Armadillo repeat proteins: beyond the animal kingdom

Juliet C. Coates

Armadillo (Arm) repeat proteins contain tandem copies of a degenerate protein sequence motif that forms a conserved three-dimensional structure. Animal Arm repeat proteins function in various processes, including intracellular signalling and cytoskeletal regulation. A subset of these proteins are conserved across eukaryotic kingdoms, and non-metazoa such as Dictyostelium and Chlamydomonas possess homologues of members of the animal Arm repeat family. Higher plants also possess Arm repeat proteins, which, like their animal counterparts, function in intracellular signalling. Notably, these plant Arm proteins have novel functions. In addition, genome sequencing has identified a plethora of Arm-related proteins in Arabidopsis.

Proteins containing Arm repeats possess tandem imperfect repeats of a sequence motif of about 42 amino acids [1,2]. The 'Armadillo' nomenclature originates from the appearance of embryos that are mutant for the Drosophila segment polarity gene armadillo, the founding member of the family [3,4]. Drosophila Arm is the homologue of mammalian β-catenin, which is required both for cell–cell adhesion and for regulating gene expression during development (reviewed in ref. [5]). Many other proteins also contain Arm repeats, and all of these are thought to share a conserved three-dimensional structure (Fig. 1). A single Arm repeat consists of three α helices [2]. Tandem Arm repeats fold together and interact extensively with one another to form a right-handed superhelix of helices, which creates a surface for protein–protein interactions [2,6–8] (Fig. 1). Arm repeat proteins are structurally related to proteins containing tandem \( \text{HEAT} \) motifs, and the two protein families probably had a common phylogenetic origin [9].

Although originally characterized in animals, proteins containing Arm repeats also exist outside the animal kingdom. Arm family proteins with known functions can be divided into several categories, with different subfamilies having characteristic sequences outside the Arm domain that contribute to protein function (Fig. 2). Searches of the Pfam (http://www.sanger.ac.uk/Software/Pfam) and MATDB (http://mips.gsf.de/proj/thal/index.html) databases, as well as BLAST searches [10] using sequences of known Arm proteins, have detected at least 80 putative proteins containing tandem Arm repeats in Arabidopsis (Fig. 3), some of which have relatives in other plant species. As in other organisms, plant proteins that contain Arm repeats can be divided into several subfamilies (Fig. 3).

The presence of Arm repeat proteins in unicellular eukaryotes, animals and plants suggests that this protein family has ancient evolutionary origins. One hypothesis is that Arm family proteins have functions in plants similar to those in other organisms, and several recent studies have shown that this suggestion is partly correct. But plant Arm family proteins have been also discovered to have novel functions that correlate with the presence of plant-specific functional groups adjacent to the Arm repeat domain.

In this review, I outline the known functions of Arm repeat proteins across kingdoms and speculate on the functions of the numerous uncharacterized Arm family members detected in the completed Arabidopsis genome sequence.

Importin-α homologues are conserved in all eukaryotes

Importin-α homologues are conserved across eukaryotic kingdoms (Fig. 2). The function of importin-α is to regulate, through nuclear pore complexes, the transport of proteins into the nucleus (reviewed in ref. [11]). Proteins containing nuclear localization signals (NLSs) are recognized by the Arm repeats of importin-α (reviewed in ref. [12]). The amino (N)-terminus (non-Arm) region of importin-α binds to importin-β, which contains \( \text{HEAT} \) motifs [13]. Importin-β interacts with the nuclear pore, allowing translocation of the importin-α/importin-β heterodimer and its NLS-containing cargo into the nucleus, where the cargo is then released by the GTP-bound form of the small GTPase Ran (reviewed in refs [11,12]).

Notably, importin-α can interact with cytoskeletal components, suggesting a way in which nuclear import might be regulated. Immunolocalization studies have demonstrated that importin-α colocalizes with both microtubules and actin microfilaments in tobacco protoplasts [14], and that both tobacco and Drosophila importin-α associate with the cytoskeleton, but only when bound to an NLS [14,15]. In addition, the \( \text{Saccharomyces cerevisiae} \) importin-α, Srp1p, associates with the actin-related protein Act2p, which is required for nuclear pore structure and function [16].

