Role of the Extensin Superfamily in Primary Cell Wall Architecture

ABSTRACT

Nearly two centuries of progress have established the major components of the plant cell wall, a composite that includes interpenetrating networks of cellulose (Payen, 1838; Schulze, 1891), microfibrils (Frey-Wyssling et al., 1948; Preston et al., 1948), pectin (Braconnot, 1825) and lignin (Payen, 1838). However, only over the last five decades has a relatively minor hydroxyproline-rich structural glycoprotein component emerged with essential roles in building and maintaining the growing primary cell wall. Here we highlight unique advances of each decade from the initial discovery of hydroxyproline (Hyp) in cell walls to the current definition of extensins as self-assembling amphiphiles that generate scaffolding networks, and where acid-base interaction – extensin pectate – may template assembly of the pectic matrix. Subsequent polymerization toughens up the wall as networks resisting both microbial and mechanical stress. At each stage we explore hypotheses arising from synthesis of emerging data with focus on structure. This review celebrates the 50th birthday of extensin.

INTRODUCTION

Protein interactions direct life processes at all levels ranging from the regulation of metabolism and nucleic acid replication to signal transduction and morphogenesis. Sophisticated extracellular matrices like those of animals and plants comprise scaffolding networks of glycoproteins and proteoglycans that interpenetrate other networks of structural polysaccharides. While polysaccharide networks are prominent in the plant extracellular matrix, glycoproteins and proteoglycans dominate the animal matrix. By weight, protein, which is largely structural, contributes up to 20% of the primary wall (Burke et al., 1974; Kieliszewski et al., 1992), yet may be essential as the loss of the structural glycoprotein network in Arabidopsis is lethal (Hall and Cannon, 2002; Cannon et al., 2008).

Since the discovery of cell wall protein in 1960 (Dougall and Shimbayashi, 1960; Lamport and Northcote, 1960b); the hydroxyproline-rich glycoprotein field, referred to generically as HRGPs, has blossomed from having only one family member (extensin) (Lamport,
1963), to three (extensin, arabinogalactan-protein [AGP]) (Yariv et al., 1962; Aspinall et al., 1969; Fincher et al., 1974), and solanaceous lectins (Allen and Neuberger, 1973)), then four (proline-rich proteins, [PRPs]) (Chen and Varner, 1985; Hong et al., 1987; Averyhart-Fullard et al., 1988; Tierney et al., 1988; Wilson et al., 1994)) ...and more. We now realize that this (super)family represents a continuum of peptide periodicity and glycosylation (Kieliszewski and Lamport, 1994); perusal of the Arabidopsis genome suggests that conserved features of HRGPs, in particular their arabinosyl-O-Hyp and arabinogalactosyl-O-Hyp glycomodules, are widespread in secreted proteins ranging from phytocyanins and systemins to fasciclins and glycerophosphodiesterases (Pearce et al., 1991; Borner et al., 2002; Johnson et al., 2003; Kieliszewski et al., 2010; Showalter et al., 2010). Furthermore, highly organized HRGP cell wall networks are ancient as they occur in algae (Thompson and Preston, 1967; Gotelli and Cleland, 1968; Miller et al., 1972; Roberts, 1974; Goodenough and Heuser, 1989).

Here we largely confine ourselves to the classical extensins, which we define as wall-located, basic, Hyp-rich structural glycoproteins with alternating hydrophilic and hydrophobic motifs whose alignment as self-assembling amphiphiles likely drives extensin network assembly. The hydrophilic motifs comprise arabinosylated X-Hyp$_n$ where X is usually Ser, and n most often 4 and occasionally up to 5 (Campargue et al., 1998) or six (Qi et al., 1995), while the hydrophobic ‘motifs’ vary, sometimes being as small as a single amino acid or dipeptide and often containing tyrosine residues as potential crosslink sites.

**THE EXTENSIN SUPERFAMILY OVER FIVE DECADES**

The findings and technologies used are presented in a historical context, building up stepwise a picture of what we know about extensins, with focus on the big question: what role does extensin and a protein network play in cell walls?

