

Regulation

It is envisaged that a single body would assign the biobarcode and maintain a public database listing the biobarcode sequences of organisms that have been released. It is clearly advantageous to industry, regulators and taxpayers that such a system be instituted because of the amount of resources saved and the protection provided. Because of the savings, it is in the public interest that such a system be universally instituted.

Acknowledgements

Useful discussions with Eithan Rubin of the Weizmann Bioinformatics Group are gratefully acknowledged. This research is supported in part by the Levin Fund. J.G. is the Gilbert de Botton Professor of Plant Sciences.

References

- Gressel, J. (2002) *Molecular Biology of Weed Control*, Taylor & Francis
- Vurro, M. *et al.*, eds (2001) *Enhancing Biocontrol Agents and Handling Risks*, IOS Press, Amsterdam
- Gressel, J. (2001) Potential failsafe mechanisms against the spread and introgression of transgenic hypervirulent biocontrol fungi. *Trends Biotechnol.* 19, 149–154
- Eparvier, A. and Alabouvette, C. (1994) Use of ELISA and GUS transformed strains to study competition between pathogenic and nonpathogenic *Fusarium oxysporum* for root colonization. *Biocontrol Sci. Technol.* 4, 35–47
- Cohen, B. *et al.* (2002) Transgenically enhanced expression of indole-3-acetic acid (IAA) confers hypervirulence to plant pathogens. *Phytopathology* 92, 590–596
- Hintz, W.E. *et al.* (2001) Development of genetic markers for risk assessment of biological control agents. *Can. J. Plant Pathol.* 23, 13–18
- Meyer, R. (1999) Development and application of DNA analytical methods for the detection of GMOs in food. *Food Control* 10, 391–399
- Huebner, P. *et al.* (1999) Quantitative competitive PCR for the detection of genetically modified organisms in food. *Food Control* 10, 353–358
- Wurz, A. *et al.* (1999) Quantitative analysis of genetically modified organisms (GMO) in processed food by PCR-based methods. *Food Control* 10, 385–389
- Lipp, M. *et al.* (2001) Validation of a method based on polymerase chain reaction for the detection of genetically modified organisms in various processed foodstuffs. *Eur. Food Res. Technol.* 212, 497–504
- Anklam, E. *et al.* (2002) Analytical methods for detection and determination of genetically modified organisms in agricultural crops and plant derived food products. *Eur. Food Res. Technol.* 214, 3–26
- Rogan, G.J. *et al.* (1999) Immunodiagnostic methods for detection of 5-enolpyruvylshikimate-phosphate synthase in Roundup-Ready soybeans. *Food Control* 10, 407–414

In Canada, for example, the insertion of a biobarcode would probably not come under regulatory scrutiny if biobarcode are introduced into plants because they are not considered to be 'plants with novel traits' if they are non-protein producing. In Europe, any introduced sequence (except antisense) seems to come under regulatory scrutiny, but after due risk assessment it is hoped that a blanket approval could be obtained for biobarcode that meet the specific criteria listed above (and possibly others).

Domains as functional building blocks of plant proteins

Bernard C-H. Lam and Eduardo Blumwald

Emerging evidence in eukaryotic systems suggests that many proteins of diverse cellular processes are made up of protein domains that are well defined in both sequence and structure. This article updates the identification of many 'classic' eukaryotic protein domains in various plant cellular processes, with particular emphasis on the non-catalytic categories. We discuss the importance of domains to plant-protein functions and cellular networking, and the emergence of plant-specific domains.

The completion of the genome sequence of the model plant *Arabidopsis* [1] has allowed many crucial questions in plant science to be investigated much more quickly. The identification of genes responsible for key *Arabidopsis* mutants has been simplified and comparisons between *Arabidopsis* and fungal and animal genomes has allowed the identification of the genetic bases of the differences between plants and other eukaryotic organisms. At the biochemical level, the study of the changes of the complete expression patterns in response to developmental or environmental signals at the RNA and protein levels have been made possible by the development

of DNA microarrays and mass-spectrometry techniques, respectively.