Unicellular Arm family proteins and their homologues

In addition to Srp1p, a second protein with Arm repeats exists in S. cerevisiae. Vac8p forms a link between intracellular membranes and the actin cytoskeleton. Vac8p co-sediments with filamentous actin and is associated with the vacuole membrane through the lipid modification of residues located N-terminal to the Arm domain [17,18]. vac8 mutants cannot target new vacuolar membrane to the bud site during cell division [17–19]. Vac8p is also a component of inter-organelle junctions between the nuclear membrane and the vacule, where it binds to and is required for the localization of the integral membrane protein Nvj1p [20]. In addition, Vac8p is
required for the fusion of vacuolar membranes within the cell [21,22].

Nd9p, an Arm repeat protein related to Vac8p, exists in the protozoan Paramecium and is required for regulated exocytic membrane fusion between the trichocyst and plasma membranes [23]. As in Vac8p, the Arm repeat region in Nd9p is thought to be involved in protein–protein interactions, whereas the membrane interaction region lies outside the repeats [23]. Whether Nd9p interacts with the actin cytoskeleton is unknown.

Recent data suggest that the regulation of the microtubule cytoskeleton by an Arm family protein is conserved among organisms as divergent as unicellular algae and mammals. The unicellular alga Chlamydomonas requires an Arm repeat protein, PF16, for cellular motility brought about by the flagellum [24,25]. PF16 localizes to the C2 microtubule in the flagellum. The PF16 protein contains eight Arm repeats and interacts specifically with the central pair of microtubules in the nine-plus-two arrangement of microtubules found in the flagellum. PF16 is required specifically for the stability of this single type of microtubule [24,25]. Deletion analysis of PF16 has shown that all eight Arm repeats are required for its function and assembly into the flagellum, indicating that the Arm repeats form a single functional unit [24].

The mammalian homologue of PF16 is a testis-specific sperm protein, SPAG6, that interacts with tubulin in vitro and colocalizes with microtubules in vivo [26,27]. SPAG6 interacts specifically with the central pair of microtubules in sperm tails, in a manner strikingly similar to that of PF16 [28]. Male mice lacking the Spag6 gene are infertile owing to a sperm motility defect, and many Spag6−/− sperm lack the central pair of microtubules in the flagellum [28]. An interaction partner of SPAG6 is PF20, a WD40 repeat protein that is also localized to the central apparatus of sperm tails. PF20 requires interaction with SPAG6 for its localization to microtubules when both proteins are heterologously expressed in COS cells [29]. Chlamydomonas PF20 localizes to the C2 microtubule in the central pair, and pf20 mutants lack the central pair of microtubules from the flagellum [30]. Thus, both the function and the interaction partners of PF16 and SPAG6 seem to have arisen early in evolution.

Other Arm family proteins that interact with microtubules

Although no obvious homologues of PF16 and SPAG6 seem to exist in higher plants, Arabidopsis has three Arm repeat proteins that are highly likely to interact with microtubules, because they possess an N-terminal kinesin domain [31] (Fig. 3). Known plant kinesins have roles in morphogenesis and development affecting cell division and cell growth (reviewed in refs [32,33]). Presumably, the Arm repeat domain of the Arabidopsis kinesin proteins interacts with target proteins, thereby localizing them to microtubules.

It is possible that Arabidopsis kinesin Arm proteins function analogously to a complex of Arm repeat proteins that interacts with microtubules in mammalian cells. Mammalian kinesin family proteins interact with an Arm family protein called SMAP (Smg-GDS-associated protein), which in turn interacts with Smg-GDS, an Arm repeat exchange factor for small GTPases [34,35] (Fig. 2). In addition, the adenomatous polyposis coli (APC) protein can, through its own Arm repeats, interact with SMAP, and this interaction is required for microtubule clustering [36]. Notably, a null mutant in the Dicyostelium Smg-GDS homologue is defective in chemotaxis and aggregation, although the mechanism underlying this defect is unclear [37].

