The WRKY superfamily of plant transcription factors

Thomas Eulgem, Paul J. Rushton, Silke Robatzek and Imre E. Somssich

The WRKY proteins are a superfamily of transcription factors with up to 100 representatives in *Arabidopsis*. Family members appear to be involved in the regulation of various physiological programs that are unique to plants, including pathogen defense, senescence and trichome development. In spite of the strong conservation of their DNA-binding domain, the overall structures of WRKY proteins are highly divergent and can be categorized into distinct groups, which might reflect their different functions.

One of the apparent fundamental principles of biological evolution is that the progression from ancient to advanced life forms is inseparably connected to an increase in regulatory capacity. Genome-sequencing efforts have provided evidence for a positive correlation between the proportion of genes involved in information processing and the complexity of organisms. More than 20% of the genes within the sequence available for the *Arabidopsis thaliana* genome appear to encode proteins that play a role in signal transduction or transcription[1], whereas only 12% of the genome of the single-celled yeast *Saccharomyces cerevisiae* contains genes of this type[2].

This increase in biological complexity coincides with the appearance or expansion of specific groups of regulator genes. One example is the nuclear-receptor-gene family, which is completely absent in yeast but highly represented in metazoan organisms[3]. The evolution of nuclear receptors is believed to be a key event in the development of intercellular communication, a prerequisite for the multicellularity of metazoans[4]. Similarly, the establishment of a complex animal body plan was driven by the amplification and divergence of ancestral homeobox genes, thereby generating a sophisticated regulatory system of functionally interconnected transcriptional regulators[5].

To meet their disparate biological requirements, plants and animals have evolved unique regulatory mechanisms. This was partly achieved by combining functional domains from pre-existing factors to build new regulators, as exemplified by the MADS-box factors, which play a central role in determining floral and organ identity in plants[6]. In addition, completely new factors have arisen and we focus here on the potential biological roles of WRKY (pronounced 'worky') proteins, a large family of transcriptional regulators that has to date only been found in plants. The abundance of information provided by the *Arabidopsis* sequencing projects is an ideal basis for comparative analysis of this superfamily within one plant species. Although their precise regulatory functions are largely unknown, the fact that these factors appear to be specific to plants, with probably up to 100 members in *Arabidopsis*, suggests that they play an important role during plant evolution.

Biochemical properties of WRKY proteins

The first WRKY cDNAs were cloned from sweet potato (*Ipomoea batatas*; SPF1), wild oat (*Avena fatua*; ABF1,2), parsley (*Petroselinum crispum*; *Pc*WRKY1,2,3) and *Arabidopsis* (ZAP1), based on the ability of the encoded proteins to bind specifically to the DNA sequence motif (T)(T)TGAC(C/T), which is known as the W box[7–10]. It has been suggested that the cognate binding site for SPF1 is different from other WRKY proteins. However, the oligonucleotide used to isolate SPF1 does have a W box in the flanking sequence[7].

The name of the WRKY family is derived from the most prominent feature of these proteins, the WRKY domain, a 60 amino acid region that is highly conserved amongst family members. The emerging picture is that these proteins are regulatory transcription factors with a binding preference for the W box, but with the potential to differentially regulate the expression of a variety of target genes. Consistent with a role as transcription factors, *Pc*WRKY1 and WIZZ (from tobacco) have been shown to be targeted to the nucleus[11,12].