However, these advances facilitated by the *Arabidopsis* genome development are also generating many more questions to be studied. Notably, fewer than 10% of the predicted *Arabidopsis* genes have been studied experimentally in an extensive manner. Moreover, more than 25% of the genes cannot be classified according to function(s) by mere sequence comparison to proteins of known roles in other organisms [1]. Although functional-genomic approaches such as the large-scale generation of T-DNA insertional mutant lines have been initiated [2], complementary approaches will probably be needed until the function of every gene product of *Arabidopsis* and other model plants is known.

Interestingly, in addition to the holistic or 'top-down' approach through mutant identification and characterization, there is currently a renaissance of plant studies starting at the level of the gene product itself (a minimalist or 'bottom-up' approach). For example, many key proteins involved in processes such as membrane transport [3] and signal transduction [4] were first identified *in silico* by searching the databases for gene products with sequence similarity to their counterparts (which were well characterized in many cases) in other model organisms such as *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans* and various mammalian species. Function prediction by amino acid sequence analysis or 'homology search' has become a common starting point of experimental design of plant research in the post-genome era.

Bernard C-H. Lam

Dept of Molecular Biology, Massachusetts General Hospital, and Dept of Genetics, Harvard Medical School, Boston, MA 02114, USA.

Eduardo Blumwald

Dept of Pomology, University of California, One Shields Ave, Davis, CA 95616, USA. e-mail: eblumwald@ucdavis.edu

Table 1. Common protein structure and domain prediction websites^a

Website	Address	Major features	Ref
InterPro	http://www.ebi.ac.uk/interpro	Interface combining various databases including Pfam, SMART, ProDom, Prosite, PRINTS and TIGRFAMS	[26]
Pfam	http://pfam.wustl.edu/	Uses Hidden Markov Models for a more sensitive search Many (>3000) domains in database, including non-signaling types Both DNA and protein input available Provides a ranking list of most observed domains Good graphical interface	[27]
ProDom	http://prodes.toulouse.inra.fr/prodom/doc/prodom.html	PSI-BLAST based ~400 domains classified Provide genome-wide scan for domain of interest, but no species-specific search	[28]
PROSITE	http://www.expasy.org/prosite/	BLAST-based linear comparison >1500 patterns and motifs in the database Provides motif and signature (for example, a sequence with less than 10 amino acids) search Also provides domain search	[29]
SBASE	http://www.icgeb.trieste.it/sbase/	>1500 domain or segment 'groups' in database for retrieval Lack of plant entries in database BLAST-based linear comparison Output delivered via e-mail	[30]
SMART	http://dylan.embl-heidelberg.de/	Uses Hidden Markov Models for a more sensitive search Specializes in signaling domains >600 defined domains in database Domain-arrangement and -architecture search Taxonomic breakdown and distribution Species-specific and genome-wide scans for domain of interest Good graphical interface	[9]

^aFor the details on the algorithms of each search method, please refer to the original publications.

However, a major pitfall of 'best hit' similarity search using software such as BLAST [5] is that, although many proteins share similar functions, they cannot be related because of their low overall sequence homology. Moreover, the predictability of a conventional BLAST search becomes limited if the 'best hit' obtained is yet another hypothetical protein or a protein of no known function. During the past few years, there has been a significant increase in the amount of specialized protein-prediction and function-search software. Some of these programs aim to identify physical landmarks of the proteins of interest. In practice, many now allow the identification of various specialized protein secondary structures such as transmembrane domains (e.g. BLOCK [6]). More recently, several programs (Table 1) have been released that are designed to identify conserved sequence segments associated with a particular function. Studies based on these programs suggest that many proteins in various organisms contain at least one such unit or domain. At least half of all the predicted genes in *Arabidopsis* contain one or more domains [1]. The use of protein-domain identification could become a crucial tool for plant-genome annotation, as well as being an important guide for future experimental plans. In this article, we discuss the possibility of identifying protein domains by comparing some 'classic' protein domains in plants and other model organisms.