**The Sixties: Cell wall protein discovery, Hyp-O-arabinosylation and evidence for soluble extensin precursors to the wall network.**

In 1960 the primary cell wall isolated from sycamore cell suspension cultures and tobacco callus (Lamport and Northcote, 1960a) contained enzymes and a Hyp-rich component (Lamport and Northcote, 1960b); it could therefore be considered as a “cell particle (or
organelle) possessing structural integrity and enzymic autonomy” (Lamport, 1964). Hyp indicated a structural protein by analogy with animal collagen where this cyclic amino acid constrains side chain rotation and yields an extended structural protein. Sycamore cells grown in $^{18}$O$_2$ showed that molecular oxygen was the direct source of the hydroxyl oxygen (Lamport, 1963) significant because the Hyp hydroxyl plays a pivotal role as a carbohydrate attachment site (Lamport, 1967) and Hyp-rich glycopeptides isolated from enzymatic digests of tomato cell walls indicated a highly glycosylated glycoprotein (Lamport, 1969). Prior to the discovery of the ER/Golgi role in protein secretion (Jamieson and Palade, 1967) early pulse-chase experiments suggested that extensin destined for the wall of sycamore cells occurred as a soluble cytoplasmic precursor (Lamport, 1965). Later pulse-chase experiments showed that macromolecular Hyp appeared in ‘membranous organelles’ before secretion to the wall (Chrispeels, 1969).

A hypothetical role in cell extension based on the structure and location of Hyp-rich glycoprotein in the primary cell wall suggested the name ‘extensin’ (Lamport, 1963). Identification of Hyp in the walls of many algae (Gotelli and Cleland, 1968) supported the hypothesis that extensins are wide-spread and may play a role in cell expansion (Thompson and Preston, 1967).

The Seventies: The Ser-Hyp$_4$ diagnostic motif, glycopeptide linkages, and links to disease responses

Facile cleavage of acid-labile (pH 1) arabinofuranoside linkages and subsequent tryptic degradation released significant amounts of Hyp-rich material from the wall (Lamport, 1974) but only a very few major peptides each containing the diagnostic Ser-Hyp$_4$ signature sequence. This was the first suggestion that extensin is a highly periodic protein, subsequently corroborated by the circular dichroism (CD) spectra of both crude extensin and extensin peptides indicating an extended left-handed polyproline-II (PP-II) helix (Lamport, 1977). The intractability of the presumed extensin network provided the impetus for developing new tools, particularly hydrogen fluoride (HF)-solvolysis deglycosylation of glycoproteins in an attempt to solubilize wall-bound extensin (Mort and Lamport, 1977) from putative glycan crosslinks (Keegstra et al., 1973). However, extensin remained insoluble indicating protein-protein crosslinking rather than protein-glycan crosslinking and confirming earlier work (Lamport, 1965).
As the structure of extensin was being determined their involvement in disease and wound responses became apparent when Esquerre-Tugay and colleagues showed that pathogens induced extensin accumulation and was correlated with disease resistance (Esquerre-Tugaye and Mazau, 1974; Esquerre-Tugaye and Lamport, 1979), while Chrispeels and colleagues showed that physical wounding induced extensin biosynthesis (Chrispeels et al., 1974). A general role for extensin in response to different stresses including senescence and abscission was corroborated and detailed (Merkouropoulos and Shirsat, 2003).

The Eighties: Extensin monomers, *in vitro* crosslinkage and *EXT* genes

The quest for salt-extractable monomeric precursors to network extensin began in the early 60s but low yields from sycamore cell suspensions (Lamport, 1965) and carrot disks (Brysk and Chrispeels, 1972; Stuart and Varner, 1980) impeded progress. Finally, substrate quantities of extensin monomers salt-eluted from intact cells of rapidly growing tomato cell suspension cultures (Smith et al., 1984) allowed detailed characterization that confirmed the remarkable periodicity of the Ser-Hyp$_4$ glycomotif and also the precursor-product relationship between monomeric extensin and the insoluble wall network. Salt-elution also implied ionic interaction between the extensin and pectin networks (Smith et al., 1984; Smith et al., 1986; Qi et al., 1995; Nunaez et al., 2009). Possession of a substantial monomeric pool enabled *in vitro* crosslinking experiments.