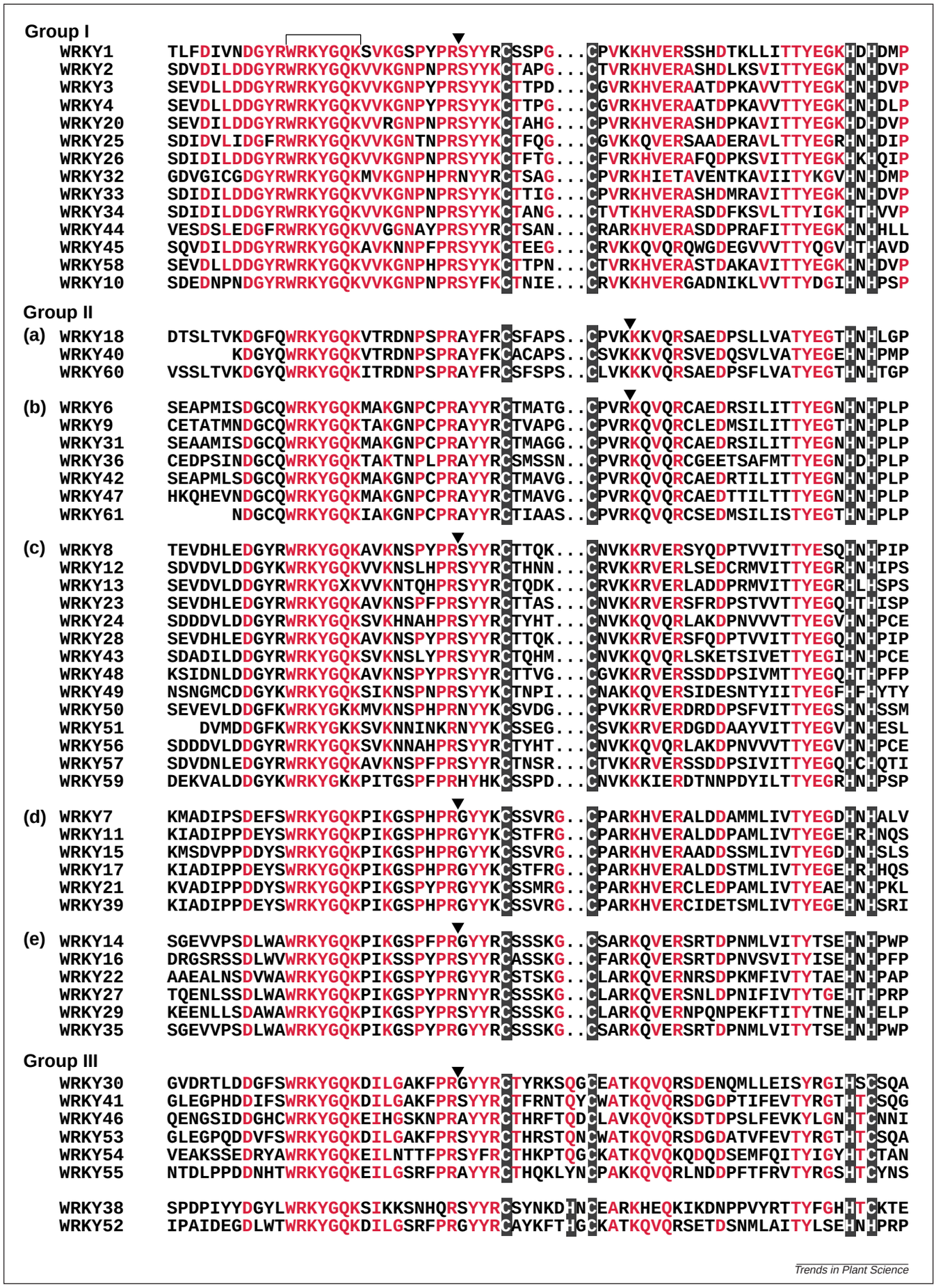
The WRKY domain and the W box

The WRKY domain is defined by the conserved amino acid sequence WRKYGQK at its N-terminal end, together with a novel zinc-finger-like motif[8] (Fig. 1). Because of the clear binding preference of all characterized WRKY proteins for the same DNA motif, it has been assumed that the WRKY domain, as their only conserved structural feature, constitutes a DNA-binding domain. Indeed, it has recently been shown that an isolated WRKY domain has sequence-specific DNA-binding activity[12]. The divalent metal chelators 1,10-*o*-phenanthroline and EDTA abolish *in vitro* DNA binding, which is taken as strong support for a zinc-finger structure within the WRKY domain[8,10,11]. However, it has not yet been proven that zinc is actually complexed in the WRKY domain. In addition, nothing is known about the function of the WRKYGQK heptapeptide stretch, the hallmark of this superfamily.

All known WRKY proteins contain either one or two WRKY domains. They can be classified on the basis of both the number of WRKY domains and the features of their zinc-finger-like motif. WRKY proteins with two WRKY domains belong to group I, whereas most proteins with one WRKY domain belong to group II (Fig. 2). Generally, the WRKY domains of group I and group II members have the same type of finger motif, whose pattern of potential zinc ligands ($\underline{C}$–X_{4-5}–$\underline{C}$–X_{22-23}–$\underline{H}$–X_1–$\underline{H}$; Fig. 1) is unique among all described zinc-finger-like motifs[13]. The single finger motif of a small subset of WRKY proteins is distinct from that of group I and II members. Instead of a C_2–H_2 pattern, their WRKY domains contain a C_2–HC motif ($\underline{C}$–X_7–$\underline{C}$–X_{23}–$\underline{H}$–X_1–$\underline{C}$; Fig. 1). Owing to this distinction, they were recently assigned to the newly defined group III. Nevertheless, experimental evidence has shown that members of all three groups bind sequence specifically to various W box elements (R.S. Cormack *et al.*, unpublished).

The two WRKY domains of group I members appear to be functionally distinct. As has been shown for SPF1, ZAP1 and *Pc*WRKY1, sequence-specific binding to their target DNA sequences is mediated mainly by the C-terminal WRKY domain[7,10,12]. The function of the N-terminal WRKY domain remains unclear. Because protein regions outside of the C-terminal WRKY domain contribute to the overall strength of DNA

1360 - 1385/00/$ – see front matter PII: S1360-1385(00)01600-9



Trends in Plant Science

Fig. 1. Left Comparison of WRKY domain sequences from *At*WRKY proteins. Sequences encoding the peptide stretch WRKYGQK were found by the BLAST programs tblastn and blastp programs[37] in genomic and EST databases. Gaps (dots) have been inserted for optimal alignment. Residues that are highly conserved within each of the major groups are in red and potential zinc ligands are highlighted in black boxes. For each (sub)group, the position of a conserved intron is indicated by an arrowhead.