Cytoskeletal functions of β-catenin and related proteins

The first Arm repeat protein to be implicated in cytoskeletal regulation in multicellular organisms was β-catenin, a component of adherens junctions in animals. Adherens junctions are points of intercellular contact, where cadherin molecules at the plasma membrane are anchored to the actin cytoskeleton via the adaptor proteins α- and β-catenin [38]. β-Catenin binds directly to the intracellular tail of cadherin through its Arm repeats [39]; the N-terminus of β-catenin in turn interacts with α-catenin [40], which interacts with actin [41]. Desmosomes - specialized adhesive structures that are prevalent in tissue types such as epithelia and link keratin intermediate filaments to desmosomal cadherins – contain plakoglobin, a close relative of β-catenin [42,43]. It has been shown that intact adherens junctions containing Arm, the Drosophila homologue of β-catenin, are necessary for cellular rearrangements that take place during morphogenesis [44]. A similar requirement is found for HMP-2, a β-catenin homologue in Caenorhabditis elegans [45].

Animals possess other Arm repeat proteins with a cytoskeletal role similar to that of β-catenin (Fig. 2). Adherens junctions also contain p120 protein and its relatives (δ-catenin, plakophilins and p0071) [46]. Desmosomes also contain members of the p120 subfamily [46]. p120 can affect the activity of Rho family GTPases, which might regulate the balance between cell motility and adhesion, via the actin cytoskeleton (reviewed in ref. [47]). The p120 family of proteins seems to be restricted to vertebrates and arthropods, suggesting that this subgroup might have evolved coordinately with the diversification of specialized types of tissue in complex animals.

The cytoskeletal function of β-catenin is conserved in a multicellular non-metazoan (Fig. 2). The Dicyostelium β-catenin-related protein Aardvark (Aar) localizes to and is necessary for the integrity of actin-containing adherens junctions during the formation of fruiting bodies [48,49]. It is not known, however, whether Aar interacts with a cadherin protein, because the only cadherin-related protein of known function in Dicyostelium, DdCad1, lacks a transmembrane domain and thus a conserved region for interaction with β-catenin [50]. The fact that β-catenin-like proteins exist both in Dicyostelium and in metazoa, and that Arm proteins that interact with the cytoskeleton are present in unicellular organisms, raises the possibility that Arm repeat proteins have a conserved cytoskeletal function in all organisms, and that proteins with similar functions also exist in higher plants.

Plant cells are surrounded by a rigid cell wall and so do not possess ‘classical adherens junctions’ linking one cell membrane directly to another. Indeed, plants do not seem
to possess any close relatives of metazoan adherens junction proteins other than Arm repeat proteins. However, adjacent plant cells are linked via membrane-lined channels, plasmodesmata, which also contain actin filaments (reviewed in ref. [51]). As yet, it has not been ascertained to what extent plasmodesmata share molecular similarities with animal junctions, but it is tempting to speculate that some plant Arm repeat proteins might turn out to be plasmodesmatal components.

**β-Catenin and Wnt signal transduction**

In addition to their cytoskeletal functions, β-catenin and its homologues act as regulators of gene expression both during development and throughout adult life. These proteins can enter the nucleus in response to extracellular signals and bind to DNA in a complex with T-cell factor (TCF) transcription factors, thus altering gene expression (reviewed in ref. [52]).

In essence, cytosolic β-catenin or Arm can be stabilized by extracellular glycoprotein signals, known as Wnts in animals, that are transduced by a seven-transmembrane receptor. This stabilization involves preventing glycogen synthase kinase-3 (GSK-3) from phosphorylating β-catenin on N-terminal serines and threonines [53]. In the absence of Wnt signals, phosphorylation of β-catenin targets it for degradation by the proteasome via interaction with a protein complex that includes the F-box/WD40 repeat protein β-Trcp (Slimb in Drosophila), as well as GSK-3, Axin and APC [52]. By contrast, stabilized β-catenin is targeted to the nucleus, where it can affect gene expression. β-Catenin and TCF transcriptional targets include regulators of cell proliferation, transcription factors required during development, and cell adhesion molecules [52].

Arm repeat proteins have not been shown to have a function in gene expression or signalling in unicellular organisms but do have a conserved signalling function in Dictyostelium. The Dictyostelium β-catenin homologue Aar is required for the differentiation of spore and prespore cell types, and for expression of the prespore cell-specific gene psA [48]. As in animals, the signalling function of Aar is dependent on Dictyostelium GSK-3 [48]; however, it is not known whether Aar interacts with TCF-like proteins during this process. Notably, Aar is more similar to proteins in Arabidopsis and Oryza than to its metazoan β-catenin counterparts, both in sequence similarity (by BLAST) and due to an F-box in Aar and in the plant proteins that is not present in metazoan β-catenin (Box 1 and see below).

