Plant Pattern-Recognition Receptor Activation and Exploration of Techniques to Engineer Novel Ligand Specificities

by

Teresa Koller

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Plant Pathology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2014

Date of final oral examination: 6/30/2014

The dissertation is approved by the following members of the Final Oral Committee:

Andrew Bent, Professor, Plant Pathology

Caitilyn Allen, Professor, Plant Pathology

Patrick Krysan, Professor, Horticulture

Jean-Michel Ané, Professor, Agronomy

Sara Patterson, Professor, Horticulture

Abstract

Pattern-recognition receptors (PRRs) localized at the plasma membrane of plant cells monitor the apoplastic space for pathogen-associated molecular patterns (PAMPs), damage-associated molecular patterns (DAMPs) and apoplastic effectors. In my thesis research I studied different aspects of the two Arabidopsis PRRs EFR and FLS2. Both EFR and FLS2 are leucine-rich repeat (LRR) receptor-like kinases (RLKs). First I cloned, sequenced and functionally tested in chimeric receptors the LRR-encoding domains of nine novel *EFR* alleles from seven Brassicaceae species. The sequences were subjected to Repeat Conservation Mapping (RCM), a computational program developed in our lab that identifies conserved patches on the predicted surface of LRR domains. This analysis revealed conserved clusters in the LRR domain of EFR, which are candidate functionally important sites. Additional work used RCM to compare the last seven LRRs of the BAK1-interacting LRR-RLKs EFR, FLS2, BRI1 and PEPR1, and revealed a conserved putative BAK1 ectodomain interaction site. However, site-directed mutagenesis of the putative BAK1 interaction site in the FLS2 ectodomain did not disrupt interaction with BAK1 and analogous mutagenesis in EFR resulted in receptor maturation defects. Hence this conserved LRR C-terminal region apparently has functions other than mediating interactions with BAK1. I then performed *In vivo* tests of the newly published FLS2-BAK1-flg22 ectodomain co-crystal structure, with the resulting data supporting models in which FLS2 signaling is initiated by formation of a ligand mediated FLS2 – BAK1 ectodomain complex.

In an effort to mold our constantly growing knowledge about the plant immune system into agriculturally useful tools to improve crop health, I also explored two approaches to

engineer novel PRRs. The first approach sought to use yeast surface display to evolve the endogenous LRR domains of EFR and FLS2 towards novel ligand binding capacities. The second approach used a LRR-based variable lymphocyte receptor (VLR) of *Petromyzon marinus* (sea lamprey, a jawless vertebrate) as template for rational design of novel ligand binding sites in FLS2. Both approaches are compromised by the subsequently discovered finding that PRR activation requires a SERK protein ectodomain as a co-receptor. The research projects presented here further elucidate PRR signaling activation and present possible technologies for the ambitious goal of engineering novel PRRs to improve crop health.

Acknowledgments

Foremost, I would like to thank Dr. Andrew Bent for giving me the opportunity to work in his group and for excellent guidance and support throughout the years. I would also like to thank the other members of my committee, Drs. Caitilyn Allen, Patrick Krysan, Jean-Michel Ané and Sara Patterson for their valuable input and feedback. I would like to thank all the past and present members of the Bent lab for their help, friendship and useful discussions. Especially I would like to thank Yangrong Cao, Xiaoli Guo, Junqi Song and Yun (Monica) Chen for sharing their expertise in molecular lab techniques. I would like to thank the faculty, staff and graduate students of the Plant Pathology Department for being supportive and creating a great work environment. I enjoyed very much being a part of such a great scientific community. I would like to thank Drs. Eric Shusta and Thomas Mallot from the Chemical and Biological Engineering Department for sharing their expertise in yeast surface display and flow cytometry. I very much appreciated Thomas's patience in answering my questions and his help with my experiments. I would like to thank my family - especially my parents, sister and brother -, my friends in Madison and Switzerland and my boyfriend Corey for their love, friendship and support.

Table of contents

Chapter 1: Introduction

The plant immune system (current paradigms)

We live in an exciting time as the molecular mechanisms of the plant immune system are being unraveled (Jones and Dangl, 2006; Bent and Mackey, 2007, Boller and Felix, 2009; Dodds and Rathjen, 2010; Takken and Goverse, 2012; Wu and Zhou, 2013; Macho and Zipfel, 2014). The new paradigm now describes how a central aspect of the plant immune system is arranged in distinct layers. The first layer consists of pattern-recognition receptors (PRRs) localized at the plasma membrane. These receptors detect pathogen-associated molecular patterns (PAMPs), damage-associated molecular patterns (DAMPs) and pathogen effectors in the plant apoplast (Figure 1A). In terms of structure, the PRRs are receptor-like kinases (RLKs) or receptor-like proteins (RLPs). Both receptor types consist of an extracellular domain for ligand perception and a transmembrane domain, but only the RLKs have an intracellular kinase domain. Depending on the type of extracellular domain, these RLKs and RLPs can further be classified. Over 21 different classes of ectodomains of RLKs have been found in Arabidopsis (Becraft, 2002). The most common type is the leucine-rich repeat (LRR) domain. The Arabidopsis genome encodes 239 LRR-RLKs and the rice genome 309 LRR-RLKs, respectively (Shiu and Bleecker, 2001; Sun and Wang, 2011). Besides LRR domains, common ectodomains of PRRs include the lysine motif (LysM) domain and the epidermal growth factor (EGF)-like domain.

The second layer in this part of the plant immune system consists of receptors localized in the cytoplasm of the plant cells. These receptors directly or indirectly detect effector proteins that are delivered into the plant cell by pathogenic microbes. The cytoplasmic receptors are

encoded by R genes. Early studies on the plant immune system led to the gene-for-gene hypothesis that recognized the high specificity between the products of R genes from plants and avirulence genes from pathogens, resulting from co-evolution of specific microbial strains and specific plant cultivars. We now know that avirulence genes encode pathogen virulence effector proteins, and that some of the characterized effectors target plant PRR pathway signaling molecules. However, the majority of effector targets are still unknown. Some effector – R protein interactions are indirect. In those cases the R proteins guard endogenous molecules in the plant cell and trigger a defense response upon disturbance of the guarded molecule. In terms of structure, most R proteins known to date carry nucleotide-binding and leucine-rich repeat domains, and are often referred to as NB-LRR proteins.

Further basics regarding these aspects of the plant immune system are reviewed in (Jones and Dangl, 2006; Bent and Mackey, 2007, Boller and Felix, 2009; Dodds and Rathjen, 2010; Takken and Goverse, 2012; Wu and Zhou, 2013; Macho and Zipfel, 2014).

Pattern-recognition receptors (PRRs)

Pattern-recognition receptors are localized at the plasma membrane and are RLK and RLP types of proteins. The Arabidopsis genome encodes more than 600 RLKs and 56 RLPs and the rice genome encodes more than 1000 RLKs and 90 RLPs, respectively (Shiu and Bleecker, 2003; Shiu et al., 2004; Dardick et al., 2007; Fritz-Laylin et al., 2005). Not all RLKs and RLPs are PRRs. Some are involved in the binding and perception of plant hormones and other

endogenous cell to cell signaling molecules. Others are regulatory proteins; thus they do not necessarily interact with a ligand but are important facilitators or suppressors of signaling activation and furthermore allow signaling cross-talk at the plasma membrane, coordinating the different signals perceived in the apoplast to ensure the proper downstream signaling. Important examples of these regulatory RLKs include the SERK family members (Chinchilla et al., 2009; Kim et al., 2013), BIR family members (Gao et al., 2009; Halter et al., 2013) and SOBIR1 (Liebrand et al., 2013).

Considering the large number of RLK genes and RLP genes found in different plant genomes, we are only at the very beginning of assigning functions to all of them. So far only a handful of PRRs have been characterized and a few ligands have been identified. PRRs characterized in **Arabidopsis** include: FLS2, EFR, CERK1, LYM1/LYM3, WAK1, ReMAX, PEPR1/2 and LecRK-1.9 (Gómez-Gómez and Boller, 2000; Bauer et al., 2001; Kunze et al., 2004; Zipfel et al., 2006; Miya et al., 2007; Wan et al., 2008; Willmann et al., 2011; Brutus et al., 2010; Jehle et al., 2013; Huffaker et al., 2006; Yamaguchi et al., 2010; Huffaker and Ryan, 2007; Choi et al., 2014; Senchou et al., 2004). PRRs characterized so far in **rice** include: Xa21, Xa26, OsFLS2, CEBiP, OsCERK1, LYP4/6 and OsWARK1 (Kaku et al., 2006; Shimizu et al., 2010; Liu, Li, et al., 2012; Song et al., 1995; Sun et al., 2004; Li et al., 2009; Takai et al., 2008). PRRs characterized so far in **tomato** include: LeEix1/2, Ve1, LeFLS2 and several Cfs (Ron and Avni, 2004; de Jonge et al., 2012; Joosten et al., 1997; Van den Ackerveken et al., 1992; Bar et al., 2010; Robatzek et al., 2007).

Many questions remain about how PRRs are activated and regulated and how the downstream signaling pathways are coordinated. These questions are currently being

addressed by scientists around the world. There are examples of homodimerization, heterodimerization and multimerization of RLKs and RLPs for signaling activation and initiation of downstream signaling upon ligand detection. Hetero- and multimerization is especially important for RLPs lacking an intracellular kinase domain. So far the RLP - RLK complexes OsCEBiP – OsCERK1 (Shimizu et al., 2010), LeSOBIR1 – Cf-4 and LeSOBIR1 – Ve1 (Liebrand et al., 2013) have been identified. Several PRRs of the LRR-RLK type have been found to interact with the SERK family LRR-RLKs to initiated downstream signaling (Figure 1B). So far the following RLK – RLK signaling initiation complexes have been identified: FLS2 – BAK1, EFR – BAK1, PEPR1/2 – BAK1 and Ve1 – BAK1 (Chinchilla et al., 2007; Heese et al., 2007; Schulze et al., 2010; Roux et al., 2011; Sun et al., 2013; Oh et al., 2010; Postel et al., 2010). LeEix1 interacts with BAK1 as well but in this case BAK1 attenuates LeEix1 signaling (Bar et al., 2010). Receptor activation through homodimerization was observed for the Arabidopsis LysM-RLK CERK1. Two AtCERK1s directly bind to one chitin molecule, resulting in a CERK1 – chitin – CERK1 signaling complex (Liu et al., 2012).

Receptor complex formation results in phosphorylation of the intracellular kinase domains of the interacting RLKs and furthermore phosphorylation of receptor-like cytoplasmic kinases (RLCKs). RLCKs play essential roles in PRR downstream signaling regulation (reviewd in Lin et al., 2013). These intracellular kinases lack extracellular domains. Some RLCKs are anchored to the plasma membrane, others are associated with the kinase domains of RLKs but can be release into the cytoplasm. 147 genes encoding RLCKs have been identified in the Arabidopsis genome and 379 in rice, respectively (Vij et al., 2008; Shiu and Bleecker, 2001). BIK1 and its homologues PBS1, PBL1 and PBL2 are well-studied RLCKs that associate with the kinase

domains of BAK1, FLS2, EFR, CERK1 and PEPR1 before PRR activation (Lu et al., 2010; Zhang et al., 2010; Liu et al., 2013). After receptor complex activation and phosphorylation of the kinase domains, BIK1 is released from the complex (Zhang et al., 2010). The precise roles of RLCKs and other types of molecules in downstream signaling are a current field of study.

Activated PRRs trigger a plethora of immune responses. Some responses are observable within seconds or minutes after pathogen exposure and others develop within days after exposure. Early molecular responses include Ca^{2+} (and other ion) influxes, MAPK phosphorylation, reactive oxygen species (ROS) burst, activation of transcription factors and subsequent expression or down regulation of hundreds of genes. An increased production of the defense-related hormones ethylene and salicylic acid is observable within minutes after pathogen exposure. Closure of stomata and callose deposition can be observed at the sites of pathogen infection (reviewed in Boller & Felix, 2009).

FLS2 and EFR ligand specificity

FLS2 and EFR both are LRR-RLK types of PRRs and both were first discovered in Arabidopsis. FLS2 detects the epitope flg22 of the bacterial flagellin (Felix et al., 1999; Bauer et al., 2001) (Figure 3A+B). The original flg22 peptide used to study FLS2 was derived from the flagellin sequence of *Pseudomonas aeruginosa (Psa)*(Figure 3E). Many bacterial species have a flg22 sequence similar or identical to *Psa*-flg22 and peptides derived from these sequences trigger similar responses in Arabidopsis as the *Psa*-flg22. However peptides derived from the

slightly altered flg22 sequences of *Ralstonia solanacearum* and *Xanthomonas campestris campestris* strain 186 do not trigger a defense response in Arabidopsis (Sun et al., 2006; Pfund et al., 2004).

FLS2 homologues have been identified in many different monocot and dicot plant species (Boller and Felix, 2009). Among different *Arabidopsis thaliana* genotypes and also among Brassicaceae relatives the extent of the immune response triggered by the flg22 peptide varies. It has been suggested that this is primarily due to variation in FLS2 protein abundance (Vetter et al., 2012).

FLS2 from Arabidopsis (AtFLS2) and FLS2 from tomato (LeFLS2) share 55% amino acid identity in the LRR domain. Both FLS2s trigger a similar defense response when exposed to flg22, but the shorter epitope flg15 (Figure 3E) activates LeFLS2 much more than AtFLS2, and the flg15 from *E. coli* is almost inactive in Arabidopsis and highly active in tomato (Robatzek et al., 2007). A different flagellin epitope, a few amino acids downstream from flg22, termed flgII-28 from *Pseudomonas syringae* pv. *tomato*, elicits a defense response in tomato plants and other solanaceous species, but not in Arabidopsis. The receptor for flgII-28 has not been identified so far (Clarke et al., 2013).

AtFLS2 and VvFLS2 from grapevine (*Vitis vinifera*) share 56% sequence identity in the LRR domain. AtFLS2 and VvFLS2 trigger a similar defense response when exposed to flg22 from *Pseudomonas aeruginose* and *Xanthomonas campestris*, however VvFLS2 has a weaker response to flg22 from *Burkholderia phytofirmans* (Trdá et al., 2014).

AtFLS2 and OsFLS2 from rice share 45% overall amino acid identity and OsFLS2 elicits

similar defense responses to *Psa*-flg22 as AtFLS2. The peptide flg22Δ2 (Figure 3E), which lacks two amino acids near the C-terminus of flg22, still binds to AtFLS2 but does not trigger a defense response in Arabidopsis (Bauer et al., 2001). However flg22Δ2 triggers a defense response in rice (Takai et al., 2008). The virulent rice pathogen *Acidovorax avenae* strain K1 has an identical flg22 sequence as *Psa*, however it does not trigger an FLS2 mediated defense response in rice. Apparently, this pathogen strain is able to mask its flg22 epitope by N-glycosylation (Hirai et al., 2011).

EFR detects the widely conserved and very abundant EF-Tu protein, a bacterial protein translation elongation factor (Kunze et al., 2004; Zipfel et al., 2006) (Figure 3C+D). EF-Tu has been identified as one of the slowest evolving proteins (Lathe and Bork, 2001) which makes it a great target for PRRs. The epitope elf18 at the N-terminus of the EF-Tu protein binds to the LRR domain of EFR (Kunze et al., 2004). Elf18 peptides derived from the EF-Tu sequences of *Pseudomonas syringae* pv *tomato* DC3000 and *Xylella fastidiosa* have a lower activity in Arabidopsis than elf18 peptides based on *Erwinia amylovora*, *Erwinia chrysanthemi* (now called *Dickeya dadantii*), *Agrobacterium tumefaciens* and *Sinorhizobium meliloti* EF-Tu (Kunze et al., 2004). EFR has only been found in members of the Brassicaceae plant family (Kunze et al., 2004).

In rice elf18 does not trigger a defense response, but a different EF-Tu epitope named EFa50 does activate defenses. The receptor for EFa50 (which was originally derived from the pathogenic bacterium *Acidovorax avenae*) has not been identified so far (Furukawa et al., 2013).

The above examples demonstrate an ongoing co-evolution of FLS2 and EFR in plants and the flg22 and elf18 epitopes in bacteria, respectively. By definition PAMPs evolved slower than effectors, however, as shown by examples above, successful pathogen species or strains find strategies to mask or alter their PAMPs. In the future we will have to re-name some PRRs and divide the term PRRs into sub-terms. Is it justified to call AtFLS2, LeFLS2 and OsFLS2 all by the same name if they do not recognize the exact same flagellin epitopes? What will we call receptors that recognize different epitopes of the same microbial molecule? Receptors can be categorized by their structures (e.g. LRR-RLK, LysM-RLP), functions (e.g. patter-recognition receptor), ligands (e.g. EFR: EF-Tu receptor), localization (e.g. RLCK: receptor-like cytoplasmic kinase), signaling partners, etc. The term pattern-recognition receptor is confusing because it contains information about the receptor function (detection of slowly evolving, broadly conserved microbial molecules), the receptor localization (plasma membrane), the receptor structure (RLKs and RLPs) and the signaling output (PTI: PAMP-triggered immunity). We will need more terms for receptors that comply only with some of the PRR characteristics. Furthermore, so far we have little information on what plant tissues and organs the different PRRs are localized in. In the future, the characterization of thousands of RLKs and RLPs in many different plant species and cultivars and the identification of their respective ligands will improve our understanding of plant biology and plant-microbe interactions.

Leucine-rich repeat (LRR) domain

Leucine-rich repeat (LRR) domains are protein - ligand interaction domains found in over 2000 different protein families in eukaryotes, prokaryotes and viruses. In 1985 Takahashi et al. first described the LRR domain. The group obtained the amino acid sequence of the 3.1S leucine-rich α_2-glycoprotein (LRG) from human plasma and observed thirteen repeats, all containing the same consensus sequence PxxLLxxxxxLxxLxLxxNxLxxL (Takahashi et al., 1985). The first crystal structure of an LRR protein – the porcine ribonuclease inhibitor – was obtained eight years later by Kobe and Deisenhofer (Kobe and Deisenhofer 1993). Ten years later, the first crystal structure of an LRR protein from plants – the polygalacturonase-inhibiting protein (PGIP) from *Phaseolus vulgaris* – was obtained by Di Matteo et al. (Di Matteo et al., 2003).

A single leucine-rich repeat consists of 20-30 amino acids (Figure 2A). The conserved hydrophobic residues fold to the inside of the protein and give the domain its typical structure. The variable and solvent exposed residues, represented as "x" in the preceding consensus, give individual LRR domains their specific function. The motif LxxLxLxxN is located at the concave face of the overall LRR domain and has a secondary beta strand structure. In many LRR proteins, the concave face is the site of protein – ligand interaction (Kobe and Kajava, 2001). Different LRR proteins vary greatly in the size of their LRR domain, ranging from 2 to up to 42 repeats (Enkhbayar et al., 2004). The Arabidopsis EFR ectodomain includes 511 LRR amino acids shaped into 21 repeats (Figure 2B). The Arabidopsis FLS2 ectodomain includes 677 LRR amino acids shaped into 28 repeats (Figure 4) and the ectodomain of the Arabidopsis SERK proteins

include slightly varying numbers of LRR amino acids but they all shape into 5 repeats (Figure 4). The typical LRR consensus sequence for plant protein extracellular LRR domains is: PxxLxxLxxLxxLxLxxNxLxGxI.

A central aspect of the study of molecular plant – microbe interactions is the study of the interaction between microbial- and plant-derived molecules. As previously mentioned, LRR domains have been found in many PRRs and R proteins. Understanding how the LRR domains of PRRs and R proteins gain specificity to their microbial ligands is a crucial element for the understanding of plant – microbe interactions. If we understand the molecular details of these interactions, we will be able to manipulate them, which ultimately will lead to the ability to engineer PRRs and R proteins able to recognize previously unrecognizable PAMPs and effectors. Adding these new receptors to the immune system of susceptible crop cultivars could significantly improve plant health.

LRR proteins play crucial roles not only in the plant immune system but also in the immune system of animals. Most famous are the LRR ectodomain containing toll-like receptors (TLRs), which are PRRs triggering innate immune responses (reviewed in Kawai and Akira, 2010). The *Toll* gene was first discovered in *Drosophila melanogaster* (Anderson et al., 1985). In chapter 2 TLRs are discussed in more detail and in chapter 4 the fascinating VLRs – LRR proteins in the adaptive immune system of jawless vertebrates – are discussed.

Research objectives

The main focus for my Ph.D. research has been to learn more about how pattern-recognition receptors in the plant immune system work, and as an application of this knowledge, to explore if we can engineer novel pattern-recognition receptors in order to strengthen the plant immune system. I conducted my research on the two PRRs EFR and FLS2.

For the research described in chapter 2, my objective was to identify sites in the LRR domain of EFR important for elf18 binding (Helft et al., 2011). I also contributed to the late stages of development of the "RCM" LRR conservation mapping program, which identifies conserved solvent exposed amino acid sites (probable active sites) in LRR domains (Helft et al., 2011). In chapter 3, my objective was to confirm that the identified conserved amino acid site in the EFR ectodomain indeed is the elf18 binding site. In chapter 4, I explored two different approaches to engineer PRRs with novel ligand specificities. In chapter 5, my objective was the identification and functional testing of the BAK1 interaction site in the FLS2 and EFR ectodomains.

References

Van den Ackerveken, G.F., Van Kan, J.A., and De Wit, P.J. (1992). Molecular analysis of the avirulence gene avr9 of the fungal tomato pathogen *Cladosporium fulvum* fully supports the gene-for-gene hypothesis. Plant J. **2**: 359–66.

Anderson, K. V, Bokla, L., and Nüsslein-Volhard, C. (1985). Establishment of dorsal-ventral polarity in the Drosophila embryo: the induction of polarity by the Toll gene product. Cell **42**: 791–8.

Bar, M., Sharfman, M., Ron, M., and Avni, A. (2010). BAK1 is required for the attenuation of ethylene-inducing xylanase (Eix)-induced defense responses by the decoy receptor LeEix1. Plant J. **63**: 791–800.

Bauer, Z., Gómez-Gómez, L., Boller, T., and Felix, G. (2001). Sensitivity of different ecotypes and mutants of *Arabidopsis thaliana* toward the bacterial elicitor flagellin correlates with the presence of receptor-binding sites. J. Biol. Chem. **276**: 45669–76.

Becraft, P.W. (2002). Receptor kinase signaling in plant development. Annu. Rev. Cell Dev. Biol. **18**: 163–92.

Bent, A.F. and Mackey, D. (2007). Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Annu Rev Phytopathol **45**: 399–436.

Boller, T. and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu Rev Plant Biol **60**: 379–406.

Brutus, A., Sicilia, F., Macone, A., Cervone, F., and De Lorenzo, G. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. Proc Natl Acad Sci U S A **107**: 9452–9457.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature **448**: 497–500.

Choi, J., Tanaka, K., Cao, Y., Qi, Y., Qiu, J., Liang, Y., Lee, S.Y., and Stacey, G. (2014). Identification of a plant receptor for extracellular ATP. Science **343**: 290–4.

Clarke, C.R., Chinchilla, D., Hind, S.R., Taguchi, F., Miki, R., Ichinose, Y., Martin, G.B., Leman, S., Felix, G., and Vinatzer, B.A. (2013). Allelic variation in two distinct *Pseudomonas syringae* flagellin epitopes modulates the strength of plant immune responses but not bacterial motility. New Phytol. **200**: 847–60.

Dardick, C., Chen, J., Richter, T., Ouyang, S., and Ronald, P. (2007). The rice kinase database. A phylogenomic database for the rice kinome. Plant Physiol. **143**: 579–86.

Dodds, P.N. and Rathjen, J.P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet **11**: 539–548.

Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T., and Matsushima, N. (2004). Structural principles of leucine-rich repeat (LRR) proteins. Proteins **54**: 394–403.

Felix, G., Duran, J.D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J. **18**: 265–76.

Fritz-Laylin, L.K., Krishnamurthy, N., Tör, M., Sjölander, K. V, and Jones, J.D.G. (2005). Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis. Plant Physiol. **138**: 611–23.

Furukawa, T., Inagaki, H., Takai, R., Hirai, H., and Che, F.-S. (2013). Two distinct EF-Tu epitopes induce immune responses in rice and Arabidopsis. Mol. Plant. Microbe. Interact. **27**: 113–124.

Gao, M., Wang, X., Wang, D., Xu, F., Ding, X., Zhang, Z., Bi, D., Cheng, Y.T., Chen, S., Li, X., and Zhang, Y. (2009). Regulation of cell death and innate immunity by two receptor-like kinases in Arabidopsis. Cell Host Microbe **6**: 34–44.

Gómez-Gómez, L. and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol. Cell **5**: 1003–1011.

Halter, T., Imkampe, J., Mazzotta, S., Wierzba, M., Postel, S., Bücherl, C., Kiefer, C., Stahl, M., Chinchilla, D., Wang, X., Nürnberger, T., Zipfel, C., Clouse, S., Borst, J.W., Boeren, S., de Vries, S.C., Tax, F., and Kemmerling, B. (2013). The leucine-rich repeat receptor kinase BIR2 is a negative regulator of BAK1 in plant immunity. Curr. Biol. **24**: 134-143.

Häweker, H., Rips, S., Koiwa, H., Salomon, S., Saijo, Y., Chinchilla, D., Robatzek, S., and von Schaewen, A. (2010). Pattern recognition receptors require N-glycosylation to mediate plant immunity. J. Biol. Chem. **285**: 4629–36.

Heese, A., Hann, D.R., Gimenez-Ibanez, S., Jones, A.M.E., He, K., Li, J., Schroeder, J.I., Peck, S.C., and Rathjen, J.P. (2007). The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. Proc. Natl. Acad. Sci. U. S. A. **104**: 12217–22.

Helft, L., Reddy, V., Chen, X., Koller, T., Federici, L., Fernández-Recio, J., Gupta, R., and Bent, A. (2011). LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains. PLoS One **6**: e21614.

Hirai, H., Takai, R., Iwano, M., Nakai, M., Kondo, M., Takayama, S., Isogai, A., and Che, F.-S. (2011). Glycosylation regulates specific induction of rice immune responses by *Acidovorax avenae* flagellin. J. Biol. Chem. **286**: 25519–30.

Huffaker, A., Pearce, G., and Ryan, C. a (2006). An endogenous peptide signal in Arabidopsis activates components of the innate immune response. Proc. Natl. Acad. Sci. U. S. A. **103**: 10098–103.

Huffaker, A. and Ryan, C.A. (2007). Endogenous peptide defense signals in Arabidopsis differentially amplify signaling for the innate immune response. Proc. Natl. Acad. Sci. U. S. A. **104**: 10732–6.

Jehle, A.K., Lipschis, M., Albert, M., Fallahzadeh-Mamaghani, V., Fürst, U., Mueller, K., and Felix, G. (2013). The receptor-like protein ReMAX of Arabidopsis detects the microbe-associated molecular pattern eMax from *Xanthomonas*. Plant Cell **25**: 2330–40.

Jones, J.D. and Dangl, J.L. (2006). The plant immune system. Nature **444**: 323–329.

De Jonge, R., van Esse, H.P., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K. V, and Thomma, B.P.H.J. (2012). Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc. Natl. Acad. Sci. U. S. A. **109**: 5110–5.

Joosten, M.H., Vogelsang, R., Cozijnsen, T.J., Verberne, M.C., and De Wit, P.J. (1997). The biotrophic fungus *Cladosporium fulvum* circumvents Cf-4-mediated resistance by producing unstable AVR4 elicitors. Plant Cell **9**: 367–79.

Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E., and Shibuya, N. (2006). Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proc. Natl. Acad. Sci. U. S. A. **103**: 11086–91.

Kawai, T. and Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat. Immunol. **11**: 373–84.

Kobe, B. and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature **366**: 751–6.

Kobe, B. and Kajava, A. V (2001). The leucine-rich repeat as a protein recognition motif. Curr. Opin. Struct. Biol. **11**: 725–32.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell **16**: 3496–507.

Lathe, W.C. and Bork, P. (2001). Evolution of tuf genes: ancient duplication, differential loss and gene conversion. FEBS Lett. **502**: 113–6.

Li, H., Zhou, S.-Y., Zhao, W.-S., Su, S.-C., and Peng, Y.-L. (2009). A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast disease resistance. Plant Mol. Biol. **69**: 337–46.

Liebrand, T.W.H., van den Berg, G.C.M., Zhang, Z., Smit, P., Cordewener, J.H.G., America, A.H.P., America, A.H.P., Sklenar, J., Jones, A.M.E., Tameling, W.I.L., Robatzek, S., Thomma, B.P.H.J., and Joosten, M.H.A.J. (2013). Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. Proc. Natl. Acad. Sci. U. S. A. **110**: 10010–5.

Lin, W., Ma, X., Shan, L., and He, P. (2013). Big roles of small kinases: the complex functions of receptor-like cytoplasmic kinases in plant immunity and development. J. Integr. Plant Biol. **55**: 1188–97.

Liu, B., Li, J.-F., Ao, Y., Qu, J., Li, Z., Su, J., Zhang, Y., Liu, J., Feng, D., Qi, K., He, Y., Wang, J., and Wang, H.-B. (2012). Lysin motif-containing proteins LYP4 and LYP6 play dual roles in peptidoglycan and chitin perception in rice innate immunity. Plant Cell **24**: 3406–19.

Liu, T., Liu, Z., Song, C., Hu, Y., Han, Z., She, J., Fan, F., Wang, J., Jin, C., Chang, J., Zhou, J.-M., and Chai, J. (2012). Chitin-induced dimerization activates a plant immune receptor. Science **336**: 1160–4.

Liu, Z., Wu, Y., Yang, F., Zhang, Y., Chen, S., Xie, Q., Tian, X., and Zhou, J.-M. (2013). BIK1 interacts with PEPRs to mediate ethylene-induced immunity. Proc. Natl. Acad. Sci. U. S. A. **110**: 6205–10.

Lu, D., Wu, S., Gao, X., Zhang, Y., Shan, L., and He, P. (2010). A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. Proc Natl Acad Sci U S A **107**: 496–501.

Macho, A.P. and Zipfel, C. (2014). Plant PRRs and the activation of innate immune signaling. Mol. Cell **54**: 263–272.

Di Matteo, A., Bonivento, D., Tsernoglou, D., Federici, L., and Cervone, F. (2006). Polygalacturonase-inhibiting protein (PGIP) in plant defence: a structural view. Phytochemistry **67**: 528–533.

Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K.A., Savino, C., De Lorenzo, G., Tsernoglou, D., and Cervone, F. (2003). The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. Proc. Natl. Acad. Sci. U. S. A. **100**: 10124–8.

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H., and Shibuya, N. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **104**: 19613–8.

Oh, M.H., Wang, X., Wu, X., Zhao, Y., Clouse, S.D., and Huber, S.C. (2010). Autophosphorylation of Tyr-610 in the receptor kinase BAK1 plays a role in brassinosteroid signaling and basal defense gene expression. Proc Natl Acad Sci U S A **107**: 17827–17832.

Pfund, C., Tans-Kersten, J., Dunning, F.M., Alonso, J.M., Ecker, J.R., Allen, C., and Bent, A.F. (2004). Flagellin is not a major defense elicitor in *Ralstonia solanacearum* cells or extracts applied to Arabidopsis thaliana. Mol. Plant. Microbe. Interact. **17**: 696–706.

Postel, S., Küfner, I., Beuter, C., Mazzotta, S., Schwedt, A., Borlotti, A., Halter, T., Kemmerling, B., and Nürnberger, T. (2010). The multifunctional leucine-rich repeat receptor kinase BAK1 is implicated in Arabidopsis development and immunity. Eur. J. Cell Biol. **89**: 169–74.

Ranjith-Kumar, C.T., Miller, W., Sun, J., Xiong, J., Santos, J., Yarbrough, I., Lamb, R.J., Mills, J., Duffy, K.E., Hoose, S., Cunningham, M., Holzenburg, A., Mbow, M.L., Sarisky, R.T., and Kao, C.C. (2007). Effects of single nucleotide polymorphisms on Toll-like receptor 3 activity and expression in cultured cells. J Biol Chem **282**: 17696–17705.

Robatzek, S., Bittel, P., Chinchilla, D., Köchner, P., Felix, G., Shiu, S.-H., and Boller, T. (2007). Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. Plant Mol. Biol. **64**: 539–547.

Ron, M. and Avni, A. (2004). The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. Plant Cell **16**: 1604–15.

Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., Malinovsky, F.G., Tör, M., de Vries, S., and Zipfel, C. (2011). The Arabidopsis leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. Plant Cell **23**: 2440–55.

Schulze, B., Mentzel, T., Jehle, A.K., Mueller, K., Beeler, S., Boller, T., Felix, G., and Chinchilla, D. (2010). Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. J Biol Chem **285**: 9444–9451.

Senchou, V., Weide, R., Carrasco, A., Bouyssou, H., Pont-Lezica, R., Govers, F., and Canut, H. (2004). High affinity recognition of a *Phytophthora* protein by Arabidopsis via an RGD motif. Cell. Mol. Life Sci. **61**: 502–9.

Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii-Minami, N., Nishizawa, Y., Minami, E., Okada, K., Yamane, H., Kaku, H., and Shibuya, N. (2010). Two LysM receptor

molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. Plant J. **64**: 204–14.

Shiu, S.H. and Bleecker, A.B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. Plant Physiol. **132**: 530–43.

Shiu, S.H. and Bleecker, A.B. (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.

Shiu, S.-H., Karlowski, W.M., Pan, R., Tzeng, Y.-H., Mayer, K.F.X., and Li, W.-H. (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell **16**: 1220–34.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C., and Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science **270**: 1804–1806.

Sun, W., Dunning, F.M., Pfund, C., Weingarten, R., and Bent, A.F. (2006). Within-species flagellin polymorphism in *Xanthomonas campestris pv campestris* and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses. Plant Cell **18**: 764–79.

Sun, X., Cao, Y., Yang, Z., Xu, C., Li, X., Wang, S., and Zhang, Q. (2004). Xa26, a gene conferring resistance to *Xanthomonas oryzae pv. oryzae* in rice, encodes an LRR receptor kinase-like protein. Plant J. **37**: 517–27.

Sun, X. and Wang, G.-L. (2011). Genome-wide identification, characterization and phylogenetic analysis of the rice LRR-kinases. PLoS One **6**: e16079.

Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.-M., and Chai, J. (2013). Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science **342**: 624–8.

Takahashi, N., Takahashi, Y., and Putnam, F.W. (1985). Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich alpha 2-glycoprotein of human serum. Proc. Natl. Acad. Sci. U. S. A. **82**: 1906–10.

Takai, R., Isogai, A., Takayama, S., and Che, F.-S. (2008). Analysis of flagellin perception mediated by flg22 receptor OsFLS2 in rice. MPMI **21**: 1635–42.

Takken, F.L.W. and Goverse, A. (2012). How to build a pathogen detector: Structural basis of NB-LRR function. Curr. Opin. Plant Biol. **15**: 375–384.

Trdá, L., Fernandez, O., Boutrot, F., Héloir, M.-C., Kelloniemi, J., Daire, X., Adrian, M., Clément, C., Zipfel, C., Dorey, S., and Poinssot, B. (2014). The grapevine flagellin receptor VvFLS2 differentially recognizes flagellin-derived epitopes from the endophytic growth-promoting bacterium *Burkholderia phytofirmans* and plant pathogenic bacteria. New Phytol. **201**: 1371–84.

Vetter, M.M., Kronholm, I., He, F., Häweker, H., Reymond, M., Bergelson, J., Robatzek, S., and de Meaux, J. (2012). Flagellin perception varies quantitatively in *Arabidopsis thaliana* and its relatives. Mol. Biol. Evol. **29**: 1655–67.

Vij, S., Giri, J., Dansana, P.K., Kapoor, S., and Tyagi, A.K. (2008). The receptor-like cytoplasmic kinase (OsRLCK) gene family in rice: organization, phylogenetic relationship, and expression during development and stress. Mol. Plant **1**: 732–50.

Wan, J., Zhang, X.C., Neece, D., Ramonell, K.M., Clough, S., Kim, S.Y., Stacey, M.G., and Stacey, G. (2008). A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis. Plant Cell **20**: 471–481.

Willmann, R., Lajunen, H.M., Erbs, G., Newman, M.-A., Kolb, D., Tsuda, K., Katagiri, F., Fliegmann, J., Bono, J.-J., Cullimore, J. V, Jehle, A.K., Götz, F., Kulik, A., Molinaro, A., Lipka, V., Gust, A.A., and Nürnberger, T. (2011). Arabidopsis lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. Proc. Natl. Acad. Sci. U. S. A. **108**: 19824–9.

Wu, Y. and Zhou, J.-M. (2013). Receptor-like kinases in plant innate immunity. J. Integr. Plant Biol. **55**: 1271–86.

Yamaguchi, Y., Huffaker, A., Bryan, A.C., Tax, F.E., and Ryan, C.A. (2010). PEPR2 is a second receptor for the Pep1 and Pep2 peptides and contributes to defense responses in Arabidopsis. Plant Cell **22**: 508–22.

Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X., Zou, Y., Gao, M., Zhang, X., Chen, S., Mengiste, T., Zhang, Y., and Zhou, J.M. (2010). Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. Cell Host Microbe **7**: 290–301.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D.G., Boller, T., and Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. Cell **125**: 749–760.

Figures

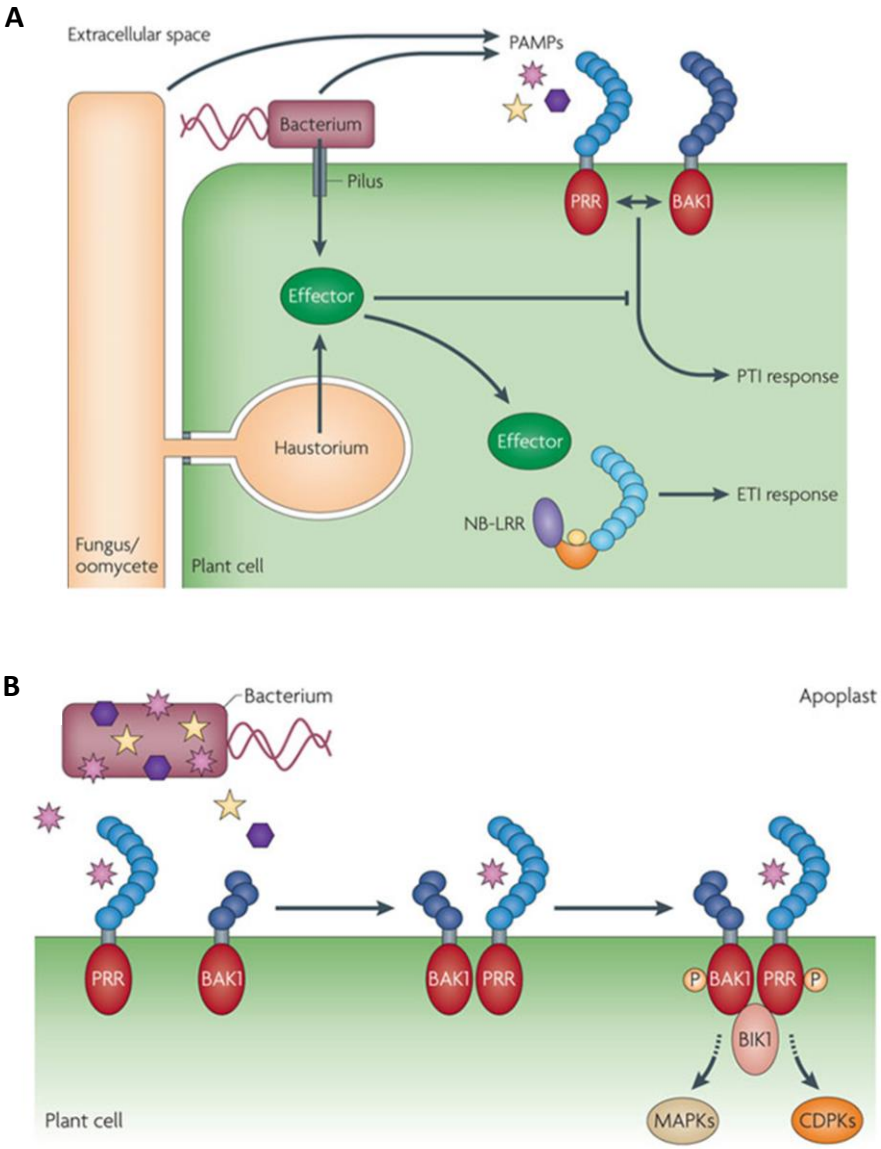

Figure 1: Pattern-recognition receptors in the plant immune system. The figures are adapted from Dodds and Rathjen, 2010. **(A)** PRRs localized at the plasma membrane detect PAMPs, DAMPs and apoplastic effectors. Intracellular R proteins detect effectors injected into the cytoplasm. **(B)** PRRs form receptor complexes with regulatory RLKs upon ligand detection to initiate downstream signaling.

