Plant bioactive peptides: an expanding class of signaling molecules

Hugo Germain, Eric Chevalier, and Daniel P. Matton

Abstract: Until recently, our knowledge of intercellular signaling in plants was limited to the so-called five classical plant hormones: auxin, cytokinin, gibberellin, ethylene, and abscisic acid. Other chemical compounds like sterols and lipids have also been recognized as signaling molecules in plants, but it was only recently discovered that peptides in plants, as in animal cells, play crucial roles in various aspects of growth and development, biotic and abiotic stress responses, and self/non-self recognition in sporophytic self-incompatibility. These peptides are often part of a very large gene family whose members show diverse, sometime overlapping spatial and temporal expression patterns, allowing them to regulate different aspects of plant growth and development. Only a handful of peptides have been linked to a bona fide receptor, thereby activating a cascade of events. Since these peptides have been thoroughly reviewed in the past few years, this review will focus on the small putative plant signaling peptides, some often disregarded in the plant peptide literature, which have been shown through biochemical or genetic studies to play important roles in plants.

Key words: signaling peptide, ligand, receptor, cell–cell communication.

Introduction

The idea that plants, like animal cells, could use peptides as signaling molecules arose 15 years ago with the discovery of systemin, a plant peptide involved in the systemic response against herbivores (Pearce et al. 1991). Other processes, such as the control of shoot apical meristem development and organ initiation, were later shown to be controlled by the CLAVATA3 (CLV3) peptide (Fletcher et al. 1999). Subsequently, genomic data for Arabidopsis thaliana revealed that the genome encodes hundreds of protein receptor kinases (Shiu and Bleecker 2001), several of these possibly binding a single or a few related peptides, suggesting that the recently discovered peptides were probably not the exception. The changing picture in plant peptide signaling is due to the emergence of peptide families. CLV3, which has been extensively studied (Fletcher et al. 1999; Trotochaud et al. 1999; Trotochaud et al. 2000; Lenhard and Laux 2003), is now known to be part of one of these families (Cock and McCormick 2001). These discoveries are due in great part to the completion of the Arabidopsis genome (The Arabidopsis Genome Initiative 2000) and to the massive number of EST sequences gathered over the last few years in various plant species.
Assigning a ligand to a receptor remains a very difficult task. Even now, putative ligands are found randomly because they are often not annotated using conventional computational algorithms. Ligand-receptor direct interaction remains state-of-the-art biochemistry work: receptors are difficult to express in soluble form and ligands are generally only available in planta at nanomolar to femtomolar concentrations. Furthermore, a synthetic ligand may not bear the proper post-translational modifications or display proper folding, resulting in nonfunctional molecules, and T-DNA insertion lines are often aphantotypic owing to the redundancy found in large family of peptides harboring very conserved functional domains.

Despite all the above-mentioned hurdles, some investigators have succeeded in their quest and a few plant peptide ligands have been assigned to their respective receptors. For examples, the mechanisms of action for CLV3, phytoinsuflo-kine, systemin, auxin, gibberellin, and abscisic acid have also been well characterized and shown to interact with specific receptor proteins (Schaller and Bleecker 1995; He et al. 2000; Inoue et al. 2001; Dharmasiri et al. 2005; Ueguchi-Tanaka et al. 2005; Razem et al. 2006). This review will thus focus on peptide families and (or) putative small signaling peptides for which a receptor or a function has not been assigned yet. Some of these putative peptides have generally been disregarded in recent reviews of the plant signaling peptide literature, although many have been partially characterized at the functional level through biochemical analyses or through the use of mutant plants. Undoubtedly, many more of these putative bioactive peptides are likely to be characterized in the next few years and may well turn out to be involved in important plant processes. In this review, the peptides presented will be grouped into five functional categories that are linked to their putative roles in planta.

**Development**

The BRICK genes

All multicellular organisms arise from one cell that divides to produce a multicellular organism with several cell types. The process by which the daughter cell acquires a different fate (cell type) is termed asymmetric cell division (Scheres and Benfey 1999). Stomata also arise from asymmetric cell divisions. To generate a normal stomata, a first asymmetric cell division must generate the guard mother cell (GMC), followed by a polarized alignment of the subsidiary cell. The GMC will become the guard cells (Gallagher and Smith 2000). The BRICK genes are ascribed small signaling peptide in plants. The BRICK1 gene (the only one yet characterized) encodes a short 252-bp cDNA and the corresponding 8 kDa peptide (Frank and Smith 2002) lacks features such as a signal peptide or the presence of conserved cysteine residues, which are often associated with diffusible peptides. The gene is highly conserved in the plant and animal kingdoms where it is involved in actin polymerization (Eden et al. 2002). It is too early to conclude that the BRICK genes are part of the same peptide family; it is possible that they represent different proteins acting in the same signaling pathway. It is nonetheless surprising that the three genes can act in a non-cell-autonomous manner, suggesting that their gene product is diffusible. The lack of a signal peptide and the fact that the gene product acts over a very short distance suggest that the peptide(s) can travel through plasmodesmata. Short-root, a transcription factor has been shown to travel (when fused to GFP) through the plasmodesmata (Nakajima et al. 2001), and proteins as big as 61 kDa have been shown to use these channels (Oparka et al. 1999). It is therefore possible that the 8 kDa peptide encoded by the BRICK1 gene either moves through the plasmodesmata or uses a non-conventional/non-classical leaderless secretion pathway as shown for some mammalian proteins (Bendtsen et al. 2004b), and acts as paracrine hormones on the surrounding cell to promote proper localization of actin patches in epidermal cells.

**CLAVATA3-like / ESR (CLE)**

CLAVATA3 (CLV3) is one of the first and best-characterized small signaling peptide in plants. The clv3 mutant causes the same phenotype as the clavata1 (clv1) mutant (Clark et al. 1995) and encodes a small peptide of 96 amino acid acids (Fletcher et al. 1999) with an N-terminal signal peptide, producing a mature peptide of 76 amino acids after cleavage (Rojo et al. 2002). Direct interaction with CLV1, although previously published, has now been retracted (Nishihama et al. 2003) (for a review on CLAVATA3s, see
Clark 2001). Although CLV3 was initially described as a protein with no similarities with other proteins from public databases, it was later found to display weak similarities (E-values ranging from 0.15 to 0.45) with the embryo surrounding region proteins (ESR) (Cock and McCormick 2001). ESR genes are expressed only in the embryo surrounding region, and three members have been found in the maize genome. All three ESRs encode small peptides of 14 kDa, which contain a signal peptide (Opsahl-Festad et al. 1997). Cell fractionation, immunocytological results, and transient expression assays all agree that the ESR proteins are secreted proteins that remain mostly in the embryo surrounding region but are also present in the basal endosperm layer (Bonello et al. 2002).

The homology between CLV3 and the ESRs is low when the overall protein is considered but high for a stretch of 14 amino acids located in the C-terminal region. BLAST searches with this highly conserved region retrieved a total of 42 related sequences from both monocots and dicots (Cock and McCormick 2001). Twenty-six are from Arabidopsis and half are located on chromosome 1 (Cock and McCormick 2001; Hobe et al. 2003; Sharma et al. 2003). Sequence conservation between all these peptides is very low (except for the C-terminal domain). Nearly all these putative peptides have a signal peptide, are short, and are intronless (except for CLV3 and CLE40), and have a very conserved C-terminal domain (Fig. 1). Since most A. thaliana CLEs originate from genomic sequences and could be pseudogenes, Sharma et al. (2003) used RT-PCR to determine whether the A. thaliana CLEs are expressed. Expression is detectable for all but one (CLE26) gene. Transient expression in leaf epidermal cells of CLV4, CLV8, and CLE40 fused to GFP revealed that these three CLEs have a functional signal peptide that directs the protein to the cell wall or the extracellular space (Sharma et al. 2003). Single amino acid changes in the CLE motif, as observed in the clv3-1 and clv3-5 mutant, disrupts CLV3 function, indicating the importance of the CLE domain for CLV3 activity (Fletcher et al. 1999).