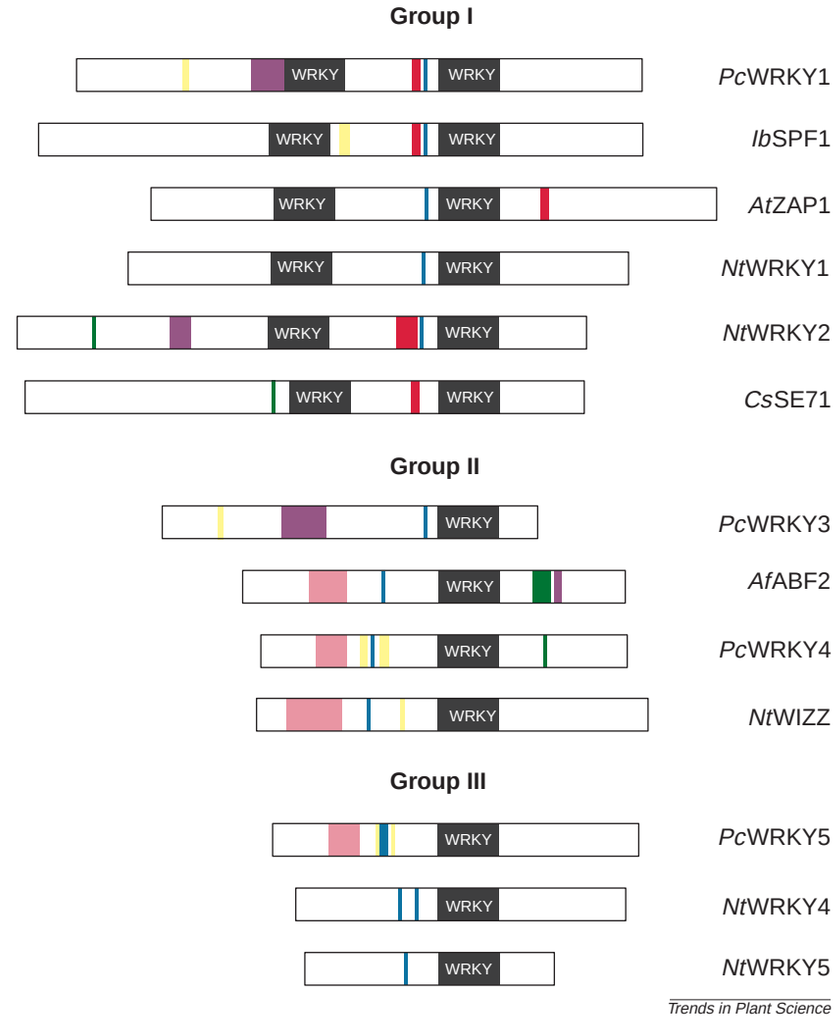


Fig. 2. Schematic representation of published full length WRKY proteins from parsley (*Pc*), sweet potato (*Ib*), *Arabidopsis* (*At*), tobacco (*Nt*), cucumber (*Cs*) and wild oat (*Af*). They are divided into three groups based on the number and type of the WRKY domains they contain. WRKY domains are black, putative basic nuclear localization signals are blue and leucine zippers are pink. Serine–threonine-rich regions are yellow, glutamine-rich regions are purple, proline-rich regions are green and acidic regions are red.

binding, the N-terminal domain might participate in the binding process, increasing the affinity or specificity of these proteins for their target sites. Alternatively, it might provide an interface for protein–protein interactions, a known function of some zinc-finger-like domains[14]; this could allow more efficient DNA binding through interactions with other DNA-associated proteins. Not unexpectedly, the single WRKY domains of group II and III family members are more similar in sequence to the C-terminal than to the N-terminal WRKY domain of group I proteins, suggesting that the C-terminal and single WRKY domains are functionally equivalent and constitute the major DNA-binding domain.

The conservation of the WRKY domain is mirrored by a remarkable conservation of the cognate *cis*-acting W box elements. These (T)(T)TGAC(C/T) sequence elements contain the invariant TGAC core, which is essential for function and WRKY binding. They mediate transcriptional responses to pathogen-derived elicitors[9,15] and are present in the promoters of many plant genes that are associated with defense[16]. Functional W boxes frequently cluster within short promoter stretches[15–17] and can act together synergistically[12]. WRKY–W box interactions have been demonstrated by numerous binding experiments, both *in vitro* and *in vivo*[8–10,12,18,19] (R.S. Cormack *et al.*, unpublished), and random binding-site selection assays have shown that the optimal binding site for ZAP1 contains the W box motif[10]. Interactions of WRKY proteins with W boxes can be regulated post-translationally, because binding of WRKY-like DNA-binding activities to W boxes in tobacco is abolished by treatment with alkaline phosphatase[18] and the protein-kinase inhibitor staurosporin[20].

In spite of the stereotypic binding preferences of WRKY proteins for W boxes, their affinities for certain types or arrangements of this element can vary (R.S. Cormack *et al.*, unpublished). Sequences flanking the invariant W box TGAC core might be partly responsible for the observed specificity. In addition, the cooperative assembly of discrete higher-order WRKY–DNA complexes at defined W box arrangements might also account for specific promoter recognition[12]. Owing to the high variability in overall protein structure, access to certain promoters would be restricted to distinct family members that fit into the three-dimensional shapes of such complexes. Indeed, WRKY proteins might be part of multimeric protein–DNA complexes. Both WRKY protein-containing nuclear extracts and purified recombinant WRKYs from tobacco lose their DNA-binding activity when treated with the protein-dissociating agent deoxycholate[18]. Furthermore, some WRKY proteins contain potential leucine zippers (LZs), structures known to allow protein dimerization. They appear to be functional in *Pc*WRKY4 and 5, because their deletion greatly reduces reporter-gene expression mediated by these proteins in yeast (R.S. Cormack *et al.*, unpublished).

Transcriptional regulation

ZAP1 and *Pc*WRKY1, 4 and 5 can activate transcription in yeast[10,12] (R.S. Cormack *et al.*, unpublished), a feature that has been confirmed for ZAP1 and *Pc*WRKY1 in plant cells[10,12]. Although it has yet to be studied in detail, the primary structures of WRKY proteins have an abundance of potential transcriptional activation or repression domains (Fig. 2). A common feature of many domains affecting transcription is the predominance of certain amino acids,

Table 1. Identified members of the *Arabidopsis* WRKY superfamily of transcription factors