Protein domains: recent definition

One of the major revelations of protein crystallography is that many proteins are made up of distinct structural units known as domains [7]. In most cases, these spatially defined domains are associated with assigned functions. Domains are usually defined by the presence and/or the combination of various protein secondary structures or specific folding of a polypeptide chain (Fig. 1a). The conservation of such protein structures is usually also seen in the conservation of the amino acid sequence of the domain (Fig. 1b). However, in many cases, the homology within a protein domain family could be low and might only be conserved in some crucial functional residues (such as active sites and ligand-binding sites) (Fig. 1b). This explains why the presence of a particular domain in a protein could be overlooked during searches, as a result of 'masking' by the unconserved region. In addition, biochemical and molecular studies during the past 20 years have suggested that many protein domains bind defined, stereotypical ligands (Fig. 1a). Moreover, many proteins are made up of more than one domain, and a particular domain might be present in various proteins of diverse cellular functions (Fig. 1c). The assembly of these structurally units or the so-called 'LEGO® blocks' could define the cellular responsibility of a protein. Hence, there is an emerging argument that protein domains are fundamental units of structure and function that make up proteins [7].

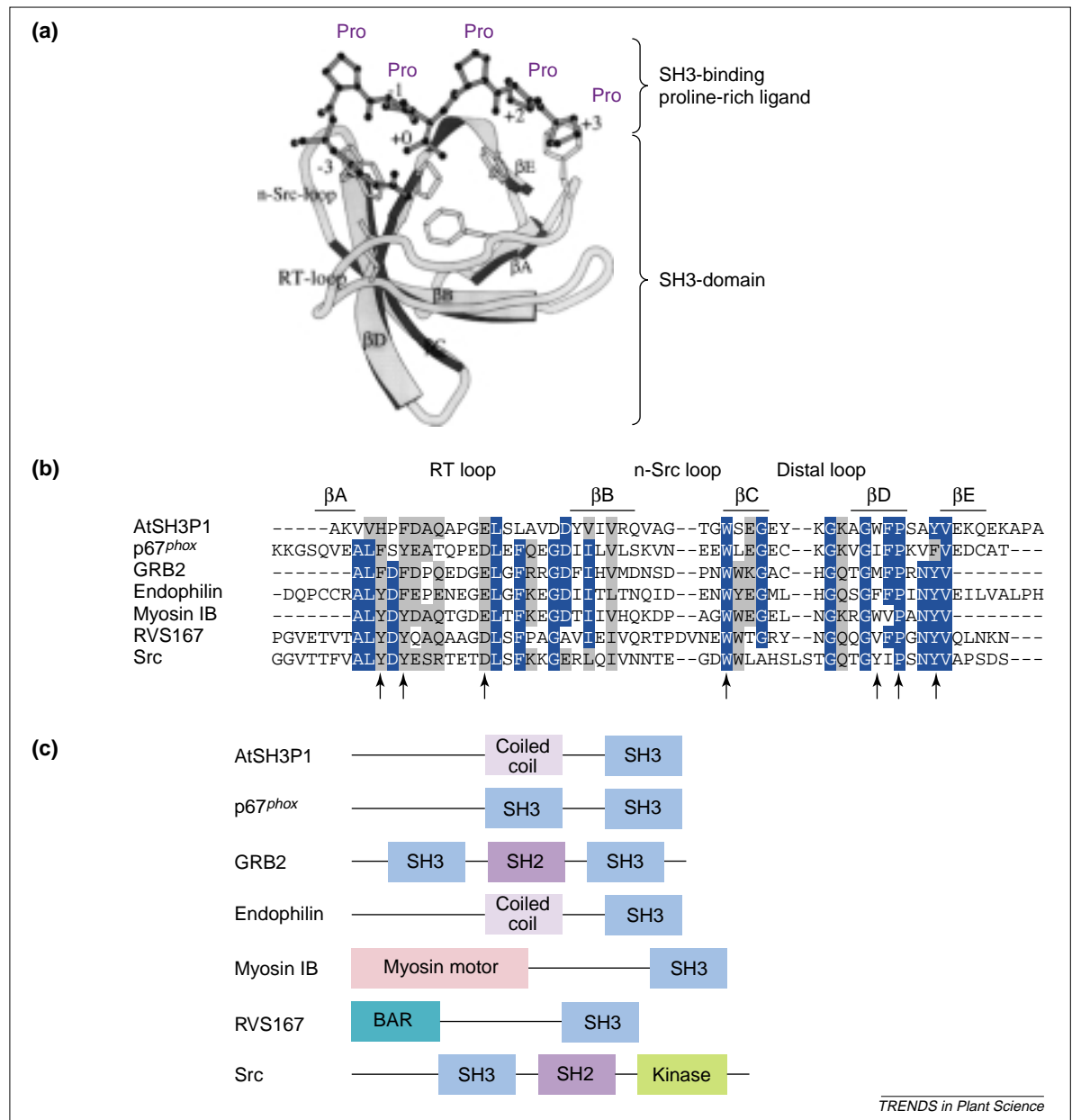


Fig. 1. The Src-homology 3 (SH3) domain as a 'classic' example of a protein-binding domain. (a) Structure of an SH3 domain coupled to a PxxP peptide. A typical SH3 domain has two anti-parallel β -sheets (β A and β B) perpendicular to each other that form a hydrophobic pocket (bottom) that could harbor the proline-rich ligand (top). Modified, with permission, from Ref. [25]. (b) Multiple sequence alignment of SH3-containing proteins including *Arabidopsis* AtSH3P1, human NADPH-oxidase subunit p67^{phox}, human growth-receptor-bound-protein 2 (GRB2), human endophilin, myosin heavy-chain IB of *Acanthamoeba castellanii*, yeast RVS167 and human Src. Sequence corresponding to the secondary structures is marked. Amino acid residues implicated in ligand binding are indicated with arrows. Grey boxes represent conserved amino acid residues and blue boxes represent identical amino acid residues. (c) Domain arrangement of the SH3-containing proteins from (b). Abbreviations: BAR, BIN1-amphiphycin-Rvs167 domain; SH2, Src-homology 2 domain.