Of these CLE genes, two have been shown to cause a developmental defect when their expression pattern is altered. One of the two is CLE19. When CLE19 expression is driven by the CaMV 35S promoter, the transgenic plants have a short-root phenotype in which the root meristem becomes fully differentiated, pistils are pin-shaped (without carpel and ovules), and flowers have disconnected vascular bundles (Fiers et al. 2004). It should be noted that T-DNA mutants are aphenotypic (Fiers et al. 2004). A suppressor screen of Brassica napus for the CLE19 phenotype revealed two loci, named SOL1 and SOL2 (suppressor of LIGAND-LIKE PROTEIN, the closest homolog to AtCLE19 from B. napus), that fully restore the wild-type phenotype (Casamitjana-Martinez et al. 2003). It should be noted that SOL1 and SOL2 do not have a phenotype of their own. The cloning of SOL1 revealed that it encodes a transmembrane Zn2+-carboxypeptidase, which is possibly involved in processing of the peptide. Although processing of CLV3 has not been demonstrated, it has been shown for other putative peptide ligands (Yang et al. 1999; Pearce et al. 2001b). The facts that CLE19 overexpression causes a defect in root meristem maintenance and SOL1 encodes a processing enzyme suggest that CLE19 is involved in a CLV-like pathway (Casamitjana-Martinez et al. 2003).

The second CLE gene that has a known phenotype is CLE40. This gene is the only CLE gene that has a similar genomic organization to CLV3, both having two introns, whereas all other CLE genes lack introns. Overall, the two sequences share very little sequence similarities, but their CLE domain is 70% identical, and both of these CLEs, unlike others, have a short stretch of amino acids C-terminal to the CLE domain (that terminates with a proline). Plants overexpressing CLE40 have phenotypes similar to the ones observed in 35S::CLV3 plants (Hobe et al. 2003). Not only are the phenotypes similar, but CLE40 can also downregulate WUSCHEL expression (like CLV3) and CLE40 requires functional CLV1 receptor complex for its shoot apical meristem activity, indicating that CLE40 when expressed at sufficient levels can substitute for CLV3 in regulating SAM cell differentiation. 35S::CLV3 and 35S::CLE40 also have the same effect on root development; both cause initial growth retardation, eventually leading to a total root growth arrest. A T-DNA insertion in the CLE40 gene causes a root-waving phenotype, indicating that CLE40 is required for normal root development and that other CLEs, albeit similar in their CLE domain, could not substitute for CLE40. Application of chemically synthesized partial peptides (p) of CLV3, CLE19 and CLE40, corresponding to their CLE domain, is sufficient to reduce root length (Fiers et al. 2005). CLV2, but not CLV1 or CLV3, is also involved in CLV3p, CLE19p, and CLE40p recognition. Synthetic CLE5p is unable to reduce root length, indicating that not all CLEs act in this pathway.

Using a motif search Olsen and Skriver (2003) found a CLE-like sequence from oesophageal gland cell library of the plant parasitic nematode Heterodera glycines (HgCLE) (Gao et al. 2001; Wang et al. 2001). This plant-infecting nematode possesses a hollow protrusable stylet capable of piercing the plant cell wall and excretes products generated by the oesophageal gland into the plant cell, thereby transforming it into a feeding cell via the action of parasitism proteins (Davis et al. 2000). An antibody directed against HgCLE revealed that the protein is present only in secretory granules of the oesophageal gland of the parasitic stage of H. glycines dissected from soybean roots, and in secretory granules that migrate into the collecting reservoir at the base of the stylet (Wang et al. 2005). Transgenic plants overexpressing HgCLE in wild-type background display floral SAM, and root phenotypes reminiscent of CLE19 and CLE40 misexpression, and also decrease WUS expression as does CLV3. 35S::HgCLE transformed in clv3-1 (a strong CLV3 mutant) background is able to partially or fully rescue the phenotype, indicating that HgCLE, like CLE19 and CLE40, can also substitute for CLV3 when expressed in the SAM (Wang et al. 2005).
genes potentially involved in infection processes between bacteria and plants or nematodes and plants, and is generally believed to be the result of horizontal gene transfer (Collmer 1998; Yan et al. 1998).

RALF
The discovery of the tomato systemin revealed that this peptide could induce the alkalinization of the medium when added to a suspension cell culture of *Lycopersicon peruvianum* (Felix and Boller 1995). Using this property as an assay for systemin in tobacco leaf extract, Pearce et al. (2001b) found a 5 kDa peptide that causes a stronger and faster alkalinization response than systemin itself (Pearce et al. 2001b). This peptide, which was called Rapid Alkalinization Factor (RALF), was also purified from tomato and alfalfa leaves (Pearce et al. 2001b). Meanwhile, Haruta and Constabel (2003) also purified a RALF homolog from poplar cell culture using a similar alkalinization assay.

RALF-like genes have now been found in more than 16 species (Pearce et al. 2001b), including 34 RALF-like genes in the *Arabidopsis* genome. The fact that this family has emerged from duplication events is strengthened by the fact that *AtRALF* copies often occur in tandem repeats in the *A. thaliana* genome and show high sequence similarity (Olsen et al. 2002). The *Nicotiana tabacum* RALF (NtRALF) is a 49-amino-acid peptide that is processed from a preproprotein of 115 amino acids. The preproprotein has a N-terminal signal peptide and a highly conserved mature peptide that is located at the C-terminus of the protein (Fig. 1). The activity of NtRALF requires the formation of disulfide bridges between Cys-18 and Cys-28 and between Cys-41 and Cys-47. All RALF-like genes identified share this common architecture. Another common feature that is conserved among preproRALFs is an acidic segment upstream of a dibasic residue motif possibly involved in the processing of the peptide (Cleland 1995; Franssen and Bisseling 2001; Olsen et al. 2002) via the action of the subtilisin-like serine protease of the kexin family. In animal cells,
these proteases are involved in the activation of peptide hormones, growth factor and viral proteins (Siezen and Leunissen 1997). Even if the cleavage specificity is unknown in plants, serine proteases are abundantly found in plants and are, therefore, good candidates for RALF processing (Siezen and Leunissen 1997; Franssen and Bisseling 2001). It should be noted that contrary to a previous report (Boller 2005), RALF has never been shown to be a glycopeptide.

The physiological effect of systemin is similar to that of erythrosin B, an inhibitor of plasma membrane proton ATPases, which causes an alkalinization of the medium, whereas suramin, an inhibitor of ligand/receptor interaction, inhibits the alkalinization of the medium. It was therefore concluded that the alkalinization of the medium mediated by systemin is the result of an inhibition of the plasma membrane proton ATPase by a receptor-mediated signaling pathway (Stratmann et al. 2000). Since RALF induces the phosphorylation of a 48 kDa MAPK and the concomitant alkalinization of the medium like systemin (Pearce et al. 2001b), and since this alkalinization is inhibited by suramin, it could act as a secreted ligand that controls cellular events by its interaction with a membrane receptor. Recently, transient expression of AtRALF in Nicotiana benthamiana leaves revealed that it localizes to the cell wall after being observed transiently in the endoplasmic reticulum (Escobar et al. 2003). Moreover, the use of a photoactivatable analog of the tomato RALF revealed two RALF interacting proteins of 120 and 25 kDa. It appears that these two proteins act in a membrane complex, since only one high affinity site ($K_D = 0.8 \text{ mmol/L}$) was characterized (Scheer et al. 2005).

The challenge now is to uncover the biological function of the RALF peptides in plants. RALF does not induce the synthesis of tobacco trypsin inhibitors (Pearce et al. 2001b) and treatments with elicitors and defense or wound hormones do not induce the transcription of RALF-like genes in cell cultures (Haruta and Constabel 2003) or intact plant tissues (Germain et al. 2005a). Moreover, the expression of some RALF-like genes in A. thaliana is not altered in ethylene-constitutive-background (ctr1) and in systemic-acquired-resistance-background (mpk4) mutants (Olsen et al. 2002). Therefore, RALF is most probably not involved in defense responses.