AtWRKY	Group	Chr.	EST[a]	Gene[a]	BAC.ORF	AtWRKY	Group	Chr.	EST[a]	Gene[a]	BAC.ORF
1	I	2	X92976/ZAP1	AC007211	F1013.1	19		4		AL049638	F16J13.90
			AI995838	AC006955	F2818					AL091613	T8E18-end
2	I	5	T44598	AB026656	MXK23	20	I	4		AL078465	T15N24.90
			AA395490			21	II	2	AA586133	U93215	T06B20.6
			N37131						T20410	AQ010529	F24C8-end
3	I	2	T45479	AC006284	T4M8.23				AI992739		
			AI099874	B77849	T29F22-end	22	II	4	T04811	AF007269	IG002N01.6
			AI993164	AL080571	F1G15-end	23	II	2	F14417	AC002337	T08I13.10
				AL096246	T16K23-end				F14438		
4	I	1	T22085	AC007576	F7A19.5	24	I	5		AB005233	MBK23
			W43265			25	I	2	T42934	AC002338	T9D9.6
			AA585810							AC004165	T27E13.1
5		5	H77044	AB011485	MXH1.P3					AL093076	T10P21-end
			H77050			26	I	5	AA585811	B09174	T30A11-end
			H77051						T22092	AB010697	M0J9.24
			AI995170						AI995443		
			AA605512			27		5		AB009055	MXC20.3
6	II	1	U75592	AC000375	F19K23.22	28	II	4		AL021713	T9A21.10
			AA650675			29	II	4		AL035394	F9D19.20
			H77127			30	III	5		AB010696	MLE8.2
			AA394951			31	II	4		AL022140	F1N20.170
			AA650826			32	I	4		AL022198	F6I18.160
			AI992388			33	I	2		AC004683	T19C21.4
7	II	4	N37775	AC005861	F23B24			2		AC005499	T6A23.33
			T20578	AL078637	T22A6.70	34	I	4		AL022223	M3E9.130
			R30038			35	II	2		AC004238	F19I3.6
			AI992658			36	II	1		AC010675	T17F3.16
8	II	5		AB010698	MPL12.9					AC010852	T12P18
9			AI998645	AQ011596	F24A12TRC-end	37				B23309	F28C5-end
				B98122	F24A12TRB-end	38	?	5		AB012244	MQJ16.9
10	II	1		AC002328	F20N2	39	II	3		AC011437	F7O18.30
11	II	4	R64846	AL080283	F3L17.120	40	II	1		AC011713	F23A5
			T88086			41	III	4		AF080120	F2P3.16
			R30283							AL049876	T22B4.50
			AI998936			42	II	4		AF076243	T26N6.6
			T22071			43		2		AC005397	T3F17.22
			T42669			44	I	2		AC005896	F3G5.5
			Z29806			45		1/3?		ATU63815	AT.I.24-4
			Z29805							AC010797	IGF-F28J7
12		2		AC003672	F16B22					AC011664	F5A18
13			F14100							AC011624	T18B3
	II	4		AL078620	F23K16.40	46	III	2		AC006526	F11C10.9
14		1	D88748	AC007060	T5I8.10	47	II	4		AF104919	T15B16.12
			T20672			48		5		AB023033	K6M13
15	II	2	Z25667	AC002391	T20D16.5	49		5		AB017070	MNL12
			T04430			50		5		AC005965	T19G15
			T43675			51		5		AB019236	MXK3
			T21472			52	III?	5		AB020744	K9E15
			H36048			53	III	4		AL078468	T32A16
			AI993841							AL035394	F9D16.280
16	II	5	AA042185	AB010693	K21C13.P1	54	III	2		AC007660	T7D17.7
			AA395309	B27842	T19B7-end	55	III	2		AC007660	T7D17.8
17		2	AA712348	AC006954	F25P17.13	56		1		AC007764	F22C12
			R90490			57				AL080748	F1L14-end
			AA067545			58	I	3		AC008261	T4P13
			AI100579			59	II	2		AC007019	F7D8.22
18	II	4	U74179	AL031004	F28M20.10	60	II	2		AC006585	F27C12.8
				AL049607	F11C18	61	II	1		AC011809	F6A14

[a]GenBank Accession no.

Abbreviations: Chr., chromosome; ORF, open reading frame; question marks denote either inconclusive group assignment or inconclusive chromosomal position.

including alanine, glutamine, proline, serine, threonine and charged amino acids[21,22]. At least two of the seven potential 'trans-regulatory' domains in *Pc*WRKY1 activate transcription in yeast[12]. The possibility that WRKY proteins possess both activator and repressor functions, as shown for the maize VP1 (Ref. 23), remains to be tested.

Complexity of the WRKY family in *Arabidopsis*

The large amount of genomic and cDNA sequences available from *Arabidopsis* yields insights into the complexity of the WRKY family in a single plant species. In total, 61 distinct ORFs potentially encoding WRKY proteins can be found in the databases to date (Table 1). With the exception of *At*WRKY1, which is identical to ZAP1 (Ref. 10), and *At*WRKY44, which is defined by the *ttg2* mutant (C.S. Johnson and D.R. Smyth, pers. commun.), none of these proteins has been described before. We encourage the use of the designations used in Table 1 in future studies to avoid the confusion often caused when multiple names are assigned to a given gene member within large families.

The *AtWRKY* genes are randomly distributed over the five chromosomes and preliminary analyses suggest that they might all be present as single copies (I. Somssich and S. Robatzek, unpublished). Many of these putative WRKY proteins are represented by ESTs showing that the corresponding genes are expressed. By the number and sequence of their WRKY domains, these proteins can be assigned to the three major groups. Given that about two-thirds of the *Arabidopsis* genome has been sequenced to date, the total number of *WRKY* genes in this species might be as high as 100.

A phylogenetic tree of the *At*WRKY proteins based on their WRKY domains (Fig. 3) clearly indicates that group II splits up into five distinct subgroups (IIa–e). The resulting refined classification is further substantiated by the presence of ten additional structural motifs that are conserved among subsets of *At*WRKY family members. Each of these motifs occurs only in certain subgroups and each subgroup seems to be best defined by combinations of such motifs. In some cases, the sequences of these motifs can reveal clues about their potential functions. In addition to peptide sequences that might serve as nuclear localization signals[24], a heptad repeat of bulky hydrophobic residues characteristic for LZs (Ref. 25) is present in some of the proteins. The heptad repeat occurs exclusively in members of subgroups IIa and IIb. Recent experiments have shown that the LZ region of *At*WRKY6 mediates dimerization (S. Robatzek and I.E. Somssich, unpublished).