Fry’s discovery of the crosslink amino acid isodityrosine (Idt) in cell wall hydrolysates (Fry, 1982) sparked speculation that Idt was the intermolecular crosslink and key to extensin network insolubilization (Fry, 1982; Lamport and Epstein, 1983). However, the insoluble extensin wall network yielded tryptic peptides that contained Idt only as a very short *intra*molecular crosslink in a highly conserved hydrophobic motif, Tyr-Xaa-Tyr (Epstein and Lamport, 1984). Nevertheless, the idea of Idt intermolecular crosslinks persisted, fueled by further evidence of *in muro* crosslinking (Cooper et al., 1987; Bradley et al., 1992). In particular Bradley et al (1992) showed that fungal elicitation of hydrogen peroxide corresponded to a rapid wall-hardening process involving a decrease in extractable extensin, emphasizing the significance of Esquerre-Tugaye’s earlier work (Esquerre-Tugaye and Mazau, 1974; Esquerre-Tugaye and Lamport, 1979), and the highly specific pI 4.6 extensin peroxidase that catalyzed *in*
extensin crosslinking (Everdeen et al., 1988; Lamport, 1989). Evidence of other extensin peroxidases appeared later (Price et al., 2003).

Finally, the diagnostic Ser-Hyp₄ peptide (Smith et al., 1986) enabled identification of the first extensin (Chen and Varner, 1985) and PRP cDNAs (Hong et al., 1987; Tierney et al., 1988; Datta et al., 1989), as a bona fide protein with the hallmark of other structural proteins, most notably collagen, which is also Hyp-rich and the major structural fibrillar protein of animals. Collagen polypeptides occur in an extended PP-II left-handed helical conformation, which was also confirmed in carrot extensin by further CD spectra (van Holst and Varner, 1984), with evidence for the role of carbohydrate in maintaining the backbone conformation (Stafstrom and Staehelin, 1986).

The Nineties: Phylogeny, glycosylation codes, crosslinking codes and synthetic genes.

Evolution conserves functional motifs. Peptide sequence motifs from gymnosperms (Fong et al., 1992; Kieliszewski et al., 1992) and dicot extensins (Smith et al., 1986; Li et al., 1990; Memelink et al., 1993), made a comparison with other advanced angiosperm groups of great interest, particularly those with a radically different growth habit like the grasses. A threonine-rich HRGP (THRGP) from maize (Kieliszewski and Lamport, 1987; Hood et al., 1988; Stiefel et al., 1988) was clearly related to dicot extensins and suggested that HRGP conserved sequence encodes both proline hydroxylation (Kieliszewski et al., 1990) and Hyp glycosylation. Another HRGP from maize contained both extensin and AGP peptide motifs and led to the formulation of the Hyp Contiguity Hypothesis: "...perhaps sequences around noncontiguous Hyp direct Hyp-arabinogalactosylation, whereas contiguous Hyp directs arabinosylation" (Kieliszewski et al., 1992). Such codes implied that extensin could readily evolve into an AGP or vice-versa as a single base change relates proline, serine and alanine codons, respectively CCX → UCX → GCX. Thus a single base change transforms contiguous to non-contiguous Hyp changing the glycosylation code and explains why members of the extensin superfamily appear as a phylogenetic continuum (Kieliszewski and Lamport, 1994); for example gum arabic glycoprotein possesses both contiguous Hyp (extensin) and non-contiguous Hyp (AGP) motifs (Qi et al., 1991).

Synthetic gene constructs confirmed the Hyp glycosylation code and enabled the design of HRGPs to elucidate post-translational codes and the function of conserved motif repeats.
(Shpak et al., 1999) in the following decade. Meanwhile the significance of putative crosslink motifs, VYK and Idt (YXY) became apparent with the availability of fast performance liquid chromatography (FPLC), notably Superose-6 columns that allowed resolution of extensin monomers, oligomers and polymers following in vitro enzymatic crosslinking of the monomers (Schnabelrauch et al., 1996) and led to the discovery of a specific extensin peroxidase.

Significantly, extensin peroxidase did not crosslink extensins like the maize THRGP (Schnabelrauch et al., 1996) that lacked the putative VYK and YXYK crosslink motifs. Some extensins like tomato P1 apparently lacking Idt motifs (Smith et al., 1984; Smith et al., 1986) were readily crosslinked suggesting VYK as intermolecular crosslink.