However, RALF-like genes are found in several EST databases (Pearce et al. 2001b; Germain et al. 2005b), and tissues expression analyses of RALF-like genes in poplar (Haruta and Constabel 2003), Arabidopsis (Olsen et al. 2002) and Solanum chacoense (Germain et al. 2005a) have revealed specific and narrow expression patterns for some RALFs and broader expression patterns for others. Therefore, RALFs could control various developmental processes such as fruit development and root elongation. In fact, two RALF-like genes are specifically expressed in fruits after pollination (Germain et al. 2005a), and the addition of a synthetic tomato RALF peptide to the medium arrest root growth of tomato and A. thaliana seeds (Pearce et al. 2001b) and decreases the elongation rate of pollen tube in vitro (Bedinger et al. 2004). A detailed analysis revealed that Arabidopsis roots grown in presence of a tomato RALF have fewer root hairs and enlarged root meristem cells (Pearce et al. 2001b). Taken together, these results suggest that RALF could be involved in the control of cell growth, perhaps through an exchange of protons across the plasma membrane. Plant cell growth has already been shown to be modulated by auxin via a similar proton pumps mechanism (Cleland 1995).

A better understanding of the biological function of RALF will await the analysis of transgenic lines. However, the high sequence similarity shared by some RALFs and the presence of numerous RALF genes per genome could obscure the effect of overexpression, knock-down or knockout lines. Promoter and in situ analyses will pinpoint the specific expression domains of each RALFs and the identification of the protease(s) that processes RALF will allow a better understanding of the activation mechanism of this pre-protein. Finally, the identification of the recently published RALF-interacting membrane proteins should lead to the deciphering of this signaling cascade.

**DEVIL/ROTUNDIFOLIA**

DEVIL1 (Wen et al. 2004) and ROTUNDIFOLIA4 (Narita et al. 2004) are two different genes that were found nearly at the same time by two independent groups, and that are part of the same gene family. DEVIL1 (DVL1, which was described first), and ROTUNDIFOLIA4 (ROT4) were both discovered by activation tagging in Arabidopsis. In the case of DVL, the phenotype caused by the insertion of the tag is pleiotropic, with a shorter stature, rounder rosette leaves, clustered inflorescences, shorter pedicels and horned fruit tips (hence the name Devil). In the case of rot4, the phenotype has rounder and shorter leaves.

Surprisingly, in both cases, the activation tag does not induce the expression of any of the nearby annotated genes. Thus, the investigators searched for short ORFs in the genomic DNA surrounding the tag. Short ORFs encoding small peptides of 51 (DVL1) or 53 (ROT4) amino acids were highly overexpressed in the tagged line when compared to the wild type plant. When a genomic fragment encoding DVL1 cloned downstream of four CaMV 35S transcriptional enhancers was reinserted in the wild type plant, it mimicked the phenotype observed in the activation tag line, thereby confirming that the previously unannotated ORF is responsible for the observed phenotype (Wen et al. 2004).

In total, 22 members of the DVL/ROT family were found (ROT4 is DVL16) and all of them encode highly homologous small protein devoid of a signal peptide (Fig. 1), and most of them (16 of 22) (Narita et al. 2004) are present in EST databases but are absent from EST or genomic sequences outside the angiosperms (Wen et al. 2004). Overexpression of various family members, DVL2, DVL3, DVL4, and DVL5, also cause a similar pleiotropic phenotype. RNAi and antisense overexpression of DVL1 are sufficient to restore the normal phenotype in the activation-tagged line, but do not confer a phenotype when transformed in wild-type plants (Wen et al. 2004), suggesting a functional redundancy between the family members.

The conserved RTF domain found at the C-terminus (Narita et al. 2004) consists of 29 amino acids, which are highly conserved and contains several basic residues and a pair of cysteines that is present in all but one member of the group (DVL21 contains only one cysteine, H. Germain, unpublished observation). The overexpression of the ROT4
peptide without the upstream N-terminal sequence (to the RTF domain) or without the downstream C-terminal sequence (to the RTF domain) is sufficient to confer the round leaves phenotype (Narita et al. 2004), while a frameshift or point mutations in the C-terminal domain (after RTF) of DVL1 fails to cause the phenotype (Wen et al. 2004). These results indicate that the RTF domain is necessary and sufficient to cause the DVL/RTF phenotype and that its integrity is needed for peptide function. The high level of conservation in the C-terminal domain is reminiscent of other families of putative signaling peptides such as CLE (Cock and McCormick 2001) and RALF (Olsen et al. 2002), which also have highly conserved C-terminal domains.

ROT/DVL family members do not have a classical signal peptide, one that can be predicted using SignalP (Nielsen et al. 1997), but ROT4 is located at the plasma membrane, possibly as part of a membrane complex or as a peripheral membrane protein (Narita et al. 2004). Because these proteins can be secreted despite the lack of a leader sequence, a process referred to as leaderless secretion or non-conventional/classical secretory pathway (Bendtsen et al. 2004a), we have used SecretomeP 1.0, a sequence-based software (Bendtsen et al. 2004b) to predict whether the peptides described in this paper and lacking a leader sequence are proteins secreted via the non-conventional secretory pathway. All DVL/ROT peptides (except DVL18) have a predicted value above threshold (0.6), indicating that they are likely secreted through a ER–Golgi independent secretory pathway (mean = 0.758, SD = 0.09, n = 22), which would explain their localization at the plasma membrane (Narita et al. 2004). Although a highly conserved cysteine pair is observed in DVL/ROT, the fact that these proteins do not go through the default secretory pathway suggests that these cysteines are not part of a disulfide bridge. Based on GFP results, these peptides do not appear to be mobile (Narita et al. 2004) and their capacity to act as signaling peptide remains to be demonstrated. One T-DNA line was found in Arabidopsis and two in rice. Although these mutants should disrupt the RTF domain, none of them display a visible phenotype, supporting the possibility of domain and (or) functional redundancy among the related family members (Narita et al. 2004). Although activation tagging, which has started to be used only recently with Arabidopsis (Weigel et al. 2000), should prove a useful method to discover small peptides that are not currently annotated in the Arabidopsis genome, care should be taken when drawing conclusion with regards to the possible function of the activated gene since the strong promoters used for activation tagging could trigger ubiquitous expression at non-physiological levels and could therefore cause an artefactual phenotype.

POLARIS

POLARIS (PLS) was identified in a promoter trap transgenic line in which reporter gene expression is specifically detected in the basal region of the embryo and the root tip (Topping et al. 1994; Topping and Lindsey 1997). PLS is also expressed in the basal part of hydra embryo and emb30 mutant, the latter lacking shoot and root meristems, indicating that PLS is not a root marker (Topping and Lindsey 1997). It was later shown that PLS is also expressed in the leaf vasculature (Casson et al. 2002). A mutant homozygote for a T-DNA insertion in the PLS gene has a short trichomeless root, whose growth responds less than wild-type plants to exogenous application of cytokinin. Overexpressing lines display a more complex leaf venation than the wild-type. Uncommonly, the initiation of the short PLS transcript (108 nt) takes place within an upstream gene. The 4.6 kDa (36 amino acids) peptide that is predicted from the 108 bp ORF has no sequence identity with any peptide in the public databases, signal peptide or cysteine pairs, and has not been shown to be mobile. Results from a complementation experiment using a PLS ORF carrying either a mutated or an unmutated ATG initiation codon (ATC) suggests that PLS could act as a peptide and not a RNA molecule. All plants lines generated with the mutant start codon are unable to rescue the pls mutant, whereas the plants with a wild-type ATG could (Casson et al. 2002). Investigation of the PLS promoter using a GUS reporter gene showed that it is inducible by functional auxins and not by non-functional auxin analogs (Casson et al. 2002) and intense GUS staining was observed at the site of lateral root initiation, a process induced by auxin (Topping and Lindsey 1997). Casson et al. (2002) therefore concluded that auxin could be a regulator of the spatial patterning of PLS expression.

