The plakin family

Conrad L. Leung[1], Ronald K. H. Liem[1], David A. D. Parry[3] and Kathleen J. Green[2,*]

[1]Department of Pathology and Department of Anatomy and Cell Biology, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

[2]Departments of Pathology and Dermatology, and the R. H. Lurie Comprehensive Cancer Center, Northwestern University Medical School, 303 E. Chicago Avenue, Chicago, IL 60611, USA

[3]Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

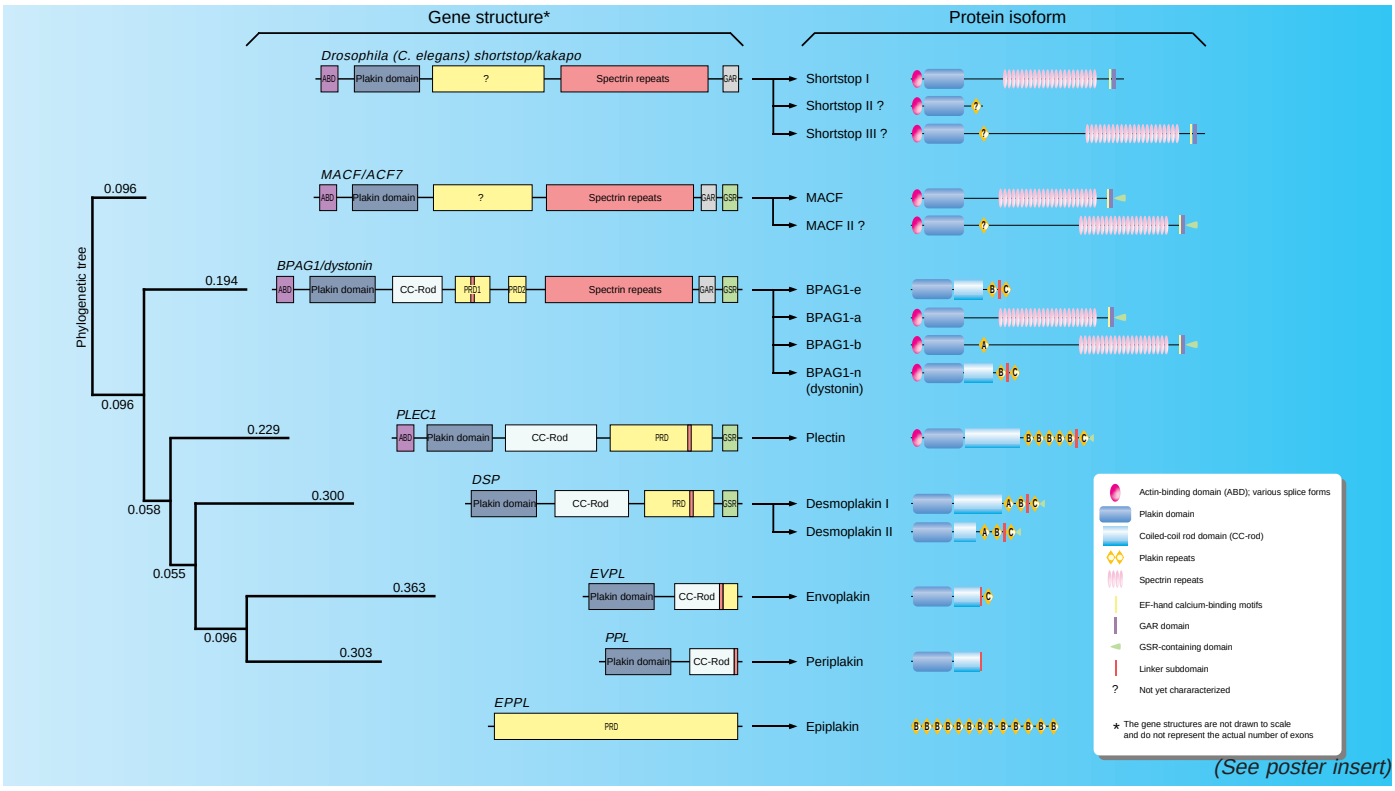
*Author for correspondence (e-mail: kgreen@northwestern.edu)

Journal of Cell Science 114, 3409-3410 (2001)

The plakins (also referred to as cytolinkers) are a family of large, modular proteins that link cytoskeletal networks to each other and to membrane-associated adhesive junctions, such as desmosomes and hemidesmosomes. Mutations in plakin family genes lead to defects in tissue integrity and function in skin, muscle and the nervous system in human and in mouse, whereas in *Drosophila* mutations lead to loss of adhesion between muscle and epidermal cells, as well as defects in neuronal outgrowth. The recent discovery of new structural features in several plakin gene loci has facilitated the construction of a family tree and model for the evolution of this family of cytoskeletal linkers.