However, discovery of the crosslinked tyrosine derivatives di-isodityrosine (di-Idt) and pulcherosine (Brady et al., 1996) resolves the issue (Figure 1): The tetra-tyrosine derivative, di-Idt, can be formed from 2 Idt residues in neighboring molecules while the tri-tyrosine derivative, pulcherosine, can be formed from Idt and Tyr (Fry, 1982; Brady et al., 1996; Brady and Fry, 1997). The apparent absence of Idt motifs in the P1 peptides isolated earlier (Smith et al., 1986) may reflect the restriction of Idt to the C-terminus YVYSSPPPPHY (SGN-U315189). Thus, abundant ‘non-Idt’ Tyr residues in P1 and the two Idt motifs at the C terminus are sufficient for peroxidatic crosslinking to give pulcherosine and some di-Idt. In addition to specific motifs and their abundance, differential localization in the wall may also influence the role of extensins (Swords and Staehelin, 1993).

The Noughties: Form and function.

Glycoproteins smothered in sugar offer technical and conceptual challenges. How can we relate structure to function? In particular, what is the role of O-Hyp glycosubstituents – highly conserved neutral oligoarabinoside glycomodules, typically tri- and tetra-arabinosylated Ser-Hyp$_4$ in extensins (Lamport et al., 1973) and the acidic arabinogalactan polysaccharide glycomodules of AGPs (Tan et al., 2004). Both are hydrophilic but their different structures imply different roles (Tan et al., 2010).

Synthetic gene technology and molecular genetics have yielded insights into the assembly of the extensin network at the molecular level and its role at the biological level: Discovery of the lethal rsh embryogenic Arabidopsis mutant corresponding to AtEXT3 showed that extensins are essential for cell plate formation, evidenced by the aberrant mutant wall
phenotype and AtEXT3 immunocytochemical localization (Hall and Cannon, 2002). At the molecular level purified AtEXT3 extensin monomers visualized by atomic force microscopy (AFM) form dendritic structures. This indicates a propensity for self-assembly driven by the alternating hydrophilic (Ser-Hyp4) and hydrophobic motifs, typically Idt (Epstein and Lamport, 1984), and is consistent with their crosslinkage in vitro by extensin peroxidase (Cannon et al., 2008). Surprisingly, AtEXT3 yielded the tri-tyrosine derivative pulcherosine as the major intermolecular crosslinked product rather than di-Idt formed in the P3-type sequence: Ser-Hyp4-Ser-Hyp-Ser-Hyp4-Tyr-Tyr-Tyr-Lys (Held et al., 2004).

This enigma was resolved by comparison of AtEXT3 (RSH) with the P3 sequences: Alignment of the Idt-forming Tyr-Tyr-Tyr motifs in the P3-type sequence forms di-Idt exclusively, because all Tyr residues align to Idt motifs hence they theoretically form di-Idt. Unlike P3, crosslinked products of AtEXT3 (Ser-Hyp4-Lys-Val-Tyr-Lys-Ser-Hyp4-Val-Lys-His-Tyr-Ser-Hyp-Val-Tyr-His)n can yield di-Idt only when the Tyr-Val-Tyr (Idt) motifs are aligned in register. However, when the AtEXT3 alignment is offset, pulcherosine is the crosslinked product as an Idt motif aligns only with single Tyr residues (His-Tyr-Ser or Val-Tyr-His) in neighboring AtEXT3 molecules (Cannon et al., 2008). Such theoretical offset AtEXT3 alignment is consistent with the in vitro crosslinking results and with the dendritic assemblies of AtEXT3 as imaged by AFM.

Self-assembly is a general feature of extensins judging by AFM imaging of a range of extensins (Figure 2). For example AFM-imaged tomato extensin P1 and the maize THRGP display similar network structures. Although THRGP is not crosslinked by extensin peroxidase (Schnabelrauch et al., 1996) we consider that hydrophilic alternating with hydrophobic motifs align monomers into segments of predetermined lengths to yield dendritic networks where distinctive N- and C-terminal sequences control segment length, and network assembly and may also confer self-sorting properties (van Esch, 2010). Thus hydrophilic arabinosylated Thr-Hyp-Hyp-Thr and Thr-Hyp-Ser-Hyp motifs of THRGP alternate with individual hydrophobic Tyr residues and the arguably hydrophobic Pro-Lys-Pro motifs (von Heijne and Blomberg, 1979; Vila et al., 1998). Disassembly of extensin networks involves cysteine endopeptidases that rapidly degrade P1- but not P3-type extensins (Helm et al., 2008); these KDEL-tailed cysteine endopeptidases are also involved in programmed cell death and intercalation of new cells.
The concept of self-assembling amphiphiles (Rapaport, 2006) very likely applies to heterophilic interactions in the wall. Acid-base interaction between extensin and pectin (Smith et al., 1984) has the potential to yield extensin pectate (Cannon et al., 2008). Thus extensin may template the orderly assembly of pectin in the cell plate, perhaps even involving some covalent extensin-pectin crosslinks (Qi et al., 1995; Nunaee et al., 2009).