Whether PLS is really a secreted signaling peptide still remains hypothetical and clarification on that matter will await additional data coming from GFP fusion or immunochemistry experiments.

Reproductive development-related peptides

IDA and IDA-like genes

Abscission is the process by which the plant sheds its leaves, flowers or fruits in a controlled manner at a special separation layer named the abscission zone (AZ). The flower generally sheds all its floral parts after they became useless following the fertilization process (Bleecker and Patterson 1997).

A new and unique mutant, Inflorescence Deficient in Abscission (ida), was identified in Arabidopsis (Butenko et al. 2003). In the ida mutant all floral parts remain attached indefinitely to the plant body after fertilization, and this mutant, unlike other abscission mutant, is sensitive to ethylene, which has previously been shown to be involved in abscission (for a review see Roberts et al. 2002). When the GUS reporter gene is fused to the IDA promoter, a very specific expression pattern is observed and staining is restricted to the surrounding area of the AZ at specific developmental stages. The break strength, which is required to remove the petals at different stages following pollination, decreases in both the ida mutant and in wild-type plants, without ever reaching zero in the ida mutant. Afterward, only in the ida mutant does the break strength progressively return to its initial value. Scanning electron microscopy revealed some differences in the appearance of the abscission zone between ida and wild-type plants. In wild-types, the cells at the fracture plane present a rounded appearance whereas the cells of the ida mutant display only flattened cavity (Butenko et al. 2003).

The ida mutant was found in a T-DNA screening for lines delayed in abscission and the identity of the mutated gene causing the ida phenotype was confirmed by complementa-
Egg apparatus 1 and transparent leaf area 1

In maize, two peptides recently identified by the Dresselhaus group have been shown to have profound effect in reproduction and leaf development (Dresselhaus et al. 2005; Marton et al. 2005). In higher plants, the pollen tube, which contains the sperm cells, grows through the transmitting tract of the style and is then believed to be guided towards the female gametophyte by unknown molecules. ZmEA1 (Zea mays Egg Apparatus 1) is a small predicted protein of 94 amino acids with a predicted transmembrane domain. It is expressed (RT-PCR and GUS staining experiments) exclusively in the egg apparatus before fertilization and it disappears after fertilization (Marton et al. 2005). Transgenic plants for the ZmEA1 gene (RNAi and antisense) display a lower seed set, but wild-type plants crossed with transgenic pollen do not show any phenotype. C-terminal GFP fusion indicated that the signal originating from the egg apparatus eventually spreads to the surface of the nucellus at the micropylar opening of the ovule, suggesting that the small protein is mobile. The authors suggest that there is proteolysis of the conserved C-terminal region from the transmembrane domain to allow secretion of the mature peptide. Putative orthologs of ZmEA1 have been found in rice, barley, millet and Tripisicum dactyloides, but apparently not in dicotyledous species (McCormick and Yang 2005). Overall peptide sequence conservation is low, but the C-terminal domain is highly hydrophobic and very well conserved. In fact, when only the conserved C-terminal domain (Fig. 1) is used for similarity searches, more than 60 independent sequences are retrieved, including both monocot and dicot species, even basal angiosperm species (Gray-Mitsumune and Matton 2006). Several methionines are present in the ZmEA1 peptide; in fact, four methionines are found in frame upstream of the putative transmembrane domain. This makes selection of the proper initiation codon difficult and produces some uncertainty about the resultant peptide sequence. Taking into account the Kozak consensus sequence (RCA AUGG, R is a purine) (Joshi et al. 1997; Kozak 1999), only the third predicted methionine would have a near Kozak consensus (3/4 outside of the AUG), while the other methionines have only one or two out of four matches. Using the most likely methionine (the 3rd one) as the translation start site would yield a peptide of 76 amino acids that contain a strongly predicted signal peptide and, after cleavage, a mature protein of 49 amino acids. This observation is also supported by the fact that the second methionine of OseAL1 has a perfect Kozak consensus sequence (4/4) and generates a peptide that also has a signal peptide. It should be noted, however, that this is not true for all EA-like proteins, including OseAL2. The fact that the authors observed that a ZmEA1-GFP fusion could be detected away from its synthesis site would in fact be better explained if the protein started from an alternative initiation codon, in a better translation initiation context, and had a bona fide signal peptide.

The other maize small putative signaling peptide, named TRANSPARENT LEAF AREA1 (TLA1), was found in a search for genes that display high expression levels and egg cell specificity (Dresselhaus et al. 2005). The transcript encoding ZmTLA1 varies in length from 310 nt to 502 nt due to the presence of seven alternative poly(A) signal in its 3′-UTR. The ORF encoded by the ZmTLA transcript is 27 amino acids in length, 21 of which are hydrophobic. Since transposon-tagged lines of ZmTLA were unavailable, the authors used antisense overexpression to investigate the function of ZmTLA. A pleiotropic phenotype is observed and its most striking features are dwarfism and the presence of yellow or transparent leaf areas (tla). The tla phenotype is observable on most plants displaying high transgene expression, and can be visible from the seedling stage, throughout the plant’s life, and is not accompanied by necrosis. A defect in anther maturation is also observed. Overexpression of sense ZmTLA is lethal in maize and Arabidopsis (from 500 000 selected progeny seedlings derived from Arabidopsis, none survive the primary leaf stage). The use of a dexamethasone-inducible promoter revealed that induction of ZmTLA-GFP in Arabidopsis causes developmental arrest. In plants that strongly overexpress ZmTLA-GFP, the elongation of hypocotyl cells, the development of primary root, and root hair differentiation are suppressed.

In plant lines expressing a ZmTLA-GFP construct, fluorescence is observed only in the apoplast (Dresselhaus et al. 2005). With appropriate controls, the authors showed that ZmTLA-GFP does not accumulate in the endoplasmic reticulum, Golgi or any other vesicle, indicating that ZmTLA is secreted into the extracellular space, but bypasses the default secretory pathway. BLAST analysis against the databases does not retrieve any similar sequences, although DNA blot analysis reveals hybridizing bands in rice and other cereals. These data taken together suggest that ZmEA1 could be a mobile signaling peptide that is involved in pollen tube guidance, at least in maize. Whether ZmEA1 binds to a receptor at the tip of the pollen tube to control its growth remains hypothetical, although pollen tube specific receptor-like kinase have already been discovered (Muschietti et al. 1998; Kim et al. 2002). As for TLA, it would represent the first plant proteolipid for which a function has been attributed.

The HT self-incompatibility modifier

In the Brassicaceae sporophytic self-incompatibility system (SSI), pollen recognition and rejection is determined by the interaction between a highly polymorphic pair composed of a small pollen secreted peptide, the SCR/SP11 ligand, and
a stigma-expressed receptor kinase (SRK) (Nasrallah 2002; Nasrallah 2005). A completely different mechanism is found in the gametophytic self-incompatibility (GSI) system of the Solanaceae, Rosaceae and Scrophulariaceae, where a polymorphic pair composed of a stylar-expressed ribonuclease (S-RNase) interacts with a pollen-expressed F-box protein, possibly targeting the S-RNase to the ubiquitin-proteasome degradation pathway (McClure 2004). Thus, in S-RNase-based GSI systems, the simplest model implies that the cytotoxic S-RNases that enter the pollen tube are targeted for degradation in a self-compatible pollination (thus the F-box protein acts as an inhibitor of non-self S-RNases by tagging them for degradation), while in a self-incompatible pollination, interaction of the S-RNase with its cognate F-box protein inhibits degradation of the S-RNase, which then can act on cellular RNAs of the pollen tube, leading to pollen tube growth arrest. Although these two proteins are necessary and sufficient for the recognition and rejection events, numerous experiments have demonstrated that other stylar factors are also necessary for the proper expression of the self-incompatibility phenotype (reviewed in McClure et al. 2000). Such factors, often considered as modifier loci, are present in the genetic background of SI plants, unlinked to the S-locus, and often lost in self-compatible (SC) relatives of self-incompatible (SI) species.