Other metazoan Arm repeat proteins might also have a dual cytoskeletal and transcriptional role. Mammalian p120, its relatives and the more distantly related plakophilins are all detected in cell nuclei, and the nuclear localization of p120 is regulated by the generation of alternatively spliced p120 isoforms with or without a nuclear export signal [54]. In addition, p120 interacts with Kaiso, a poxvirus and zinc-finger (POZ) transcription factor [55], which inhibits the interaction between Kaiso and DNA [56]. Of interest, Arabidopsis possesses a protein (At5g13060) that comprises Arm repeats and a ‘Broad-complex Tramtrack Bric-a-brac’ (BTB)/POZ DNA-binding domain (Fig. 3). Perhaps this protein has an analogous transcriptional role to that of the p120-Kaiso complex found in mammalian cells. Similarly, Arabidopsis At1g08320 contains a basic zipper (bZIP) domain in addition to Arm repeats (Fig. 3). bZIP domains mediate both protein dimerization and transcriptional regulation, suggesting that At1g08320 is also likely to be a transcription factor or part of a transcriptional complex.

**PHOTOPERIOD RESPONSIVE 1 functions in light and gibberellin signalling**

Recent work by Amador et al. [57] is the first to demonstrate the movement of Arm repeat plant proteins into the nucleus in response to extracellular signals. The PHOTOPERIOD RESPONSIVE 1 (PHOR1) protein of potato (Solanum tuberosum) possesses seven Arm repeats located downstream from a U-box motif (also referred to as a CPI domain [57]). U-box proteins, so called because of their homology to S. cerevisiae UFD2, are predicted to be components of the cellular ubiquitination machinery, which targets proteins for proteolytic degradation (reviewed in refs [58,59]).

PHOR1 mRNA is upregulated under short-day conditions, which are required for potato tuberization. PHOR1 mRNA also undergoes diurnal variation, with levels increasing after the transition to light. In short days, a second peak in PHOR1 mRNA expression is seen at dusk [57]. Inhibition of PHOR1 through the expression of an antisense construct results in early and increased tuberization under short-day conditions and also produces shorter stem height [57].

PHOR1 functions in the gibberellin (GA) hormone signalling pathway. PHOR1 antisense lines are less sensitive than wild-type plants to exogenously applied GA, as measured by both stem elongation and the expression of GA target genes, whereas plants overexpressing PHOR1 show an enhanced response to GA application and reduced sensitivity to an inhibitor of GA biosynthesis [57]. In transformed tobacco cells, a fusion protein of PHOR1 and green fluorescent protein (GFP) is present in both the nucleus and the cytosol. Notably, the distribution of PHOR1-GFP changes in response to GA. PHOR1-GFP translocates to the nucleus transiently in response to exogenous GA, and is predominantly cytosolic in cells treated with an inhibitor of GA biosynthesis [57]. Deletion analysis of PHOR1 domains has shown that nuclear targeting of PHOR1-GFP requires the Arm repeat domain, whereas cytosolic localization seems to be mediated by the U-box motif [57].

**ARC1 signalling in the Brassica self-incompatibility response**

A protein related to PHOR1, Arm repeat containing 1 (ARC1), has been characterized in Brassica [60]. ARC1 was identified as a specific binding protein for the kinase domain of Brassica S locus receptor kinases (SRKs). SRKs are transmembrane serine/threonine kinases required for the self-incompatibility (SI) response that prevents a plant self-fertilizing. arc-1 is expressed specifically in the stigma, the female part of the plant to which pollen containing the male gametes, binds [60]. In vitro binding of ARC1 to SRKs is dependent on phosphorylation, and
ARC1 is also phosphorylated by the SRKs themselves [60]. Inhibition of ARC1 through expression of an antisense construct results in a partial loss of SI [61], which suggests that it is a positive regulator of the SI response. ARC1 can enter the nucleus and contains functional nuclear localization and nuclear export signals [62]. Like PHOR1, ARC1 contains a U-box.