Hydrophilic AGPs may also behave as self-assembling amphiphiles that are inserted into the membrane in an orderly fashion (Gens et al., 2000), where they are initially bound by their C-terminal hydrophobic GPI-lipid anchor (Oxley and Bacic, 1999; Svetek et al., 1999; Borner et al., 2002). However, Hyp-arabinogalactans of AGPs likely play a role that differs from neutral Hyp-arabinosides of extensin. Acidic Hyp-AGs (Lamport et al., 2006) cover the plasma membrane and could potentially chelate calcium (Tan et al., 2010), thus allowing AGPs to act as a calcium reservoir. We know that phospholipase C releases membrane-bound AGPs as soluble periplasmic AGPs that are then incorporated into the growing wall as putative pectic plasticizers (Lamport, 2001; Lamport et al., 2006). We also know that the β-D-glucosyl Yariv reagent inhibits expansion growth (Jauh and Lord, 1996), and that this reagent strongly associates with AGPs, potentially negating the role of AGPs as a plasticizer in muro and thus inhibiting expansion growth. Taking what we know about extensins as crosslinkers and AGPs as putative plasticizers, speculatively one might view extensins and AGPs as the ‘Yin & Yang’ of cell extension – negative and positive regulators, respectively.

**QUESTIONS AND HYPOTHESES FOR FURTHER STUDY**

*Why are some extensins not covalently crosslinked by extensin peroxidase* (Kieliszewski and Lamport, 1994; Schnabelrauch et al., 1996)? This raises the deeper problem of their biological significance, if any. *Do they have a role that does not require them to be crosslinked or are they residual features of evolution*? The first extensins, typified by Chlamydomonas formed a weak non-covalent wall lattice consisting entirely of Hyp-rich glycoproteins notably non-crosslinking (NCL) and secreted via Golgi vesicles to the cell surface. Cytokinesis in these early protists involves a cleavage furrow that precedes the cell plate division mechanism of later chlorophytes where intracellular fusion of Golgi vesicles with their Hyp-rich crosslinking extensin (CL) cargo form a glycoprotein network or scaffold strengthened by further addition of...
cellulosic and pectic networks. Hence an extensin network with covalent crosslinks was a
delayed development and led to strong walls that enabled terrestrial colonization and turgor-
driven growth. NCL extensins are still represented in land plants such as the maize THRGP (not
crosslinked by extensin peroxidase (Schnabelrauch et al., 1996) and tobacco extensin HRPnt3
expressed during initiation of lateral root meristems (Keller and Lamb, 1989; Vera et al., 1994).
Both are amphiphiles and basic, therefore with the potential for self-assembly (Figure 2) of non-
covalent networks. Regarding a role for NCL extensins, hypothetically, they may organize self-
assembly of the cell plate while CL extensins increase the tensile strength of the primary cell
wall.

What is the role of P3 type extensin? The occurrence of P3 type extensin in the first
vascular plants implies a possible role for these extensins in vascular development consistent
with the earlier suggestion of P3 as a ‘brace’ protein (Smith et al., 1986). It is interesting that CL
P3-type extensins can now be traced to both branches of the first vascular plants, Lycophytes and
ferns, notably in Selaginella (an extensin of 319 residues, accession XP 002961156) and a 210-
residue extensin in the maidenhair fern, Adiantum (Uchida et al., 1998). Both show conservation
of the repetitive 16-residue motif SPPPSPPPPPPPYPXYK first identified in tryptic peptides
isolated from tomato cell walls, as extensin P3 (Smith et al., 1986); accession CAA39215.1
(Showalter et al., 1991). Idt motifs of a P3 extensin analog are crosslinked in vitro by extensin
peroxidase to form the tetrameric tyrosine derivative di-Idt (Held et al., 2004).