One such candidate for a modifier gene was originally cloned in Nicotiana alata, and named HT (McClure et al. 1999). The NaHT gene was cloned based on a differential screen between stylar expressed mRNAs from SC Nicotiana plumbaginifolia and a SC accession of N. alata that is defective in S-RNase expression, but competent to express SI (Murfett et al. 1996). Antisense NaHT plants with reduced level of the HT protein, but with normal levels of S-RNases are either fully of partially self-compatible, when pollen tube growth is measured in the style of these transgenic plants (McClure et al. 1999). This strongly suggested that the NaHT gene encodes a modifier factor that is necessary for the SI reaction to occur.

Phylogenetic analyses of numerous NaHT homologues isolated in various solanaceous species clearly demonstrated that two different HT isoforms exist (O’Brien et al. 2002). In Solanum chacoense, where the expression of both isoforms is suppressed either through antisense or RNA interference, only the ScHT-B isoform, the most closely related to the NaHT gene (now renamed HT-B), is involved in transforming a SI plant to a SC plant (O’Brien et al. 2002). Suppression of the HT-A isoform has no effect on the self-incompatibility phenotype, although both isoforms share extensive sequence similarities and are coordinately expressed in styles, as for the S-RNases (O’Brien et al. 2002). All the HT proteins share some common features (Fig. 1). First, a highly conserved N-terminal region that is predicted to be a signal peptide that upon cleavage would release an approximately 75–80 amino acid peptide (approx. 8.5 kDa). Experimental evidence from N-terminal sequencing of a partially-purified extract of the peptide from N. alata indicates that the predicted signal sequence is indeed cleaved (McClure et al. 1999). Second, and similar to the RALF peptide family (Pearce et al. 2001a), a dibasic residue motif is also found in the mature HT protein, suggesting that the HT protein could be subjected to further processing. Furthermore in N. alata, an antiserum raised against the NaHT protein also detects a faster migrating fragment, that corresponds to a protein fragment cleaved at the dibasic residue motif position as determined from N-terminal sequencing (in that region QKI is found in N. alata instead of QKK as in all other species determined). This fragment would correspond to an approximately 50-amino-acid peptide of roughly 5.5 kDa if no further post-translation modification occurs; however, it migrates at a higher molecular weight, suggesting that the fragment released undergoes modification (McClure et al. 1999). Third, all HT homologues possess a C-terminal region composed of consecutive stretches of 15–20 aspartic and asparagine and acidic residues, flanked by conserved cysteine motifs (CX2-CXC and CX3-CC), which are probably involved in disulfide bridges. The HT proteins are structurally similar to most of the previously described peptides, and could thus be also considered as putative signaling peptides. Furthermore, no direct interaction between the HT-B protein and the S-RNase could be detected either biochemically (McClure et al. 1999) or in a two-hybrid system (O’Brien et al. 2002), suggesting that HT proteins and S-RNases do not interact directly. Apart from its role in self-incompatibility barrier breakdown, only one other phenotype could be associated with the loss of HT-B expression. In RNAi ScHT-B lines displaying severely reduced HT-B mRNA levels, flower senescence and stylar abscission are markedly retarded following a SI pollination, suggesting that HT-B may also have a more general role in floral abscission, like the IDA peptide described previously. This is supported by a link between HT-B mutant plants and ethylene, a plant hormone involved in senescence and abscission (M. O’Brien and D.P. Matton, unpublished results). Plants with reduced levels of HT-B (either transgenic lines or WT accessions that have a naturally low level of HT-B transcripts) and treated with ethylene produced small parthenocarpic-like fruits that never developed to a size greater than an ovary 4–5 d after pollination.

**Defense-related peptides**

**Systemin-like gene family**

Tobacco and tomato plants release proteinase inhibitors as a defense mechanism against herbivores (or wounding) (Green and Ryan 1972). The mediator of this defense response, systemin (in tomato), is a small mobile 18-amino-acid peptide active at femtomolar concentrations (Pearce et al. 1991). It is also found in potato, nightshade, and pepper plants (Constabel et al. 1998), and it binds to a receptor kinase of the leucine-rich repeat family (Scheer and Ryan 2002). Synthetic systemin is as active as the native systemin (for reviews on systemin see Ryan et al. 2002; Torii 2004; Wang and He 2004). Neither EST nor polypeptide corresponding to tobacco systemin could be found despite the fact that tobacco displays a very similar response to wounding, inducing the accumulation of proteinase inhibitors. However, the use of an alkalinization assay enabled the identification of two tobacco peptides that have effects similar to the tomato systemin. The two peptides, named tobacco hydroxypoline-rich systemins (TobHypSysI and TobHypSysII) are both 18 amino acids long, contain hydroxyprolines, and are linked to pentose residues (Pearce et al. 2001b).
al. 2001a). Notably, synthetic TobHypSysI and TobHypSysII are 10,000 times less active than their native counterpart, indicating that the sugar decoration is important for the peptide function. Both peptides are encoded as a prohormone by the same 165 amino acids coding mRNA, a phenomenon novel for a plant peptidic hormone. Tomato also contains similar systemin-like peptides, named TomHypSysI, TomHypSysII, and TomHypSysIII, which are, respectively, 18, 20, and 15 amino acids long, are glycosylated, contain hydroxyprolines, and, as for the tobacco HypSys, are all derived from the same precursor protein (Pearce and Ryan 2003). In situ mRNA hybridization revealed that TomHypSys is synthesized in phloem parenchyma cells in response to wounding, systemic, or methyl jasmonate (Narvaez-Vasquez et al. 2005). A GFP fusion of TomproHypSys demonstrated that the precursor protein localizes to the cell wall of epidermal cells, whereas transmission electron microscopy revealed that the precursor protein of TomproHypSys is found in the cell wall matrix of vascular parenchyma cells (Narvaez-Vasquez et al. 2005).

Striking differences between the original systemin and these systemin-like peptides include the presence of an N-terminal signal peptide, hydroxylation of proline residues, and glycosylations. These features, all absent from systemin, indicate that the peptides must go through the secretory pathway (hydroxylation of proline residues and glycosylation are post-translational modifications), whereas systemin does not have a signal peptide and is believed to be synthesized on cytosolic ribosome (Pearce and Ryan 2003). It should be noted that although tobacco does not have a true systemin ortholog, it responds to systemin application in an alkalization assay when tobacco plants are engineered to overexpress the systemin receptor kinase, suggesting that downstream signaling mechanisms are conserved (Scheer et al. 2003). Therefore, these newly identified peptides could be classified as a small protein family (even though they are coded by only one transcript) that are functionally related to the original systemin, and that may also bind to a receptor kinase.

**Physiological modulators**

**Plant atrial natriuretic peptide (pANP)**

In mammals, the natriuretic peptide (NP) was originally discovered in an extract of rat atria in 1981 (de Bold et al. 1981). The atrial natriuretic peptide is encoded by a 152-amino-acid precursor protein, the propreANP. Mature ANP is obtained after removal of the signal peptide (proANP) and cleavage of the propeptide at position 98 and deletion of the two carboxy-terminal amino acids, resulting in the C-terminal ANP peptide (99–126), which is then circularized through the formation of a disulfide bond between amino acids 7 and 23. It should be mentioned that the N-terminal peptide is also processed into three distinct peptides, proANP 1–30, proANP 31–67, and proANP 79–98, which share similar properties with regards to their capacity to cause vasodilation and increase cellular cyclic guanosine 3’,5’-monophosphate (cGMP) (Vesely et al. 1987).