Most proteins domains defined to date [7] are concerned with intracellular signaling. The first wave of protein crystallography identified several key catalytic domains and included the kinase domain (not discussed here because of the vast amount of

literature for plant catalytic domains). Protein domains of the second generation, of which there are hundreds, are mainly noncatalytic. According to the type of ligand, domains can be divided into protein binding, lipid binding and nucleic acid binding (Table 2).

Protein-binding domains

Two of the most studied protein-binding domains stemmed from the *src* oncogene products. In the late 1980s, these noncatalytic domains [Src-homology 2 (SH2) and Src-homology 3 (SH3)] were identified in the non-kinase-coding region of this cytoplasmic tyrosine kinase by their presence in non-kinase-like and noncatalytic proteins with diverse cellular functions [8]. Members within each domain family are structurally conserved and bind to a short, defined peptide in a 'lock-and-key fashion', which is crucial for most protein-binding domains. In particular, the SH2 domain recognizes the phosphotyrosine residue and

Table 2. Number of genes predicted to possess known protein domains in various model organisms^a

	<i>Arabidopsis thaliana</i>	<i>Caenorhabditis elegans</i>	<i>Drosophila melanogaster</i>	<i>Homo sapiens</i>	<i>Saccharomyces cerevisiae</i>
Predicted genes	25 498	19 099	13 601	~30 000	5651
Protein binding					
PDZ	21	84	104	298	2
PTB	0	23	33	68	0
SH2	2	79	61	215	1
SH3	8	74	136	438	25
WW	14	23	38	104	6
Lipid binding					
ENTH	30	15	9	18	8
PH	43	106	118	446	29
PX	9	11	23	73	15
Nucleic acid binding					
AP2	198	0	0	0	0
FH	0	22	24	65	5
HALZ (homeobox)	143	137	208	323	10
MADS	141	3	7	6	4

^aNumbers predicted by architecture analysis of SMART [12] without amendment. Because SMART returns all GenBank entries with sequence containing a particular domain, redundancy exists. The number of hits presented by each domain in this table might therefore be slightly higher than that suggested in other recent publications (e.g. [19]) in which redundant records were eliminated.

three to five neighboring residues of a receptor tyrosine kinase (RTK). This is important for recruiting downstream SH2-containing effectors upon RTK activation. By contrast, the SH3 domain binds to a short proline-rich peptide with a core PxxP signature (where x represents any amino acid). The SH3 domain is present in many scaffolding proteins participating in tyrosine-kinase signal transduction, as well as other processes such as vesicle trafficking.