An additional common feature that is found in the *WRKY* genes is the existence of an intron within the region encoding the C-terminal WRKY domain of group I members or the single WRKY domain of group II and III members. This intron position is highly conserved, being localized after the codon encoding arginine that is N terminal to the zinc-finger-like motif (Fig. 1). Strikingly, in all the genes encoding subgroup IIa and IIb members, the position of this intron is exactly 16 codons further towards the C terminus. In spite of the phylogenetic distance of their WRKY domains, members of all three groups have been shown to recognize W box elements, indicating that this is a general feature of the entire superfamily.

A few *At*WRKY proteins do not fit neatly into any one (sub)group. For example, *At*WRKY10, which carries only one WRKY domain, appears to be more related to group I (Fig. 3). This might be explained by the secondary loss of the N-terminal WRKY domain. Furthermore, based on the pattern of cysteine and histidine residues within their WRKY domains (Fig. 1), *At*WRKY38 and *At*WRKY52 could either belong to group III or represent members of a novel group (Fig. 3).

Biological roles of WRKY factors

One of the most challenging questions concerns the regulatory processes governed by WRKY proteins. Clues might come partly from gene expression studies. Because many *WRKY* genes are themselves transcriptionally regulated, their distinct expression patterns might yield hints as to the regulatory functions of the encoded factors under particular biological conditions. In addition, a full understanding of the biological roles of these factors will require the identification of the target genes whose expression they affect.

Expression behavior of WRKY genes

Current data point to many WRKY proteins having a regulatory function in the response to pathogen infection and other stresses. Effective plant defense against pathogenic microorganisms is associated with the concerted activation of a large variety of genes, occurring in several temporally distinct waves[26]. Increased levels of WRKY mRNA, protein and DNA-binding activity have been reported to be induced by infection with viruses[19], bacteria (A. Dellagi and P. Birch, pers. commun.) or oomycetes[12], by fungal elicitors[9,20] (R.S. Cormack *et al.*, unpublished), and by signaling substances such as salicylic acid[18]. In addition, *WRKY* gene expression has been shown to be upregulated in response to wounding[11] (S. Robatzek and I.E. Somssich, unpublished) and upon local mechanical stimulation of plant protoplasts[27]. Induced WRKY mRNA accumulation is often extremely rapid and transient, and seems not to require *de novo* synthesis of regulatory factors[9,11] (R.S. Cormack *et al.*, unpublished). This immediate–early expression behaviour indicates a role for the WRKY proteins in regulating subsequently activated secondary-response genes, whose products carry out the protective and defensive reactions.

Comparative expression studies with several *AtWRKY* genes also suggest that certain family members have a role in the regulation of senescence (S. Robatzek and I.E. Somssich, unpublished). Transcript levels of *AtWRKY4*, *6*, *7* and *11* are enhanced in senescent leaves. In transgenic *Arabidopsis* plants, an *AtWRKY6* promoter–*GUS* reporter gene is strongly activated in senescent leaves as well as in response to infection by pathogenic bacteria. As several genes are known to be highly expressed during both leaf senescence and defense, we might expect the existence of common regulatory mechanisms between these two physiological processes[28].

Inspection of plant databases has revealed the existence of more than 500 WRKY ESTs identified from various tissue sources, including roots, leaves, inflorescences, abscission zones, seeds and vascular tissue, as well as from drought- or salt-stressed, or pathogen-infected tissue. Thus, *WRKY* genes appear to be expressed in numerous cell types and under different physiological conditions and could therefore participate in the control of a wide variety of biological processes.

Targets of WRKY regulation

As suggested by the general binding preference of WRKY proteins for W boxes, genes containing these promoter elements are likely targets of WRKY factors, and these include the *WRKY* genes themselves[12] as well as a large variety of defense-related genes of the *PR* type[16,18]. Additionally, gibberellic acid-induced expression of the wild-oat *α-Amy2/54* gene[8] and activation of the barley *HvLox1* gene in response to the defense and wound signaling molecule jasmonic acid[29] also appear to involve WRKY–W box interactions. Furthermore, a role has been suggested for SPF1 in the sucrose- or polygalacturonic-acid-induced expression of genes coding for sporamin and β-amylase in sweet potato[7]. However, as mentioned earlier, uncertainties about the