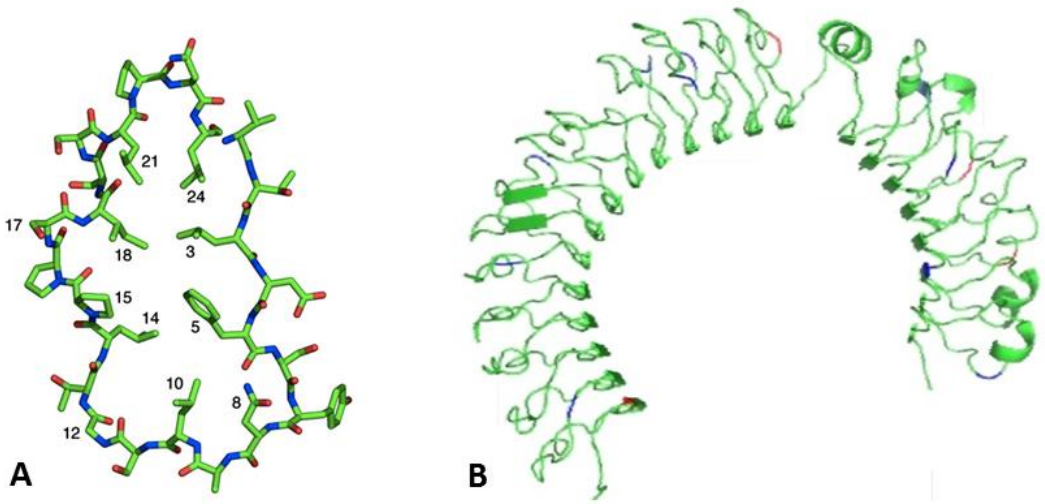

Figure 2: The leucine-rich repeats (LRR) domain. (A) adopted from (Di Matteo et al., 2006). Structure of a single LRR: the hydrophobic consensus residues are driven to the protein interior. The more variable residues reside on the protein surface. **(B)** adopted from (Häweker et al., 2010). Bioinformatics model (using "Swiss-Model") of the EFR LRR domain consisting of 21 repeats.

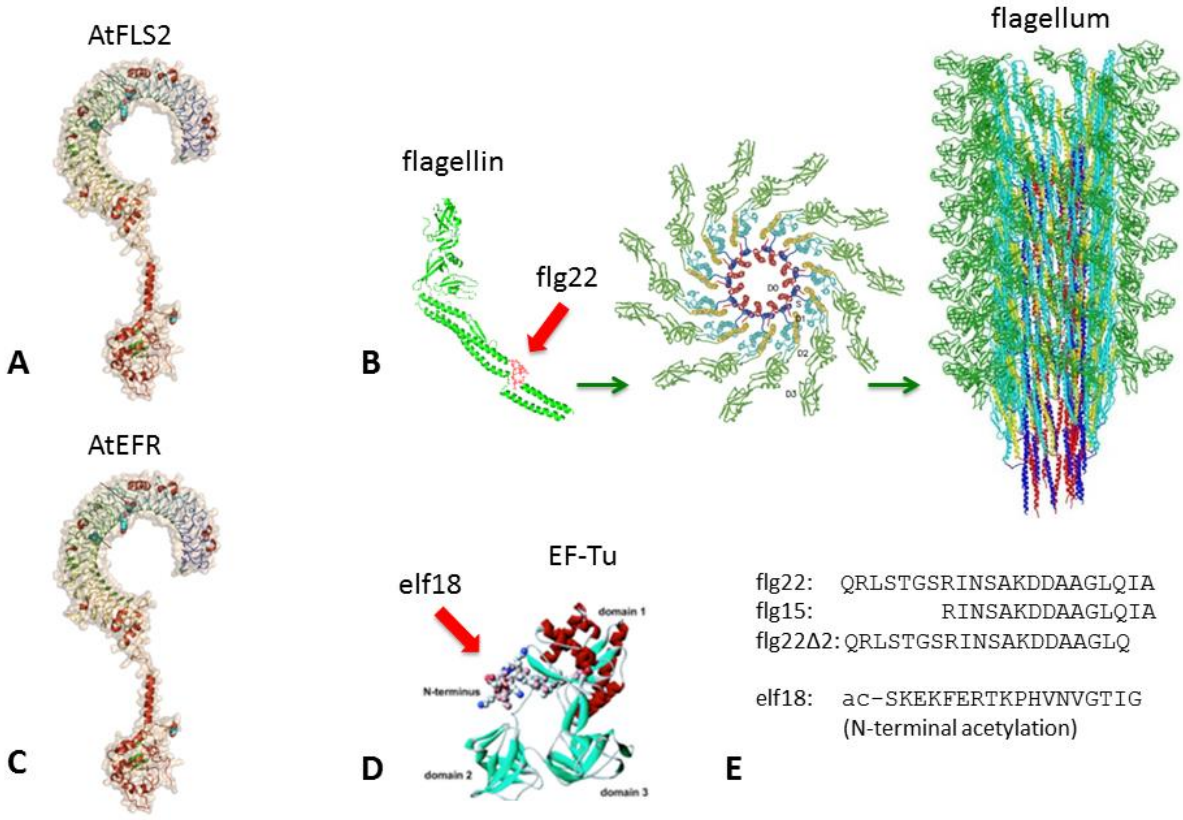

Figure 3: FLS2 detects flg22 and EFR detects elf18. FLS2 **(A)** and EFR **(C)** are illustrated by using the TLR3 models adopted from (Ranjith-Kumar et al., 2007). FLS2 detects the flg22 epitope of flagellin **(B)**. Flagellin monomers fold into multimeric structures, which form the bacterial flagellum. EFR detects the epitope elf18 **(D)** of the bacterial elongation factor EF-Tu. **(E)** Peptide sequences.

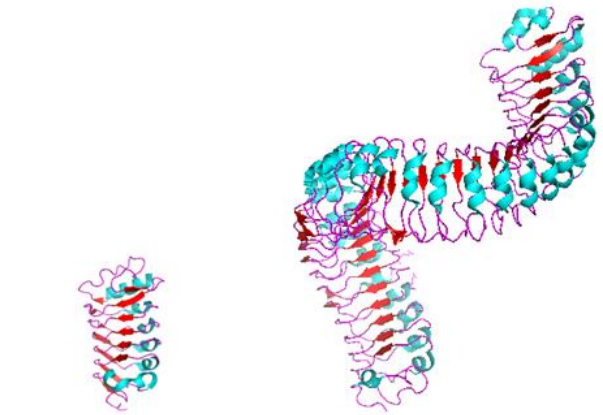

Figure 4: The ectodomains of SERK1 and FLS2. The ectodomain of the regulatory LRR-RLK SERK1 (PDB ID: 4LSC) has 5 LRRs, The ectodomain of FLS2 (PDB ID: 4MN8) has 28 LRRs.

Chapter 2: Identification of conserved solvent exposed amino acids in the ectodomain of EFR using conservation mapping

The work presented in this chapter is a brief extract summarizing and commenting on my contributions to the published research paper:

Helft, L., Reddy, V., Chen, X., **Koller, T**., Federici, L., Fernandez-Recio, J., Gupta, R. and Bent, A. (2011) LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains. *PloS One*, 6, e21614.

A copy of the Helft et al. 2011 paper, which provides a formal description of the broader project, is appended at the end of this chapter.

Sequences obtained in this study have been deposited at Genbank under accession numbers JN002095-JN002103.

Abstract

Many pattern recognition receptors and R proteins in the plant immune system, and other plant receptor proteins, contain large leucine-rich repeat (LRR) domains. In several cases a direct interaction of immune system LRR domains with a danger signal or a pathogen-derived ligand has been demonstrated. We studied the specificity and plasticity of the LRR domains of EFR and FLS2, which detect segments of bacterial EF-Tu and flagellin proteins, respectively. Based on the hypothesis that amino acids involved in ligand interaction are evolutionarily more conserved while other solvent exposed residues in the LRR domain are less conserved, we used our recently released Repeat Conservation Mapping computational tool to discover functional sites within the LRR domain of EFR. EFR LRR domains from nine Brassicaceae species were cloned and sequenced. Capacity for detection of elf18 (the recognized epitope of EF-Tu) by these putative EFR LRR domains was confirmed by transient expression in *Nicotiana benthamiana* of domain-swap full-length *EFR* derivatives carrying these Brassicaceae-derived *EFR* LRR-encoding domains in place of the corresponding Arabidopsis sequence. The remainder of these chimeric genes carried Arabidopsis *EFR* DNA sequence. These nine functional EFR LRR sequences were then subjected to repeat conservation mapping to discover conserved amino acid clusters on the predicted surface of the folded protein. The EFR sequences were also used to further document the behavior of the RCM computer program. Several conserved sites on the inferred surface of the EFR LRR domain were identified. Thus in addition to EF-Tu/elf18 ligand binding sites, we hypothesize that conservation mapping identified sites important for other LRR functions that may include receptor multimerization, interactions with co-receptors, and/or receptor glycosylation, processing and subcellular localization.

Introduction

The Repeat Conservation Mapping bioinformatics computer program (RCM) was developed in our laboratory (Helft et al., 2011) (Figure 2). The program predicts functional sites on LRR domains by identifying conserved amino acid clusters on the predicted surface of LRR domains, via comparisons of the amino acid sequences of orthologous proteins from genetically distinct organisms (such as different plant species). The method was validated using known LRR-ligand crystal structures, and was used to discover previously unknown functional sites (Helft et al., 2011). The program is available online at www.plantpath.wisc.edu/RCM. I investigated EFR functional sites as part of that work.

Hypotheses for T. Koller studies:

- Different species of the plant family Brassicaceae contain functional EFR orthologs
- Comparing the amino acid sequence of different EFR orthologs can identify clusters of spatially adjacent amino acids that were conserved while other parts of the LRR domain diverged during species evolution
- Amino acid clusters evolutionarily conserved among EFR orthologs are important for receptor function
- Only a few amino acids of the huge LRR domain play a primary, specificity-determining role in ligand binding
- The elf18 binding site is the most conserved region and is located in the concave face of the LRR domain

- The plant immune system can utilize chimeric receptors containing the LRR domain of a different plant species

Methods

Methods are described in detail in Helft et al. 2011.

Synopsis of study approach and results

EFR seems to exist only in members of the Brassicaceae (Kunze et al., 2004). Several individual plants of each Brassicaceae species used for this study were tested in the ROS assay. Some plants exhibited significantly stronger ROS bursts than others (Figure 1). The EFR LRR domain encoding part of the genes from some of the positively responding species were cloned and sequenced (*Brassica rapa, Brassica napus, Brassica aucheri, Erysimum raulinii*, *Eruca sativa, Enarthrocarpus arcuatus* and *Biscutella auriculata*). In the cases of *Erysimum raulinii* and *Enarthrocarpus arcuatus* two different alleles were obtained from each species because of polyploidy.

To confirm the function of these alleles, chimeric EFR receptors based on Arabidopis EFR were generated in which the wild type EFR LRR domain of Arabidopsis was replaced by the putative EFR LRR domains from the different Brassicaceae species. The nine chimeric genes were transiently expressed in *Nicotiana benthamiana*, a plant species lacking an endogenous EFR. All of the nine chimeric receptors triggered an oxidative burst in *N. benthamiana* leaves

upon treatment with the ligand elf18 and were thus shown to be functional EFR LRR domains (Table 1).

We then used the ten EFR LRR domain alleles (the nine newly sequenced ones and EFR from Arabidopsis) to generate conservation maps using the RCM program. We compared all ten sequences (Figure 3A), two sequences from the most distantly related species *Eruca sativa* and *Erysimum raulinii* (Figure 3B) and two similar sequences which were the two *EFR* alleles obtained from *Erysimum raulinii* (Figure 3C).

Discussion

EFR orthologs from Brassicaceae species

Most of the 32 Brassicaceae accessions tested in this study elicited an ROS burst when exposed to elf18 (Figure 1B). Quantitatively the ROS burst varied between different species as well as between different plants of the same species. This was probably to a large extent due to the set-up of the experiment. Plants were tested on different days and at different plant ages. However, some differences might be due to EFR protein abundance, as was shown to be a mayor determinant for the quantity of flg22 triggered immune responses mediated by FLS2 in different Arabidopsis accessions (Vetter et al., 2012). Another possibility might be a variation of affinity of the Brassicaceae EFRs to elf18. However, in this study, we did not test EFR protein abundance and EFR – elf18 affinity.

We cloned, sequenced and functionally tested EFR LRR domains from seven different Brassicaceae species. For the functional testing we used the chimeric receptor approach. We generated nine chimeric receptors and they all were functional in *N.benthamiana* (Figure 2). Thus we successfully integrated the chimeric receptors into the *N.benthamiana* immune system. From studies performed in Arabidopsis we know that EFR interacts with BAK1 after exposure to elf18 (Chinchilla et al., 2007). Similarly FLS2 interacts with BAK1 upon exposure to flg22 (Chinchilla et al., 2007; Heese et al., 2007) and the receptor complex is mediated by an initial interaction of the FLS2 and BAK1 ectodomains mediated by the ligand flg22 (chapter 5 and Sun et al., 2013). Thus it is very likely that a similar elf18 mediated EFR – BAK1 ectodomain complex initiates EFR signaling. It is interesting to note that all the Brassicaceae EFR LRR domains tested in this study probably were able to interact with the ectodomains of BAK1 from *N.benthamiana*. Thus the NbBAK1 seems to be a flexible signaling partner.

So far, elf18 perception has only been found in members of the huge Brassicaceae family (Kunze et al., 2004). This is surprising considering the great PAMP characteristics of EF-Tu. It is a very abundant, essential and slow evolving microbial protein. A recent study in rice found a different EF-Tu epitope – named EFa50 – which elicits defense responses in rice (Furukawa et al., 2013). The receptor for EFa50 has not been identified so far. Different types of EF-Tu receptors might exist in different plant families. We are only at the beginning of understanding how plant families/species/cultivars gain and loose PRRs during the courses of evolution, domestication and plant breeding.

Repeat conservation maping (RCM)

Former Bent lab graduate student Laura Helft generated an initial EFR conservation map using the EFR sequences from Arabidopsis and *Brassica rapa* as input. Now with the additional nine functionally tested EFR LRR domain sequences at hand, we were able to generate a more precise conservation map and furthermore we used the ten sequences to test the changes in the RCM program outputs, when different numbers of sequences, resulting from closely or distantly related species, were used as program inputs.

RCM using ten EFR LRR domain sequences

Conservation mapping of the ten sequences revealed several potential active sites in the EFR ectodomain (Figure 3A). We hypothesize the distinct conserved pattern located on repeats 15-17 in the concave face to be the elf18 binding site. Former graduate student Laura Helft has shown that mutations at this site disrupt receptor function and has shown that “control” mutations in areas not predicted to be the functional site do not disrupt EFR function. The region right below the concave face of repeats 15-17 could be the BAK1 interaction site. In both the FLS2-ECD – flg22 – BAK1-ECD interaction (Sun, Li, et al., 2013) and the BRI1-ECD – BL – SERK1/BAK1-ECD interaction (Santiago et al., 2013; Sun, Han, et al., 2013) the respective ligands mediate the interaction of the two LRR domains. In both cases the co-receptor interaction site is located right below the ligand binding site in the ectodomains of the main receptors.

Alternatively to a single elf18 binding site, is possible that the binding site is more spread out over several LRRs. One part of elf18 might bind to the concave face of LRRs 10-11

and another part to LRRs 15-17, which then could generate the BAK1 interaction platform. This would comply with the address-message principle that was proposed for the flg22 - FLS2 interaction even before the crystal structure of the receptor complex was available (address: binding to FLS2, message: interaction with BAK1). However, Laura's mutations in repeats 10-11 did not interfere with receptor function (Helft et al., 2011).

Our conservation map revealed additional conserved regions outside the concave face. Some of these sites might be important for EFR homodimerization. Using a co-immunoprecipitation assay, Sun et al. detected FLS2 homodimerization prior to flg22 binding (Sun et al., 2012). FLS2 might be present in homodimers at the plasma membrane and a sub pool might disassociate after flg22 binding. A similar scenario could be true for EFR. However in the crystal structure of flg22 free FLS2ΔLRR2-6, no FLS2 dimerization was observed (Sun, Li, et al., 2013). Thus at this point in time it is unclear if FLS2 and EFR are present as homodimers at the plasma membrane.

The EFR ectodomain might interact with so far unidentified molecules. In future studies the conserved, solvent exposed sites in the EFR ectodomain uncovered by conservation mapping could be further characterized.

RCM using many or few EFR orthologs, from closely or distantly related species

Additional RCM maps were generated using smaller subsets of the ten EFR sequences, based on the overall relatedness of different Brassicaceae species (Couvreur et al., 2010). One analysis used EFR sequences from two of the most distantly related species of our sample set, *E. sativa* and *E. raulinii* (Figure 3B), while another analysis used two different EFR alleles

obtained from the species *E. raulinii* (Figure 3C). The maps of functionally similar LRRs from closely related sequences emphasize amino acid clusters containing less common amino acid residues that score highly in Blosum matrices, whether or not they reside at functional sites. The maps made from the ten functionally confirmed EFR sequences more reliably emphasize evolutionary conservation of residue clusters (Figure 3A). This latter maps also offers better resolution of conserved and diversified areas, but maps made from just two sequences from more distantly related species were sufficient to highlight the main features also present on the maps made from the ten sequences.

The advantages of the RCM program are discussed in more detail in (Helft et al., 2011).

References

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature **448**: 497–500.

Couvreur, T.L.P., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A., and Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Mol. Biol. Evol. **27**: 55–71.

Furukawa, T., Inagaki, H., Takai, R., Hirai, H., and Che, F.-S. (2013). Two distinct EF-Tu epitopes induce immune responses in rice and Arabidopsis. Mol. Plant. Microbe. Interact. **27**: 113–124.

Heese, A., Hann, D.R., Gimenez-Ibanez, S., Jones, A.M.E., He, K., Li, J., Schroeder, J.I., Peck, S.C., and Rathjen, J.P. (2007). The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. Proc. Natl. Acad. Sci. U. S. A. **104**: 12217–22.

Helft, L., Reddy, V., Chen, X., Koller, T., Federici, L., Fernández-Recio, J., Gupta, R., and Bent, A. (2011). LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains. PLoS One **6**: e21614.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell **16**: 3496–507.

Santiago, J., Henzler, C., and Hothorn, M. (2013). Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. Science **341**: 889–92.

Sun, W., Cao, Y., Jansen Labby, K., Bittel, P., Boller, T., and Bent, A.F. (2012). Probing the Arabidopsis flagellin receptor: FLS2-FLS2 association and the contributions of specific domains to signaling function. Plant Cell **24**: 1096–113.

Sun, Y., Han, Z., Tang, J., Hu, Z., Chai, C., Zhou, B., and Chai, J. (2013). Structure reveals that BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. Cell Res. **23**: 1326–9.

Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.-M., and Chai, J. (2013). Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science **342**: 624–8.

Vetter, M.M., Kronholm, I., He, F., Häweker, H., Reymond, M., Bergelson, J., Robatzek, S., and de Meaux, J. (2012). Flagellin perception varies quantitatively in Arabidopsis thaliana and its relatives. Mol. Biol. Evol. **29**: 1655–67.

Figures

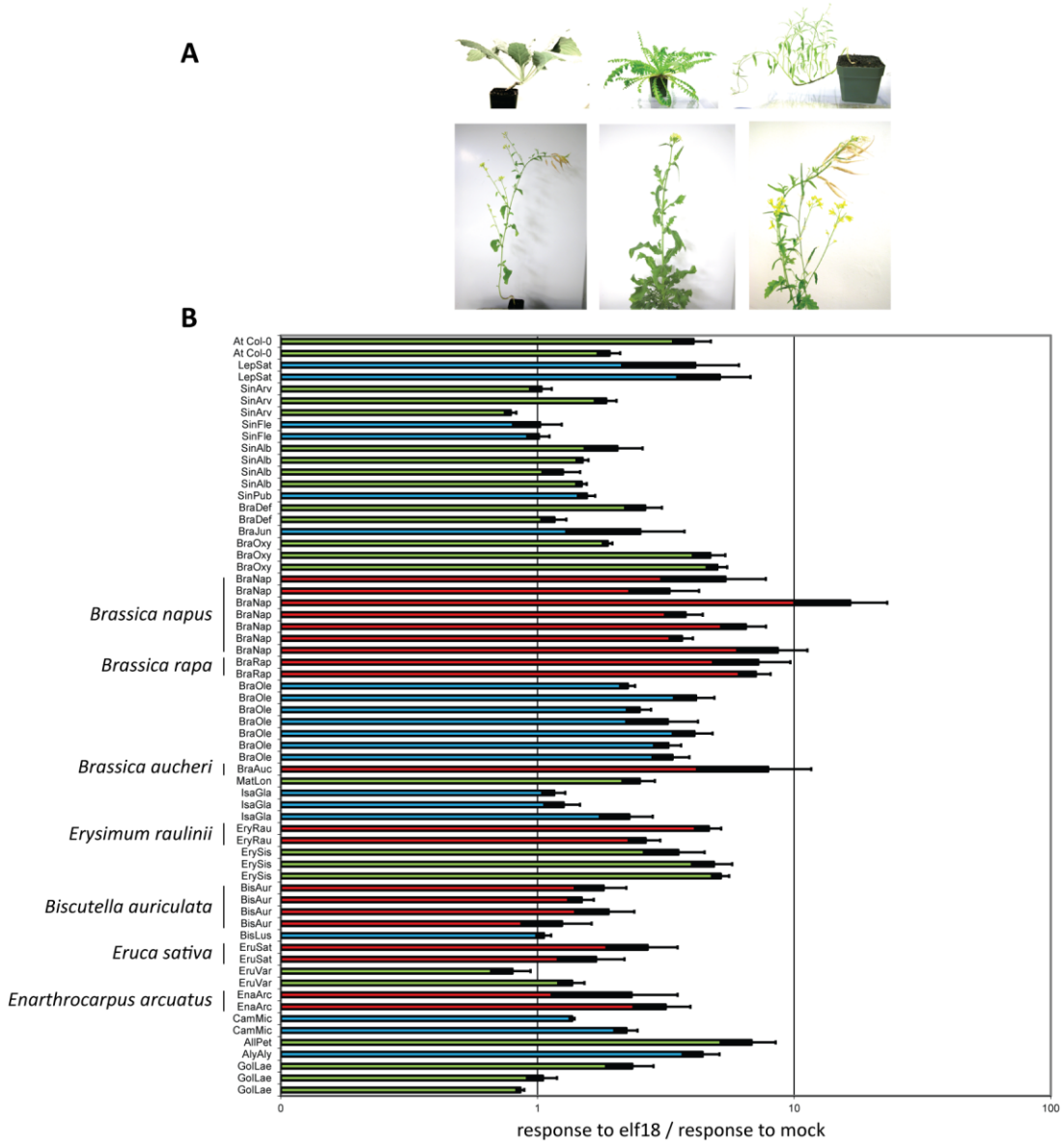

Figure 1: Testing of various Brassicaceae species for response to elf18 (presence of EFR). (A) Pictures taken by T.Koller of some of the Brassicaceae species tested in this study. **(B)** Pooled data from several ROS assay experiments. Each bar represents one plant accession tested on one day. For simplicity only the species names (six letter codes) and not the accessions are indicated. Several leaf discs per plant accession were inoculated with elf18 or mock and ROS burst (luminescence) was measured. The response from each leaf disc to elf18 was divided by the average response of the mock-treated leaf discs from the same plant accession of the same experiment. Plant accessions for which response to elf 18 / response to mock > 1 are elf18-responsive. For better legibility of the graph, the colors of the bars alter between species. The bars in red represent species that were chosen for sequencing of their putative endogenous *EFR*.

Chimeric EFR with LRR of…	Mean ROS production after elf18 treatment/mock treatment	Number of leaf discs (n)	Grouping (90% confidence)
BraRap 50-1	9.2	12	A
EnaArc 52-3	8.3	12	A
EnaArc 52-1	12	9	A
BraNap 66-2	6.7	12	A
EryRau 65-5	6.5	12	A
EryRau 65-4	7.8	12	A
EruSat 61-2	5.7	12	A
BisAur 68-1	11.3	9	A
BraAuc 60-1	8.1	9	A
Control: *N. benthamina* leaves unmodified			
Sample 1	0.9	12	B
Sample 2	1	9	B

Table 1: The EFR LRR orthologs are functional. Production of reactive oxygen species (ROS) in *N. benthamiana* leaf discs expressing the chimeric EFRs after treatment with elf18 and mock, respectively. Unmodified *N. benthamiana* leaf discs served as control. All chimeric EFRs were functional: ROS production after elf18 treatment was much higher than after mock treatment (response >> 1 and group A). In contrast unmodified *N. benthamiana* leaf discs did not show a difference in response to elf18 and mock (response ≈ 1 and group B).

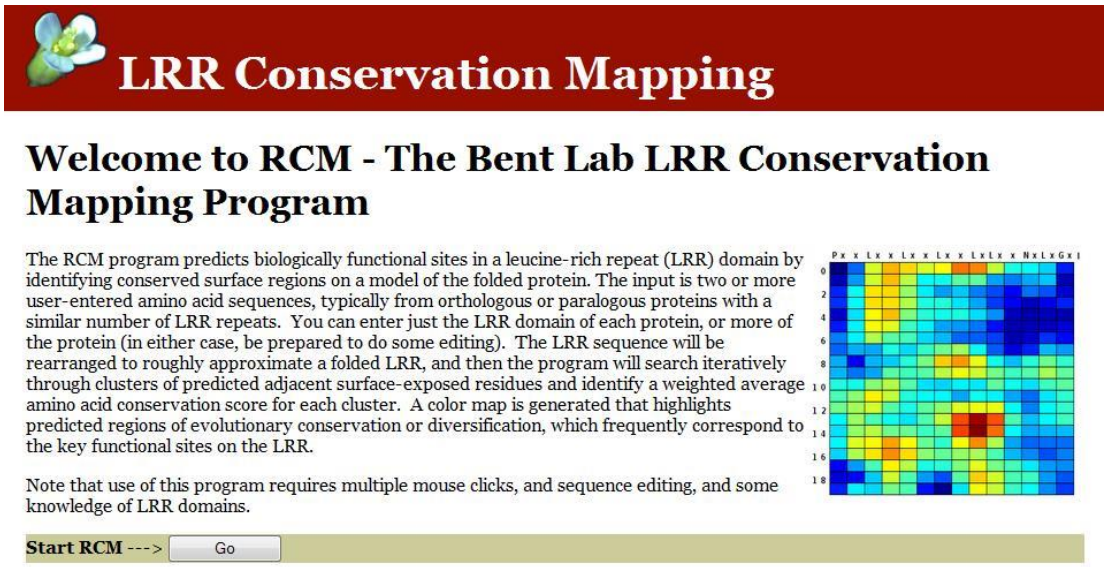

Figure 2: The LRR Conservation Mapping Program. The program is available online at www.plantpath.wisc.edu/RCM. It generates heat maps of the repeat domains using a color scale ranging from blue (non-conserved) to red (conserved).

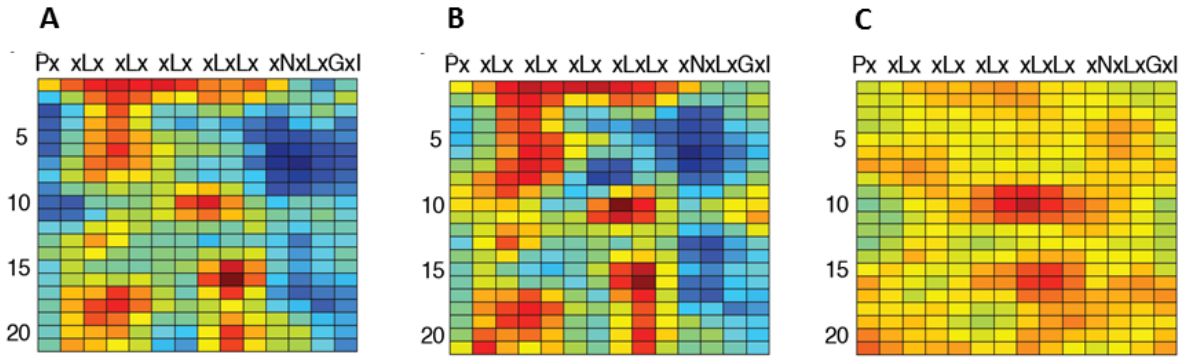

Figure 3: Conservation mapping of EFR reveals potential functional sites of the LRR domain. **(A)** RCM map for ten Brassicaceae EFR LRR orthologs. **(B)** RCM map of two EFR LRR orthologs, from distantly related *Eruca sativa* and *Erysimum raulinii*. **(C)** RCM map of two EFR LRR orthologs, both from *Erysimum raulinii*. Each box represents one amino acid on the predicted surface of the LRR domain, each line represents one LRR. Heat map color scale ranges from blue (non-conserved) to red (conserved).

Published Research Article

PLOS ONE 2011

LRR Conservation Mapping to predict functional sites within protein leucine-rich repeat domains

Laura Helft[1], Vignyan Reddy[2], Xiyang Chen[3¤], Teresa Koller[3], Luca Federici[4], Juan Fernández-Recio[5], Rishabh Gupta,[2] Andrew Bent[1,3*]

[1] Program in Cellular and Molecular Biology, [2] Department of Electrical and Computer Engineering, [3] Department of Plant Pathology, University of Wisconsin - Madison, Madison, WI USA, [4] Dipartimento di Scienze Biomediche, Università 'G. d'Annunzio', Chieti, Italy, [5] Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain

* Corresponding Author: afbent@wisc.edu

¤ Current Address of XC: Department of Computer Science, Purdue University, West Lafayette, IN USA

Contribution Teresa Koller: Isolation and functional testing of all Brassicaceae *EFR* sequences except for *AtEFR* and *BnEFR1*. The sequences were used to generate Figures 8, 9 and 10.

Abstract

Computational prediction of protein functional sites can be a critical first step for analysis of large or complex proteins. Contemporary methods often require several homologous sequences and/or a known protein structure, but these resources are not available for many proteins. Leucine-rich repeats (LRRs) are ligand interaction domains found in numerous proteins across all taxonomic kingdoms, including immune system receptors in plants and animals. We devised Repeat Conservation Mapping (RCM), a computational method that predicts functional sites of LRR domains. RCM utilizes two or more homologous sequences and a generic representation of the LRR structure to identify conserved or diversified patches of amino acids on the predicted surface of the LRR. RCM was validated using solved LRR+ligand structures from multiple taxa, identifying ligand interaction sites. RCM was then used for *de novo* dissection of two plant microbe-associated molecular pattern (MAMP) receptors, EF-TU RECEPTOR (EFR) and FLAGELLIN-SENSING 2 (FLS2). *In vivo* testing of *Arabidopsis thaliana* EFR and FLS2 receptors mutagenized at sites identified by RCM demonstrated previously unknown functional sites. The RCM predictions for EFR, FLS2 and a third plant LRR protein, PGIP, compared favorably to predictions from ODA (optimal docking area), Consurf, and PAML (positive selection) analyses, but RCM also made valid functional site predictions not available from these other bioinformatic approaches. RCM analyses can be conducted with any LRR-containing proteins at www.plantpath.wisc.edu/RCM, and the approach should be modifiable for use with other types of repeat protein domains.

Introduction

The conservative nature of evolution causes selection of stable structures that nevertheless are modifiable for operation in varied processes. Proteins carrying repetitive domains such as leucine rich repeats (LRRs), ankyrin repeats or tetratricopeptide repeats are one common solution to these evolutionary demands [1-3]. A single class of repeat domain can interact with a wide array of chemically distinct ligands, yet each particular repeat protein shows high specificity for particular ligands. Consensus amino acid motifs have been identified for these repeat domains [1-5]. The consensus amino acids within a particular type of repeat provide a regular, stable scaffold to the domain, while non-consensus residues within the repeat allow variability in function [1]. The characteristic structure formed by a particular type of repetitive motif is identifiable by comparison of multiple solved protein structures, and can be used to predict the overall configuration of other protein domains sharing that repetitive motif. However, methods to query the variable portions of these repeat domains, to identify and understand the sites that control specialized functions, are less well developed.

LRR domains are a protein-ligand interaction domain found in many types of prokaryotic, eukaryotic, and viral proteins, including ubiquitin ligases, hormone receptors, enzyme inhibitors, and immune receptors in plants and animals [1,6-12]. As annotated by Pfam, more than 500 different proteins encoded by the human genome contain LRRs, and there are over 1000 types of LRR-containing proteins in individual plant species such as *Arabidopsis thaliana* or rice [4]. The ubiquity of this domain may be due to its ability to interact with a wide range of substrates including proteins, nucleic acids, lipids, and small molecule hormones. It is

particularly compelling that in jawless vertebrate adaptive immune systems, antibodies are produced by shuffling hypervariable LRR repeats [13,14]. The resultant receptors recognize a wide range of substrates with high affinity. Any single LRR domain can potentially interact with several different molecules, either simultaneously or asynchronously.

There are multiple sub-families of LRR domain types, each with slightly different consensus amino acid motifs, but all share the tendency to form a solenoid structure with approximately 20-30 amino acids per repeat where each repeat forms one turn of the helix (Figure 1A; an example of an LRR consensus motif is xxLxxLxxLxxLxxNxLxGxIP). The solenoid is curved so that a convex and a concave face are created. The concave face is largely composed of β-strands, and often forms the ligand binding site for LRR domains [15]. Deviations from the repeat consensus can result in intervening segments that interrupt adjacent solenoid regions, providing structural flexibility to the LRR. An LRR domain may be relatively brief (e.g. two or three repeats), or quite long (30 repeats or more). The consensus residues (mostly leucine or other hydrophobic amino acids) are usually buried in the core of the solenoid, while residues at the variable positions are predominantly solvent exposed (Figure 1A). The LRR structure efficiently creates a large surface-to-volume ratio protein domain that tolerates a wide variety of surface compositions, encoded in a condensed genomic space [16].

LRRs play a central role in the receptors that mediate two major branches of the plant immune system [17]. Many plant cell surface receptors for microbe- or pathogen-associated molecular pattern (MAMP/PAMP) molecules, which confer recognition of conserved microbial molecular motifs, contain LRRs as the bulk of their extracellular domain and a protein kinase

intracellular domain [18-20]. The transmembrane LRR-kinase family of receptors is expansive in plants (over 200 different protein types in *Arabidopsis*, over 400 in rice [21]), and their functions extend well beyond immunity to include prominent roles in plant growth and development pathways [22,23]. Plant MAMP receptors carry clear structural and functional analogies to (but apparent evolutionary independence from) animal MAMP receptors, the Toll-like receptors (TLRs), which also contain LRR domains [24]. LRRs are also central to recognition specificity in the large, diverse family of plant intracellular nucleotide binding (NB)-LRR proteins known as resistance (or "R") proteins. These R proteins initiate strong defense responses upon recognition of specific pathogen effector molecules or upon recognition of a host protein alteration caused by a specific pathogen effector [11]. Plant R proteins are structurally similar to animal nucleotide-binding leucine-rich repeat (NLR) proteins that play significant recognition roles in the mammalian immune response [25-28].

Two plant LRR-kinase MAMP receptors that have been a particular focal point for research are ELONGATION FACTOR TU (EF-Tu) RECEPTOR (EFR) and the flagellin receptor FLAGELLIN-SENSING 2 (FLS2) [29-33]. FLS2 orthologs can be found in a wide range of monocotyledonous and dicotyledonous plant species, whereas EFR appears to be limited to the mustard family Brassicaceae. FLS2 can be activated by flagellin or by synthetic peptides that represent the highly conserved minimal recognition domain of flagellin, such as the 22 amino-acid flg22 peptide [29]. Similarly, EFR responds to conserved peptides from the recognized domain of bacterial EF-Tu, such as the elf18 peptide [32]. There is evidence that the LRR domains of these proteins directly interact with MAMP ligands [32,34-36], but the large size of their LRR domains (22 repeats for EFR, 28 for FLS2) leaves open the possibility that these LRRs

also mediate interaction with other ligands, co-receptor proteins and/or cofactors. FLS2, EFR and other MAMP receptors can be significant barriers to microbial infection [32,37-42]. Intriguingly, transgenic tomato plants expressing *Arabidopsis* EFR are more resistant to the plant pathogenic bacteria *Ralstonia solanacearum* and *Xanthomonas campestris* pv. *vesicatoria*, implying that many of the mechanisms for downstream defense signaling from MAMP receptors are conserved across diverse plant species [38,43].

Protein functional site prediction is already possible using a number of computational methods, each of which has certain advantages and limitations. Functional sites are often detectable as the sets of amino acids that have been most conserved or diversified among a set of homologous proteins [44-46]. Some computational methods require only primary sequences to look for conserved codons or amino acids within a sequence alignment, such as database searching for protein motifs (e.g., [47]) or alignment with homologous sequences (e.g., CLUSTAL). In contrast, positive/purifying selection analysis (e.g., Ka/Ks or dN/dS ratio; [46]) looks for amino acids that have undergone selection, either purifying or diversifying (positive), based on identification of non-neutral, selected substitutions at the nucleotide level among a group of homologous sequences. However, the above methods do not identify functional groupings of residues that are nearby in the folded protein but dispersed in the primary amino acid sequence. Some studies of positive selection on LRRs have manually approximated structural proximity to improve their power (e.g.,[48]). Other computational approaches, such as Consurf [49], optimal docking area (ODA) [50], and conserved functional group (CFG) analysis [45], do use protein structural information obtained experimentally (X-ray or NMR) or by homology modeling. CFG and Consurf search homologous input sequences for conserved

groupings of amino acids on the surface of the folded protein. ODA models desolvation energy, i.e. it searches for continuous surface patches that undergo favorable energy change when buried during a modeled protein-protein association; a low ODA value indicates a location that is predicted to interact with a ligand or another protein. These latter programs offer valuable paradigms but are limited to proteins with a crystal structure or a reliable homology model, and CFG and Consurf have constraints related to handling of hydrophobic and/or repeat motif residues that limit their efficacy for analysis of LRR domains.

In hundreds of important proteins, LRR domains contain the ligand specificity region or other functional sites, and there is significant interest in identification and manipulation of these sites. In the present study we developed the Repeat Conservation Mapping (RCM) program that predicts functional sites in LRR domains by identifying, among a group of homologous LRR-containing proteins, the patches of predicted spatially adjacent residues on the surface of the LRR that exhibit the greatest conservation or greatest divergence. The RCM method utilizes linear amino acid sequences as input, along with existing generalized LRR structure principles, to predict likely adjacency of residues in the LRR. It then identifies regional conservation scores for predicted surface residues based on conservation of that residue and its proximal surface amino acids. We validate the method using previously solved co-crystal structures for LRR with ligand, and through discovery and *in vivo* validation of previously unknown functional sites of the plant EFR and FLS2 MAMP receptors, for which structures are not currently available. The RCM program can be run using the publicly accessible web server at http://www.plantpath.wisc.edu/RCM, and the source code is openly available via a GNU general public license.

Methods

Repeat Conservation Mapping program

RCM analyses can be conducted at a publicly accessible web interface served from a Linux-based virtual machine. The RCM web site runs a set of linked php files that draw upon PHP, Perl, HTML, Python and C scripts, including local implementations of functions from ClustalW2 [51,52], LRRScan [53] and MATLAB (The Mathworks, Inc., Natick, MA). The code is available under GNU general public license v3; local installation is not recommended for most end-users.

Figure 1B and the following briefly describe the method. The linear amino acid sequences of two or more LRR domains are aligned and compared, and a conservation score is determined for each amino acid position. Characteristic properties of LRR domains are then used to generate a generic super-helical structural model of the LRR domain being queried, which places amino acids in their likely relative locations to each other in a folded protein. A sliding window analysis then determines a center-weighted conservation score for all possible groups of adjacent amino acids that are predicted to reside on the surface of the helix. Typically a 5 x 5 matrix of 25 amino acids, spanning 5 LRR repeats, is queried to derive each regional conservation score. A colored heat-map of the result directs investigator attention to the most conserved (or most divergent) regions. Alternative score-weighting systems for the sliding window analysis are available to users; adjustment of these values significantly impacts the resulting conservation map. The default weighting values achieve a useful balance between the excessive smoothing (loss of resolution) that occurs with less center-weighting, and the loss

of site identification (similarity to the initially calculated individual residue conservation scores) that occurs with greater center-weighting. Readers are referred to a more detailed step-by-step description of conservation mapping in Text S1.

For positive selection analysis, sequence alignment and phylogenetic trees were constructed using MEGA 4.0 [54]; information from these trees were used to calculate positive selection using the CODEML module in PAML [46]. Optimal docking analysis (ODA) was implemented as described in [55]. The current web module of Consurf was used [49].

