins BE, Werner-Allen JW, Yan A, Zhou P (2012) Rapid protein global fold determination using Ultrasparse sampling, high-dynamic range artifact suppression, and time-shared NOESY. J Am Chem Soc 134:18619–18630

Crespi HL, Rosenberg RM, Katz JL (1968) Proton magnetic resonance of proteins fully deuterated except for H-leucine side chains. Science 161:795–796

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Dodevski I, Nucci NV, Valentine KG, Sidhu GK, O'Brien ES, Pardi A, Wand AJ (2014) Optimized reverse micelle surfactant system for high-resolution NMR spectroscopy of encapsulated proteins and nucleic acids dissolved in low viscosity fluids. J Am Chem Soc 136(9):3465–3474

Doreleijers JF, Sousa da Silva AW, Krieger E, Nabuurs SB, Spronk CA, Stevens TJ, Vranken WF, Vriend G, Vuister GW (2012) CING: an integrated residue-based structure validation program suite. J Biomol NMR 54(3):267–283

Engelke J, Rüterjans H (1999) Recent developments in studying the dynamics of protein structures from and ^{15}N and ^{13}C relaxation time measurements. In: Krishna NR, Berliner L (eds) Structure computation and dynamics in protein NMR. Kluwer Academic/Plenum, New York

Ernst RR, Bodenhausen G, Wokaun A (1987) Principles of nuclear magnetic resonance in one and two dimensions. Oxford University Press, Oxford

Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) NMR analysis of a 900K GroEL–GroES complex. Nature 418:207–211

Gossert AD, Jahnke W (2016) NMR in drug discovery: a practical guide to identification and validation of ligands interacting with biological macromolecules. Prog NMR Spectrosc 97:82–125

Gottstein D, Kirchner DK, Güntert P (2012) Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. J Biomol NMR 52(4):351–364

Güntert P, Buchner L (2015) Combined automated NOE assignment and structure calculation with CYANA. J Biomol NMR 62:453–471

Hafsa NE, Arndt D, Wishart DS (2015) CSI 3.0: a web server for identifying secondary and supersecondary structure in proteins using NMR chemical shifts. Nucleic Acids Res 43(W1):W370–W377

Hagn F, Etzkorn M, Raschle T, Wagner G (2013) Optimized phospholipid bilayer Nanodiscs facilitate high-resolution structure determination of membrane proteins. J Am Chem Soc 135 (5):1919–1925

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319(1):209–227

Hus JC, Salmon L, Bouvignies G, Lotze J, Blackledge M, Brüschweiler R (2008) 16-fold degeneracy of peptide plane orientations from residual dipolar couplings: analytical treatment and implications for protein structure determination. J Am Chem Soc 130(47):15927–15937

Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. J Biomol NMR 52(4):315–327

Ishima R, Nagayama K (1995) Protein backbone dynamics revealed by quasi spectral density function analysis of amide N-15 nuclei. Biochemistry 34:31623171

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57

Keeler J (2010) Understanding NMR spectroscopy, 2nd edn. Wiley, Chichester

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220 (4598):671–680

Korzhnev DM, Billeter M, Arseniev AS, Orekhov VY (2001) NMR studies of Brownian tumbling and internal motions in proteins. Prog NMR Spectrosc 38(3):197–266

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8(4):477–486

Levitt MH (2008) Spin dynamics: basics of nuclear magnetic resonance, 2nd edn. Wiley, Chichester

Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. Protein Sci 15(12):2795–2804

Meirovitch E, Shapiro YE, Polimeno A, Freed JH (2010) Structural dynamics of biomacromolecules by NMR: the slowly relaxing local structure approach. Prog Nucl Magn Reson Spectrosc 56:360405

Nováček J, Žídek L, Sklenář V (2014) Toward optimal-resolution NMR of intrinsically disordered proteins. J Magn Reson 241:41–52

Ohki S, Takeuchi M, Mori M (2011) The NMR structure of stomagen reveals the basis of stomatal density regulation by plant peptide hormones. Nature Communications 2:512. https://doi.org/10.1038/ncomms150

Orekhov VY, Jaravine VA (2011) Analysis of non-uniformly sampled spectra with MultiDimensional decomposition. Prog Nucl Magn Reson Spectrosc 59:271–292

Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci U S A 94:12366–12371

Primrose WU (1993) Sample preparation. In: Roberts GCK (ed) NMR of macromolecules, a practical approach. Oxford University Press, Oxford
Putter I, Barreto A, Markley JL, Jardetzky O (1969) Nuclear magnetic resonance studies of the structure and binding sites of enzymes, X. preparation of selectively deuterated analogs of staphylococcal nuclease. Biochemistry 64:1396–1403
Riek R, Wider G, Pervushin K, Wüthrich K (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. Proc Natl Acad Sci U S A 96 (9):4918–4923
Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics 23:381–382
Rule GS, Hitchens TK (2006) Fundamentals of protein NMR spectroscopy. Springer, Dordrecht
Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158
Schwieters CD, Kuszewski J, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:66–74
Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR 56:227–241
Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci U S A 105(12):4685–4690
Stanek J, Koźmiński W (2010) Iterative algorithm of discrete Fourier transform for processing randomly sampled NMR data sets. J Biomol NMR 47(1):65–77
Sun S, Gill M, Li Y, Huang M, Byrd RA (2015) Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms. J Biomol NMR 62(1):105–117
Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc 125 (45):13868–13878
Vallurupalli P, Bouvignies G, Kay LE (2012) Studying "invisible" excited protein states in slow exchange with a major state conformation. J Am Chem Soc 134(19):8148–8161
Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8 (1):52–56. 29
Wand AJ, Ehrhardt MR, Flynn PF (1998) High-resolution NMR of encapsulated proteins dissolved in low-viscosity fluids. Proc Natl Acad Sci U S A 95(26):15299–15302
Williamson MP, Havel TF, Wüthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. J Mol Biol 182(2):295–315
Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222(2):311–333
Worley B (2016) Convex accelerated maximum entropy reconstruction. J Magn Reson 265:90–98
Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York
Yadav DK, Lukavsky PJ (2016) NMR solution structure determination of large RNA-protein complexes. Prog Nucl Magn Reson Spectrosc 97:57–81
Yao J, Dyson HJ, Wright PE (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. FEBS Lett 419(2–3):285–289
Ying J, Delaglio F, Torchia DA, Bax A (2016) Sparse multidimensional iterative Lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. J Biomol NMR 68(2):101–118
Zhang HY, van Ingen H (2016) Isotope-labeling strategies for solution NMR studies of macromolecular assemblies. Curr Opin Struct Biol 38:75–82

Chapter 13
Structural Biology Using Electron Microscopy

Chikara Sato, Takao Shinkawa, Mari Sato, Masataka Ohashi, Mitsuru Ikeda, Masaaki Kawata, Kazuhiro Mio, and Masanori Koshino

13.1 Introduction

The number of structures solved using electron microscopy (EM) is increasing rapidly; the number registered in the Protein Data Bank (PDB; http://www.rcsb.org/pdb/) is over 3100 volumes and 900 models in September 2015 and is increasing year by year (Morikawa et al. 2015; Lawson et al. 2016). The recent acceleration is attributable to several factors: improved cryo-transmission electron microscopes (cryo-TEMs), newly developed electron detectors (direct electron detection camera:

C. Sato (✉)
Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki Prefecture, Japan
e-mail: ti-sato@aist.go.jp

T. Shinkawa
BioNet Laboratory Inc., Tokyo, Japan

M. Sato
Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

M. Ohashi
Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki Prefecture, Japan

BioNet Laboratory Inc., Tokyo, Japan

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_13

DDC), the development of moderate biochemical purification protocols, structurally stable detergent substitutes including amphipols and nanodisks, the construction of new algorithms, increased computational ability (Beck et al. 2012; Liao et al. 2013; Zhou 2011), and the funding of many national projects. Among them, the developments in DDC and cryo-TEM have been critical, pushing the resolution limit to the atomic level in 2013 by Cheng's group (Cao et al. 2013; Liao et al. 2013). At this time, almost 50 years have already passed since De Rosier and Klug showed three-dimensional (3D) helical reconstruction from two-dimensional (2D) projections obtained by a transmission electron microscope (TEM) (De Rosier and Klug 1968).

Single particle analysis (SPA) has been developed by many researchers, starting from highly symmetric large protein complexes and spherical viruses (Bottcher et al. 1997; Frank 2006; van Heel et al. 2000) and ribosomes (Frank et al. 1995; Stark et al. 1995), and further applied to smaller asymmetric proteins and membrane proteins. The analysis of ribosomes and viruses has an advantage of the higher contrast and/or the higher point symmetry. The attainable resolution by SPA using cryo-TEM has reached to 2.2 Å (Bartesaghi et al. 2015) and further to better than 2Å recently (Borgnia et al. 2016).

13.2 Materials

Instruments and image processing software necessary for SPA are listed in the following.

M. Ikeda
Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki Prefecture, Japan

BioNet Laboratory Inc., Tokyo, Japan

Terabase Inc., Okazaki, Aichi, Japan

M. Kawata
Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

K. Mio
Advanced Operando-Measurement Technology Open Innovation Laboratory (OPERANDO-OIL) and Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan

M. Koshino
Nano Materials Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

13.2.1 Cryo-TEM

13.2.1.1 General

Cryo-TEM was developed specifically to realize stable observation of samples embedded in vitrified ice (Fig. 13.1) without frosting. In cryo-TEM, the electron beam generated at the electron gun at the top is accelerated down to the microscope column, radiates through the sample, and impinges on the detector (Fig. 13.2). Because electrons are scattered by air, the microscope column is kept at high vacuum, $10^{-5} \sim 10^{-7}$ Pa. The enlarged image of the specimen obtained can be checked using the phosphor screen at the bottom of the column or recorded on a detector (Sects. 13.2.2 and below).

13.2.1.2 Manufacturers

The decision as to which cryo-TEM to purchase for SPA is a big step that requires careful consideration. There are at least three leading cryo-TEM manufacturers to be consulted.

- JEOL Ltd., 1–2, Musashino 3-chome Akishima, Tokyo 196–8558, Japan, (http://www.jeol.co.jp/en/).
- Hitachi High-Technologies Corporation, 24–14, Nishi-Shimbashi 1-chome, Minato-ku, Tokyo 105–8717, Japan, (http://www.hitachi-hitec.com).
- FEI, Thermo Fisher Scientific Materials & Structural Analysis Division, 5350 NE Dawson Creek Drive, Hillsboro, Oregon 97,124, USA (http://www.fei.com).

High-resolution SPA can be achieved not only by high-performance cryo-TEM itself but also by continuous attention to maintenance of stable environments surrounding TEM, including a temperature-controlled room with soundproof walls

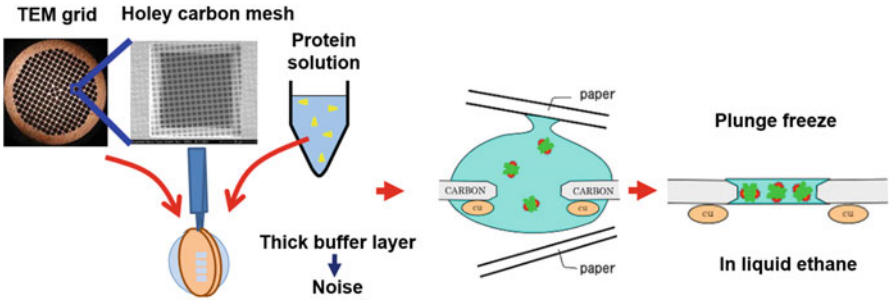


Fig. 13.1 Sample preparation for cryo-transmission electron microscopy (cryo-TEM). In this figure, membrane proteins (green) are associated with membrane lipids and detergent (red). A 3-μl droplet of the purified protein solution was applied to a holey carbon grid. The grid was partially blotted with filter paper to remove excess solution, immediately plunged into liquid ethane chilled by liquid nitrogen and transferred into the cryo-TEM

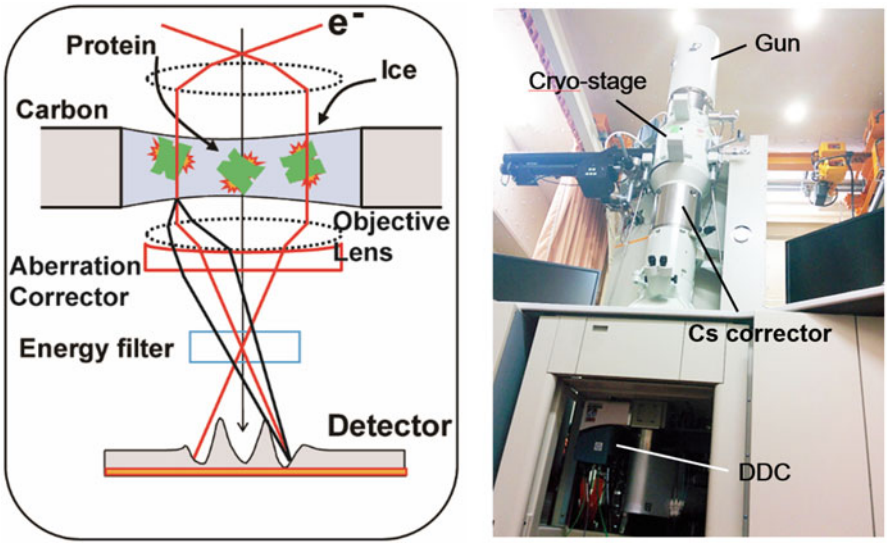


Fig. 13.2 Schematic diagram of cryo-TEM for structural biology. The cryo-transferred sample grid was kept under liquid nitrogen temperature in the cryo-TEM. Low-dose (<47 electrons per $Å^2$) images were recorded by DDC typically at a magnification of 20,000–40,000 and with a 200–300 kV acceleration voltage

and ceiling, a magnetic field canceler, a vibration isolator, and so on. It is very important and critical to develop measures to establish acceptable environments before and after cryo-TEM installation.

Design of cryo-TEM is directly related to the image quality as well as productivity of mass data collection for SPA. A top-entry specimen stage has more stability than a side-entry holder as the former places the specimen inside the TEM column whereas the latter has one end inside the TEM column and the other end outside the microscope. The latter is also more influenced by outer pressure variations or possible vibrations from the cryogenic dewar attached to the outer end. It is still possible to achieve atomic resolution by side-entry type cryo-holders. Side-entry type cryo-holders are available not only from TEM manufacturers but also from other manufacturers such as Gatan, Oxford Instruments, and Fischione Instruments. Side-entry type cryo-holders are more flexible than top-entry stages as they can be installed on microscopes with different pole piece gaps suitable for high-resolution imaging, analytical TEM, and 3D tomography. Some types of cryo-TEM are designed to incorporate the advantages of both top-entry and side-entry holders.

13.2.1.3 Automation

As large numbers of data sets are required for SPA, many researchers have demanded an automatic sample loader system as well as an automatic image acquisition system

for mass data collection. Titan Krios from FEI and CRYO ARM 200 (JEM-Z200FSC n.d.) from JEOL, for example, provide such an automatic sample loading and data acquisition system, accelerating SPA for 3-D reconstruction.

Development of an automatic specimen loading system, in principle, requires the participation of a primary TEM manufacturer. A cryo-EM with an automatic specimen loader system, initially developed by FEI and recently followed by JEOL, can store a dozen of cryo-samples in a magazine kept below freezing point in a vacuum and transfer one of them without exposing "any possible contamination."

Development of automatic data acquisition software, on the other hand, has involved both cryo-EM companies and third-party developers. Some available up-to-date software programs are, for example, EPU by FEI, Jeol automated data acquisition system (JADAS) by JEOL, SeralEM by Prof. Mastronardes group (University of Colorado, Boulder), and Liginon by National Resource for Automated Molecular Microscopy (NRAMM). To control the microscope using by software, it is necessary to access the specimen stage control, to acquire images with regard to the specimen position, beam alignment, and defocus condition, and to feed back the corrected alignment to the microscope. There is still room for improvement of both hardware and software, feedback from users playing an important role in this.

13.2.1.4 Resolution Limit

Based on Abbe's principle (Abbe 1873), the attainable resolution does not basically exceed a half wavelength used for imaging. In optical microscopy (OM), a resolution of ca. 200 nm has been achieved with aberration-free objective lens and has been improved far more with recent developments in super-resolution microscopy. The spatial resolution of TEM, on the contrary, is much better than that of OM due to the shorter wavelength of electrons, for example, 49 pm at 60 kV and 1 pm at 300 kV, compared with that of visible light (350 nm<). The actual spatial resolution of TEM has not yet reached a half wavelength of electrons but has become smaller and smaller, less than 100 pm at 60 kV (Linck et al. 2016; Sasaki et al. 2010) and 50 pm at 300 kV (Kisielowski et al. 2008) owing to the development of aberration-correctors (Haider et al. 1998), which have become very popular in material science EM.

In SPA of cryo-TEM, on the other hand, a resolution of far better than 2 Å has never been attained, which is mainly due to the limited electron dose tolerant for fragile protein samples, the low signal-to-noise ratio (S/N) typical of ice-embedded protein, and a large defocus condition easily obstructing the proper correction of the phase contrast transfer function (CTF) at higher frequency (= high-resolution information), which will be described in the following Sect 13.3. In SPA of protein particles, for example, a number of projected images with similar orientation are classified and averaged to improve the S/N, and in addition, noise or random information from vitrified ice is canceled out.

The acceleration voltage employed for SPA in cryo-TEM varies by users, typically ranging from 200 to 300 kV. The higher acceleration voltage enables electrons

to be transmitted through thicker ice layer. The target size of protein particles differs from sample to sample: a typical viruses ca. 100 nm or less in diameter, whereas some proteins for drug design including G protein-coupled receptor (GPCR) are much smaller than viruses, from a few nm to several tens of nm in size. To obtain a high S/N ratio of cryo-TEM images, it is critical to tune the thickness of the ice layer so as to make it as thin as possible but not so thin that it affects the target protein. The larger the target size of protein becomes, the higher the acceleration voltage is needed. The higher acceleration voltage (higher transmitting power) means less interaction with the specimen, resulting in a low S/N ratio of the images. If we tune the thickness of ice so that it is just somewhat larger than the size of the target proteins or complexes, it will be advantageous to use lower acceleration voltage because more electrons will contribute to signal (contrast) with the same dose of electrons. There is a controversy as to which acceleration voltage should be used in terms of radiation damage: a lower acceleration voltage produces more radiolysis, whereas a higher acceleration voltage causes atomic displacement and sputtering that is not negligible above the threshold energy of element (Hobbs 1984; Egerton 2004). Roughly speaking, elastic scattering contributes to diffraction contrast whereas inelastic scattering results in the excitation of electrons leading to the formation of phonons, ions, radicals, and so on. This is one of the reasons that the zero-loss energy filtered TEM image has a clearer contrast. The ratio of elastic/inelastic scattering is indifferent to the acceleration voltage (Egerton 2011), which indicates that a much higher contrast can be obtained at lower acceleration voltage as more electrons contribute to elastic scattering with the same electron dose irradiated on the sample.

13.2.2 Electron Detecters

There is a clear correlation between the emergence of DDC and the increasing number of published reports on high-resolution 3D structural determination in biological samples (Schroder 2015). People are aware of the critical importance of DDC, especially in its high-sensitive imaging, where both new hardware and software technologies have been implemented.

Before DDC, a traditional recording device commonly used for SPA was highly sensitive photographic film (e.g., Kodak electron microscope film SO-163) combined with a film digitizer (e.g., Leaf Scan L45 scanner). The recording size is comparable with the TEM screen, and ideally, the resolution should be close to the grain size in the coat of polyester base. A relatively long exposure time, however, easily caused image blurring due to specimen stage drift, beam-induced motion of materials or other factors, even in a low-dose condition (Brilot et al. 2012). In addition, the development of the film requires a certain level of experience. From this practical background, real-time imaging is demanded so as to provide feedback on the imaging condition from acquired data.

Easier access to cryo-TEM images is made possible by a charge-coupled device (CCD) several megapixels in size. Traditionally, CCD has a scintillator on its surface, converting electrons to photons that are transferred by fiber optics or coupled lens to detectors where photons are finally converted to charge (signals). The intermediate process causes the possible loss of signal, noise mixing, and larger point spread function (PSF). In principle, CCD requires readout (= beam blanking) time for electrons not to irradiate the detector, and thus, temporal resolution is limited by the size of the data transfer as well as the speed of the shutter.

Advances in complementary metal–oxide–semiconductor (CMOS) have resulted in CMOS taking over the market of CCD as one of the major manufactures decided to stop its CCD production and shifted to CMOS (Allied 2015). Each pixel of CMOS contains both a photodetector and an active amplifier that are addressed and readout individually, enabling no beam blanking and fast acquisition. There is a type of CMOS camera that possesses a scintillator similar to CCD, such as the OneView Camera by Gatan and the TemCam-F series by TVIPS. That could be the only choice, for low voltage TEM because DDC does not yield enough signals. The DDC has also been developed on the basis of a CMOS sensor but without a scintillator: direct conversion of electrons to charge is realized. A modern thin DDC used for TEM generally has high detective quantum efficiency (DQE), higher sensitivity, a very small PSF, and a rapid readout, leading to short exposure time (Grigorieff 2013; Jin et al. 2008; Li et al. 2013b; Milazzo et al. 2010).

Another benefit of DDC is electron counting, the methodology of which is based on the quantum effect and super-resolution microscopy. As described in the literature on the Gatan K2 Summit (Li et al. 2013a, b), the virtual pixel size is four times larger than the real pixel size when identifying the position of electrons to one quarter of each pixel with the "super-resolution" mode. The technology exactly matches the requirement of SPA researchers for the recording of a large number of particles in a large viewing area with higher pixel resolution. The electron dose rate used for electron counting also meets the range of imaging conditions of SPA, typically less than 50 electrons $Å^{-2}$. A dose rate less than 8 electron per real pixel can achieve, i.e., 2 electrons $pixel^{-1}$ s^{-1} on a virtual pixel with a maximum acquisition rate of 400 frames per second (fps). In addition to DDC, mass data transfer and its processing are realized by additional independent hardware such as a digitizer, a processor, and an interface PC server in the total DDC system, in which basic noise reductions such as gain normalization and dark current subtraction are implemented in addition to the motion correction.

As for the software, the motion correction of particles is a key for high-contrast and high-resolution imaging without image blurring, based on image alignment and averaging technique. It greatly enhances the initial quality of images for SPA.

Major DDC products are listed as follows:

K2 summit and K3: Gatan Inc., 5794 W. Las Positas Blvd., Pleasanton, CA 94588, USA.

Falcon 2 and 3: FEI, Thermo Fisher (As described above).

DE-Series: Direct Electron, 13,240 Evening Creek Dr. S., Ste. 311, San Diego, CA 92128, USA.

13.2.3 Ice-Embedded Sample Preparation System

For thin ice embedding, a protein solution is applied to an EM grid with a perforated carbon film (holey grid). Excess liquid is blotted away using filter paper, and the thin aqueous layer (<500 nm or less) is immediately frozen by plunging the grid into liquid nitrogen-cooled ethane slush (Fig. 13.1, right) either manually or using a plunge freezer system. Adequate thickness of the buffer layer relative to the protein particle size is critical to obtain high S/N images of the particles. This process can be done manually, or with a robot hand to enhance the reproducibility of the thickness. Hand paper blotting removes part of the protein solution on the carbon grid to realize adequate thickness for a clear image of the protein particles against the ice background. The sample grid is immediately plunged frozen. Program-controlled robot-arm movements in a controlled environment can perform this process more systematically. The major products using robot-arm paper blotting and plunge freezing are listed as follows.

Vitrobot (FEI).
EM GP Automatic Plunge Freeze (Leica).
CP3 (Gatan).

