

Minireview

The WD-40 repeat

Loesje van der Voorn and Hidde L. Ploegh*

Division of Cellular Biochemistry, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received 25 May 1992; revised version received 15 June 1992

An amino acid sequence motif, called the WD-40 repeat, has been found as a repeat in a large variety of proteins that do not share any obvious functional properties. At present, the function of the repeated motif is not known for any of these proteins. Interestingly, recent experiments in yeast indicate that several proteins containing the WD-40 repeat are genetically associated with members of the TPR-family, a protein family that is characterized by the presence of another repeated motif of unknown function: the tetratricopeptide repeat. It is conceivable that proteins containing the WD-40 repeat interact physically with members of the TPR-family via their respective repeated motifs.

WD-40 repeat; G protein β -subunit; Protein family; TPR-gene family

Many proteins contain repeated motifs that may be considered as building blocks, representing structural elements or functional domains. The deduced amino acid sequence of the β -subunit of heterotrimeric G-proteins revealed that, except for the first 50 N-terminal amino acids, the β -subunit is comprised of 7 segments that show sequence similarity to each other and are arranged in tandem [1]. This repeated motif, referred to as the WD-40 repeat [2], is characterized by the presence of a number of amino acids, conserved not only with respect to the type of side chain, but also in their spacing. A similar motif has been found in several other proteins. Based on the presence of the WD-40 repeat, these proteins may constitute a WD-40 family.

Members of the WD-40 family include: (i) the β -subunit of guanine nucleotide regulatory proteins (G-proteins), a component of the heterotrimeric complex that transduces signals from transmembrane receptors to a variety of second messenger generating effectors [1,2]; (ii) STE4, a functional G-protein β -subunit homologue in yeast (*Saccharomyces cerevisiae*), involved in a signal pathway that controls the response to mating pheromone [3]; (iii) CDC4, a component of the yeast nuclear cytoskeleton, required at the late G₁/S phase boundary of the cell cycle [4,5]; (iv) CDC20, a yeast gene product required for several microtubule-dependent processes at multiple stages in the cell cycle [6–8]; (v) Enhancer of Split, the product of one of the neurogenic genes in

Drosophila [9]; (vi) 12.3, the product of a gene in the chicken MHC-locus [10]; (vii) PRP4, a stable component of yeast U4/U6 small nuclear ribonucleoprotein particle (snRNP) [11–13]; (viii) PRP17, another protein involved in pre-mRNA splicing in yeast [13]; (ix) AER2/TUP1, a transcriptional repressor in yeast [14–15]; (x) MS11, a negative regulator of the RAS-cAMP pathway in yeast [16]; (xi) coronin, a component of the actin/myosin complex of the slime mold *Dictyostelium discoideum* [17]; (xii) PWP1, a yeast protein, containing periodic tryptophan residues [18]; (xiii) Clbp, a *Chlamydomonas* protein of unknown function [19]; (xiv) MAK11, the apparently membrane associated product of an essential yeast gene, necessary for the maintenance of killer M1 double-stranded RNA [20]; and (xv) AAC3, the deduced product of a developmentally regulated transcript in *Dictyostelium discoideum* that contains long AAC repeats (reading frame specifies glutamine in this case) [21].

A recent search of Protein Sequence Databases (SWISS, release 18.0 and NBRF/PIR, release 29.0) [22,23] identified only one other protein containing more than one region homologous to the WD-40 repeat, namely, ligninase, a lignin peroxidase, secreted by the fungus *Phanerochaete chrysosporium* [24,25]. It should be noted, however, that the second motif in ligninase displays a rather weak homology to the WD-40 consensus sequence. An alignment of all WD-40 repeats identified and a WD-40 consensus sequence based on this alignment are shown in Fig 1.

The WD-40 repeat can be divided in 2 relatively conserved elements, A and B, spaced by regions variable in both sequence and length (Fig. 1A). The most important features of the repeat are the LxGH in part A and

*Present address: Department of Biology, MIT Center for Cancer Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

Correspondence address: H.L. Ploegh, Department of Biology, MIT Center for Cancer Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. Fax: (1) (617) 2539 891.

A.WD-40 repeat
consensus

	(A)	(B)
(n1)	LxGHxxxIxkx Φ x δ	(n2) Φ SGGxDxx Φ xIWD δ
	F L	TAA N C LFN
	V	S VY

B.