Homology Modeling

The FLS2 homology model was obtained using the structure of PGIP2 from Phaseolus vulgaris as a template (PDB ID 1OGQ) [56]. Both proteins (as well as EFR) belong to the plant extracellular LRR protein subfamily [1] characterized by the same consensus sequence in the LRR domain, i.e. xxLxLxxNxLt/sGxIPxxLxxLxxL. Furthermore, the LRR domain is capped at the N-terminus and at the C-terminus by two small cysteine-rich domains, which are also evolutionarily conserved among PGIP and the RLK receptors FLS2 and EFR [57]. However, PGIP2 contains 10 LRRs matching the above consensus while FLS2 is characterized by 28 complete repeats, hence four separate alignments were manually prepared and used for homology modeling. The N-terminal and first 9 LRR repeats of PGIP2 were manually aligned to the N-terminal and first 9 LRR repeats of FLS2. The PGIP LRR domain was separately aligned to FLS2 repeats 7-15 and 13-21. A final alignment encompassed FLS2 repeats 20-28 plus the C-terminal flanking region. These alignments were used, together with the relevant PGIP2 coordinates, to obtain four independent partial models using Modeller, version 9.1 [58]. Twenty models were

obtained from each alignment and the lowest energy models were selected according to the Modeller objective function. The four partial models obtained by this strategy are partially overlapping with two or three LRRs in common between consecutive ones. This allowed superposition of the models in their overlapping regions and merging of coordinates to obtain a single full-length model using the SSM algorithm as implemented in the program COOT [59]. This model was further energy minimized within Modeller to obtain a final model that encompasses residues 1 to 744 of FLS2. The geometrical quality of the model was very good as judged with PROCHECK [60] with 98.1 % of residues lying in allowed regions of the Ramachandran plot, 1.2 % in generously allowed regions and only 0.6% (4 residues) in the non allowed regions. The EFR LRR domain was aligned to FLS2 using ClustalW and the alignment was manually adjusted, when necessary, to match the plant extracellular LRR consensus sequence. An EFR homology model, encompassing residues 1-576, was calculated and energy minimized within Modeller using the FLS2 homology model as a template. The final model has a very good geometry with 96.8 % residues in allowed regions of the Ramachandran plot, 2.4 % in generously allowed regions and 0.8 % (four residues) in non-allowed regions.

EFR and FLS2 Constructs

The protein coding sequence of *Arabidopsis thaliana EFR* up to but not including the stop codon, along with native promoter sequence (1091 bp upstream of the start codon = 1074 bp upstream of the transcription start site), was amplified from Col-0 accession genomic DNA using the primers (CACCGGGTTTTTGTTTATTCAAAGATGGG and CATAGTATGCATGTCCGTATTTAACATCC) and cloned into pENTR/D-TOPO (Invitrogen).

Mutations to *EFR* were made in this construct via PCR as in [35], using mutagenic primers and a high-fidelity DNA polymerase (typically Pfu Ultra II (Stratagene)), followed by DpnI treatment to digest template and transformation of the linear product into *E. coli*. Mutations were verified by DNA sequencing. Site-directed randomizing mutagenesis was performed as described in [35] using mutagenic ~30 nt PCR primers in which only one codon was mutagenized using the degenerate codon NNB. Similarly, double-alanine mutants were created with a mutagenic primer of ~30 nt with two selected codons mutated to alanine. To clone BrEFR1, primers (CACCATGAAGCCGTTTCTTTCAATTGCTCTACTCATG and CATTGTATGCATGTCCGCGCTTAACATCC) were designed to amplify the full-length best hit of *Arabidopsis* EFR, minus stop codon for fusion to C-terminal tag, for insertion into pENTR/D-TOPO. To clone *EFR* from other species, genomic DNA was extracted and the *EFR* LRR-encoding domain was amplified using primers based on the *EFR* sequences of *Arabidopsis* and *Brassica rapa*. Restriction enzyme sites (SbfI and FseI) flanking the LRR domain were engineered by site-directed mutagenesis into the *Arabidopsis EFR* promoter+coding region in the pENTR/D-TOPO vector. The *Arabidopsis EFR* LRR-encoding domain was then cut out of the vector. The LRR-encoding domains of the Brassicaceae *EFR* genes were amplified using species-specific primers with the restriction enzyme sites at the 5'-ends and cloned into TOPO vectors, then cut out of the TOPO vectors and ligated into the *Arabidopsis EFR* gene lacking the LRR domain. Sequences in pENTR/D-TOPO were moved, by LR Clonase II reaction (Invitrogen), into the Gateway vector pGWB13 (if using native EFR promoter) or pGWB14 (if using CaMV 35S promoter), fusing the EFR amino acid sequence to a C-terminal HA tag [61]. Constructs were then moved into *Agrobacerium tumefaciens GV3101* by electroporation and used to transform homozygous *Arabidopsis efr* -

plants (SALK 068675c) using the floral dip method or for transient expression in *Nicotiana benthamiana* via *Agrobacterium* infiltration. Site-directed randomizing mutagenesis of FLS2 was as described in [35]. Constructs were transformed into homozygous *fls2-101* plants. All new DNA and derived amino acid sequences are deposited at Genbank under accession numbers JN002095-JN002103.

Receptor function assays

Seedling growth inhibition assays were performed as described [35], with appropriate selection for transgenic seedlings prior to use. ROS assays were performed on leaf discs taken from 4- to 8-week old transgenic *Arabidopsis* or from 4- to 6-week old *N. benthamiana* leaves infiltrated two days prior with *A. tumefaciens* containing an *EFR*, *FLS2* or corresponding empty vector construct. Leaf discs were floated on 1% DMSO overnight and then treated with 1μM peptide (or no peptide in the case of mock) in the presence of 1μg/mL luminol and 1μg/mL horseradish peroxidase. Luminescence was measured on a Synergy HT Microplate Reader (Bio-Tek) for 30 minutes following addition of peptide [32,62,63]. For callose deposition assays, seedlings were grown for 5 days on 0.5x MS agar and then transferred to liquid 0.5x MS (500 μl per well in 24-well plate) with elf18 or flg22 peptide at the indicated concentrations. After 24 hours in liquid, seedlings were fixed with 2% formaldehyde/5% acetic acid/60% ethanol (FAA), cleared overnight in 95% ethanol, stained with 0.01% aniline blue and viewed under an epifluorescence microscope to visualize callose deposits [62,63]. For receptor protein detection, six to eight 3-week old seedlings were ground in 2x SDS buffer (2mL buffer per g tissue), boiled, and centrifuged to remove particulates. 50 μl per sample was separated by SDS-

PAGE, blotted onto a PVDF membrane, and detected using an anti-HA antibody conjugated to horseradish peroxidase (Roche), made visible using the ECL Plus Kit (Amersham).

Results

Rationale and description of LRR conservation mapping approach

LRRs have a regular structure in which a single repeat forms one turn within the overall super-helical structure. Across numerous solved LRR structures, the LRR consensus residues form the buried core of this configuration (e.g., Figure 1A) [1]. Because of the regularity of these LRR structures, two assumptions can be made for conservation mapping: 1) consensus residues are not on the protein surface; and 2) the repetitive structure of LRRs allows prediction of relative amino acid positions in the tertiary structure without requiring a crystal structure or a detailed homology model of the protein in question. Assumption 1 allows elimination of the highly conserved but functionally less revealing consensus residues from the analysis; assumption 2 allows prediction and assessment of spatially adjacent groups of surface residues that are not adjacent in the primary amino acid sequence. Following this rationale, we devised and implemented Repeat Conservation Mapping (RCM), a set of algorithms to identify predicted functional sites of LRR domains. RCM accomplishes this by identifying the extent of conservation of different amino acid patches on the predicted surface of LRR domains (see also program description in Text S1).

As an example, the lower-right element of Figure 1B shows RCM output for an extensively studied plant LRR domain-containing protein, POLYGALACTURONASE INHIBITOR PROTEIN (PGIP). PGIP was the first LRR-containing plant protein to have its structure solved [56]. In bean (*Phaseolis vulgaris*), there are four PGIPs, designated PGIP1-4, which have varied but overlapping specificities for different polygalacturonases (PGs) [64]. The RCM map for these four paralogs highlights several divergent and conserved patches (Figure 1B). Interestingly, the three positions in PvPGIP2 that, when mutated to alanine, have a significant negative impact on inhibitor activity [55], are all located in divergent patches identified by conservation mapping, perhaps indicating that these residues are also responsible for differences in receptor specificity. Indeed, a single amino acid that can switch specificity between PGIP1 and PGIP2 is also found in this divergent region [65].

Validation: Conservation mapping highlights functionally significant regions of LRR domains

RCM highlights, among a group of homologous proteins, regions that are highly conserved or highly divergent on the surface of the LRR domain. To verify that RCM identifies significant functional sites, we utilized RCM to analyze all proteins in PDB that, as of October 2010, had a structure for an LRR domain interacting with another protein or a ligand, and for which at least one functional homolog could be identified. Two examples, RIBONUCLEASE INHIBITOR (RI), and TRANSPORT INHIBITOR-RESPONSIVE 1 (TIR1) are discussed here; the maps and analyses of the other nine protein groups are provided as Figure S1. For all of these RCM analyses, proteins were compared to at least three high-scoring homologs identified through

BLAST and/or literature searches. An amino acid was considered to be interacting with ligand if identified in the article(s) accompanying the crystal structure(s) [66-76].

RI is one of the most extensively studied LRR-containing proteins, with many reported crystal structures and mutagenesis studies (for a review, see [77]). RI prevents ribonucleases (RNases) from acting by binding to their catalytic domains with femtomolar affinities. There are two solved structures of RI bound to ribonucleases, human RI complexed with angiogenin, and porcine RI complexed with RNaseA. Binding of angiogenin also induced homodimerization of RI. These two RIs, along with RI from rat and mouse, were used to generate the conservation map in Figure 2A. Residues involved in any of the three interactions (hRI-angiogenin, sRI-RNAse A, hRI-hRI) are highlighted in Figure 2A with asterisks. Although many ligand contact sites could be seen in the crystal structure, kinetic analyes of mutants in these regions demonstrated that mutation of the locations identified by RCM as most conserved, that is, the residues in the β-strand, β-turn region of LRRs 10, 12, and 14-16, have the largest impact on ligand binding [77].

TIR1 is an auxin receptor in *Arabidopsis* and other plants; it ubiquitinates target proteins when it binds auxin hormones in a pocket formed by the LRR domain of the protein and the cofactor inositol-6 phosphate ($InsP_6$). To investigate recognition of auxin by these receptors, we constructed a conservation map of *Arabidopsis* TIR1 with its five paralogs. In this case, the five *Arabidopsis* paralogs (AUXIN SIGNALNG F-BOX (AFB) 1-5) have all been implicated in auxin signaling [78-81]. Again, RCM successfully highlighted two patches on the surface of the LRR (Figure 2B): one where TIR1 binds auxin, and a second site where TIR1 binds auxin as well as the cofactor $InsP_6$ [68]. For comparison, it can be valuable to see the individual residue

conservation scores for the RI and TIR1 protein sets (Figure S2), which are obtained after alignment of primary amino acid sequences, but before sliding-window calculation of regional conservation scores that are shown in Figure 2. Visual inspection of the RCM maps for the other validation proteins (Figure S1) again indicates successful identification by RCM of sites involved in LRR+ligand interactions, in follicle stimulating hormone receptor, glycoprotein 1b alpha, Skp2, Slit, and TLR3. Poor success was obtained for TLR1, TLR2, TLR4 and TLR6.

However, the ligand specificity is not known for many of the homolog sequences that were used, and they may not have been appropriate proteins to compare via RCM (see Discussion).

Continuing with the above validation, combined data were analyzed for all conservation maps generated for the above proteins with known LRR+ligand structures. The regional conservation scores for residue positions known to be involved in LRR+ligand interactions were significantly higher than the set of all other residue scores in each RCM map generated (Student's T-test, p-value 0.005 or less). We also ranked the scores for each map into deciles and then determined the distribution of regional conservation scores, and individual residue conservation scores, for known LRR+ligand contact positions. Despite the presence of many potentially misleading TLR comparisons in the dataset, the distribution of scores for LRR+ligand interaction residues is weighted towards the highest ten percent of the RCM scores on their respective maps (Figure 3). Importantly, the proportion of scores in this decile increases when weighted pairwise comparisons are utilized (see Step 4 in Methods), and the proportion increases further when regional conservation scores are used (from 18% to 20% to 25% of

interacting residues, respectively; Figure 3). 45% of all interacting residues appeared in the top 20% of RCM scores in this analysis. The remaining high-scoring sites identified by RCM may be involved in other functional processes that are not detected in the available LRR+ligand crystal structures, such as interaction with other ligands, cofactors or co-receptor proteins (see Discussion). Comparison of homologous proteins with diversified functions is also addressed below.

Mutations of AtEFR at conserved locations identified by RCM often disrupt receptor function

RCM was used as a *de novo* discovery tool in experiments on EFR, the EF-Tu receptor of plants, for which there is no known structure or ligand binding site. To identify an EFR homolog for comparison, the derived amino acid sequence of *Arabidopsis* EFR was used with BLASTP to query all publicly available sequences from the Brassica Genome Gateway as of April 2009, and two high-scoring matches from *Brassica rapa* were identified. One of these sequences was cloned and then transiently expressed in *Nicotiana benthamiana* leaf mesophyll tissues, which do not otherwise respond to the EF-Tu-based elf18 peptide [32]. The defense mechanisms initiated by EFR or FLS2 include production of reactive oxygen species, release of other directly antimicrobial compounds, and production of callose and lignin as part of a cell wall strengthening response that limits pathogen penetration [82]. Plant seedlings undergoing chronic defense activation display inhibition of growth after multiple days of exposure to a recognized MAMP; this is a sensitive and widely used assay for FLS2 or EFR activation [31,32]. The cloned *Brassica rapa EFR* homolog, designated Br*EFR1*, conferred recognition of 1uM elf18

in an ROS assay (Figure S3). Br*EFR1* subsequently was shown to complement an *Arabidopsis ef*$^{-}$ mutant, rescuing the ability of the plants to respond to elf18 in ROS and seedling growth inhibition assays (Figure S3 and data not shown).

RCM was used to create a conservation map of AtEFR with BrEFR1 (Figure 4A). The map highlights several small patches of conservation on the surface of EFR, the highest scoring (most conserved) of which appear on the concave face of the receptor. The convex face also contains patches of conservation (the third and fourth columns of Figure 4A). There is a large patch of divergence in LRRs 3 through 10.

Hypothesizing that the ortholog map predicts functionally significant locations, we constructed a series of double-alanine mutants within EFR along the predicted concave face (also known as the β-strand, β-turn region [1]). For each construct, we changed two of the five variable residues within a single repeat's β-strand, β-turn region to alanine. Alleles were made that mutate sites of conservation, or as controls, sites lacking conservation. The double-alanine mutants were tested in stable transgenic *Arabidopsis efr*$^{-}$ plants. Of the fourteen constructs tested in a seedling growth inhibition assay, four of them had no detectable response to peptide, and four constructs had a response that was weaker than a wild-type response (Figure 5). Seven of these mutants were in conserved locations on the RCM map made from the EFR orthologs, including each of the four conserved patches of the β-strand, β-turn region (Figures 4, 5). The other two alleles with mutations in conserved locations did not detectably alter function. Importantly, only one of the five mutants in any of the poorly conserved regions resulted in a discernible difference in receptor function, with response slightly less than a wild-

type receptor (Figures 4, 5). These results were further supported by ROS assays in *N. benthamiana* (not shown) and ROS and callose assays in *Arabidopsis* (Figure S4). All of the EFR double-alanine mutant proteins, whether functional or not, were still present at functional levels in plants, as detected by Western blot (Figure S5).

Conservation mapping of EFR and its paralogs reveals a divergent β-strand, β-turn region

Use of paralogs rather than orthologs was further examined. The four *Arabidopsis* paralogs most similar to EFR were identified by BLAST using the *Arabidopsis* EFR LRR sequence to query the *Arabidopsis* genome. An RCM map of EFR and these paralogs was then generated (Figure 4B). Some areas are generally conserved across these paralogs, in particular along one 'shoulder' of the convex face of the LRR that is also conserved among EFR orthologs. However, a large portion of the concave face of the LRR region is highly divergent in this paralog map. We hypothesized that the divergent β-strand, β-turn region is responsible for recognizing distinct ligands in these receptors, since 1) the concave face of LRR regions are most often implicated in ligand binding [15]; and 2) knock-out of *EFR* results in a plant completely insensitive to elf18 [32], implying that these EFR paralogs are not capable of EF-Tu recognition. Based on this map, we performed site-directed randomizing mutagenesis on *EFR* to create libraries of concave face mutations at amino acid positions that are highly divergent (predicted to impact function), or conserved (also predicted to impact function), or neither highly conserved nor highly divergent (not predicted to impact function). Each allele library contains different mutations at a single position. These libraries were introduced into *Arabidopsis efr*$^{-}$

plants that were then tested for response to 100nM elf18 in a seedling growth inhibition assay. Most libraries were significantly impacted in their ability to recognize 100nM elf18 (as compared to a wild-type receptor response) regardless of level of conservation, but all libraries still had several functional clones (Figure 6). The most significantly impacted library (LRR 2.3 (Asn 103)) was also identified in the double-alanine mutagenesis and is conserved in both RCM analyses performed (Figure 4). Overall, mutagenesis of regions conserved among EFR orthologs often broke function while mutagenesis of regions divergent among EFR paralogs had a less severe impact.

Conservation mapping highlights an additional region of functional significance in AtFLS2

RCM was further tested as a discovery tool through work with a second protein, FLS2. FLS2 has been deeply studied but there is no solved structure or definitively determined extracellular ligand binding site [33]. To identify additional important regions of the FLS2 LRR we created various RCM maps of AtFLS2, for example in comparison to eight Brassicaceae orthologs from plants that respond to flg22 (Figure 7A; see also [35]). Within the β-strand/β-turn region there were two main areas of conservation. One of these was located at LRRs 9-13, in agreement with previous findings that this is an important region for flg22 binding and recognition [35]. Another conserved area was persistently observed at LRRs 22-26. To investigate their role in flg22 perception, five solvent-exposed residues in this region were mutagenized by site-directed randomizing mutagenesis. These libraries were used to transform mutant *Arabidopsis fls2-101* plants that were then tested for response to flg22 peptide in a

seedling growth inhibition assay. Alteration of Y629 significantly impacted flg22 responsiveness, and alteration of S633 also caused a detectable depletion of FLS2 activity (Figure 7B).

RCM using many or few EFR and FLS2 orthologs, from closely or distantly related species

Changes in RCM output for a given protein, in response to varied types of input, was further examined after isolation of additional EFR orthologs from other Brassicaceae. EFR homologs were isolated by PCR from plant accessions that exhibited a response to elf18. A function in EF-Tu sensing was confirmed for nine homologous LRR domains (one each from *Brassica aucheri, Brassica rapa, Brassica napus, Eruca sativa, Biscutella auriculata* and two *EFR* sequences each from *Enarthrocarpus arcuatus* and *Erysimum raulinii*), using ROS assays for responsiveness to elf18 after transient expression in *N. benthamiana* leaves (data not shown). These confirmed homologs and the EFR sequence from *Arabidopsis thaliana* Col-0 were analyzed by RCM. The resulting map (Figure 8A) is notably similar to the map generated with only two EFR sequences (Figure 4). However, the map generated with ten sequences more clearly delineates the most consistently conserved clusters, which included the concave face regions confirmed to be required for EFR function (Figures 4-6).

Additional RCM maps were generated using smaller subsets of these EFR proteins based on the overall relatedness of different Brassicaceae species [83]. One analysis used EFR sequences from two of the most distantly related species of our sample set, *E. sativa* and *E. raulinii* (Figure 8B), while another analysis used two different EFR sequences obtained from a

single *E. raulinii* plant (Figure 8C). A similar exercise was performed using FLS2 sequences from seven closely or distantly related Brassicaceae species (Figure S6). The maps of functionally similar LRRs from very closely related sequences (Figures 8 and S6) emphasize amino acid clusters containing less common amino acid residues that score highly in Blosum matrices, whether or not they reside at functional sites. The maps made from seven or ten functionally confirmed sequences more reliably emphasize evolutionary conservation of residue clusters. These latter maps also offer better resolution of conserved and diversified areas, but maps made from just two sequences from more distantly related species were sufficient to highlight the main features also present on the maps made from larger numbers of proteins (Figures 8 and S6, see also Figure 4). As with the map of *Arabidopsis* EFR and its paralogs, these ortholog maps also highlight regions outside of the concave face that have been conserved, presumably due to their functional significance.

Comparison of conservation mapping to other computational methods

Other methods that predict important amino acid residues within a protein, beyond the BLAST or Pfam-like methods that identify conserved motifs in the primary sequence, include positive/purifying selection analysis [46], optimal docking area (ODA) calculation [55], and Consurf [49]. Because these three approaches employ different criteria for identifying functionally significant residues, we searched for important residues of EFR, FLS2 and PGIP using these three methods and compared their results to those generated by RCM (Figures 9 and S7). When a method required multiple sequences as input, we used the same sequences

for each analysis (the Brassicaceae orthologs described above for EFR and FLS2, or the Fabaceae PGIP sequences from [55]). Homology models of EFR and FLS2 were created using the crystal structure of PGIP as template [56]. To allow comparison, the RCM results of Figure 4A are also presented on a homology model of EFR (Figure 9D).

As previously reported, integration of positive selection analysis and ODA identified several residues on the surface of PGIP that were important for specificity toward different PG's [55]. Unsurprisingly, most surface residues of EFR and FLS2, when compared to functional Brassicacea orthologs, were under purifying selection ($-\log(\omega)>0$). Only two residues of the LRR of EFR showed evidence of positive selection, neither of which is predicted to be on the concave face of the receptor (although both are on the surface of the protein) (Figure 9A). Similarly, two residues of the LRR of FLS2 are under positive selection, predicted to be on the side of the LRR and solvent-exposed (Figure S7).

Using ODA, a patch of potential ligand binding residues for EFR was identified on the β-strand, β-turn region of LRRs 17-18 (Figure 9B), overlapping with a region strongly identified by RCM. Two patches of predicted binding sites were detected for FLS2: a patch three to five surface residues wide in LRRs 1-5 occurring partially on the convex face and partially on the side of the LRR; and a patch primarily in the β-strand region of LRRs 10-15 (Figure S7). This second patch was previously shown to be important for flg22 recognition [35]. Consurf analysis of EFR highlighted many conserved areas; the most highly conserved residues were found in the second and third columns of the surface of the repeats as shown in Figure 4, and in the β-strand β-turn region (Figure 9C). Because the similarity of the Brassicacea FLS2 sequences is high

(greater than 80%), Consurf ranked almost all residues as either highly conserved or highly divergent (Figure S7).

To investigate overlap between RCM output and outputs from the other computational methods, we looked for a correlation between our scoring method and positive selection analysis, ODA, and Consurf analysis (Figure 10). No strong correlation between any of the three computational methods and RCM was observed ($R^2 < 0.80$). However, all methods could identify important functional residues. T-tests for differences between the set of scores for residues tested by mutation in which function was disrupted (Figure 10, black symbols), in comparison to the scores for residues for which function was retained after mutation (Figure 10, white symbols), revealed significant differences ($p < 0.05$) for RCM, ODA, Consurf and positive selection analyses. However, of all the methods, RCM had the most significant difference ($p<.0001$). The ability of RCM to perform at least as well as other contemporary, well-accepted computational methods, in conjunction with its need for only two homologous sequences and no homology model, highlights the utility of this method.

Discussion

The RCM method identifies regions on the predicted surface of LRR domains that have been conserved over the course of evolution. Conserved regions on the surface of folded proteins often correspond to key functional sites such as ligand binding sites or enzyme catalytic sites [44,45]. In many cases where solved LRR+ligand crystal structures and appropriate homologs were available, RCM correctly identified the ligand binding regions of

LRR domains. RCM also enabled discovery of previously unknown functional sites on the surface of EFR and FLS2. A strong trend was observed when RCM was used to direct mutational studies: significant impacts on function were frequently observed for mutations in regions that RCM predicted as conserved, while impacts were rare for mutations in non-conserved regions. This demonstrates the utility of RCM for *de novo* prediction of functional sites.

The RCM analyses of proteins for which there are LRR+ligand crystal structures demonstrated that conservation mapping can work for orthologs or paralogs, as long as they share similar functions. For example, creating a conservation map of RI orthologs highlighted functionally important regions previously demonstrated to be important for common function among these proteins. Mapping TIR1 and its *Arabidopsis* AFB1-5 paralogs, as well as PGIP1-4, demonstrated the utility of RCM for comparisons of functionally related paralogs. TIR1 functional sites again mapped to conserved regions of the protein. Irregularities in an LRR domain must be considered when generating and interpreting the results of RCM maps. Still, RCM performed well over irregular loop regions. In TIR1, for example, a loop region in the second LRR is involved in binding auxin, and RCM correctly identified this region as a functional site.

In contrast to comparison of orthologs, functionally important sites of PGIP could be seen in divergent regions highlighted by the program, as might be expected since PGIPs display varied specificity and inhibition mechanisms towards different polygalacturonases [84,85]. This reinforces the point that careful choice of input sequences is crucial to gaining meaningful output data. It will be interesting in the future to map plant NB-LRR proteins that exhibit direct

recognition of changing pathogen ligands, where the most divergent LRR sites might be predicted to be the sites of ligand recognition, and to compare these results to those obtained using Ka/Ks (positive selection) analyses [e.g., 86,87].

Although it was encouraging that 45% of all interaction residues from LRR+ligand structures appeared in the top 20% of RCM scores, some interaction residues received low conservation scores in RCM, indicating that these positions are not conserved. These amino acids could be responsible for differences in specificity of the compared proteins rather than any shared activity. Even in cases where two related proteins have extremely similar functions, some variable residues are to be expected at functional sites as these could be responsible for fine-tuning of receptor function, such as through ligand interaction kinetics. As a separate issue, there were many residues that scored highly in RCM that did not interact with ligand in the available crystal structure. While these could be false positives, many of them are likely responsible for functions not detectable in the crystal structure (i.e., interaction with molecules not present in the crystal structure). The regions highlighted by RCM but not matching known LRR+ligand interaction sites are intriguing targets for future study.

In LRR domains, disruption of single solvent-exposed residues very frequently does not disrupt function [e.g., 35,88]. We were pleased to find seven different double-alanine mutants of EFR that were each severely disrupted in function when compared to a wild-type receptor. These mutants were in four distinct conserved regions on the surface of EFR, one of which overlaps with a region identified by [36] as important for EFR function. A full-length protein can be detected for these different EFR LRR mutation alleles (Figure S5), but that leaves many other

possible activities that could be disrupted. Some of these patches may fold together to form a shared ligand-binding domain, but alternatively, three of the four sites may be involved in other processes such as receptor dimerization, interaction with co-receptors, or protein localization. Experiments are in progress to investigate the functions that are disrupted when these different regions are mutated. One prediction is that different function-blocking mutations in the same region (such as in the concave face of LRRs 11-12, or LRRs 16-17) will cause the same type of functional disruption.

Interestingly, although seven of the nine double-alanine mutants made in conserved EFR regions disrupted function, two of the mutants behaved in a manner identical to wild-type. This demonstrates that not all residues in a conserved region are essential for function. It is important to note that the score placed at any residue position is a regional conservation score for a 5x5 window, and may be elevated due to the presence of multiple conserved residues surrounding a relatively less conserved center residue. Examination of the single-residue conservation scores (as in Figure S2) may help to identify the most significant residues in a conserved region. Alternatively, the conserved regions not functionally disrupted by mutations may be functionally important in ways that would be not detected using the ligand that we utilized. As a third option, it is also possible that RCM detects a certain amount of 'noise' that represents relatively unselected stochastic conservation.

Prediction of functionally significant EFR sites was originally performed with only two sequences. When we identified a further seven functional orthologs of EFR and created a conservation map with these additional sequences, the same sites were identified, probably

with more precision. Use of two sequences could be misleading when the two sequences are very similar (see also the FLS2 maps of Figure S6). Additional sequences may be a means of increasing the power or reliability of RCM, but even a pairwise comparison can be informative.

We chose to focus on the β-strand, β-turn region of EFR because of the propensity of ligand binding sites to occur in this region [15]. However, in a few cases crystal structures have identified important residues outside the concave face of the protein, such as in the case of TLR3. The regions outside the concave face that were highlighted by the present conservation mapping effort may also be important for receptor function. For example, [89] recently identified a site of glycosylation, N143, on the convex face of the LRR of EFR that, when mutated, results in a severely impaired receptor. This residue occurs in the fourth repeat within the large conserved patch identified in both the ortholog and paralog conservation maps of EFR (Figure 4).

We were surprised that the most divergent region RCM identified among related but functionally distinct EFR paralogs, centered on the concave face of the fifth through ninth repeats, contained residues with an intermediate rather than a strong impact elf18 perception (Figure 4B and Figure 6). We infer that these paralogs are functionally distinct from EFR because none of the four paralogs confer elf18 recognition when they remain wild-type in an *efr*$^{-}$ mutant. This failure to see a strong impact on EFR function, when testing many of the site-directed random mutations of single residues in this most divergent region, may have occurred because the divergent sites we tested are more important for the gain of function of one or more of the other paralogs while being relatively unimportant for the responsiveness of EFR to

EF-Tu. It is equally possible that these divergent regions carry many residues that each make only modest contributions to elf18 recognition, with tolerance for alternative amino acid side chains. Alternatively, as was discussed for the double-alanine mutations, insignificant residues within functional regions may have been chosen for testing in our non-exhaustive mutational screen, or it may be unselected chance that the LRRs of EFR and its paralogs have diverged in this area. If these regions are under no substantial selective pressure one might expect that they would show intermediate coloration in the RCM image rather than deep blue. The deep blue sites on the map of Figure 4B are in fact quite diverse; the set of five EFR paralogs carried at least three different amino acids at most of the mutated positions, and in many cases five. As a separate matter, in the future it will be interesting to test the large conserved region that occurs on the convex face of the LRR region, especially since this region is conserved among both orthologs and paralogs of EFR.

Repeat conservation mapping offers advantages over other methods that could be used to identify functional regions. RCM only requires two homologous sequences to compare, whereas positive selection analysis and Consurf are best done with at least seven sequences [46,49]. Most methods that utilize information about the spatial proximity of residues in a folded protein, such as ODA, CFG and Consurf, require a solved crystal structure or a homology model. However, LRR domains can be very challenging to crystallize, and generation of a valid homology model is also highly challenging. The validation experiments conducted with RCM suggest that use of a generic LRR structural model with removal of consensus residues is sufficient to successfully predict areas of conservation or divergence.

There were significant instances of overlap in the functional regions predicted by ODA and RCM. However, the two methods rely on very different concepts and they also identified non-overlapping regions in response to the same input data. This included functionally confirmed regions such as the LRR11 region of EFR and the LRR 23 region of FLS2 that ODA did not identify. Consurf was not highly informative when used with our homology-modeled FLS2 or EFR (Figures 9 and S7), possibly because of the proximity of many LRR consensus sequence residues near the LRR surface. The CFG program, while previously shown by us to be useful for some parts of LRRs [35], is not optimal because it removes hydrophobic residues from consideration even if they are on the protein surface. CFG also takes into account residues that are in the consensus of many LRRs, such as the N and P residues that are largely buried and highly conserved for structural reasons rather than driving functional differences between LRRs. By assessing variable-position rather than consensus-position residues of the LRR, RCM may benefit from a focus on the residues that are most likely to impact function for reasons other than overall protein structure and integrity.

Future improvements to RCM are anticipated. Output of results onto a generic 3D LRR model as a pdb file may be useful, for viewing and manipulation in PyMOL, Swiss-PdbViewer or similar programs. The LRR repeat models developed in an RCM run currently require hand-curation, but greater automation of LRR matrix building is anticipated, for instance through use of HMMER (http://hmmer.janelia.org) to build consensus logos of the proteins being compared. RCM at this stage is limited to use with LRR domains, but because RCM utilizes typical repeat domain structures rather than precise spatial data, the method is likely to be adaptable to other repeat proteins that have a known repetitive structure. This includes armadillo,

tetratricopeptide, and ankyrin repeats. Adaptation of conservation mapping to ankyrin repeats could be particularly insightful, as these domains are the focus of artificial evolution towards novel ligands [90].

The repeat conservation mapping approach is a predictive method that can be utilized to identify key similarities and differences among groups of homologous LRR-containing proteins. The RCM program predicts functional sites in LRR domains, which may facilitate basic structure-function studies or *in vitro* protein evolution toward modified functions for these widespread and biologically significant domains. The current implementation is functional for LRRs, but it should be modifiable for use with other repeat-containing protein domains.

Acknowlegements

The authors thank Dan Huttenlocher for initial contributions to the RCM programming approach, and Mark Craven for helpful conversations.

References

1. Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. Curr Opin Struct Biol 11: 725-732.
2. Schmitz-Linneweber C, Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 13: 663-670.
3. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) The ankyrin repeat as molecular architecture for protein recognition. Protein Sci 13: 1435-1448.
4. Finn R, Mistry J, Tate J, Coggill P, Heger A, et al. (2009) The Pfam protein families database. Nucl Acids Res 38: D211-222.
5. Sigrist C, Cerutti L, De Castro E, Langendijk-Genevaux P, Bulliard V, et al. (2010) PROSITE, a protein domain database for functional characterization and annotation. Nucl Acids Res 38: D161-166.
6. Napier R (2004) Plant hormone binding sites. Annals Botany 93: 227-233.
7. Mercer A, Fleming S, Ueda N (2005) F-box-like domains are present in most poxvirus ankyrin repeat proteins. Virus Genes 31: 127-133.
8. Van Loy T, Vandersmissen H, Van Hiel M, Poels J, Verlinden H, et al. (2008) Comparative genomics of leucine-rich repeats containing G protein-coupled receptors and their ligands. General Comparative Endocrinology 155: 14-21.
9. Shanmugam V (2005) Role of extracytoplasmic leucine rich repeat proteins in plant defence mechanisms. Microbiol Res 160: 83-94.
10. Padmanabhan M, Cournoyer P, Dinesh-Kumar SP (2009) The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. Cell Microbiol 11: 191-198.
11. Eitas TK, Dangl JL (2010) NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Curr Opin Plant Biol 13: 472-477.
12. Herrin BR, Cooper MD (2010) Alternative adaptive immunity in jawless vertebrates. J Immunol 185: 1367-1374.
13. Tasumi S, Velikovsky CA, Xu G, Gai SA, Wittrup KD, et al. (2009) High-affinity lamprey VLRA and VLRB monoclonal antibodies. Proc Natl Acad Sci U S A 106: 12891-12896.
14. Velikovsky CA, Deng L, Tasumi S, Iyer LM, Kerzic MC, et al. (2009) Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen. Nature Struct & Mol Biol 16: 725-730.
15. Bella J, Hindle KL, McEwan PA, Lovell SC (2008) The leucine-rich repeat structure. Cell Mol Life Sci 65: 2307-2333.
16. Kajander T, Cortajarena AL, Regan L (2006) Consensus design as a tool for engineering repeat proteins. Methods Mol Biol 340: 151-170.
17. Jones JD, Dangl JL (2006) The plant immune system. Nature 444: 323-329.

18. Boller T, Felix G (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu Rev Plant Biol 60: 379-406.

19. Nicaise V, Roux M, Zipfel C (2009) Recent advances in PAMP-triggered immunity against bacteria: pattern recognition receptors watch over and raise the alarm. Plant Physiol 150: 1638-1647.

20. Schwessinger B, Zipfel C (2008) News from the frontline: recent insights into PAMP-triggered immunity in plants. Curr Opin Plant Biol 11: 389-395.

21. Shiu SH, Bleecker AB (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. Plant Physiol 132: 530-543.

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

23. Morillo SA, Tax FE (2006) Functional analysis of receptor-like kinases in monocots and dicots. Curr Opin Plant Biol 9: 460-469.

24. Takeda K, Akira S (2005) Toll-like receptors in innate immunity. Int Immunol 17: 1-14.

25. Ye Z, Ting J (2008) NLR, the nucleotide-binding domain leucine-rich repeat containing gene family. Curr Opin Immunol 20: 3-9.

26. Rafiqi M, Bernoux M, Ellis J, Dodds P (2009) In the trenches of plant pathogen recognition: Role of NB-LRR proteins. Semin Cell Dev Biol 20: 1017-1024.

27. Shaw M, Reimer T, Kim Y, Nuñez G (2008) NOD-like receptors (NLRs): bona fide intracellular microbial sensors. Curr Opinion Immunol 20: 377-382.

28. Proell M, Riedl S, Fritz J, Rojas A, Schwarzenbacher R (2008) The Nod-like receptor (NLR) family: a tale of similarities and differences. PLoS ONE 3: e2119.

29. Felix G, Duran JD, Volko S, Boller T (1999) Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J 18: 265-276.

30. Kunze G, Zipfel C, Robatzek S, Niehaus K, Boller T, et al. (2004) The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell 16: 3496-3507.

31. Gómez-Gómez L, Boller T (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell 5: 1003-1011.

32. Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JD, et al. (2006) Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. Cell 125: 749-760.

33. Chinchilla D, Boller T, Robatzek S (2007) Flagellin signalling in plant immunity. Curr Topics Innate Immunity: 358-371.

34. Chinchilla D, Bauer Z, Regenass M, Boller T, Felix G (2006) The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant Cell 18: 465-476.

35. Dunning FM, Sun W, Jansen KL, Helft L, Bent AF (2007) Identification and mutational analysis of Arabidopsis FLS2 leucine-rich repeat domain residues that contribute to flagellin perception. Plant Cell 19: 3297-3313.

36. Albert M, Jehle AK, Mueller K, Eisele C, Lipschis M, et al. (2010) Arabidopsis thaliana pattern recognition receptors for bacterial elongation factor Tu and flagellin can be combined to form functional chimeric receptors. J Biol Chem 285: 19035-19042.

37. Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JD, et al. (2004) Bacterial disease resistance in Arabidopsis through flagellin perception. Nature 428: 764-767.

38. Lacombe S, Rougon-Cardoso A, Sherwood E, Peeters N, Dahlbeck D, et al. (2010) Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. Nature Biotechnology 28: 365-369.

39. Wan J, Zhang XC, Neece D, Ramonell KM, Clough S, et al. (2008) A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis. Plant Cell 20: 471-481.

40. Kishimoto K, Kouzai Y, Kaku H, Shibuya N, Minami E, et al. (2010) Perception of the chitin oligosaccharides contributes to disease resistance to blast fungus Magnaporthe oryzae in rice.

41. Zeidler D, Zahringer U, Gerber I, Dubery I, Hartung T, et al. (2004) Innate immunity in Arabidopsis thaliana: lipopolysaccharides activate nitric oxide synthase (NOS) and induce defense genes. Proc Natl Acad Sci U S A 101: 15811-15816.

42. Forsyth A, Mansfield J, Grabov N, Sinapidou E, de Torres M, et al. (2010) Genetic Dissection of Basal Resistance to Pseudomonas syringae pv. phaseolicola in accessions of Arabidopsis. Mol Plant-Microbe Interactions 23: 1545-1552.

43. Zhou L, Cheung M, Zhang Q, Lei C, Zhang S, et al. (2009) A novel simple extracellular leucine-rich repeat (eLRR) domain protein from rice (OsLRR1) enters the endosomal pathway and interacts with the hypersensitive-induced reaction protein 1. Plant Cell Environm 32: 1804-1820.

44. Suzuki Y (2004) Three-dimensional window analysis for detecting positive selection at structural regions of proteins. Mol Biol Evol 21: 2352-2359.

45. Innis CA, Anand AP, Sowdhamini R (2004) Prediction of functional sites in proteins using conserved functional group analysis. J Mol Biol 337: 1053-1068.

46. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591.

47. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, et al. (2002) The Pfam protein families database. Nucl Acids Res 30: 276-280.

48. Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, et al. (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91: 821-832.

49. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucl Acids Res 38: W529-533.

50. Fernández-Recio J, Totrov M, Skorodumov C, Abagyan R (2005) Optimal docking area: a new method for predicting protein-protein interaction sites. Proteins Struct Function Bioinformatics 58: 134-143.

51. Wilbur W, Lipman D (1983) Rapid similarity searches of nucleic acid and protein data banks. Proc Natl Acad Sci U S A 80: 726-730.

52. Myers E, Miller W (1988) Optimal alignments in linear space. Comput Appl Biosci 4: 11-17.

53. Dolan J, Walshe K, Alsbury S, Hokamp K, O'Keeffe S, et al. (2007) The extracellular leucine-rich repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns. BMC Genomics 8: 320-343.

54. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

55. Casasoli M, Federici L, Spinelli F, Di Matteo A, Vella N, et al. (2009) Integration of evolutionary and desolvation energy analysis identifies functional sites in a plant immunity protein. Proc Natl Acad Sci U S A 106: 7666-7671.

56. Di Matteo A, Federici L, Mattei B, Salvi G, Johnson KA, et al. (2003) The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. Proc Natl Acad Sci U S A 100: 10124-10128.

57. van der Hoorn RA, Wulff BB, Rivas S, Durrant MC, van der Ploeg A, et al. (2005) Structure-function analysis of Cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. Plant Cell 17: 1000-1015.

58. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779-815.

59. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126-2132.

60. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures.

61. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng 104: 34-41.

62. Adams-Phillips L, Briggs AG, Bent AF (2009) Disruption of poly(ADP-ribosyl)ation mechanisms alters responses of Arabidopsis to biotic stress. Plant Physiol 152: 267-280.

63. Gómez-Gómez L, Felix G, Boller T (1999) A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana. Plant J 18: 277-284.

64. D'ovidio R, Raiola A, Capodicasa C, Devoto A, Pontiggia D, et al. (2004) Characterization of the complex locus of bean encoding polygalacturonase-inhibiting proteins reveals subfunctionalization for defense against fungi and insects. Plant Physiol 135: 2424-2435.

65. Leckie F, Mattei B, Capodicasa C, Hemmings A, Nuss L, et al. (1999) The specificity of polygalacturonase-inhibiting protein (PGIP): a single amino acid substitution in the solvent-exposed beta-strand/beta-turn region of the leucine-rich repeats (LRRs) confers a new recognition capability. EMBO J 18: 2352-2363.

