Research

Are plant formins integral membrane proteins?

Fatima Cvrčková

Address: Department of Plant Physiology, Faculty of Sciences, Charles University, Viničná 5, CZ 128 44 Praha 2, Czech Republic.
E-mail: fatima@natur.cuni.cz

Published: 27 April 2000
Genome Biology 2000, 1(1):research001.1–001.7
The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2000/1/1/research/001
© Genome Biology.com (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

Background: The formin family of proteins has been implicated in signaling pathways of cellular morphogenesis in both animals and fungi; in the latter case, at least, they participate in communication between the actin cytoskeleton and the cell surface. Nevertheless, they appear to be cytoplasmic or nuclear proteins, and it is not clear whether they communicate with the plasma membrane, and if so, how. Because nothing is known about formin function in plants, I performed a systematic search for putative Arabidopsis thaliana formin homologs.

Results: I found eight putative formin-coding genes in the publicly available part of the Arabidopsis genome sequence and analyzed their predicted protein sequences. Surprisingly, some of them lack parts of the conserved formin-homology 2 (FH2) domain and the majority of them seem to have signal sequences and putative transmembrane segments that are not found in yeast or animal formins.

Conclusions: Plant formins define a distinct subfamily. The presence in most Arabidopsis formins of sequence motifs typical of transmembrane proteins suggests a mechanism of membrane attachment that may be specific to plant formins, and indicates an unexpected evolutionary flexibility of the conserved formin domain.

Background

Some mechanisms involved in cell morphogenesis, such as membrane vesicle transport, are conserved at least among crown eukaryotes (metazoa, fungi and plants) [1,2], whereas others, such as those involving extracellular structures or the precise roles of different Rho-like GTPases [3], are not. Yet other cellular processes, such as cytokinesis, often recruit conserved proteins to accomplish superficially dissimilar tasks (for example, budding, cleavage or phragmoplast-based cell division of plant cells) [4]. For many morphogenetic mechanisms, the question of evolutionary conservation remains unresolved because available information is limited to one or a few model organisms. For example, this is the case for the molecular mechanisms that ensure the communication between the cytoskeleton and the surface of the cell. However, the recent increase in the data available from a number of genome projects allows wide-ranging searches for homologs of known components of signaling and morphogenetic pathways. The results of such searches can lead both to experimentally testable hypotheses and to general conclusions regarding the evolution of morphogenetic processes.

Formins, also known as formin homology (FH) proteins, are proteins implicated in cellular and organismal morphogenesis of both metazoa and fungi. On the cellular level, they are involved in the establishment and maintenance of cell and/or tissue polarity [5,6], in cytokinesis [4] and in the positioning of the mitotic spindle [7]. They interact directly or indirectly with actin, profilin, Rho-like GTPases [5,6,8,9–11], the yeast Spa2 protein and septins [12,13], proteins containing SH3 or WW domains [10,14], dynein and microtubules [7,15–17]. The yeast formin homolog encoded by BNI1 is localized to the cell periphery and participates in positioning cortical actin patches towards distinct regions of the plasma membrane.
The possibility remains that plant formin homologs have a modular structure within the FH2 domain at the gene level, and that at least some of the FH2-related sequences within predicted introns are vestiges of exons lost by mutation.

Proline-rich regions corresponding to FH1 were identified in all *Arabidopsis* formins. Surprisingly, there are two such regions in *AtFORMIN* 2, 6 and 8 — a feature not observed in the non-plant formins examined (listed in Materials and methods). Neither motifs corresponding to FH3 nor coiled-coil regions flanking FH1 (common but not ubiquitous in non-plant formins [10]) were found. The structure of FH2, the overall protein size (smaller than most non-plant formins) and the domain layout of *Arabidopsis* formins therefore show possible plant-specific features (Figure 2). This idea is supported by the topology of an evolutionary tree that consistently places *Arabidopsis* formins in a branch separate from other members of the formin family (Figure 3).

