Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1774 (2007) 1069–1078
Biochimica et Biophysica Acta
BBA
www.elsevier.com/locate/bbapap

Review

Structural similarities and functional diversity of eukaryotic discoidin-like domains

A. Kiedzierska, K. Smietana, H. Czepczynska, J. Otlewski *

Faculty of Biotechnology, University of Wroclaw, Str. Tamka2, 50-137 Wroclaw, Poland

Received 10 May 2007; received in revised form 2 July 2007; accepted 17 July 2007
Available online 24 July 2007

Abstract

The discoidin domain is a ~150 amino acid motif common in both eukaryotic and prokaryotic proteins. It is found in a variety of extracellular, intracellular and transmembrane multidomain proteins characterized by a considerable functional diversity, mostly involved in developmental processes. The biological role of the domain depends on its interactions with different molecules, including growth factors, phospholipids and lipids, galactose or its derivatives, and collagen. The conservation of the motif, as well as the serious physiological consequences of discoidin domain disorders underscore the importance of the fold, while the ability to accommodate such an extraordinarily broad range of ligand molecules makes it a fascinating research target. In present review we characterize the distinctive features of discoidin domains and briefly outline the biological role of this module in various eukaryotic proteins.

Keywords: β-barrel; Cell adhesion; Coagulation factor; Discoidin

The discoidin domain (DS domain, also called discoidin-like domain, discoidin motif, FA58C or F5/8C) is a structural and functional motif found in various proteins, both eukaryotic and prokaryotic [1]. It was first identified in discoidin protein (DS), discovered in 1981 in an amoeba *Dictyostelium discoideum* and described as a lectin with high affinity for galactose and modified galactose residues [1]. Subsequently, similar domains have been detected in many extracellular and membrane proteins, including blood coagulation factors, enzymes, receptors and proteins involved in neural development [2–5]. A homology search by SMART (Simple Modular Architecture Research Tool [6]) reveals the discoidin motif in more than 100 eukaryotic and 300 prokaryotic proteins.

The DS domain comprises ca. 150 amino acids and shows a considerable functional diversity. The biological role of the domain is based on its interactions with a variety of molecules, such as growth factors, phospholipids and neutral lipids, galactose with its derivatives and collagens [5,7–9]. Many of the DS domain-containing proteins are involved in cellular adhesion, migration or aggregation events, mostly associated with organogenesis (vasculogenesis and angiogenesis) and other developmental processes [3,10,11]. Examples of such proteins are neuropilins and neurexins, involved in the nervous system development [3,4,12], and tyrosine kinase receptors, DDR1 and DDR2 (discoidin domain receptor family), which regulate cell adhesion, proliferation and extracellular matrix remodeling [13]. Several DS domain proteins, for example sperm–egg adhesion protein/milk

Abbreviations: ACLP, aortic carboxypeptidase-like protein; BTB, bric-a-brac, tramtrack, broad-complex; Carb, carboxypeptidase; CFV/VIII, coagulation factors V and VIII; CNS, central nervous system; CTLH, C-terminal domain to LisH; CUB, complement factor C1s/C1r, urchin embryonic growth factor, bone morphogenetic protein homology; DEL1, developmental endothelial locus-1; DDR, discoidin domain receptor; DS domain, discoidin-like domain (FA58C, F5/8C, C2 domain of coagulation factors V and VIII); DS, discoidin; ECM, extracellular matrix; EGF, epidermal growth factor; ER, endoplasmatic reticulum; GO, galactose oxidase; LamG, laminin G; LisH, N-terminal domain in Lissencephaly 1; LH, lisH with CTLH; MAM, meprin, A5, tyrosine phosphatase (mu) homology domain; mkln1, muskelin; MMP, matrix metalloproteinse; Npn1, neuropilin-1; Nrx IV, neurexin IV; PKC, protein kinase C; RS1, retinoschisin; RTK, tyrosine kinase receptor; Sco, scospondin; SED1/MFG, sperm–egg adhesion protein/milk fat globule; SMART, Simple Modular Architecture Research Tool

* Corresponding author. Fax: +48 71 3752608.
E-mail address: otlewski@protein.pl (J. Otlewski).

1570-9639/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2007.07.007

fat globule (SED1/MFG) and DDRs, are overexpressed in breast carcinomas [13,14]. The pathological consequences of mutations in the DS domain or changes in the levels of DS domain proteins highlight the importance of this protein motif in cell functioning [15,16].

1. Arrangement of DS domains in proteins

The discoidin domain can be found in proteins either singly or in repeats [3]. A tandem of DS domains was first discovered in the C-terminal portion of the blood coagulation factor VIII [17] and later in other proteins with similar DS domain arrangement, such as DEL-1 (Developmental Endothelial Locus-1) [18] and SED1 [19].