Plakin family members are built from combinations of the following modules: a calponin-type actin-binding domain (ABD); a plakin domain harboring several α-helical bundles; a heptad-repeat-containing coiled-coil rod domain (CC-rod); a plakin-repeat domain (PRD) generally thought to have intermediate filament (IF)-binding properties, a function that in some cases requires an associated linker (L) subdomain; a spectrin repeat (SR)-containing rod domain; two EF-hand calcium-binding motifs; a Gas2-homology region called the GAR domain; and a domain containing GSR (Gly-Ser-Arg) repeats. The PRD comprises varying numbers of repeating unit subdomains each of 176 residues, which are termed A, B or C, depending on their degree of similarity to each other. The L subdomain usually resides between the B and C repeats. For some plakins, the GAR domain and GSR-containing domain can associate directly with microtubules (MTs) and define a novel MT-binding domain (MTBD). The family tree depicted here was derived on the basis of sequence alignment of the most conserved region of the N-terminal plakin domain, which is so named because it is shared by all family members except for one - epiplakin. Epiplakin is a recently described protein that lacks plakin and dimerization domains but contains 13 B subdomains characteristic of the IF-associated proteins in the family, which are connected by conserved linking sequences. These conserved linker sequences do not exhibit homology to the L subdomain described above, the latter which is indicated as a red bar in the figure. The functions of the plakin domain have not been fully elucidated. However, in desmoplakin (DP), it contains targeting sequences for desmosomes; in BPAG1 (Bullous Pemphigoid Antigen 1) and plectin, it has targeting sequences for hemidesmosomes; in certain products of the BPAG1/dystonin locus, the sequences target it to microtubules, and in periplakin, they target the cortical actin network.

To perform the phylogenetic analysis, we aligned primary sequences of the plakin domains of human plakins by using the ClustalW program. The aligned sequences were then imported into MacVector 7.0, and we built the phylogenetic tree by using the neighbor joining method. The illustrated phylogram was the only 'Best'



(See poster insert)

tree obtained, and MACF (MT actin-cross linking factor) was selected as the root. A bootstrap analysis was also employed to test the reliability of the tree. The numbers that are printed close to the branches represent relative evolutionary distances.

The ancestral gene for plakins is hypothesized to resemble the *shortstop*/*kakapo* gene found in *Drosophila* and *C. elegans*. The *shortstop*/*kakapo* locus contains exons encoding the following domains: an ABD, a plakin domain, a domain that has some similarity to the mammalian PRDs (denoted ?), an SR-containing rod domain, and a MTBD. Various splice forms of *shortstop*/*kakapo* have been reported and predicted by gene analysis programs (GenBank accession number AAG22268), but, of the forms shown in this poster, only Shortstop I has been proven to exist at the protein level. In flies, Shortstop/Kakapo is concentrated at the end of MT bundles, which associate with integrins at the basal surface of epidermal muscle attachment cells, and it may also link to cortical actin. Because *Drosophila* does not have cytoplasmic IFs, the domain that has limited homology to the PRD of vertebrates is likely to have other functions. Note that this region in *C. elegans*, an organism that does have IF, is more closely related to the PRDs of desmoplakin and plectin, particularly when compared specifically with the 38-residue motifs (see below).