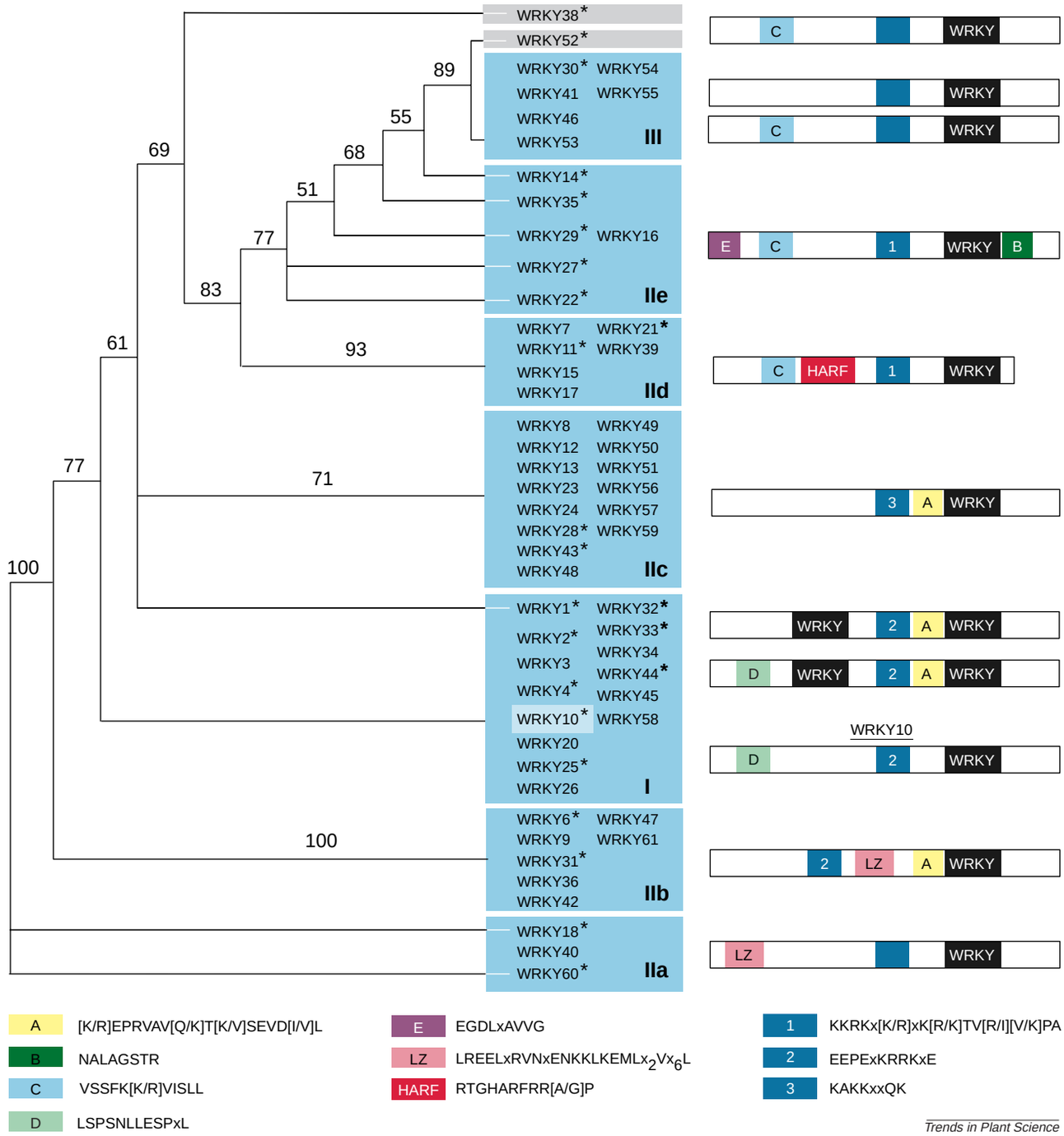


Fig. 3. Phylogenetic analysis of 58 members of the *At*WRKY family. Amino acid sequences from the single WRKY domain of group II and III members or the C-terminal WRKY domain of group I members were aligned using PileUp (Wisconsin Package Version 10.0, Genetics Computer Group, Madison, WI, USA). The diagram shows the most parsimonious tree constructed using PAUP 3.1.1 (Smithsonian Institution, Washington, DC, USA) to perform a heuristic search with a pre-aligned reduced data set including only representatives of each *At*WRKY(sub)group (indicated by asterisks). Based on the results of additional PAUP 3.1.1 runs with extended data sets, further members of each (sub)group were added to the figure. The tree shown is unrooted and has a consistency index of 0.808. The numbers above the branches are bootstrap values from 100 replicates. *At*WRKYs that consistently clustered together are grouped in blue boxes. Members of subgroups IIa and IIe, are not lined up at the branches of separate subtrees but nevertheless share significant similarities in their WRKY domains. Higher-order branches on the right-hand side of the tree representing relationships within each (sub)group were not highly reproducible and were therefore eliminated. White extensions of branches within the blue boxes indicate that the branch leads only to one distinct *At*WRKY. Family members that cannot be unequivocally assigned to a defined (sub)group are highlighted by gray boxes. Conserved primary structural features of the *At*WRKY family outside the WRKY domains were identified using MEME (Ref. 38; http://www.sdsc.edu/MEME) and are shown below the tree. Schematic representations of typical members of each (sub)group are shown on the right: WRKY domains are indicated by black boxes; motifs 1, 2 and 3 are basic stretches that might be nuclear localization sequences; additional basic motifs not detected by MEME are shown by blue boxes without numbers; LZ indicates potential leucine zipper structures that were also predicted by the COILSCAN and COIL (Wisconsin Package Version 10.0)[39] programs.

exact binding site of SPF1 means that more work is required to establish its role *in vivo*. To date, W boxes have been described as positive *cis*-acting elements upregulating transcription. However, in the case of the *Arabidopsis PR1* gene, the basal and salicylic acid-induced expression levels might be negatively regulated by W boxes[17]. SNI1, a negative regulator of *PR* gene expression, was recently identified in a genetic screening for second-site suppressors of the *Arabidopsis* mutation *npr1* (Refs 30,31). Interestingly, SNI1, which is nuclear localized, contains no obvious DNA-binding domain. One possible mode of SNI1 action would involve interaction with WRKY factors bound to the W box[31].

The involvement of WRKY factors in regulating part of the defense program is further substantiated by a large-scale expression profiling study (J. Dangl and R.A. Dietrich, pers. commun.). Using a DNA microarray with 10 000 *Arabidopsis* ESTs, a group of 25 genes, including *PR1*, was identified whose expression responded coordinately to various pathogens as well as to other defense-inducing conditions. Within the first kilobase of their promoters, these genes shared only the W box motifs (TTGAC), with on average four copies, which were often clustered. By contrast, the promoters of a control set of genes not coordinately regulated with *PR1* contained, on average, less than two W boxes.