Why are some walls Hyp-poor? Tissue differences exist in all species, however some
plants possess relatively little Hyp regardless of the tissue. Again a simple answer suggests that
some evolutionary lines, notably grasses have founded mechanical support systems largely
involving non-HRGP structural proteins (Kieliszewski and Lamport, 1987; Kieliszewski et al.,
1990) but perhaps retain the primary role of extensin as a self-assembling amphiphile with a
templating role at cytokinesis.

Why are there so many extensins and related HRGP hybrids and chimeras? Arabidopsis
extensins are the best characterized genomically with up to 63 potential extensins of which 20
are very likely extensins (Cannon et al., 2008), 12 are shorter potential extensins, and 31 are
extensin chimeras and hybrid extensins (Showalter et al., 2010). Multiple extensins point to either multiple functions or similar functions with differential extensin expression.

Judging from their tissue specific expression, it is argued that extensins and proline-rich proteins are ‘tailored to the tissue’ during embryogenesis (Zhang et al., 2008) and throughout development (Fowler et al., 1999); this includes root hair formation (Bucher et al., 2002) which also involves a LRX1, a chimera of extensin and leucine-rich repeat protein (Baumberger et al., 2001). Furthermore the identification of gene regulatory elements (Guzzardi et al., 2004) is consistent with the suggestion that: “selective activation of genes encoding specific structural proteins provides a mechanism for precise morphogenetic control of cell wall architecture during cellular differentiation.” (Keller and Lamb, 1989). Given the known role(s) of extensins in cell wall assembly, cell shape and size, disease resistance and quite possibly in reproductive isolation and speciation (Lee et al., 2007) raises the question of mechanism for each extensin molecule. Stress induces specific extensins (Merkouropoulos et al., 1999). There is a correlation between extensin expression and walls that withstand tensile stress such as hypocotyls (Shirsat et al., 1996), seed coats (Cassab et al., 1985), and root hairs that exemplify cells exposed to extremes of osmotic stress – unlike pollen tubes, they possess an abundance of extensin (Bucher et al., 1997). Association of extensins with extension growth is increasingly well-documented (Cleland and Karlsnes, 1967; Sadava and Chrispeels, 1973; Roberts and Shirsat, 2006; Gille et al., 2009).

The evolutionary paradigm shows the versatility of functional adaption: structures initially selected for one role then recruited later to serve another quite different role. While it is beyond the scope of this review to catalog all extensin chimeras and hybrids, potato lectin is a good example of the extensin Ser-Hyp\textsubscript{n} motif recruited to act as a spacer between two lectin domains (Van Damme et al., 2004).

What is the role PRPs that lack obvious Ser-Hyp\textsubscript{4} glycomodule repeats, yet are closely related to the extensins? PRPs are highly basic, minimally arabinosylated and share variations of the Val-Tyr-Lys motif (i.e. Pro-Hyp-Val-Tyr-Lys) (Averyhart-Fullard et al., 1988; Tierney et al., 1988; Iannetta et al., 1993; Bernhardt and Tierney, 2000). Presumably PRPs are covalently crosslinked into the wall network (Bradley et al., 1992; Frueauf et al., 2000) although direct evidence is lacking. Nevertheless PRPs are associated with particular cell types exemplified by
root nodules and stomatal guard cells (Menke et al., 2000). More recent work using mutants atprp2 and atprp4 show that AtPRP2 and AtPRP4 are required for stomatal guard cell function (Carter and Tierney, 2010). Significantly, predictions of secondary structure using COUDES software (Fuchs and Alix, 2005) indicate that the repetitive motifs of AtPRP2 and AtPRP4 form a coil of beta-turns. This suggests an elastic protein contributes to the elasticity of guard cell walls.

**Why are PRPs not glycosylated like AGPs?** The Hyp-contiguity hypothesis predicts contiguous Hyp residues as sites of arabinosylation and clustered non-contiguous Hyp residues as sites of arabinogalactosylation. These predictions readily fit experimental data from extensins and AGPs (Shpak et al., 1999; Shpak et al., 2001; Zhao et al., 2002; Tan et al., 2003) as AGP and extensin analogs can be ‘designed’ to produce predictable glycosylation patterns when expressed in transformed plants. Although tightly clustered non-contiguous Hyp residues (e.g. Ala-Hyp-Ala-Hyp repeats) are reliably arabinogalactosylated, the clustering of the non-contiguous Hyp/Pro residues in AGPs can be fairly loose in that the non-contiguous Hyp residues can be separated by as many as 3 or 4 intervening residues and still be sites of arabinogalactosylation (Zhao et al., 2002).