Various carbon film EM grids are commercially available for this purpose. Quantifoil holey carbon film grids (Quantifoil Micro Tools GmbH) are popular for SPA in cryo-EM. As carbon film itself is hydrophobic, an etching device makes the film hydrophilic by glow discharge, which is indispensable for the protein solution to evenly spread on the surface of the TEM grid. Such etching instruments, sometimes capable of metal coating, can be purchased from companies specializing in EM equipment.

13.2.4 Image-Based 3D Reconstruction Through Statistical Analysis

Three-dimensional reconstruction from EM images was first studied using helical assemblies of macromolecules (De Rosier and Klug 1968) because the regular arrangements of those protein complexes amplify necessary signals for reconstruction, even using noisy projections. Later, numerous reconstructions were reported for various types of single particles, such as icosahedral viruses (Bottcher et al. 1997), asymmetric ribosome complex (Frank and Agrawal 2000), and isolated protein molecules. Later, using an algorithm for helical reconstruction (Ogura et al. 2014) developed from a SPA algorithm (Ogura and Sato 2006), a higher resolution structure of microtubule (Yajima et al. 2012) and its complex with motor protein kinesin 5 (Morikawa et al. 2015) were determined.

For SPA 3D reconstruction, several software packages are available:

- RELION2 (http://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page) (Scheres 2012).
- XMIPP (http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/WebHome) (Marabini et al. 1996; Scheres et al. 2008; Sorzano et al. 2004),
- FREALIGN (http://grigoriefflab.janelia.org/frealign) (Grigorieff 2007),
- IMAGIC (https://www.imagescience.de/imagic.html) (van Heel et al. 1996, 2011),
- Spider (http://spider.wadsworth.org/spider_doc/spider/docs/spider.html) (Frank et al. 1996; Shaikh et al. 2008),
- EMAN1 and EMAN2 (http://blake.bcm.edu/emanwiki/EMAN) (Ludtke et al. 1999; Tang et al. 2007),
- Eos (http://www.yasunaga-lab.bio.kyutech.ac.jp/Eos/index.php/Main_Page) (Yasunaga and Wakabayashi 1996),

Computational power can also limit the efficiency of the analysis, especially when the size of the target protein is very large, or the target resolution is very high. To address these issues, parallel processing using a computer cluster has been employed. A message passing interface (mpi) efficiently accesses multiple cores and CPUs, accelerating the speed of calculations. Recently, graphics processing unit (GPU) computing such using Compute Unified Device Architecture (cuda) by NVIDIA Co. has rapidly become popular due to its lower initial cost and faster speed, typically several times to several hundred times faster, compared with high-end CPUs only. Relion 2, for example, uses both mpi and cuda technology, running on the Linux system, shortening the time of data processing of SPA, important for practical use.

13.3 Methods

13.3.1 Purification of Proteins or Protein Complexes

Because cryo-TEM images all the solubilized molecules, isolated proteins should be pure with minimum degradation. When the target is a membrane protein, it is recommended that an affinity purification step be performed at an early stage of the purification process. The number of purification steps should be as limited as possible to minimize the sample loss due to adsorption of the membrane proteins by purification materials including gels and filters and housings. For any proteins or protein complexes, it is recommended that a size exclusion chromatography (SEC) step can be carried out immediately before the sample is absorbed by the EM grid. This step efficiently removes molecules with different mobilities due to partial denaturation, degradation, or subunit dissociation, thereby ensuring that it is homogeneous. In the following, we discuss the purification of membrane-integrated proteins known to be difficult to purify and make several suggestions.

13.3.1.1 Purification of Membrane Proteins

Membrane proteins such as receptors, ion channels, and pumps have at least one, and sometimes more than seven, membrane spanning regions with extracellular and intracellular domains. Membrane proteins extracted from the cell membrane using detergents are sometimes unstable and denature easily. The best detergent for protein extraction is not always the best to ensure stability: the detergent for solubilization from the membrane can be different from the detergent for the following purification. To find a detergent suitable for the target protein, protein aggregation can be used as a marker to detect inadequate detergents. For example, fluorescence-detection SEC (FSEC) using a fluorescence tag on the target protein is effective (Kawate and Gouaux 2006), as additional peaks with higher molecular weight appear if there is significant denaturation or aggregation.

13.3.1.2 Intermediate SEC to Screen for Adequate Purification Conditions

Careful comparison of the protein peak during purification using standard SEC is also available to detect protein aggregation, a peak shift to higher molecular weight being a sign of aggregation. Solubilized proteins should be handled in aqueous solutions containing detergent at least three times above the critical micelle concentration (CMC) to minimize denaturation. The addition of glycerol or sucrose (up to 50%) to the protein sample sometimes helps to minimize absorption loss and unstability of proteins, especially during chromatography and ultrafiltration.

13.3.1.3 Combination of Purification Steps for Membrane Proteins

A simple combination of affinity purification followed by SEC could be idealistic for membrane protein purification (Sato et al. 2001). Polyclonal antibodies against cytoplasmic tails or linkers in the target proteins are ideal for immuno-affinity chromatography (Sato et al. 2001) because, in our experience, finding monoclonal antibody clones appropriate for affinity-column purification is sometimes difficult, because the affinity of most monoclonal antibodies to the target protein is relatively low. The detergent can be replaced with another amphipathic reagent, for example, Amphipol (Anatrace), which is more structurally uniform and facilitates the following image processing.

13.3.1.4 Negative Staining to Check Protein Quality

To monitor the condition of the proteins, negatively stained TEM is quick and effective, and the degree of protein aggregation can be also checked. In this method,

the protein is absorbed by the carbon film of an EM grid, stained by a heavy salt solution, and dried. It gives a negative contrast in TEM (Bremer et al. 1992). Those who are not familiar with negatively stained TEM of detergent-rich protein samples are recommended to observe protein aggregation in aqueous buffer using the atmospheric scanning electron microscope (ASEM) named Clair Scope (Nishiyama et al. 2010), as described in Sect. 13.4.2.

13.3.2 Protein Particle Embedding into a Thin Ice Layer for Cryo-EM

To obtain a better S/N ratio, the thickness of the vitrified ice layer is important when imaging particles at high contrast. To avoid production of crystalline ice in the layer, the protein solution should be rapidly cooled using liquid ethane or propane or their mixture cooled by liquid nitrogen (<−195.8 °C).

To prepare samples in vitrified ice, approximately 1 - 3 μl of aqueous solution containing target proteins is applied to an EM grid with a perforated carbon film (holey grid) (Fig. 13.1 left). Excess liquid is blotted away using filter paper. The blotting time should be longer when the protein concentration is higher and can be even longer according to the addition of glycerol (Fig. 13.1). At the moment when the residual protein suspension spans most of the small holes in the perforated carbon film, this thin aqueous layer is rapidly cooled by plunging the grid into liquid nitrogen-cooled liquid ethane. The thickness of the buffer layer (recommended to be <30 nm) can be measured using electron energy loss spectroscopy (EELS) using energy-filtered TEM (FTEM).

This process can be done manually or in a more controlled manner using a systematic robot-arm paper blotting system combined with a plunge freezer. The proteins become embedded in vitrified ice in a close-to-native state (Adrian et al. 1984; Taylor and Glaeser 1976).

13.3.3 Cryo-Transfer

After the ice embedding, the frozen sample grid must be transferred to a cryo-EM without frosting (Henderson 2004). This cryo-transport is a critical step for the success of cyro-electron microscopy. A subsystem attached to a blotting/plunge freezer system (Sect 13.2.3) or an antifrost protection system combined with a side-entry cold-holder could help to transport a cryo-sample successfully. The cryo-transfer instrument could protect the cryo-sample from moisture and transfer it to the cryo-EM at liquid nitrogen temperature. Fully manual transport can also be performed using a sample immersed in liquid N_2 in a styrofoam box. For imaging, a grid or multigrids can be stored in a cryo-TEM for a certain period. Each grid is

observed in the cooled stage of a cryo-EM-specified microscope at liquid-nitrogen or liquid-helium temperature (Fujiyoshi et al. 1991), or in a side-entry cold-holder developed for standard TEM.

13.3.4 Data Processing

The following data processing can be performed using one of the image processing packages for SPA (Sect 13.2.4): RELION (Scheres 2012), XMIPP (Marabini et al. 1996; Scheres et al. 2008; Sorzano et al. 2004), FREALIGN (Grigorieff 2007), IMAGIC (van Heel et al. 1996, 2011), Spider (Frank et al. 1996; Shaikh et al. 2008), EMAN (Ludtke et al. 1999; Tang et al. 2007), and Eos (Yasunaga and Wakabayashi 1996) or their combination. The analytical strategy and required hardware are various. Here, we outline the basic concept and analytical flow of SPA (Fig. 13.3).

13.3.4.1 Contrast Transfer Function

Experimentally obtained TEM images are modulated by the contrast transfer function (CTF) that is determined mainly by the parameters of the microscope, including the acceleration voltage used and the spherical aberration constant, Cs. Thus, the image processing starts from CTF correction. Because organic materials such as proteins (1.37 g/ml) have just slightly higher density than vitrified ice, they are observed under defocus conditions of usually reaches more than several thousand Å to enhance the contrast. Together with these effects and other effects derived through the weak-phase approximation in the elastic scattering (Frank 1992), the image formation involved in TEM is always accompanied by a modulation, i.e., the CTF. It should be corrected as this is critical for the interpretation of spatial frequencies beyond the first zero of the CTF. In the Fourier transform of the observed image, the CTF is multiplied in its spatial frequency component, consisting of phase contrast and amplitude contrast. In practice, before CTF correction, actual defocus value in each micrograph should be determined first. In the calculation of the Fourier transform, the circular modulation of the image's spatial frequency component represents phase contrast effects, known as the Thon ring. From the sinuous modulation of the power spectrum, defocus values can be determined. Phase contrast is then corrected basically by a filter to compensate for phase flipping in certain spatial frequency components by the CTF. Some modulations caused by the CTF, e.g., those resulting from mal-adjusted astigmatism and stage drift, prevent high-resolution reconstruction. Inspection of the thon rings is, thus, an effective way to identify and remove of the micrographs that are affected by such modulations. Accurate CTF correction could also contribute to improvement of the following particle picking accuracy (Frank 2006).

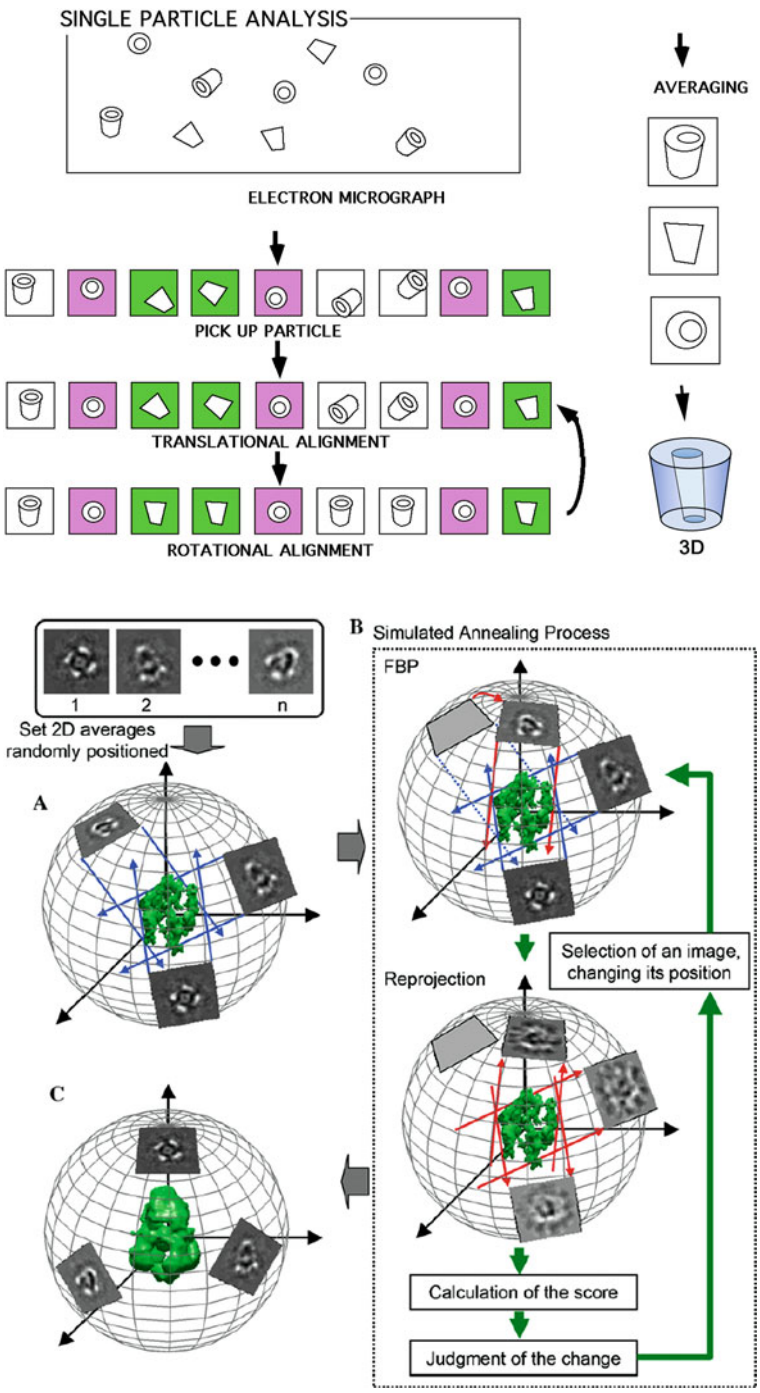


Fig. 13.3 (Upper) Schematic workflow for an example of 3D reconstruction by SPA. Because protein images taken by Cryo-TEM have a low S/N ratio (top), the particles are cropped from the micrographs to create an image library for image processing. A particle in each image is

13.3.4.2 Particle Picking

Currently, hundreds of thousands or even more than a million particles (Fischer et al. 2015; Shen et al. 2017) are used for structure determination at near-atomic or atomic resolution from cryo-EM images. Therefore, a relatively large number of particles have to be chosen from the raw images or CTF-corrected images. If the target protein complex has a high point symmetry, e.g., an icosahedral virus shell, the number of particles required is less. Overall, an automatic particle picking algorithm is generally required except when manual picking is applied for the entire particle picking. This is mainly because the centering of the particle, which can only be performed manually, is still important for most of SPA algorithms.

Image quality of a protein particle affects both manual and automatic particle picking results. Images taken with a relatively small defocus contain the high-frequency information essential for high-resolution analysis, but the contrast is low and faint. Projections are difficult to distinguish from the noisy background. Motion correction of multiple images taken by a DDC camera enables the use of much clearer particles (Li et al. 2013b) although it has a limit due to the low electron dose. When the buffer layer is thick relative to particle size, an energy filter eliminating inelastic electron scattering is effective to reduce the background noise. The traditional "focal pair" technique can also be employed to facilitate particle pickup; pairs of images are recorded, the first is taken close to focus and the second with a larger under focus. The higher contrasted second image is used only to determine the coordinates of particles, and particles are chosen from the first image for the following image processing.

Automated pickup programs are also being developed not only to minimize the processing period but also to realize smaller defocus imaging. Present automatic particle picking algorithms can be classified into reference-free and reference-based algorithms. In case of reference-based algorithms, training data are usually collected by interactive manual particle picking.

Most straight forward reference-based picking algorithms are multireference alignment (MRA)-based particle picking. In this method, manually picked particles can be directly used as references for particle identification using MRA. The strategy is simple, although it can be caught by an accidental resembling of particles with heavy noise; the highest CCF does not always imply particles. It can be mitigated by

Fig. 13.3 (continued) translationally aligned by centering (second row) and then rotationally aligned (third row). This cycle is repeated, and the aligned particles are classified in the library. Similarly orientated molecules in a same group are averaged to create clearer projections. Averaged projections are attached to a sphere at appropriate coordinates, and an initial 3D structure is constructed in the center. The structure is optimized by repeating the whole cycle. (Lower) A new 3D reconstruction method, i.e., echo-correlated 3D reconstruction with simulated annealing (SA) proposed in Ogura and Sato (2006). Euler angles of class average images are initially set randomly. One of the class average images is selected, shifted, and rotated randomly on the sphere. A new 3D volume is reconstructed by FBP, evaluated using correlations between the reprojections and the average images employing SA, and iterated as described (Ogura and Sato 2006)

the multireference multiple alignment (MRMA) method that considers not only the highest correlation peak but also its surrounding peaks (Kawata and Sato 2013). Other supervised-learning algorithms can be also efficiently applied for picking various types of particles. For example, we have developed a neural-network (NN)-based particle pickup program (Ogura and Sato 2004a) and applied it to membrane proteins. Reference-free particle-picking algorithms, namely, unsupervised-learning algorithms, have also been developed using various strategies. For example, the auto-accumulation method was developed using simulated annealing, which can be applied to pick membrane protein particles (Ogura and Sato 2004b). After purification, some proteins in a sample may be partly truncated by proteolysis or denatured. Such particles cannot be used for the reconstruction and ideally should be manually excluded from the particle library in advance. When this is not realistic, they are excluded in the following image processing.

For primary 3D reconstruction, it is important to reduce background noise and improve the signal-to-noise ratio of each projection: The selected particle images are aligned using MRA and sorted into classes using image-classification algorithms to yield average images of homogeneous 2D images.

13.3.4.3 Image Alignment

In general, the orientation of a structure in 3D space is defined in terms of three variables, usually in Euler angles. The in-plane rotation of images should be aligned first during the alignment, resulting in two undetermined variables for the particle direction that should be addressed in the classification in the next section. In the alignment, each image is scanned for position shift and rotation repeatedly to fit one of the references. If possible, references should not be noisy raw images but clearer images including averages because MRA should not be performed between two very noisy images.

If a low-resolution structure of the target protein or a homolog has already been determined, its projections could be directly used as references for MRA. The structure can be used as a 3D reference for the following 3D reconstruction aiming at higher resolution. If the atomic model has been determined using crystals, a low-resolution volume is produced from its low-pass filtering to avoid bias from the model. Using its projections as references, the raw images can be aligned using MRA. If no 3D structural information is available for MRA, the first reference can be the total average of the raw images in the library as recommended in Imagic V (van Heel et al. 1996). Alternatively, candidates for MRA reference could be acquired during an autopickup procedure. For example, after training a three-layer neural network with particle images, information such as Eigen images is automatically stored as weights in the NN. Because these weights reflect particle images that trained the NN, they can be combined to create clear reference images for MRA (Ogura and Sato 2004a). Strategies to avoid subjective selection of specific images as references are called reference-free alignments, and they have been developed in various ways (Ogura and Sato 2004b; Penczek et al. 1992; van Heel et al. 2000).

During MRA, the highest cross-correlation function (XCF) coefficient might not always indicate the best match between two particles due to the high noise. It is proposed that this limit can be improved by the multireference multiple alignment (MRMA) method, which considers not only the peak height but also the information surrounding the candidate peak and subpeaks, and selects the best peak in MRA (Kawata and Sato 2007), as it is applied for reference-based particle picking (Sects. 13.3.4.2).

13.3.4.4 Classification and 2D Averaging for Noise Reduction

Because images are already aligned rotationally and translationally, pixel data in each image can be directly evaluated using various image classification algorithms. Using Eigen value mathematics, the principal component analysis of multivariate statistical analysis (MSA) has been used to classify particles into groups of similar images. For this analysis, the corresponding analysis capable of scale-invariant comparison of the image data set is a major choice, and usually followed by hierarchical descendent clustering, which finds groups of images in a "bottom-up" manner (van Heel 1989).

Because highly accurate MSA of large-scale image data is CPU intensive, other classification algorithms, including neural network classification, have been developed for SPA. For example, SOM-based classification (Pascual et al. 2000) and a modified growing neural gas (GNG) network method (Ogura et al. 2003) have been developed, realizing both accurate and less CPU intensive calculations. To realize more accurate classification in a limited time, further development of a novel classification method is required taking advantage of modern high-performance computing including GPU.

Obtained image groups and their 2D averages will reflect the two direction angle variables (β, γ) of the Euler angle (α, β, γ); they could also reflect other factors including structural variations and molecular motions. Obtained image members in the same group are aligned and averaged to form the class average with less noise. During the calculation, well-determined average images can be chosen as references, though the exclusion should be as minimum as possible so as not to miss minor classes. 2D average images produced through this process are called "characteristic views," a representative set of images observed in different orientations, and are used for the following 3D reconstruction.

13.3.4.5 Reconstruction

The Euler angle of each particle average needs to be determined to create an initial 3D structure. If an atomic model of the target protein or its homolog has already been determined using crystals, a low-resolution structure produced from its low-pass filtering can be employed as a 3D reference, avoiding bias from the model. If a low-resolution structure of the protein has already been determined, the structure can be directly used as a reference to optimize the structure aiming at

higher resolution. For example, using similarities between the reference projections and 2D averages, the Euler angles of 2D averages can be determined, and the initial 3D structure can be reconstructed.

If no structural information is available. The random conical tilt method can also be used to generate initial models (Frank et al. 1978; Radermacher et al. 1986). Intuitively, this method is analogous to stereoscopic observation by two eyes. Tilted (30–60°) and untilted EM images are sequentially taken from the same grid area. If the protein has a preferred orientation, the untilted images are used only to sort classes representing the characteristic views and to determine the azimuth angles, whereas the tilted images are used for reconstruction. This technique is especially suitable for proteins that have a strong preferred orientation. Because the angle of stage tilting cannot cover the area close to 90°, the reconstruction is often incomplete due to missing data around the 90° tilt (the missing cone), which reflects a lack of structure factor in the conical region of the Fourier space (missing cone). This method requires tilting stage equipment for the microscope.

Angular reconstitution methods without stage tilting have also been proposed based on particle images in random orientation, raising challenges of computationally estimating the particle orientations in the presence of noise. For this posterior determination of the Euler angles of projections, the common-line method (Crowther 1971; van Heel 1987) uses the central section theorem, which states that any two projections of a given 3D structure share a unique line, a common central line, in Fourier space. This direct synogram approach was successfully applied to structures with high point symmetry, although the robustness is relatively low for very noisy images of asymmetric or heterogeneous molecules. We have developed a novel reference-free 3D reconstruction system using simulated-annealing algorithms, avoiding any one-dimensional (1D) compression of 2D averages for the Euler angle assignment (Ogura and Sato 2006) (Fig. 13.3 lower). Starting from an initial 3D volume generated by back projecting the randomly oriented 2D averages on a sphere, the structure is optimized by evaluating the correlation coefficient between the reprojections of the volume and the average images. The local minimums during the calculation are overcome using the simulated annealing.

13.3.4.6 Structure Refinement

The initial model generated here becomes the starting model for the second round of MRA (Fig. 13.3 upper). The Euler angle of each projection must be assigned to refine the structure by employing a refinement method. For example, the projection matching finds the adequate Euler angle by directly scanning the XCF between each average image and reprojections of the obtained 3D structure. Such cycles of alignment and reconstruction are repeated until the 3D structure converges. Using maximum likelihood estimation, Carazo et al., established a method to optimize the multiple structural variants starting from multiple initial 3D structures (Scheres et al. 2009). Scheres also established an elegant Bayesian inference-based 3D optimization method that allows structural determination of multiple structures at high resolution (Scheres 2012).

13.3.4.7 Assessment and Interpretation of Reconstruction

The Euler angles of the projections must be examined for their coverage of Fourier space, using a scatter map of Euler angles (β, γ). This must cover a wide area of the sphere. The obtained 3D reconstruction is then examined with regard to volume and resolution. The expected volume is calculated based on its molecular weight and the protein density (1.37 g/ml). In surface rendering of the structure, the isosurface value of the density is chosen so as to give the approximate target volume. In general, slightly smaller volumes will show more precise structural features.