The only WRKY mutant so far described is *transparent testa glabra 2* (*ttg2*), which is based on a transposon insertion within *AtWRKY44/TTG2* (C.S. Johnson and D.R. Smyth, pers. commun.). In *ttg2*, the number of trichomes and their branching is reduced, as is anthocyanin pigmentation of the seed coat, together with a loss of mucilage. This pleiotropic phenotype resembles that of *ttg1*, which is defective for a regulatory protein of the WD40-repeat type[32]. *At*WRKY44/TTG2 and TTG1 might therefore act in the same regulatory cascade, controlling a common set of genes. The extensive use of reverse genetics to obtain additional tagged *WRKY* mutants, as well as the generation of *WRKY* promoter–reporter gene and WRKY overexpressor lines, will allow us to gain a more comprehensive understanding of the various biological roles of WRKY proteins. Furthermore, inducible expression systems[33] could be used for controlled temporal overexpression of *WRKY* transgenes in their own loss-of-function background; combined with methods of large-scale gene expression profiling (e.g. differential display, DNA chips), this should facilitate the identification of defined WRKY target genes. In a similar way, the *Arabidopsis NAP* gene was identified as a target of the APETALA3–PISTILLATA transcription factor dimer[34].

Conclusions

WRKY proteins have only recently been identified as a new family of transcription factors. In *Arabidopsis*, this family appears to be nearly as complex as the well-known MYB family[35], but it is restricted to the plant kingdom. This suggests that *WRKY* genes originated concurrently with the major plant phyla. Current information suggests that WRKY factors play a key role in regulating the pathogen-induced defense program. The exposure of plants to a wide variety of biotic or abiotic stresses connected with their sessile, autotrophic lifestyle could be one major factor in the enormous expansion of the *WRKY* family during evolution. In addition, the extensive metabolic changes associated with the establishment of defense responses[26] or senescence[36] might require an elaborate regulatory system.

WRKY proteins also seem to be involved in other plant-specific processes, such as trichome development and the biosynthesis of secondary metabolites. Thus, they appear to participate in controlling the expression of a plethora of genes. As with other large gene families, the problem of functional redundancy will complicate genetic attempts to determine the role of individual WRKY proteins. Comparative studies in lower plants (e.g. ferns, mosses and algae) can give clues to whether *WRKY* gene diversification correlates with increasing developmental and metabolic pathway complexity. Furthermore, generating *Arabidopsis* knock-out lines that affect several members of individual subgroups might help to 'wrky' matters out.

Acknowledgements

We thank Hiroshi Sano (NAIST, Japan); David R. Smyth (Monash University, Australia); Zhixiang Chen (University of Idaho, USA); Jeff Dangl (University of North Carolina, USA); Robert Dietrich (Novartis, Research Triangle, USA); Alia Dellagi and Paul Birch (Scottish Crop Research Institute, UK), for providing preprints of unpublished data; and Klaus Hahlbrock for critical reading of the manuscript and continuous support.

References

1 Bevan, M. *et al.* (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* 391, 485–488
2 Mewes, H.W. *et al.* (1997) Overview of the yeast genome. *Nature* 387, 7–8
3 Clarke, N.D. and Berg, J.M. (1998) Zinc fingers in *Caenorhabditis elegans*: finding families and probing pathways. *Science* 282, 2018–2022
4 Laudet, V. *et al.* (1992) Evolution of the nuclear receptor gene superfamily. *EMBO J.* 11, 1003–1013
5 Gellon, G. and McGinnis, W. (1998) Shaping animal body plans in developmental and evolution by modulation of *Hox* expression patterns. *BioEssays* 20, 116–125
6 Riechmann, J.L. and Meyerowitz, E.M. (1997) MADS domain proteins in plant development. *Biol. Chem.* 378, 1079–1101
7 Ishiguro, S. and Nakamura, K. (1994) Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5′ upstream regions of genes coding for sporamin and β-amylase from sweet potato. *Mol. Gen. Genet.* 244, 563–571
8 Rushton, P.J. *et al.* (1995) Members of a new family of DNA-binding proteins bind to a conserved *cis*-element in the promoters of *α-Amy2* genes. *Plant Mol. Biol.* 29, 691–702
9 Rushton, P.J. *et al.* (1996) Interaction of elicitor-induced DNA binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *EMBO J.* 15, 5690–5700
10 de Pater, S. *et al.* (1996) Characterization of a zinc-dependent transcriptional activator from *Arabidopsis*. *Nucleic Acids Res.* 24, 4624–4631
11 Hara, K. *et al.* (2000) Rapid systemic accumulation of transcripts encoding a tobacco WRKY transcription factor upon wounding. *Mol. Gen. Genet.* 263, 30–37
12 Eulgem, T. *et al.* (1999) Early nuclear events in plant defense: rapid gene activation by WRKY transcription factors. *EMBO J.* 18, 4689–4699
13 Berg, J.M. and Shi, Y. (1996) The galvanization of biology: a growing appreciation for the roles of zinc. *Science* 271, 1081–1085
14 Mackay, J.P. and Crossley, M. (1998) Zinc fingers are sticking together. *Trends Biochem. Sci.* 23, 1–4
15 Fukuda, Y. and Shinshi, H. (1994) Characterization of a novel *cis*-acting element that is responsive to a fungal elicitor in the promoter of a tobacco class I chitinase gene. *Plant Mol. Biol.* 24, 485–493
16 Rushton, P.J. and Somssich, I.E. (1998) Transcriptional control of plant genes responsive to pathogens. *Curr. Opin. Plant Biol.* 1, 311–315
17 Lebel, E. *et al.* (1998) Functional analysis of the regulatory sequences controlling *PR-1* gene expression in *Arabidopsis*. *Plant J.* 16, 223–233
18 Yang, P. *et al.* (1999) A pathogen- and salicylic acid-induced WRKY DNA-binding activity recognizes the elicitor response element of tobacco class I chitinase gene promoter. *Plant J.* 18, 141–149
19 Wang, Z. *et al.* (1998) An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with plant defense. *Plant J.* 16, 515–522