The possible presence of ANP in planta was first assessed by Vesely and Giordano (1991). They used a radioimmunoassay, in which the antibodies raised against animal ANP (amino acids 99–126), proANP (amino acids 1–30) and the mid-portion of the N-terminus (proANP 31–67) were used to detect the presence of these peptides in the monocot plant *Dracena godseffiana*. All three antibodies detected a peptide, and gel permeation chromatography revealed that the recognized peptides were at the expected molecular weight.

In a follow-up paper, Vesely and collaborators showed that rat proANP 1–30, proANP 31–67, and proANP 79–98, but not the ANP (99–126), could increase solute flow, solute uptake and transpiration (Vesely et al. 1993). Why ANP did not produce the same physiological changes as the N-terminal peptides still remains unclear. The authors also showed the immunodetection of ANP, proANP 1–30, and proANP 31–67 from a wide range of very diversified plant species, including mosses, ferns, gymnosperms, and a protist (*Euglena*).

Transpiration rate and gas exchange in plant leaves are controlled via stomatal opening and closing. The state of the stomatal aperture is itself dependent on the turgor pressure of the two guard cells that surround the stomata, which is regulated by water and solute fluxes. Gehring et al. (1996) took advantage of this turgor-based mechanism to demonstrate that rat ANP (rANP) could cause stomatal opening in plants (i.e., that rANP can cause solute movement in cells) (Gehring et al. 1996). They demonstrated that rANP could cause stomatal opening at micromolar concentration in three different plant species. It was later shown (Pharmawati et al. 1998) that stomatal opening is dependent on two factors: (1) the rANP structure, since the formation of disulfide bridges is necessary, and (2) rANP signaling occurs through a cGMP mechanism (shown using cGMP signaling inhibitors), as previously demonstrated in animals.

Plant ANPs (PNPs) were found by Billington et al. (1997) using immunoaffinity chromatography to purify ANP antiserum-specific epitope from a plant extract, followed by size exclusion separation of the eluate (Billington et al. 1997). The resultant fractions were assayed for their capacity to induce stomatal opening. This method allowed the authors to find functional immunoreactive plant ANPs (irPNPs), which were shown to modulate the ATP-dependent proton gradient in vesicles (proton gradient is used as the driving force for K+ uptake, which leads to increased cell turgor) (Maryani et al. 2000), cell turgor in *Solanium tuberosum* mesophyll cell protoplast (Maryani et al. 2001), and cGMP levels in a manner similar to that observed previously for rANP (Pharmawati et al. 1998; Pharmawati et al. 1999).

Using N- and C-terminal partial protein sequences obtained from immunoaffinity-purified plant natriuretic peptide immunonanalogues (irPNPs) from *Solanium tuberosum*, (Lu-didi et al. 2002) successfully identified and eventually isolated a plant gene encoding irPNP. The AtPNP-A gene encodes a small 126 amino acids protein with an N-terminal signal peptide that had not been previously annotated in the databases. AtPNP displays weak overall sequence identity with human ANP, although significant similarity (50%; and 25% identity) is found between residues 33–66 of AtPNP and the mature human ANP. This section of the peptide (33–66) is also sufficient to confer solute uptake by protoplast (Morse et al. 2004). The AtPNP-A gene shares weak sequence similarities with the expansins, although it lacks a carbohydrate-binding domain. Purified recombinant full-
length AtPNP-A protein, full AtPNP-A without the signal peptide, and AtPNP-A (33–66) could all induce stomatal opening, and recombinant AtPNP-A without the signal peptide causes ion (K+, Na+ and H+) fluxes in A. thaliana roots (Morse et al. 2004).

The data presented above strongly suggest that PNP act as a plant signaling peptide. PNP, like several plant signaling peptide (see Table 1), is encoded as a prepropeptide containing a signal peptide; it can act at nanomolar concentrations (Morse et al. 2004), and is present in xylem exudates (Mar-yani et al. 2003), which are associated with transport rather than protein synthesis. It is also possible that PNP undergoes additional processing since amino acids 33 to 66 are sufficient to induce stomatal opening (Morse et al. 2004). It should be noted that a PNP-like sequence is also found in the bacterium Xanthomonas axonopodis (Nembaware et al. 2004). The activity of the encoded protein has not been demonstrated yet, but such a case of apparent molecular mimicry would not be without precedent in the plant signaling peptides (Olsen and Skriver 2003). NPs would represent the first known conserved hormonal system that is shared between animals and plants. The sequence similarity and homology in the signaling mechanism strongly support that this mechanism, involved in homeostasis regulation, may have evolved before the last common ancestor of the two kingdoms, 1.6 billions years ago (Meyerowitz 1999).

**ENOD40**

Two independent groups discovered ENOD40 nearly at the same time. Yang et al. (1993) found two soybean clones (GmENOD40-1, GmENOD40-2) through a cDNA library screen of root and nodule cDNAs, and the ENOD40 clones yielded a nodule-specific signal. Kouchi and Hata (1993) also found a soybean clone, which is identical to ENOD40-2. In both cases, expression of the gene was induced in nodules after infection of the legume host with a Rhizobium spp., and was therefore named ENOD. In situ hybridization with ENOD40 transcript revealed that following bacterial inoculation of the roots, ENOD40 transcript could be detected in the nodule (Kouchi and Hata 1993; Yang et al. 1993; Asad et al. 1994; Crespi et al. 1994; Matvienko et al. 1994), as well as in other tissues (Asad et al. 1994; Fang and Hirsch 1998). ENOD40 has now been reported (in papers or GenBank) in at least 18 species, including legumes and non-legumes, monocots, and dicots. The patterns of expression, as well as the discovery of ENOD40 in non-legume species and in monocot, confirms that ENOD40 is not only involved in symbiotic interaction and nodule development, but also in a more general developmental process conserved throughout the plant kingdom.

Since ENOD40 was discovered, researchers have debated whether it is active as a peptide or as a RNA molecule. It is now accepted that the ENOD40 transcript, which is about 700 bp long, encodes very short ORFs. Region 1, which is located in the 5’ terminus of the transcript, encodes a short and highly conserved oligopeptide of 12–13 amino acids in legumes and of 10 amino acids in non-legume (van de Sande et al. 1996). The consensus sequence of the region I-encoded peptide is W-X$_4$-HGS. Region 2 is highly conserved at the nucleotide level, but lacks a discernible ORF or coding capacity in monocots (Kouchi et al. 1999; Larsen 2003) and is therefore not believed to be biologically active.

Solid evidences regarding the function of ENOD40 as a peptide have only recently started to accumulate. Rohrig et al. (2002) performed affinity purification of nodule extracts using biotinylated ENOD40 (ORF1) and identified a 93 kDa protein, that after MALDI-TOF analysis, was determined to be nodulin 100, a subunit of the enzyme sucrose synthase, one of the most abundant nodule proteins (Rohrig et al. 2002). A point mutation in the synthetic peptide abolishes binding to nodulin 100, thereby confirming the binding specificity. Hardin et al. (2003) demonstrated that sucrose synthase serine 170 could be phosphorylated both in vitro and in vivo by CDPK II. Knowing that the ENOD40 peptide could bind to sucrose synthase, the authors tested whether the presence of the ENOD40 peptide affects S170 phosphorylation by CDPK II. The dodecapeptide encoded by region I could fully antagonize (97%) S170 phosphorylation (Hardin et al. 2003). When sucrose synthase is phosphorylated at position 170 (p170-SUS), it is found in the same fraction as the proteasome (unlike the non-phosphorylated form), suggesting that this phosphorylation event could target sucrose synthase to the proteasome-mediated degradation pathway (Hardin et al. 2003). This antagonistic reaction mediated by the ENOD40 peptide against the phosphorylation event on S170 of sucrose synthase is mediated via disulfide bond formation between cysteine 4 of ENOD40 peptide and cysteine 264 of the sucrose synthase subunit (Rohrig et al. 2004). The authors also showed that the cleavage activity, but not the synthesis activity, of sucrose synthase increases following ENOD40 covalent binding. Finally, the last piece of evidence that supports the role of ENOD40 as a peptide comes from a combination of transgenic experiments and the use of a synthetic peptide. Transgenic plants of Arabidopsis over-expressing ENOD40 and protoplasts transiently overexpressing ENOD40 have a subtle, but nonetheless statistically significant, smaller cell size when measured by flow cytometry; the same phenomenon is observed when non-transformed protoplasts are treated with an ENOD40 synthetic peptide (Guzzo et al. 2005).