Some PRPs and maize THRGP also contain clustered non-contiguous Hyp, however, these Hyp residues either remain non-glycosylated or are arabinosylated. For example, Douglas Fir proline-rich HRGP is a PRP that undergoes arabinosylation only on contiguous Hyp residues, while the major peptide repeat motif also contains non-contiguous Hyp/Pro residues that are clustered yet remain non-glycosylated (underlined): Lys-Pro-Hyp-Val-Hyp-Val-Ile-Pro-Pro-Hyp-Val-Val-Lys-Pro-Hyp-Val-Tyr-Lys-Pro-Hyp-Val-Hyp-Val-Ile-Pro-Pro-Hyp-Val-Val-Lys-Pro-Hyp-Val-Tyr-Lys-Ile-Pro-Pro/Hyp-Val-Ile-Lys-Pro. There are other examples: some essentially non-glycosylated PRPs are almost entirely variations of the repeat (Pro-Hyp-Val-Tyr-Lys-Pro-Hyp-Val-Tyr-Lys) that also contains clustered non-contiguous Hyp, with only 4 amino acids separating the Hyp residues also remain non-glycosylated (Marcus et al., 1991); maize THRGP possesses Thr-Hyp-Ser-Hyp repeats that undergo arabinosylation only; some dicot extensins, like tomato P1 also contain loosely clustered non-contiguous Hyp/Pro residues that are never arabinogalactosylated: Ser-Hyp4-Thr-Hyp-Val-Tyr-Lys-Ser-Hyp4-Val-Lys-Pro-Tyr-His-Pro-Thr-Hyp-Val-Tyr-Lys. Why? Earlier work indicates that arabinogalactosylation of clustered
non-contiguous Hyp is influenced by amino acid context: Ala-Hyp and Ser-Hyp repeats were consistently and extensively arabinogalactosylated whereas Thr-Hyp and Val-Hyp repeats were sites of arabinosylation and arabinogalactosylation (Tan et al., 2003). Thus the biased and distinctive amino acid compositions and sequences of extensins and PRPs (rich in Lys and Tyr, low in Ala and Gly) compared to the AGPs (rich in Ala and Gly and low in Tyr and Lys) also hint that sequence environment in addition to the arrangement of Hyp residues influences whether or not a single Hyp residue is arabinosylated, arabinogalactosylated or remains non-glycosylated. Future work will undoubtedly refine the Hyp contiguity code.

**What is the precise role of the highly conserved Ser-Hyp4 glycomodule?** The Hyp-β-L-arabinofuranoside linkage has been conserved since Chlamydomonas (Miller et al., 1972; Bollig et al., 2007). Thus the conservation of the Ser-Hyp4 glycomodule implies a crucial function, yet the module is uncharged and chemically unreactive under physiological conditions. It does have H-bonding potential and presents a unique shape: The Ser residues are α- monogalactosylated, and all Hyp residues are arabinosylated with chains 3-5 residues long. The structures are unusual: In contrast to the all α-3 and 5-linked arabinofuranose and the β-galactopyranose of the AGPs, arabinosides of the Ser-Hyp4 glycomodule are all β-2 linked except the fourth residue (at the non-reducing end), which is α-3 linked (Akiyama et al., 1980). Thus module shape and H-bond presentation creates a unique surface for homophilic and/or heterophilic interactions, perhaps including galactose-binding lectins or lectin modules.

**Are there practical applications for wall protein networks?** A rapid *in muro* response to pathogen attack leads to insolubilization of extensins within minutes (Bradley et al., 1992), mirrored at the genomic level by upregulation of specific extensins containing numerous Idt motifs (Showalter et al., 1991; Zhou et al., 1992; Wycoff et al., 1995). This creates a potentially highly crosslinked defensive network welded together by extensin peroxidase and reactive oxygen species generated by fungal elicitors (Brady and Fry, 1997). Such a clear link between disease resistance and extensin remains to be exploited by molecular pathologists.