As mentioned above, U-box proteins are thought to target proteins for degradation [58,59]. Recent work has shown that ARC1 functions in vitro as an E3 ubiquitin ligase that can interact both with the ubiquitination machinery and with target proteins that become ubiquitinylated [62]. This E3 ligase activity requires the U-box and is also likely to function in vivo. ARC1 protein colocalizes with the proteasome in tobacco cells cotransformed with active SRK; this colocalization requires the ARC1 U-box.

Brassica pistils (comprising the stigmas and associated female structures) pollinated with self-incompatible pollen have increased levels of ubiquitinated proteins as compared with non-pollinated or compatibly pollinated controls. This increase in ubiquitination, however, is lost in plants expressing an ARC1 antisense construct. In addition, treating pistils with proteasome inhibitors blocks the SI response, allowing self-incompatible pollen to germinate and to grow [62]. Thus, the apparent function of ARC1 in the SI response is to target several proteins for degradation by the proteasome, specifically on activation of the SRK. It is likely that these target proteins, which usually would be required for aspects of fertilization, bind to the Arm repeats. ARC1 could shuttle some of its target proteins from the nucleus to the proteasome, or it might have an additional, independent nuclear function.

**Potential proteasomal functions for other plant Arm family proteins?**

By far the largest Arm repeat protein subgroup in Arabidopsis is the one containing relatives of PHOR1 and ARC1, which have a predicted U-box upstream of the Arm domain. Some relatives were identified by Azevedo et al. [58,59] and some by Amador et al. [57], but searches of the Arabidopsis proteome by BLAST and Pfam suggest that Arabidopsis possesses at least 35 PHOR1- and ARC1-related proteins (Fig. 3). By analogy to ARC1 and PHOR1, it is likely that Arabidopsis U-box proteins have diverse functions and interaction partners, but at least some might enter the nucleus in response to extracellular signals or mediate the regulated degradation of target proteins.

In addition to the large family of proteins with Arm repeats and a U-box, a few Arabidopsis Arm proteins contain other domains present in proteins that interact with the ubiquitination machinery [59]. At5g02880 possesses a predicted ‘homologous to the E6-AP C-terminus’ (HECT) domain (Fig. 3), and At2g44900 and At3g60350 both contain a predicted F-box – a motif of 40–50 amino acids that was originally identified in cyclin F [63–65]. Known proteins with HECT domains and F-boxes function as E3 ubiquitin ligases, which target interacting proteins for ubiquitination and subsequent degradation by the proteasome. Given that targeted proteasomal degradation of proteins seems to be integral to plant physiology and development, for example, in hormone signalling and responses to light (reviewed in ref. [59]), and because Arm repeats function as sites of protein–protein interaction, many Arabidopsis Arm proteins are likely to have important and diverse roles in targeted protein degradation throughout the life of the plant.

**Potential ‘Wnt-like’ signalling functions for plant Arm family proteins?**

It has been suggested that, other than GSK-3, there are no Wnt signalling pathway homologues in Arabidopsis [66,67]. But BLAST searches based on the Dictyostelium Aar protein sequence have identified two genes, At2g44900 and At3g60350, that encode proteins with the most similarity in structure and sequence to Aar, Arm or β-catenin of any Arabidopsis proteins (Box 1). These two proteins, for which I propose the names Arabidillo-1 and Arabidillo-2 (for ‘Arabidopsis Armadillo’) consist of a core of at least nine Arm repeats, several putative sites for GSK-3 phosphorylation and an N-terminal F-box (Box 1) [64,65]. An apparent homologue of Arabidillo-1 and Arabidillo-2 also exists in Oryza (AAG60190).

The presence of a putative F-box in Aar, Arabidillo-1 and Arabidillo-2 suggests that these proteins might target other proteins for destruction by ubiquitin-mediated proteolysis. In addition, there is evidence that F-box-containing proteins can themselves be degraded autocatalytically when bound to the ubiquitination machinery in yeast, animal cells and plants (ref. [68] and references therein). In metazoa, phosphorylated β-catenin is targeted for destruction by β-Trcp, a protein that contains an F-box and WD40 repeats [69]. Aar, Arabidillo-1 and Arabidillo-2 all possess an integral F-box, which could perhaps allow them to be targeted to the proteasome directly in the absence of a stabilizing signal, bypassing the need for the more complex destruction mechanism found in animal systems. Arabidopsis does, however, also possess two F-box/WD40 repeat proteins, At3g52030 and At5g21040 [64,65,70]. It has yet to be determined whether Arabidillo-1 and Arabidillo-2 are regulated in a manner analogous to the regulation of β-catenin – that is, either by constitutive degradation in the absence of a stabilizing signal or by GSK-3 phosphorylation.