66. Papageorgiou AC, Shapiro R, Acharya KR (1997) Molecular recognition of human angiogenin by placental ribonuclease inhibitor—an X-ray crystallographic study at 2.0 A resolution. EMBO J 16: 5162-5177.

67. Kobe B, Deisenhofer J (1996) Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. J Mol Biol 264: 1028-1043.

68. Tan X, Calderon-Villalobos L, Sharon M, Zheng C, Robinson C, et al. (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446: 640-645.

69. Fan Q, Hendrickson W (2005) Structure of human follicle-stimulating hormone in complex with its receptor. Nature 433: 269-277.

70. Celikel R, McClintock R, Roberts J, Mendolicchio G, Ware J, et al. (2003) Modulation of α-thrombin function by distinct interactions with platelet glycoprotein Ibα. Science 301: 218-221.

71. Hao B, Zheng N, Schulman BA, Wu G, Miller JJ, et al. (2005) Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. Mol Cell 20: 9-19.

72. Morlot C, Thielens N, Ravelli R, Hemrika W, Romijn R, et al. (2007) Structural insights into the Slit-Robo complex. Proc Natl Acad Sci U S A 104: 14923-14928.

73. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, et al. (2007) Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell 130: 1071-1082.

74. Kang J, Nan X, Jin M, Youn S, Ryu Y, et al. (2009) Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity 31: 873-884.

75. Park B, Song D, Kim H, Choi B, Lee H, et al. (2009) The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. Nature 458: 1191-1195.

76. Liu L, Botos I, Wang Y, Leonard J, Shiloach J, et al. (2008) Structural basis of toll-like receptor 3 signaling with double-stranded RNA. Science 320: 379-381.

77. Dickson K, Haigis M, Raines R (2005) Ribonuclease inhibitor: structure and function. Prog Nucl Acid Res Mol Biol 80: 349-374.

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

79. Dharmasiri N, Dharmasiri S, Weijers D, Lechner E, Yamada M, et al. (2005) Plant development is regulated by a family of auxin receptor F box proteins. Dev Cell 9: 109-119.

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

81. Walsh T, Neal R, Merlo A, Honma M, Hicks G, et al. (2006) Mutations in an auxin receptor homolog AFB5 and in SGT1b confer resistance to synthetic picolinate auxins and not to 2, 4-dichlorophenoxyacetic acid or indole-3-acetic acid in Arabidopsis. Plant Physiol 142: 542-552.

82. Zipfel C (2009) Early molecular events in PAMP-triggered immunity. Curr Opin Plant Biol 12: 414-420.

83. Couvreur TL, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, et al. (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Mol Biol Evol 27: 55-71.

84. Federici L, Di Matteo A, Fernández-Recio J, Tsernoglou D, Cervone F (2006) Polygalacturonase inhibiting proteins: players in plant innate immunity? Trends Plant Sci 11: 65-70.

85. Sicilia F, Fernández-Recio J, Caprari C, De Lorenzo G, Tsernoglou D, et al. (2005) The polygalacturonase-inhibiting protein PGIP2 of Phaseolus vulgaris has evolved a mixed mode of inhibition of endopolygalacturonase PG1 of Botrytis cinerea. Plant Physiol 139: 1380-1388.

86. Dodds PN, Lawrence GJ, Catanzariti AM, Teh T, Wang CI, et al. (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc Natl Acad Sci U S A 103: 8888-8893.

87. Krasileva KV, Dahlbeck D, Staskawicz BJ (2010) Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. Plant Cell 22: 2444-2458.

88. Clark LB, Viswanathan P, Quigley G, Chiang YC, McMahon JS, et al. (2004) Systematic mutagenesis of the leucine-rich repeat (LRR) domain of CCR4 reveals specific sites for binding to CAF1 and a separate critical role for the LRR in CCR4 deadenylase activity. J Biol Chem 279: 13616-13623.

89. Häweker H, Rips S, Koiwa H, Salomon S, Saijo Y, et al. (2010) Pattern recognition receptors require N-glycosylation to mediate plant immunity. J Biol Chem 285: 4629-4636.

90. Binz H, Stumpp M, Forrer P, Amstutz P, Plückthun A (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 332: 489-503.

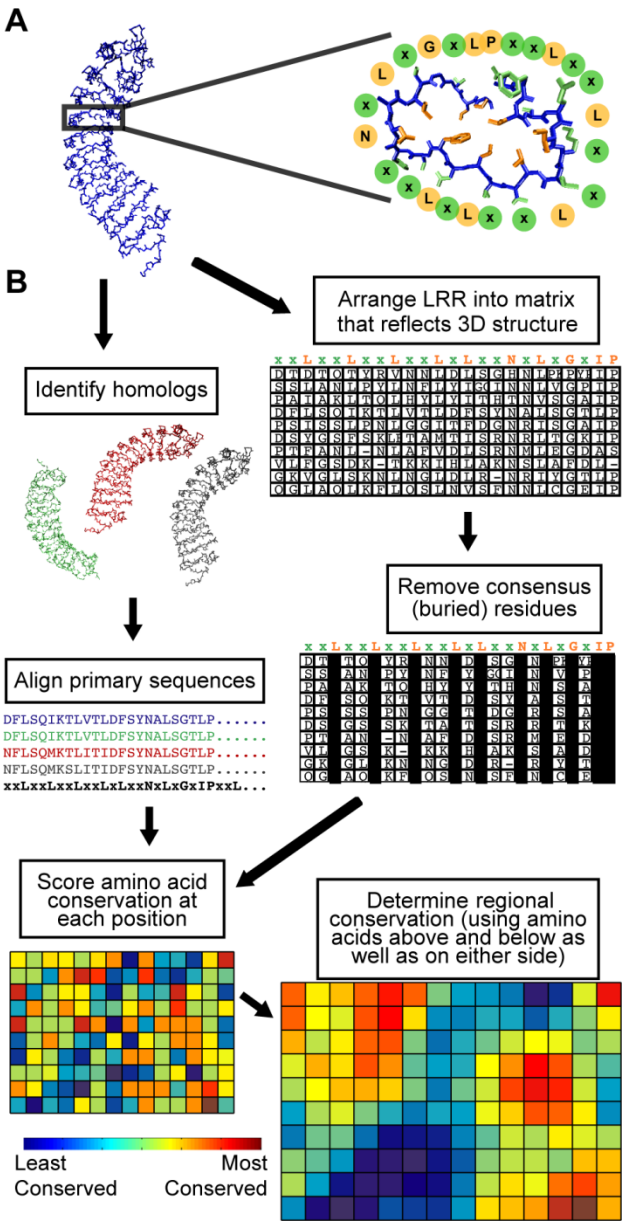

Figure 1: LRR structure and an outline of conservation mapping procedure. (A) Left: a representative LRR domain (left), *P. vulgaris* PGIP2, which forms the regular spiral pattern typical of LRRs. Right: a single 24 amino acid repeat of the LRR, surrounded by circles designating the residues of the LRR consensus amino acid sequence (`xxLxxLxxLxxLxLxxNxLxGxIP`). Note that the consensus residue side chains (orange) form the core of the protein, whereas the variable residues (green) are solvent-exposed. **(B)** Schematic representation of the conservation mapping procedure, using the example of PGIP1-4. See Methods and Text S1 for a detailed description of the procedure.

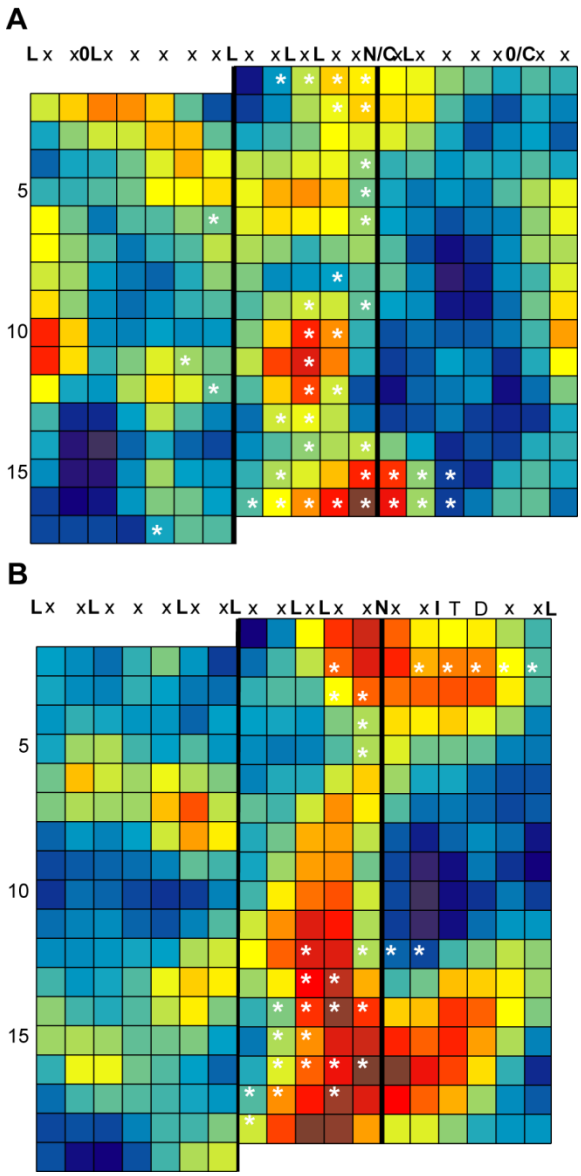

Figure 2: Validation of RCM by mapping LRR domains for which there are solved crystal structures with ligand. For all maps, each row represents a single repeat of the LRR, with each colored box representing a solvent exposed (non-consensus) amino acid position. Each column corresponds to a position within the LRR consensus sequence, as denoted at the top of each map. The color in each box reports the center-weighted regional conservation score for the 5x5 set of boxes that centers on that box (see text for details); dark red indicates the most conserved regions and blue indicates the most divergent regions (see scale bar in Fig. 1). Bold black vertical lines delineate the five residues in each row that comprise the β-strand, β-turn region (the convex face of the LRR domain). White asterisks were added after RCM and indicate amino acid positions that are LRR-ligand contact points in solved crystal structures. **(A)** RCM output for Ribonuclease inhibitor (RI) from human, rat, mouse and pig ribonuclease inhibitor (RI) LRRs. **(B)** RCM output for TIR1 and AFB1-5 (auxin receptors) from *Arabidopsis*.

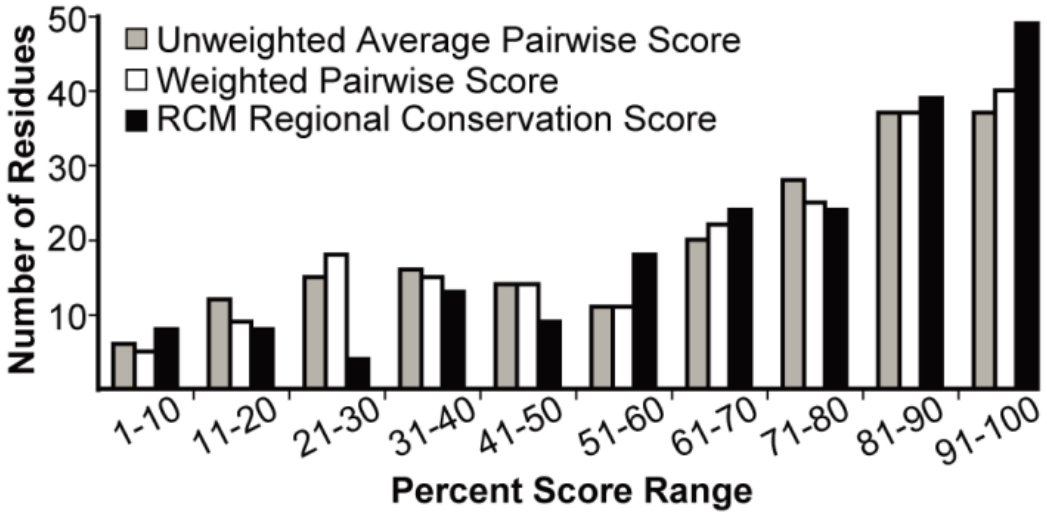

Figure 3: Residues involved in intramolecular interactions receive high scores in conservation mapping. For each map in Figures 2 and S1 (eleven protein families), all scores were arranged in descending order and then divided into deciles. Then, across all maps, the number of occurrences of residue positions marked by asterisks (LRR+ligand contact points) were tallied for each decile. Histogram shows the frequency with which three different types of scores fell into a particular decile. Score types are raw average pairwise BLOSUM65 score (grey), weighted average pairwise BLOSUM65 scores (see "Step 4" in Methods) prior to adjustment based on nearby amino acids (white), and regional conservation score that serves as final RCM output (black). Note that a subset of these protein comparisons involved homologs with unclear and possibly divergent rather than fully overlapping functions, in which case LRR-ligand contact points might be predicted to score as divergent rather than as conserved (see text).

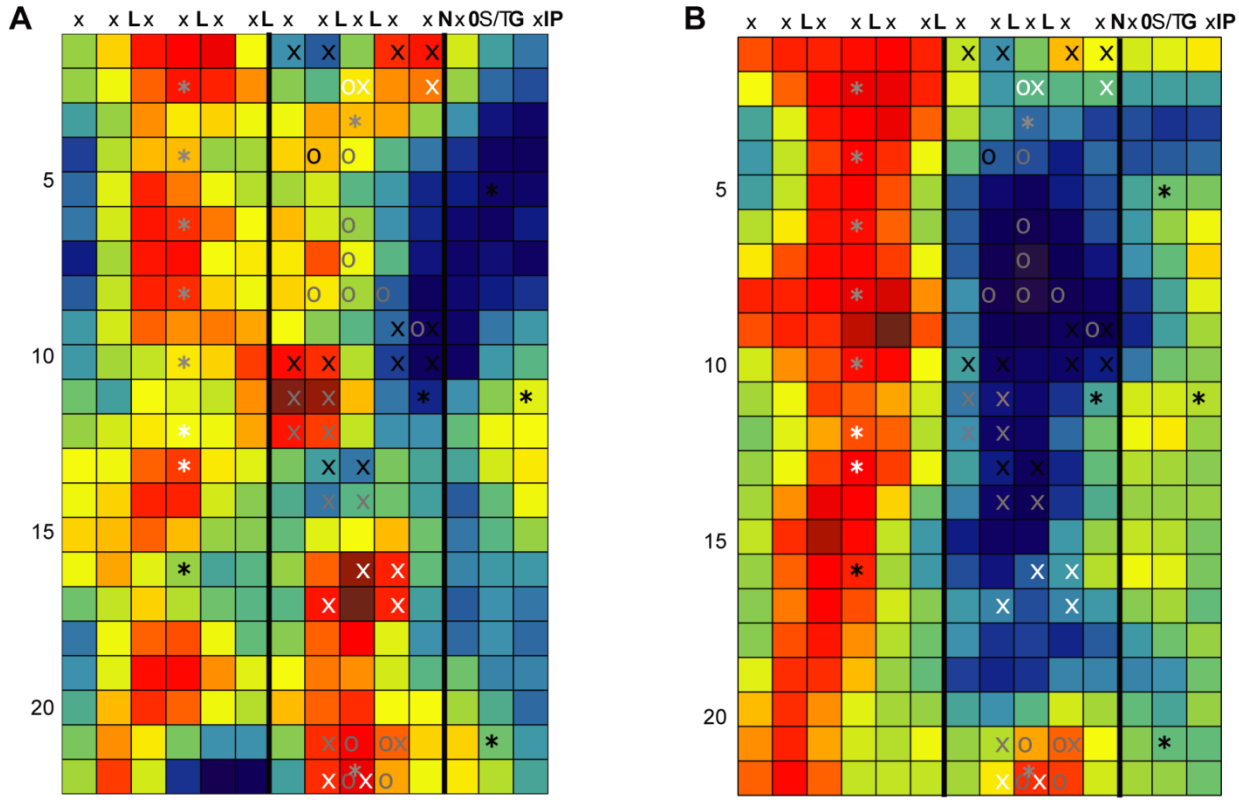

Figure 4: Conservation mapping of EFR. RCM maps of LRR domains (as described in Figure 2) depicting: **(A)** Conservation of AtEFR and a *Brassica rapa* homolog shown to have EFR activity. **(B)** Conservation mapping of AtEFR and its four most closely related *Arabidopsis* EFR paralogs, which fail to confer elf18 recognition in an *Arabidopsis efr*$^{-}$ mutant. The pairs of X symbols mark sites of double-alanine mutations, and O symbols mark sites of site-directed randomizing mutagenesis. The asterisks mark predicted sites of N-glycosylation as reported in [89]. Both (A) and (B) show the same set of X, O and * symbols. Based on data from Figures 5 and 6 and [89], symbols are white if the mutation disrupts EFR function, black if it does not, and grey for a partial/intermediate impact (less responsive than 95% of positive controls transformed with wild-type *EFR* but more responsive than 95% of empty-vector *efr*$^{-}$ negative controls; or statistically different from both wild-type and empty vector controls).

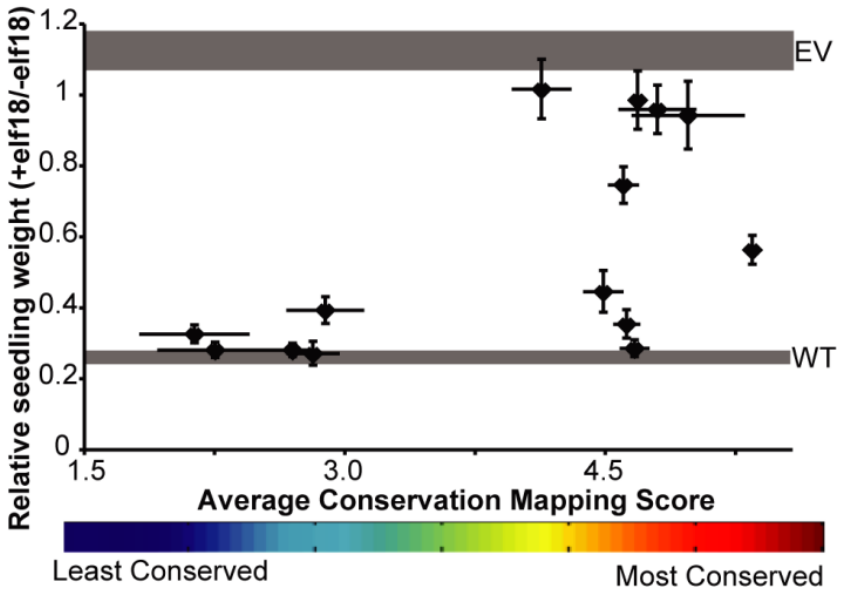

Figure 5: Functional testing of double-alanine mutagenesis alleles of EFR reveals that function-blocking mutations map to sites that RCM scores as conserved sites. A series of *EFR* alleles that each encode two alanine mutations within the β-strand/β-turn region of a single repeat (locations shown in Figure 4) were introduced into *efr*$^-$ plants and transgenic T1 seedlings were tested for their ability to respond to 100nM elf18 in a seedling growth inhibition assay. Y-axis: Average weight of seedlings (+/- standard error), as compared to a no-peptide control for each line, for at least three replicate experiments with at least eight transgenic seedlings per genotype per treatment. Controls are weight (mean +/- standard error) of *efr*$^-$ seedlings transformed with wild-type *EFR* (WT, lower grey band) or empty vector (EV, upper grey band), as determined within the same experiments. X-axis: The average RCM regional conservation score for the two positions mutated in any given construct (x-axis bar ends for each symbol are at the RCM scores for the two positions mutated in each construct).

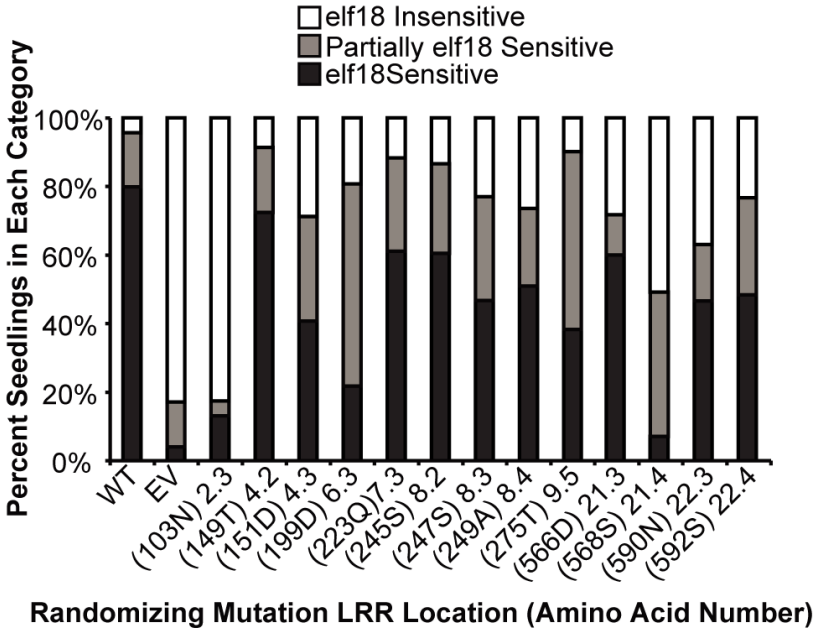

Figure 6: Site-directed, randomizing mutagenesis of EFR sites that are divergent in EFR paralogs reveals many partial-impact LRR sites. Seedling growth inhibition assays for EFR function were performed as in Figure 5 except that seedlings were treated with 1uM elf18. A minimum of 65 T1 seedlings were tested for each allele library. Plants were classified as elf18-sensitive if their response was in the range of 95% of positive controls transformed with wild-type *EFR*, as elf18-insensitive if response was in the range of 95% of negative controls transformed with empty vector, or as partially sensitive if they fell within both ranges. WT: wild-type; EV: empty vector; allele number codes reflect the repeat number and the β-strand/β-turn position of the mutagenized amino acid (for example, 2.3 is the third x position of LxxLxLxxN in the second repeat), and parentheses enclose the amino acid number and letter.

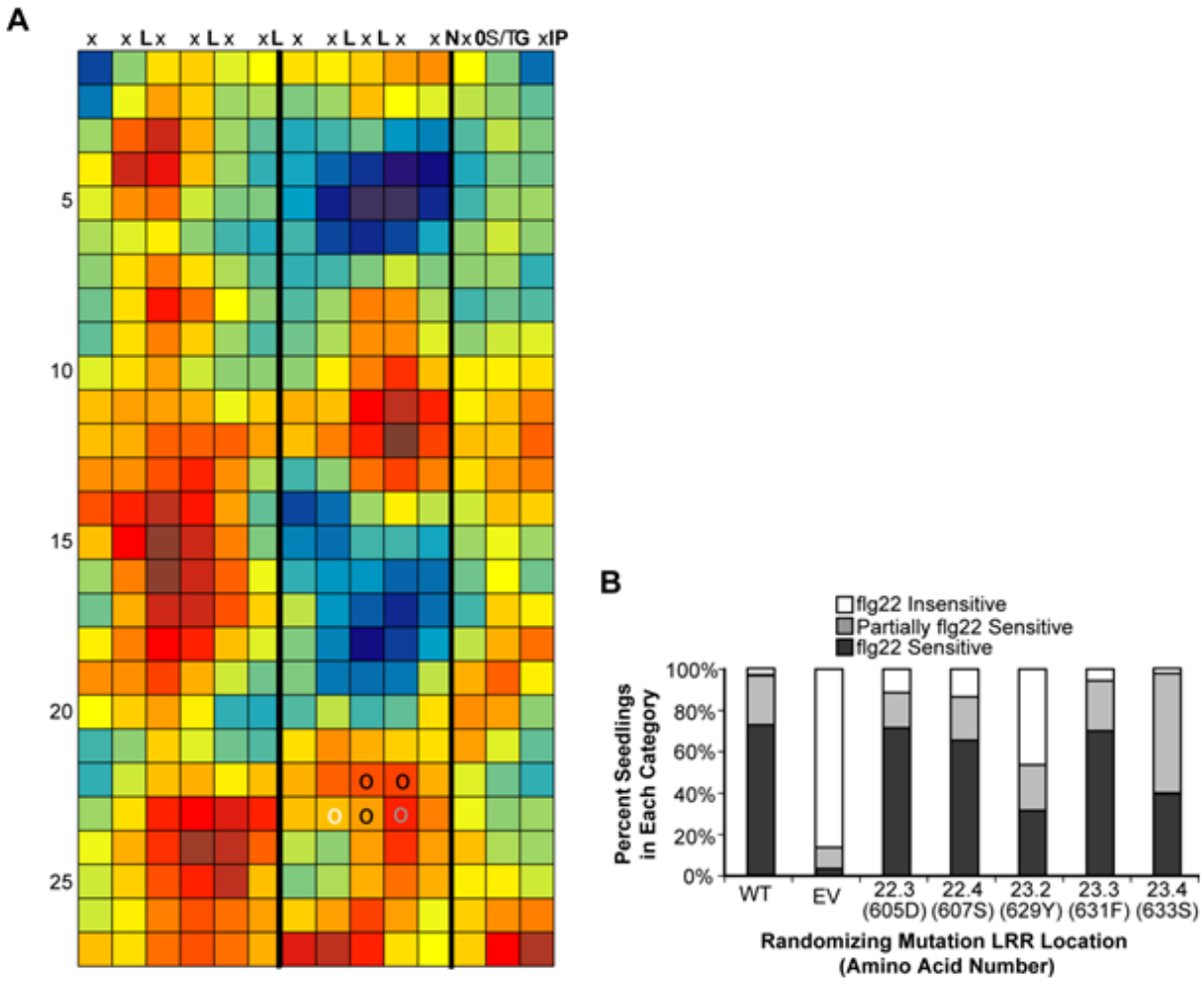

Figure 7: Site-directed, randomizing mutagenesis of five FLS2 residues in a region predicted by RCM to be functional identifies functional sites. (A) RCM map of eight Brassicaceae FLS2 orthologs. **(B)** Response to flg22 in T1 seedlings carrying FLS2 mutagenized at sites predicted by RCM. Each library represents a pool of *fls2-101* seedlings, each with one FLS2 construct carrying a random mutation at the position indicated. WT: wild-type; EV: empty vector; allele numbers as in Figure 6, and parentheses enclose the amino acid number and letter (indexed both by amino acid number and LRR position). Seedlings were scored as sensitive, insensitive, or partially sensitive to 1uM flg22 in a seedling growth inhibition assay by comparison to 95% confidence intervals for data from *fls2-101* plants transformed with a wild-type *FLS2* or empty vector control, assayed on the same day. At least 65 plants per library were tested over 3-4 independent experiments, and the results are pooled.

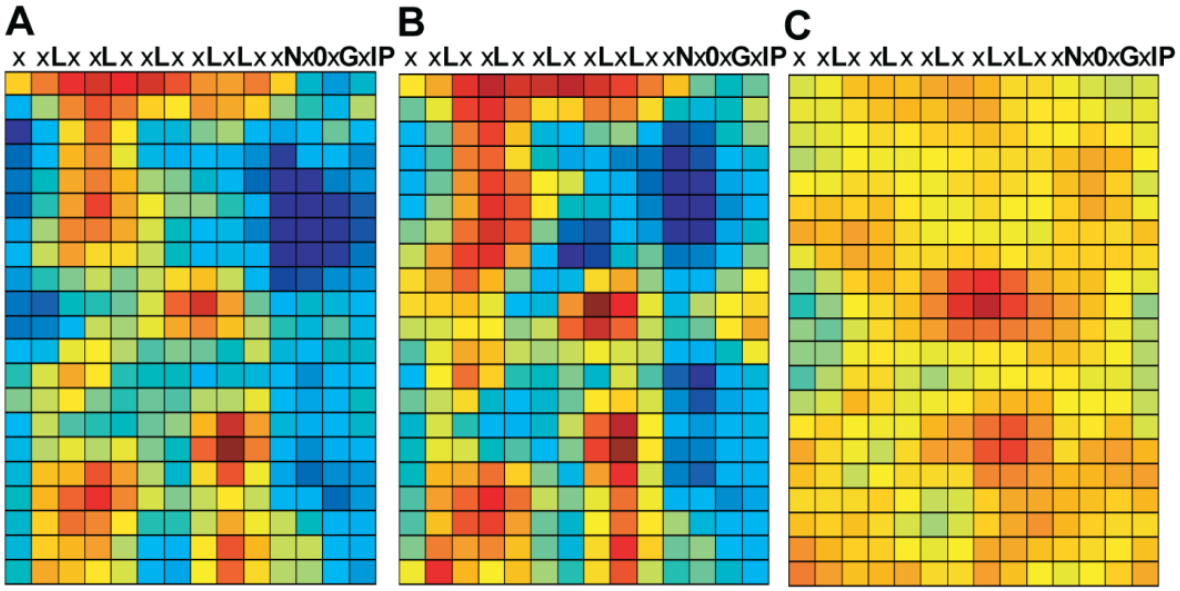

Figure 8: Conservation mapping using different EFR ortholog sets. **(A)** RCM map for ten Brassicaceae EFR orthologs. All sequences were confirmed as functional orthologs. **(B)** RCM map of two EFR orthologs, from distantly related *Eruca sativa* and *Erysimum raulinii*. **(C)** RCM map of two EFR orthologs, both from *Erysimum raulinii*. Maps are as described in Figure 2.

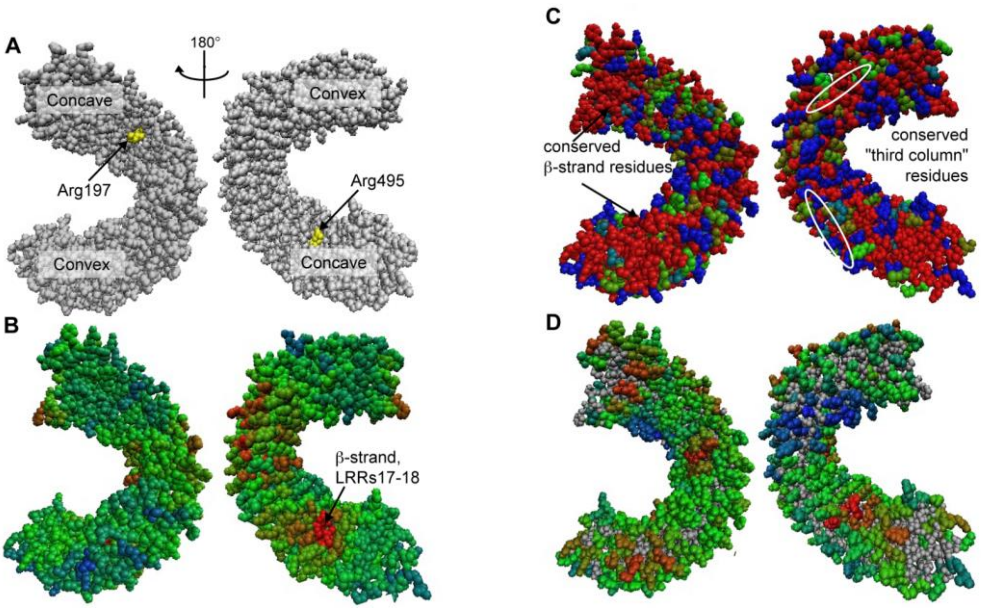

Figure 9: Analysis of EFR and orthologs by other computational methods. Results displayed on a homology model for **(A)** PAML positive/purifying selection analysis, **(B)** ODA analysis, **(C)** Consurf, and **(D)** RCM. For (A), residues undergoing positive selection (PP>.95 for at least one model) are highlighted in yellow. In (B), ODA exclusively utilized the homology model to make its functional predictions. For Consurf (C), derived amino acid sequence data for ten orthologous LRR domains from diverse Brassicaceae species were used (see text). To allow comparison, (D) presents the RCM data of Figure 4A on a homology model. Color scale ranges from "not predicted to be a binding site/not conserved" (blue) to "predicted to be a binding site/highly conserved" (red). All maps are shown in two views with 180^{o} rotation between right and left columns; concave and convex faces of the LRR are indicated in (A). See Figure S7 for analysis of FLS2 using the same methods, and [55] for ODA and positive selection analyses of PGIP.

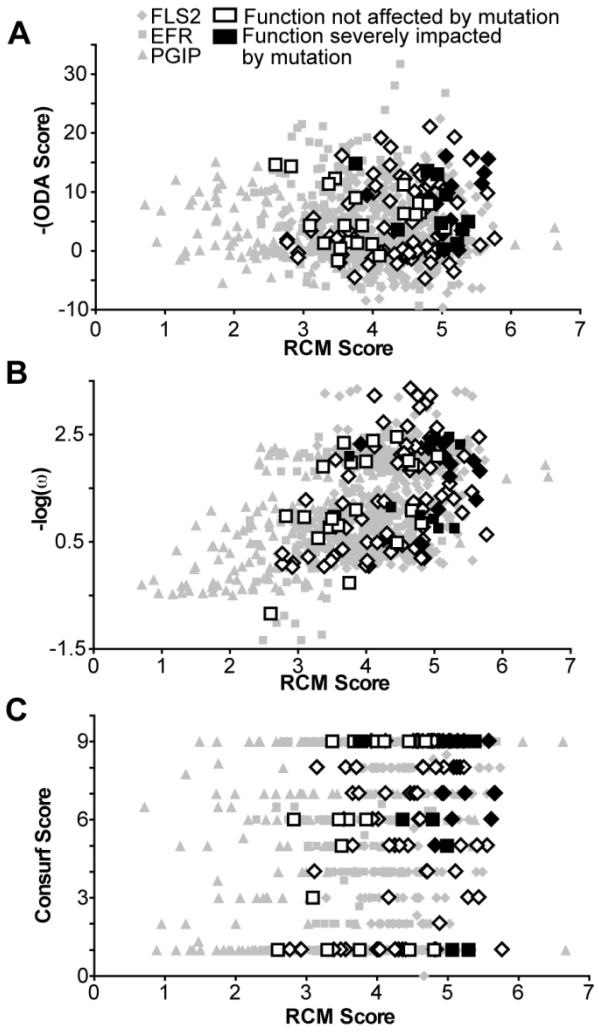

Figure 10: RCM analysis of FLS2, EFR, and PGIP performs as well as other computational methods. Y-axis: scores for each amino acid according to **(A)** positive selection analysis (nonsynonymous/synonymous substitution rates); **(B)** optimal docking area (scale is negative so that more likely docking sites are higher along y-axis); and **(C)** Consurf; all are graphed with respect to RCM score on the x-axis. Residues that were mutationally tested in the present study or in [35] are indicated as having no significant impact on function (white) or a significant impact on function (black) when mutated. Functional residues (in black) scored significantly higher than non-functional residues (in white) in all analyses completed (T-test; $p<.05$).

Chapter 3: EFR – elf18 binding study using yeast surface display

Introduction

Yeast surface display (YSD)

Yeast surface display (YSD) is a platform for the study and manipulation of proteins (Boder and Wittrup, 1997). YSD can be used to build protein libraries, or most commonly, variant forms of one type of protein (Zaccolo et al., 1996; Swers et al., 2004). These libraries may then be subjected to functional studies and *in vitro* protein evolution. A major application of YSD is the engineering of antibodies to increase their affinity, specificity and stability (Boder et al., 2000; Kieke et al., 1997, 1999; Shusta et al., 2000; Cho et al., 1998; Chao et al., 2006). Compared to phage display, YSD has the advantage that the proteins of interest are produced in eukaryotic cells. This can be important for post-translational modifications, e.g. glycosylation and disulfide isomerization. In contrast to the yeast two hybrid method and phage display, the proteins of interest go through the entire secretory pathway. These advantages make YSD the method of choice for the study of complex extracellular eukaryotic proteins (or protein domains). For receptor proteins, an added advantage is that soluble ligand-binding kinetics and equilibria can be measured (Boder and Wittrup, 1998). Flow cytometry is often used as a readout in YSD studies, and fluorescence-activated cell sorting can be used to capture sub-pools of a protein library that are enriched for yeast cells carrying the DNA constructs that encode proteins with desired functional behaviors (Boder and Wittrup, 1997).

Hybrid LRR technique

The hybrid LRR technique was developed for crystallization of the LRR domain of Toll like receptors (Jin and Lee, 2008). In this technique, N-and/or C-terminal regions of VLR proteins such as VLRB.61 (PDB ID: 2O6R) are appended to the LRR of interest. As fusion sites, the highly conserved motif LxxLxLxxN is used. Applying this technique it was possible to crystalize TLR2, TLR4 (Figure 1), TLR5 and TLR6 (Kang et al., 2009; Kim et al., 2007; Yoon et al., 2012). The technique enhances the soluble expression of the LRR domain while maintaining its ligand binding ability (Figure 1A). To our knowledge, the hybrid LRR technique has not been used to crystalize plant proteins so far.

Objective

The objective of this work was to determine the elf18 binding site in the ectodomain of EFR. So far no crystal structure of the EFR ectodomain alone or together with ligand elf18 is available. Häweker et al. studied EFR N-glycosylation and identified residue N143 in the EFR ectodomain to be important for EFR function (Häweker et al., 2010). In our lab we used conservation mapping and site-directed mutagenesis (see chapter 2) to identify the putative elf18 binding site. Conservation mapping identified two distinct conserved regions in the concave face of the EFR LRR domain.

In this study we sought to display several EFR LRR domain fragments on yeast in order to determine which fragment binds elf18. The fragments LRR 8-21 and LRR 13-21 contain the

conserved regions identified using conservation mapping (Figure 6E). We hypothesize that EFR LRRs 13-19 are sufficient to bind elf18.

Methods and results

Display of EFR and FLS2 LRR domains

In collaboration with Professor Eric Shusta (Department of Chemical and Biological Engineering, UW Madison), who is a worldwide leader in yeast surface display technology, we worked to adapt this technique for display of EFR and FLS2 LRR domains. We used the plasmid pCT4 from the Shusta lab to display our constructs. As positive control we used pCT4-ESOP3 from the Shusta lab. The pCT4 plasmid was originally from (Tachibana and Stevens, 1992). Wenxian Sun (former Bent lab postdoc) generated the plasmid pCT4-FLS2 and I generated the plasmid pCT4-EFR. The constructs contain an N-terminal c-myc tag and a C-terminal HA tag (Figure 2). FLAG tagged elf22 and flg22 peptides were ordered from Genscript. We decided to use elf22 (22 amino acids) instead of elf18 in order to have a flexible linker between the eliciting peptide and the FLAG tag.

The FLS2 and EFR LRR domains displayed on yeast but the display was very low compared to the positive control construct ESOP3 (Figure 3, Figure 4). Binding of the ligands elf22 and flg22 could not be detected (Figure 4A-D). The displayed constructs were cut off the yeast cells, subjected to SDS-PAGE and detected by Western Blot. Two bands of the expected sizes were detected corresponding to the EFR and the FLS2 constructs (Figure 4E).

Display of VLRB.61 capped EFR LRR domain

In order to improve the display of our constructs and to be able to detect ligand binding, the hybrid LRR technique was then implemented. We obtained the cDNA of VLR.B61 from the RIKEN BioResource Center in Japan. The EFR LRR domain was capped immediately N-terminal to the LRR with the N-terminus of VLRB.61, using splice overlap extension PCR. The VLR cap improved the display significantly (Figure 5B). But surprisingly, the EFR LRR domain with VLR N-terminal cap was able to bind the wrong ligand flg22 (Figure 5D). A low amount of elf22 binding was observable but only when 3 x more elf22 was added then the standard amount (Figure 5C).

To ensure the biological activity and the specificity of the epitope tagged ligands FLAG-flg22 and elf22-FLAG, leaf discs from Col-0, *fls2*$^{-}$ and *efr*$^{-}$ Arabidopsis plants were exposed to the ligands and subsequent ROS bursts were measured. FLAG-flg22 elicited an ROS burst in Col-0 and *efr*$^{-}$ plants and elf22-FLAG elicited an ROS burst in Col-0 and *fls2*$^{-}$ plants, respectively (Figure 5E). Thus the FLAG-tagged ligands were active and specific.

Display of EFR LRR domain fragments

Three EFR LRR domain fragments were generated using splice overlap extension PCR. The motif LxxLxLxxN was used as fusion site. Fragment 1 contained LRRs 1-9, fragment 2 contained LRRs 8-21 and fragment 3 contained LRRs 13-21 (Figure 6E).

The three EFR fragments displayed better than the full length EFR LRR domain (Figure 6). The best display was obtained with fragment 3 LRR 13-21 (Figure 6 C+D). However no elf22 binding was detected on any of the three fragments (Figure 6 A - C). As with the full-length LRR

experiments described above, when the wrong ligand flg22 was added to the well-displayed EFR LRR fragment 3, there was substantial non-specific ligand binding (Figure 6D).

Protocol YSD

Yeast strain EBY100 was transformed with the pCT4 plasmids containing the gene construct of interest and grown for 2 days at 28°C on SD-CAA (-ura, -trp) plates. One colony from each genotype was grown overnight at 30°C in liquid SD-CAA medium. Cells were pelleted by centrifugation, resuspended in SG-CAA medium (containing galactose for induction of gene expression by the galactose promoter), and grown on an orbital shaker at 20°C for 20h to 40h. The cells were pelleted by centrifugation, washed in PBS-BSA and incubated for 30 min on ice with the primary antibody in PBS-BSA. The cells were again pelleted and washed and then incubated for 30 min on ice with the secondary antibody in PBS-BSA, this time while kept in the dark. Then the cells were pelleted, washed and resuspended in PBS-BSA. A flow cytometer was used to analyze the cells and the programs Summit and WinMDI were used to analyze the blots.