The resolution of an EM reconstruction is mostly assessed in terms of the agreement of two independent reconstructions. The particle library is randomly divided in half, and the two halves are independently reconstructed to compare the two reconstructions. For example, two reconstructions from the even- and odd-numbered projections are compared in 3D Fourier space, and the differences are determined over different shells. In Fourier shell correlation (FSC), a threshold of 0.5 is a conservative criterion used to evaluate SPA (Bottcher et al. 1997). However, a threshold of 0.143, which comes from the corresponding threshold used in X-ray crystallography, is also popular (Rosenthal and Henderson 2003).

As an example of analysis of a small asymmetric membrane protein, a voltage-sensitive sodium channel was reconstructed. Voltage-sensitive membrane channels, the sodium channels, the potassium channels, and the calcium channels work together to amplify, transmit, and generate electric pulses in higher forms of life. Sodium channels of almost 200 kDa play a principal role in neural excitation. The amino acid sequence of a sodium channel has four repeats, each of which possesses 6 transmembrane segments (Fig. 13.4Ba). Muscle-type sodium channel molecules purified from Electrophorus electricus electroplax in the Amazon River were imaged by cryo-electron microscopy and reconstructed (Sato et al. 2001).

Longitudinal views of the molecule (Fig. 13.4A) show a quasi-fourfold symmetric molecule that has a 30 Å wide horizontal band without peripheral holes, which could represent the transmembrane domains (white lines). Thus, the volume ratio of the upper and lower regions suggests that they are the extracellular and cytoplasmic domains of the channel molecule (Fig. 13.4Ba), which was confirmed by an antibody binding (Sato et al. 2001). Sections through the 3D density map reveal the specific internal structure of the voltage-sensitive sodium channel (Fig. 13.4Bb–e). For smaller membrane proteins, the TRIC cation channel, as small as 99 kDa, was reconstructed by SPA using negatively stained particle images (Yazawa et al. 2007).

Docking of X-Ray Structure

When a higher resolution X-ray substructure or a partial structure is available, it should be fitted to the EM volume. The fitted structures could be utilized to interpret the molecular mechanism. For example, the whole EM map of the oxidative stress sensor Keap1 was docked with the DC-domain structure alone (Fig. 13.5). Keap1 is

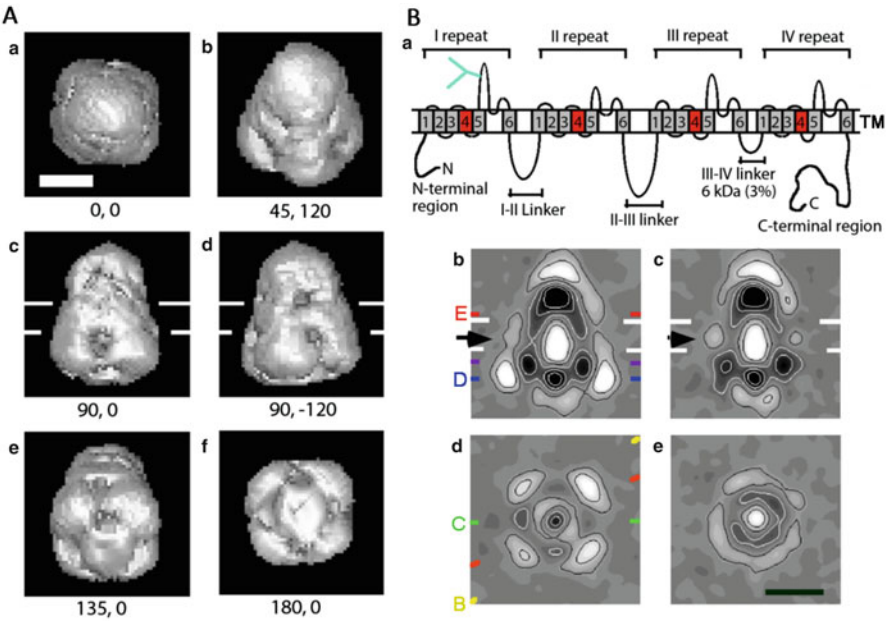


Fig. 13.4 Single-particle reconstruction of voltage-sensitive sodium channel. (**A**): Surface representations display the outer shapes of the protein at viewing angles defined by the Euler angles (noted below each column). **a**; Top view. **b**; View at oblique angle. **c–d**; Two side views rotated by 120° interval around the vertical axis. **e**; View at oblique angle. **f**; Bottom view. White bars delineate the lipid bilayer region. (**B**): **a**; Repeats, loops, and terminal extensions predicted from the amino acid sequence of the sodium channel. The S4 segments are colored in red. **b**; Axial section of the 3D volume along the diagonal indicated by the yellow line (B) in d. **c**; Axial section along the green line (C) in d. **d, e**; Perpendicular sections at the positions marked in b by lines colored blue (D), and red (E). High densities are shown in bright shades. White bars delineate the lipid bilayer. Black arrows indicate the constriction between the cytoplasmic (lower) and extracellular (upper) cavities. Scale bar indicates 50 Å. (Modified from Sato et al. 2001 with permission)

a dimer molecule with a molecular weight of almost 140 kDa. Under unstressed conditions, Keap1 binds transcription factor Nrf2 and results in rapid degradation of Nrf2 through the proteasome pathway. Under stressed conditions, Keap1 releases transcription factor Nrf2, and free Nrf2 induces transcription of Phase 2 detoxifying enzymes in the nucleus. Mouse Keap1 whole molecules were imaged by negatively stained TEM and reconstructed by SPA showing a twofold quasi-symmetric cherry-like structure (Ogura et al. 2010). The structure of the DC domain alone determined using X-ray crystallography (PDB 1X2J) (Padmanabhan et al. 2006) was fitted to the 3D volume: the DC domain was fitted to the globular volumes. This result suggests that they are DC domains (Ogura et al. 2010) that can bind NRF2 (Fig. 13.5). The additional volume surrounding each DC domain is interpreted to include the IVR domain possessing stress-sensing cysteines. The close location between IVR and DC suggests that the oxidative stress binding to cysteines could cause conformational change of the neighboring DC domain, resulting in the release of NRF2. Free NRF2

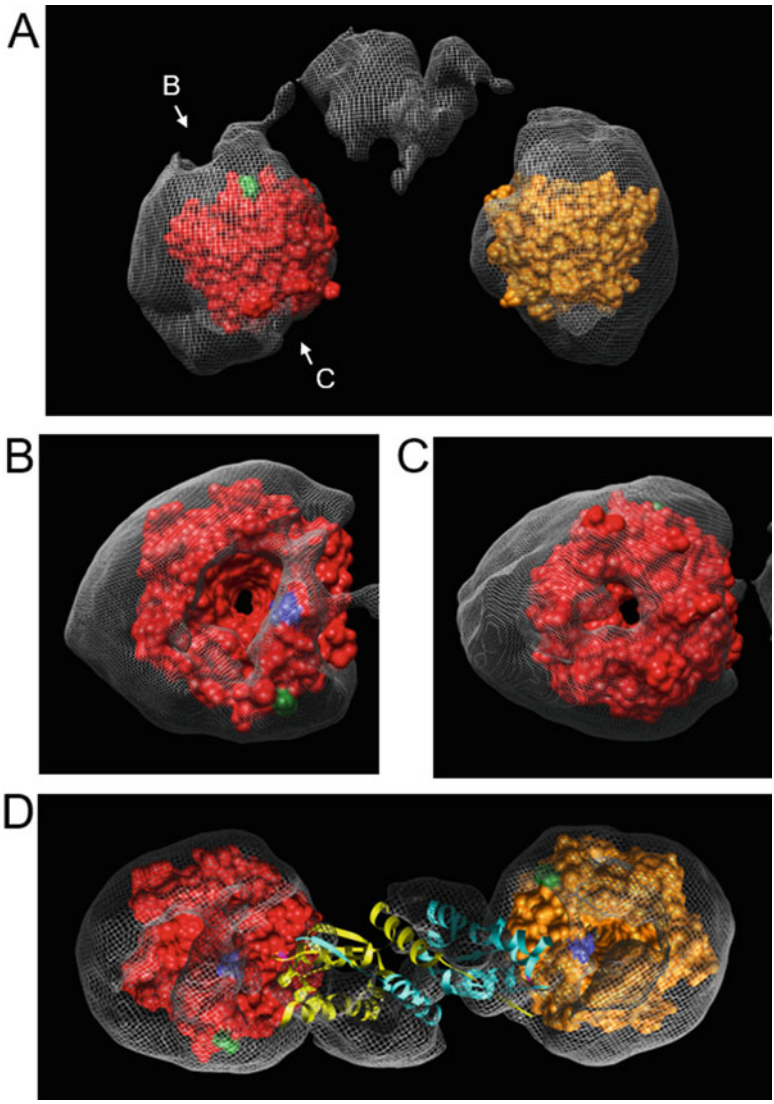


Fig. 13.5 Molecular fitting of the DC crystal structure to the Keap1 single particle reconstruction. (**A**) Atomic model of the Keap1-DC domain solved by X-ray analysis (Orange or Red) (Padmanabhan et al. 2006), fitted into the EM density map (Ogura et al. 2010). (**B**) View from direction "B" as indicated in A. (**C**) View from direction "C" as indicated in A. (**D**) Top view. The X-ray model of a BTB homolog, the LRF BTB-zinc finger protein (PDB accession code 2NN2 (Stogios et al. 2007)), nearly fitted with the stem-like region. The small exit of each tunnel closely coincides with that in the X-ray model. Extra density surrounding the DC domain is composed of the IVR domain and part of the BTB domain. The distance between the two Nrf2 binding sites at the bottom of the DC domains is approximately 80 Å. The N and C termini of the DC domain are shown by blue and green, respectively. (Modified from Ogura et al. 2010 with permission)

can induce transcription of phase 2 detoxifying enzymes to protect the body from electrophiles, including carcinogens and peroxides, such as H_2O_2.

In addition to manual fitting, several automatic fitting programs are available for rigid body fitting and flexible fitting (Trabuco et al. 2008). Wriggers' group successfully developed NN-based fitting algorithms and enables highly efficient molecular fitting by Situs (Wriggers 2010).

13.4 Other Electron Microscopies Under Hydrophilic Conditions

13.4.1 Tomography

For 3D reconstruction of larger biological structures including cells and tissues, electron tomography (ET) or serial thin sectioning is required. Cryo-ET is a modern structure determination method used under hydrophilic conditions, which can be applied to macromolecular complexes in cells. In cryo-ET, a series of EM images is taken at low dose by tilting the sample stage in the microscope column, and the 3D structure is reconstructed from the images (Fig. 13. 6a) (Baumeister 2002). Therefore, the thickness of the frozen sample should be thin enough to transmit electrons (e.g., <500 nm). For ideal reconstruction, images are taken from all directions. However, it is not usually possible using cryo-EM, and the missing angle is called the "missing wedge" if single axis tomography is employed. This method diverged to scanning transmission electron microscope (STEM) tomography. This method can change the scanning focus depth depending on the tilt of the sample plane (Brandt et al. 2010) and has potential for observation of thicker samples.

Tomography has been combined with classification and averaging technique of SPA. Murata et al. reported that interactions between marine cyanophage P-SSP7 and cyanobacterium Prochlorococcus MED4 were reconstructed using cryo-ET (Fig. 13.6) (Murata et al. 2017). The 3D structures of infected bacteria were visualized with attached phages. The subtomogram of the attached phages was clipped from the whole tomogram, and these 3D structures were classified and averaged using SPA techniques. It revealed that the three main conformations are characterized by different angles between the phage tail and the cell surface, which suggest the mechanism of infections (Murata et al. 2017). These results suggest the huge possibility of the application of total cryo-ET/SPA analysis for the comprehensive understanding of cells.

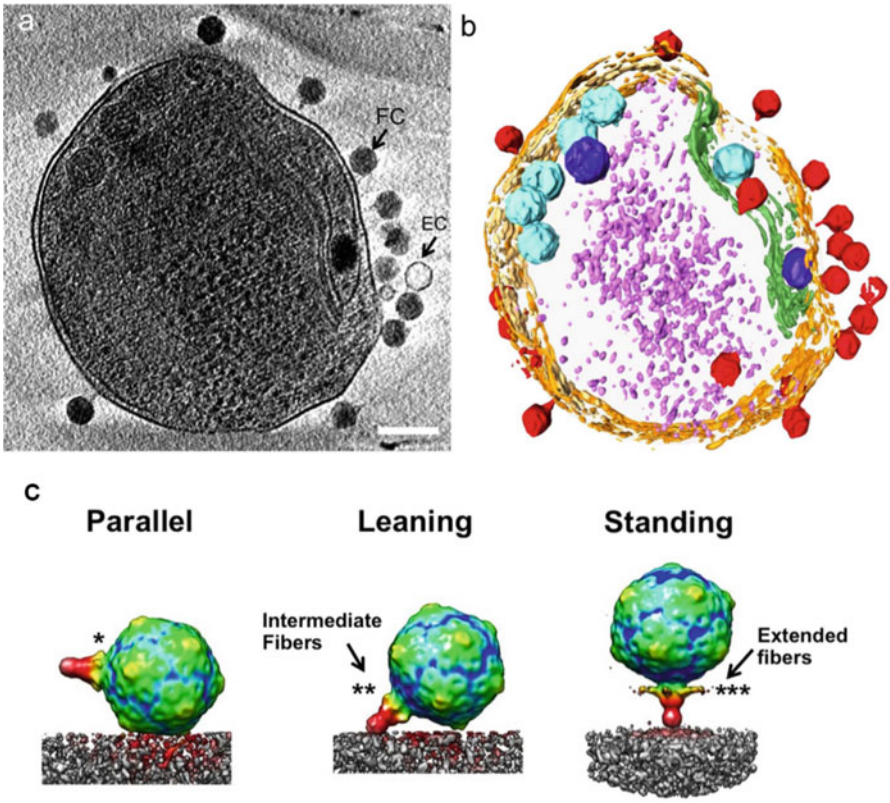


Fig. 13.6 ET and subtomogram averaging. Interactions between cyanobacterium Prochlorococcus MED4 and marine cyanophage P-SSP7 visualized by ET. (**a**) Slice through a reconstructed tomogram, and (**b**) corresponding annotation highlighting the organelle. (**c**) P-SSP7 isosurface models derived from the averages composed of particles classified by tail fiber conformation (***extended, **intermediate, and *folded). Structural change of the tail fibers is suggested to be correlated with the stages of adsorption. (Modified from (Murata et al. 2017) with permission)

13.4.2 In-Liquid Observation of Protein Localization and Crystals by ASEM

In standard EM, the sample must be observed in a vacuum, which means that it must be frozen or dried. Atmospheric scanning electron microscopy (ASEM) was developed to realize direct observation of cells or protein complexes in an aqueous environment. The ASEM is an inverted scanning electron microscope (SEM) (Fig. 13.7a). Inside of the column is under a vacuum being sealed at the top by an electron transparent SiN-film window in the base of the open ASEM specimen dish (Fig. 13.7b). The sample in this dish is under atmospheric pressure and can be in liquid. The electron beam at an acceleration voltage of 5-30 kV of the inverted SEM is projected up the column through the SiN film onto the sample, and the

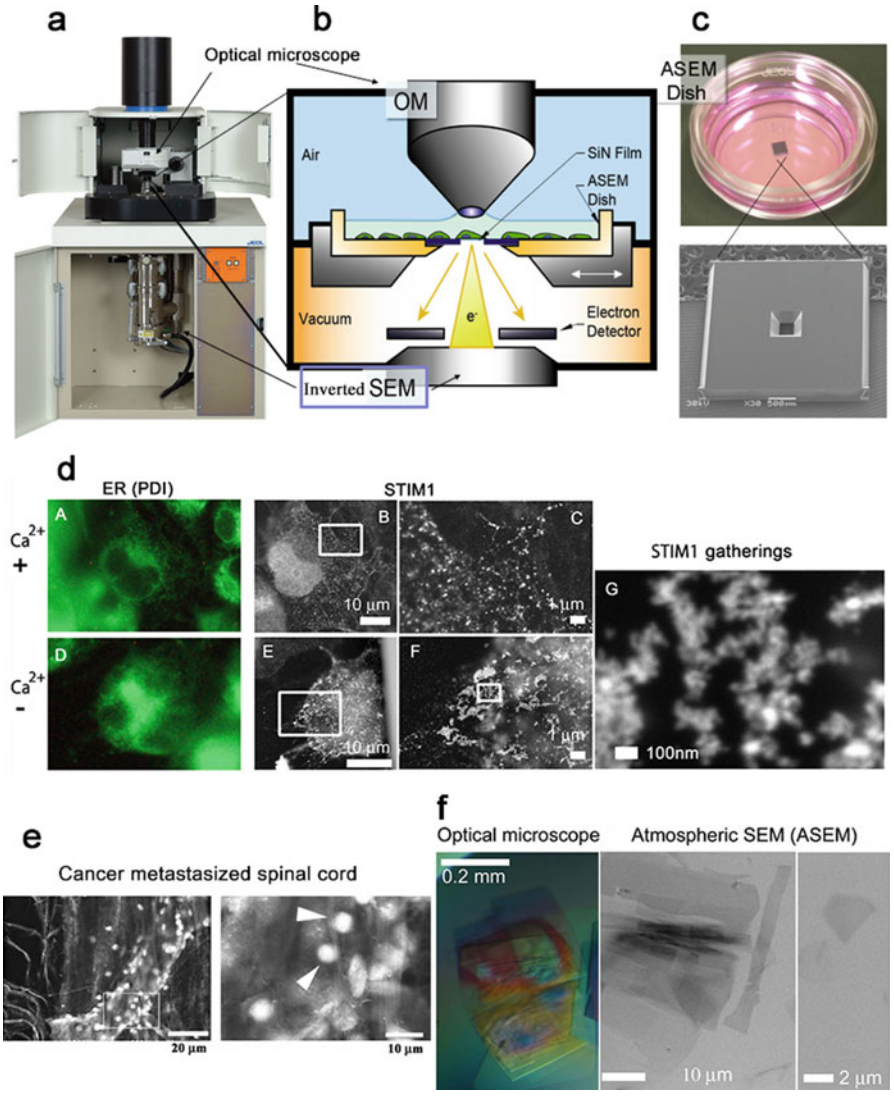


Fig. 13.7 ASEM observation of cells, tissues and protein crystals. (**a**–**c**) The Atmospheric scanning electron microscopy (ASEM) as used for CLEM in solution (Nishiyama et al. 2010). (**d**) Dynamic rearrangement of STIM1 in response to Ca^{2+} store depletion. STIM1-expressed COS7 cells with (lower) and without thapsigargin treatment (upper) were aldehyde-fixed, immuno-labeled for STIM1, and observed in radical scavenger solution (Maruyama et al. 2012a). (**e**) Spinal cord metastasized by breast cancer cells. Spinal cord tissues were aldehyde-fixed, and stained with both Pt-blue and PTA. Breast cancer cells are indicated by arrows (Memtily et al. 2015). (**f**) Direct observation of protein 3D microcrystals in crystallization buffer without staining (Maruyama et al. 2012b). (Modified from Maruyama et al. 2012a, b, Memtily et al. 2015 with permission)

backscattered electrons are collected (Fig. 13.7b) (Nishiyama et al. 2010). The observable sample depth is 1–3 μm, and the resolution obtained when imaging a sample in solution is 8 nm near the SiN film. The inverted SEM and a light microscope (LM) positioned above the sample (Fig. 13.7a, b) are aligned, allowing correlative light-electron microscopy (CLEM).

ASEM realizes high-throughput immuno-EM because it does not require time-consuming hydrophobic treatment and resin embedding otherwise necessary for samples to endure the vacuum of a usual immuno-EM. Various kinds of primary cells including neurons, megakaryocyte, and ES cells have been cultured on the ASEM dish and labeled in situ (Hirano et al. 2014). In Fig. 13.7d, CLEM using ASEM allowed to visualized molecular super complex formation by the Ca^{2+} sensor STIM1 of the endoplasmic reticulum in COS7 cells in response to Ca^{2+} store depletion (Maruyama et al. 2012a). Moreover, ASEM can be used to observe tissue immersed in aqueous solution in the ASEM dish and could be used in intra-operated cancer diagnosis (Fig. 13.7e) (Memtily et al. 2015; Nishiyama et al. 2010).

ASEM can also be used to directly observe protein micro-3D crystals in a crystallization buffer with or without staining (Fig. 13.7f) (Maruyama et al. 2012b). To realize in situ check of micro-crystals in a crystallization chamber, an ASEM dish with a standard crystallization chamber has been developed to allow the observation of crystallization in crystallization buffer (Maruyama et al. 2012b).

Acknowledgements This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas, Sparse modeling (to C.S.), by CREST (to C.S.), by a Grant-in-Aid for Scientific Research from JSPS (15 K14499) (to C.S.), by grants from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) (to C.S.), by a Grant-in-Aid from CANON (to C.S.), by a Strategic grant to NISP from AIST (to C.S.) and by a Grant-in-Aid from AIST (to C.S).