As in the non-plant formins, the amino-terminal portions of all *Arabidopsis* formins are divergent, although there is 63% identity between *AtFORMIN* 1 and 4 in the overlapping parts of their sequences. Analysis of *AtFORMIN* sequences with SMART [26,27] revealed no previously characterized domains outside the FH2 region. However, putative amino-terminal membrane insertion signals (signal peptides) followed by a segment highly likely to be membrane-spanning and a variable number of possible transmembrane domains were found in *AtFORMIN* 1, 2, 4, 6 and 8. A possible membrane insertion signal was also identified in *AtFORMIN* 5 by one of the two methods used (see Materials and methods, and Figures 2, 4). The length of predicted signal peptides suggests that they may represent membrane anchors rather than secretion signals [28]. A putative transmembrane segment was also found in the apparently amino-terminally truncated sequence of *AtFORMIN* 3. In contrast, no signal peptides were found in 12 fungal and animal formins listed in Materials and methods, although transmembrane-like segments were observed in some. Surprisingly, the putative transmembrane segment lies between the two Pro-rich regions in *AtFORMIN* 2, 6 and 8. Obviously, only the cytoplasmic one of these two motifs can act as a conventional FH1 domain. Its size ranges from 106 to 423 amino acids, with proline content of 13 to 41% and multiple stretches of five to nine consecutive proline residues. This structure roughly corresponds to that of previously characterized FH1 domains [10]. Interestingly, the FH1 domains of *AtFORMIN* 2, 7 and 8 are extremely rich in serine (up to 20%) and contain stretches of up to seven consecutive serine residues.

The other proline-rich domain of *AtFORMIN* 2, 6 and 8 is predicted to be exposed to a non-cytoplasmic compartment. Given that polyproline stretches are characteristic for a class of structural cell-wall proteins known as extensins [29], it is tempting to speculate about a possible role for this domain in communication between formins and structures within the cell wall. Apart from this, few predictions of function can
be made on the basis of the sequence data. Although formins are well conserved with respect to their molecular structure, we do not know the extent of their conservation within signaling or structural modules [21]. As the relationships between protein structure, module structure and biological function are far from straightforward [30], we can at present neither prove nor exclude the possibility that plant formins contribute to similar functional modules to their animal and fungal counterparts. The question of whether these proteins have a direct role in cytokinesis, in mitotic spindle localization, or in some other cellular process, possibly involving cytoskeleton rearrangement or cell-surface growth, will have to be answered experimentally.

**Conclusions**

A systematic search of the available *Arabidopsis* genomic and cDNA sequences revealed the presence of eight genes encoding proteins that define a novel subfamily of the formin family. At least six out of eight *Arabidopsis* formins appear to be integral membrane proteins. This indicates a mechanism of membrane localization that may be specific to...
Figure 1
Alignment of the FH2 domain of selected formins and definition of the subdomain modules. Subdomain modules (a–j) are marked in color. Red dots denote the position of introns (not shown in MFORMIN, for which only mRNA sequence is available). The consensus line shows 80% consensus of the EMBL DS39866 alignment. Numbers indicate positions within the sequence and the size of unaligned insertions; residues corresponding to unambiguous consensus and/or shared by all Arabidopsis formins are highlighted. For gene terminology see Table 1 and Materials and methods.

Figure 2
Domain structure of Arabidopsis and selected yeast and animal formins. Letters denote subdomain modules within FH2 as defined in Figure 1. Only the ‘highly likely’ membrane-spanning segments are shown.
plants and functionally related to a possible role for formins in the communication between the plant cell and extra-cellular structures.

Materials and methods

Identification of Arabidopsis formin homologs and protein sequence prediction

The initial search for formin homologues in the non-redundant Arabidopsis thaliana protein (NRAT) database, performed using the PatMatch program [31,32] with the query pattern L-x-x-G-N-x-M-N, yielded three potential formin homologs — AtFORMIN1 to AtFORMIN3. AtFORMINs 2 to 8 were found by a TBLASTN 2.0 search [33,34] in GenBank, using the predicted protein sequence of AtFORMIN 1 as query (P(N) values in the range of 5.8 × 10^{-227} to 1.3 × 10^{-11}). Known members of the formin family (a human diaphanous homolog and Drosophila melanogaster capuccino) were found in the same search (P(N) values 1 × 10^{-21} and 1.3 × 10^{-13}, respectively), verifying the statistical significance of the initial PatMatch results.

Membrane insertion signals (anchors)

| Membrane peptides | Descriptions |
|-------------------|--------------|
| AtFORMIN1 78 AVLITAASTLLVAVGFFFCCLQ 98 |
| AtFORMIN2 157 TAVISIAAAALLLAFFVIFLI 177 |
| AtFORMIN3 157 AVASTAVLFVVMALFFVFCF 177 |
| AtFORMIN4 80 AVLITAASTLLVAVGFFFCFLVH 100 |
| AtFORMIN5 109 IVISVGTIILMLSLAFLLY 129 |
| AtFORMIN8 108 LLIVALAASVSSALVLLAL 128 |

Figure 4

Putative membrane anchors and transmembrane domains of Arabidopsis formins. Aliphatic (I, L, V), aromatic (F, H, W, Y) and other potentially hydrophobic (A, C, G, K, M, R, T) amino acids are highlighted.