Discussion

Display of EFR and FLS2 LRR domains

The LRR domains of EFR and FLS2 did not display well on yeast. However, the protein constructs were present on the yeast cells because they were detectable by western blot.

Display of VLRB.61 capped EFR LRR domain

The VLRB.61 cap at the N-terminus of the EFR LRR domain significantly improved the display. Surprisingly, it did not significantly improve elf22 binding. The addition of 3x more elf22 resulted in some detectable ligand binding. Disappointingly, the addition of flg22 to the VLR capped EFR LRR domain resulted in much greater ligand binding that the 3x amount of elf22. This was unexpected because flg22 did not bind substantially to yeast cells when no or only little construct was displayed. The data suggested that the VLRB-EFR construct, rather than the flg22 ligand, was sticky (prone to non-specific binding). Hence, although the elf22 binding that was detected may suggest some promise for the YSD approach for study of the EFR LRR domain, the project was deemed a lower priority at that juncture in my Ph.D. thesis research.

Display of EFR LRR domain fragments

The EFR LRR domain fragments displayed better than the EFR full-length LRR domain. However none of the three fragments were able to bind elf22 to a significant extent. Surprisingly, when flg22 was added to fragment 3, binding was observed. Thus it was the same unexpected result as was obtained when flg22 was added to the VLRB capped EFR. It was encouraging to see that the VLR cap improved the EFR LRR domain display, but it was dis-encouraging that VLRB-EFRLRR and EFRLRR13-21 both were binding flg22 instead of elf22. One explanation might be that glycosylation defects of the EFR LRRs somehow made the construct sticky and unable to bind the right ligand elf22. N-glycan processing in the Golgi apparatus is kingdom specific and results in different oligosaccharide structures attached to glycoproteins in plants, yeast and mammals. Much effort has been made in engineering yeast (and plants) with

the human N-glycosylation pathway in order to produce therapeutic glycoproteins (Wildt and Gerngross, 2005; Hamilton et al., 2003; Arico et al., 2013; Gomord et al., 2010). Correct N-glycosylation is important for EFR function while it is less important for FLS2 function (Häweker et al., 2010; Sun et al., 2012). Thus yeast surface display might not be a good method to study EFR – elf18 interaction. At this point in time, we cannot confirm that the highly conserved cluster in the EFR LRR domain identified in chapter 2 is the elf18 binding site.

Further YSD optimization ideas

It may be useful to close with a list of further YSD optimization ideas that were relevant to the above project, in case future researchers wish to further pursue this approach.

- Optimize the temperature and time for induction of the proteins of interest
- Optimize the ligand binding buffers (I experimented with using a specific FLS2 – flg22 binding buffer)
- Mutate the glycosylation sites
- Mutate Lys-Arg (KR) sites. KR sites are potential proteolysis sites in yeast
- Optimize fluorescence labelling. I detected a cross reactivity: the secondary antibody anti-mouse PE cross reacted with the primary antibody anti-HA (3F10) from rat. Each antibody needs to be optimized in terms of amount, incubation time and temperature.
- Express a co-receptor such as BAK1 in the yeast surface display system in a way that can allow receptor/ligand/co-receptor interaction.

Acknowledgments

We would like to thank our collaborators Eric Shusta and Thomas Mallot, Chemical and Biological Engineering at UW Madison for sharing their expertise in yeast surface display and providing plasmids, conjugated antibodies and other resources.

References

Arico, C., Bonnet, C., and Javaud, C. (2013). N-glycosylation humanization for production of therapeutic recombinant glycoproteins in *Saccharomyces cerevisiae*. Methods Mol. Biol. **988**: 45–57.

Boder, E.T., Midelfort, K.S., and Wittrup, K.D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc. Natl. Acad. Sci. U. S. A. **97**: 10701–5.

Boder, E.T. and Wittrup, K.D. (1998). Optimal screening of surface-displayed polypeptide libraries. Biotechnol. Prog. **14**: 55–62.

Boder, E.T. and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. **15**: 553–7.

Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D. (2006). Isolating and engineering human antibodies using yeast surface display. Nat Protoc **1**: 755–768.

Cho, B.K., Kieke, M.C., Boder, E.T., Wittrup, K.D., and Kranz, D.M. (1998). A yeast surface display system for the discovery of ligands that trigger cell activation. J. Immunol. Methods **220**: 179–88.

Gomord, V., Fitchette, A.-C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., and Faye, L. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant Biotechnol. J. **8**: 564–87.

Hamilton, S.R., Bobrowicz, P., Bobrowicz, B., Davidson, R.C., Li, H., Mitchell, T., Nett, J.H., Rausch, S., Stadheim, T.A., Wischnewski, H., Wildt, S., and Gerngross, T.U. (2003). Production of complex human glycoproteins in yeast. Science **301**: 1244–6.

Häweker, H., Rips, S., Koiwa, H., Salomon, S., Saijo, Y., Chinchilla, D., Robatzek, S., and von Schaewen, A. (2010). Pattern recognition receptors require N-glycosylation to mediate plant immunity. J. Biol. Chem. **285**: 4629–36.

Jin, M.S. and Lee, J.O. (2008). Application of hybrid LRR technique to protein crystallization. BMB Rep **41**: 353–357.

Kang, J.Y., Nan, X., Jin, M.S., Youn, S.J., Ryu, Y.H., Mah, S., Han, S.H., Lee, H., Paik, S.G., and Lee, J.O. (2009). Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity **31**: 873–884.

Kieke, M.C., Cho, B.K., Boder, E.T., Kranz, D.M., and Wittrup, K.D. (1997). Isolation of anti-T cell receptor scFv mutants by yeast surface display. Protein Eng. **10**: 1303–10.

Kieke, M.C., Shusta, E. V, Boder, E.T., Teyton, L., Wittrup, K.D., and Kranz, D.M. (1999). Selection of functional T cell receptor mutants from a yeast surface-display library. Proc. Natl. Acad. Sci. U. S. A. **96**: 5651–6.

Kim, H.M., Park, B.S., Kim, J.I., Kim, S.E., Lee, J., Oh, S.C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O.J., and Lee, J.O. (2007). Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. Cell **130**: 906–917.

Shusta, E. V, Holler, P.D., Kieke, M.C., Kranz, D.M., and Wittrup, K.D. (2000). Directed evolution of a stable scaffold for T-cell receptor engineering. Nat. Biotechnol. **18**: 754–9.

Sun, W., Cao, Y., Jansen Labby, K., Bittel, P., Boller, T., and Bent, A.F. (2012). Probing the Arabidopsis flagellin receptor: FLS2-FLS2 association and the contributions of specific domains to signaling function. Plant Cell **24**: 1096–113.

Swers, J.S., Kellogg, B.A., and Wittrup, K.D. (2004). Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. Nucleic Acids Res. **32**: e36.

Tachibana, C. and Stevens, T.H. (1992). The yeast EUG1 gene encodes an endoplasmic reticulum protein that is functionally related to protein disulfide isomerase. Mol. Cell. Biol. **12**: 4601–11.

Wildt, S. and Gerngross, T.U. (2005). The humanization of N-glycosylation pathways in yeast. Nat. Rev. Microbiol. **3**: 119–28.

Yoon, S.I., Kurnasov, O., Natarajan, V., Hong, M., Gudkov, A. V, Osterman, A.L., and Wilson, I.A. (2012). Structural basis of TLR5-flagellin recognition and signaling. Science (80-.). **335**: 859–864.

Zaccolo, M., Williams, D.M., Brown, D.M., and Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J. Mol. Biol. **255**: 589–603.

Figures

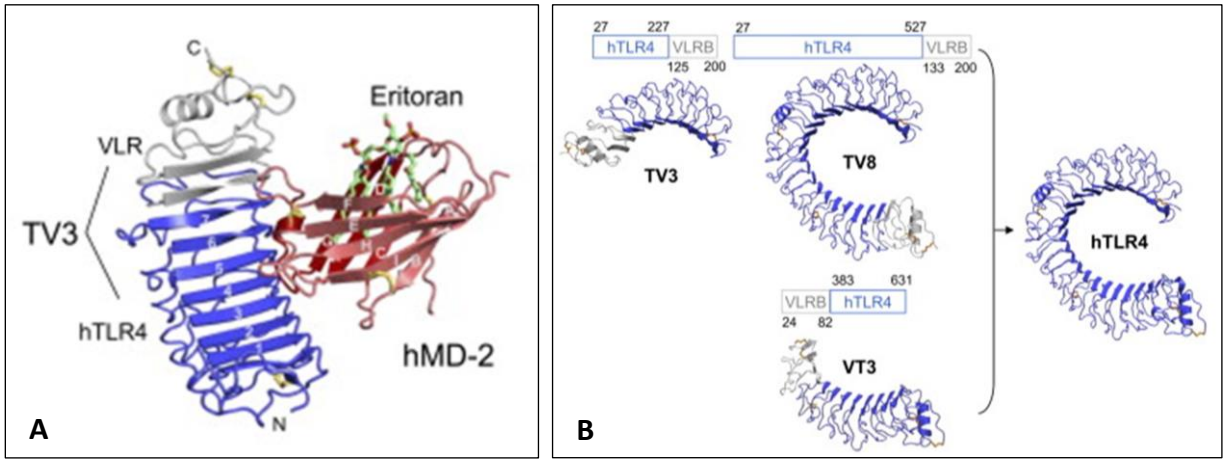

Figure 1 adopted from Kim et al. (2007): TLR ectodomain fragments. **(A)** A TLR4 ectodomain fragment is sufficient to bind the ligand MD-2. **(A+B)** The 3 TLR4 ectodomain fragments are capped with VLRB.61 (grey).

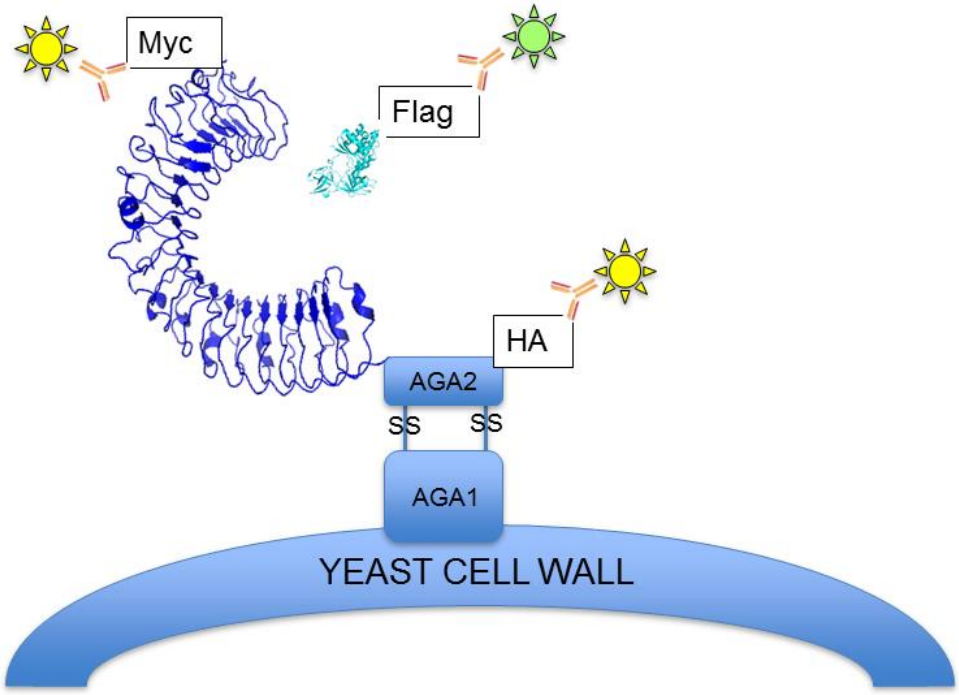

Figure 2: Cartoon of LRR domain displayed on yeast. The displayed constructs have an N-terminal c-Myc tag and a C-terminal HA tag. The ligands are FLAG tagged. Different fluorophore conjugated antibodies are used to visualize the constructs and the ligands. The constructs are anchored to the cell wall with disulfide bonds between the cell wall component AGA1 and the construct component AGA2.

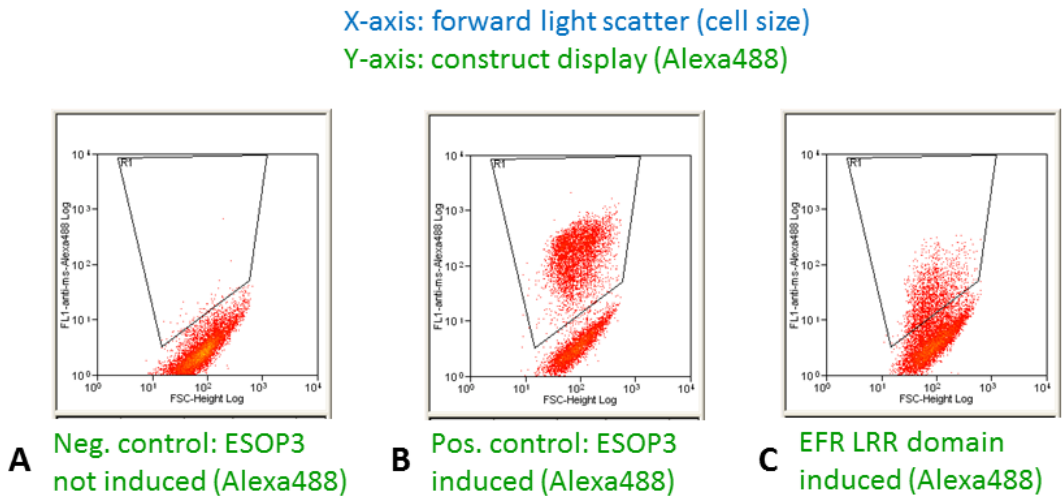

Figure 3: Display of the wild-type EFR LRR domain. Section R1 is based on the negative control, where no construct is displayed. Fluorescence outside R1: cell autofluorescence. Fluorescence inside R1: Alexa488. **(A)** No construct displays, autofluorescence only. **(B)** ESOP3 displays well. **(C)** Low display of the wild-type EFR LRR domain.

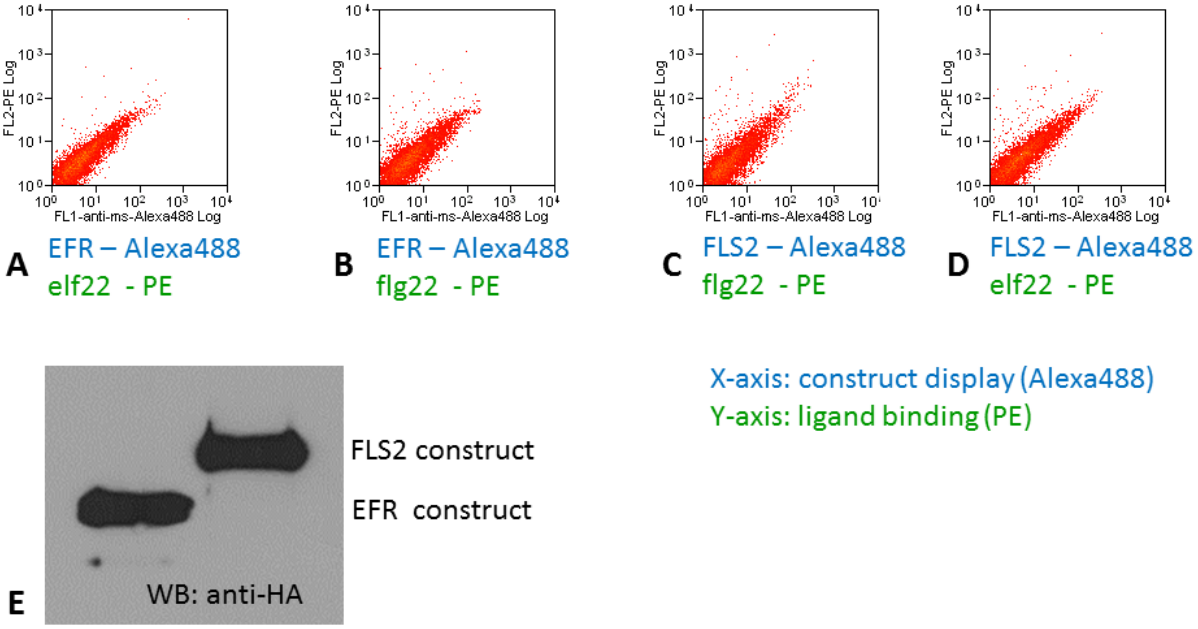

Figure 4: Low display of EFR and FLS2 LRR domains and no detectable ligand binding. (A) EFR does not display well and no elf22 binding is detectable. **(B)** EFR does not display well. No un-specific flg22 binding. **(C)** FLS2 does not display well and no flg22 binding is detectable. **(D)** FLS2 does not display well. No un-specific elf22 binding. **(E)** The EFR and FLS2 constructs were cut off the yeast cells and subjected to SDS-PAGE. Low protein levels of the constructs were detectable by Western Blot.

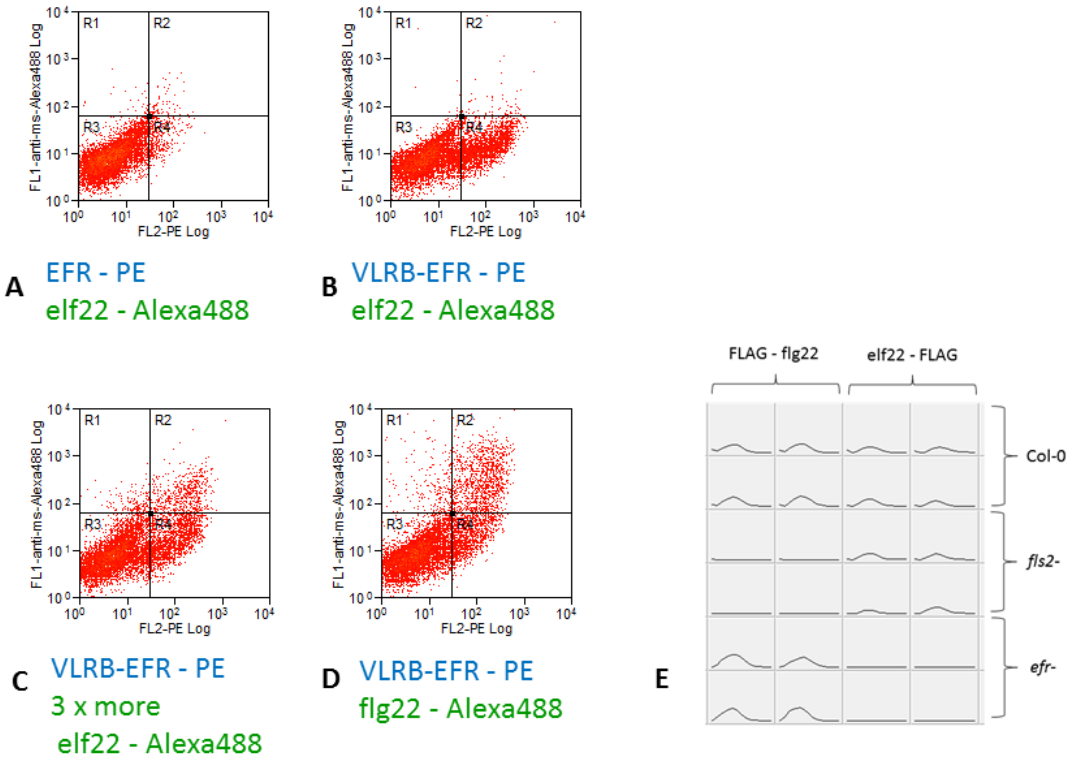

Figure 5: VLRB capped EFR LRR domain displays better than the uncapped EFR. X-axis: construct display (PE). Y-axis: ligand binding (Alexa488). Quadrants: R3: non-specific signal (cell autofluorescence). R4: construct display, no ligand binding. R1: ligand binding to yeast cells. R2: ligand binding to construct. **(A)** Wild-type EFR does not display well and no elf22 binding is detectable. **(B)** VLRB capped EFR displays well, but no elf22 binding is detectable. **(C)** VLRB capped EFR displays well and some elf22 binding is detectable when 3x more ligand than standard amount is added. **(D)** VLRB capped EFR displays well and flg22 (at standard amount) binding is detectable. **(E)** ROS assay demonstrates activity and specificity of the FLAG tagged ligands in Col-0, *fls2*$^{-}$ and *efr*$^{-}$ Arabidopsis plants. X-axis: time scale 0 – 30 min after addition of ligand. Y-axis: luminescence.

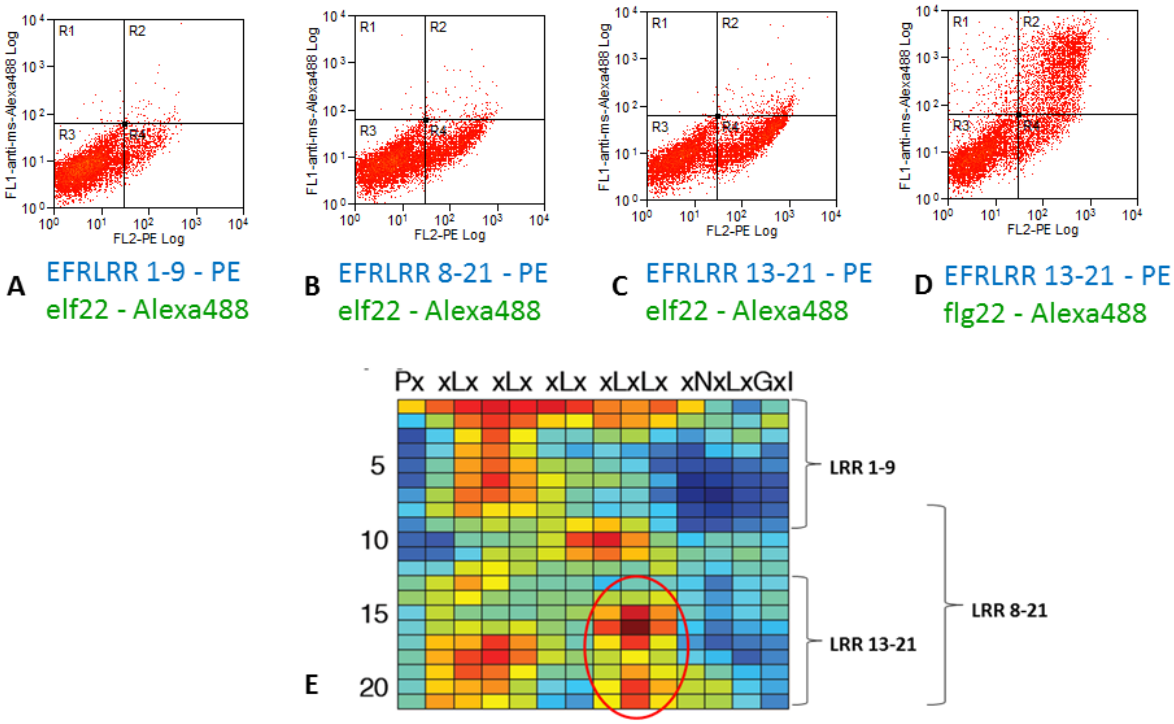

Figure 6: Fragment EFRLRR 13-21 displays well but has higher affinity to flg22 than elf22. X-axis: construct display (PE). Y-axis: ligand binding (Alexa488). Quadrants: R3: non-specific signal (cell autofluorescence). R4: construct display, no ligand binding. R1: ligand binding to yeast cells. R2: ligand binding to construct. **(A)** Fragment EFRLRR 1-9 does not display well and no elf22 binding is detectable. **(B)** Fragment EFRLRR 8-21 displays well but no elf22 binding is detectable. **(C)** Fragment EFRLRR 13-21 displays well but no elf22 binding is detectable. **(D)** Fragment EFRLRR 13-21 displays well and flg22 binding is detectable. **(E)** EFR LRR domain repeat conservation map (RCM) (see chapter 2). The red circle indicates the conserved site in the concave face. The repeats included in the three EFR LRR domain fragments are indicated.

Chapter 4: Towards engineering pattern-recognition receptors with novel ligand specificities

Abstract

Genetic modification is a powerful technique to improve the immune system of agriculturally important susceptible crop cultivars. So far genetic modification to improve plant health has mainly been used to add exogenous genes to susceptible cultivars. There are only a few examples where an endogenous R gene was engineered toward improved pathogen perception. In this study we explored two techniques to engineer pattern-recognition receptors with novel ligand recognition abilities. A first approach used yeast surface display. The idea was to create EFR and FLS2 LRR domain mutant libraries and screen them for novel ligand specificities. The wild type EFR and FLS2 LRR domains did not display well on yeast. Display of the EFR LRR domain was improved by capping the N-terminus with the VLRB.61 N-terminus. However, no elf18 ligand binding could be detected in these experiments, and thus we did not proceed with the generation of mutant libraries. In our second approach, we generated chimeric receptors for expression in plants using full-length FLS2 and EFR as scaffolds that carried targeted LRR amino acid changes. The templates for these amino acid changes were the ligand binding sites of variable lymphocyte receptors (VLRs). VLRs are the equivalents of antibodies, but are derived from the unique adaptive immune system of jawless vertebrates and mainly consist of LRRs. As a prototype, we created different versions of a chimeric receptor made of FLS2 and VLRA.R2.1, a well-characterized VLR able to bind hen egg white lysozyme (HEL) with high affinity. The chimeric receptors were expressed in *Nicotiana benthamiana* and *Arabidopsis thaliana*. The proteins accumulated, but no defense response was observed when the transformed plants were inoculated with the novel ligand HEL. EndoH assay revealed

glycosylation defects of the maturing chimeric proteins; and thus, most likely the receptor proteins were degraded in the Golgi, before they reached the plasma membrane. So far we were not able to integrate the chimeric receptors into the plant immune system. The alteration of FLS2 towards novel ligand specificity is further complicated by the recent discovery that BAK1 is a co-receptor of FLS2 for flg22 perception. In the future *in silico* techniques might be used for a precise ectodomain signaling complex design, based on the available crystal structures of FLS2-ECD together with BAK1-ECD and flg22 and VLR ligand binding residues, to engineer FLS2 to recognize novel ligands.

Introduction

New approaches to strengthen the plant immune system

Classical plant protection agents sprayed on crops aim at killing off pathogens and pests. A different approach of keeping crops healthy is to strengthen the plant's immune system. Missing components, e.g. genes encoding receptors able to detect crucial effectors, can be added to susceptible plants of interest by genetic modification. Targeted genetic modifications can offer very subtle and specific change to the plant genome, and circumvent the problems of classical breeding, where susceptible and resistant plant cultivars are crossed, which not only introduces the wanted resistance characteristics to the progeny plants, but also many unwanted traits. Genetic engineering of plants is a relatively new and very promising approach to improve crop health. The best known and widely used genetic modification for plant

protection is the introduction of genes encoding *Bacillus thuringiensis* (Bt) toxins to corn, soybean and cotton cultivars (reviewed in Sanahuja et al., 2011). The toxins kill insect larvae feeding on plant leaves. Thus instead of the farmers spraying the toxins onto the crops, the plants produce the toxins themselves. A similar approach was used by Ko et al. by adding a gene encoding an antimicrobial compound attacin E from moth to the genome of the *Erwinia amylovora* susceptible apple cultivar 'Galaxy' (Ko et al., 2000). The genetically modified apples are not commercially available. Other examples of genetically modified plants with improved disease resistance include plant cultivars with added genes encoding PRRs or R genes (see examples below), added genes encoding viral coat proteins (see examples below) or added resistance genes with unknown functions (e.g. Cook et al., 2012; D. M. Horvath et al., 2012; Tripathi et al., 2010). The added genes were obtained from different sources e.g. from a different cultivar of the same plant species, from a wild relative of the same plant genus, from a plant species of a different plant family or from an entirely different organism belonging to a different kingdom.

Examples of genetically modified plants expressing a viral coat protein

In the 1990s squash and papaya plants were successfully engineered towards virus resistance. The transgenic squash cultivars express coat proteins from the cucumber mosaic virus (CMV), watermelon mosaic virus 2 (WMV 2) and zucchini yellow mosaic virus (ZYMV) and are resistant to CMV, WMV2 and ZYMV (Tricoll et al., 1995). The transgenic papaya cultivars express the coat protein of papaya ringspot virus (PRV) and thus are resistant to PRV (Lius et al., 1997). The transgenic squash and papaya cultivars were approved for commercial use, as were

potato lines with both Colorado Potato Beetle and potato leafroll virus resistance (Kaniewski and Thomas, 1998). What at the time probably seemed like the start of a new era of commercially available crops with engineered disease resistance, was unfortunately rather already the end of it. None of the plants subsequently genetically modified for enhanced disease resistance have been successfully commercialized and widely used (does not apply to genetically modified insect resistant Bt crops).

Examples of genetically modified plants carrying exogenous R proteins

Resistance to the stem rust fungus *Puccinia graminis f. sp. tritici* was improved in barley by expressing the gene *Rpg1* from barley cultivar Morex in the susceptible barley cultivar Golden Promise (Horvath et al., 2003). *Rpg1* does not encode a typical NB-LRR protein, but a cytoplasmic receptor-like kinase with two tandem kinase domains (Nirmala et al., 2006).

The NB-LRR encoding gene *Rxo1* was originally found in maize. Transgenic rice plants expressing *Rxo1* are more resistant to *Xanthomonas oryzae pv. oryzicola* (Zhao et al., 2005).

The NB-LRR encoding gene *RB* originally found in the wild potato species *Solanum bulbocastanum* was expressed in cultivated potato (*Solanum tuberosum*). The leaves of the genetically modified potato plants showed enhanced resistance to the late blight causing pathogen *Phytophthora infestans* (Song et al., 2003; Halterman et al., 2008; Bradeen et al., 2009).

The NB-LRR encoding gene *Mr5* originally found in crab apple enhances resistance to the fire blight causing bacterium *Erwinia amylovora* when expressed in the susceptible apple cultivar 'Gala' (Broggini et al., 2014).

Often the addition of a single R protein is not enough to improve plant health over an extended time period. Fast breakdown of resistance has been observed in many cases (mainly in crops where the R gene was added by classical breeding), because of selection of resistant pathogen strains, not carrying the recognizable effectors (McDonald and Linde, 2002). Thus a promising approach is the stacking of several R genes. An example would be the stacking of the genes *Sr35* (Saintenac et al., 2013), *Sr33* (Periyannan et al., 2013) and *Sr2* (Ayliffe et al., 2013) into a high yield wheat cultivar. The three genes are involved in conferring resistance to the wheat rust isolate Ug99, which is a current thread to world wheat production (Singh et al., 2011).

Examples of genetically modified plants carrying exogenous PRRs

In 2010 Lacombe et al. transferred the PRR EFR from Arabidopsis to tomato and *N.benthamiana* plants. The transgenic tomato and *N.benthamiana* plants were more resistant to pathogenic bacteria from different genera (Lacombe et al., 2010).

Fradin et al. expressed Ve1, a LRR-RLP from tomato able to recognize *Verticillium dahliae* and *Verticillium albo-atrum* race 1 in Arabidopsis. *Ve1*-trangenic Arabidopsis plants were resistant to *Verticillium dahliae* and *Verticillium albo-atrum* race1 (Fradin et al., 2011).

Engineering PRRs and R proteins with novel ligand specificities

Drugs (excluding antibiotics) that alleviate human diseases often aim at the inhibition of specific proteins in the human cells. Typical drug targets are protein kinases and G-protein-coupled receptors (Overington et al., 2006). Drug design often aims at finding agonistic or antagonistic receptor ligands. So far it is not possible to alter human receptors by gene therapy. However, in plants we have this possibility. We can engineer plant proteins and express them in the plant cultivar of interest. Engineering plants for improved disease resistance so far has been done primarily by adding exogenous genes to the plant genome of interest (see examples above), and less by engineering endogenous plant proteins to favorably alter their properties. Similarly, plants engineered for improved biofuel characteristics (e.g. altered lignin content, altered starch compositions) are modified by the addition or deletion of genes encoding crucial enzymes involved in the pathway of interest (e.g. lignin synthesis, starch synthesis), and not by enzyme engineering (Santelia and Zeeman, 2011; Sonnewald and Kossmann, 2013; Zeng et al., 2014). There are very few examples in the literature of engineered R proteins with novel ligand binding abilities and so far there are no examples of engineered PRRs.

<u>Examples of engineered R proteins:</u>

In 2006 Farnham and Baulcombe used error-prone PCR to create a mutant library of the LRR domain of *Rx*. *Rx* encodes an NB-LRR protein that recognizes the coat protein of a potato virus X (PVX) strain. By screening the library they obtained *Rx* alleles that encoded receptors able to recognize the coat protein from an additional PVX strain and furthermore the coat protein from the poplar mosaic virus (PoMV) (Farnham and Baulcombe, 2006). Harris et al. then further

evolved some of the obtained full-length *Rx* variants by random mutagenesis. They obtained *Rx* alleles with mutations in the N-terminal domain which improved receptor signaling (Harris et al., 2013).

R3a is an NB-LRR protein found in potato that recognizes the *Phytophthora infestans* RXLR-effector $AVR3a^{KI}$ but does not recognize a similar effector $AVR3a^{EM}$ which differs in two amino acids (Armstrong et al., 2005; Bos et al., 2006; Bos et al., 2009). In 2014 Segretin et al. screened a library of R3a variants obtained by random mutagenesis and obtained several R3a variants that now recognized both $AVR3a^{KI}$ and $AVR3a^{EM}$ (Segretin et al., 2014). One of the obtained R3a variants additionally was able to recognize an AVR3a homolog from *Phytophthora capsici*.

Stirnweis et al. discovered two crucial amino acids in the ARC2 subdomain of the wheat powdery mildew resistance protein Pm3 responsible for the switch from narrow to broad resistance spectrum. The resistance spectrum of Pm3f was broadened by changing the two amino acids in the ARC2 domain (Stirnweis et al., 2014).

In the future modern genome editing tools like TALEN and CRISPR could be used to edit endogenous plant genes (Belhaj et al., 2013; Schornack et al., 2013; Shan et al., 2013). This would be an incredibly subtle change to the plant genome of interest and would circumvent the problems of gene insertion at random loci in the genome and of insertion of exogenous DNA, which is often looked upon skeptically by the public.

Objective

The objective of this study was to explore techniques to generate PRRs with novel ligand specificities. We were aware of the difficulty of this undertaking, and currently there are no reports of a PRR successfully engineered for novel specificity. In a first approach we explored yeast surface display as a technique to evolve the ectodomains of EFR and FLS2 towards novel ligand recognition abilities. As a second approach we explored the use of exogenous LRR domains from VLRs to generate chimeric receptors.

Approach 1: In vitro evolution of EFR and FLS2 using yeast surface display

Project outline

I intended to use the yeast surface display method for two separate research projects: the EFR – elf18 binding study (chapter 3) and the evolution of EFR and FLS2 ligand binding capacities (the present chapter 4). I have decided to include all the yeast surface data in chapter 3 in an effort to make this thesis more reader-friendly. Some of the experimental work discussed in the present chapter 4 concerns results that are presented in chapter 3.

The following is an outline of the *in vitro* evolution of EFR and FLS2 ligand binding capacities project as envisioned:

Step 1: Create plasmids to display the wild type EFR and FLS2 LRR domains on yeast (see chapter 3)

Step 2: Do error-prone PCR mutagenesis on the cDNAs encoding the *EFR* and *FLS2* LRR domains to create libraries of approx. 10^7 different alleles. Express each allele in a separate yeast cells. The yeast cells then display the different protein domain variants, which results in EFR and FLS2 LRR protein libraries.

Step 3: Use fluorescence-activated cell sorting (FACS) to select protein variants with favorable ligand binding phenotypes. Any ligand of interest can be used to screen the libraries.

Step 4: Perform several rounds of error prone PCR and FACS to enrich the libraries with allele/protein variants of interest (see lists below).

Step 5: Isolate the final alleles of interest and cloned them back into their original receptor gene in place of the original LRR domain.

Step 6: Express the generated recombinant genes in *N.benthamiana* and Arabidopsis. Tested the transformed plants for defense responses when treated with the correspondent receptor ligands.

Protein variants of interest

EFR LRR domain library:

- LRR domains with higher affinity for elf18 than wild type EFR
- LRR domains able to bind the elf18 peptides based on the EF-Tu sequences from *P.syringae* pv *tomato* DC3000 and *Xylella fastidiosa*. The elf18 peptides based on the EF-TU sequences of these species have very low activities in Arabidopsis (Kunze et al., 2004)

FLS2 LRR domain library:

- LRR domains with higher affinity for flg22 than wild type FLS2
- LRR domains able to bind the previously unrecognizable flg22 peptides based on the flagellin sequences of *Erwinia amylovora, Ralstonia solanacearum* and/or *Xanthomonas campestris campestris* strain 186. The flg22 peptides based on the flagellin sequences of these species have very low activities in Arabidopsis (Pfund et al., 2004; Sun et al., 2006).

Discussion

As discussed in chapter 3, I was able to display the EFR LRR domain on yeast by capping it with a VLR cap. However, no specific binding to elf18 was detected. With the achieved EFR display, it would be possible to proceed to create the mutant libraries (Eric Shusta personal communication). We theoretically could create EFR LRR domains with novel ligand binding abilities; however, the ligand binding ability might disappear when the LRR domain alleles later would be expressed in *N.benthamiana* and Arabidopsis, because the plant proteins would have a different *N*-glycan decoration than the yeast proteins had. The different N-glycosylation could impact ligand binding abilities. Correct N-glycosylation is particularly important for EFR processing and function, but it is less important for FLS2 processing and function (see chapter 5 and Haweker et al., 2010; Sun et al., 2012).

We now know that flg22 not only binds to FLS2, but additionally to BAK1 (Sun et al., 2013). The FLS2 – flg22 – BAK1 ECD complex presumably enables phosphorylation of the intracellular

kinase domains which initiates downstream signaling. A similar EFR-ECD – elf18 – BAK1-ECD complex might initiate EFR signaling. Thus, this significantly complicates receptor evolution using yeast surface display.

In order to change FLS2 ligand binding ability, one possibility would be to only display and evolve the ligand binding LRRs instead of the full-length LRR domain. This would keep the BAK1 interaction site and potentially other important sites in the LRR domains (e.g. potential dimerization sites) intact. As shown by the FLS2-ECD – flg22 crystal structure, there are no glycans involved in ligand perception. Thus, displaying only the ligand binding LRRs has the additional advantage of less glycosylation sites than the full-length FLS2 LRR domain. Remaining glycosylation sites could be mutated prior to yeast display. Thus creating a novel ligand binding site should be feasible. However, the generation of a functional receptor is a bigger challenge now that we know that the ligand has to interact with BAK1 and most likely other SERK proteins as well. Further studies need to address the question of how specific the ligand – SERK protein interaction has to be in order to activate the receptor complex e.g. is a slight touch of ligand and SERK enough to initiate downstream signaling or does the interaction need to be very fine-tuned for instance to position the intracellular kinase domains into an very specific position for correct phosphorylation?

Currently, there is no crystal structure available for the EFR-ECD together with elf18 (and BAK1-ECD). Furthermore, EFR is very susceptible to glycosylation defects. Thus, at this point in time it is even more challenging to engineer EFR towards novel ligand recognition abilities using yeast surface display.

Approach 2: VLR chimeric receptors

Introduction

Variable lymphocyte receptors (VLRs)

The adaptive immune system of ancient jawless vertebrates (hagfish and lamprey) does not rely on antibodies and T-cell receptors for antigen recognition, but instead uses variable lymphocyte receptors (VLRs) that are generated by combinatorial assembly of LRR cassettes (Figure 1A) (Herrin and Cooper, 2010). Best estimates are that a repertoire of > 10^{17} distinct antigen receptors can be generated, which is sufficiently diverse to recognize most, if not all, pathogens (Herrin and Cooper, 2010). The VLRs are membrane-bound proteins. They are structured as follows: an N-terminal cap (LRRNT), an 18 amino acid LRR1, then multiple 24 amino acid LRRV modules, a connecting peptide (CP) LRR, a C-terminal cap (LRRCT), and a stalk region rich in threonine and proline residues. A single VLR contains between 0 – 9 LRRV modules and each module has the consensus sequence xLxxLxxLxLxxNxLxxLPxxxFx (Pancer et al., 2004; Alder et al., 2005).

Studies of the ancient immune system of hagfish and lamprey have improved our understanding of the evolution of adaptive immunity (Boehm et al., 2011; Kasahara and Sutoh, 2014). Furthermore, efforts are ongoing to use VLRs as alternatives to immunoglobulin antibodies in biotechnological applications such as diagnostics, therapeutics and medical research. Ligand affinities of VLRs are comparable to ligand affinities of immunoglobulins (Han

et al., 2008). However, VLRs have the advantages of being composed of a single polypeptide chain and of having a lower molecular weight. Wezner-Ptasińska at al. and Lee et al. designed protein scaffolds based on VLRs (Wezner-Ptasińska et al., 2011; Lee et al., 2012). The scaffolds termed “dVLR” and “Repebody”, respectively, now can be engineered toward novel ligand specificities. Both “dVLRs” and “Repebodies” exhibit high thermodynamic and pH stabilities and high soluble expressions in bacteria: characteristics that make them great candidates for biotechnological applications. Two “Repebodies” have been successfully engineered, one able to bind hen egg lysozyme and the other able to bind myeloid differentiation protein-2 (Lee et al., 2012).