References

Abbe E (1873) Über einen neuen Beleuchtungsapparat am Mikroskop [about a new illumination apparatus to the microscope]. Arch Mikrosk Anat 9:469–480

Adrian M, Dubochet J, Lepault J, McDowall AW (1984) Cryo-electron microscopy of viruses. Nature 308:32–36

Allied (2015 https://www.alliedvision.com/jp/news/news-archive/detail/news/interview-sony-announces-end-of-ccd-sensor-production.html. Allied vision news

Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S (2015) 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science 348:1147–1151

Baumeister W (2002) Electron tomography: towards visualizing the molecular organization of the cytoplasm. Curr Opin Struct Biol 12:679–684

Beck F, Unverdorben P, Bohn S, Schweitzer A, Pfeifer G, Sakata E, Nickell S, Plitzko JM, Villa E, Baumeister W, Forster F (2012) Near-atomic resolution structural model of the yeast 26S proteasome. Proc Natl Acad Sci U S A 109:14870–14875

Borgnia M J, Banerjee S, Merk A, Matthies D, Bartesaghi A, Rao P, Pierson J, Earl L A, Falconieri V, Subramaniam S, Milne J L S (2016) Using Cryo-EM to Map Small Ligands on

Dynamic Metabolic Enzymes: Studies with Glutamate Dehydrogenase. Molecular pharmacology 89: 645–51
Bottcher B, Wynne SA, Crowther RA (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386:88–91
Brandt F, Carlson LA, Hartl FU, Baumeister W, Grunewald K (2010) The three-dimensional organization of polyribosomes in intact human cells. Mol Cell 39:560–569
Bremer A, Henn C, Engel A, Baumeister W, Aebi U (1992) Has negative staining still a place in biomacromolecular electron microscopy? Ultramicroscopy 461:85–111
Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N (2012) Beam-induced motion of vitrified specimen on holey carbon film. J Struct Biol 177:630–637
Cao E, Liao M, Cheng Y, Julius D (2013) TRPV1 structures in distinct conformations reveal activation mechanisms. Nature 504:113–118
Crowther RA (1971) Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. Philos Trans R Soc Lond Ser B Biol Sci 261:221–230
De Rosier DJ, Klug A (1968) Reconstruction of three dimensional structures from electron micrographs. Nature 217:130–134
Egerton R F, Li P, and Malac M (2004) Radiation Damage in the Tem and Sem. Micron 35 (6):399–409
Egerton RF (2011) Electron energy-loss spectroscopy in the electron microscope, 3rd edn. Springer, New York
Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H (2015) Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. Nature 520:567–570
Frank J (1992) Electron tomography: three-dimensional imaging with the transmission electron microscope. Plenum Press, New York
Frank J (2006) Three-dimensional Electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state. Oxford University Press, New York
Frank J, Agrawal RK (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature 406:318–322
Frank J, Goldfarb W, Eisenberg D, Baker TS (1978) Reconstruction of glutamine synthetase using computer averaging. Ultramicroscopy 3:283–290
Frank J, Zhu J, Penczek P, Li Y, Srivastava S, Verschoor A, Radermacher M, Grassucci R, Lata RK, Agrawal RK (1995) A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. Nature 376:441–444
Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116:190–199
Fujiyoshi Y, Mizusaki T, Morikawa K, Yamagishi H, Aoki Y, Kihara H, Harada Y (1991) Development of a superfluid-helium stage for high-resolution electron-microscopy. Ultramicroscopy 38:241–251
Grigorieff N (2007) FREALIGN: high-resolution refinement of single particle structures. J Struct Biol 157:117–125
Grigorieff N (2013) Direct detection pays off for electron cryo-microscopy. elife 2:e00573
Haider M, Uhlemann S, Schwan E, Rose H, Kabius B, Urban K (1998) Electron microscopy image enhanced. Nature 392:768–769
Henderson R (2004) Realizing the potential of electron cryo-microscopy. Q Rev Biophys 37:3–13
Hirano K, Kinoshita T, Uemura T, Motohashi H, Watanabe Y, Ebihara T, Nishiyama H, Sato M, Suga M, Maruyama Y, Tsuji NM, Yamamoto M, Nishihara S, Sato C (2014) Electron microscopy of primary cell cultures in solution and correlative optical microscopy using ASEM. Ultramicroscopy 143:52–66

Hobbs LW (1984) Radiation effects in analysis by tem. In: Quantitative Electron microscopy. SUSSP Publications, Edinburgh, pp 399–445

JEM-Z200FSC, https://www.jeol.co.jp/en/products/detail/JEM-Z200FSC.html

Jin L, Milazzo AC, Kleinfelder S, Li S, Leblanc P, Duttweiler F, Bouwer JC, Peltier ST, Ellisman MH, Xuong NH (2008) Applications of direct detection device in transmission electron microscopy. J Struct Biol 161:352–358

Kawata M, Sato C (2007) A statistically harmonized alignment-classification in image space enables accurate and robust alignment of noisy images in single particle analysis. J Electron Microsc 56:83–92

Kawata M, Sato C (2013) Multi-reference-based multiple alignment statistics enables accurate protein-particle pickup from noisy images. Microscopy 62:303–315

Kawate T, Gouaux E (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14:673–681

Kisielowski C, Freitag B, Bischoff M, van Lin H, Lazar S, Knippels G, Tiemeijer P, van der Stam M, von Harrach S, Stekelenburg M, Haider M, Uhlemann S, Muller H, Hartel P, Kabius B, Miller D, Petrov I, Olson EA, Donchev T, Kenik EA, Lupini AR, Bentley J, Pennycook SJ, Anderson IM, Minor AM, Schmid AK, Duden T, Radmilovic V, Ramasse QM, Watanabe M, Erni R, Stach EA, Denes P, Dahmen U (2008) Detection of single atoms and buried defects in three dimensions by aberration-corrected electron microscope with 0.5-a information limit. Microsc Microanal 14:469–477

Lawson, C. L., Patwardhan, A., Baker, M. L., Hryc, C., Garcia, E. S., Hudson, B. P., Lagerstedt, I., Ludtke, S. J., Pintilie, G., Sala, R., Westbrook, J. D., Berman, H. M., Kleywegt, G. J. & Chiu, W. (2016). Nucleic Acids Res. 44:D396–D403

Li X, Zheng SQ, Egami K, Agard DA, Cheng Y (2013a) Influence of electron dose rate on electron counting images recorded with the K2 camera. J Struct Biol 184:251–260

Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y (2013b) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods 10:584–590

Liao M, Cao E, Julius D, Cheng Y (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature 504:107–112

Linck M, Hartel P, Uhlemann S, Kahl F, Muller H, Zach J, Haider M, Niestadt M, Bischoff M, Biskupek J, Lee Z, Lehnert T, Borrnert F, Rose H, Kaiser U (2016) Chromatic aberration correction for atomic resolution TEM imaging from 20 to 80 kV. Phys Rev Lett 117:076101

Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol 128:82–97

Marabini R, Masegosa IM, San Martin MC, Marco S, Fernandez JJ, de la Fraga LG, Vaquerizo C, Carazo JM (1996) Xmipp: an image processing package for electron microscopy. J Struct Biol 116:237–240

Maruyama Y, Ebihara T, Nishiyama H, Suga M, Sato C (2012a) Immuno EM-OM correlative microscopy in solution by atmospheric scanning electron microscopy (ASEM). J Struct Biol 180:259–270

Maruyama Y, Ebihara T, Nishiyama H, Konyuba Y, Senda M, Numaga-Tomita T, Senda T, Suga M, Sato C (2012b) Direct observation of protein microcrystals in crystallization buffer by atmospheric scanning electron microscopy. Int J Mol Sci 13:10553–10567

Memtily N, Okada T, Ebihara T, Sato M, Kurabayashi A, Furihata M, Suga M, Nishiyama H, Mio K, Sato C (2015) Observation of tissues in open aqueous solution by atmospheric scanning electron microscopy: applicability to intraoperative cancer diagnosis. Int J Oncol 46:1872–1882

Milazzo AC, Moldovan G, Lanman J, Jin L, Bouwer JC, Klienfelder S, Peltier ST, Ellisman MH, Kirkland AI, Xuong NH (2010) Characterization of a direct detection device imaging camera for transmission electron microscopy. Ultramicroscopy 110:744–747

Morikawa M, Yajima H, Nitta R, Inoue S, Ogura T, Sato C, Hirokawa N (2015) X-ray and Cryo-EM structures reveal mutual conformational changes of kinesin and GTP-state microtubules upon binding. EMBO J 34:1270–1286

Murata K, Zhang Q, Gerardo Galaz-Montoya J, Fu C, Coleman ML, Osburne MS, Schmid MF, Sullivan MB, Chisholm SW, Chiu W (2017) Visualizing adsorption of Cyanophage P-SSP7 onto marine Prochlorococcus. Sci Rep 7:44176

Nishiyama H, Suga M, Ogura T, Maruyama Y, Koizumi M, Mio K, Kitamura S, Sato C (2010) Atmospheric scanning electron microscope observes cells and tissues in open medium through silicon nitride film. J Struct Biol 169:438–449

Ogura T, Sato C (2004a) Automatic particle pickup method using a neural network has high accuracy by applying an initial weight derived from eigenimages: a new reference free method for single-particle analysis. J Struct Biol 145:63–75

Ogura T, Sato C (2004b) Auto-accumulation method using simulated annealing enables fully automatic particle pickup completely free from a matching template or learning data. J Struct Biol 146:344–358

Ogura T, Sato C (2006) A fully automatic 3D reconstruction method using simulated annealing enables accurate posterioric angular assignment of protein projections. J Struct Biol 156:371–386

Ogura T, Iwasaki K, Sato C (2003) Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking. J Struct Biol 143:185–200

Ogura T, Tong KI, Mio K, Maruyama Y, Kurokawa H, Sato C, Yamamoto M (2010) Keap1 is a forked-stem dimer structure with two large spheres enclosing the intervening, double glycine repeat, and C-terminal domains. Proc Natl Acad Sci U S A 107:2842–2847

Ogura T, Yajima H, Nitta R, Hirokawa N, Sato C (2014) New simulated annealing approach considering helix bending applied to determine the 8.8A structure of 15-protofilament microtubules. J Struct Biol 188:165–176

Padmanabhan B, Tong KI, Ohta T, Nakamura Y, Scharlock M, Ohtsuji M, Kang MI, Kobayashi A, Yokoyama S, Yamamoto M (2006) Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. Mol Cell 21:689–700

Pascual A, Barcena M, Merelo JJ, Carazo JM (2000) Mapping and fuzzy classification of macromolecular images using self-organizing neural networks. Ultramicroscopy 84:85–99

Penczek P, Radermacher M, Frank J (1992) Three-dimensional reconstruction of single particles embedded in ice. Ultramicroscopy 40:33–53

Radermacher M, Wagenknecht T, Verschoor A, Frank J (1986) A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of E. coli. J Microsc 141:RP1–RP2

Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J Mol Biol 333:721–745

Sasaki T, Sawada H, Hosokawa F, Kohno Y, Tomita T, Kaneyama T, Kondo Y, Kimoto K, Sato Y, Suenaga K (2010) Performance of low-voltage STEM/TEM with delta corrector and cold field emission gun. J Electron Microsc 59(Suppl 1):S7–13

Sato C, Ueno Y, Asai K, Takahashi K, Sato M, Engel A, Fujiyoshi Y (2001) The voltage-sensitive sodium channel is a bell-shaped molecule with several cavities. Nature 409:1047–1051

Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519–530

Scheres SH, Nunez-Ramirez R, Sorzano CO, Carazo JM, Marabini R (2008) Image processing for electron microscopy single-particle analysis using XMIPP. Nat Protoc 3:977–990

Scheres SH, Valle M, Grob P, Nogales E, Carazo JM (2009) Maximum likelihood refinement of electron microscopy data with normalization errors. J Struct Biol 166:234–240

Schroder RR (2015) Advances in electron microscopy: a qualitative view of instrumentation development for macromolecular imaging and tomography. Arch Biochem Biophys 581:25–38

Shaikh TR, Gao H, Baxter WT, Asturias FJ, Boisset N, Leith A, Frank J (2008) SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat Protoc 3:1941–1974

Shen H, Zhou Q, Pan X, Li Z, Wu J, Yan N (2017) Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. Science 355(6328):eaal4326

Sorzano CO, Marabini R, Velazquez-Muriel J, Bilbao-Castro JR, Scheres SH, Carazo JM, Pascual-Montano A (2004) XMIPP: a new generation of an open-source image processing package for electron microscopy. J Struct Biol 148:194–204

Stark H, Mueller F, Orlova EV, Schatz M, Dube P, Erdemir T, Zemlin F, Brimacombe R, van Heel M (1995) The 70S Escherichia coli ribosome at 23 a resolution: fitting the ribosomal RNA. Structure 3:815–821

Stogios PJ, Chen L, Prive GG (2007) Crystal structure of the BTB domain from the LRF/ZBTB7 transcriptional regulator. Protein Sci 16:336–342

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol 157:38–46

Taylor KA, Glaeser RM (1976) Electron microscopy of frozen hydrated biological specimens. J Ultrastruct Res 55:448–456

Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure 16:673–683

van Heel M (1987) Angular reconstitution - a posteriori assignment of projection directions for 3-D reconstruction. Ultramicroscopy 21:111–123

van Heel M (1989) Classification of very large electron microscopical image data sets. Optik 82:114–126

van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. J Struct Biol 116:17–24

van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A (2000) Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys 33:307–369

van Heel M, Portugal R, Rohou A, Linnemayr C, Bebeacua C, Schmidt R, Grant T, Schatz M (2011) Four-dimensional cryo electron microscopy at quasi atomic resolution: 'IMAGIC 4D'. In: Arnold E, Himmel DM, Rossmann MG (eds) Crystallography of biological macromolecules. Wiley, New York, pp 624–628

Wriggers W (2010) Using situs for the integration of multi-resolution structures. Biophys Rev 2:21–27

Yajima H, Ogura T, Nitta R, Okada Y, Sato C, Hirokawa N (2012) Conformational changes in tubulin in GMPCPP and GDP-taxol microtubules observed by cryoelectron microscopy. J Cell Biol 198:315–322

Yasunaga T, Wakabayashi T (1996) Extensible and object-oriented system Eos supplies a new environment for image analysis of electron micrographs of macromolecules. J Struct Biol 116:155–160

Yazawa M, Ferrante C, Feng J, Mio K, Ogura T, Zhang M, Lin PH, Pan Z, Komazaki S, Kato K, Nishi M, Zhao X, Weisleder N, Sato C, Ma J, Takeshima H (2007) TRIC channels are essential for Ca2+ handling in intracellular stores. Nature 448:78–82

Zhou ZH (2011) Atomic resolution cryo electron microscopy of macromolecular complexes. Adv Protein Chem Struct Biol 82:1–35

Chapter 14
Biological Small-Angle X-Ray Scattering (SAXS)

Tomáš Klumpler

14.1 Introduction

Biological small angle X-ray scattering (SAXS) allows rapid determination of size and low-resolution structure of biological macromolecules in solution. In a typical SAXS experiment, the sample is placed in quartz capillary, exposed with X-ray beam, and the scattered intensity is recorded at the detector as a function of scattering angle (Fig. 14.1). Radially averaged scattering images called SAXS curves are used to determine average particle size, shape reconstructions, unraveling the quaternary architecture of complexes, modeling of molecular flexibility, and more. SAXS allows characterizing macromolecules in solution, i.e., close to their native and biologically relevant conditions. Development of the software tools available and advances of synchrotron and "home" X-ray sources brought SAXS to routine work flow of number of structural biologists. It is a low-resolution technique, but in combination with other techniques such as X-ray crystallography and nuclear magnetic resonance, the SAXS becomes powerful tool of the structural analysis of biological macromolecules.

As all other parts of the Method Guidance section of this book, even this one serves as a brief introduction to the method, scattering of neutrons and number of modifications of SAXS method will be not discussed. To develop a deeper understanding of the biological SAXS, our reader should continue with excellent textbooks or reviews available. Special attention is addressed to clear and comprehensive textbook written by "founding father" of the modern biological SAXS (Svergun 2013). Importantly, own personal experience with SAXS data acquisition and analysis is most valuable.

T. Klumpler (✉)
Central European Institute of Technology, Masaryk University, Brno, Czech Republic
e-mail: klumpler@sci.muni.cz

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_14

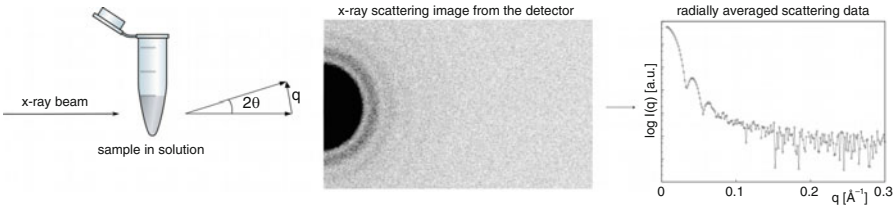


Fig. 14.1 Schematic representation of SAXS experiment. Biological sample in solution is placed into collimated X-ray beam, isotropic scattering pattern is recorded at the two-dimensional detector and radially averaged to one-dimensional curve called SAXS profile, curve or data. For biological SAXS, the scattered intensity is plotted as a function of scattering vector $q = (4\pi \sin \theta)/\lambda$ [$Å^{-1}$], where the intensity is plotted on the logarithmic scale using artificial units [a.u.] or absolute units [cm^{-1}]

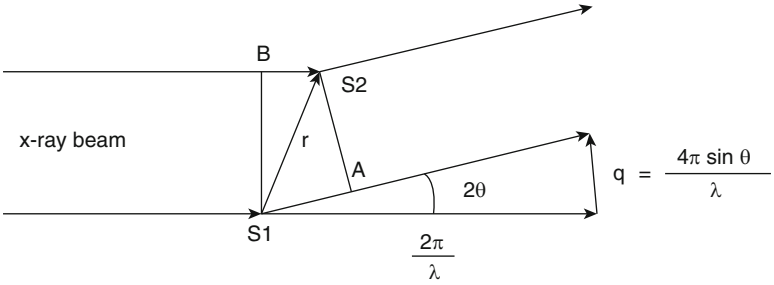


Fig. 14.2 Elastic scattering from two scattering centers. X-ray beam scattered from two scattering centers *S*1 and *S*2, toward the scattering angle 2θ. The magnitude of the incoming wave vector is $k = 2\pi/\lambda$, and the scattering vector magnitude is defined as $q = (4\pi \sin \theta)/\lambda$. The nature of the interference of the scattered waves is dependent on the phase shift of the scattered waves caused by path length difference of the X-ray photons. The path length difference $S1A - S2B = 2r \sin(\theta)$ and the phase shift defined as $\phi = 2\pi\ 2r \sin(\theta)/\lambda$ or $\phi = qr$

14.1.1 Basic Principles

Scattering of X-ray photons occurs as a consequence of their interaction with electrons in the sample. When the studied object is irradiated by a monochromatic plane wave, all atoms primarily through electrons scatter photons to all directions with the same probability. Thus, all atoms could be considered as sources of spherical waves with the same wavelength λ as the incident wave (elastic scattering). The magnitude of the wave vector of the incoming beam is described as $k = |\,k\,| = 2\pi/\lambda$. In case of elastic scattering, the magnitude of the wave vector k of the incident wave equals the magnitude of the scattered waves $k' = |k'| = k$, and the scattering vector q is then defined as momentum transfer $q = k' - k$. From Fig. 14.2, the relationship could be derived: $|q| = q = (4\pi \sin \theta)/\lambda$, and the dimension of the scattering vector q is of reciprocal length, usually [$Å^{-1}$] or [nm^{-1}]. In structural biology field, the scattering vector is denoted as "q" or "s" and with the same meaning, rarely is defined as $(2\pi \sin \theta)/\lambda$. Similarly as in case of X-ray

crystallography, the resulting scattered amplitude in certain scattering angle is dependent nature of interference of waves from multiple scattering centers (atoms), i.e., on the phase difference ϕ of scattered spherical waves from these scattering centers. In Fig. 14.2, there is an example of scattering from two scattering centers $S1$ and $S2$ with fixed position in space and graphically described the phase difference of two secondary scattered spherical waves. It is obvious that the phase difference ϕ is dependent on the path difference of the traveling radiation, with dependence $\phi = qr.$, where r is the distance between $S1$ and $S2$. For scattering angles 2θ, where both scattered waves are perfectly in phase, the resulting wave amplitude will be the addition of the two interfering scattered waves. Such scattering centers obey Bragg's law, see Chap. 11. In crystalline samples, this adding effect is multiplied by regularly ordered scattering centers, and at those directions, diffraction spots in X-ray crystallography are then recorded. In case of noncrystalline sample, such as macromolecules in ideal solution, the interference pattern recorded at the detector is isotropic, and the scattered intensity decays from the center of the beam (place, where the primary X-ray beam hits the detector) toward the higher angles (Fig. 14.1). The second case is typical for biological SAXS studies.

Scattered intensity (number of photons) is recorded at the detector as a function of scattering vector $I(q)$. In general, the scattering intensity is the sum of scattering from all irradiated particles weighted by abundance and their volume and electron density contrast square. To extract structural information from scattering pattern, the solution sample of macromolecules should be monodisperse and dilute. In ideal solution, when particles of the same size and shape are not interacting with each other, the angular dependence of scattered intensity could be expressed as:

$$I(q) = N(\Delta\rho V)^2 P(q)$$

where N is the number of particles, $\Delta\rho$ is the electron density contrast of the particle and bulk solvent, V is the volume of the particle, and $P(q)$ is the spherically averaged form factor function (Svergun 2013). Form factor dependence on the shape of the particle of uniform electron density could be expressed as integral through particle volume V:

$$P_{\mathrm{j}}(q) = \left\langle \left| \Delta\rho \int_V \mathrm{e}^{iqr} dV \right|^2 \right\rangle$$

where $\Delta\rho$ is the electron density contrast, q is the scattering vector, and r is the distance between scattering centers and the brackets symbolizing the spherical average. Thus, form factor is related to the electron density of the scattering particle by Fourier transformation. The direct extraction of the particle shape from scattering data is principally not possible, as scattering data are not collected to infinite angle and are recorded with experimental error. The indirect Fourier transform of the scattering $I(q)$ has been proposed (Svergun 1992), resulting pair distance distribution function $P(r)$ and represents all scattering pair distances in the particle weighted by the distance square, see Sect. 14.3.5.

14.2 Biological SAXS Experiment

Basic scheme of SAXS experiment setup is shown in Fig. 14.1. Solution sample in glass/quartz capillary is placed into collimated X-ray beam, and the scattered intensity is recorded at the detector. For structural biology studies, the X-ray wavelengths λ in range 0.9–2.0 Å are typically employed. The X-ray sources used for SAXS are in principle the same as X-ray sources used for X-ray crystallography, see Chap. 11. The macromolecules of interest not only scatter and absorb the X-rays part of the beam path in the experimental hutch is evacuated. The outer diameter of capillary is usually approximately 1.0 mm, which is an optimal ratio between scattering and absorption of aqueous solutions. The sample stage with the capillary is usually temperature controlled. The data acquisition times at third generation synchrotron sources are usually in range of seconds or minutes, whereas at the modern rotating anode home sources are in range of tens of minutes or hours, depending on the beam intensity. As the data acquisition is relatively fast, automatic sample changers are exploited to deliver the samples to the stationary capillary from temperature-controlled storage stage. At synchrotron SAXS beamlines allowing acquisition times short enough (~1 s), the capillary could be connected to the size exclusion chromatography, and the scattering data are recorded as the separated fractions are leaving the column continuously (SEC-SAXS). Online SEC-SAXS experiments are valuable for nonstable complexes, samples with tendency to aggregate or to separate multimeric mixtures in dynamic equilibrium. Time-resolved SAXS studies are possible at specialized synchrotron beamlines, where processes as protein folding, kinetic, and dynamic are studied with millisecond time resolution (Svergun 2013).

14.2.1 Sample Requirements

For successful biological SAXS experiments, it is crucial to measure samples that are as pure as possible, monodisperse, and remain free of aggregation during the course of an experiment. A comprehensive user guide for sample and buffer requirements for biological X-ray and neutron scattering experiments was released by Nature Protocols (Jeffries 2016). The purity of a SAXS sample has to be higher than that required for X-ray crystallography or NMR spectroscopy due to the summed nature of the scattering contributions made in a SAXS profile from every atom in a sample. For example, a 15 kDa protein purified to 98% that contains 2% of a 100 kDa protein (that does not interact with the 15 kDa protein or itself) will generate a forward scattering intensity ($I(0)$) that is almost twice what is expected from a 100% pure 15 kDa sample. In this instance, the high molecular weight contaminant, although presented at a low concentration, dominates the scattering, and ruins meaningful data interpretation. Prior to synchrotron measurement, the quality of the sample should be carefully inspected. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in

combination with native-PAGE, size-exclusion chromatography (SEC), multiangle or right-angle laser-light scattering (MALLS/RALS), and dynamic light scattering (DLS) should be used to evaluate the purity, state, and stability of a protein in solution, as well as assessing the effects of altering sample environments and monitoring the effects of sample handling histories.

Size exclusion chromatography (SEC) allows more precise quantitative analyses of the components present within a sample. SEC reveals the size and oligomerization state of a protein, information on the concentration or time-dependent stability of a protein oligomer or complex. Dynamic light scattering (DLS) offers a quick and simple way to evaluate sample monodispersity and monitors the effects of concentration or time-dependent aggregation using a minimal amount of sample (2–10 µl). Furthermore, DLS can save an investigator time when screening solvent conditions (salt, pH, temperature) maintain a sample in a desired state for SAXS analysis. In general, only monodisperse samples, i.e., solution of macromolecules or complexes with monomodal size distribution determined by DLS are suitable for further SAXS analysis (Jeffries 2016).

Sample volume required for one SAXS measurement is usually in range of 10–30 µl. Minimal required protein concentration for sufficient signal-to-noise ratio is inversely proportional to the molecular mass of the protein in the sample. For example, 15 kDa protein will produce sufficient signal at concentration 2 mg/ml, whereas 300 kDa protein could be obtained SAXS data with similar signal-to-noise ratio at 0.1 mg/ml. Usually, concentration series of minimum three concentrations are measured to reveal plausible concentration effects. Concentration-dependent interparticle interaction as attraction or repulsion may bias the SAXS data particularly at low q region. Example of typical concentration series could be 10, 5, and 2.5 mg/ml.