Intron positions in the genomic sequences were determined (or confirmed) using the NetGene2 server [25]. Translation of the DNA sequences was performed on the SIB ExPASy WWW server [35,36]. Only the longest predicted ORFs were subjected to further analysis.

Sequence alignment and domain structure analysis

All sequence comparisons were done on a set of 20 metazoan, yeast and plant formin sequences. These were FUGU, Fugu rubripes formin homolog gb|AAC34395.1; LFORMIN, mouse lymphocyte-specific formin gb|AAD01273; BNR1, yeast Bnr1 protein sp|P40450; BNI1, yeast Bni1 protein sp|P14832; FHOS, human formin-like protein gb|AAD39906.1; CAENO, Caenorhabditis elegans formin homolog gb|AAB42354.1; CAPPU, D. melanogaster Cap-puccino gb|AAC46925.1; P140MDIA and P134MDIA2, mouse Diaphanous homologs gb|AAC53280 and gb|AAC71771.1; DIA_DROME, D. melanogaster Diaphanous sp|P48608; CYK1, C. elegans Cyk1 assembled from gb|AAA81161.1 and gb|AAC17501.1; MFORMIN, mouse formin sp|Q05860; and AtFORMIN 1 to 8. Protein sequences were aligned with the aid of MACAW [37], using the Gibbs sampler and segment pair algorithms, BLOSUM45 matrix. Only blocks with P < 10^{-7} were considered. No homology to FH3 as defined by Petersen et al. [23] or to the amino-terminal conserved region [10] was revealed by this tool, whereas the FH2 domain was readily identified. Non-aligned parts of the sequence within the FH2 domain were adjusted manually. Consensus of the resulting alignment of FH2 (deposited in the EMBL alignment database, accession number D839866) has been calculated for each subdomain separately (see Figure 1) by the method of Brown and Lai [38,39].

The SMART program [26,27] was used to examine predicted protein sequences for the presence and location of known
sequence domains, putative secretion signals, transmembrane segments, coiled-coil motifs and low sequence complexity regions (usually representing proline-rich FH1 domains whose location was confirmed by visual inspection). Prediction of signal peptides by the neural network (NN) method [28]) was independently verified by a hidden Markov model-based (HMM) method on the SignalP 2.0 server [40,41]. Results of both methods were in agreement, with the exception of AtFORMIN5, which was predicted to be membrane-anchored by NN but cytoplasmic by HMM.

Construction of the evolutionary tree

The tree (Figure 3) was calculated from the three FH2 subdomains present in all formins studied, using programs from the PHYLIP package [42,43] version 3.573. An input file was prepared by joining subdomains a, c and h and was used to produce a bootstrapped data set by SEQBOOT with 500 sampling cycles. Distances were calculated using PROTDIST with the PAM distance matrix, and the results were used for tree construction using the neighbor-joining method [44] by NEIGHBOR. The consensus tree was determined by CONSENSE and plotted using DRAWTREE.

Acknowledgements

This work has been supported by the Grant Agency of the Czech Republic Grant 204/98/0482 and by the Czech Ministry of Education Program J13/98:11100003. I thank J. Flegr for helpful discussion.