SH2 domains have not been found in higher plants. Only two entries (Table 2) encoding an SH2 domain could be predicted from the *Arabidopsis* genome using programs such as SMART [9]. Although a short segment within the *Arabidopsis GAI* (*Gibberellin Insensitive*) gene and its maize and wheat orthologs was shown to be weakly related to animal SH2 domains [10], there has been no indication that the *GAI* gene product binds phosphotyrosine. It is likely that the absence of this important signaling domain in plants is closely related to the lack of RTKs in *Arabidopsis* [1]. In support of this observation, no phosphotyrosine-binding (PTB) domain (another major mammalian signaling domain) is present in *Arabidopsis*.

By contrast, the SH3 domain has been recently identified in an *Arabidopsis* protein family implicated in vesicle trafficking [11]. The ligand preference of the plant SH3 domain appears to be similar to that in other species. Interestingly, like yeast, significantly fewer SH3 domains could be predicted in *Arabidopsis* genome than in those of animals (Table 2), and all of the *Arabidopsis* SH3-containing proteins appear to be members of a single protein family. No proteins similar to other eukaryotic signaling SH3-containing proteins (such as growth-receptor-bound protein 2) have been found. In addition, no cytosolic subunits of

NADPH oxidase (p40^{phox}, p47^{phox}, p67^{phox}), another major SH3-containing protein group, could be predicted from the *Arabidopsis* genome. An interesting inference from the comparison of such functional domains as SH3 is that, although housekeeping events such as vesicle trafficking can be conserved among various organisms, processes such as signal transduction can be divergent. Hence, each organism might use its own repertoire of domains in the non-housekeeping processes.

Other well-documented protein-binding signaling domains such as the WW domain (which also binds a short proline-rich peptide) and the PDZ domain are present in *Arabidopsis*. However, neither their binding specificity nor their cellular role has been implicated experimentally. In addition, it is yet to be discovered whether the specific animal or yeast signaling pathways associated with these domains are also present in plants.

Lipid-binding domains

Far fewer lipid-binding domains have been recognized than protein-binding domains. However, these domains have aroused a rapidly growing interest because of their importance in the precise targeting of proteins involved in signaling and other cellular processes. The pleckstrin-homology (PH) domain, the 'classic' lipid-binding domain, was first identified in the cytoskeletal protein pleckstrin [12,13]. Like SH3 domains, PH domains are present in proteins of diverse functions, including the Btk tyrosine kinases, phospholipase C, phosphoinositide 3-kinase, the SOS guanine-nucleotide-exchange factor and the GTPase dynamin [8].

A distinct feature of PH domains is that they can bind to either proteins or phospholipids. *In vitro* binding