The exceptions to these "quantity guides" have to be amended on a case-by-case basis. If needed, less or more concentrated samples are measured to obtain good quality data not affected by concentration effects. Preparing 100–300 µl of biological sample as a minimum provides sufficient material to handle with confidence and to increase the probability to obtain relevant SAXS data.

14.2.2 Buffer Requirements

Important step in the SAXS data reduction process is the accurate subtraction of the scattering contributions made by the solvent from the sample scattering to "reveal" the scattering from a macromolecule or complex of interest. Inaccuracies in the solvent subtraction will result in the incorporation of additional solvent terms to the resulting, reduced, scattering profile that can cause significant perturbations in the structural parameters derived from the data.

Most of common biological buffers can be used for SAXS data collection. Recommended concentration of salts is less than 1 M. Glycerol in concentration 5–10% *v/v* has a protective and stabilizing effect and in higher amount could

diminish the SAXS contrast. Concentration of glycerol should not exceed 30% *v/v*. Detergents should be used in concentration less than critical micelle concentration.

The buffer should be identical to the buffer in the macromolecular sample. As rule of thumb, either a dialysis or SEC solvent exchange should be the last preparative step prior to SAXS experiment to obtain an equivalent sample solvent blank. The easiest way to match a solvent with a sample is to perform dialysis of the sample against a buffer of choice and collect scattering data from both the dialyzed sample and the postdialysis buffer. If the dialysis is impractical, (e.g., self-aggregation over time), the size exclusion chromatography (SEC) should be used to obtain matched solvent blank from the "macromolecular-free" fractions that have passed through the column. In case of use of an expensive, perishable cofactor (e.g., NADPH) that cannot be wasted in the preparation of liters of dialysis or SEC buffer, it is recommended to prepare small volume of the additive as a concentrated stock solution and add equivalent volume of the additive to the sample and to the matched solvent (after dialysis or SEC) immediately prior to collecting data.

14.3 SAXS Data Analysis

The scattering pattern from biological samples is recorded at the two-dimensional detector usually. To normalize the scattering to the incident beam intensity, the open beam intensity is measured, and the transmission factor is calculated for each sample or buffer solution. In case of ideal solution of macromolecules, the scattering pattern is isotropic and could be radially averaged. First, the position of the direct beam is located, and then, the recorded intensities of all pixels with the same distance to the center of the direct beam are averaged. Thus, the scattering data from two-dimensional detector are reduced to one-dimensional curve. Typically, SAXS data sets are stored as text files containing three columns (scattering vector, scattering intensity, standard deviation of scattering intensity). To obtain better data statistics of the SAXS data for further analysis, multiple exposures of sample and buffer could be averaged or merged.

Subtraction of the scattering contribution from the solvent allows obtaining the scattering data corresponding to scattering from biomacromolecules of interest. Importance of accurate buffer subtraction is discussed in Sect. 14.2.1. Radially averaged data of the buffer are subtracted from averaged sample data, normalized to sample concentration, and the subtracted curve is used for further analysis.

14.3.1 Radiation Damage Check

Ionizing radiation such as X-ray could damage the biomacromolecules and affect their structure during the SAXS experiment. Majority of the radiation damage of the biological samples is caused by free radicals produced by radiolysis of water.

Radiation damage causes aggregation, fragmentation, or partial unfolding of the biomacromolecules, which could be observed in SAXS data. It is necessary to pick up only SAXS curves not affected by radiation damage for further analysis. Usually, multiple exposure frames (e.g., 20) are recorded at the detector, whereas the sample is in the capillary, and the radiation check is performed by comparing unsubtracted scattering curves, where the curves recorded earlier serve as a reference. If the curves recorded later differ significantly, then those should be neglected from subsequent data averaging. Addition of ~5% *v/v* glycerol into sample buffer and/or oscillation of the sample in the capillary during data acquisition serve well as a radiation protection.

14.3.2 Guinier Analysis

Guinier analysis is one of the first steps in SAXS data evaluation, following initial data processing steps as averaging, buffer subtraction, and concentration normalization, etc. Guinier analysis provide information as radius of gyration of the particle, sample condition (monodispersity, aggregation, repulsion), and forward scattering intensity extrapolation, which is proportional to molecular mass of the biomacromolecule. André Guinier showed that in very low angles, the intensity decay is proportional to radius of gyration regardless the particle shape (Guinier 1939). For monodisperse globular particles, the Guinier approximation is given by $I(q) = \$ $\$I(0) \exp.\left(-Rg^2q^2/3\right)$ where $I(0)$ is extrapolated intensity at zero angle. Radius of gyration Rg is mechanic size parameter describing the distribution of mass of the particle. Rg could be defined as root-mean-square distances of the excess electron density to the center of gravity of the particle.

Guinier analysis is performed in Guinier plots, where the scattered intensity on natural logarithmic scale is plotted as a function of scattering vector square. In the Guinier region (limited to maximal scattering vector $s < 1.3/Rg$), the scattering intensity could be fitted by straight line. The slope of this line is proportional to particle Rg, and by extrapolation to zero angle, the forward scattering intensity $I(0)$ is obtained. If the Guinier plot in the Guinier region is not linear, sample is considered to be aggregated or affected by interparticle repulsion. Scattering data from aggregated samples should not be further analyzed, and attention should be focused on sample preparation. Note, linear Guinier region is not a proof of monodispersity of the sample: oligomeric mixtures or samples of complexes containing free subunits could exhibit linear Guinier behavior and intermediate values of Rg and $I(0)$.

14.3.3 Porod Volume

Excluded volume of the studied particles could be determined from the scattering data. Günther Porod shows the asymptotic decay of the scattering intensity at high q range as $I(q) \sim q^{-4}$ (Porod 1951). The integral of

$$Q = \int_0^\infty q^2[I(q) - K]\,dq$$

is called Porod invariant Q, where K is a constant determined to ensure the asymptotic intensity decay proportional to q^{-4} at higher s range. Porod invariant is related to the volume Vp of the particle by

$$V_\mathrm{p} = \frac{2\pi^2 I(0)}{Q}$$

where $I(0)$ is the extrapolated forward scattering intensity, see Sect. 14.3.1. The Porod volume is informative for well-folded macromolecules, whereas Porod volume estimation of flexible macromolecules could be affected by not accurate determination of Porod invariant. As a rule of thumb, the Porod volume of well-folded protein macromolecules is proportional to molecular weight by MM $\approx V\mathrm{p} \times 0.625$, see Sect. 14.3.6.

14.3.4 Kratky Plot Analysis

Flexibility or compactness of the biomacromolecule could be qualitatively evaluated by inspection of Kratky plot (Glatter and Kratky 1982), where $q^2 \times I(q)$ is plotted as a function of q (Fig. 14.3). According to Porod law approximation, the scattering intensity of a compact, globular particles decay proportionally to q^{-4}, which could be observed as a bell-shaped curve in the Kratky plot. Scattering intensity of unfolded macromolecules as intrinsically disordered proteins (IDP) decays slower, e.g., random chain proportionally to q^{-2}, which could be observed in the Kratky plot as plateau followed by monotonic increase. Scattering intensity of partially unfolded macromolecules as multidomain proteins with flexible linkers exhibits intermediate behavior in the Kratky plot.

Estimation of the folding state by inspection of the Kratky plot is routine step of SAXS data. The Kratky plot analysis is used for detection of flexibility, in folding/unfolding experiments. Note, scattering intensity of rigid but elongated particles decays slower (proportionally up to q^{-1}); thus, "flexible-like" shape of scattering data in Kratky plot should be considered as indication, rather than the proof of flexibility.

14.3.5 Pair-Distance Distribution Function

Indirect Fourier transform of the scattering data results to the pair-distance distribution function of the single macromolecule. The pair-distribution function $P(r)$ describes the distribution of distances between pairs of points (atoms) within the

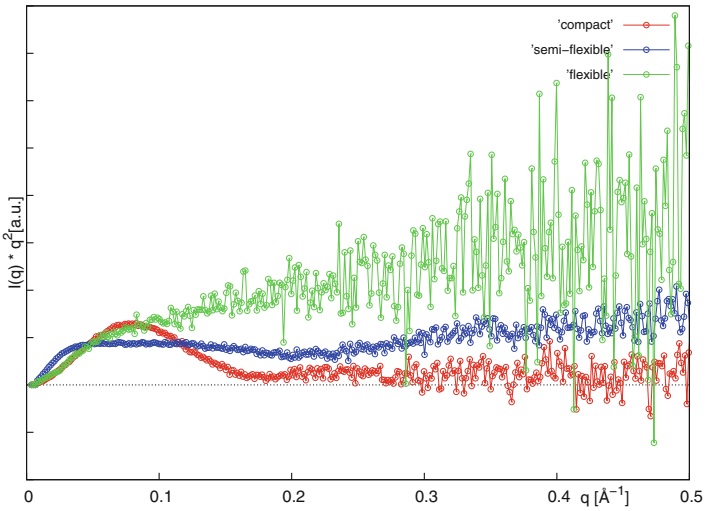


Fig. 14.3 Kratky plot and evaluation of flexibility. The scattering intensity of a compact, globular particles (in red) decays approximately as q^{-4}, which could be observed as a bell-shaped curve in the Kratky plot (q^2 $I(q)$ vs q). Scattering intensity of unfolded macromolecules (in green) decays slower, which could be observed in the Kratky plot as plateau followed by monotonic increase. Scattering intensity of multidomain proteins with flexible linkers exhibits intermediate behavior in the Kratky plot

macromolecule. By defining correct $P(r)$, the maximal chord length of the particle (Dmax) is obtained. The $P(r)$ is used for shape restoration experiments using the ab initio modeling programs. In ideal case of monodisperse solution of not interacting homogenous particles, the pair-distance distribution function is related to scattering intensity by

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) \frac{\sin\,(qr)}{qr} dq$$

where $I(q)$ is the scattering intensity, q is the scattering vector, and r is the distance in real space. To solve this equation, the precise scattering intensity measurement in angular range from zero to infinity is needed. In practice, the scattering intensity is measured in limited angular range and containing inherent statistical and systematic errors. The indirect Fourier methods were developed to overcome these problems using regularized scattering data and iterative parameterization (Svergun 1992). By definition, the $P(r)$ function starts smoothly from zero at $P(0)$ and should terminate smoothly to zero at $r = D$max. Deviation from zero value at $P(0)$ could be caused by incorrect background subtraction. Examples of theoretical $P(r)$ function of basic geometrical bodies of the same Dmax are shown in Fig. 14.4.

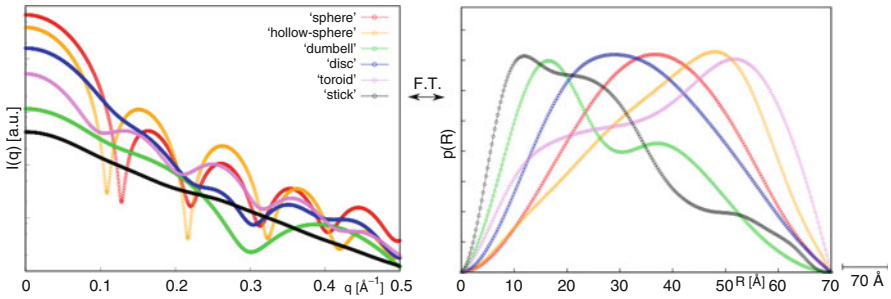


Fig. 14.4 Pair-distance distribution function. Simulated solution scattering profiles (left) of characteristic geometrical bodies and corresponding pair-distance distribution functions (right). Note, all geometrical bodies are of the same *D*max. See text for details

14.3.6 Molecular Mass Estimation

Most straightforward way to estimate the molecular mass from SAXS data is to use the relation between molecular mass *MM* and volume of the proteins, given by partial specific volume of proteins determined experimentally as 0.7245 cm^3/g. Approximate protein volume is given by

$$V[\mathring{A}^3] = \frac{0.7245[cm^3 g^{-1}] \times 10^{24}[\mathring{A}^3 cm^{-3}] \times MM[gmol^{-1}]}{6.023 \times 10^{23}}$$

that gave a relationship:

$$MM[\mathrm{Da}] = V_\mathrm{p}[\mathring{A}^3] \times 0.833$$

where *V*p is Porod volume. The experimental praxis shows that the factor of 0.833 leads to overestimation of *MM* and more appropriate is the factor ~0.625. This approach used for scattering data from well-folded monodisperse protein solutions results in *MM* estimation with error less than 20%. Such a precision is sufficient for rapid estimation of the oligomeric state or to distinguish the complex formation from mixture of its subunits. Similar approximation for nucleic acids is given by

$$MM[\mathrm{Da}] \approx V_\mathrm{p}[\mathring{A}^3]$$

Different way to estimate the *MM* of protein of interest is to use the Guinier extrapolation of forward scattering intensity *I*(0) of protein standard, as BSA (bovine serum albumin) or lysozyme:

$$MM_\mathrm{p} = \frac{I(0)_\mathrm{p}}{c_\mathrm{p}} \frac{MM_\mathrm{st}}{I(0)_\mathrm{st}/c_\mathrm{st}}$$

where MM_p and MM_{st} are the molecular masses of protein and protein standard, $I(0)p$ and $I(0)st$ are the forward scattering of the protein and standard, respectively, and c_p and c_{st} are the concentrations of the protein and protein standard. This approach requires two SAXS measurements and precise concentration determination of the protein and standard solution.

There is the number of other ways of *MM* estimation, as using forward scattering *I* (0) on the absolute scale (Mylonas and Svergun 2007) and volume of correlation (Rambo and Tainer 2013).

14.4 SAXS Modeling

SAXS modeling is based on comparison of theoretical and experimental scattering profiles. The theoretical scattering profiles are computed from molecular models generated ab initio or using atomic structures obtained by X-ray crystallography, nuclear magnetic resonance, molecular dynamics, and homology modeling. The theoretical SAXS profile could be computed from all pairwise interatomic distances by the Debey formula (Debey 1915), defined in general form as scattering from *N* spheres:

$$I(q) = \sum_{i=1}^{N_{sph}} F_i^2(q) + 2 \sum_{i=1}^{N_{sph}-1} \sum_{j=i+1}^{N_{sph}} F_i(q) F_j(q) \frac{\sin\left(qr_{ij}\right)}{qr_{ij}}$$

where $Fi(q)$ and $Fj(q)$ are the form factors of the spheres, rij is the distance between centers of the spheres i and j. Comparison and fitting of the theoretical and experimental SAXS data sets are evaluated by the χ^2 test:

$$\chi^2 = \frac{1}{N_p} \sum_{k=1}^{N_p} \left[\frac{I_{exp}(q_k) - c.I_{calc}(q_k)}{\sigma_k} \right]^2$$

where Iexp and Icalc are the experimental and theoretical intensities, respectively, σk of standard deviation k-th data point. In general, the algorithms used for SAXS modeling generate a model of the macromolecule, and the theoretical scattering profile is fitted to the experimental profile by minimization of the χ^2. In multiple cycles, the algorithms modify the generated model and evaluate the fit of scattering profiles. Structural modifications leading to higher χ^2 values are neglected, whereas changes improving the fit are kept till any additional modifications of the model do not improve the fit.

Still developing field of SAXS-based modeling and interpretation of SAXS data allow to study mixtures (Konarev et al. 2003), conformational changes (Panjkovich 2016), and adding missing fragments to crystal structures (Petoukhov et al. 2012), and analyze intrinsically disordered proteins (Ozenne et al. 2012) and more. Here, SAXS-based ab initio, rigid body, and ensemble modeling is introduced.

14.4.1 Ab Initio Modeling

The ab initio modeling does not require any initial information about the macromolecular system. Current ab initio shape reconstruction algorithms produce models composed of densely packed "beads" to describe low-resolution shape of the macromolecule in solution from its scattering data. Beads are representing spherical pieces of electron density of certain radius (dummy atoms) or spherical approximation of electron density of average amino acid (dummy residues). The widely used programs DAMMIN and GASBOR are introduced here as representatives of SAXS-based ab initio modeling software.

In the program DAMMIN (Svergun 1999), the search volume (sphere with diameter ~*D*max) is filled with the number of dummy atoms with fixed spatial positions. Each dummy atom is indexed to belong to either the solvent phase or the macromolecular phase, where the initial configuration is random. Then, in the number of cycles, the index of randomly selected bead is changed, and theoretical scattering of new configuration of dummy atom model is computed using spherical harmonics and compared with the experimental data. Additional penalties (disconnectivity, looseness, etc.) are introduced to score the reconfiguration of dummy atom model in respect to biological relevance. The simulated annealing technique is used to decide whether new configuration will be accepted. In the beginning, the simulated annealing temperature is high; thus, even a "worse" solution could be accepted. In each other cycle, the annealing temperature is cooled down; thus, the probability of accepting "worse" solution is decreasing, and algorithm is continuing until no improvement of the model is observed. In successful DAMMIN run, the theoretical scattering fits the experimental data.

The program GASBOR (Svergun et al. 2001) uses the dummy residues approach. Dummy residue corresponds to averaged amino acid weighed by their abundance; thus, it could be used for shape reconstruction of proteins only. The centers of the dummy acids correspond to potential Cα atom and in the model are separated by distance of 0.38 nm, as in naturally occurring protein macromolecules. The number of dummy residues is specified by user, as is known from primary sequence of the protein. The dummy residues model in GASBOR is surrounded by dummy waters representing the first hydration shell (Fig.14.5). The reconstruction algorithm is similar to dummy atoms approach as implemented in DAMMIN. The search model is randomly modified and scored by fit of the simulated scattering (computed by Debye formula) versus the experimental scattering data, and several penalties are introduced to enhance the biological relevance.

The resulting ab initio models are written in pdb format (Berman et al. 2000) and could be understood as low-resolution electron density map. The shape reconstruction from scattering data is inherently ambiguous; thus, multiple runs of the ab initio modeling programs could result in models with more or less different shapes. To evaluate the diversity of ab initio models, multiple runs (e.g., 10–20 runs) of ab initio modeling are performed, and resulting models are superimposed and compared.

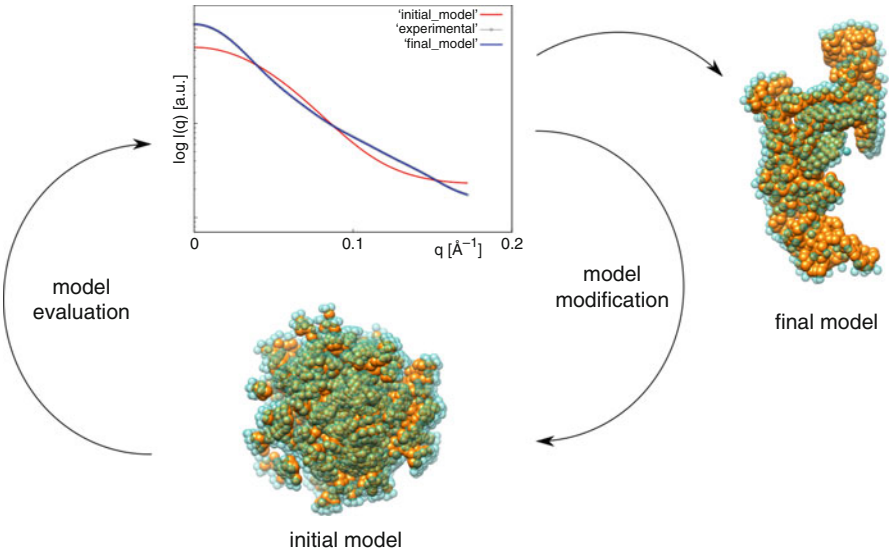


Fig. 14.5 Dummy residues ab initio modeling. A scheme of iterative process of shape reconstruction. Starting from random configuration of dummy residues, through the number of cycles, the final model is reconstructed by minimizing the scoring function. Theoretical scattering from final model fits the experimental data

Program suite DAMAVER performs alignment of the models, selects the most typical model, and identifies the outliers. After alignment, the filtered model composed of dummy atoms (residues) with highest occupancy is created. Filtered model could serve as an input for final round of ab initio modeling. Resulting refined model represents the shape of the particle in solution. The diversity of ab initio models generated from the same data set could be used to estimate the resolution of SAXS models by Fourier Shell Correlation, similarly as in cryo-electron microscopy (Tuukkanen et al. 2016). Usually, the resolution of SAXS models is in range of 10–60 Å.

14.4.2 Rigid Body Modeling

In the SAXS rigid body modeling, high-resolution atomic models are used to reveal the structure of macromolecular complexes. In the simplest case, the theoretical scattering of the atomic model is directly compared with experimental SAXS data. This way could be quickly validated structures obtained by X-ray crystallography, nuclear magnetic resonance or modeled by molecular dynamics, homology modeling, and molecular docking. Poor fit (see Sect. 14.4) indicates conformational changes caused by crystal packing, incorrect oligomeric state determination, or incorrect structure determination in general. The computation of the theoretical

scattering must take into account not only the atomic model but also the scattering of the solvent as well. The electron density of hydration layer significantly differs from the electron density of the bulk water. In general, the scattering intensity could be expressed as

$$I(q) = \left\langle \left| A_{\mathrm{a}}(q) - \rho_{\mathrm{s}} A_{\mathrm{s}}(q) + \delta\rho_{\mathrm{b}} A_{\mathrm{b}}(q) \right|^2 \right\rangle_{\Omega}$$

where $A_{\mathrm{a}}(q)$, $A_{\mathrm{ex}}(q)$, and $A_{\mathrm{h}}(q)$ are the scattering amplitudes from the atomic model in vacuo, from the excluded volume, and from the hydration shell, respectively. The scattering profiles computation is performed using the Debey formula (Debey 1915) or the multipole expansion of partial amplitudes (Svergun et al. 1995). The hydration shell is approximated as border layer of effective thickness and electron density (Svergun et al. 1995), dummy waters (Schneidman-Duhovny et al. 2010), molecular dynamics simulation (Grishaev 2010), and other methods. In Fig. 14.6, the scattering contribution of the solvent is illustrated, as it is implemented in CRYSOL (Svergun et al. 1995).

In more complicated cases, structure of the macromolecular complex is reconstructed from atomic models of its components. Atomic models of components are rotated and translated in respect to each other to reconstruct the structure of the complex. Such a translational and rotational search is performed by program SASREF (Pethoukov and Svergun 2005), which is assumed that full atomic structures of the all components are available. In case of hybrid modeling, as is implemented in CORAL (Pethoukov et al. 2012), missing fragments, linkers, or C- and N- termini of the polypeptide chains are modeled as dummy residues, see Sect. 14.4.1. Similarly as in ab initio modeling, in rigid body modeling, multiple cycles of model modifications are evaluated, simulated annealing global minimization is employed, and additional penalties are introduced to score the biological relevance.