References

1. Žarký V, Cvrčková F: Small GTPases in the morphogenesis of yeast and plant cells. In Molecular Mechanisms of Signalling and Membrane Transport. Edited by Wirtz KWA. Berlin: Springer, 1997, 75-88.
2. Sanderfoot AA, Raikhel N: The specificity of vesicle trafficking: coat proteins and SNAREs. Plant Cell 1999, 11:629-642.
3. Li H, Wu G, Ware D, Davis KR, Yang Z: Membrane-transport coat proteins and SNAREs. Plant Cell 1999, 11:629-642.
4. Field C, Li R, Oegema K: Membrane-transport coat proteins and SNAREs. Plant Cell 1999, 11:629-642.
5. Evangelista M, Blundell K, Longtine MS, Chow CJ, Adames N, Pringle JR, Peter M, Boone C: Localisation of Bni1p at the bud site and regulation of the actin cytoskeleton during polarized morphogenesis. J Mol Biol 1998, 276:126-139.
6. Zeller R, Haramis AG, Zuniga A, McGuigan C, Dono R, Davidson G, Cvršková cZ, Peter M, Boone C: Localisation of Bni1p at the bud site and regulation of the actin cytoskeleton during polarized morphogenesis. J Mol Biol 1998, 276:126-139.
7. Heil-Chapdelaine R, Adames N, Cooper JA: Formin: the connection between microtubules and the cell cortex. J Cell Biol. 1999, 144:809-811.
8. Frazer J, Field C: Actin cytoskeletons are FH proteins located in the cytoplasm of prokaryotic and eukaryotic cells. Cell Biol 1997, 177:84-98.
9. Fujiwara T, Tanaka K, Mino A, Takai Y: Interaction between the FH1 and FH2 domains of formin Bni1p. J Mol Biol 1998, 199:221-229.
10. Wawszczak A, Blundell PE, de la Pompa JL, Zeller R: The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. Genes Dev 1992, 6:14-28.
11. de la Pompa JL, James D, Zeller R: The limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. Genes Dev 1992, 6:14-28.
12. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: From molecular to modular cell biology. Nature 1999, 402 Suppl:C7-C52.
13. Miller RK, Matheos D, Rose MD: The limb deformity proteins during avian neurulation and sense organ development. Dev Biol 1995, 206:156-167.
14. Petersen J, Nielsen O, Engelbrecht J, Rouze P, Brunak S: Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. Nucleic Acids Res 1996, 24:3439-3452.
15. Schultz J, Milpetz F, Bork P, Ponting C: SMART, a simple modular architecture research tool: Identification of signalling domains. Proc Natl Acad Sci U S A 1998, 95:5857-5864.
16. SMART - simple modular architecture research tool [http://smart.embl-heidelberg.de/].
17. Nielsen O, Engelbrecht J, Brunak S: Identification of prokaryotic and eukaryotic signal peptides and prediction of cleavage sites. Prot Eng 1997, 10:6.
18. Keller B: Structural cell wall proteins. Plant Physiol 1993, 101:1127-1130.
19. Hebsgaard SM, Kornning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S: Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. Nucleic Acids Res 1996, 24:3439-3452.
20. de la Pompa JL, James D, Zeller R: The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. Genes Dev 1992, 6:14-28.
21. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: From molecular to modular cell biology. Nature 1999, 402 Suppl:C7-C52.
22. Petersen J, Nielsen O, Engelbrecht J, Rouze P, Brunak S: Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. Nucleic Acids Res 1996, 24:3439-3452.
23. Schultz J, Milpetz F, Bork P, Ponting C: SMART, a simple modular architecture research tool: Identification of signalling domains. Proc Natl Acad Sci U S A 1998, 95:5857-5864.
24. SMART - simple modular architecture research tool [http://smart.embl-heidelberg.de/].
25. Nielsen O, Engelbrecht J, Brunak S: Identification of prokaryotic and eukaryotic signal peptides and prediction of cleavage sites. Prot Eng 1997, 10:6.
26. Keller B: Structural cell wall proteins. Plant Physiol 1993, 101:1127-1130.
27. Hebsgaard SM, Kornning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S: Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. Nucleic Acids Res 1996, 24:3439-3452.
28. Schultz J, Milpetz F, Bork P, Ponting C: SMART, a simple modular architecture research tool: Identification of signalling domains. Proc Natl Acad Sci U S A 1998, 95:5857-5864.
29. SMART - simple modular architecture research tool [http://smart.embl-heidelberg.de/].
30. Nielsen O, Engelbrecht J, Brunak S: Identification of prokaryotic and eukaryotic signal peptides and prediction of cleavage sites. Prot Eng 1997, 10:6.
31. Keller B: Structural cell wall proteins. Plant Physiol 1993, 101:1127-1130.
32. PatMatch [http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl].
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215:403-410.
34. Gish W, States DJ: Identification of protein coding regions by database similarity search. Nat Genet 1993, 5:266-272.
35. Appel RD, Baloch A, Hochstrasser DF: A new generation of information retrieval tools for biologists: the example of the ExPaSy WWW server. Trends Biochem Sci 1994, 19:259-260.
36. ExPaSy Molecular Biology Server [http://www.expasy.org/].
37. Schuler GD, Altschul SF, Lipman DJ: A workbench for multiple alignment construction analysis. Prot Struct Funct Genet 1999, 36:130-132.
38. Ponting C, Aravind L: START: a lipid-binding domain in STAR, HD-ZIP and signalling proteins. Trends Biochem Sci 1999, 24:130-132.
39. Consensus server [http://www.bork.embl-heidelberg.de/Alignment/consensus.html].
40. Nielsen H, Krogh A: Prediction of signal peptides and signal anchors by a hidden Markov model. In Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB 6). Menlo Park, California: AAAI Press, 1998, 122-130.
41. Signal P v2.0b2 World Wide Web Prediction Server [http://www.cbs.dtu.dk/services/SignalP-2.0/].
42. Felsenstein J: PHYLIP — Phylogeny Inference Package (Version 3.2). Cladistics 1989, 5:164-166.
43. PHYLIP [http://evolution.genetics.washington.edu/phylip.html].
44. Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987, 4:406-425.