Hong et al. selected several VLRs specific for different types of biomedically relevant glycans. These VLRs were named “Lambodies” and can now be used for diagnostics and biomedical research (Hong et al., 2013).

Several crystal structures of VLRs together with their ligands have been solved. VLRA.R2.1 (Figure 1C) and VLRB.2D both bind hen egg white lysozyme (HEL), but they recognize different HEL epitopes (Deng et al., 2010; Velikovsky et al., 2009). VLR.RBC36 binds H-trisaccharide (Han et al., 2008) and VLR4 binds BclA (Kirchdoerfer et al., 2012). Additionally there is a crystal structure available in the protein data bank (PDB) of a synthetic, unpublished VLRB (PDB ID: 4J4L) together with a ligand.

Chimeric receptors

Several recent studies demonstrate the feasibility of constructing functional chimeric receptors for plant cells containing the ligand binding domain of one receptor and the kinase domain of another receptor. Examples: Kishimoto et al. created chimeric receptors composed of CEBiP and Xa21, two PRRs from rice (Kishimoto et al., 2010). Brutus et al. created chimeric receptors composed of EFR and FLS2 and furthermore of EFR and WARK1 (Brutus et al., 2010). Albert et al., 2010 created chimeric receptors composed of EFR and FLS2 as well (Albert et al., 2010). Mueller et al. created chimeric receptors composed of FLS2 from Arabidopsis and FLS2 from tomato (Mueller et al., 2012). In chapter 2 we created chimeric receptors using the LRR domains from different Brassicaceae EFR and the transmembrane and kinase domains of EFR from Arabidopsis.

In the examples above, the chimeric receptors were mainly used as a biotechnological tool to study receptor function, specificity and signaling.

Objective

The objective of this study was to create PRRs with novel ligand specificities by using a chimeric receptor approach. We combined gene parts from Arabidopsis FLS2 and *Petromyzon marinus* (sea lamprey) VLRA.R2.1 and expressed the recombinant genes in Arabidopsis and *N. benthamiana*. The long-term goal is to immunize lamprey with an antigen from a plant

pathogen of interest, harvest the antigen binding VLR and engineer the VLR ligand binding site in a plant PRR in order to generate a novel receptor able to recognize the antigen of interest.

Methods

Domain swap of the EFR LRR domain with the VLRA.R2.1 LRR domain

The synthetic 771 bp VLRA.R2.1 gene was ordered from GenScript. The codons were optimized for Arabidopsis and the gene was flanked by the restriction enzyme sites Sbf1 and Fse1. The pGWB13 – EFR (Sbf1, Fse1) plasmid from chapter 2 was used to cut out the entire LRR domain from Arabidopsis EFR (AtEFR) and the VLRA.R2.1 LRR domain was ligated in its place, as in Figure 2A. The final construct contained the signal peptide encoding sequence from AtEFR, the N - and C - terminal LRR caps and all LRRs from VLRA.R2.1 and the transmembrane and kinase domains from AtEFR. The binary plasmid pGWB13-EFR-VLR.R2 was used for *N. benthamiana* transient leaf transformation and for stable Arabidopsis Col-0 and *efr*$^{-}$ transformation. This chimeric gene was named chimera H.

Exchange of FLS2 LRRs with VLRA.R2.1 LRRs

Overlap-extension PCR was used to make the chimeric genes. **Chimera E**: FLS2 LRRs 22 to 28 were replaced by the nine VLRA.R2.1 LRRs and the VLRA.R2.1 α-helix and loop were added at the C-terminus of the FLS2 LRR domain. **Chimera F**: FLS2 LRRs 8 to 14 were replaced by the nine VLRA.R2.A LRR and the α-helix and the loop were added.

Exchange of single FLS2 amino acids with single VLRA.R2.1 amino acids

Three synthetic gene parts (sgp) were ordered from GenScript: sgp A (612bp), sgp C (675bp) and sgp G (609 bp). Overlap-extension PCR was used to make the chimeric genes. **Chimera A**: sgp A was used. The sgp was placed in lieu of FLS2 LRRs 7 to 15. The VLRA.R2.1 α-helix was not included and the VLRA.R2.1 loop was engineered onto FLS2 LRR 15. **Chimera B**: sgp G and C were used. The sgps were placed in lieu of FLS2 LRRs 13-21. The VLRA.R2.1 α-helix and loop were both included. **Chimera C**: sgp C was used. The sgp was placed in lieu of FLS2 LRRs 21-28, the very last LRRs of FLS2, so the VLRA.R2.1 α-helix and loop were included by placing them at the FLS2 LRR domain terminus. **Chimera G**: sgp G was used. The sgp was placed in lieu of FLS2 LRRs 13-21. The VLRA.R2.1 α-helix was not included and the VLRA.R2.1 loop was engineered onto FLS2 LRR 21.

The chimeric genes were cloned into pCR8/GW/TOPO plasmids (Invitrogen). LR clonase (Invitrogen) was used to transfer the chimeric genes into the binary plasmids pGWB14: 35S, C-3xHA for transient transformation of *N.benthamiana* leaves and stable transformation of Arabidopsis Col-0 and *fls2*$^{-}$ plants. Furthermore the chimeric genes were transferred into pUC-GW14: 35S, C-3xHA for Arabidopsis Col-0 and *fls2*$^{-}$ protoplast transformation.

Results

Three different strategies were used to create the chimeric receptors (Figure 2). The first strategy was a drastic change by replacing the whole ectodomain of EFR with VLRA.R2.1

(Figure 2A). The second approach was more subtle. Nine LRRs of FLS2 were replaced by the nine LRRs of VLRA.R2.1. The replacement was done on two different sites in the FLS2 LRR domain, once at LRR 8 to 14 and the other time at the LRR domain C-terminal LRRs (Figure 2B). Even a more subtle change was used in the third strategy were single solvent exposed amino acids in the concave face of FLS2 were replaced by the ligand binding and the ligand binding adjacent amino acids of VLRA.R2.1. Four different chimeras were generated by altering insertion site and integration strategy of the α-helix and the loop (Figure 2C). Altogether seven different chimeric genes were generated. The genes were expressed in *N. benthamiana* and all chimeras were detectable by Western blot (Figure 3). ROS assay was performed on the transiently transformed *N. benthamiana* leaves. Adding the new ligand HEL did not elicit an ROS burst (Figure 4). The chimeras were further tested in Arabidopsis protoplasts, but none of the tested chimeras exhibited a response to HEL (data not shown). EndoH assay in transformed *N. benthamiana* leaves and Arabidopsis seedlings revealed a glycosylation defect of the chimeric proteins (Figure 5), which suggests that the proteins are degraded in the Golgi before they can reach the plasma membrane.

Discussion and future directions

The chimeric receptors all accumulated in transiently transformed *N.benthamiana* leaves and stably transformed Arabidopsis plants. However, the receptors were not successfully integrated into the immune system of these plant species. The EndoH assay results suggest that the chimeric receptors were not correctly glycosylated, which most likely resulted in the

degradation of the proteins before they reached the plasma membrane. Even the most subtle changes of the third strategy resulted in a degradation of the chimeric receptors. Thus the changes were probably still too drastic to be accepted by the protein quality control chaperones. Alternatively our protein design resulted in misshaped LRR domains. We did not use computer modeling to test our design *in silico*.

It would be interesting to test if the chimeric receptors can bind the novel ligand HEL. We did not perform a binding study. If the receptors are not localized correctly at the plasma membrane, the addition of radiolabeled ligand would not give us any conclusive results. We would have to perform the binding study in an *in vitro* system. We did investigate using surface plasmon resonance (SPR), Co-IP or yeast surface display but did not pursue with the setup of experiments.

The chimeric FLS2 – VLRA.R2.1 receptors were designed and generated before the co-crystal structure of FLS2 ECD – flg22 – BAK1 ECD (Sun et al., 2013) was available. Now the design of chimeric receptors could be done more precisely by altering the flg22 binding amino acids. The crystal structure reveals an interaction of BAK1 with the ligand flg22, thus in addition to engineering a novel ligand binding site in FLS2, an investigator would have to ensure the novel ligand can interact with BAK1 as well. It might be exceedingly difficult to find a ligand that fits between FLS2 and BAK1. However, computer modeling can be used to do some investigation by *in silico* modeling of the FLS2 ECD – new ligand – BAK1 ECD complex. If the needed amino acid changes in the FLS2 ectodomain are subtle enough, and *in silico* models predict a correctly folded LRR domain, it may be possible to engineer a chimeric receptor that passes through endogenous protein quality control mechanisms and localizes correctly at the

plant cell plasma membrane. Such a project may be feasible, for example, to seek engineered recognition of only slightly different flg22 domains, from flagellins of pathogenic bacteria that presently escape detection by known FLS2 proteins.

References

Albert, M., Jehle, A.K., Mueller, K., Eisele, C., Lipschis, M., and Felix, G. (2010). *Arabidopsis thaliana* pattern recognition receptors for bacterial elongation factor Tu and flagellin can be combined to form functional chimeric receptors. J Biol Chem **285**: 19035–19042.

Alder, M.N., Rogozin, I.B., Iyer, L.M., Glazko, G. V, Cooper, M.D., and Pancer, Z. (2005). Diversity and function of adaptive immune receptors in a jawless vertebrate. Science **310**: 1970–1973.

Armstrong, M.R., Whisson, S.C., Pritchard, L., Bos, J.I.B., Venter, E., Avrova, A.O., Rehmany, A.P., Böhme, U., Brooks, K., Cherevach, I., Hamlin, N., White, B., Fraser, A., Lord, A., Quail, M.A., Churcher, C., Hall, N., Berriman, M., Huang, S., Kamoun, S., et al. (2005). An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. Proc. Natl. Acad. Sci. U. S. A. **102**: 7766–71.

Ayliffe, M., Singh, D., Park, R., Moscou, M., and Pryor, T. (2013). Infection of *Brachypodium distachyon* with selected grass rust pathogens. Mol. Plant. Microbe. Interact. **26**: 946–57.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods **9**: 39.

Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M., and Cooper, M.D. (2012). VLR-based adaptive immunity. Annu. Rev. Immunol. **30**: 203–20.

Bos, J.I.B., Chaparro-Garcia, A., Quesada-Ocampo, L.M., McSpadden Gardener, B.B., and Kamoun, S. (2009). Distinct amino acids of the *Phytophthora infestans* effector AVR3a condition activation of R3a hypersensitivity and suppression of cell death. Mol. Plant. Microbe. Interact. **22**: 269–81.

Bos, J.I.B., Kanneganti, T.-D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M.R., Birch, P.R.J., and Kamoun, S. (2006). The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in Nicotiana benthamiana. Plant J. **48**: 165–76.

Bradeen, J.M., Iorizzo, M., Mollov, D.S., Raasch, J., Kramer, L.C., Millett, B.P., Austin-Phillips, S., Jiang, J., and Carputo, D. (2009). Higher copy numbers of the potato RB transgene correspond to enhanced transcript and late blight resistance levels. Mol. Plant. Microbe. Interact. **22**: 437–46.

Broggini, G.A.L., Wöhner, T., Fahrentrapp, J., Kost, T.D., Flachowsky, H., Peil, A., Hanke, M.-V., Richter, K., Patocchi, A., and Gessler, C. (2014). Engineering fire blight resistance into the

apple cultivar “Gala” using the FB_MR5 CC-NBS-LRR resistance gene of *Malus × robusta 5*. Plant Biotechnol. J. **12**: 728-733

Brutus, A., Sicilia, F., Macone, A., Cervone, F., and De Lorenzo, G. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. Proc Natl Acad Sci U S A **107**: 9452–9457.

Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A.M., Wang, J., Hughes, T.J., Willis, D.K., Clemente, T.E., Diers, B.W., Jiang, J., Hudson, M.E., and Bent, A.F. (2012). Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science **338**: 1206–9.

Deng, L., Velikovsky, C.A., Xu, G., Iyer, L.M., Tasumi, S., Kerzic, M.C., Flajnik, M.F., Aravind, L., Pancer, Z., and Mariuzza, R.A. (2010). A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey. Proc Natl Acad Sci U S A **107**: 13408–13413.

Farnham, G. and Baulcombe, D.C. (2006). Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. Proc. Natl. Acad. Sci. U. S. A. **103**: 18828–33.

Fradin, E.F., Abd-El-Haliem, A., Masini, L., van den Berg, G.C.M., Joosten, M.H.A.J., and Thomma, B.P.H.J. (2011). Interfamily transfer of tomato Ve1 mediates *Verticillium resistance* in Arabidopsis. Plant Physiol. **156**: 2255–65.

Halterman, D.A., Kramer, L.C., Wielgus, S., and Jiang, J. (2008). Performance of transgenic potato containing the late blight resistance gene RB. Plant Dis. **92**: 339–343.

Han, B.W., Herrin, B.R., Cooper, M.D., and Wilson, I.A. (2008). Antigen recognition by variable lymphocyte receptors. Science **321**: 1834–1837.

Harris, C.J., Slootweg, E.J., Goverse, A., and Baulcombe, D.C. (2013). Stepwise artificial evolution of a plant disease resistance gene. Proc. Natl. Acad. Sci. U. S. A. **110**: 21189–94.

Häweker, H., Rips, S., Koiwa, H., Salomon, S., Saijo, Y., Chinchilla, D., Robatzek, S., and von Schaewen, A. (2010). Pattern recognition receptors require N-glycosylation to mediate plant immunity. J. Biol. Chem. **285**: 4629–36.

Herrin, B.R. and Cooper, M.D. (2010). Alternative adaptive immunity in jawless vertebrates. J Immunol **185**: 1367–1374.

Hong, X., Ma, M.Z., Gildersleeve, J.C., Chowdhury, S., Barchi, J.J., Mariuzza, R.A., Murphy, M.B., Mao, L., and Pancer, Z. (2013). Sugar-binding proteins from fish: selection of high affinity “lambodies” that recognize biomedically relevant glycans. ACS Chem. Biol. **8**: 152–60.

Horvath, D.M., Stall, R.E., Jones, J.B., Pauly, M.H., Vallad, G.E., Dahlbeck, D., Staskawicz, B.J., and Scott, J.W. (2012). Transgenic resistance confers effective field level control of bacterial spot disease in tomato. PLoS One **7**: e42036.

Horvath, H., Rostoks, N., Brueggeman, R., Steffenson, B., von Wettstein, D., and Kleinhofs, A. (2003). Genetically engineered stem rust resistance in barley using the Rpg1 gene. Proc. Natl. Acad. Sci. U. S. A. **100**: 364–9.

Kaniewski, W.K. and Thomas, P.E. (1998). Field testing resistance of transgenic plants. Methods Mol. Biol. **81**: 497–507.

Kasahara, M. and Sutoh, Y. (2014). Two forms of adaptive immunity in vertebrates: similarities and differences. Adv. Immunol. **122**: 59–90.

Kirchdoerfer, R.N., Herrin, B.R., Han, B.W., Turnbough, C.L., Cooper, M.D., and Wilson, I.A. (2012). Variable lymphocyte receptor recognition of the immunodominant glycoprotein of *Bacillus anthracis* spores. Structure **20**: 479–86.

Kishimoto, K., Kouzai, Y., Kaku, H., Shibuya, N., Minami, E., and Nishizawa, Y. (2010). Perception of the chitin oligosaccharides contributes to disease resistance to blast fungus *Magnaporthe oryzae* in rice. Plant J **64**: 343–354.

Ko, K., Norelli, J.L., Reynoird, J.-P., Boresjza-Wysocka, E., Brown, S.K., and Aldwinckle, H.S. (2000). Effect of untranslated leader sequence of AMV RNA 4 and signal peptide of pathogenesis-related protein 1b on attacin gene expression, and resistance to fire blight in transgenic apple. Biotechnol. Lett. **22**: 373–381.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell **16**: 3496–507.

Lacombe, S., Rougon-Cardoso, A., Sherwood, E., Peeters, N., Dahlbeck, D., van Esse, H.P., Smoker, M., Rallapalli, G., Thomma, B.P., Staskawicz, B., Jones, J.D., and Zipfel, C. (2010). Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. Nat Biotechnol **28**: 365–369.

Lee, S.-C., Park, K., Han, J., Lee, J., Kim, H.J., Hong, S., Heu, W., Kim, Y.J., Ha, J.-S., Lee, S.-G., Cheong, H.-K., Jeon, Y.H., Kim, D., and Kim, H.-S. (2012). Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. Proc. Natl. Acad. Sci. U. S. A. **109**: 3299–304.

Lius, S., Manshardt, R.M., Fitch, M.M.M., Slightom, J.L., Sanford, J.C., and Gonsalves, D. (1997). Pathogen-derived resistance provides papaya with effective protection against papaya ringspot virus. Mol. Breed. **3**: 161–168.

McDonald, B.A. and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. **40**: 349–79.

Mueller, K., Bittel, P., Chinchilla, D., Jehle, A.K., Albert, M., Boller, T., and Felix, G. (2012). Chimeric FLS2 receptors reveal the basis for differential flagellin perception in Arabidopsis and tomato. Plant Cell **24**: 2213–24.

Nirmala, J., Brueggeman, R., Maier, C., Clay, C., Rostoks, N., Kannangara, C.G., von Wettstein, D., Steffenson, B.J., and Kleinhofs, A. (2006). Subcellular localization and functions of the barley stem rust resistance receptor-like serine/threonine-specific protein kinase Rpg1. Proc. Natl. Acad. Sci. U. S. A. **103**: 7518–23.

Overington, J.P., Al-Lazikani, B., and Hopkins, A.L. (2006). How many drug targets are there? Nat. Rev. Drug Discov. **5**: 993–6.

Pancer, Z., Amemiya, C.T., Ehrhardt, G.R., Ceitlin, J., Gartland, G.L., and Cooper, M.D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature **430**: 174–180.

Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., Deal, K., Luo, M., Kong, X., Bariana, H., Mago, R., McIntosh, R., Dodds, P., Dvorak, J., and Lagudah, E. (2013). The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. Science **341**: 786–8.

Pfund, C., Tans-Kersten, J., Dunning, F.M., Alonso, J.M., Ecker, J.R., Allen, C., and Bent, A.F. (2004). Flagellin is not a major defense elicitor in *Ralstonia solanacearum* cells or extracts applied to Arabidopsis thaliana. Mol. Plant. Microbe. Interact. **17**: 696–706.

Saintenac, C., Zhang, W., Salcedo, A., Rouse, M.N., Trick, H.N., Akhunov, E., and Dubcovsky, J. (2013). Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. Science **341**: 783–6.

Sanahuja, G., Banakar, R., Twyman, R.M., Capell, T., and Christou, P. (2011). *Bacillus thuringiensis*: a century of research, development and commercial applications. Plant Biotechnol. J. **9**: 283–300.

Santelia, D. and Zeeman, S.C. (2011). Progress in Arabidopsis starch research and potential biotechnological applications. Curr. Opin. Biotechnol. **22**: 271–80.

Schornack, S., Moscou, M.J., Ward, E.R., and Horvath, D.M. (2013). Engineering plant disease resistance based on TAL effectors. Annu. Rev. Phytopathol. **51**: 383–406.

Segretin, M.E., Pais, M., Franceschetti, M., Chaparro-Garcia, A., Bos, J.I., Banfield, M.J., and Kamoun, S. (2014). Single amino acid mutations in the potato immune receptor R3a expand response to *Phytophthora* effectors. Mol. Plant. Microbe. Interact.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., and Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. **31**: 686–8.

Singh, R.P., Hodson, D.P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., Herrera-Foessel, S., Singh, P.K., Singh, S., and Govindan, V. (2011). The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. Annu. Rev. Phytopathol. **49**: 465–81.

Song, J., Bradeen, J.M., Naess, S.K., Raasch, J.A., Wielgus, S.M., Haberlach, G.T., Liu, J., Kuang, H., Austin-Phillips, S., Buell, C.R., Helgeson, J.P., and Jiang, J. (2003). Gene RB cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proc. Natl. Acad. Sci. U. S. A. **100**: 9128–33.

Sonnewald, U. and Kossmann, J. (2013). Starches--from current models to genetic engineering. Plant Biotechnol. J. **11**: 223–32.

Stirnweis, D., Milani, S.D., Jordan, T., Keller, B., and Brunner, S. (2014). Substitutions of two amino acids in the nucleotide-binding site domain of a resistance protein enhance the hypersensitive response and enlarge the PM3F resistance spectrum in wheat. Mol. Plant. Microbe. Interact. **27**: 265–76.

Sun, W., Cao, Y., Jansen Labby, K., Bittel, P., Boller, T., and Bent, A.F. (2012). Probing the Arabidopsis flagellin receptor: FLS2-FLS2 association and the contributions of specific domains to signaling function. Plant Cell **24**: 1096–113.

Sun, W., Dunning, F.M., Pfund, C., Weingarten, R., and Bent, A.F. (2006). Within-species flagellin polymorphism in *Xanthomonas campestris pv campestris* and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses. Plant Cell **18**: 764–79.

Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.-M., and Chai, J. (2013). Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science **342**: 624–8.

Tricoll, D.M., Carney, K.J., Russell, P.F., McMaster, J.R., Groff, D.W., Hadden, K.C., Himmel, P.T., Hubbard, J.P., Boeshore, M.L., and Quemada, H.D. (1995). Field evaluation of transgenic squash containing single or multiple virus coat protein gene constructs for resistance to cucumber mosaic virus, watermelon mosaic virus 2, and zucchini yellow mosaic virus. Bio/Technology **13**: 1458–1465.

Tripathi, L., Mwaka, H., Tripathi, J.N., and Tushemereirwe, W.K. (2010). Expression of sweet pepper Hrap gene in banana enhances resistance to *Xanthomonas campestris* pv. *musacearum*. Mol. Plant Pathol. **11**: 721–31.

Velikovsky, C.A., Deng, L., Tasumi, S., Iyer, L.M., Kerzic, M.C., Aravind, L., Pancer, Z., and Mariuzza, R.A. (2009). Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen. Nat. Struct. Mol. Biol. **16**: 725–30.

Wezner-Ptasińska, M., Krowarsch, D., and Otlewski, J. (2011). Design and characteristics of a stable protein scaffold for specific binding based on variable lymphocyte receptor sequences. Biochim. Biophys. Acta **1814**: 1140–5.

Zeng, Y., Zhao, S., Yang, S., and Ding, S.-Y. (2014). Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. Curr. Opin. Biotechnol. **27**: 38–45.

Zhao, B., Lin, X., Poland, J., Trick, H., Leach, J., and Hulbert, S. (2005). A maize resistance gene functions against bacterial streak disease in rice. Proc. Natl. Acad. Sci. U. S. A. **102**: 15383–8.

Figures

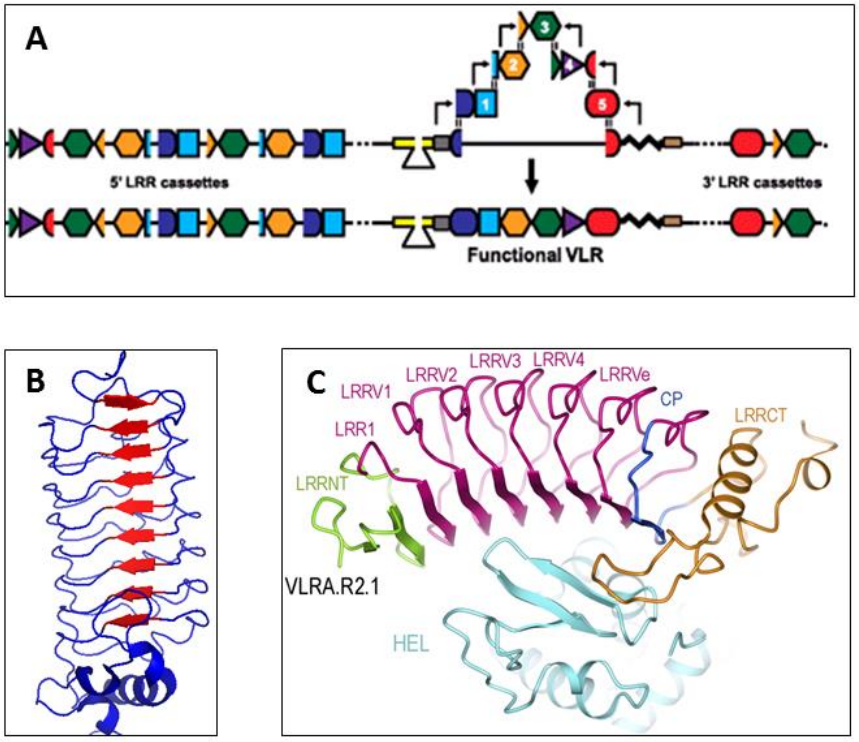

Figure 1: Variable lymphocyte receptors (VLRs). (A) adapted from Herrin and Cooper 2010: The VLR gene scaffold is flanked by hundreds of LRR gene cassettes. In nascent lymphocytes, the functional VLR genes are generated by combinatorial assembly of LRR cassettes into the VLR gene scaffold. **(B)** VLRA.R2.1 (PDB ID: 3M18). The concave face consists of nine beta sheets. **(C)** adapted from Deng et al 2010: VLRA.R2.1 bound to its ligand HEL. The LRRs are labeled. The LRRCT consists of the LRR cap with the alpha helix and the loop. The loop is an important part of the HEL binding interface.

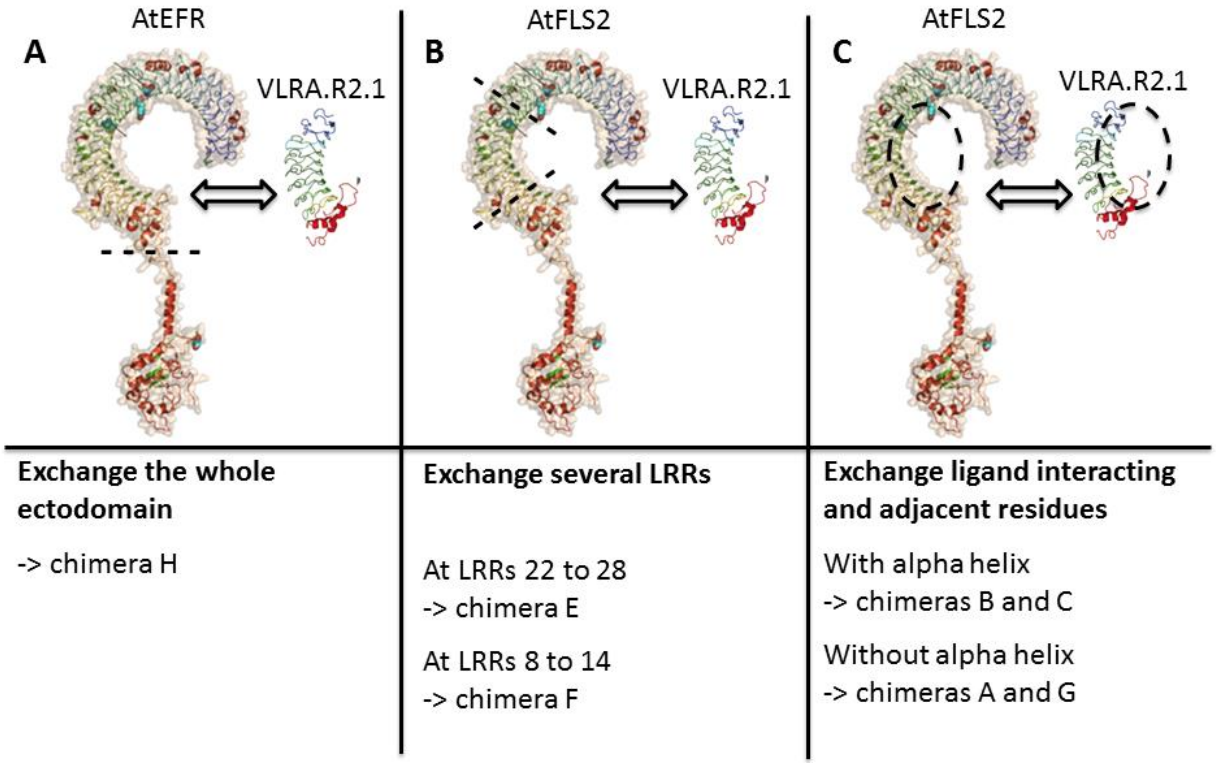

Figure 2: Design of the chimeric receptors. (A) LRR domain swap of EFR with VLRA.R2.1. Only the LRR domain was exchanged, the AtEFR signal peptide was retained. **(B)** Exchange of FLS2 LRRs with VLRA.R2.1 LRRs. **(C)** Exchange of single FLS2 amino acids with single VLRA.R2.1 amino acids.

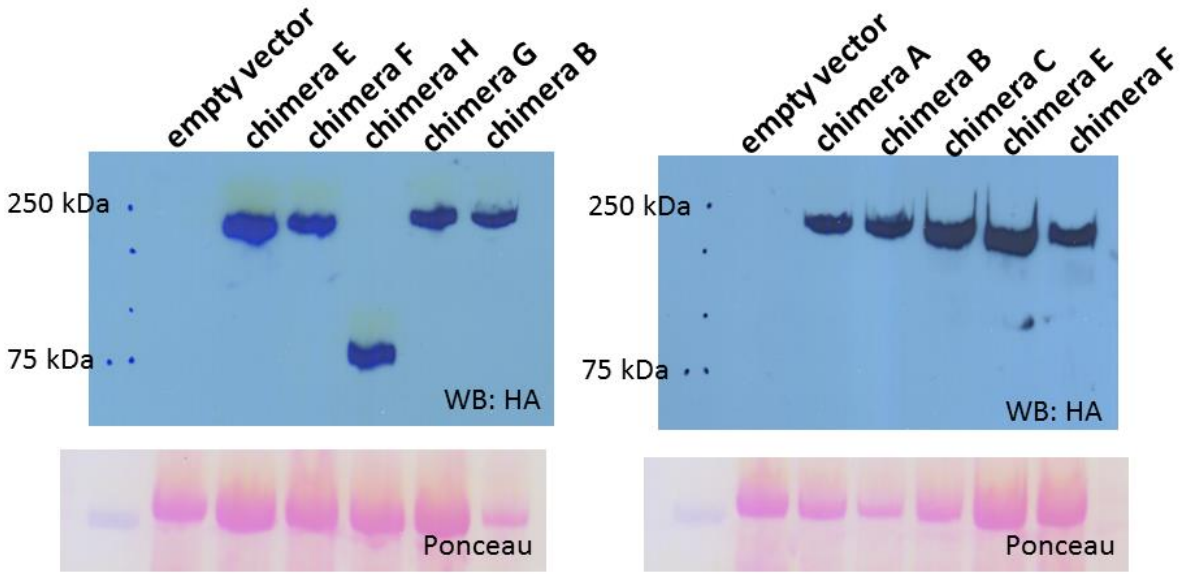

Figure 3: Chimeric proteins accumulate. The chimeric genes were transiently expressed in *Nicotiana benthamiana* leaves. All chimeras carried an HA tag at the C-terminus of their kinase domains. Western blot was performed using anti-HA antibody. Chimera H is much smaller because of only 9 LRRs in the ectodomain. All the other chimeras have 28 LRRs in their ectodomain. Ponceau staining indicates total protein on membranes (rubisco bands).

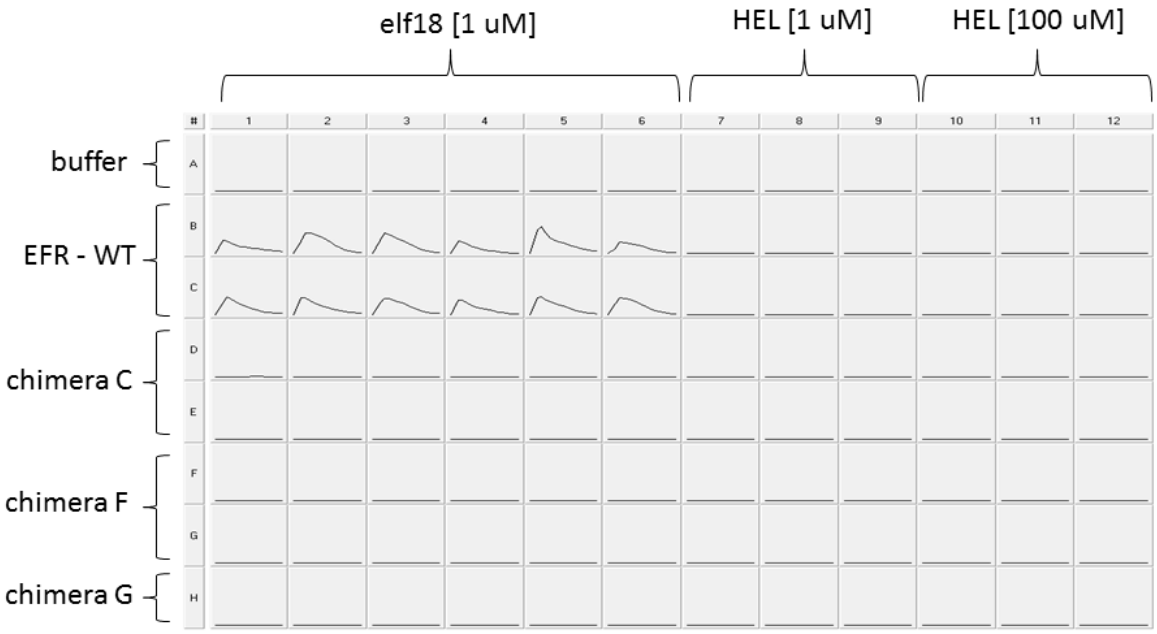

Figure 4: Chimeras C, F and G are not functional. ROS assay was performed with transiently transformed *Nicotiana benthamiana* leaf discs. X-axis: time scale 0 to 30 min after addition of ligand. Y-axis: luminescence. Wild type EFR (EFR-WT) was used as positive control of *Nicotiana benthamiana* transformation and ROS assay. Buffer: not transformed *Nicotiana benthamiana* leaf discs (the leaf was only inoculated with buffer, no *Agrobacterium*). Elf18 was added to chimeras C, F and G as negative controls of ROS burst. HEL was added at two different concentrations. The *Nicotiana benthamiana* leaf discs transformed with chimeras C, F and G did not trigger an ROS burst when 1 µM or 100 µM HEL were added.

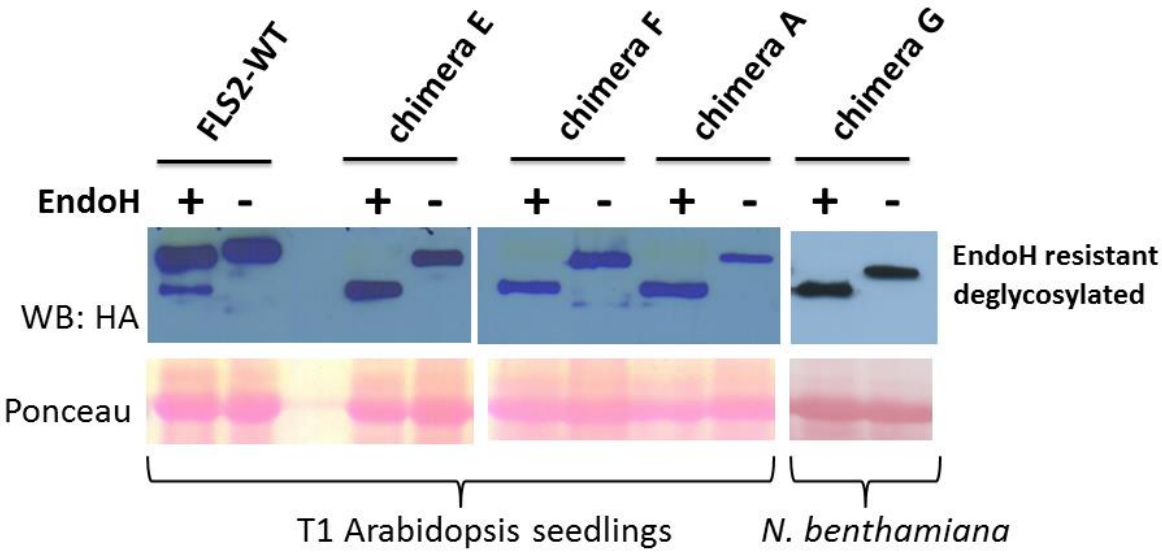

Figure 5: Glycosylation defects of chimeric proteins. EndoH assay was performed using stably transformed T1 Arabidopsis seedlings to test chimeras A, E and F and wild-type FLS2 (FLS2-WT) as positive control. EndoH assay was performed using transiently transformed *N.benthamiana* leaves to test chimera G. In FLS2-WT samples (seedling protein extract) treated with EndoH, both the EndoH resistant (= glycosylated) and the EndoH sensitive (= deglycosylated) protein pools are visible. In the chimera samples (protein extracts) treated with EndoH, the EndoH resistant (= glycosylated) protein pools are missing. Ponceau staining indicates total protein on membranes.

Chapter 5: FLS2-BAK1 extracellular domain interaction sites required for defense signaling activation

This chapter was submitted to PLOS ONE on August 23rd 2014, for consideration as a Research Article.

Authors: Teresa Koller and Andrew F Bent

Contributions: Teresa Koller performed all experiments. Teresa Koller and Andrew Bent analyzed the data and wrote the manuscript.

Abstract

Signaling initiation by receptor-like kinases (RLKs) at the plasma membrane of plant cells often requires regulatory leucine-rich repeat (LRR) RLK proteins such as SERK or BIR proteins. The present work examined how BAK1 (SERK3) builds signaling complexes with the microbe-associated molecular pattern (MAMP) receptor FLS2. We first, using *in vivo* methods that validate separate findings by others, demonstrated that flg22 (flagellin epitope) ligand-initiated FLS2-BAK1 extracellular domain interactions can proceed independent of intracellular domain interactions. We then explored a candidate SERK protein interaction site in the extracellular domains (ectodomains; ECDs) of the significantly different receptors FLS2, EFR (MAMP receptors), PEPR1 (damage-associated molecular pattern (DAMP) receptor), and BRI1 (hormone receptor). Repeat conservation mapping revealed a cluster of conserved solvent-exposed residues near the C-terminus of models of the folded LRR domains. However, site-directed mutagenesis of this conserved site in FLS2 did not impair FLS2-BAK1 ECD interactions, and mutations in the analogous site of EFR caused receptor maturation defects. Hence this conserved LRR C-terminal region apparently has functions other than mediating interactions with BAK1. *In vivo* tests of the subsequently published FLS2-flg22-BAK1 ECD co-crystal structure were then performed to functionally evaluate some of the unexpected configurations predicted by that crystal structure. In support of the crystal structure data, FLS2-BAK1 ECD interactions were no longer detected in *in vivo* co-immunoprecipitation experiments after site-directed mutagenesis of the FLS2 BAK1-interaction residues S554, Q530, Q627 or N674. In contrast, *in vivo* FLS2-mediated signaling persisted and was only minimally reduced, suggesting residual FLS2-BAK1 interaction and the limited sensitivity of co-immunoprecipitation data relative to *in*

vivo assays for signaling outputs. However, Arabidopsis plants expressing FLS2 with the Q530A+Q627A double mutation were impaired both in detectable interaction with BAK1 and in FLS2-mediated responses, lending overall support to current models of FLS2 structure and function.

Introduction

Plants use pattern-recognition receptors (PRRs) as a first layer of defense against pathogens [1,2]. In order to engineer plants with improved pathogen recognition abilities, it is important to understand the molecular details underlying the interaction of PRRs not only with their ligands but also with their co-receptors, immediate downstream targets and other partner proteins that facilitate appropriate signaling. Several PRRs have been identified in different plant species [reviewed in 2,3]. PRRs are localized at the plasma membrane where they monitor the apoplastic space for microbe-associated molecular patterns (MAMPs), damage-associated molecular patterns (DAMPs) and apoplastic effectors. Most known PRRs are receptor-like kinases (RLKs) or receptor-like proteins (RLPs). Both receptor types consist of an extracellular domain for ligand perception and a transmembrane domain, but only the RLKs have an intracellular kinase domain. Two of the best characterized PRRs, FLS2 and EFR [3,4], carry large extracellular domains (ECDs, ectodomains) that predominantly consist of a leucine-rich repeat (LRR) domain [5,6]. The genomes of Arabidopsis and other plants each encode hundreds of LRR receptor-like kinases (LRR-RLKs) with 4 to 28 repeat units of the LRR [7].

Receptors typically exhibit high specificity for ligands with which they interact, but cells also contain co-receptors and regulatory proteins that function together with receptors and do not necessarily exhibit specificity for only a single type of ligand [8,9]. These co-receptors and regulatory proteins can be important facilitators or suppressors of signaling activation. They also allow signaling crosstalk at the plasma membrane, helping to coordinate appropriate downstream signaling in the presence of diverse endogenous and exogenous extracellular ligands. Important examples of regulatory/co-receptor RLKs include the SERK family members [8,10], BIR family members [11,12] and SOBIR1 [13].