14.4.3 Ensemble Modeling

Structural characterization of proteins containing substantial-disordered regions, proteins with multiple domains connected by flexible linkers, or intrinsically disordered proteins (IDP) is very challenging. As these systems do not exist in one conformation in solution, the SAXS-based ab initio or rigid body modeling reconstructions will not reflect the flexible nature of the studied macromolecule. In the ensemble modeling, as is implemented in program EOM (Ensemble Optimization Method, Bernadó et al. 2007), extensive pools of models with random conformation of macromolecules are generated, and genetic algorithms are employed to select subset of models. The sequence of the studied polypeptide serves as an input for random pool generation. If known, atomic models of the polypeptide chain substructures could be utilized as rigid bodies. The number of conformations in the

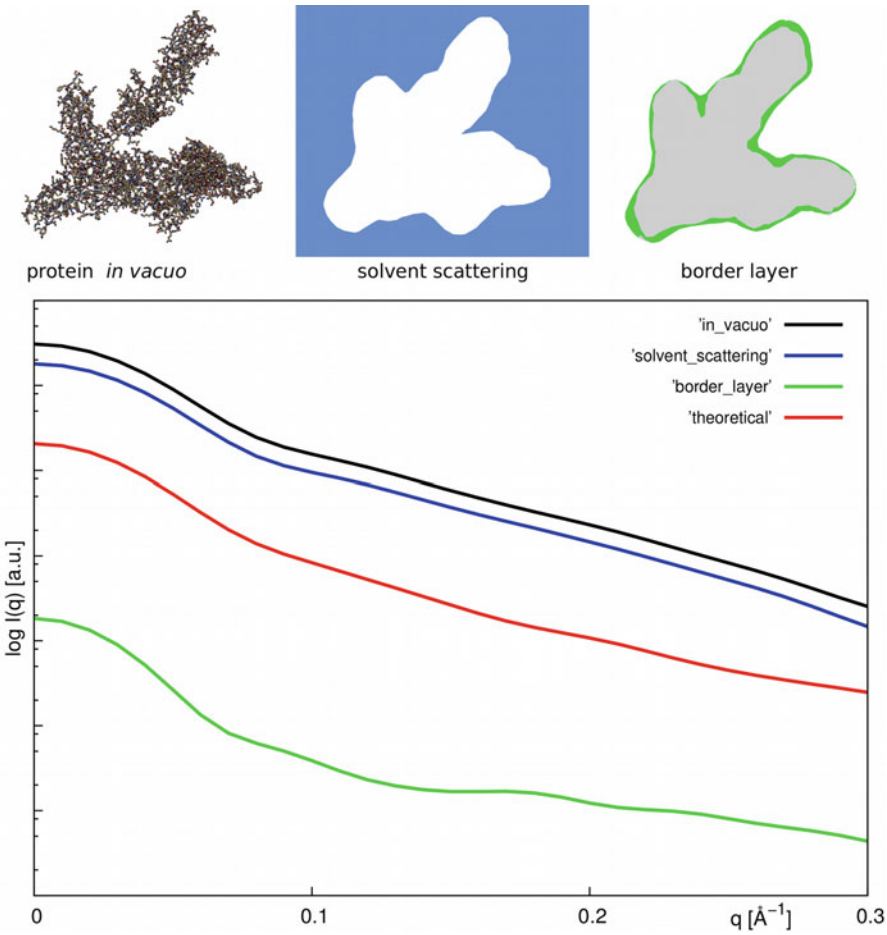


Fig. 14.6 Theoretical scattering. Computed scattering profile of the protein in solution (red) and contribution of border layer and excluded volume solvent scattering, as is approximated by CRYSOL. See text for details

random pool is sufficiently big, to sample the conformational space of the polypeptide, usually in range of 10^4. Theoretical scattering curve is precomputed using CRYSOL (see Sect. 14.4.2) for each conformation in the pool. The selection and optimization of the ensemble are performed by genetic algorithm procedure in a level of scattering curves. Theoretical scattering of the given ensemble is computed as sum of scattering contribution:

$$I(q_i) = \sum_{k=1} v_k i_k(q_i)$$

where i_k are the scattering intensities and v_k is the volume fraction of the individual k-th component. Initially, random subsets of scattering curves (chromosomes) consisting of number of scattering profiles (genes) undergo genetic algorithm operations as mutation and crossing in number of generations. Mutation means random exchange of genes from chromosome for genes from the pool, whereas crossing is exchange of genes between two chromosomes of the same generation. During multiple cycles, the ensembles are scored by the quality of the fit to the experimental data, and the best chromosomes are selected for the next generation. After 1000–5000 generations, ensemble with best fit to the experimental data is generated, and individual conformers are reconstituted. Importantly, selected conformers (and their volume fractions) are not representing discrete conformation of the macromolecule in solution or the most populated conformers. The optimized ensemble describes the nature of flexibility, size, and shape of the macromolecule in solution, which could be evaluated by comparison of *Rg* and *Dmax* distribution of the ensemble and the random pool (Svergun 2013).

14.5 Critical Assessment of SAXS Data

Summed nature of scattering data requires careful sample preparation, quality control, and accurate buffer subtraction. Scattering contribution from nonspecific aggregates, impurities, or not-matching buffer could significantly affect the data analysis. Comprehensive guidelines for good-quality biological sample preparation could be found in Jeffries 2016. Inherent ambiguity of the SAXS modeling demands special caution and critical assessment of SAXS models and interpretation of data. Freely accessible archive SASBDB (Valentini et al. 2015) has been established to facilitate the access to SAXS and SANS data sets, experimental details, and models. Recent recommendations for biological SAXS and SANS modeling, released by Trewhella et al. 2017, were prepared in consultation with International Union of Crystallography, Worldwide Protein Data Bank, and Small-Angle Scattering Validation Task Force. It highlights the importance of the accuracy and transparency of data acquisition and clear presentation of SAXS experiment and modeling. The risk of over-fitting or misinterpreting of the SAXS data should be minimized by cross validation and integrative approaches of structural biology.

References

Berman HM (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242

Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129:5656–5664

Debye PZ v R (1915) Ann Phys 351:809–823

Glatter O, Kratky O (1982) Structure analysis by small-angle X-ray and neutron scattering. Academic Press, New York

Grishaev A, Guo L, Irving T, Bax A (2010) Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. J Am Chem Soc 132:15484–15486

Guinier A (1939) La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. Ann Phys 11:161–237

Jeffries CM et al (2016) Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. Nat Protoc 11:2122–2153

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) *PRIMUS*: a windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr 36:1277–1282

Mylonas E, Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. J Appl Crystallogr 40:s245–s249

Ozenne V et al (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28:1463–1470

Panjkovich A, Svergun DI (2016) Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. Phys Chem Chem Phys 18:5707–5719

Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89:1237–1250

Petoukhov MV et al (2012) New developments in the *ATSAS* program package for small-angle scattering data analysis. J Appl Crystallogr 45:342–350

Porod G (1951) Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen: I. Teil. Kolloid-Z 124:83–114

Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. Nature 496:477–481

Schneidman-Duhovny D, Hammel M, Sali A (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res 38:W540–W544

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 25:495–503

Svergun D, Barberato C, Koch MHJ (1995) *CRYSOL* – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28:768–773

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 76:2879–2886

Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80:2946–2953

Svergun DI, Koch MHJ, Timmins PA, May RP (2013) Small angle x-ray and neutron scattering from solutions of biological macromolecules. Oxford University Press

Trewhella J et al (2017) 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. Acta Crystallogr Sect Struct Biol 73:710–728

Tuukkanen AT, Kleywegt GJ, Svergun DI (2016) Resolution of *ab initio* shapes determined from small-angle scattering. IUCrJ 3:440–447

Valentini E, Kikhney AG, Previtali G, Jeffries CM, Svergun DI (2015) SASBDB, a repository for biological small-angle scattering data. Nucleic Acids Res 43:D357–D363

Chapter 15
Computational Molecular Modeling Techniques of Biomacromolecular Systems

Jozef Hritz and Arnost Mladek

15.1 Introduction

Molecular modeling (MM) is a collection of techniques (the most often computational) for analyzing, representation, manipulation, and simulation of 3D molecular structures. Techniques of computational physics and chemistry enable to calculate physicochemical properties of a system and to study its time evolution. Importance of these techniques is underlined by the Nobel prizes in chemistry awarded to Martin Karplus, Michael Levitt, and Arien Warshel in 2013 for "Multiscale models for complex chemical systems."

In the presence, it is not possible to use methods of quantum mechanics (QM) for systems containing a large number of atoms (e.g., proteins and nucleic acids) or for the large timescale simulations. The reason is the enormous computational demand for such type of computations. In many cases (when the electron movement does not play significant role), it is possible to use methods of molecular mechanics based on simple, Newtonian physics-based empirical force fields for such atomic/molecular systems. Approaches of molecular mechanics for small molecules are in several ways different from those for proteins; therefore, they will be described separately.

J. Hritz (✉) · A. Mladek
CEITEC MU, Masaryk University, Brno, Czech Republic
e-mail: jozef.hritz@ceitec.muni.cz

J. Hejátko, T. Hakoshima (eds.), *Plant Structural Biology: Hormonal Regulations*,
https://doi.org/10.1007/978-3-319-91352-0_15

15.2 Structure and Topology of Molecules

15.2.1 Geometry of Biomolecules

Geometry of a molecule is defined by a determination of its atom positions. For this purpose, the most used representation is the common Cartesian coordinate system (x, y, z coordinates).

A dihedral (torsion) angle *a-b-c-d* is defined as the angle between two planes determined by atoms *a-b-c* (Fig. 15.1, atoms 9-1-2, blue rectangle) and plane *b-c-d* (Fig. 15.1, atoms 1-2-3, orange rectangle).

The success or failure of whole molecular simulation critically relies on the quality of the initial molecular structure construction. It is, thus, worthy to spend sufficient time for this particular step. It is especially crucial point for complex structures such as proteins or nucleic acids. The three most common approaches for the construction of 3D protein structures/complexes are

1. Usage of the databases of experimentally derived atomic coordinates of biomacromolecules if their structure has already been determined. The most common is the Brookhaven (pdb) database (http://www.rcsb.org/pdb/). The structures deposited in the pdb database are most often determined by the X-ray crystallography and NMR. The Brookhaven data files have standardized format (.pdb), of which header comprises some general information about the protein such as the official name, references, resolution of the crystal structure, remarks about the secondary structure of protein, etc. The atomic coordinates are listed in adjacent to the header. Atoms belonging to the standard amino acid residues are labeled as "ATOMS." The rest of atoms are labeled as "HETATOMS."

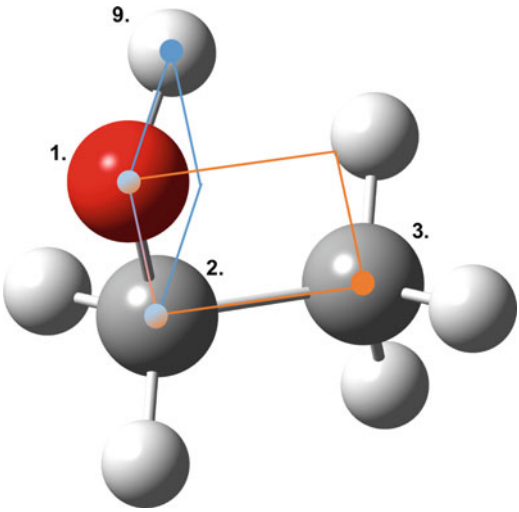


Fig. 15.1 3D structure of ethanol. Dihedral angle 9-1-2-3, i.e., angle between planes 9-1-2 (blue rectangle) and 1-2-3 (orange rectangle). The intersection of the planes (i.e., the axis of the rotation) goes through atoms 1 and 2

HETATOMS can either belong to nonstandard amino acids or, in the case of complexes, to the ligand molecules involved in the ligand–protein interactions.
2. Homology (comparative) modeling of 3D protein structure in cases in which the structures of homologous proteins are known. This approach is described in more detail in Sect. 15.2.2.
3. Construction of biomolecular complexes by the docking procedure if the structures of individual parts are known or can be constructed.

For analysis and manipulation of molecular structures, computational visualization is very worthy, mainly for the complicated structures such as proteins, because it enables to create 3D projections (e.g., by programs like CHIMERA, PYMOL, or VMD) and rendering programs (such as POVRAY), which enable to create high quality 3D static pictures (see Fig. 15.1) or video sequences.

15.2.2 Homology (Comparative) Modeling

Homology modeling exploits the structural similarities between proteins to construct the unresolved 3D protein structure using the known structure of another protein (as the template). To do this, it is first necessary to decide which protein to use as the template and then to determine how to match the amino acids in the known structure. The template is often chosen as the known 3D structure of a protein of which sequence is the most similar to the protein with unknown structure by searching sequence database. In some cases, the backbone conformation can be simply transferred from the template to the unknown 3D protein structure. Sometimes, the conformation of loop regions joining the appropriate parts of the core is solved separately, to have a low internal energy and not to have any unfavorable interactions with the rest of the molecule. Once a protein backbone conformation has been derived, it is then necessary to assign conformations of the side chains (rotameric states).

Many popular approaches (e.g., SCWRL, MODELLER) are based on the hypothesis that a side-chain conformation is mainly determined by the local main chain conformation. This method uses side-chain backbone library, (i.e., lists of values of torsional angles of side chains and their probability of occurrence for residues at given values of backbone) and explores these conformations to minimize side-chain backbone clashes and side chain–side chain clashes. The initial structures obtained from the comparative modeling can be rather high in energy; thus, energy minimization (structure relaxation) is often performed to obtain the final structure. Once a protein model has been constructed, it is important to examine it for validity. A simple test is to generate Ramachandran map, to find out whether the amino acid residues occupy the energetically favorable regions.

15.2.3 Empirical Force Fields

Molecular mechanics uses analytical, differentiable, and relatively simple potential energy functions, for describing the interactions between a set of atoms specified by their Cartesian coordinates. Derivation of this potential energy functions comes from the Born–Oppenheimer approximation that uses the fact that atom's nuclei are much heavier than electrons and move on a timescale that is about two orders of magnitude longer than that of the electrons. Therefore, according to the Born–Oppenheimer approximation, it is sensible to regard the nuclei as frozen or fixed as far as the electronic part of the problem is concerned. By using this approximation, it can be shown that the motion of the nuclei will depend on the quantity depending only on the coordinates of the nuclei, not electrons.

This quantity is called interatomic potential. Analytical form of the interatomic potential is usually obtained by a fitting procedure attempting to obtain as accurately as possible representation of the real potential. The most successful fitting procedures, having general utility, lead to terms in the potential that can be ascribed to chemically meaningful interactions. For example, molecular mechanics potentials have typically simple analytic terms that can be interpreted as an energetic penalty for deviations from standard bond lengths, bond angles, and bond dihedral angles (bonded interactions between atoms connected by chemical bond), together with simple analytic terms for long-range nonbonded interactions: electrostatic (described by the well-known Coulomb term) and van der Waals interactions (described by the Lennard-Jones term). Empirical force fields used in molecular mechanics methods have three important components:

I. Set of *atom types* defining the chemical environment of an atom, characterized by the atom number, hybridization, and neighboring environment. For example, the common protein AMBER99SB force field uses five "standard" atom types for oxygen and several other types for various water molecule models beyond the common TIP3P water model:

O	carbonyl group oxygen (C = O)
OW	oxygen in TIP3P water
OH	oxygen in hydroxyl group (−OH)
OS	ether and ester oxygen
O2	carboxyl and phosphate group oxygen

II. Functional form: the set of equations (called the potential functions) used to generate the potential energies and their derivatives – the forces.
III. Parameter sets used in these potential functions.

The functional forms describe the potential energy dependence on the atoms configuration. The acting force is then calculated as the negative value of the potential energy gradient. The potential functions can be divided into three parts (1–3):

1. Nonbonded interactions are pair additive:

$$V\left(\vec{r_1}, \vec{r_2}, \ldots, \vec{r_n}\right) = \sum_{i<j} V_{ij}\left(r_{ij}\right) \tag{15.1}$$

The nonbonded interactions contain a repulsion term, a dispersion term, and a Coulomb term. The repulsion and dispersion terms are usually combined in the Lennard-Jones (or 6–12 interaction) potential. In addition to that, partially charged atoms interact electrostatically through the Coulomb term. The Lennard-Jones interaction is described by two different forms connected via two combination rules. The first form of the Lennard-Jones potential is as follows:

$$V_{\mathrm{LJ}}\left(r_{ij}\right) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \tag{15.2}$$

where r_{ij} is the distance between atoms i and j. The parameters $C_{ij}^{(12)}$ and $C_{ij}^{(6)}$ depend on the respective atom types. They are calculated using the following combination rules:

$$C_{ij}^{(12)} = \sqrt{C_{ii}^{(12)} C_{jj}^{(12)}} \tag{15.3a}$$

$$C_{ij}^{(6)} = \sqrt{C_{ii}^{(6)} C_{jj}^{(6)}} \tag{15.3b}$$

The parameters $C_{ii}^{(12)}$ and $C_{ii}^{(6)}$ are included in the force field parameter sets. The alternative form of the Lennard-Jones potential is

$$V_{\mathrm{LJ}}\left(r_{ij}\right) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{15.4}$$

with the following combination rule:

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj}\right) \tag{15.5a}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{15.5b}$$

where the parameters σ_{ii} and ε_{ii} are part of the force field parameter sets.

The Coulomb interaction between two (partially) charged particles is given by the Coulomb law:

$$V_{\mathrm{C}}\left(r_{ij}\right) = \frac{1}{4\pi\varepsilon_0\varepsilon_{\mathrm{r}}}\frac{q_i q_j}{r_{ij}} \tag{15.6}$$

where ε_0 is the dielectric constant of vacuum, ε_{r} is the relative permittivity, and q_i, and q_j are partial atom charges. In molecular mechanics, calculation of the electrostatic interaction energy mostly uses approach with point charges located at the

center of the atom positions. This produces a very simplified picture of the molecular electron distribution, but on the other hand, it is computationally very efficient approach. The application of these methods allows the rapid computation of electrostatic energies, even for macromolecules with more than several thousand atoms.

2. Bonded interactions are covalent bond stretching (2-body), angle bending (3-body), and dihedral angle (4-body) interactions. The bond stretching between two covalently bonded atoms i and j is usually represented by a harmonic potential:

$$V_{\mathrm{b}}\left(r_{ij}\right) = \frac{1}{2}k_{ij}^{\mathrm{b}}\left(r_{ij} - r_{ij}^{0}\right)^{2} \tag{15.7}$$

where k_{ij}^{b} is the force constant of the given bond and r_{ij}^{0} is the equilibrium value of the bond length. The bond angle bending between three covalently bonded atoms i, j, and k is also represented by a harmonic potential of the angle θ_{ijk}:

$$V_{\mathrm{a}}\left(\theta_{ijk}\right) = \frac{1}{2}k_{ijk}^{\theta}\left(\theta_{ijk} - \theta_{ijk}^{0}\right)^{2} \tag{15.8}$$

where k_{ijk}^{θ} is the bending force constant, and θ_{ijk}^{0} is the equilibrium angle value. Improper dihedrals are special type of dihedral interactions used to force atoms to remain in a plane (e.g., planar aromatic rings) or to prevent transition to a configuration with the opposite chirality. It is also represented by a harmonic potential of the dihedral angle:

$$V_{\mathrm{id}}\left(\varepsilon_{ijkl}\right) = k_{ijkl}^{\varepsilon}\left(\varepsilon_{ijkl} - \varepsilon_{ijkl}^{0}\right)^{2} \tag{15.9}$$

where k_{ijkl}^{ε} is the respective force constant, and ε_{ijkl}^{0} is the equilibrium value of the given dihedral angle. The proper dihedral interaction is typically described by a periodic function:

$$V_{\mathrm{d}}\left(\phi_{ijkl}\right) = k_{\phi}\left[1 + \cos\left(n\phi_{ijkl} - \phi_{ijkl}^{0}\right)\right] \tag{15.10}$$

where k_{ϕ} is the dihedral force constant, ϕ_{ijkl} is the angle between the $i - j - k$ and the $j - k - l$ planes, ϕ_{ijkl}^{0} is the phase angle (offset), and n is the multiplicity.

The origin of necessity of special treatment of 1–4 interactions is that in the chain of covalently bonded atoms, the interactions of atom i with atoms $i + 1$ and $i + 2$ are mainly quantum mechanical, and they cannot be modeled by the Lennard-Jones potential. Instead, it is assumed that these interactions are adequately modeled by a harmonic bond and angle terms (Eqs. 15.7 and 15.8). For the interactions of atom i with atom $i + 3$, the normal Lennard-Jones repulsion is sometimes too strong; therefore, for some of these interactions, the Lennard-Jones repulsion is reduced in the selected force field, which is implemented by keeping a separate list of 1–4 and normal Lennard-Jones parameters. In other force fields, such as OPLS (Kaminski

et al. 2001), the standard Lennard-Jones parameters are reduced by a factor of two in the case of 1–4 interaction, but in that case also the Coulomb interaction is scaled.

3. Special interactions such as position restraints and distance restraints are used for imposing restraints on the motion of the system, either to avoid disastrous deviations or to include knowledge from the experimental data. Position restraints are used to restrain particles to fixed reference positions $\vec{R}_i$. The following form is used:

$$V_{pr}\left(\vec{r}_i\right) = \frac{1}{2}k_{pr}\left|\vec{r}_i - \vec{R}_i\right|^2 \tag{15.11}$$

where k_{pr} is the respective force constant.

Distance restraints add a penalty to the potential energy when the distance between specified pairs of atoms deviates from the reference distance. The potential form in GROMOS force field (Hermans et al. 1984; van Gunsteren and Berendsen 1987) is quadratic below a specified lower bound (r_0) and between two specified upper bounds (r_1, r_2) and linear beyond the largest bound (r_2):

$$V_{dr}\left(r_{ij}\right) = \begin{cases} \frac{1}{2}k_{dr}\left(r_{ij} - r_0\right)^2 & \text{for} \quad r_{ij} < r_0 \\ \frac{1}{2}k_{dr}\left(r_{ij} - r_1\right)^2 & \text{for} \quad r_1 \le r_{ij} < r_2 \\ \frac{1}{2}k_{dr}\left(r_2 - r_1\right)\left(2r_{ij} - r_2 - r_1\right) & \text{for} \quad r_2 < r_{ij} \\ V_{dr}\left(r_{ij}\right) = 0 & \text{for} \quad r_0 \le r_{ij} < r_1 \end{cases} \tag{15.12}$$

where k_{dr} is the force constant.

Distance restraints are normally used to impose experimental data, such as nuclear overhauser effect (NOE), signals from 2D NMR studies. NOE signals provide distances between atom pairs in a molecule. Distance restraints limit distances during a molecular mechanics geometry optimization (Sect. 15.3) or molecular dynamics (Sect. 15.5) calculation. To be useful, molecular mechanics potentials have to balance the simplicity of the analytical form of the potential (so that the energy and its derivatives can be rapidly evaluated, even for large molecules such as proteins) with the accuracy of describing important characteristics of the exact potential energy function.

Thus, different force fields can have different functional forms of the potential energy. For example, the bond stretching is in GROMOS and AMBER force fields (usually used for large biomacromolecules) described by only quadratic terms. On the other hand, MM+ force field (used for small organic molecules) (Burkert and Allinger 1982) has both a quadratic and a cubic stretch term in the potential energy expression and contains also other additional terms such as explicit parameters for the description of hydrogen bonds.

The most computational consuming part in the calculation of the potential energy is the contribution from the nonbonded interactions. For the system containing

N atoms, there are $N(N-1)/2$ atom pairs. The number of atom pairs interacting by nonbonded interactions can be significantly reduced by applying a cutoff radius (or interactional radius) R_c beyond which no interactions are computed. Due to the r^{-6} distance dependence of van der Waals interactions, cutoff radius of 1 nm is usually used in the practice. For electrostatic interactions, cutoff radius of at least 1.4 nm should be used.

15.2.4 Periodic Boundary Conditions

Molecular modeling is usually performed in a simulation box. How large system is necessary to simulate for correct reproduction of properties of real system? Box containing 10^{24} water molecules can be considered as the example of macroscopic system. In such system, there are 6×10^{16} water molecules in the surface box area, and it means that for one molecule in the surface area, there are 2×10^{7} molecules inside the box. On the other hand, in the typical simulation system containing 1000 water molecules, there are more than 500 molecules in the box surface area, what is more than half of all molecules in the system. If it is not our purpose to simulate the behavior of water molecules in the close region to the box walls, then we have to remove the effect of the box walls presence. One of the possible ways how to eliminate surface effects is to introduce periodic boundary conditions, where the simulated box is surrounded by neighboring boxes, which are the copies of the original box. In the two dimensional case of the rectangular periodic boundary conditions, there are eight neighboring boxes (Fig. 15.2), in the three-dimensional case.