DS domains are often accompanied by the CUB module (first observed in complement factor C1s/C1r, also called urchin embryonic growth factor or bone morphogenetic protein homology). This combination has already been identified in 53 different proteins [6]. Other domains often present in DS domain-containing proteins are: RTK (Tyrosine Kinase Receptor), Peptidase M14, laminin G2 and EGF (Epidermal Growth Factor) domains. Occasionally, DS domains are also accompanied by BTB (bric-a-brac, tramtrack, broad-complex), LisH (N-terminal domain in Lis1), kelch, PKD motifs, and VWD repeats [6] (Fig. 1).

The DS domain-containing proteins are very often implicated in protein oligomerization events. For example in retinoschisin (RS1) the octamerization required for protein function in the cell adhesion depends solely on the DS domain [15]. Other domains mediating oligomerization are frequently found within the same polypeptide chain as the DS domain, which may result in enhanced binding strength. A cooperation between the DS module and a β-propeller region is important in self-association of muskelin molecules [20]. The dimerization required for signal transduction and/or ligand recognition [9] can depend on ligand binding by DS domain, as was observed in case of DDR1 and collagen-mediated dimer formation [21].

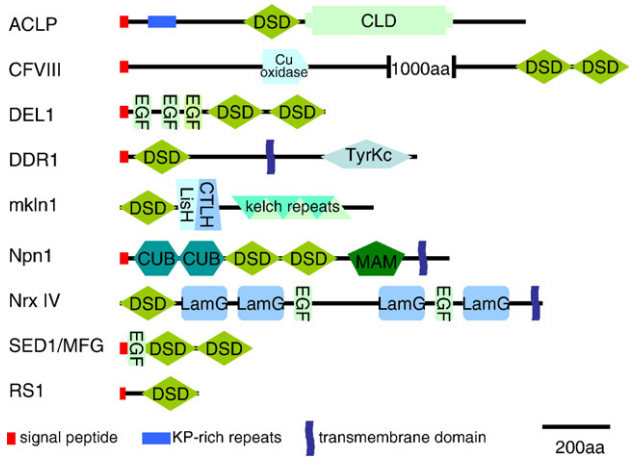

Fig. 1. Domain arrangement in proteins containing DS domain (DSD). Colors and shapes represent different domain types.

2. Similarities between DS domains

The DS domain is composed of a highly variable N-terminal fragment of about 40 amino acids and a relatively conserved C-terminal part. Sequence analysis of known DS domains specifies 47 highly conserved residues (Fig. 2). These include two cysteines forming the disulphide bridge that links the domain's N- and C-terminus in most eukaryotic and some of prokaryotic DS domains. Another disulphide bond is located near the C-terminus of the second DS domain in SED1 protein. A structural homology study has shown that tryptophanes are the most conserved amino acid residues throughout the DS domain family [22]. This may be connected with their role in maintaining a stable hydrophobic core of the DS domain β-barrel. Other highly conserved residues are glycines and prolines responsible for the formation of β-turns.

The high degree of sequence similarity in DS domains in prokaryotes and eukaryotes suggests that all these domains evolved from a common ancestor. The variety of proteins with the DS domain can be attributed to multiple transmissions of this module between prokaryotes and eukaryotes [3] (Fig. 3). Probably, each DS domain evolved to best accommodate its natural ligand, hence the diversity is seen mostly within the binding region. It could also be connected with the fact, that there is no published data on DS domains being able to bind more than one type of partner molecules.

There are very few studies concerning DS domains binding properties. The dissociation constants measured for the interactions between DDRs and collagen [23] and between discoidin and galactose [24] are within the range of 10^{-6} to 10^{-8} M. These reports were not confirmed for other DS domain–protein and DS domain–sugar interactions, since most reports are based only on qualitative assays, like immunoprecipitation [25], ELISA [26], immunoblotting [9], surface plasmon resonance [21]. It is also unclear, whether lipids and their derivatives can be bound with similarly high affinities.

The DS domain sequences often contain postranslational modification sites, enabling their phosphorylation (DDR [27]), sumoylation (muskelin [Kiedzierska A. and Czepczynska H., unpublished data]) and/or glycosylation (factor VIII [28]). The lack of conservation observed within the modification sites can be explained by the substantial diversity of DS domains, as well as their presence in both intra- and extracellular portions of the proteins. The biological role of the modifications and their possible regulatory function remains mostly unelucidated.

3. Structure–function relation in DS domains

The first structural information on a DS family protein was the crystal structure of galactose oxidase (1gof). To date, nineteen X-ray or NMR structures of DS domains have been deposited in the PDB. The DS fold consists of eight β-strands arranged in a compact β-barrel and three flexible loops ("spikes") forming ligand binding surface (Fig. 4A). All eukaryotic DS domains have their termini close together, enforced by the disulphide bridge formed between the two highly conserved N- and C-terminal cysteines. Such architecture is common for modules often found in multidomain proteins, as it facilitates domain shuffling and