SERK proteins have been identified in many different plant species. In Arabidopsis the family consists of five members (SERK1, SERK2, SERK3/BAK1, SERK4 and SERK5). They all have five LRRs in their ectodomain, share high overall sequence similarity and have redundant functions to various degrees. SERK proteins (mainly SERK3, also known as BAK1) have been shown to be involved in plant immunity in Arabidopsis, tomato and rice, through interactions with the receptors FLS2, EFR, PEPR1, PEPR2, Xa21, Ve1 and Eix1 [14–19]. The BAK1 co-receptor also contributes to somatic embryogenesis [20,21] and to plant development through interaction with the brassinolide hormone receptor BRI1 [22,23]. Despite impressive progress, much remains unknown about how the SERK proteins participate in all these different cell signaling tasks, and about the spatial expression of SERK proteins [24]. Studies of the SERK proteins are impeded by the redundant functions among family members and by pleiotropic effects when multiple SERK proteins are knocked out. As an example, *bak1*$^{-}$ Arabidopsis plants only have partially disrupted FLS2 signaling outputs [14,15,25,26]. A possible means of circumventing this problem of SERK functional redundancy, adopted in the present study, is to

identify the specific SERK interaction site of a partner receptor and then mutate that site. If all SERKs interact with a specific receptor at similar amino acids, this approach should impair the interaction of the receptor with all SERK family members.

Recent X-ray crystallography studies provided detailed insight into the interaction of the ectodomain of BAK1 with the ectodomains of FLS2 and BRI1 [27,28], and the interaction of the ectodomains of SERK1 and BRI1 [29]. In all three cases the respective ligand promotes interaction between the ectodomains of the main receptors (FLS2 and BRI1) and the SERK co-receptors (BAK1 and SERK1). The ligand binds to the LRR domain of the main receptor, but the LRR domain of the SERK co-receptor also has multiple direct contacts with the ligand. It is surprising to see these fine-tuned co-receptor/ligand interactions, considering how many different known and potential unknown receptors and ligands BAK1 and SERK1 are able to interact with. Similar residues of the BAK1 and SERK1 ectodomains are involved in their interactions with FLS2 and BRI1. However, the residues on FLS2 and BRI1 ectodomains predicted to be used for the interactions with their SERK co-receptors are very different, not only in sequence but also in their location within the receptor LRR domain [27–29]. In BRI1 the residues interacting with co-receptors are located at the island domain, the last LRR, and the juxtamembrane domain, all close to the transmembrane domain. However, in FLS2 the BAK1-interacting residues in the crystal structure are located 108 - 300 amino acids from the predicted transmembrane domain, at repeats #18 to 26 of the LRR domain. This predicts a relatively recumbent orientation for the FLS2 ectodomain, bent down toward the plasma membrane (see Figure 9A).

FLS2 mediates perception of bacterial flagellin protein, an abundant MAMP, and FLS2 recognizes in particular a ~20 amino acid region that is relatively conserved across flagellins from diverse Gram-negative bacteria [1,30]. Many aspects of FLS2 structure and function have been characterized [reviewed in 28]. There is a third surprising feature of the FLS2-flg22-BAK1 ECD co-crystal structure [27]. Most research regarding FLS2 utilizes as ligand, in place of flagellin protein, a 22 amino acid "flg22" peptide whose sequence matches the recognized domain of *Pseudomonas aeruginosa* flagellin, or utilizes other small peptides based on similar sequences from various bacteria [30,32,33]. The FLS2-flg22-BAK1 ECD co-crystal structure predicts a tight pocket for the flg22 peptide, which may not be compatible with (allow sufficient space or sufficient ligand flexibility for) analogous binding of flg22 domains embedded within full-length flagellin proteins (discussed below).

In this study we first explored the possibility that a relatively universal SERK interaction site has evolved in the LRR domains of different SERK-interacting LRR-RLKs. We also showed that flg22-dependent FLS2 interaction with BAK1 occurs via the FLS2 extracellular domains – a result subsequently shown by alternative methods by Sun *et al.* (2103). We then performed site-directed mutagenesis and functional testing of predicted LRR-RLK receptor/SERK co-receptor interaction residues, and obtained *in vivo* evidence that supports models suggested by the recently published receptor/co-receptor co-crystal structures of truncated FLS2 and BAK1. The overall goal of this study was to furnish a more clear understanding of the requirements for formation of a signaling-competent plant basal immune system MAMP receptor – an understanding that may be essential to allow future engineering of PRRs with broadened or otherwise improved performance.

Methods

Arabidopsis and *Nicotiana benthamiana* transformation

The floral dip method was used to stably transform Arabidopsis *fls2*$^-$ and *efr*$^-$ plants. T1 seedlings were selected on 0.5x MS plates containing 25 mg/L kanamycin and 25 mg/L hygromycin. Leaves of 4-week-old *Nicotiana benthamiana* plants were infiltrated with *Agrobacterium tumefaciens* GV3101 containing the binary plasmids [34]. Proteins were harvested two days after *Agrobacterium tumefaciens* infiltration.

Co-immunoprecipitation

Transiently transformed leaf tissue from *Nicotiana benthamiana* was infiltrated with 1 µM flg22 or 1 µM elf18, or with water for mock infiltration. After 2 minutes the leaf tissues were blotted dry and frozen in liquid N_2. Then 200 mg of tissues were ground in 200 µl protein extraction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.5% Triton X-100, 1x plant protease inhibitor cocktail (Sigma-Aldrich)). After centrifugation 300 µl supernatant was incubated with 3 µl 9E10 anti-myc antibody (Sigma-Aldrich or Covance) and rotated at 4°C for 1h. 50 µl Protein A (Thermo Scientific) was added and the tubes were rotated at 4°C for an additional 2h. After 3x washing with protein extraction buffer and 1x washing with ddH_2O the beads were resuspended in 60 µl loading buffer and boiled at 95°C for 5 min. After centrifugation the supernatant was separated on two 8% SDS-PAGE gels. For protein detection the antibodies anti-HA-HRP, anti-myc rabbit and goat-anti-rabbit-HRP (Sigma-Aldrich) were used.

Conservation mapping

Mapping of conserved regions of predicted LRR surfaces was performed using the Repeat Conservation Mapping (RCM) program at www.plantpath.wisc.edu/RCM [35], with heat map coloration range set to the minimal and maximal conservation scores of the data within each figure. The LRR domain sequences were obtained from The Arabidopsis Information Resource (TAIR) website at www.arabidopsis.org and from the National Center for Biotechnology Information (NCBI) website at www.ncbi.nlm.nih.gov. The FLS2 non-Brassicaceae sequences used to generate Figure 2B were: *Populus trichocarpa* (XP_002305701.1); *Vitis vinifer*a (XP_002272319.2)**;** *Glycine max* (XP_003532650.1); *Lotus japonicus* (AER60531.1); *Ricinus communis* (XP_002519723.1); *Sorghum bicolor* (XP_002448543.1); *Oryza sativa Japonica* (CAE02151.2); *Oryza sativa Indica* (CAH68341.1); *Hordeum vulgare* (BAJ89141.1); *Brachypodium distachyon* (XP_003581675.1).

Site-directed mutagenesis

Point mutations were generated according to the QuikChange mutagenesis kit (Agilent Technologies) on pENTR plasmids (Invitrogen) containing FLS2, FLS2-NoKinase or EFR with 35S or native promoters [36]. Gateway LR Clonase II (Invitrogen) was used to transfer the construct into the binary plasmids pGWB13 or pGWB14 [37].

EndoH assay

Leaf tissues (60 mg) from Arabidopsis T1 plants or from transiently transformed *Nicotiana benthamiana* plants were ground in 2x SDS buffer and boiled for 5 min at 95°C. After centrifugation for 10 min at 14000 rpm at 4°C supernatants were digested with Endoglycosidase H (New England BioLabs) as per manufacturer's suggestion and separated on 8% SDS-PAGE gel. Proteins were detected using anti-HA-HRP antibody (Sigma-Aldrich).

Seedling growth inhibition

T1 Arabidopsis seedlings were grown for 6 days on 0.5x MS plates with 25 mg/L kanamycin and 25 mg/L hygromycin and 200 mg/L cefotaxime. 24 seedlings per genotype, representing 24 independent transformation events, were transferred to 24-well-plates containing 1 ml 0.5x MS liquid media per well. 12 seedlings per genotype were grown for 14 days in wells containing 1 µM flg22 and 12 seedlings per genotype were grown for 14 days in wells containing only 0.5 x MS. Seedlings were then blotted dry and weighed. The weight of each flg22-treated seedling was divided by the average weight of the mock treated seedlings of the same genotype from the same experiment, prior to determination of experiment means and standard errors.

Oxidative burst

Seven leaf discs were taken from six-week-old T1 Arabidopsis plants and incubated overnight in 1% DMSO solution. Peptide solution was added to the leaf discs and luminescence

was measured by a plate reader for 0 – 30 min after addition of flg22 peptide. For measurement each leaf disc was in 100 µl peptide solution containing 0.5 µl 2 mg/ml horseradish peroxidase, 0.5 µl 2 mg/ml luminol in DMSO and 1 µM flg22.

Results and Discussion

Extracellular domain of FLS2 can mediate interaction with BAK1 in the presence of flg22

Full-length FLS2 and BAK1 do not detectably interact until exposure to flg22 or similar flagellin ligands, at which time interaction is immediately observed [14,15,38]. Flg22-elicited immune signaling then requires phosphorylation events among the respective kinase domains [26,38,39]. We hypothesized that the FLS2-BAK1 interaction is mediated not only intracellularly by the respective kinase domains, but also by interaction of the ectodomains. To test this we used a truncated FLS2 carrying the N-terminal ~70% of the protein including the LRR and transmembrane domains but not the predicted intracellular domains (*FLS2-NoKinase-HA*; [36]). *FLS2-NoKinase-HA* was expressed in *Nicotiana benthamiana* together with a plasmid encoding a full-length, epitope tagged *BAK1-Myc*. The transiently transformed leaves were treated with flg22 and co-immunoprecipitation experiments were performed. BAK1 and FLS2-NoKinase interact in the presence of flg22, indicating that the kinase domain of FLS2 is not needed for interaction with BAK1 *in planta* (Figure 1).

Sun *et al.* 2013 also showed ECD mediation of FLS2-BAK1 interaction [27]. Their work utilized *in vitro* mixing experiments with purified recombinant proteins, or mutated BAK1 expressed in Arabidopsis protoplasts. Our results with mutated FLS2, tested in transgenic whole

plants with *FLS2* expressed under control of *FLS2* promoter sequences, are complimentary and in agreement with the results of Sun *et al.* (2103), and reveal that intracellular/kinase domain interactions of these proteins are not required for flg22-stimulated FLS2-BAK1 interaction. It is also interesting to note the previously published finding that FLS2-FLS2 interaction occurs *in planta*, with either full-length FLS2 or FLS2-NoKinase constructs [36]. At least some FLS2 exists *in planta* in FLS2-FLS2 complexes, prior to and after flagellin or flg22 exposure. FLS2-BAK1 interaction after exposure to flg22 did not appreciably deplete the overall presence of co-immunoprecipitable FLS2-FLS2 complexes [36]. Hence findings that FLS2 and BAK1 interact via LRR domains suggest either that FLS2-FLS2 interactions utilize a different side or face of the FLS2 LRR than the region that interacts with BAK1, or that different sub-pools of FLS2 are at any given moment interacting with FLS2 or BAK1. The results of Albert *et al.* (2013) and Cao *et al.* (2013) are also relevant to these updated models of PRR receptor - co-receptor structure/function [39,40]. Those studies demonstrated that *in planta* responses to flg22 are retained when hybrid FLS2 and BAK1 proteins are expressed in which the kinase domains of FLS2 and BAK1 have been reciprocally swapped [40], and that flg22-mediated FLS2-BIK1 disassociation and FLS2-BAK1 association still occur when FLS2 kinase domain mutations are present that block defense signaling. Schulze *et al.* 2010 and Schwessinger *et al.* 2011 showed that kinase-dead BAK1 still interact with FLS2, but impair FLS2 signaling [26,38]. The evidence increasingly indicates that interactions of the FLS2 and BAK1 extracellular domains are a first step in flg22 perception that can proceed relatively independent of intracellular domain structural or functional interactions.

Identification of a conserved region in the C-terminal LRRs of BAK1-interacting receptors

The SERK family members have been shown to interact with several different transmembrane LRR-RLKs involved in plant immunity and development. It is not known if the SERK interaction sites of these receptors evolved independently or originate from a common and potentially conserved SERK interaction site. We hypothesized the latter and also hypothesized that, to facilitate spatial proximity of potentially interacting extracellular domains, the relatively small ectodomains of SERK proteins would interact near the C-terminal end of the large LRR ectodomains of those partner receptors. Using Repeat Conservation Mapping [35] we searched the last seven repeats of the LRRs of the known Arabidopsis BAK1-interacting proteins FLS2 (28 total repeats in the LRR domain), EFR (21 LRRs), BRI1 (25 LRRs) and PEPR1 (26 LRRs), looking for the patch of solvent-exposed amino acids in this region that is most conserved across the four proteins. A conserved region of interest was identified (Figure 2A). Separately, we compared the solvent exposed amino acids of the whole LRR domains of eleven non-Brassicaceae FLS2s (Figure 2B). Both conservation maps revealed a conserved region at a similar location in the C-terminal LRRs. We hypothesized that this may be a somewhat universal site for interaction with SERK proteins.

No disruption of FLS2-BAK1 interaction by mutations in the FLS2 LRR domain C-terminal region conserved among EFR, PEPR1, BRI1 and multiple FLS2s

Site-directed mutagenesis was carried out to alter residues in the identified conserved LRR C-terminal region of FLS2 and EFR (Figure 2; Figure 7A-E). D557 and S559 mutations in FLS2 were included as control mutations located in LRR sites analogous to N704/S706 and

D728/S730, but outside of the conserved LRR C-terminus. The amino acids were replaced with similar yet bulkier residues in order to impair interactions. The resulting full-length receptors were expressed in *N. benthamiana* and co-immunoprecipitation experiments were then carried out, using BAK1-Myc for pull-down in the presence and absence of the corresponding ligands flg22 and elf18 in the case of FLS2 or EFR, respectively. The mutations in FLS2 did not abolish the interaction with BAK1 in the presence of flg22 (Figure 3A). As is common for agroinfiltration experiments, variable levels of expression were observed for any single transgene-encoded FLS2 protein across replicates within or between experiments, but none of the mutant proteins was reproducibly present at levels different from transgene-encoded wild-type FLS2. To ensure that interaction of the kinase domains of FLS2 and BAK1 was not masking non-interaction of mutated FLS2 and BAK1 ectodomains, the same mutations were also placed into FLS2-NoKinase constructs. In these FLS2-NoKinase variants the mutations again did not prevent interaction with BAK1 (Figure 3B).

Mutations in the conserved C-terminal LRR region of EFR cause EFR glycosylation/maturation defects

Mutations analogous to those of the preceding section were also engineered into *EFR*. These LRR domain C-terminal region mutations (Figure 2; Figure 7C, D) did cause disruption of interaction with BAK1 in the presence of elf18 (Figure 4A). This *in vivo* result could be attributable to direct impacts of the mutations on EFR-BAK1 interaction, or to defects in maturation and delivery of newly synthesized EFR out of the endoplasmic reticulum (ER) and golgi. Endoglycosidase H (EndoH) analyses were therefore conducted. EndoH cleaves

incomplete glycosylation modifications present on proteins that have not successfully passed through the ER and related endomembrane systems [41,42]. On the other hand, mature glycosylated proteins that are delivered to their functional location typically carry EndoH-resistant glycosylation [41,42]. Treatment of the EFR protein extracts with EndoH revealed defects in the mutated EFR proteins, both in *N. benthamiana* and in stable transgenic Arabidopsis *efr*- plants expressing transgene *EFR* constructs driven by native *EFR* promoter sequences (Figure 4B, C). The mutations we generated in FLS2 full-length and FLS2-NoKinase did not result in glycosylation defects (Figure 8A, B). Häweker *et al.* 2010 [42] and Sun at al. 2012 [36] showed that single amino acid changes in glycosylation sites in the EFR ectodomain result in protein degradation and several studies reported the importance of intact glycosylation enzymes for successful processing and function of EFR [43–47]. FLS2 is less sensitive to mutations in glycosylation sites [36,42]. The N590Q+S592T mutations that we placed in EFR are indeed in a Nx(S/T) predicted glycosylation site [48]. However, the EFR mutations D566E+S568T and D566F are not, yet they still disrupted correct EFR processing.

Taken together, the above results suggest that functional roles of the LRR C-terminal conserved domain of BAK1-interacting proteins (Figure 2) do not serve as a universal SERK protein interaction site. However, in EFR the integrity of this site is important for correct protein processing.

FLS2 mutations in proposed FLS2-BAK1 ECD interaction residues disrupt FLS2-BAK1 interaction in the presence of flg22

While the above work was in progress the crystal structure of FLS2-flg22-BAK1 ECD became available [27]. That important work identified in detail the interaction sites of the FLS2 and BAK1 ectodomains. Because the data are for *in vitro* crystallized protein complexes of isolated LRR domains, they may or may not capture the most functionally prominent *in vivo* configurations. Sun *et al.* [27] therefore functionally tested BAK1 mutations, and also tested FLS2 mutations in repeats #9, 11, 14 and 15 of the LRR that are predicted to mediate interaction with flg22. The sites on FLS2 predicted to mediate interaction with BAK1 did not receive mutational testing. In order to test *in vivo* the significance of these FLS2 BAK1-interaction residues, which are likely to also mediate interaction of FLS2 with other SERK proteins, we performed site-directed mutagenesis on *FLS2-NoKinase* and full-length *FLS2*.

For FLS2 amino acids predicted in the crystal structure to form FLS2-BAK1 interaction sites [27], we changed single residues to alanine (small and relatively inactive) or to tryptophan (bulky). In addition to the single mutations we made two *FLS2-NoKinase* constructs with double mutations and one full-length *FLS2* construct with a double mutation. We had previously shown that the FLS2-NoKinase used in this work performed similarly to FLS2-full-length in flg22-dependent BAK1 co-immunoprecipitation experiments (Figure 3A, B). In *in vivo* tests of the newly predicted FLS2-BAK1 interaction sites, mutation of FLS2 residues Q530, S554, Q627 or N674 to tryptophan disrupted the flg22-stimulated interaction of FLS2-NoKinase with BAK1 (Figure 5A, B). The interaction was disrupted as well when FLS2 Q530 and N674 were changed to alanine (Figure 5A, B). However, the FLS2 S554 and Q627 single mutations to alanine had

much less impact on flg22-dependent interaction with BAK1 (Figure 5A, B), suggesting a stronger role for Q530 and N674 than S554 or Q627 in mediating FLS2-BAK1 interaction. The double alanine mutation Q530A+Q627A and the double tryptophan mutation S554W+Q627W disrupted BAK1 interaction as well (Figure 5C). The presence of abundant EndoH-insensitive bands suggested that FLS2 maturation had proceeded successfully for each of the representative FLS2 mutants S554A, Q627W and Q530A+Q627A (Figure 8C).

Arabidopsis *fls2*$^-$ plants carrying FLS2-Q530A+Q627A have impaired FLS2-mediated signaling outputs

To investigate if mutations in predicted FLS2 BAK1-interaction residues not only disrupt FLS2-BAK1 interactions in co-immunoprecipitation experiments but also have an impact on FLS2 signaling, we made the analogous single mutations and one of the double mutations in full-length *FLS2s.* We then tested FLS2 signaling in stably transformed *fls2*$^-$ Arabidopsis plants containing the mutated and HA-tagged full-length *FLS2*s under control of native *FLS2* promoter sequences. The two most widely used assays for FLS2 signaling were utilized: ROS burst assays and seedling growth inhibition assays [1]. Surprisingly, *in vivo* FLS2-mediated signaling persisted and was only minimally reduced in plants expressing most single-mutant forms of FLS2 (Figure 6A, B), including mutants that exhibited no detectable flg22-induced co-immunoprecipitation with BAK1 (Figure 5A, B). As a general trend across the multiple independent transgenic lines tested for each *FLS2* construct, mutations to alanine allowed stronger FLS2 signaling than mutations to tryptophan (Figure 6A, B). The results suggest that reduced-affinity or more transient interactions of FLS2 and BAK1 occur with many of the FLS2 mutants described in

Figures 5 and 6, and that those interactions are sufficient for flg22-stimulated FLS2 signaling even if the stability of FLS2-BAK1 interactions is reduced below levels detectable in standard co-immunoprecipitation experiments. Although some FLS2 signaling capacity was still conferred by FLS2 constructs mutated at single predicted FLS2 BAK1-interaction sites, FLS2-mediated signaling was significantly impaired in plants carrying FLS2 with the double mutation Q530A+Q627A (Figure 6C, D).

LRRs are a protein structure evolved to display widely varying surface amino acid combinations on a relatively invariant scaffold [5,6]. A previous study of over 1200 FLS2 LRR mutations of predicted LRR solvent-exposed residues at and adjacent to flg22 binding sites, carrying changes to all possible amino acids (i.e., not just to alanine), found that the vast majority of LRR surface mutations do not disrupt FLS2 function [49]. Hence the structural alterations caused by the FLS2 mutations of the present study are likely to be highly local. Their disruption of FLS2-BAK1 interactions detected via co-immunoprecipitation supports the relevance of the FLS2-flg22-BAK1 configuration in the published co-crystal structure. Mutation of FLS2 residues D557 and S559, which reside close to but outside of the BAK1-interaction residues in the solved crystal structure ([27], Figure 7), did not disrupt flg22-stimulated FLS2-BAK1 co-immunoprecipitation (Figure 3). The functional disruption of signaling caused by the presumably additive effect of two alanine substitutions in FLS2 Q530A+Q627A provides further *in vivo* functional evidence indicating the requirement for this site both for FLS2-BAK1 interaction and for flg22 induction of FLS2-dependent immune signaling. Our results also indicate that, if SERK proteins other than BAK1 make residual contributions to FLS2 activation

(as is suggested in the literature [14,15,25,26]), the FLS2 Q530A+Q627A mutations are sufficient to disrupt functional signaling mediated by those interactions as well.

Closing Observations

In this study we explored the idea of a universal SERK protein interaction site in the C-terminal repeats of the LRR ectodomains of receptors known to interact with SERK proteins. However, mutagenesis of a possible BAK1 interaction site in the ectodomains of FLS2 and EFR did not confirm this hypothesis. The subsequently available FLS2-flg22-BAK1 and BRI1-brassinolide-SERK1 extracellular domain crystal structures [27] [29], and the mutational studies in the present work, instead suggest a fine-tuned interaction unique for each receptor/ligand/co-receptor complex. SERK1 and BAK1 use similar residues to interact with the BRI1 and FLS2 ectodomains, respectively. However, the SERK-interacting residues in the ectodomains of BRI1 and FLS2 are very different in terms of both the amino acid identities and their location along the large LRR macromolecule, and thus may have evolved separately.

The LRR surface region exhibiting conservation between FLS2, EFR, PEPR1 and BRI1 (Figure 2A) spans four repeats of the LRR, but overlaps with the larger region highlighted in Figure 2B that is conserved across diverse FLS2 proteins and spans the final seven repeats of the LRR. Within the larger conserved region, the residues that are further to the left as shown in Figure 2B (or Figure 7E) encompass the BAK1 interaction site, but the residues on the right do not. The present study detected no impact of mutations in FLS2 in the Figure 2A conserved region, which in FLS2 is the same as the bottom right of the larger conserved region of Figure 2B. A previous

study from our group [35] reported little or no functional impact of mutations in the upper-right area of the conserved region (the darkest red/most conserved area of Figure 2B). In that study, libraries of changes to all possible amino acids were made at the four FLS2 residues D605, S607, F633 and S634, directly above the N704/S706 and D728/S730 residues targeted in this study but in repeats #22 and 23 [35]. Hence it is intriguing that this right side of the region highlighted in Figure 2B, which lies along the concave β-strand surface of repeats #21-27, is highly tolerant of mutations despite being relatively conserved across FLS2 proteins from diverse plant species. It remains of interest to discover the function of this portion of the FLS2 LRR.

As a separate but related matter, it is intriguing that the set of BAK1-interacting residues of FLS2 lie not only within regions highly conserved across FLS2 proteins from diverse plant species (e.g., Q627 and N674, Figure 7E), as might be expected, but also outside of conserved regions (e.g., Q530 and S554, Figure 7E). Figure 2B and Figure 7E show regions of LRR surface residue conservation in a comparison among FLS2s from non-Brassicaceae plant species (see Methods). But even in maps of conservation among FLS2s only from Brassicaceae species (see for example [35]), the region around Arabidopsis FLS2 residues Q627 and N674 is strongly conserved while Q530, S554 and adjacent BAK1-interacting residues [27] are in an LRR surface region that is less conserved. This raises the hypothesis that there is a functionally relevant diversification of SERKs and/or this upper portion of the SERK-interaction site of FLS2, even across Brassicaceae species.

The relevance of the FLS2-flg22-BAK1 co-crystal structure to actual configurations of the protein complex within plant cells would gain stronger support if more features of the crystal

structure were reconciled with other findings regarding plant FLS2s and flagellin detection. We noted in the Introduction the concern that the co-crystal, made with flg22 peptide, may not allow enough space for docking of a full-length flagellin protein at the appropriate location. Figure 9 shows hypothetical alignments of the FLS2-flg22-BAK1 ECD structure (PDB ID: 4MN8) with the structure of one Salmonella flagellin protein (PDB ID: 3A5X), placing the flg22 region of 3A5X near the apparent flg22 binding sites of FLS2 and BAK1 while attempting to minimize co-occupancy of the same space by two different molecules. The FLS2 LRR, which is notably lacking in 'loop-out' or non-LRR-consensus regions, is likely to be relatively inflexible. Flagellin monomers in solution (not polymerized with other flagellins to form flagella-like structures) are likely to be more flexible than shown, particularly in the region of the flg22 residues that form a less ordered linker between two alpha-helical regions [50,51] (see Figure 9F). Nevertheless, space-filling models (e.g., Figure 9D) demonstrate the difficulty of docking a large flagellin onto the requisite FLS2 LRR sites while also allowing space for BAK1 and not allowing co-occupancy of identical space. Importantly, even in hypothesized configurations (not shown) that might allow space for a more flexible full-length flagellin to interact with FLS2 and BAK1, the flg22 residues within a flagellin protein are apparently constrained in ways that would restrict simultaneous interaction with the majority of the FLS2 LRR surface residues that interact with the elongated flg22 in the published FLS2-flg22-BAK1 co-crystal structure (e.g., Figure 9E). FLS2, flagellins and BAK1 may associate *in vivo* in configurations that depart significantly from the co-crystal structure. However, numerous aspects of the published FLS2-flg22-BAK1 co-crystal structure are substantiated by experimental evidence ([27]; references therein; present study). Hence we consider it equally likely that the published FLS2-flg22-BAK1 co-crystal is essentially

correct in representing *in vivo* configurations, and predict that flagellin proteins within plants must be fragmented rather than intact in order to form the FLS2-flagellin-BAK1 complexes that elicit plant innate immune system activation.

In the future, it also will be interesting to compare more receptor/ligand/co-receptor signaling complexes in order to learn more about the signaling crosstalk mediated by regulatory RLKs at the plasma membrane. As one example, a ligand-mediated EFR-BAK1 ectodomain complex is likely to initiate EFR signaling. Interestingly, when EFR from Arabidopsis was transferred to *Nicotiana benthamiana* or tomato (which lack an endogenous EFR) it triggered an elf18-activated immune response, indicating functional interaction of AtEFR with SERK proteins from *Nicotiana benthamiana* and tomato [52]. Thus one or more SERK proteins apparently carry sufficient structure-function plasticity to interact with different receptors even from diverse plant species, while complying with the fine-tuned sequence constraints of the resulting receptor/ligand/co-receptor complexes. For future engineering of PRR receptors with novel ligand specificities it will be important to ensure presence of an intact SERK protein interaction site in the ectodomain of the PRR, close to or overlapping with the ligand binding site, and ensure that the co-receptors also can form PRR/ligand/co-receptor complexes with the novel ligands for which new recognition specificity is sought.

Acknowledgments

We thank Stephen Mosher and Adam Bayless for crucial reading of the manuscript.

References

1. Boller T, Felix G (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu Rev Plant Biol 60: 379–406. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19400727.

2. Wu Y, Zhou J-M (2013) Receptor-like kinases in plant innate immunity. J Integr Plant Biol 55: 1271–1286. Available: http://www.ncbi.nlm.nih.gov/pubmed/24308571.

3. Gómez-Gómez L, Boller T (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell 5: 1003–1011. doi:S1097-2765(00)80265-8 [pii].

4. Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JDG, et al. (2006) Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts *Agrobacterium*-Mediated Transformation. Cell 125: 749–760. doi:10.1016/j.cell.2006.03.037.

5. Kobe B, Kajava A V (2001) The leucine-rich repeat as a protein recognition motif. Curr Opin Struct Biol 11: 725–732. Available: http://www.ncbi.nlm.nih.gov/pubmed/11751054.

6. Bella J, Hindle KL, McEwan PA, Lovell SC (2008) The leucine-rich repeat structure. Cell Mol Life Sci 65: 2307–2333. Available: http://www.ncbi.nlm.nih.gov/pubmed/18408889.

7. Lehti-Shiu MD, Zou C, Hanada K, Shiu S-H (2009) Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. Plant Physiol 150: 12–26. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2675737&tool=pmcentrez&rendertype=abstract.

8. Chinchilla D, Shan L, He P, de Vries S, Kemmerling B (2009) One for all: the receptor-associated kinase BAK1. Trends Plant Sci 14: 535–541. Available: http://www.ncbi.nlm.nih.gov/pubmed/19748302.

9. Liebrand TWH, van den Burg H a, Joosten MH a J (2014) Two for all: receptor-associated kinases SOBIR1 and BAK1. Trends Plant Sci 19: 123–132. Available: http://www.ncbi.nlm.nih.gov/pubmed/24238702.

10. Kim BH, Kim SY, Nam KH (2013) Assessing the diverse functions of BAK1 and its homologs in arabidopsis, beyond BR signaling and PTI responses. Mol Cells 35: 7–16. Available: http://www.ncbi.nlm.nih.gov/pubmed/23269431.

11. Gao M, Wang X, Wang D, Xu F, Ding X, et al. (2009) Regulation of cell death and innate immunity by two receptor-like kinases in Arabidopsis. Cell Host Microbe 6: 34–44. Available: http://www.ncbi.nlm.nih.gov/pubmed/19616764.

12. Halter T, Imkampe J, Mazzotta S, Wierzba M, Postel S, et al. (2013) The Leucine-Rich Repeat Receptor Kinase BIR2 Is a Negative Regulator of BAK1 in Plant Immunity. Curr Biol. Available: http://www.ncbi.nlm.nih.gov/pubmed/24388849.

13. Liebrand TWH, van den Berg GCM, Zhang Z, Smit P, Cordewener JHG, et al. (2013) Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. Proc Natl Acad Sci U S A 110: 10010–10015. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3683720&tool=pmcentrez&rendertype=abstract.

14. Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nürnberger T, et al. (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature 448: 497–500. Available: http://www.ncbi.nlm.nih.gov/pubmed/17625569.

15. Heese A, Hann DR, Gimenez-Ibanez S, Jones AME, He K, et al. (2007) The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. Proc Natl Acad Sci U S A 104: 12217–12222. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1924592&tool=pmcentrez&rendertype=abstract.

16. Bar M, Sharfman M, Ron M, Avni A (2010) BAK1 is required for the attenuation of ethylene-inducing xylanase (Eix)-induced defense responses by the decoy receptor LeEix1. Plant J 63: 791–800. Available: http://www.ncbi.nlm.nih.gov/pubmed/20561260.

17. Fradin EF, Zhang Z, Juarez Ayala JC, Castroverde CDM, Nazar RN, et al. (2009) Genetic dissection of Verticillium wilt resistance mediated by tomato Ve1. Plant Physiol 150: 320–332. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2675724&tool=pmcentrez&rendertype=abstract.

18. Chen X, Zuo S, Schwessinger B, Chern M, Canlas PE, et al. (2014) An XA21-Associated Kinase (OsSERK2) regulates immunity mediated by the XA21 and XA3 immune receptors. Mol Plant. Available: http://www.ncbi.nlm.nih.gov/pubmed/24482436.

19. Postel S, Küfner I, Beuter C, Mazzotta S, Schwedt A, et al. (2010) The multifunctional leucine-rich repeat receptor kinase BAK1 is implicated in Arabidopsis development and immunity. Eur J Cell Biol 89: 169–174. Available: http://www.sciencedirect.com/science/article/pii/S0171933509003306.

20. Schmidt ED, Guzzo F, Toonen MA, de Vries SC (1997) A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. Development 124: 2049–2062. Available: http://www.ncbi.nlm.nih.gov/pubmed/9169851.

21. Hecht V, Vielle-Calzada JP, Hartog M V, Schmidt ED, Boutilier K, et al. (2001) The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol 127: 803–816. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=129253&tool=pmcentrez&rendertype=abstract.

22. Li J, Wen J, Lease KA, Doke JT, Tax FE, et al. (2002) BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell 110: 213–222. Available: http://www.ncbi.nlm.nih.gov/pubmed/12150929.

23. Nam KH, Li J (2002) BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell 110: 203–212. Available: http://www.ncbi.nlm.nih.gov/pubmed/12150928.

24. Belkhadir Y, Yang L, Hetzel J, Dangl JL, Chory J (2014) The growth-defense pivot: crisis management in plants mediated by LRR-RK surface receptors. Trends Biochem Sci. Available: http://www.ncbi.nlm.nih.gov/pubmed/25089011.

25. Roux M, Schwessinger B, Albrecht C, Chinchilla D, Jones A, et al. (2011) The Arabidopsis leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. Plant Cell 23: 2440–2455. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3160018&tool=pmcentrez&rendertype=abstract.

26. Schwessinger B, Roux M, Kadota Y, Ntoukakis V, Sklenar J, et al. (2011) Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. PLoS Genet 7: e1002046. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3085482&tool=pmcentrez&rendertype=abstract.

27. Sun Y, Li L, Macho AP, Han Z, Hu Z, et al. (2013) Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science 342: 624–628. Available: http://www.ncbi.nlm.nih.gov/pubmed/24114786.

28. Sun Y, Han Z, Tang J, Hu Z, Chai C, et al. (2013) Structure reveals that BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. Cell Res 23: 1326–1329. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3817550&tool=pmcentrez&rendertype=abstract.

29. Santiago J, Henzler C, Hothorn M (2013) Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. Science 341: 889–892. Available: http://www.ncbi.nlm.nih.gov/pubmed/23929946.

30. Felix G, Duran JD, Volko S, Boller T (1999) Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J 18: 265–276. Available: http://www.ncbi.nlm.nih.gov/pubmed/10377992.

31. Robatzek S, Wirthmueller L (2012) Mapping FLS2 function to structure: LRRs, kinase and its working bits. Protoplasma 21. Available: http://www.ncbi.nlm.nih.gov/pubmed/23053766.

32. Sun W, Dunning FM, Pfund C, Weingarten R, Bent AF (2006) Within-species flagellin polymorphism in *Xanthomonas campestris pv campestris* and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses. Plant Cell 18: 764–779. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1383648&tool=pmcentrez&rendertype=abstract.

33. Mueller K, Bittel P, Chinchilla D, Jehle AK, Albert M, et al. (2012) Chimeric FLS2 Receptors Reveal the Basis for Differential Flagellin Perception in Arabidopsis and Tomato. Plant Cell 24: 2213–2224. Available: http://www.ncbi.nlm.nih.gov/pubmed/22634763.

34. Tai TH, Dahlbeck D, Clark ET, Gajiwala P, Pasion R, et al. (1999) Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato. Proc Natl Acad Sci U S A 96: 14153–14158. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24206&tool=pmcentrez&rendertype=abstract.

35. Helft L, Reddy V, Chen X, Koller T, Federici L, et al. (2011) LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains. PLoS One 6: e21614. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3138743&tool=pmcentrez&rendertype=abstract.

36. Sun W, Cao Y, Jansen Labby K, Bittel P, Boller T, et al. (2012) Probing the Arabidopsis flagellin receptor: FLS2-FLS2 association and the contributions of specific domains to signaling function. Plant Cell 24: 1096–1113. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3336135&tool=pmcentrez&rendertype=abstract.

37. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng 104: 34–41. Available: http://www.ncbi.nlm.nih.gov/pubmed/17697981.

38. Schulze B, Mentzel T, Jehle AK, Mueller K, Beeler S, et al. (2010) Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. J Biol Chem 285: 9444–9451. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20103591.

39. Cao Y, Aceti DJ, Sabat G, Song J, Makino S-I, et al. (2013) Mutations in FLS2 Ser-938 dissect signaling activation in FLS2-mediated Arabidopsis immunity. PLoS Pathog 9: e1003313. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3630090&tool=pmcentrez&rendertype=abstract.

40. Albert M, Jehle AK, Fürst U, Chinchilla D, Boller T, et al. (2013) A two-hybrid-receptor assay demonstrates heteromer formation as switch-on for plant immune receptors. Plant Physiol 163: 1504–1509. Available: http://www.ncbi.nlm.nih.gov/pubmed/24130196.

41. Maley F, Trimble RB, Tarentino a L, Plummer TH (1989) Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. Anal Biochem 180: 195–204. Available: http://www.ncbi.nlm.nih.gov/pubmed/2510544.

42. Häweker H, Rips S, Koiwa H, Salomon S, Saijo Y, et al. (2010) Pattern recognition receptors require N-glycosylation to mediate plant immunity. J Biol Chem 285: 4629–4636. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2836068&tool=pmcentrez&rendertype=abstract.

43. Farid A, Malinovsky FG, Veit C, Schoberer J, Zipfel C, et al. (2013) Specialized roles of the conserved subunit OST3/6 of the oligosaccharyltransferase complex in innate immunity and tolerance to abiotic stresses. Plant Physiol 162: 24–38. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3641206&tool=pmcentrez&rendertype=abstract.

44. Li J, Zhao-Hui C, Batoux M, Nekrasov V, Roux M, et al. (2009) Specific ER quality control components required for biogenesis of the plant innate immune receptor EFR. Proc Natl Acad Sci U S A 106: 15973–15978. Available: http://www.ncbi.nlm.nih.gov/pubmed/19717464.

45. Liu Y, Li J (2013) A conserved basic residue cluster is essential for the protein quality control function of the Arabidopsis calreticulin 3. Plant Signal Behav 8: e23864. Available: http://www.ncbi.nlm.nih.gov/pubmed/23425854.

46. Su W, Liu Y, Xia Y, Hong Z, Li J (2012) The Arabidopsis homolog of the mammalian OS-9 protein plays a key role in the endoplasmic reticulum-associated degradation of misfolded receptor-like kinases. Mol Plant 5: 929–940. Available:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3399701&tool=pmcentrez&rendertype=abstract.

47. Von Numers N, Survila M, Aalto M, Batoux M, Heino P, et al. (2010) Requirement of a homolog of glucosidase II beta-subunit for EFR-mediated defense signaling in Arabidopsis thaliana. Mol Plant 3: 740–750. Available: http://www.ncbi.nlm.nih.gov/pubmed/20457640.

48. Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem 54: 631–664. Available: http://www.ncbi.nlm.nih.gov/pubmed/3896128.

49. Dunning FM, Sun W, Jansen KL, Helft L, Bent AF (2007) Identification and mutational analysis of Arabidopsis FLS2 leucine-rich repeat domain residues that contribute to flagellin perception. Plant Cell 19: 3297–3313. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2174712&tool=pmcentrez&rendertype=abstract.

50. Yonekura K, Maki-Yonekura S, Namba K (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature 424: 643–650. Available: http://www.ncbi.nlm.nih.gov/pubmed/12904785.

51. Maki-Yonekura S, Yonekura K, Namba K (2010) Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. Nat Struct Mol Biol 17: 417–422. Available: http://www.ncbi.nlm.nih.gov/pubmed/20228803.

52. Lacombe S, Rougon-Cardoso A, Sherwood E, Peeters N, Dahlbeck D, et al. (2010) Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. Nat Biotechnol 28: 365–369. Available: http://www.ncbi.nlm.nih.gov/pubmed/20231819.

Figures

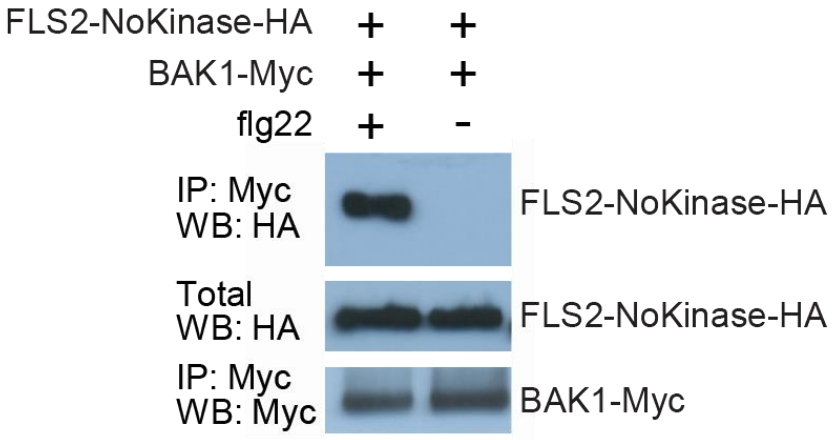

Figure 1: Extracellular domain of FLS2 can mediate interaction with BAK1. Co-immunoprecipitation experiments performed using *35S-FLS2-NoKinase-HA* (construct lacking the FLS2 intracellular domain) and *35S-BAK1-Myc* transiently expressed in *Nicotiana benthamiana* leaves by agroinfiltration. Samples were prepared for SDS-PAGE two days after agroinfiltration, two minutes after flg22 or water (mock) was infiltrated into leaves. IP: antibody used for immunoprecipitation prior to SDS-PAGE; WB: antibody used for immunodetection on protein blot; crude: SDS-PAGE and blotting of total (crude extract) protein samples. The experiment was repeated three times with similar results.