MACF (originally known as ACF7) is the vertebrate homolog of Shortstop/Kakapo. MACF binds to MTs through the GAR and GSR-repeat domains and to actin through the ABD; it has been postulated to facilitate actin-MT interactions at the cell periphery and to anchor the MT network to cell junctions. Although no PRD has been reported to be present in MACF, a brief analysis of the completed human genome sequence revealed that a putative exon that has an open reading frame homologous to the PRD of other plakins is located between exons encoding the plakin domain and the SR-containing rod domain of MACF. This domain, denoted ? and probably comprising multiple repeats, is predicted to be present in the putative gene product MACF II, which has not yet been identified at the protein level.

The gene structure of *BPAG1*/*dystonin* is similar to that of *shortstop*/*kakapo*, except that two additional exons have been inserted downstream of the plakin domain and upstream of the PRD-like domain. These two exons encode a CC-rod domain and a domain that has high homology to PRD1, which here we term PRD2. The *BPAG1*/*dystonin* locus gives rise to a number of tissue-specific forms generated from alternative promoters and by differential splicing. The major neuronal isoform of BPAG1 is a homolog of MACF and is termed BPAG1-a (previously MACF2). In addition, other neuronal isoforms of BPAG1 (one of which, BPAG1-n, is shown here) are also proposed to link neuronal IFs to the actin cytoskeleton. Loss of the neuronal isoforms leads to sensory neuron degeneration in the mutant mouse *dystonia musculorum*. The epithelial form, BPAG1-e (or BP230), anchors keratin IFs to hemidesmosomal junctions and is important for epidermal integrity. The muscle specific isoform is called BPAG1-b and contains interaction domains that can potentially associate with all cytoskeletal filament networks.

The rest of the plakin family members do not contain spectrin repeats and a GAR domain that can bind to MTs, and are therefore more divergent from *Shortstop/Kakapo*. As a result, these other plakins, including plectin, desmoplakin, envoplakin and periplakin, adopt protein structures that resemble BPAG1-e, instead of MACF, BPAG1-a or BPAG1-b. However, plectin still retains the N-terminal ABD. In addition, unusual complexity within the plectin N-terminus upstream of and including the ABD (not shown here) is generated by alternative splicing. In contrast, desmoplakin exists in only two spliced variants, including a longer form found in all desmosomes and a form that has a shortened rod domain and a more restricted tissue distribution. The primary sequence of the plakin PRD is relatively conserved, although these domains vary in size among the family members. Both plectin and desmoplakin have been demonstrated to interact directly with IF types from a variety of families through their PRDs, which consist of variable numbers of subdomains each constructed of tandem repeating motifs of 38 residues organized into a larger subdomain of 176 residues. Desmoplakin contains three subdomains (A, B and C), whereas BPAG1 and plectin contain two (B and C) and six subdomains (5Bs and C), respectively. Sequences downstream of the C subdomain in desmoplakin also play an important role in promoting and regulating interactions with keratin IFs. It is thought that desmoplakin functions primarily to anchor IFs to the plasma membrane through desmosomes in epithelial and cardiac muscle cells. In contrast, plectin, which also harbors an ABD and interacts with MTs, probably through the GSR-repeat domain, is widespread in its tissue and subcellular distributions and has been called a universal cytolinker. Its absence in humans and mice leads to both skin and muscle defects.

The farthest relatives of MACF are envoplakin and periplakin, proteins found in desmosomes and epithelial cell envelopes. These two proteins are highly homologous to each other and might function as partners. The putative PRD of envoplakin contains only one C subdomain, whereas periplakin contains no complete subdomains but retains the L sequence (denoted in red) that overlaps a region in DP harboring a partial repeat and ~50 residues in plectin necessary and sufficient for binding to vimentin and keratin 8/18 IFs. Furthermore, periplakin polypeptides containing this domain are capable of associating with filament networks in transfected cells.

As the human genome is further mined, we expect that additional sequences linked to the plakin loci, as well as novel plakin family members, will emerge. An example is the recently described protein epiplakin, identified as an autoantigen in a subepidermal blistering disease, which falls into the plakin family on the basis of the high homology of its 13 repeat domains to the B subdomains in the PRDs of plakin family members. The potential involvement of epiplakin in associating with IFs or other cytoskeletal elements has yet to be determined.

Cell Science at a Glance on the Web
Electronic copies of the poster insert are available in the online version of this article at jcs.biologists.org. JPEG and PDF files (see supplemental material) can be downloaded for printing or use as slides.