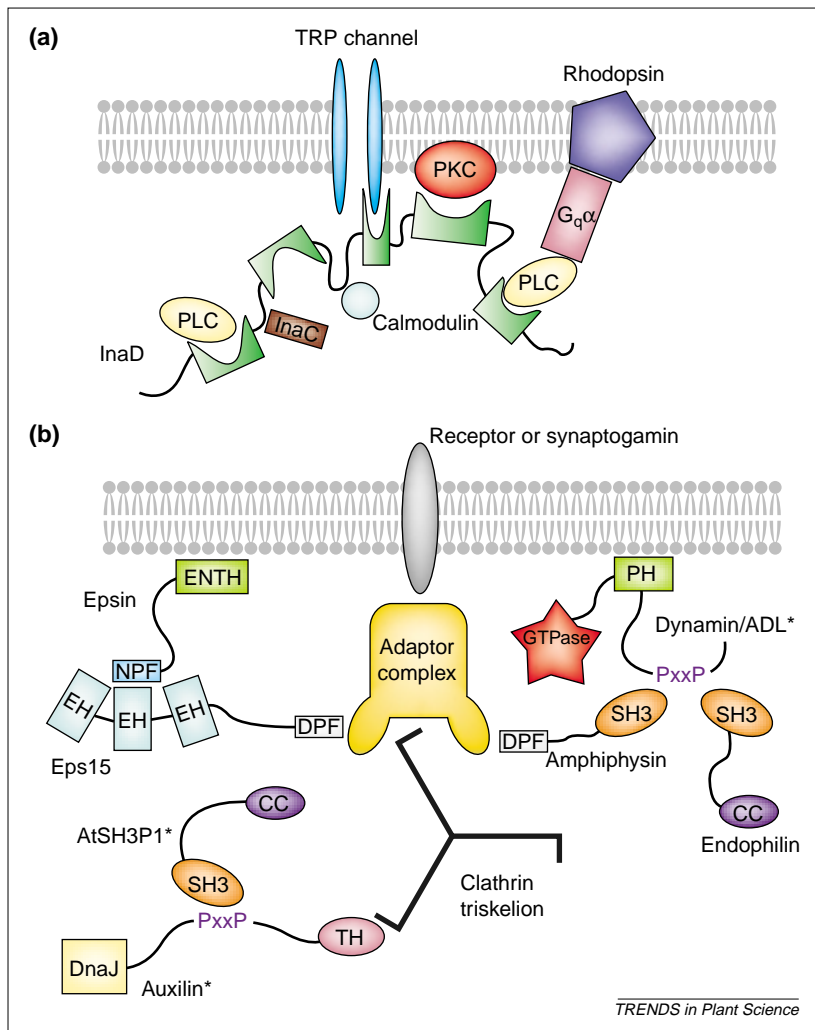


Fig. 2. Involvement of modular proteins or 'LEGO' blocks in cellular processes. (a) The *Drosophila* InaD protein contains five PDZ domains (green) that recruit various signaling proteins that control TRP-channel activity during phototransduction. Abbreviations: G_qα, G_q-protein α subunit; PKC, protein kinase C; PLC, phospholipase C. (b) Modular accessory proteins provide enzymatic activity or scaffolding that regulate the attachment of clathrin and its adaptor proteins to membrane receptors. Protein- and lipid-binding domains (Table 2) are present in these accessory proteins. Asterisks represent proteins studied in plants. Abbreviations: CC, coiled-coil domain; DPF, Asp-Pro-Phe motif; EH, epsin-homology domain; ENTH, epsin-N-terminus-homology domain; GTPase, dynamin GTPase domain; NPF, Asn-Pro-Phe motif; PH, pleckstrin-homology domain; PxxP, SH3-binding proline-rich motif; SH3, Src-homology 3 domain.

assays suggested that the PH domain binds to the β or γ subunits of heterotrimeric G proteins [14]. By contrast, PH domains are the first and the most studied protein-lipid-interaction domains. Most of the PH domains identified to date bind phosphoinositides but, because of the low sequence homology between various PH domains, the different PH domains have different specificities to each type of phosphoinositide [15].

A few PH-domain-containing proteins have been identified in plants. A carrot phosphatidylinositol 4-kinase was shown to contain a PH domain with affinity for phosphatidylinositol 4-monophosphate [16]. More recently, two *Arabidopsis* dynamin-like proteins, ADL3 [17] and ADL6 [18], have been identified that each has a PH domain. Other common lipid-binding domains such as the Phox (PX) domain and the epsin-N-terminus-homology (ENTH) domain could

also be predicted from the *Arabidopsis* genome but they have not been studied extensively.