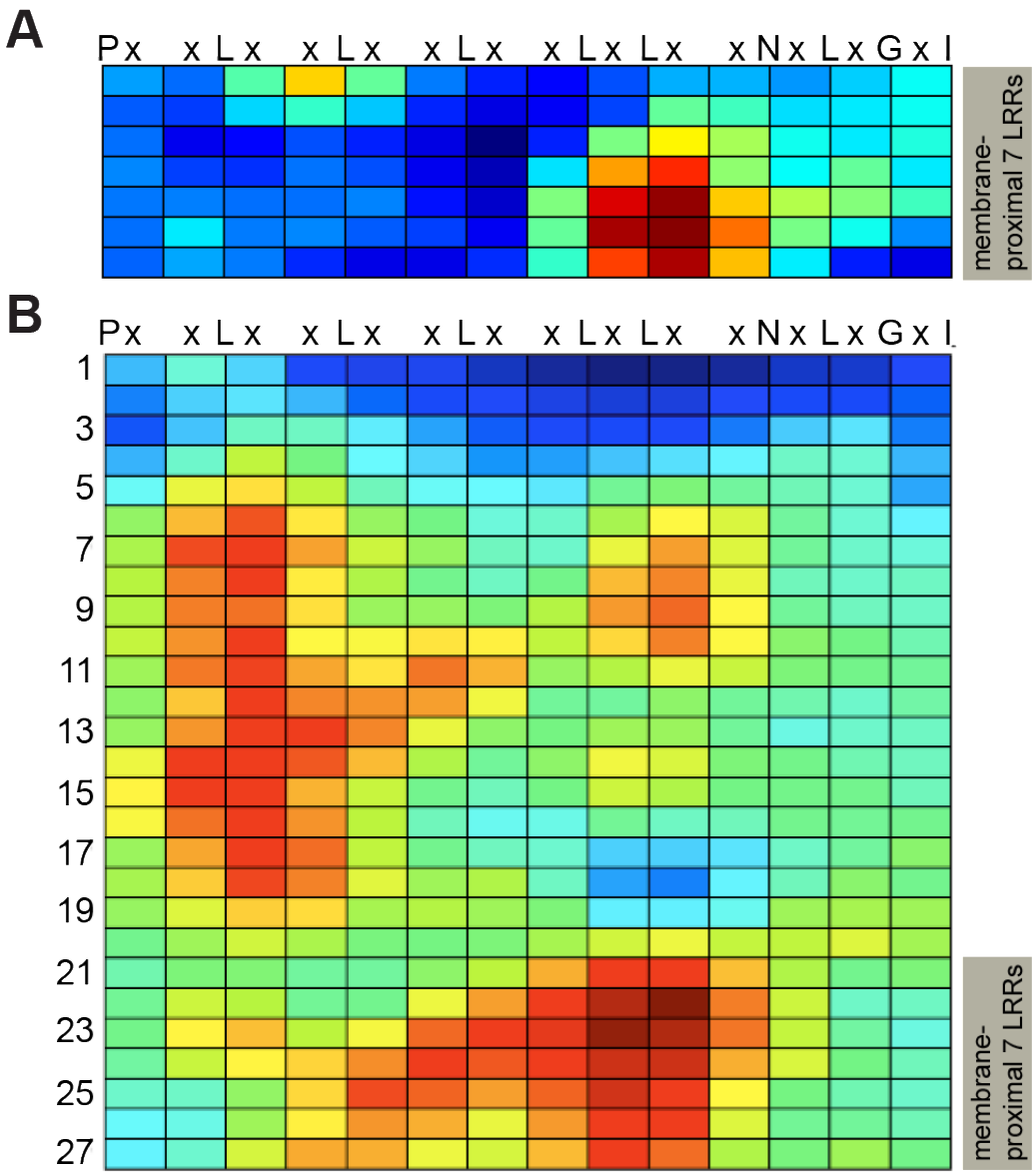

Figure 2: Repeat Conservation Mapping reveals conserved region near C-terminus of LRR domains of FLS2, EFR, BRI1 and PEPR1. Each row represents one leucine-rich repeat (LRR) and each square represents one solvent-exposed "x" amino acid position (as per LRR consensus sequence shown at the top). Conservation score at each amino acid position is center-weighted score for the cluster of 15, 20 or 25 predicted solvent-exposed LRR amino acids surrounding that site; blue: least conserved, red: most conserved. For FLS2, the seven rows of (A) are the same repeats (same residues) as rows 21-27 of (B). **(A)** Conservation map generated by comparing the most C-terminal seven repeats of the LRR sequences of the BAK1 interacting proteins FLS2, EFR, BRI1 and PEPR1. **(B)** Conservation map generated by comparing the entire FLS2 LRR domain sequences from eleven non-Brassicaceae plant species.

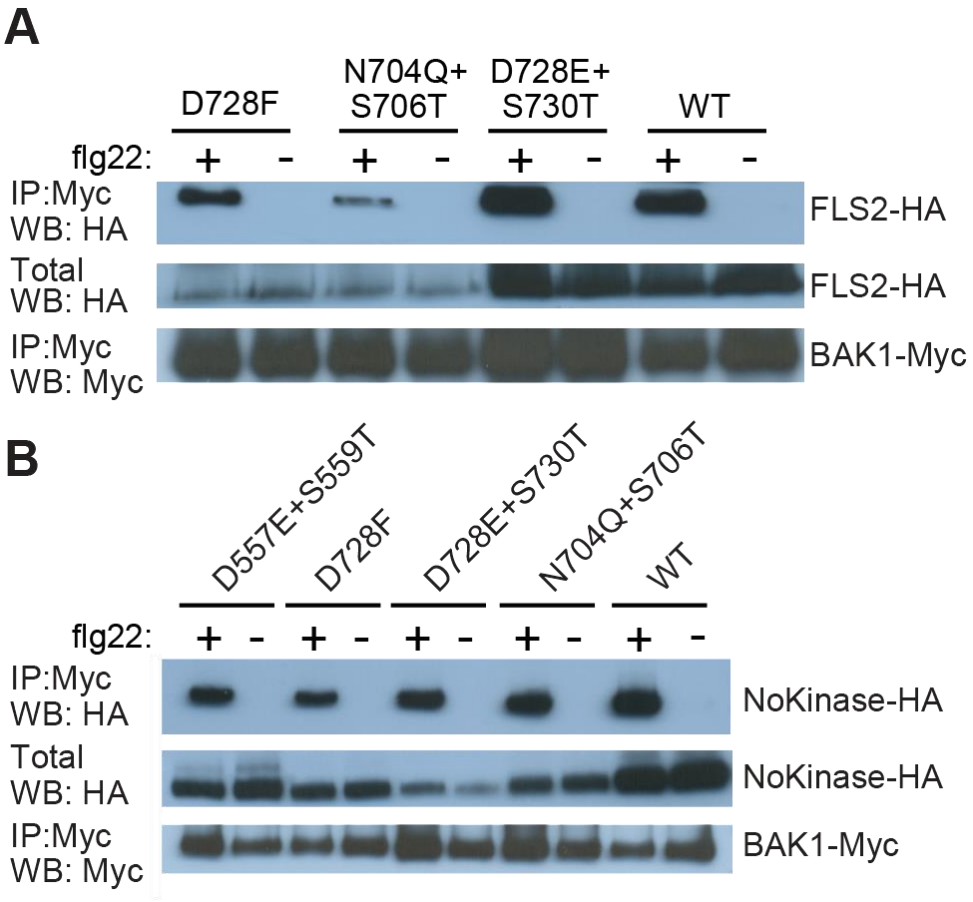

Figure 3: Mutations in the conserved C-terminal region of the FLS2 LRR domain did not have an impact on BAK1-FLS2 or BAK1-FLS2-NoKinase interaction. (A) Co-immunoprecipitation experiments performed using full-length P_{FLS2}-*FLS2-HA*, with mutations as indicated or WT (no mutations), and *35S-BAK1-Myc*. **(B)** Co-immunoprecipitation experiments performed using *35S-FLS2-NoKinase-HA*, with mutations as indicated or WT (no mutations), and *35S-BAK1-Myc*. All samples in (A) and (B) are from *Nicotiana benthamiana.* Labeling as in Figure 1.

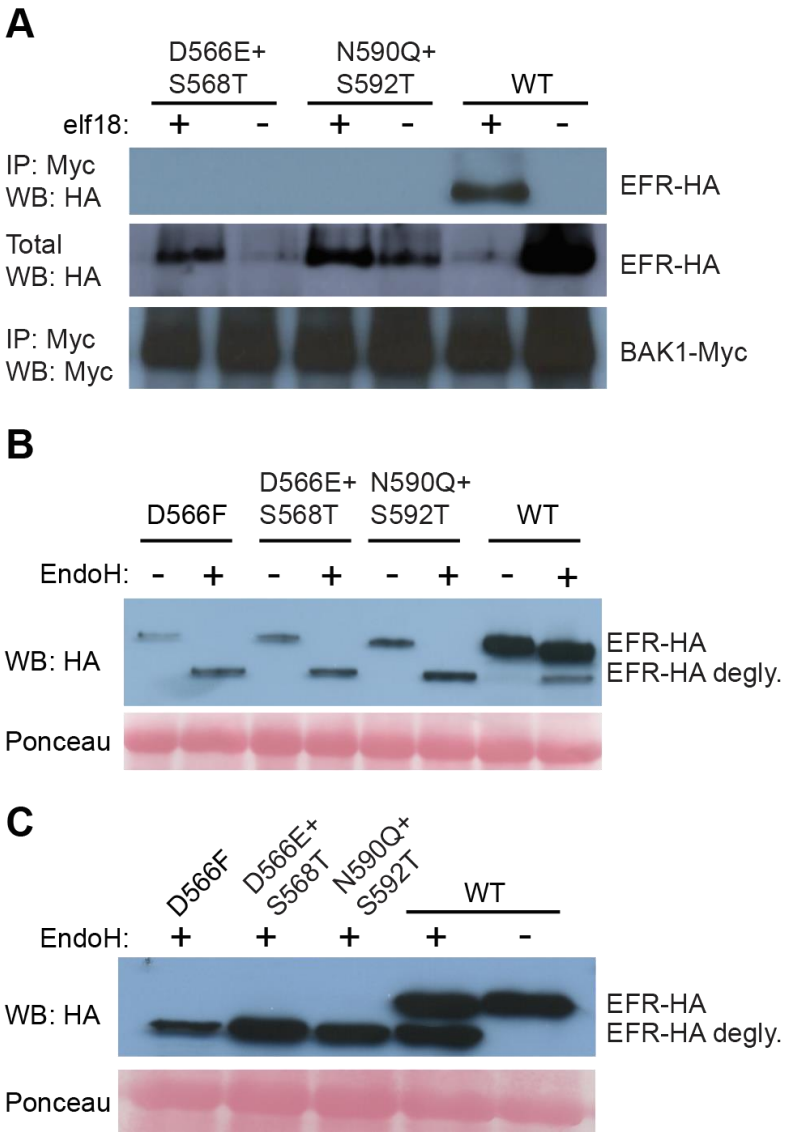

Figure 4: Mutations in the conserved C-terminal region of the EFR LRR domain disrupt EFR glycosylation and interaction with BAK1 in the presence of elf18. (A) Co-immunoprecipitation experiments performed using P_{EFR}*-EFR-HA* with mutations as indicated or WT (no mutations), and *35S-BAK1-Myc*, in *Nicotiana benthamiana.* **(B, C)** Protein extracts from plants expressing P_{EFR}*-EFR-HA* with mutations as indicated, or WT (no mutations), not digested or digested with endoglycosidase H (EndoH). Samples in (B) are from *Nicotiana benthamiana*, samples in (C) are from stably transformed *efr* Arabidopsis leaves. EndoH-resistant (mature) EFR is present in the EndoH-treated EFR wild type (WT) samples but is not detected for EFRs carrying the indicated mutations. Degly.: EFR pool deglycosylated by EndoH. Labeling as in Figure 1. Ponceau: blots treated with Ponceau stain to confirm even loading of total protein.

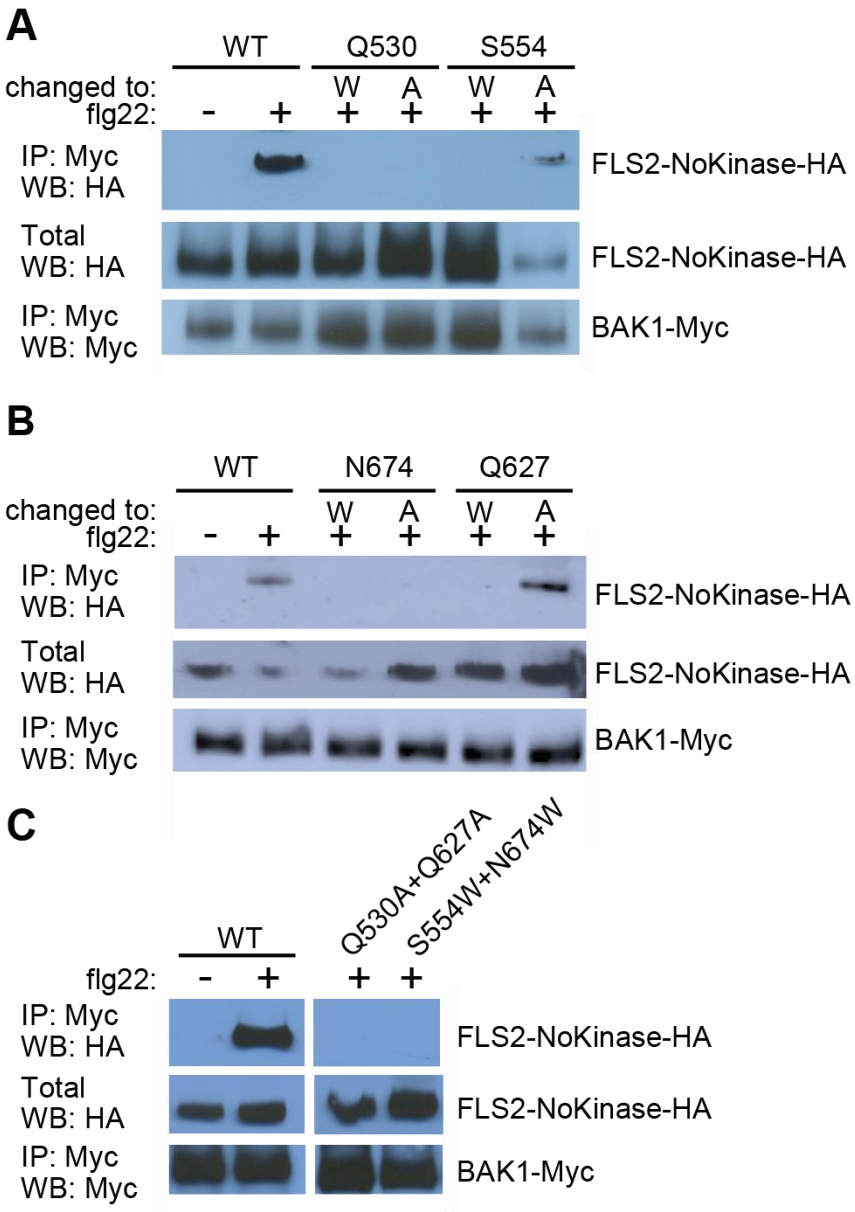

Figure 5: FLS2 residues Q530, S554, Q627 and N674 are important for FLS2-BAK1 ectodomain interaction in the presence of flg22. Co-immunoprecipitation experiments performed in *N. benthamiana* with *35S-FLS2-NoKinase-HA* with mutations as indicated or WT (no mutations), and with *35S-BAK1-Myc*. Flg22-dependent interaction between FLS2-NoKinase and BAK1 not detected for **(A)** FLS2 carrying Q530A, Q530W or S554W mutations, **(B)** FLS2 carrying N674A, N674W or Q627W mutations, or **(C)** FLS2 carrying Q530A+Q627A or S554W+N674W double mutations. Labeling as in Figure 1.

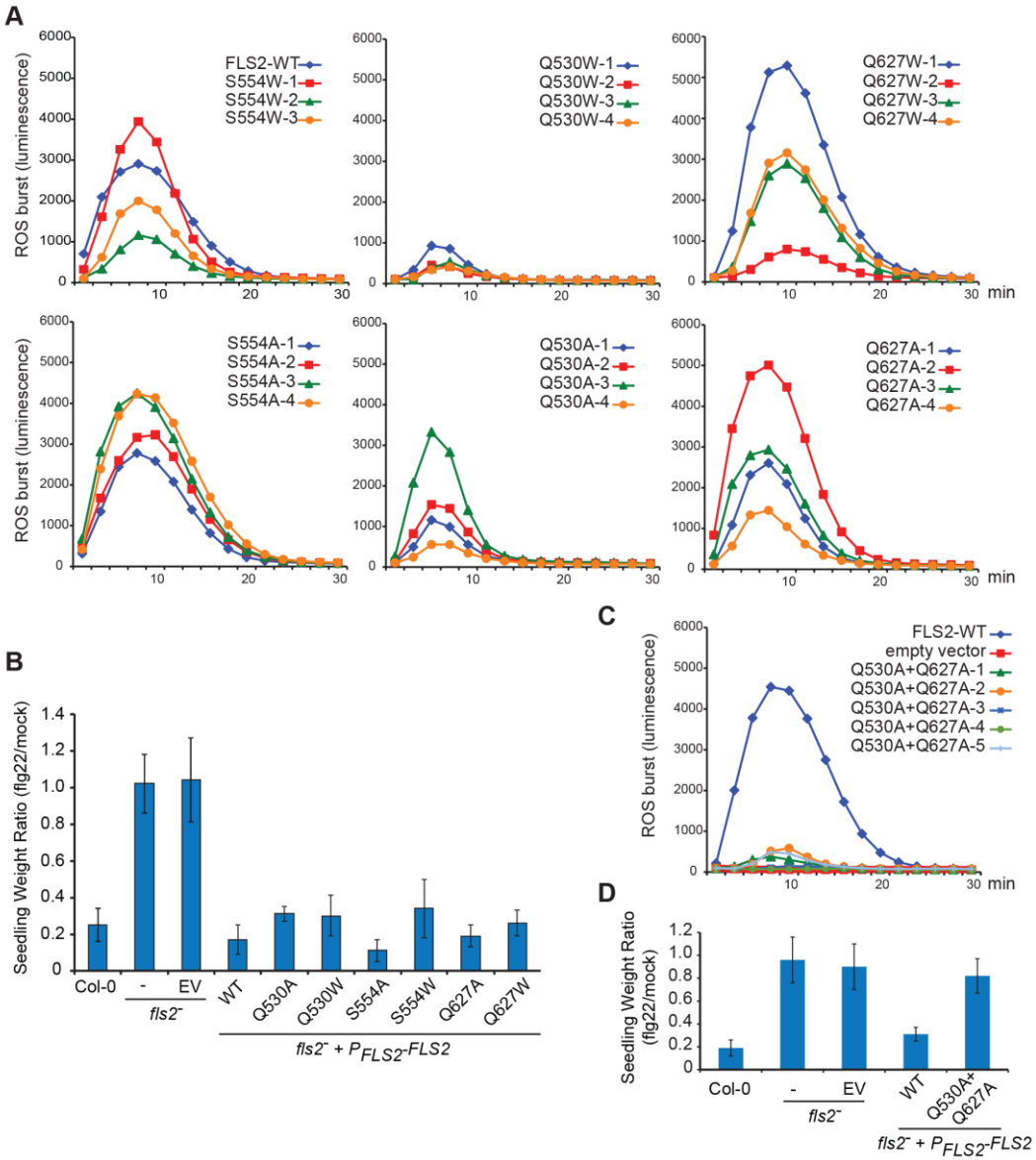

Figure 6: FLS2 signaling output impaired to various degrees in Arabidopsis *fls2*⁻ plants expressing FLS2 mutations that impact FLS2-BAK1 interaction. (A) Reactive oxygen species (ROS) production in response to flg22 in Arabidopsis Col-0 *fls2*⁻ plants stably transformed to express full-length FLS2 proteins carrying single mutations as noted, under control of *FLS2* promoter sequences. For each mutation, ROS production was recorded for 30 min. and the average for seven separately monitored leaf discs is shown for each of four independent transgenic lines (or three lines for S554W). WT: Average ROS response for six independent *fls2*⁻ transformants expressing wild-type FLS2 (42 total leaf discs for WT), from same experiment. **(B)** FLS2-mediated seedling growth inhibition (SGI) in response to flg22, for plant lines as in (A). Mean and std. error of mean shown for six to eight independent transformants for each *FLS2* construct. **(C)** ROS experiment as in (A), except with five independent lines expressing FLS2 Q530A+Q627A double mutations. **(D)** Seedling growth experiment as in (B), except with twelve independent lines expressing Q530A+Q627A double mutations.

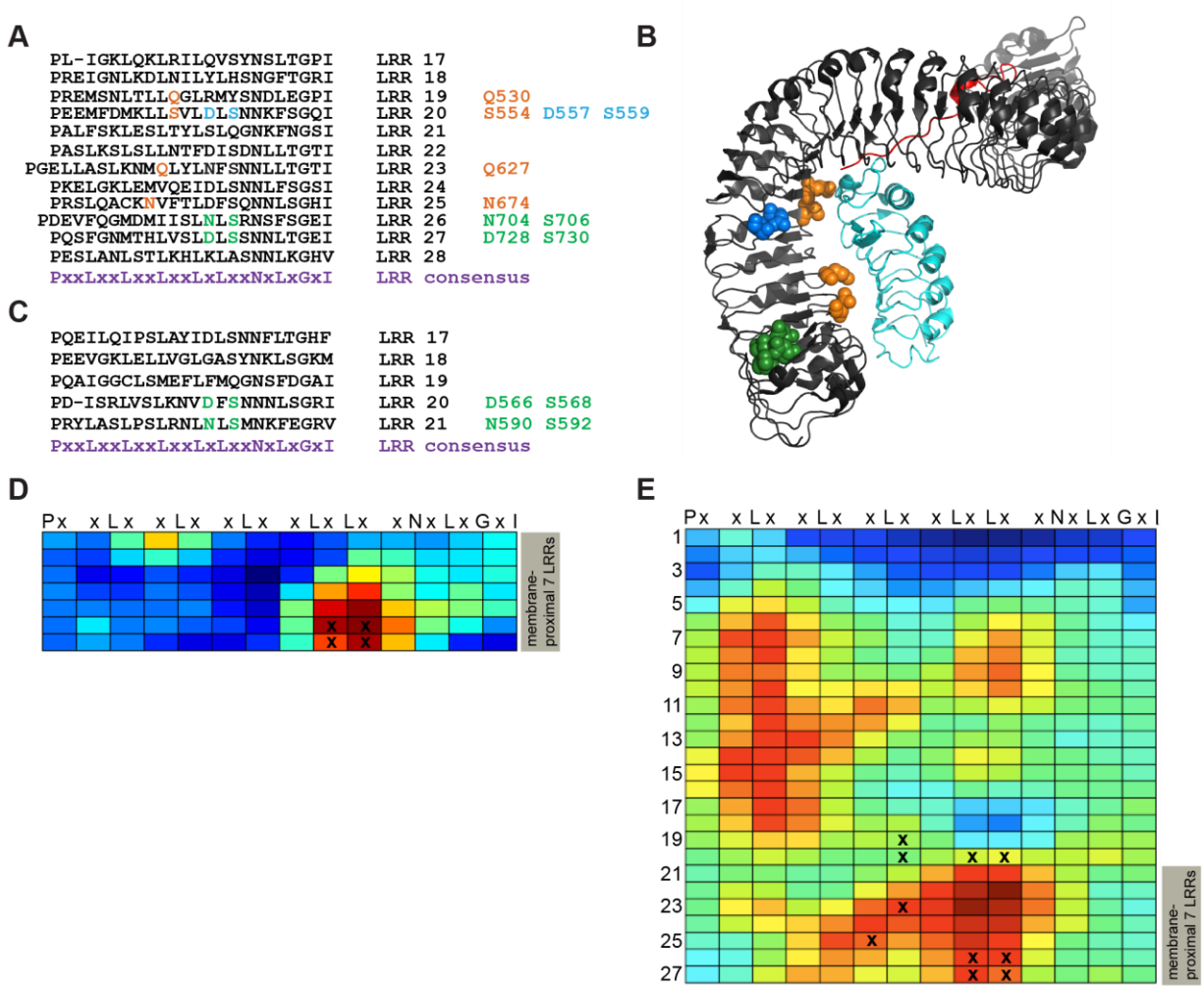

Figure 7: Mutation sites in FLS2 and EFR ECDs. (A) Sites subjected to site-directed mutagenesis in the FLS2 LRR domain. Only repeats 17-28 are shown (FLS2: total 28 repeats). Green: mutation sites in the conserved LRR domain C-terminus (see also (B, D, E)). Blue: "control" mutations; sites similar to N704/D728 and S706/S730 but outside of the conserved region; blue "control" sites also adjacent to but outside of FLS2 BAK1-interaction site. Orange: mutation sites based on FLS2 BAK1-interaction sites in the FLS2-flg22-BAK1 ECD co-crystal structure. **(B)** Mutation sites as described in (A), using same color scheme as in (A). Structure is PDB ID: 4MN8 with FLS2 backbone as black ribbon, BAK1 backbone as light blue ribbon, and flg22 backbone as red ribbon. Space-filling spheres show side-chains only for mutagenized sites. **(C)** Mutation sites in the EFR LRR domain. Only repeats 17-21 are shown (EFR: total 21 repeats). Green: mutation sites in the conserved LRR domain C-terminus (see also (D)). **(D, E)** Regional LRR surface conservation maps from Arabidopsis FLS2, EFR, PEPR1 and BRI1 (D) or eleven non-Brassicaceae FLS2s (E), as shown and described in Figure 2, with x's at the FLS2 and EFR LRR domain amino acid positions described above that were subjected to site-directed mutagenesis in the present study.

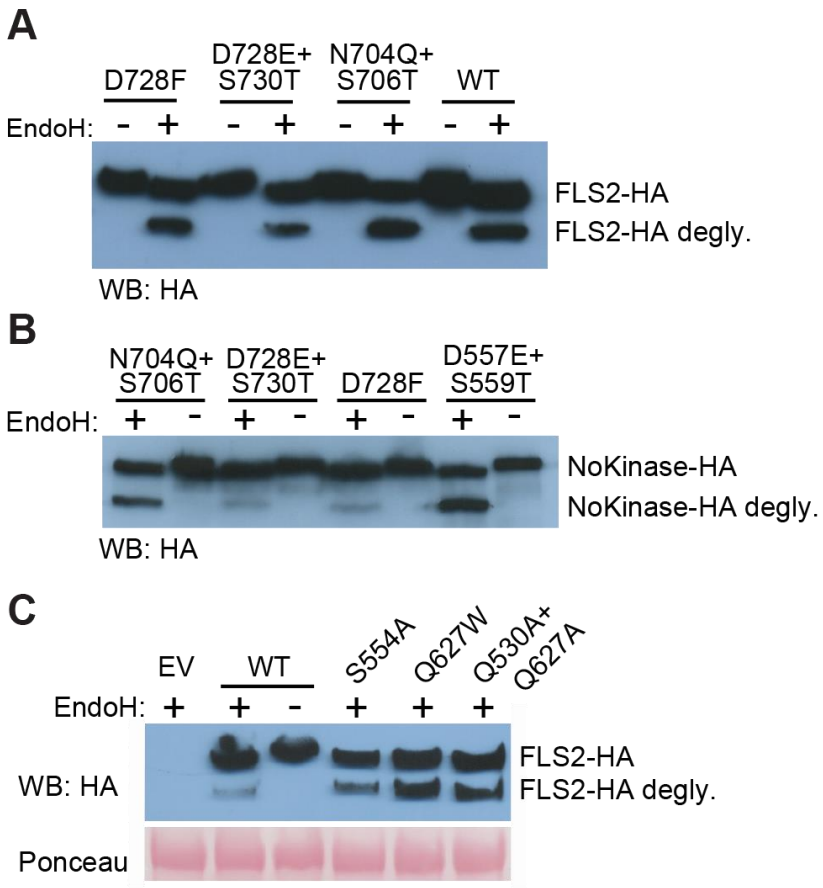

Figure 8: EndoH assay reveals no glycosylation defects in mutated FLS2 and FLS2-NoKinase. **(A)** Protein extracts from Arabidopsis *fls2*$^-$ leaves carrying P_{FLS2}-*FLS2-HA* with mutations as indicated or WT (no mutations), not digested (-) or digested (+) with endoglycosidaseH (EndoH). An EndoH-resistant protein pool (characteristic of mature glycosylated proteins) is visible in all EndoH-treated samples. **(B)** Protein extracts from *Nicotiana benthamiana* carrying *35S-FLS2-NoKinase-HA* with mutations as indicated, not digested (-) or digested (+) with endoglycosidaseH (EndoH). An EndoH-resistant protein pool is visible in all EndoH-treated samples. Mutations D557E+S559T were included as control mutations located in sites of a single LRR repeat analogous to D728E+S730T, but outside of the conserved LRR C-terminus. **(C)** Protein extracts from Arabidopsis *fls2*$^-$ seedlings carrying P_{FLS2}-*FLS2-HA* with mutations as indicated or WT (no mutations), digested with EndoH. An EndoH-resistant protein pool is visible in all EndoH-treated samples except for the empty vector (EV) control. Ponceau stained blot shows similar loading of total protein in all lanes including EV negative control. Degly.: FLS2 pools deglycosylated by EndoH.

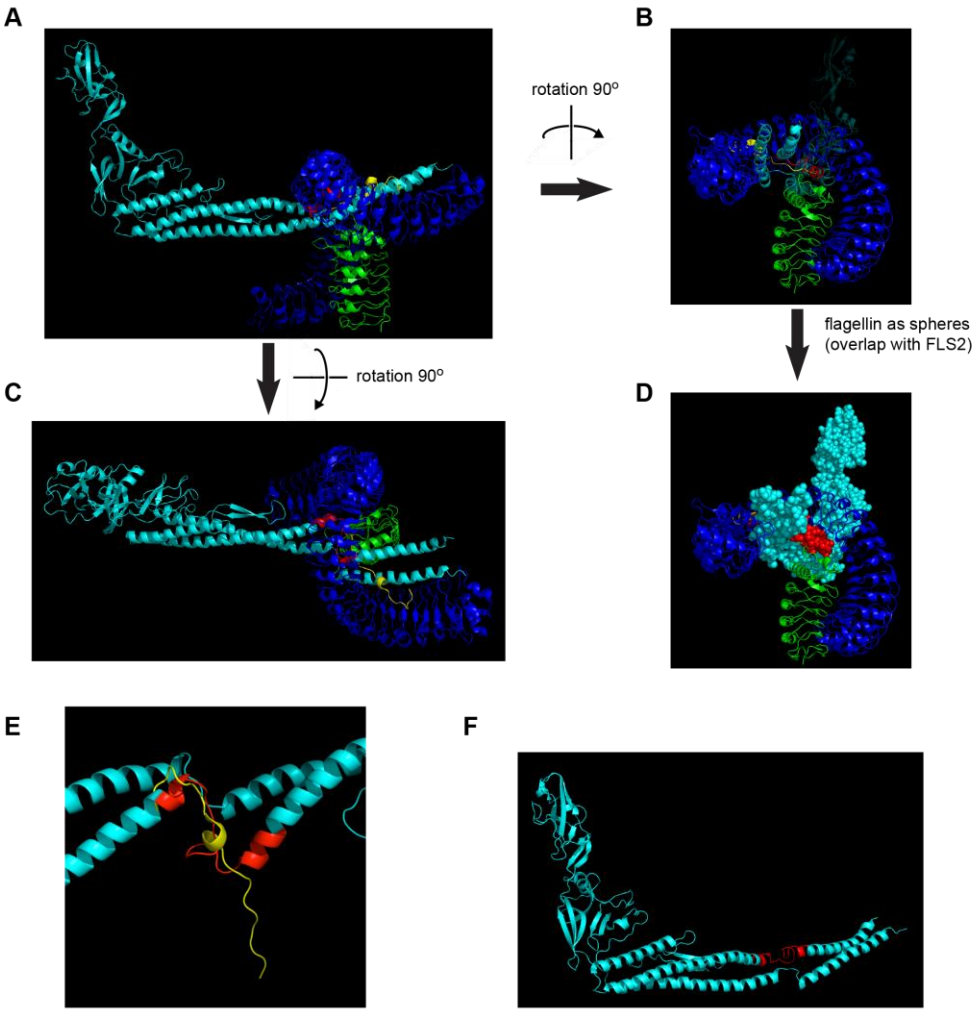

Figure 9: Hypothetical docking of full-length flagellin structure (PDB ID: 3A5X) to FLS2-flg22-BAK1 structure (PDB ID: 4MN8) illustrates minimal space for flagellin inside FLS2 LRR, and constraints to flg22 contact with FLS2 LRRs #3-15 if flg22 region is held within full-length flagellin. (A), (B), (C) Flagellin, hypothetically positioned so that flg22 residues within full-length flagellin are near the flg22 binding sites of FLS2 and BAK1. PDB structures 3A5X and 4MN8 superimposed at same scale; (B) and (C) are 90° rotated views of (A). Light blue: flagellin; red: flg22 residues within flagellin (3A5X). Dark blue: FLS2 LRR; green: BAK1 LRR; yellow: flg22 co-crystallized with FLS2 and BAK1 LRRs (4MN8). **(D)** Same view as (C), with space-filling representation of flagellin to more clearly illustrate impossible overlap of flagellin and FLS2 residues in same spatial locations in this arrangement (and other arrangements) of 3A5X and 4MN8. FLS2 and BAK1 side-chains omitted for clarity. **(E)** PyMol alignment of flg22 (yellow, in structure 4MN8) and flg22 region within flagellin (red, in structure 3A5X). Lower portions of flg22 in (E) (the yellow residues that are not proximal to red residues) are the N-terminal 7 residues of flg22 that associate with FLS2 LRRs #3-7 (FLS2 and BAK1 not shown, for clarity). **(F)** Full length flagellin (PDB structure 3A5X) colored as in (A) but shown by itself, showing that flg22 region forms a less-ordered hinge region between flanking pairs of alpha-helical bundles.

Chapter 6: Conclusions and future directions

In the past years significant progress has been made in the understanding of pattern-recognition receptors (PRRs) in plant and animal immune systems. New plant PRRs and microbial ligands have been discovered. PRR homologs have been identified and characterized in various plant species. New insights have been reported about receptor gene expression and gene methylation patterns, post-translational receptor modifications, ligand specificities, interactions with co-receptors and regulatory proteins, receptor signaling activation and attenuation, signaling crosstalk at the plasma membrane, receptor auto- and trans-phosphorylation, interactions with cytoplasmic kinases, and receptor recycling, endocytosis and degradation. These mechanisms have been elucidated using various model PRRs but also receptor-like proteins and receptor-like kinases at the plasma membrane that are not involved in plant immunity. Taking a closer look at a specific receptor of interest, many questions remain. My thesis research focused on the roles of the EFR and FLS2 ectodomains in ligand perception and signaling activation, and furthermore I explored approaches to engineer PRRs with novel ligand specificities.

In the study described in chapter 5, I showed that the FLS2 - BAK1 ectodomain complex initiates receptor signaling. The complex forms only in the presence of the ligand flg22. The subsequent availability of the FLS2LRR – flg22 – BAK1LRR co-crystal structure was a milestone for FLS2 research and had a major impact on my research projects. In chapter 5 I discussed the current knowledge about FLS2 signaling initiation. A list of some of the most important remaining questions about FLS2 ligand perception and signaling activation might include the following:

- Does full-length flagellin bind to FLS2, or does FLS2 more typically detect flagellin fragments containing the flg22 epitope?
- Is FLS2 present in a homodimer before flg22 binding?
- Is the BAK1 interaction site on FLS2 masked before flg22 binding?
- Do all SERK proteins interact with the FLS2 ectodomain at the same residues?
- How similar is FLS2 signaling in plant species other than Arabidopsis?
- Are there currently unidentified PRRs that can recognize a flagellin epitope different than flg22?
- Does the relative overall positioning of the FLS2 and BAK1 kinase domains change after flg22 binding?

Similar questions remain about EFR. In the study described in chapter 2 I cloned and sequenced nine new *EFR* LRR domain alleles from different Brassicaceae species. By comparing the sequences using Repeat Conservation Mapping, we identified several conserved solvent exposed amino acid clusters. Laura Helft, a former graduate student in our lab, performed site-directed mutagenesis in the ectodomain of EFR and tested signaling outputs of the mutated receptor. She identified several mutations that impaired receptor function. I generated additional site-directed mutations in the EFR ectodomain in order to identify a potential BAK1 interaction site. Some of the mutations that Laura or I made impaired receptor localization. Thus we have not yet been able to confirm the functions of all of the conserved amino acid clusters identified by Repeat Conservation Mapping. Future studies could aim to further elucidate EFR receptor homodimerization before and after elf18 binding, identification of the amino acids potentially required for interaction with the SERK protein ectodomains and

confirmation of the ligand binding amino acid cluster. Furthermore it would be interesting to know if full-length EF-Tu binds to EFR or only an EF-Tu fragment containing the N-terminal elf18 epitope.

Our current picture of the receptor landscape at the plant plasma membrane is incomplete. In the future, many more receptors and ligands will be identified. How many and what types of receptors are localized at the plasma membrane at the same time? How many of these receptors are PRRs/hormone receptors/regulator proteins/other types of receptors? How does the receptor landscape change in different cell types and different plant tissues and organs? How does it change between the same cell types in different plant species? How is signaling cross-talk at the plasma membrane coordinated and carried out? Are the SERK proteins the major players in the signaling cross-talk, or are there many remaining unidentified regulatory proteins? How often are the regulatory proteins that impact receptor activity co-ligand receptors? How specific are the ligand – co-ligand receptor interactions? How important are the small cytoplasmic kinases for signaling cross-talk?

Many of these questions might have to be answered before we successfully can integrate engineered PRRs with completely novel ligand perception capacities into the plant immune system. In the study described in chapter 5 we embarked on the adventure of engineering novel plant PRRs. This complicated task opened our eyes to all the meticulous details of receptor signaling activation. Each study aimed at the elucidation of PRR signaling activation reveals a new layer of complexity. Nevertheless, in our study we explored two techniques for the engineering of PRR with novel ligand perception abilities: yeast surface display for endogenous receptor evolution and VLRs as templates for novel ligand binding site

design. The following paragraphs summarize some general conclusions that can be drawn from our experiments.

Yeast-surface display is more applicable for receptors not sensitive to changes in glycosylation. Unfortunately in the studies involving yeast surface display I focused more on EFR than FLS2. EFR is much more sensitive to glycosylation defects than FLS2.

Using the hybrid LRR technique it was possible to improve the display of EFR. This technique applied to FLS2 could be powerful in improving FLS2 display and then proceed to FLS2 *in vitro* evolution.

It should be possible to engineer FLS2 towards novel ligand binding capacities. However to generate a functional novel receptor, it must be able not only to bind a new ligand but to recruit and position BAK1 correctly in order to form a signaling competent complex. This remains a substantial challenge for all but the most conservative changes in FLS2 ligand recognition capacity.

In the second approach we used VLRs as templates to generate novel ligand binding sites on the FLS2 and EFR ectodomains. We designed the receptors, generated the recombinant genes and expressed them in *N.benthamiana* and Arabidopsis plants and protoplasts. The proteins accumulated, but the receptors were not functional. Novel chimeric receptors must be tested for proper localization at the plasma membrane, for ligand binding, co-receptor interaction and correct signaling outputs. Thus we face the same challenges as in the first approach. Are VLRs appropriate templates to use for the design of novel ligand binding sites on plant PRR ectodomains? At this point in time we cannot answer this question because we did not perform ligand binding studies on our chimeric receptors. Testing our chimeric receptors

for ligand binding would have required additional *in vitro* approaches as our receptors were not localized properly at the plasma membrane *in vivo*.

Work is in progress by many scientists to create VLR scaffolds that can be engineered toward ligand specificities of interest. First attempts of such scaffolds were named “dVLRs”, “Repebodies” and “Lambodies”. Thus, we are just at the beginning of using the vast potential of the tractable VLRs for biotechnological applications.

Our knowledge about the plant immune system and its complexity is constantly growing. In the future it might be possible to successfully design and engineer novel PRRs and to integrate them into the immune system of agriculturally important plant cultivars. Improving plant immune systems not by adding genes encoding toxins, but instead by adding genes encoding receptors, might provide an environmentally friendly and societally acceptable approach to improving plant health.