20 Fukuda, Y. (1997) Interaction of tobacco nuclear proteins with an elicitor-responsive element in the promoter of a basic class I chitinase gene. *Plant Mol. Biol.* 34, 81–87
21 Triezenberg, S.J. (1995) Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.* 5, 190–196
22 Hanna-Rose, W. and Hansen, U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* 12, 229–234
23 Hoecker, U. *et al.* (1995) Integrated control of seed maturation and germination programs by activator and repressor functions of Viviparous-1 of maize. *Genes Dev.* 9, 2459–2469
24 Garcia-Bustos, J. *et al.* (1991) Nuclear protein localization. *Biochim. Biophys. Acta* 1071, 83–101
25 Landschulz, W.H. *et al.* (1988) The leucine zipper: a hypothetical structure common to a new class of DNA-binding proteins. *Science* 240, 1759–1764
26 Somssich, I.E. and Hahlbrock, K. (1998) Pathogen defense in plants – a paradigm of biological complexity. *Trends Plant Sci.* 3, 86–90
27 Gus-Mayer, S. *et al.* (1998) Local mechanical stimulation induces components of the pathogen defense response in parsley. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8398–8403
28 Quirino, B.F. *et al.* (1999) Diverse range of gene activity during *Arabidopsis thaliana* leaf senescence includes pathogen-independent induction of defense-related genes. *Plant Mol. Biol.* 40, 267–278
29 Rouster, J. *et al.* (1997) Identification of a methyl-jasmonate-responsive region in the promoter of a lipoxygenase-1 gene expressed in barley grain. *Plant J.* 11, 513–523
30 Cao, H. *et al.* (1997) The *Arabidopsis NPR1* gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* 88, 57–63
31 Li, X. *et al.* (1999) Identification and cloning of a negative regulator of systemic acquired resistance, SNI1, through a screen for suppressors of *npr1-1*. *Cell* 98, 329–339
32 Walker, A.R. *et al.* (1999) The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* 11, 1337–1349
33 Gatz, C. and Lenk, I. (1998) Promoters that respond to chemical inducers. *Trends Plant Sci.* 3, 352–358
34 Sablowski, R.W.M. and Meyerowitz, E.M. (1998) A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*. *Cell* 92, 93–103
35 Martin, C. and Paz-Ares, J. (1997) MYB transcription factors in plants. *Trends Genet.* 13, 67–73
36 Gan, S. and Amasino, R.M. (1997) Making sense of senescence. *Plant Physiol.* 113, 313–319
37 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
38 Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* (Altmann, R., ed.), pp. 28–36, AAAI Press
39 Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.* 21, 375–382

Thomas Eulgem, Paul Rushton, Silke Robatzek and Imre Somssich* are at the Max-Planck-Institut für Züchtungsforschung, Abteilung Biochemie, Carl-von-Linné-Weg 10, D-50829 Köln, Germany; Thomas Eulgem is currently at the Dept of Biology, 108 Coker Hall CB#3280, University of North Carolina, Chapel Hill, NC 27599-3280, USA.

*Author for correspondence (tel +49 221 5062310; fax +49 221 5062313; e-mail somssich@mpiz-koeln.mpg.de).

Plant one-carbon metabolism and its engineering

Andrew D. Hanson, Douglas A. Gage and Yair Shachar-Hill

The metabolism of one-carbon (C_1) units is vital to plants. It involves unique enzymes and takes place in four subcellular compartments. Plant C_1 biochemistry has remained relatively unexplored, partly because of the low abundance or the lability of many of its enzymes and intermediates. Fortunately, DNA sequence databases now make it easier to characterize known C_1 enzymes and to discover new ones, to identify pathways that might carry high C_1 fluxes, and to use engineering to redirect C_1 fluxes and to understand their control better.

One-carbon (C_1) metabolism is essential to all organisms. In plants, it supplies the C_1 units needed to synthesize proteins, nucleic acids, pantothenate and many methylated molecules[1]. Fluxes through C_1 pathways are particularly high in plants that are rich in methylated compounds such as lignin, alkaloids and betaines because methyl moieties make up several percent of their dry weight[2]. Transfers of C_1 units are also central to the massive photorespiratory fluxes that occur in all C_3 plants[3]. In spite of the fundamental significance of these roles, and the interest in the metabolic engineering of lignin[2], betaines[4] and photorespiration[3], there is much that is not understood about the enzymes, pathways and regulatory mechanisms of plant C_1 metabolism. In part this is because of the obstacles that C_1 metabolism presents for classical biochemistry and genetics: its enzymes can be of low abundance and/or exist as several isoforms, mutants are lacking, and its key intermediates – C_1 substituted folates – are labile and hard to quantify. Fortunately, classical approaches to C_1 metabolism can now be complemented by genomics-driven approaches that exploit the fast-growing DNA sequence databases. Accordingly, this review has three aims:

- To illustrate how genomics-based approaches are advancing our knowledge of plant C_1 biochemistry.
- To bring together biochemical and genomics-derived data to show which C_1 pathways might operate in plants, and where they operate in the cell.
- To examine progress towards engineering C_1 metabolism.

Nucleotide sequence information – from genomes, cDNAs and ESTs – can be used to complement biochemical approaches in several ways. Because most enzymes of C_1 metabolism are highly

1360 - 1385/00/$ – see front matter PII: S1360-1385(00)01599-5