Sucrose synthase, which is a very abundant enzyme in nodules, could be modified by ENOD40 to catalyze the conversion of sucrose into glucose (and fructose), thereby providing the energy and carbon-based molecule required by the bacterioid in the nodule and making ENOD40 a key regulator of sink strength (Rohrig et al. 2002). The expression of ENOD40 in other tissues could still be associated with sucrose synthase since sucrose is also used as a precursor of cellulose synthesis. Notably, ENOD40 expression is observed in rapidly growing (and dividing) tissues such as meristems, fruits and ovules (Asad et al. 1994; Fang and Hirsch 1998; Flemetakis et al. 2000), in which the balance between cellulose synthesis and sucrose degradation must be tightly regulated. Consistent with the above mentioned results, we have also observed that ENOD40 mRNA levels are strongly induced following fertilization and in young developing Solanum chacoense fruits (H. Germain and D.P. Matton, unpublished results). Based upon the data mentioned above, it seems that ENOD40 can act as a peptide (although a possible action as a RNA molecule cannot be ruled out), but there is a lack of evidence with regard to its
Peptides with no known function

The large NCR gene family

The NCR (nodule cysteine-rich) gene family represents (to our knowledge) the largest family of small putative signaling peptides. Members of the family were discovered as early nodulin (ENOD) genes in Pisum sativum more than 15 years ago and were believed to be putative metal-binding protein (Scheres et al. 1990; Kardailsky et al. 1993). These genes were expressed during the development of indeterminate nodule. Members of the family were later discovered as genes expressed during nodule development through EST sequencing (Gyorgyey et al. 2000). Bioinformatic searches revealed that the family was much larger than previously believed when more than 300 cysteine cluster proteins (CPP) were identified (Fedorova et al. 2002; Mergaert et al. 2003), several of which were part of the NCR family. All members have mRNA levels that are significantly increased in developing nodules, encode small peptides (approx. 70 amino acids), have a signal peptide at the N-terminal of the protein, possess a conserved cysteine motif. All members are found in species forming indeterminate nodules and none are found in Arabidopsis or other legume species outside the galegoid group. Sequence identity between the family members is quite variable, ranging from 20% to 70%, but intriguingly, the signal peptide is highly conserved even between distantly related members and for at least two members, the peptide is targeted to the extracellular space. NCR genes are not only abundant at the genomic level, but are also highly expressed and estimated to represent 4.6% of the total nodule transcriptome. Macroarray analyses revealed that two genes, a calmodulin-like protein (containing a signal peptide) and a signal peptide peptidase are co-expressed with NCRs. In animal cells, signal peptide peptidase catalyzes intramembrane proteolysis of some signal peptides after they have been cleaved from the preprotein (Weihofen et al. 2002).

Recently, members of the family were also found in a non-galegoid species (soybean) using other BLAST algorithms (Graham et al. 2004), and the family was even extended further to a large variety of plant species using successive iteration of the hidden Markov model algorithm (Silverstein et al. 2005) (referred to as defensin-like genes). The extended family refers to short proteins with a peptide signal and conserved cysteine motifs. This rather large description also includes LCR (low-molecular-weight cysteine-rich) and SCRL (S-focus cysteine-rich-like) proteins (Vanoos- thuyse et al. 2001), which were identified as homologs of pollen coat proteins (Doughty et al. 2000), SCR/SP11 (Watanabe et al. 2000), and S-protein homologs (Ride et al. 1999).

Until now no functions have been ascribed to the nodule-specific NCR gene family or to any of its members (except SP11/SCR and their alleles). Given the fact that there are hundreds of members in the family, some of which display high homology among themselves, it is unlikely that phenotype searches through knockout, knock-down or overexpressing mutant lines will be successful. As suggested by Mergaert et al. (2003), it is possible, given the expression pattern and the sequence variability of NCRs (and others) that they may act as a wide-spectrum antimicrobial cocktail or that they could be involved in the complex dialog between plants and bacteria, which is triggered during the course of nodule formation or other processes.

4 kDa peptide (leginsulin)

Barbashov and Egorov (1990) found that insulin from pig kidney could inhibit DNA synthesis in human cells, and identified a soybean protein capable of competing with insulin and insulin growth factor-I (IGF-1) for binding to a plasma membrane protein (mediating this inhibition of DNA synthesis); the soybean protein immunoreacted with the anti-insulin antibody, suggesting that plants would also express insulin-like proteins. A soybean protein that binds to insulin and insulin-like growth factor was then identified and was shown to have orthologs with similar binding activity in different species, including cowpea, mung bean, azuki bean and lupin (Komatsu and Hirano 1991).

This protein, the basic 7S globulin (Bg), is a cysteine-rich glycoprotein composed of two subunits (16 and 27 kDa) (Kagawa and Hirano 1989), which is released when soybean seeds are immersed in hot water (40–60 °C) and is believed to have tyrosine kinase activity (Komatsu et al. 1994), although no domains characteristic of protein kinases are found in this protein. Since the tyrosine kinase activity detected was measured in a crude extract preparation, it is strongly suggested that activity is derived from a contaminating or co-migrating protein. Using affinity chromatography with immobilized Bg as a bait, Watanabe et al. (1994) eluted and identified by mass spectrometry a 4 kDa peptide (which was originally named leginsulin and later renamed 4 kDa peptide) from a soybean radicle extract. The binding between the 4 kDa peptide and Bg is inhibited with 0.1 nmol/L insulin, confirming that the 4 kDa peptide and insulin are structural analogs. Edman degradation revealed that the 4 kDa peptide contains 37 amino acids, six of which are cysteine residues, expected (by mass spectrometry) to form three disulfide bonds (Watanabe et al. 1994). The 357-bp ORF, which was identified from a soybean cDNA library, encodes a 119-amino-acid prepropeptide with a signal peptide and a predicted cleavage site between amino acids 19 and 20 (IEA-AD) (H. Germain, personal observation, SignalP 3.0), while amino acids 25–56 encode the 4 kDa peptide (Watanabe et al. 1994). The native peptide, but not the oxidized form, has a strong stimulatory effect on Bg phosphorylation activity (Watanabe et al. 1994), supporting the importance of disulfide bonds in the integrity of the protein.