Introducing of periodic boundary conditions has two consequences. The first one is that when the atom leaves the simulation (central) box at one side, it enters the box with identical velocity at the opposite side at the translated image position (Fig. 15.2).

The second consequence is that atoms, which distance from one of the box walls is less than the cutoff interaction radius, will interact with the atoms in the neighbor cells. It removes the problem of simulation in the simple box, where molecules "felt" vacuum behind the box walls, and also molecules can pass through the walls, so molecules "do not feel" the presence of box walls. In practice, the convention of minimal image is used, which requires that atom interacts only with the closest image of the other atom. It means that when calculating the forces acting on the atom j in the central box, only the interaction with the nearest image of atom i is taken into account (Fig. 15.2).

The minimal image convention, i.e., that only interactions between nearest images should be evaluated, constrains the box dimension: the shortest box edge has to be longer that double of cutoff interaction radius R_c.

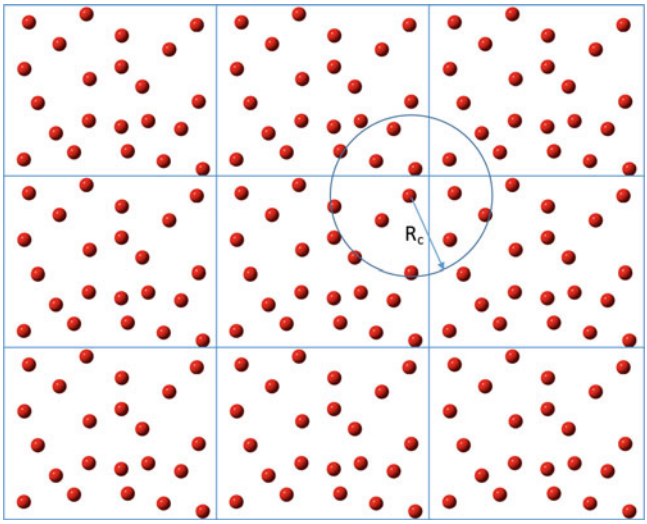


Fig. 15.2 Periodic boundary conditions (PBC) in 2D. Circle around atom (j) indicates the region within cutoff radius R_c. It is shown that atom j interacts only with the nearest image of the particle i. It can be also seen that particle j leaving the central box through the right wall will come back to the central box through the left wall

$$R_c < \frac{1}{2}\min(a, b, c) \tag{15.13}$$

This condition also ensures that atom will not interact with its own image in one of neighboring boxes. In the opposite case, the crystal would be simulated what would be an artifact of computation of a liquid or gas. The most important advantage of usage of periodic boundary conditions is that it effectively removes the boundary effects requiring only slight computational demand.

15.3 Geometry Optimization (Energy Minimization)

The Potential energy of biomolecular systems is often represented in the potential energy surface (PES) representation. Even for relatively simple systems, potential energy is multidimensional function of atomic coordinates.

For example, potential energy of dialanine (Fig. 15.3) is the function of 84 Cartesian, respectively, $3N - 6$, i.e., 78 internal coordinates. In cases where certain subparts are quite rigid, it is sometimes possible to select only few coordinates (or collective variables) along which the most relevant energy changes occur (e.g., ϕ and ψ dihedral angles for dialanine in Fig. 15.3.).

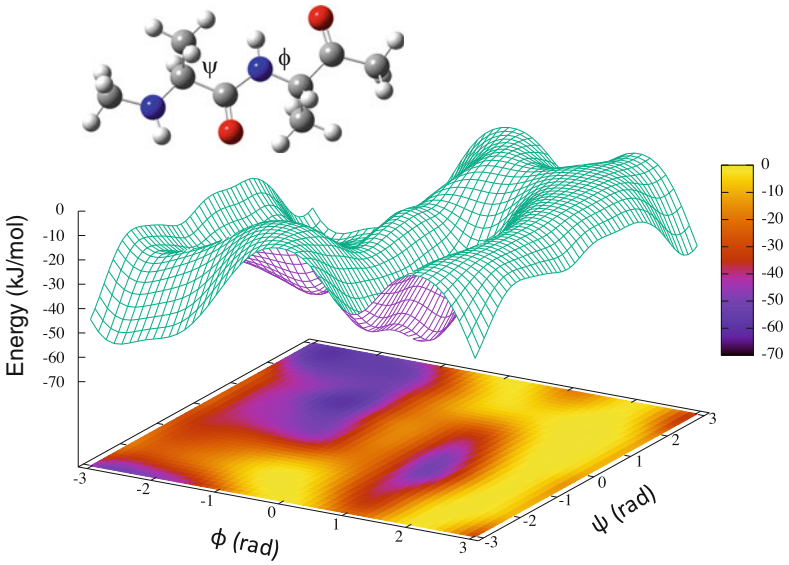


Fig. 15.3 Potential energy surface (kJ/mol) of a dialanine molecule (top) as a function of protein backbone ϕ and ψ torsional angles

From the whole hypersurface of potential energy, only few regions are of practical interest. The minima of a potential energy correspond to the stable conformations of the system or to the different chemical species. Minimum with the lowest potential energy is called the global minimum. Other minima are called local. Minima of potential energy are searched by energy minimization algorithms. Assuming that the potential energy V is function of coordinates x_1, x_2, …, x_{3N} than the condition for energetic minimum is as follows:

$$\frac{\partial V}{\partial x_i} = 0 \quad \text{and} \quad \frac{\partial^2 V}{\partial x_i^2} > 0 \tag{15.14}$$

For analytically defined potential energy surfaces, the most utilized function minimization methods are gradient methods (using the first-order derivatives). Gradient methods require the knowledge of functional form and its first derivative values. Two most popular gradient optimization methods are

(i) *The steepest descents method* is one of the simplest methods of optimization. The search for the configuration with small potential energy is performed by "moving" along the actual energy gradient in sequence of small steps (in cartesian space). The steepest descent performs well far from a minimum but converges slowly near a minimum. The principal problem with the steepest descent is that the successive directions of search are not conjugate directions (see below). Searching along the successive directions of the steepest descent might, for example, keep skipping back and forth across a long narrow valley.

A conjugate direction would be along the long narrow valley. The conjugate gradients methods are a considerable improvement over steepest descent methods and involve the search down conjugate directions rather than the steepest descent directions.

(ii) *The conjugate gradients method* searches along directions corresponding to the local quadratic approximation of the function. In the steepest descent method described in the previous paragraph, the first derivatives are recomputed at each step but not essentially looking at how these first derivatives are changing. The basis of the conjugate gradient methods is that the history of the first derivatives allows a more intelligent search direction than just the direction of the steepest descent. These methods search in conjugate gradient directions rather than simple steepest descent direction.

When applying energy minimization (EM) algorithms, the minimum energy configuration of a system is searched along the gradient of the potential energy through the configuration space. Because these procedures enable only downhill searching over the energy hypersurface, EM algorithms yield only a "closest" local minimum energy configuration that is usually not far from the initial one and a crossing of energy barriers is impossible. A more efficient way to find low energy configurations of large biomacromolecules is to apply molecular dynamics (MD, Sect. 15.5) that samples the phase space much better than EM algorithms. The available kinetic energy may be used to pass over energy barriers of k_bT order (k_b is the Boltzmann's constant and T is the absolute temperature). It has been shown that MD at elevated temperatures can be used to generate a variety of configurations (van Gunsteren and Berendsen 1984). Therefore, MD searches a larger part of configuration space for the energy minimum and generally ends up in the lower energy minimum than an ordinary energy minimizer does (Lautz et al. 1987). However, the application of MD starting from a highly strained, very high potential energy configuration is not recommended, because the immediate conversion of potential energy into kinetic energy will raise the temperature to unacceptably high values. In that case, one should first perform a number of EM steps to reduce the high potential energy of the system and release the strain. In the cases where close nonbonded contacts, stretched bond lengths or bent bond angles in a molecular system lead to the high potential energy, ten to fifty EM steps generally suffice to reduce the potential energy to values which are normal at room temperature.

15.4 Force Fields for Proteins

Due to protein size comprising hundreds or thousands atoms, classical mechanics is method of choice for the study of proteins. The molecular mechanics calculations are based on empirical force fields, described in Sect. 15.2.3. Force fields for protein modeling differ in some aspects from the force fields used for small molecules. Besides the specific parameterization for proteins, protein force fields are usually

simplified in the comparison to the force fields of small molecules. For example, some force fields (e.g. GROMOS) utilize united atom approach, in which non-polar hydrogens are not treated explicitly, but are included into the descriptions of the heavy atom to which they are bonded. Other type of simplification consists of the introducing of non-explicit solvent by means a distance dependent dielectric constant. Other simplifications established in the protein force fields are comprehensively described in (van Gunsteren 1991; Lavery and Rivail 1991). The most commonly used force fields in the protein molecular modeling are AMBER (Weiner et al. 1984, 1986), CVFF (Dauber-Osguthorpe et al. 1988), CHARMM (MacKerell Jr et al. 1998), GROMOS (Hermans et al. 1984), and OPLS (Kaminski et al. 2001).

15.5 Molecular Dynamics (MD)

Molecular modeling techniques described in the previous chapters such as quantum calculations of atomic charge distribution and energy minimization techniques describe only static characteristics of molecules. However, a molecular system at a room temperature is far from being static. A system of interacting atoms traverses multiple minima of the PES. One would like to know the multidimensional distribution function of all atomic coordinates and its evolution in time. In practice, only parts of the configuration space can be searched for relevant low energy conformations. The computer simulation technique of molecular dynamics provides the possibility to scan that part of configuration space that is accessible to the molecular system at the given temperature. MD simulations allow calculating the trajectory of all atoms in the simulated system during a period of time, where the time evolution of a set of interacting atoms is followed by integrating their equations of motion. By following the dynamics of a molecular system in space and time, we can obtain information about molecular geometries and energies, mean atomic fluctuations, local fluctuations (like formation/breakage of hydrogen bonds), protein/ligand binding, free energies, and the nature of various types of concerted motions. Classical mechanics is sufficiently precise for the most of practical cases, with the exception of those, where is not possible to neglect QM effects.

15.5.1 Molecular Dynamics Protocol

The basic idea of MD program is very simple: the forces acting on the atoms in the system are calculated, and in an iterative procedure, the atoms are moved in small time steps. There are five main steps in the MD algorithm:

1. Reading of parameters specifying the conditions of simulation run (e.g., temperature, type of thermostat and barostat, the length of simulation step, the number of steps, i.e., the total simulation length, restraints, etc.)

2. Initialization of system – determination of the initial positions and velocities of atoms
3. Calculation of net force acting on each of atoms
4. Integration of the Newton equations of motion
5. Analysis and visualization of MD simulation

Until the time of simulation is not exhausted, steps 3 and 4 are repeated. Simulation parameters from the first step are applied in the steps 2–5 which will be described in the next paragraphs.

15.5.2 Initialization of the Molecular System

At the beginning of a simulation, it is necessary to assign the initial position and velocity to each atom. It is possible to use initial homogeneous or random distribution of positions of small molecules. The initial atoms velocities are usually assigned on the basis of Maxwell–Boltzmann distribution of velocities for desired temperature:

$$p(v_{ik}) = \left(\frac{m_{\mathrm{i}}}{2\pi k_{\mathrm{B}} T}\right)^{\frac{1}{2}} \exp\left(-\frac{m_{\mathrm{i}} v_{ik}^2}{2k_{\mathrm{B}} T}\right) \tag{15.15}$$

where $p(v_{ik})$ is the probability that atom i with the mass m_{i} at temperature T has velocity v_{ik} in the direction k (axes x, y, z). Obtained velocities usually lead to the nonzero total translational momentum, which can be removed using the following relation:

$$v'_{ik} = v_{ik} - \sum_{j=1}^{N} \frac{v_{jk} m_j}{N m_i} \tag{15.16}$$

where N is the number of atoms, and m_j is the mass of atom j.

15.5.3 Calculation of Acting Force

Section 15.2 describes the calculation of the potential energy for the determined atoms positions. Acting force $\vec{F}_{ij}$ is the negative gradient of the corresponding pair potential energy $V(r_{ij})$:

$$\vec{F}_{ij} = -\nabla V(r_{ij}) \tag{15.17}$$

For example, the corresponding force to the Lennard-Jones potential energy given by Eq. (15.2) is

$$\vec{F}_{ij} = -\nabla V_{\mathrm{LJ}}(r_{ij}) = \left(12\frac{C_{ij}^{(12)}}{r_{ij}^{13}} - 6\frac{C_{ij}^{(6)}}{r_{ij}^{7}}\right)\frac{\vec{r}_{ij}}{r_{ij}} \tag{15.18}$$

and the corresponding force to the Coulombic potential energy given by Eq. (15.6) is

$$\vec{F}_{ij} = -\nabla V_{C}(r_{ij}) = \left(\frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_i q_j}{r_{ij}^{2}}\right)\frac{\vec{r}_{ij}}{r_{ij}} \tag{15.19}$$

It is enough to calculate the force for each atom pair only once, because when we know the force $\vec{F}_{ij}$, then according to the third Newton law of motion, $\vec{F}_{ij} = -\vec{F}_{ji}$. The net force acting on the atom i is the sum of pair interactions coming from other atoms. The number of pair interactions is proportional to N^2. Due to the computational demand, it is usual to take into account only the forces within a certain interaction radius (Sect. 15.2.4).

15.5.4 Integration of the Newton's Equations of Motion

Analytical solution of general motion of mass points is known for two body problem. In the simulations of real systems containing thousands particles, a numerical integration of equations of motion is used. In the simulation, the whole time of simulation is divided into the short time segments Δt during which the acceleration, i.e., the acting force is almost constant, and for such case, one can use the high-school physics equations:

$$\vec{r}\,(t+\Delta t) = \vec{r}\,(t) + \Delta t\,\vec{v}\,(t) + \frac{1}{2}(\Delta t)^2\,\vec{a}\,(t) \tag{15.20}$$

$$\vec{v}\,(t+\Delta t) = \vec{v}\,(t) + \Delta t\,\vec{a}\,(t) \tag{15.21}$$

where $\vec{r}$ is the positional vector, $\vec{v}$ is the velocity, and $\vec{a}$ is the acceleration. The disadvantage of such algorithm is its low numerical precision, because of adding small and large values. One of the several algorithms that remove the mentioned disadvantage is so-called leap-frog algorithm. The name leap-frog comes from the fact that positions and velocities in this algorithm are not calculated in the same time. The position in the time $t + \Delta t$ and the velocity in the time $t + \frac{1}{2}\Delta t$ are expressed as follows:

$$\vec{r}\,(t+\Delta t) = \vec{r}\,(t) + \Delta t\,\vec{v}\left(t+\frac{1}{2}\Delta t\right) \tag{15.22}$$

$$\vec{v}\left(t+\frac{1}{2}\Delta t\right)=\vec{v}\left(t-\frac{1}{2}\Delta t\right)+\Delta t\,\vec{a}\,(t) \tag{15.23}$$

The disadvantage of the leap-frog algorithm is that positions and velocities are not calculated at the same time. It means that it is not possible to determine the part of kinetic energy of the total energy in the same time when the atom positions are known and from which potential energy can be calculated. In each simulation, we face very important and practical question, which is the choice of integration time Δt. The choice of too short integration step allows simulating only restricted time of simulation. On the other hand, too long integration step can lead to the instabilities and artifacts due to strong repulsion between adjacent atoms. It can be nicely understood in the example of two atoms moving to each other. In reality as well as in the simulations with proper integration step, these atoms will continually repeal each other. For long integration step, the atoms can come from the distance in which they interact with each other only slightly to very close distance, where it is a very strong repulsion force between atoms. In next step, this strong repulsive force will "shoot away" atoms from each other. Such behavior in the presented simulation with too long time step is in the contrary with the reality. The time step is determined by the fastest movement in the system. For molecules, it is the high-frequency vibration of hydrogen, and in practice, integration time step of 0.5 fs is a good choice. In many cases, where it is not necessary to reproduce molecular vibrations, it is possible to use bond constraints algorithms (e.g., SHAKE (Ryckaert et al. 1977), LINCS (Hess et al. 1997)) with the integration step 2 fs (2.10^{-15} s).

15.5.5 *Molecular Dynamics at Constant Temperature and Pressure*

The molecular dynamic discussed in the previous sections is a technique for studying the time evolution of a classical system of N particles in constant volume. In such simulations, the total energy and the total momentum of the system are conserved constants of motion (it is not completely true because of numerical imperfections over millions of calculations). Simulations described above with the unchanged volume and number of particles are very close to the microcanonical; namely, the constant NVE ensemble. However, it is often more convenient to perform simulations in other ensembles (e.g., in the case of biomolecular simulations: NVT or NPT, where N is the number of particles, V is volume, T is temperature, and P is pressure). Simulations at constant temperature are used for systems, for which it is necessary to control temperature or simulate properties dependent on the temperature change (e.g., phase transition, protein folding, or unfolding). The relationship between the macroscopic quantity – temperature (T) and the microscopic quantity – the average kinetic energy $\langle E_{\mathrm{k}}\rangle_{\mathrm{NVT}}$ for the system without internal constraints is

$$\langle E_{\mathrm{k}} \rangle_{\mathrm{NVT}} = \frac{3}{2} N k_B T \tag{15.24}$$

where N is the number of particles in the system and k_{B} is the Boltzmann constant. The kinetic energy of whole system can be calculated as sum of kinetic energies of all particles:

$$E_{\mathrm{k}} = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \tag{15.25}$$

From the equality of right sides of Eqs. (15.24) and (15.25), for temperature, we will obtain expression:

$$T_{\mathrm{c}} = \sum_{i=1}^{N} \frac{m_i v_i^2}{3 N k_{\mathrm{B}}} \tag{15.26}$$

where T_{c} is the current temperature. The most simple way of how to keep the constant temperature in the system is to multiply all velocities in Eq. (15.26) by scaling factor λ changing the current temperature T_{c} to the required temperature T_{r}. The relationship between a current and required temperature is

$$T_{\mathrm{r}} = \sum_{i=1}^{N} \frac{m_i (\lambda v_i)^2}{3 N k_{\mathrm{B}}} = \lambda^2 \sum_{i=1}^{N} \frac{m_i v_i^2}{3 N k_{\mathrm{B}}}, \tag{15.27}$$

and by expression of the scaling factor λ from this equation, we obtain

$$\lambda = \sqrt{\frac{T_{\mathrm{r}}}{T_{\mathrm{c}}}}. \tag{15.28}$$

Thus, the simple way how to keep the system with the required temperature T_{r} is to calculate the current temperature T_{c} according the relation (15.26), from the relation (15.28) to calculate the scaling factor and finally to multiply velocities of all particles in the system by this factor. From Eq. (15.27), we get for the difference of required and current temperature using the relation:

$$\Delta T = T_{\mathrm{r}} - T_{\mathrm{c}} = \left(\lambda^2 - 1\right) T_{\mathrm{c}} \tag{15.29}$$

Disadvantage of temperature scaling method is that it does not allow the temperature fluctuations around the required temperature in the system that are known to be important for the biomolecular systems. For this reason, the next very often used way how to control the temperature in the system is the Berendsen thermostat (Berendsen et al. 1984) coupling a system with the external thermostat with temperature T_{ter}. Particle velocities are multiplied in each step by such factor λ, for which the rate of temperature change is proportional to the difference of temperatures of reservoir and

system. This leads to the differential equation for the time evolution of the temperature $T(t)$:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{\text{ter}} - T(t)) \tag{15.30}$$

where τ is coupling parameter, which value determines how strong is reservoir coupled with the system. For the temperature difference between two-time moment, separated by the time Δt:

$$\Delta T(t) = \frac{\Delta t}{\tau}(T_{\text{ter}} - T(t)) \tag{15.31}$$

By expression of parameter λ from the equality of right sides of Eqs. (15.29) and (15.31), we get

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{\text{ter}}}{T(t)} - 1\right)} \tag{15.32}$$

From this equation that can be seen that, if τ is large, then the coupling is weak. For the small τ, the coupling is very strong, and when τ equals to the time step Δt, then algorithm of Berendsen thermostat comes to the simple velocity scaling method described at the beginning of this section (Eq. (15.32) will have the same form as Eq. (15.28)). The typical value of τ in the biomolecular simulations is about 1 ps, i.e., value about thousand times larger than the elementary simulation step. Disadvantage of both temperature scaling and Berendsen thermostat methods is that they do not generate canonical ensemble. The examples of methods that generate canonical ensemble are Anderson (1980) and Nose-Hoover (1984, 1985) thermostat. Their description can be found in example of the book of Frenkel and Smith (2002).

Standard laboratory conditions correspond more closely to the NPT (isobaric, isothermal) ensemble; therefore, in many cases, we also need to control the pressure of the system. The pressure of the system at constant N, V, and T can be expressed by the virial equation. Assuming the pairwise additive interactions, it can be written (see, e.g., Hansen and McDonald 1986) as

$$P = \frac{1}{V}\left[Nk_{\text{B}}T + \frac{1}{d}\left\langle \sum_{i<j} \vec{F}\left(\vec{r_{ij}}\right)\vec{r_{ij}} \right\rangle\right] \tag{15.33}$$

where d is dimensionality of the system, $\vec{r_{ij}}$ is relative position vector, and $\vec{F}\left(\vec{r_{ij}}\right)$ is the force between particles i and j. The constant pressure of the system can be maintained by the change of volume of simulated box by scaling factor λ. Methods for the pressure control are very similar to those for the temperature control (Eqs. 15.28 or 15.32). Typical value of the coupling constant for the Berendsen barostat is about 0.5 ps. The only qualitative difference in temperature and pressure coupling is the possibility to apply the scaling factor either isotropically, i.e., with the

same scaling factor in all three dimensions of the simulated box or anisotropically, where scaling factor is calculated for each direction independently. Anisotropic pressure scaling is suitable for systems with preferential orientation, e.g., oriented fibrils, liquid crystals, and biological membranes.