Nucleic-acid-binding domains

Although many eukaryotic protein-binding and lipid-binding domains could be identified in *Arabidopsis* and other higher plants, this is not the case for nucleic-acid-binding domains. By contrast, ~45% of the *Arabidopsis* transcription factors are unique to plants and many of the DNA-binding domains that make up these proteins are not present in other eukaryotes [19]. One example is the DNA-binding domain of the large transcription factor family AP2/EREBP (APETELA2/ethylene-responsive-element protein), which has a unique three-stranded β-sheet structure and associates with a non-palindromic sequence [20]. In addition, some DNA-binding domains, such as the MADS box (Table 2), are more highly represented in plants than in other eukaryotes. In many cases, the transcription factors containing such domains also consist of different protein-domain combinations.

Although structural work aimed at elucidating the relationships between the plant-specific nucleic-acid-binding domains and their DNA ligands is scarce, transcription factors remain among the most studied plant proteins, and many of these domains were defined according to mutational and functional analysis [21]. In particular, AP2/EREBP proteins play crucial roles in floral development and ethylene responses, and MADS-box proteins are involved in floral-organ identity determination [22]. The predicted structures and the associated functions of these DNA-binding domains suggest that plants developed specific regulatory mechanisms at the transcriptional level for various unique physiological and developmental processes.

Multidomain or modular proteins: the LEGO® analogy

The functions associated with each protein domain are as diverse as their structure. By the addition of extra domains, the physiological complexity of a protein could be greatly enhanced. In most eukaryotes, many proteins involved in processes such as signaling, gene regulation and vesicle trafficking are modular or multidomain in nature (Fig. 2). In the case of signal transduction, modular proteins provide coordination (Fig. 2a). For example, in addition to protein-binding domains, many adaptor proteins involved in signaling pathways in the mammalian systems contain an SH2 domain, a PTB domain or a lipid-binding domain. The phosphotyrosine- or lipid-binding domains help to recruit downstream effectors to the membrane, mediated through the other protein-binding domains of the adaptor proteins. Moreover, the specificity of domain-ligand interaction can also provide the 'lock-and-key' mechanisms that act as specific switches for the signaling pathways [23]. In the case of gene regulation, the DNA-binding domain of many transcription factors lies next to a protein-binding

domain. Oligomerization through this protein–protein interaction could act as a switch for the transcription activation, mediated through the DNA-binding domain [21]. Finally, many modular proteins play the role of scaffolding in various signal-transduction pathways and vesicle trafficking (Fig. 2b). For example, in clathrin-mediated endocytosis, a few accessory proteins associated with the clathrin coat contain one or more SH3 domains and other protein-binding domains such as the epsin-homology (EH) domain. These proteins not only help to recruit other proteins of the complex that regulate clathrin coat formation but also contribute to the maintenance of the structural integrity of this uniform protein coat [24].

Perspective

In this article, we have only characterized a few ‘classic’ examples of protein domains. Major efforts in future plant-protein-domain research should involve the systematic identification of well-defined protein domains (several hundreds) from other eukaryotes, the identification of protein domains unique to plants and the establishment of database(s) specialized in plant protein domains. At the database or computational level, the identification of known protein domains in other model organisms should contribute to a better annotation of the genome database of *Arabidopsis* (which has recently been expanded by the addition of annotations based on InterPro) as well as other model plants. This will aid in the assignment of function to many of the proteins with unknown functions. At the experimental level, the understanding of domain–ligand interactions

could become a crucial ‘reference guide’ when performing protein-networking experiments such as the protein–protein interaction trap. Technically, the use of protein domains as ‘baits’ has been shown to be as powerful as the use of full-length proteins and might provide higher specificity [11]. Also, knowledge of the precise ligand preference of a particular domain might help us quickly to eliminate any false positives obtained from an interaction screen.

Although most immediate research will be based on ‘forward’ domain prediction approach (domain-homology search followed by ligand recognition), ‘reverse’ approaches could be important for identifying novel plant protein domains. For example, high-throughput screening (such as *in vitro* binding assays on protein chips and phage displays) using a specified ligand (peptide, protein, lipid or DNA) could characterize the consensus sequences of novel plant domains. These, together with an increased interest in plant-protein-structure studies through crystallography and nuclear magnetic resonance should help to build up the repertoire of plant-specific protein domains. To cater for such an increase of information, a database dedicated specifically to plant protein domains would be highly desirable for more-sensitive domain scanning of novel plant-gene products.