**Does proline hydroxylation play a subtle regulatory in addition to a structural role in plant development?** This review concludes by returning to the initial observation that molecular
oxygen is the direct source of the Hyp hydroxyl group and may have a physiological implication: “…and because the $K_m$ (for oxygen) of hydroxylases is much larger than that of cytochrome oxidase it now becomes important to investigate the possibility of a direct effect of oxygen tension…” (Lamport, 1963). Such oxygen sensing by specific PHDs is well-characterized in animals (Appelhoff et al., 2004). The corresponding oxygen sensor of plants remains unknown but plants also contain multiple plant prolyl hydroxylases (Tiainen et al., 2005), therefore we hypothesize their likely role as an oxygen sensor, and thus to the survival of waterlogged crops and marshland flora that enhance gaseous diffusion to roots by constructing aerenchyma (Kende et al., 1998; Jackson and Armstrong, 1999).

ACKNOWLEDGEMENTS

We thank the National Science Foundation for funding (MJK, Award IOS 0622394; MCC, Award IOS 0622428).
**FIGURE LEGENDS**

**Figure 1.** Amino acid structures of tyrosine derivatives. A, the diphenylether, isodityrosine (Idt). B, the tri-tyrosine, pulcherosine. C, the tetra-tyrosine, di-isodityrosine (di-Idt).

**Figure 2.** AFM images of self-assembling extensins. *Top*, Tomato P1 (PI). *Bottom*, Maize Threonine-rich HRGP (THRGP). Proteins were dissolved in ddH2O to a concentration of 10 μg/ml. Sixty μl of the protein solutions were deposited on to a freshly cleaved highly ordered pyrolytic graphite for 5 min then blotted dry. The graphite surface was rinsed with 100 μl of ddH2O and then dried under N2. AFM imaging was carried out on a MFP-3D AFM system (Asylum Research, Santa Barbara, CA) using AC mode in air. The P1 image is a 256×256 pixel scan while the THRGP image is a 512×512 pixel scan. For further details and comparison with AFM images of an Arbidopsis extensin, see (Cannon et al., 2008).
REFERENCES


Braconnot H (1825) Nouvelles observations sur l’acide pectique. Annales de Chimie et de Physique 30: 96


Carter JP, Tierney ML (2010) AtPRP2 and AtPRP4 are structural wall proteins that are necessary for stomatal function. In Cell Wall Meeting XII, Porto, Portugal, p Abstract T14


Lamport DTA (1969) The isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymic degradation of primary cell walls. Biochemistry 8: 1155-1163


Lamport DTA, Northcote DH (1960a) The use of tissue cultures for the study of plant-cell walls. Biochem J 76: 52P


Schnabelrauch LS, Kieliszewski MJ, Upham BL, Alizedeh H, Lamport DTA (1996) Isolation of pl 4.6 extensin peroxidase from tomato cell suspension cultures and
identification of Val-Tyr-Lys as putative intermolecular cross-link sites. Plant J. 9: 477-489


Swords KMM, Staehelin LA (1993) Complementary immunolocalization patterns of cell wall hydroxyproline-rich glycoproteins studied with the use of antibodies directed against different carbohydrate epitopes. Plant Physiol. 102: 891-901


Figure 1. Amino acid structures of tyrosine derivatives. A, the diphenylether, isodityrosine (Idt). B, the tri-tyrosine, pulcherinosine. C, the tetra-tyrosine, di-isodityrosine (di-Idt).
Figure 2. AFM images of self-assembling extensins. *Top*, Tomato P1 (P1). *Bottom*, Maize Threonine-rich HRGP (THRGP). Proteins were dissolved in ddH$_2$O to a concentration of 10 ug/ml. Sixty ul of the protein solutions were deposited on to a freshly cleaved highly ordered pyrolytic graphite for 5 min then blotted dry. The graphite surface was rinsed with 100 ul of ddH$_2$O and then dried under N$_2$. AFM imaging was carried out on a MFP-3D AFM system (Asylum Research, Santa Barbara, CA) using AC mode in air. The P1 image is a 256×256 pixel scan while the THRGP image is a 512×512 pixel scan. For further details and comparison with AFM images of an Arabidopsis extensin, see (Cannon et al., 2008).