15.6 Enhanced Sampling Methods

An experiment is usually carried out on a macroscopic sample containing a huge number of interacting atoms and molecules which sample enormous number of *conformations*. To obtain meaningful results from computer simulations in agreement with the experiment data, one has to involve all possible conformations (or, strictly speaking, *microstates*) and the system may theoretically occupy with a certain probability. Note that one particular *configuration*, i.e., positions of all atoms at a given time along with their instant momenta represent one *phase state*, a point within a multidimensional *phase space*. Indeed, in statistical mechanics, the theory underlying molecular dynamics and Monte Carlo simulations, the *ensemble averages* correspond to the experimental observables of a macrostate that we are interested in like chemical shifts, fluorescence spectra, free energies of dissociation, catalytic reaction rates, temperature, etc. An ensemble average is defined as the mean of a quantity that is a function of the individual microstates of the system. Even though each microscopic state can be different at the atomic level, the ensemble average calculated over all possible microstates corresponding to a particular macrostate represents a stable characterization of the system. The *ensemble average* may be written as

$$\langle X \rangle_{\text{ens}} = \sum_{r^N, p^N} X\left(r^N, p^N\right) \rho\left(r^N, p^N\right) \tag{15.34}$$

where $X(r^N, p^N)$ stands for the observable of interest expressed as a function of positions r^N and momenta $p^N$of N particles, in the case of classical MD simulations of all the atoms system is composed of. Note that the summation runs over all possible values of the variables r^N and p^N, or, in other words, over all possible system states. The second term in Eq. (15.34) is the *ensemble probability density* reflecting the intuitive fact that each microstate of the system is adopted with different probability. The ensemble probability density reads as

$$\rho\left(r^N, p^N\right) = \frac{1}{Z} \mathrm{e}^{\frac{-E\left(r^N, p^N\right)}{k_{\mathrm{B}} T}} \tag{15.35}$$

where $E(r^N, p^N)$ is the total energy of the system including kinetic and potential energy as a function of positions and momenta, T is the absolute temperature of the system, k_{B} is the Boltzmann constant, and Z is the *partition function* defined as

$$Z = \sum_{r^N, p^N} \mathrm{e}^{\frac{-H(r^N, p^N)}{k_\mathrm{B} T}} \tag{15.36}$$

The main bottleneck of the statistical mechanics application is the calculation of the partition function that is computationally extremely challenging even for small real-life systems. One can rigorously (analytically) derive the partition function only for the simplest "toy models" like *harmonic oscillator*, a theoretical system composed of two vibrating particles bound by a simple parabolic potential, or a *rigid rotor*, the simplest model system of rotating diatomic molecules. For more complex systems like those of solvated biomacromolecules, the partition function summation running over all positions and momenta needs to be calculated with lower accuracy and using numerical mathematical methods. The difficulty in evaluating the partition function lies in the fact that one has calculated all system microstates along with their respective energies. Note that up to this point, no approximations have been in the derivation of Eqs. (15.34, 15.35, and 15.36). Therefore, provided that we are able to generate all possible microstates (for the sake of simplicity you can think of protein conformations) of the system and compute their energies with ultimate accuracy, we would have everything at our disposal to calculate the desired partition function out of which various thermodynamic properties can be derived. Basically, the quality of the calculated physical observables as measured by the agreement with the experimental data is determined by two aspects: (i) the accuracy with which we are able to calculate the total energy of the system (especially the many-body potential energy of the interacting atoms), and (ii) the number of microstates we take into account in our calculation. The enhancement of the energy calculations accuracy implies to utilize either more elaborate force fields covering many-body effects inherently neglected in contemporary simplified pair-wise additive force fields or even more rigorous and computationally demanding quantum mechanics-based techniques, which is beyond the scope of this chapter. To improve the later aspect, one has to generate as much microstates of the system as possible to sample the whole phase space. As the weight of each microstate $w(i)$ decays exponentially with its energy, i.e.,

$$w(i) \sim \mathrm{e}^{\frac{-E_i}{k_\mathrm{B} T}} \tag{15.37}$$

We aim to include primarily the low-energy states with higher probability of occurrence. On the contrary, the high-energy (i.e., unstable) states are significantly less probable to occur, and hence, involvement of these has a rather negligible impact on the ensemble averages. As a result, we do not aim to consider all imaginable states of the molecular system, which would be, however, mathematically correct but practically intractable, but to include as many states as possible especially within the energy low-lying regions. Even though the restriction to the low-energy states makes the calculation significantly more simple, exploration of the

complex multidimensional conformation space and location of all respective domains remain far from being straightforward.

In MD simulations, the microstates of the ensemble are calculated sequentially in time that is called *trajectory*, so to evaluate an ensemble average, the simulation must pass through all possible states corresponding to the specific thermodynamic constraints. Another way, typical for MD simulations, is to determine the *temporal average* of the observable, $\langle X \rangle_{\text{time}}$, instead of the ensemble average $\langle X \rangle_{\text{ens}}$:

$$\langle X \rangle_{\text{time}} = \frac{1}{N} \sum_{i=0}^{N-1} X\left(r^N(t_i), p^N(t_i)\right) \tag{15.38}$$

where $X(r^N(t_i), p^N(t_i))$ is the value of the observable X at a time step i expressed as a function of positions r^N and momenta $p^N$of the atoms. The dilemma appears to be that one can calculate time averages using molecular dynamics simulation, but the experimental observables are assumed to be ensemble averages as outlined above. Resolving this leads us to one of the most fundamental axioms of statistical mechanics, the *ergodic hypothesis,* which states that the time average equals the ensemble average in case of infinite simulations:

$$\langle X \rangle_{ens} = \langle X \rangle_{\text{time}} \tag{15.39}$$

The basic idea is that if one allows the system to evolve in time *indefinitely* that system will eventually pass through all possible states. One goal, therefore, of a molecular dynamics simulation is to generate enough representative conformations such that this equality is satisfied. If this is the case, experimentally relevant information concerning structural, dynamic, and thermodynamic properties may then be calculated using a feasible amount of computer resources. Because the simulations are of fixed duration, one must be certain to sample a sufficient amount of phase space.

15.6.1 Sampling Problem

As outlined in Sect. 15.5, the sense of MD simulations is to obtain positions and momenta or velocities of the interacting atoms at the present time t based on these quantities one-time increment in the past, i.e., at time $t - \Delta t$. The main difficulty of the MD simulations limiting their ability to reproduce experimental data is that the MD trajectories are *not ergodic* and leave many relevant regions of the conformation space unexplored. This stems mainly from the separation of the high-probability meta-stable regions by low probability, i.e., high-energy transition areas and the inherent difficulty of sampling a multidimensional configuration space by embedding it into a one-dimensional MD trajectory. Indeed, biological systems often have many local minima separated by high-energy barriers, and this can limit complete

sampling as simulations can get stuck in local energy minima. This is of particular concern for protein folding and protein docking simulations.

To overcome the sampling problem, various elaborate techniques improving exploration of the configuration space have been devised. The main concern of any enhanced sampling methods is guaranteeing that the statistical weights of the microstates along the trajectory are known and correct (or at least correctable) while simultaneously ensuring that as much of the relevant regions of the configuration space are sampled. In the following paragraphs, we will comment on the most popular enhanced sampling methods that we will divide into three groups:

(i) *Collective variable biasing methods*
(ii) *Tempering and simulated annealing methods*
(iii) *Hybrid methods*

15.6.2 Collective Variable Biasing Methods

- Umbrella sampling

Umbrella sampling attempts to overcome the sampling problem by modifying the potential function to enhance the sampling of events hindered by large free energy barriers. Umbrella sampling was developed by Torrie and Valleau (1974, 1977) based on related previous work of McDonald and Singer (1967, 1969). A bias, an additional potential energy term, is applied to the system to ensure efficient sampling along the *reaction coordinate,* a path connecting two end-point states separated by an energy barrier, e.g., bound and unbound states of a dimeric protein. This can either be aimed at in one simulation or in different simulations (windows), the distributions of which energetically overlap. The effect of the bias potential to connect energetically separated regions in phase space gave rise to the name *umbrella sampling* (Abrams and Bussi 2014).

The bias potential w_i of window i is an additional potential energy term, which depends on the reaction coordinate only. Ideally, the bias potential is chosen such that sampling along the whole range of the reaction coordinate is uniform. Therefore, the optimal bias potential is the negative value of the actual (unbiased) free energy profile, which would yield a truly uniform distribution. However, the actual free energy profile is obviously not known, it is what we aim to calculate with umbrella sampling (Kaestner 2011).

MD simulation of the biased system provides the biased (unrealistic) distribution along the reaction coordinate. Under certain conditions, discussion of which is beyond the scope of this chapter, one can derive the global unbiased distributions along the reaction coordinate out of which respective free energy profile as a function of the reaction coordinate can be calculated, the property called potential of mean force (PMF). Series of umbrella sampling simulations can be analyzed using the weighted histogram analysis method (WHAM) (Kumar et al. 1992), which determines optimal weights for combining the simulations.

- Metadynamics and Local Elevation

The aim of metadynamics (Laio and Parrinello 2002) and local elevation (Huber et al. 1994) techniques is to add "memory" into the sampling process, thus, preventing oversampling of local energy minima. This is made possible by adding an external history-dependent bias potential that is a function of the selected collective variables to the Hamiltonian of the system. This potential can be written as a sum of Gaussians deposited along the system trajectory in the collective variables space to discourage the system from revisiting configurations that have already been sampled in previous time steps. In other words, the effect of the metadynamics bias potential is to push the system away from local minima into visiting new regions of its phase space. Furthermore, in the long time limit, the bias potential converges to minus the free energy as a function of the selected collective variables (Abrams and Bussi 2014).

We imagine a simple model system having two energy minima named A and B that are separated by an energy barrier. In a standard MD simulation, the system would remain stuck in the initial minimum (A) because the barrier is larger than the thermal fluctuations at the given temperature (Barducci et al. 2001). Instead, in metadynamics simulations, Gaussians are deposited causing the underlying bias potential to grow, until eventually the system is pushed out of the basin A into the second energy basin B. The natural and more convenient escape route is to pass the lowest barrier and fall into the second basin. Here, the Gaussians accumulation starts again. The system is trapped in B until the underlying free energy basin is completely filled. At this point, the system freely diffuses in the region between the first two minima resembling a random walk on the flattened free energy surface. At this point, the underlying free energy surface can be reconstructed based on the summation of the deposited biasing Gaussians (Spiwok et al. 2015).

15.6.3 Tempering and Simulated Annealing Methods

- Parallel tempering replica-exchange

Significant improvement of sampling can be achieved at elevated temperatures. Intuitively, systems at higher temperatures can overcome energy barriers easily, compared with a situation at low temperatures. However, temperature does not change only free energy barriers, but it also changes free energy values of individual minima. For example, the native structure of a protein represents its global free energy minimum at biological temperatures. However, the unfolded structure becomes the global free energy minimum at temperatures above protein's melting point.

Fortunately, there is a way to exploit higher temperatures to enhance sampling at biological temperatures, which is represented by parallel tempering method. The method has been first introduced as a variant of Monte Carlo method (Swendsen and Wang 1986). Wider application on biological systems has been made possible by

development of parallel tempering molecular dynamics simulation (Sugita and Okamoto 1999). The parallel tempering scheme starts with choosing the temperatures. Values of temperature must cover a wide range including the biologically relevant ones as well as temperatures when sampling is significantly enhanced. In parallel tempering, the studied system is simulated in parallel replicas, each at one of the selected temperatures (Abrams and Bussi 2014). At certain time intervals, for example, every picosecond, energies of neighboring replicas are compared. Obviously, a replica simulated at a lower temperature tends to have lower potential energy than the one simulated at higher temperature. However, as temperature differences between replicas are small, it may happen that the potential energy of the high-temperature replica is lower. In this case, coordinates of replicas are exchanged and simulation of low temperature continues at higher temperature and vice versa (probability of exchange is 1). If not, replica exchange probability is calculated from the potential energy difference as

$$P = \exp\left[\left(E_i - E_j\right)\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)\right] \tag{15.40}$$

where E_i and E_j are energies of the i-th and j-th replica, respectively. Similarly, T_i and T_j are respective temperatures of the i-th and j-th replica (Spiwok et al. 2015). The computer then generates a random number in the range from 0 to 1. If this number is lower than the calculated probability, replicas are exchanged and the simulation continues. At the end of the simulation, it is possible to collect all coordinate snapshots at a certain temperature. The exchange criterion is designed in the way that such collected snapshots sample the studied system canonically at a given temperature. This means that it is possible to calculate probabilities of different conformational families and the free energy surface, not only for the biological temperature. All of the independent simulations should experience at least one variable exchange event, so will be discontinuous but will follow a proper Boltzmann distribution.

- Simulated annealing

Simulated annealing is method suitable for cases, where many thermal accessible minima are present. In such cases, gradient methods often lead to the system in local instead of global minima. The method of simulated annealing has its analogy in physics, respectively, in a technology process of a production of large crystals. The basic feature is its slow controlled annealing, during which system should relax to the ideal crystal, which corresponds to the global minimum of free energy. Simulated annealing (Kirkpatrick et al. 1983) is computational method, which mimics this process with the aim to search for optimal solution (e.g., global minimum of energy) for the systems with large number of freedom. Simulated annealing is successfully used for optimization of biomolecular system and their complexes with ligands. Simulated annealing utilizes the fact that according to the Boltzmann distribution, states with lower energy are higher populated compared with states with the lower energy. At zero temperature, the system should occupy only the state with minimum

energy. In practice, simulated annealing is realized by slow decrease of temperature during which the thermal equilibrium is obtained at each temperature. Simulated annealing is often used in combination with molecular dynamics described in the previous sections. At the initialization, the initial temperature is set to sufficiently high temperature (typically up to 1000 K). This temperature corresponds to the temperature of thermostat (e.g., Berendsen thermostat), and in the next steps, the system is thermal equilibrated. In the next step, temperature of the thermostat will be slightly decreased, and again molecular dynamics will be applied for obtaining the thermal equilibrium. This process is repeated until the required temperature is obtained (e.g., 0 K).

15.6.4 Hybrid Methods

The description of methods outlined above indicates that different methods are suitable for studying different systems and processes. For example, metadynamics makes it possible to precisely model certain process defined by collective variables, for example, a small conformational change in the active site of a large enzyme. Unfortunately, sampling of only few CVs can be actively enhanced (Abrams and Bussi 2014). On the other hand, parallel tempering enhances sampling of all degrees of freedom through elevated temperatures. Choice of collective variables before the start of the simulation, which may be highly nontrivial, is not necessary. Unfortunately, the size of the simulated system is significantly limited because a huge number of replicas must be used for large systems. There are, however, several ways of how to combine advantages of both approaches.

- Hamiltonian replica-exchange method

The replica-exchange molecular dynamics methods (REMD) have emerged as one of the most widely used techniques to enhance conformational sampling. The idea of the REMD group of methods is to simulate several copies called replicas of a given molecular system in parallel under different conditions. At predefined time intervals, neighboring pairs of replicas are exchanged with a specific transition probability. The REMD methods basically differ in the chosen parameter that changes among the simulated replicas. In many REMD studies, the simulation temperature represents the parameter (see parallel tempering methods) (Abrams and Bussi 2014). As the temperature-based REMD method suffers from the need of high numbers of replicas to be efficient, the Hamiltonian replica exchange (H-REMD) is often method of choice (Meli and Colombo 2013). The H-REMD methods is grounded on the consideration that, because the different parallel simulations do not interact (i.e., independent simulations), there is no need to use the same Hamiltonian for all of the replicas. In H-REMD, the different replicas are usually simulated at a constant temperature, whereas the Hamiltonian of the system is used as the replica coordinate. Although the standard part of the Hamiltonian (the kinetic and the potential energy) is usually common for all simulated replicas,

the additional bias term is replica dependent. The switching probability α_{ij} is determined based on the energy difference following the Metropolis criterion:

$$\alpha_{ij} = \min\left\{1, \mathrm{e}^{-\frac{H_i(q_j)+H_i(q_i)}{k_\mathrm{B}T_i}+\frac{-H_j(q_i)+H_j(q_j)}{k_\mathrm{B}T_j}}\right\} \tag{15.41}$$

where Hamiltonian H is defined as the sum of the force field and the bias potential. There are various ways of how to modify the Hamiltonian. For protein–protein interactions, it is often advantageous to complement the standard Hamiltonian with a distance restrain potential (Oostenbrink et al. 2012; de Ruiter and Oostenbrink 2013). To overcome energy barriers, it possible to use variable soft-core potentials for the interactions that contribute most to the energy barriers (Hritz and Oostenbrink 2008). It should be stressed, however, that the setup of the replica-coordinate parameters may be nontrivial, and it is often necessary to go through trial and error tuning process. Another possibility is to utilize some kind of optimization algorithm to get most appropriate parameter values (Hritz and Oostenbrink 2007).

- Parallel tempering-metadynamics

Metadynamics and parallel tempering were successfully combined in a parallel tempering metadynamics (Bussi et al. 2006). The system under study is simulated using multiple replicas at different temperatures in line with the classical parallel tempering. Replica exchanges are attempted in regular intervals with an exchange criterion similar to the one given in Eq. (15.7). However, this criterion contains the potential energy as well as the bias potential. Free energy surfaces can be calculated for different temperatures as a negative imprint of the bias potential.

- Bias-exchange metadynamics

Another combination of metadynamics and replica exchange methods is the technique called bias exchange metadynamics (Piana and Laio 2007). The system is again simulated in multiple replicas. Simulation in each replica is biased by a one-dimensional bias potential with a single collective variable. The number of collective variables is, therefore, equivalent to the number of replicas (an additional "neutral" replica without bias potential may be also introduced). Replicas can occasionally swap their coordinates, i.e., the system biased along one collective variable becomes biased along another CV. The criterion for replica exchange is calculated from the bias potentials. The method makes it possible to calculate a series of one-dimensional free energy surfaces for each CV.

15.7 Analysis and Validation of Generated Computational Data

The molecular dynamics simulation is in the many aspects very similar to the real experiment. It is the reason, why it is often called as computational experiment or experiment in the white and why the calculation of averages of properties of our

interest is very similar. In the case of experimentally measured value of the property A, it is the average value for the duration of experiment. In principle, the real average value would be gained in the measurement in the limit of infinitively time duration:

$$A_{\text{average}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} \vec{A} \left(\vec{r^N}(t), \vec{p^N}(t) \right) dt \tag{15.42}$$

We know from the experience that for the determination of the average value with sufficient precision, measurement with finite time duration is sufficient (this time has to be much longer than the relaxation time of the measured quantities). Similarly, in the MD simulation with the high enough number of steps M, we can "measure" many properties of a system as the average value of discrete values in the individual steps:

$$A_{\text{average}} = \frac{1}{M} \sum_{i=1}^{M} A\left(\vec{r^N}_i, \vec{p^N}_i \right) \tag{15.43}$$

For example, the internal energy U is possible to calculate as the average of the energies in the individual steps in simulation:

$$U = E_{\text{average}} = \frac{1}{M} \sum_{i=1}^{M} E_i\left(\vec{r^N}_i, \vec{p^N}_i \right) = \frac{1}{M} \sum_{i=1}^{M} \left[V_i\left(\vec{r^N}_i \right) + \sum_{j=1}^{N} \frac{\vec{p_{ij}}^{\,2}}{2m_i} \right] \tag{15.44}$$

Another important property often calculated in biomolecular simulations is the root mean square deviation (RMSD) of atoms in molecule at the time t_1 with respect to the reference structure at the time t_2

$$\text{RMSD}(t_1, t_2) = \left[\frac{1}{N} \sum_{i=1}^{N} \left\| \vec{r_i}\,(t_1) - \vec{r_i}\,(t_2) \right\|^2 \right]^{\frac{1}{2}} \tag{15.45}$$

where $\vec{r_i}\,(t)$ is the position of atom i at time t.

We already described in Sect. 15.5.5 that the concept of ensemble averages is used also for the determination of temperature (15.26) and pressure (15.33) of the simulated systems. It should be noted that to obtain data that can be compared with the experiment, one needs to ensure that the MD simulation setup (e.g., choice of thermostat and/or barostat, etc.) would provide desired ensemble. For example, the simple velocity rescaling algorithm maintaining the desired system temperature does not generate correct canonical ensemble.

Acknowledgments The financial contribution made by the Ministry of Education, Youths and Sports of the Czech Republic within special support paid from the National Programme for Sustainability II funds, project CEITEC 2020 (LQ1601), is gratefully acknowledged.

References

Abrams C, Bussi G (2014) Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. Entropy 16:163–199

Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 72:2384

Barducci A, Bonomi M, Parrinello M (2001) Metadynamics. WILEY Int Rev Comput Sci 1:826–843

Berendsen HJC, Postma JPM, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Burkert U, Allinger NL (1982) Molecular mechanics. In: ACS monograph, vol 177. American Chemical Society, Washington, DC

Bussi G, Gervasio FL, Laio A, Parrinello M (2006) Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. J Am Chem Soc 128:13435–13441

Dauber-Osguthorpe P, Robert VA, Osguthorpe DJ, Hagler AT (1988) Structure and energetics of ligand-binding to proteins - Escherichia-coli dihydrofolate reductase trimethoprim, a drug-receptor system. Protein Struct Funct Genet 4:31–47

de Ruiter A, Oostenbrink C (2013) Protein–ligand binding from distancefield distances and hamiltonian replica exchange simulations. J Chem Theory Comput 9:883–892

Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications. Academic Press, Boston

Hansen JP, McDonald IR (1986) Theory of simple liquids, 2nd edn. Academic Press, London

Hermans J, Berendsen HJC, van Gunsteren WF, Postma JPM (1984) A consistent empirical potential for water-protein interactions. Biopolymers 23:1513–1518

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472

Hoover WG (1985) Canonical dynamics: equilibrium phase-space distributions. Phys Rev A 31:1695–1697

Hritz J, Oostenbrink C (2007) Optimization of replica exchange molecular dynamics by fast mimicking. J Chem Phys 127:204104

Hritz J, Oostenbrink C (2008) Hamiltonian replica exchange molecular dynamics using soft-core interactions. J Chem Phys 128:144121

Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. J Comput Aided Mol Des 8:695–708

Kaestner J (2011) Umbrella sampling. WILEY Int Rev Comput Sci 1:932–942

Kaminski G, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474–6487

Kirkpatrick C, Gelatt D Jr, Vecchi MP (1983) Optimalization by simulated annealing. Science 220:671–680

Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13:1011–1021

Laio A, Parrinello M (2002) Escaping free energy minima. Proc Natl Acad Sci U S A 99:12562–12566

Lautz J, Kessler H, Kaptein R, van Gunsteren WF (1987) Molecular dynamics simulations of cyclosporin A: the crystal structure and dynamic modelling of a structure in Apolar solution based on NMR data. J Comput Aided Mol Des 1:219–241

Lavery R, Rivail J-L (1991) In: Smith J (ed) American Institute of Physics (A.I.P.) conference proceedings 1991, vol 239. New York, pp 131–146

MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith

JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616

McDonald IR, Singer K (1967) Machine calculation of thermodynamic properties of a simple fluid at supercritical temperatures. J Chem Phys 47:4766–4772

McDonald IR, Singer K (1969) Examination of the adequacy of the 12–6 potential for liquid argon by means of Monte Carlo calculations. J Chem Phys 50:2308–2315

Meli M, Colombo G (2013) A Hamiltonian replica exchange molecular dynamics (MD) method for the study of folding, based on the analysis of the stabilization determinants of proteins. Int J Mol Sci 14:12157–12169

Nosé S (1984) A unified formulation of the constant temperature molecular-dynamics methods. J Chem Phys 81:511–519

Oostenbrink C, de Ruiter A, Hritz J, Vermeulen NPE (2012) Malleability and versatility of Cytochrome P450 active sites studied by molecular simulations. Curr Drug Metab 13:190–196

Piana S, Laio A (2007) A bias-exchange approach to protein folding. J Phys Chem B 111:4553–4559

Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341

Spiwok V, Sucur Z, Hosek P (2015) Enhanced sampling techniques in biomolecular simulations. Biotechnol Adv 33:1130–1140

Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151

Swendsen RH, Wang JS (1986) Replica Monte Carlo simulation of spin glasses. Phys Rev Lett 57:2607–2609

Torrie GM, Valleau JP (1974) Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid. Chem Phys Lett 28:578–581

Torrie GM, Valleau JP (1977) Non-physical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. J Comput Phys 23:187–199

van Gunsteren WF, Berendsen HJC (1984) Computer simulation as a tool for tracing the comformational differences between proteins in solution and in the crystalline state. J Mol Biol 176:559–564

van Gunsteren WF, Berendsen HJC (1987) Gromos-87 manual. Biomos BV, Groningen

van Gunsteren WF (1991) Computer simulation of biomolecular systems: overview of timesaving techniques. In: Advances in biomolecular simulations. American Institute of Physics (A.I.P.), New York

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784

Weiner SJ, Kollman PA, Nguyen DT et al (1986) An all atom force-field for simulations of proteins and nucleic-acids. J Comput Chem 7:230–252