Note added in proof

The Center for Eukaryotic Structural Genomics (University of Wisconsin-Madison) has recently established a service to elucidate the 3D-structure of *Arabidopsis* proteins. For more information, please visit the website (<http://www.uwstructuralgenomics.org/>).

References

- The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
- Bouché, N. and Bouchez, D. (2001) *Arabidopsis* gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.* 4, 111–117
- Apse, M.A. *et al.* (1999) Salt tolerance conferred by overexpression of a vacuolar Na⁺/H⁺ antiport in *Arabidopsis*. *Science* 285, 1256–1258
- Tena, G. *et al.* (2001) Plant mitogen-activated protein kinase signaling cascades. *Curr. Opin. Plant Biol.* 4, 392–400
- Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
- Henikoff, S. and Henikoff, J.G. (1994) Protein family classification based on searching a database of blocks. *Genomics* 19, 97–107
- Bork, P. *et al.* (1997) Cytoplasmic signaling domains: the next generation. *Trends Biochem. Sci.* 22, 296–298
- Pawson, T. (1995) Protein modules and signalling networks. *Nature* 373, 573–580
- Schultz, J. *et al.* (1998) SMART a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5857–5864
- Peng, J. *et al.* (1999) ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature* 400, 256–261
- Lam, B.C.-H. *et al.* (2001) Role of SH3 domain-containing proteins in clathrin-mediated vesicle trafficking in *Arabidopsis*. *Plant Cell* 13, 2499–2512
- Haslam, R.J. *et al.* (1993) Pleckstrin domain homology. *Nature* 363, 309–310
- Mayer, B.J. *et al.* (1993) A putative modular domain present in diverse signaling proteins. *Cell* 73, 629–630
- Touhara, K. *et al.* (1994) Binding of G-protein $\beta\gamma$ -subunits to pleckstrin homology domain. *J. Biol. Chem.* 269, 10217–10220
- Lemmon, M.A. and Ferguson, K.M. (2000) Signal-dependent membrane targeting by pleckstrin homology (PH) domains. *Biochem. J.* 350, 1–8
- Stevenson, J.M. *et al.* (1998) A phosphatidylinositol 4-kinase pleckstrin homology domain that binds phosphatidylinositol 4-monophosphate. *J. Biol. Chem.* 273, 22761–22767
- Mikami, K. *et al.* (2000) A novel *Arabidopsis thaliana* dynamin-like protein containing the pleckstrin homology domain. *J. Exp. Bot.* 51, 317–318
- Jin, J.B. *et al.* (2001) A new dynamin-like protein, ADL6 is involved in trafficking from the trans-Golgi network to the central vacuole in *Arabidopsis*. *Plant Cell* 13, 1511–1525
- Riechmann, J.L. *et al.* (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110
- Allen, M.D. *et al.* (1998) A novel mode of DNA recognition by a β -sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* 17, 5484–5496
- Liu, L. *et al.* (1999) Transcription factors and their genes in higher plants: functional domains, evolution and regulation. *Eur. J. Biochem.* 262, 247–257
- Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol.* 3, 423–434
- Scott, J.D. and Pawson, T. (2000) Cell communication: the inside story. *Sci. Am.* 282, 72–79
- Marsh, M. and McMahon, H.T. (1999) The structural era of endocytosis. *Science* 285, 215–220
- Kuriyan, J. and Cowburn, D. (1997) Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* 26, 259–288
- Apweiler, R. *et al.* (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29, 37–40
- Bateman, A. *et al.* (2000) The Pfam protein families database. *Nucleic Acid Res.* 28, 263–266
- Corpet, F. *et al.* (1998) The ProDom database of protein domain families. *Nucleic Acid Res.* 26, 323–326
- Hofmann, K. *et al.* (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27, 215–219
- Vlahovicek, K. *et al.* (2002) The SBASE protein domain library, release 9.0: an online resource for protein domain identification. *Nucleic Acid Res.* 30, 273–275