Arabidopsis possesses ten homologues of GSK-3. Some of these are expressed in a tissue-specific manner and some have known developmental functions – for example, in flower development and hormone signalling (reviewed in ref. [66]). In addition, the Arabidopsis GSK3 homologue BIN-2 (also known as UCU1 or DWF-12) regulates the stability and nuclear localization of BES-1 and BZR-1 proteins in response to brassinosteroid signals, in a similar manner to the regulation of β-catenin homologues by Wnt signalling [71–73]. Although BES-1 and BZR-1 are not members of the Arm repeat family, it remains to be determined whether plant Arm proteins can be regulated in the same way.

**Concluding remarks**

Members of the Arm repeat protein family are present in all eukaryotic kingdoms. The structure of the Arm repeat...
domain allows these proteins to have many interaction partners and functions in the cell. Although most extensively studied in animals, some Arm family members clearly have non-animal homologues, with functions that have been conserved throughout evolution.

Recent studies have shown that plants possess many proteins containing Arm repeats. In Arabidopsis, the largest subgroup of Arm proteins is that of proteins containing an associated U-box – a motif associated with proteasomal functions. This U-box/Arm family seems to be unique to higher plants as there are no detectable proteasomal functions. This U-box/Arm family contains an associated U-box – a motif associated with the ubiquitination of target proteins during the SI response. Despite having novel functions, plant and non-plant Arm family members have similarities in their regulation – for example, their nuclear accumulation in response to extracellular signals and their regulation by phosphorylation.

With the large amount of genomic information now available, it will be possible to discover the role of the many uncharacterized Arm repeat proteins in various systems, for example, by investigating their intracellular localization and their loss-of-function phenotypes. Future studies of plant Arm repeat proteins are likely to provide insights into the mechanisms of many aspects of plant cell biology and development.

Acknowledgements

This work was supported by a Gatsby Charitable Foundation Postdoctoral Research Fellowship and was done in Jim Haseloff’s laboratory. I thank Adrian Harwood, Jim Haseloff and Mark Tester for critically reading the manuscript.

References

4 Riggelman, B. et al. (1989) Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. Genes Dev. 3, 96–113
16 Yan, C. et al. (1997) A role for the divergent actin gene, ACT2, in nuclear pore structure and function. EMBO J. 16, 3572–3586
18 Wang, Y.X. et al. (1998) Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole. J. Cell Biol. 140, 1063–1074
20 Pan, X. et al. (2000) Nucleus-vacuole junctions in Saccharomyces cerevisiae are formed through the direct interaction of Vac8p with Nvj1p. Mol. Biol. Cell 11, 2445–2457
22 Vet, M. et al. (2001) Vac8p release from the SNARE complex and its palmitoylation are coupled and essential for vacuole fusion. EMBO J. 20, 3145–3155
34 Wedaman, K.P. et al. (1996) Sequence and submolecular localization of the 115Kd accessory subunit of the...
37 Vithalani, K.K. et al. (1998) Identification of darlin, a Dictyostelium protein with Armadillo-like repeats that binds to small GTPases and is important for the proper aggregation of developing cells. Mol. Biol. Cell 9, 3095–3106
38 Ozawa, M. et al. (1989) The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. EMBO J. 8, 1711–1717
63 Bai, C. et al. (1996) SKP1 connects cell cycle regulators to the ubiquitin proteasomal machinery through a novel motif, the F-box. Cell 86, 263–274
73 Zhao, J. et al. (2002) Two putative BIN2 substrates are nuclear components of brassinosteroid signaling. Plant Physiol. 130, 1221–1229
Fig. 1. Arm repeat proteins have a conserved three-dimensional structure. Three-dimensional structures of the Arm repeat regions of mouse β-catenin (left) and yeast (S. cerevisiae) importin-α (right). Both proteins form a similar helical bundle of α helices, which are depicted as green cylinders. Importin-α is shown in a complex with the nuclear localization signal (NLS) peptide from *Xenopus* Nucleoplasm (yellow strand), which fits into the groove formed by the superhelix of helices. Figures were generated using Cn3D software (http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml) from the structures 3BCT (β-catenin, 457 amino acids [2]) and 1EE5 (importin-α, 424 amino acids [8]) in the Protein Data Bank (http://www.rcsb.org/pdb/). The N- and C-terminus of each Arm repeat region is marked.