At physiological levels the 4 kDa peptide is thought to have a role in cellular proliferation. Transgenic callus over-expressing the 4 kDa peptide divide much more rapidly than wild-type. In a similar manner, when the peptide is added to carrot cell suspension cultures, they divide faster than wild-type suspensions in a concentration-dependent manner (Ya-
<table>
<thead>
<tr>
<th>Name of peptide</th>
<th>Receptor if known</th>
<th>Precursor size (in amino acid)</th>
<th>Final size (in amino acid)</th>
<th>Signal peptide</th>
<th>Post-translationnal modification</th>
<th>Expression</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAVATA3</td>
<td>CLAVATA1</td>
<td>96</td>
<td>76</td>
<td>Yes</td>
<td>No</td>
<td>Specific to layer 1 and 2 of shoot meristem central zone</td>
<td>Apical meristem maintenance</td>
</tr>
<tr>
<td>Systemin</td>
<td>SR160</td>
<td>200</td>
<td>18</td>
<td>No</td>
<td>No</td>
<td>Parenchyma cells of vascular bundles</td>
<td>Induced by wounding</td>
</tr>
<tr>
<td>Phytosulfokine</td>
<td>PSKR</td>
<td>80</td>
<td>5</td>
<td>No</td>
<td>Yes, sulfated</td>
<td>Ubiquitous</td>
<td>Cellular dedifferentiation and proliferation</td>
</tr>
<tr>
<td>SCR/SP11</td>
<td>SRK</td>
<td>83</td>
<td>59</td>
<td>Yes</td>
<td>Disulfide bridges possible</td>
<td>Anther, pollen coat</td>
<td>Pollen determinant of self-incompatibility</td>
</tr>
<tr>
<td>BRICK</td>
<td>Not identified</td>
<td>62</td>
<td>62</td>
<td>No</td>
<td>No</td>
<td>Throughout the plant</td>
<td>Formation of epidermal cell lobe and actin patches involved in subsidiary mother cell polarization</td>
</tr>
<tr>
<td>CLE/ESR</td>
<td>Not identified</td>
<td>75–125, average 95</td>
<td>approx. 70</td>
<td>Yes (most)</td>
<td>Unknown</td>
<td>The various family members cover all plant tissues</td>
<td>The various family members cover all plant tissues but individual show specificity</td>
</tr>
<tr>
<td>RALF</td>
<td>Not identified</td>
<td>115</td>
<td>49</td>
<td>Yes</td>
<td>N-terminal processing, disulfide bridges</td>
<td>The various family members cover all plant tissues</td>
<td>Stop root growth, induces alcalinization of the media, C-terminus highly conserved</td>
</tr>
<tr>
<td>DEVIL/ROTUNDIFO-LIA</td>
<td>Not identified</td>
<td>51</td>
<td>51</td>
<td>No</td>
<td>No</td>
<td>Most member ubiquitous</td>
<td>Pleiotropic, RTF domain in C-terminus is highly conserved</td>
</tr>
<tr>
<td>POLARIS</td>
<td>Not identified</td>
<td>36</td>
<td>36</td>
<td>No</td>
<td>Unknown</td>
<td>Root tip and basal part of the plant and leaf vasculature</td>
<td>Knockout has short roots, overexpressing line has more complex leaf venation</td>
</tr>
<tr>
<td>IDA</td>
<td>Not identified</td>
<td>77</td>
<td>50</td>
<td>Yes</td>
<td>No</td>
<td>IDA is flower specific to abscission zone. IDA-like showed specificity to other tissue</td>
<td>High pl, C-terminal PIP motif highly conserved</td>
</tr>
<tr>
<td>Egg apparatus 1</td>
<td>Not identified</td>
<td>94</td>
<td>94</td>
<td>No</td>
<td>No</td>
<td>Exclusively in the maize egg apparatus</td>
<td>Has a putative transmembrane domain, C-terminus highly conserved</td>
</tr>
<tr>
<td>HT</td>
<td>Not identified</td>
<td>100</td>
<td>75</td>
<td>Yes</td>
<td>Possible N-terminal processing and glycosylation</td>
<td>Style specific</td>
<td>Self-incompatibility modifier</td>
</tr>
<tr>
<td>TobSys</td>
<td>Not identified</td>
<td>165</td>
<td>18</td>
<td>Yes</td>
<td>Glycosylated, hydroxylated</td>
<td>Cell wall matrix of vascular parenchyma cells</td>
<td>Encoded on a single mRNA, effects similar to systemin</td>
</tr>
<tr>
<td>TomSys</td>
<td>Not identified</td>
<td>146</td>
<td>20, 18, 15</td>
<td>Yes</td>
<td>Yes</td>
<td>Peptide isolated in xylem exudate</td>
<td>Regulate solute flow</td>
</tr>
<tr>
<td>Plant natriuretic peptide</td>
<td>Not identified</td>
<td>126</td>
<td>27</td>
<td>Yes</td>
<td>Disulfide bridges possible</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
mazaki et al. 2003). Hanada et al. (2003) used size exclusion chromatography and surface plasmon resonance, combined with alanine-scanning mutagenesis, to investigate the binding mechanism between the 4 kDa peptide and Bg. Their finding revealed that Bg purifies and binds to the 4 kDa peptide as a 80 kDa dimer. The surface plasmon resonance experiment showed that the $K_d$ of the interaction between wild-type 4 kDa peptide and Bg was $1.86 \times 10^{-8}$ mol/L (or 18.6 nmol/L) (Hanada et al. 2003). Alanine scanning revealed that a mutation in the C-terminus of the peptide has a higher impact on the binding interaction than a mutation in the N-terminus, suggesting that this region of the hormone-like peptide is more important for the protein–protein interaction. It was later shown that Bg has two sites in the 27 kDa subunit that interacts with the 4 kDa peptide and that the same region of the Bg protein is recognized by the animal insulin (Hanada and Hirano 2004). A Bg ortholog from carrot is glycosylated at four sites and the interaction of this protein with the 4 kDa peptide is 20-fold higher when the protein is deglycosylated (Shang et al. 2005). Furthermore, these glycosylations are necessary for the secretion of the 43 kDa protein (the 16 and 27 kDa subunits). It is therefore possible that glycosylation could facilitate the proper folding of the protein and must later be removed once the protein as been secreted (Shang et al. 2005).

The 4 kDa peptide contains a signal peptide, is synthesized from a larger precursor, and has pairs of disulfide bonds, features that make it a good candidate to act as a small putative signaling peptide. On the other hand, the Bg protein has no known downstream effectors, and no clear biological activities have been reported for these 7S globulin proteins, except as seed storage proteins, suggesting that its earlier classification as a receptor-like protein by Watanabe et al. (1994) is questionable. Functional analyses in transgenic plants overexpressing or impaired in the expression of the 4 kDa peptide should provide some details about its role in plants.

**Conclusion**

The known putative signaling peptides already add up to a few hundred peptides and surely many more will be discovered. It was initially thought that each receptor kinase would bind one ligand. Since $CLE40$ can act as a signal in the SAM through $CLV1$, and $CLV3$, $CLE19$ and $CLE40$ all cause root meristem consumption (Fiers et al. 2004), and the solanaceous systemin receptor can also bind brassinosteroids (Scheer et al. 2003), there are doubts about this one-to-one hypothesis. This, combined with the conspicuous observation that the members of several peptide families (RALF, IDA, CLE, DVL/ROT, HT, Egg apparatus) have highly conserved C-terminal domains, shared by all the members within a family, suggest that related peptides within a family that have most probably arisen through gene conversion and (or) duplication events, may have acquired spatial and temporal specificity and interact with a more limited number of receptors.

The field of peptide signaling and the number of putative signaling peptides in plants has seen tremendous growth over the last decade. It has gone from its birth, with the discovery of systemin, to the point where several orphan puta-
tive ligands have now been found either through biochemical, genetical or database mining. Improved prediction algorithms should allow more peptides to be found from the available databases, but success will be hampered by the fact that short cDNAs are still (for good reasons) being discarded during library construction and often end-up being underrepresented. Another bioinformatics approach enabling identification of small peptide or non-coding RNAs is comparison of large EST sets with annotated genomic resources, allowing the retrieval of ESTs that correspond to unannotated genes (Riano-Pachon et al. 2005). New methods such as mass spectrometry of apoplastic fluids or sap have already started to yield fruitful results (Hoffmann-Benning et al. 2002; Boudart et al. 2005) and not only permits the finding of new peptides, but also provides invaluable information with regards to their post-translational modification. In the case where large peptide families are involved, functional characterization will have to rely on the production of double or triple knockout mutants (or even more) or targeted interference of conserved domains. High throughput biochemical approaches involving bioassays or high density protein arrays could speed-up the process of assigning peptides to their binding partners.

Acknowledgements

We thank members of the Matton laboratory for sharing unpublished results. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Canada Research Chair program. H.G. is the recipient of Ph.D. fellowships from NSERC and from Le Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT, Québec). E.C. is the recipient of a Canada Research Chair program.

References