![Fig. 1](image)

Fig. 2. Conservation of Arm repeat proteins in eukaryotes. Shown are the subfamilies of Arm repeat proteins with known functions. Green boxes indicate Arm repeats, other coloured boxes represent sequences specific to each protein subfamily. Proteins are drawn roughly to scale apart from APC, which varies from 1067 to 2845 amino acids among species. Each Arm repeat is about 42 amino acids. A brief description of the function of each protein is given on the right: functions known to be not associated with the Arm domain are listed in red; those associated with the Arm domain are listed in blue. Proteins are arranged according to their presence in various eukaryotic kingdoms: orange background, proteins found in all kingdoms; yellow background, proteins found in both animals and protists or unicellular algae; blue background, proteins found only in animals; light green background, proteins found only in plants; mauve background, proteins found only in fungi; grey background, proteins found only in protists. Although no function for importin-α in protists has been reported, there is a distinct homologue of importin-α in *Dictyostelium* (EMBL accession code AC116030), which has been included to illustrate the conservation of this protein across all eukaryotic kingdoms.

![Fig. 2](image)
Fig. 3. Arm repeat proteins in *Arabidopsis*. Putative genes encoding proteins with tandem Arm repeats have been compiled from the *Arabidopsis* genome. *Arabidopsis* Arm repeat proteins can be subdivided on the basis of their homology with each other and with proteins from other organisms. Orange background, importin-α family; yellow background, proteins with a kinesin domain; light blue background, predicted U-box-containing proteins similar to PHOR1/ARC1; dark blue background, proteins with similarity to Arm, β-catenin or Aar; grey background, proteins containing a predicted motif in addition to the Arm repeats — namely, a HECT domain (Pfam PF00632), a BTB/POZ domain (PF00651), a serine/threonine kinase domain (PF00069), a bZIP domain (PF00170), a leucine-rich repeat (LRR) domain (PF00560) and a patatin-like phospholipase domain (PF01734); mauve background, proteins with about 20 Arm repeats (two of which also have a C-terminal C2 domain); light green background, proteins containing Arm repeats but no other recognizable protein motifs. In the different categories of protein, green boxes indicate Arm repeats and other coloured boxes represent additional domains. This collection of proteins is the result of BLASTP and PSI-BLAST searches (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) done using Aar, β-catenin, Arm and Vac8p protein sequences to identify similar proteins in *Arabidopsis*, combined with searches of the Pfam database (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00514). All proteins have been cross-referenced with the MATDB database (http://mips.gsf.de/proj/thal/db/search/search_frame.html) to ascertain gene accession codes.
Box 1. *Arabidopsis* possesses proteins related to Aar, Arm and β-catenin

*Arabidopsis* proteins At2g44900 and At3g60350 (named Arabidillo-1 and Arabidillo-2, respectively) show similarity to *Dictyostelium* Aar (PSI-BLAST first iteration score, 10⁻⁸¹). A representation of these proteins is shown in Figure Ia, where green boxes represent Arm repeats, yellow circles represent putative sites for GSK-3 phosphorylation ([S/T]-x-x-x-[S/T], where S is serine, T is threonine and x is any amino acid), and grey rectangles represent predicted F-boxes.

Below this representation (Fig. Ib) is an alignment of the three F-box sequences, done first by ClustalW [74] and followed by slight refinement by eye. The amino acid positions of the sequences in the full-length proteins are shown. The Pfam F-box consensus sequence is shown in yellow.

Below the F-boxes (Fig. Ic) is an alignment of the nine Arm repeats of Arabidillo-1 and Arabidillo-2 and repeats 2–10 of *Dictyostelium* Aar done by ClustalW as above. The amino acid positions of the sequences in the full-length proteins are shown. Each repeat is outlined in green, and the positions of the three α helices (H1, H2, H3) in each Arm repeat [2] are indicated by black lines above the alignment. Identical residues between proteins are shaded in red, conserved residues are shaded in blue.