CDC4	(377-409)	376	LRGHMTSVITCLOF.EDNYVITGADDKMIRVYDS
CDC4	(418-450)	8	LSGHDGGVWALKYA.HGGILVSGSTDRTVRVWDI
CDC4	(459-494)	8	EEGHNSTVRCIDIV*NIKYIVTGSRENTLHVWKG
CDC4	(526-557)	31	LRGHMASVRTVSG..HGNIYVSGSYDNTLIVWDV
CDC4	(566-599)	8	LSGHTDRYSTIYDHERKRCISASMDTTIRTWDL
CDC4	(628-659)	28	LQGHNTALVGLIRL..SDKFLVSAADGSIKRWDA
TB	(51-84)	50	LRGHLAKIYAMHWGTD SRLV SASQDGKLI LWD S
TB	(112-126)	8	VACGGLDNICSTYNI
TB	(139-171)	8	LAGHTGYLSCCREL.DDNQIVTSSGDTTCALWDI
TB	(180-213)	8	FTGHTGDVMSISLAPDTRLFVSGACDASAKLWDV
TB	(222-255)	8	FTGHESDINATCFEPNGNAFATGSDBATCRIFDI
TB	(285-299)	10	LLAGYDDFNQNVWDA
TB	(308-340)	8	LAGHDNRVSCLGVTDDGMAVATGSDWDSFEKLNN
STE4	(88-121)	87	LKGNHNNKISDFRWSRDSKRI LSASQDGFMLIWD S
STE4	(149-163)	8	VASAGLNNNCTIYRV
STE4	(177-209)	13	PKGHTCYISDIET.DNAHILTAGSDMTCAIWDI
STE4	(243-257)	14	FASCGSDGYTYIWD S
STE4	(285-299)	8	IVAGSDNGAIINMYDL
STE4	(364-378)	45	MYSCYTDIGGVVWDV
STE4	(387-420)	8	LECHGGRVTGVRSSPDGLAVCTGSDWDSMKIWD S
12.3	(11-45)	10	LKGNHNGWVTOIATT*FPDMILSASRDKTIIMWKL
12.3	(59-92)	13	LRGASHFVSDVVISSDGGQFALSGSWDGTIRLWDI
12.3	(101-134)	8	FVGHTKDVLSVAFSSDNQIVSGSRDKTIKLNWT
12.3	(165-179)	11	IVSCGWDKLYKVWNL
12.3	(188-221)	8	HIGHTGYLNTVTVSPDGSICASGGKDGQAMLWDI
12.3	(247-261)	6	WLCAATGPSIKTWDL
12.3	(298-311)	17	IFAGYTDNLVRVWDV
E(sp1)	(429-461)	428	TLSHGEVVCAYTIS.NPTRYVYTGGKGCYKIVWDI
E(sp1)	(494-508)	13	LIVGGEASNISTWDL
E(sp1)	(538-552)	10	CFSCCSDGNIAVWDL
E(sp1)	(561-594)	8	EQGHTDGASCIDISPDGSRILWTGGLDNTVRSWDT
E(sp1)	(644-676)	49	LHLHESCVLSLRFAACGKWFVSTGKDNLINAWRT
PRP4	(261-294)	260	IVGHERRISDVKYHPSGKFISASHDMTWRLWDA
PRP4	(303-336)	8	QEGHDKGVFSLSFQCDGSLYC9GGHPSLSMLWDI
PRP4	(345-378)	8	LAGHSKPIYTVAWSPNGYQVATGGGCGIINWDT
PRP4	(410-424)	12	LVSCGYDNLNVYSS
PRP4	(433-465)	8	LAGHTDKIISLDSNNSHFLVSGGWERSIKLWN
AER2	(458-472)	457	LATGAEDRLIRLWDI
AER2	(481-514)	8	EQGHEQDIYSLDYFPGDKLVSGSGERTYRVWDI
AER2	(542-556)	8	LAAGSLDRAVRWDS
AER2	(572-605)	15	GTGHKDSVYSVVEPTRDGSVVSGLDRSVKLNWL
AER2	(626-659)	20	YIGHKDFVLSVATTQND EYLSGSKLRGYLFWDK
AER2	(667-707)	7	LQSHRNSVISVAVA*EYNVFATGSGDCKARTWKY
COR	(75-109)	74	FNGHKSAYLDIAFH*NENLVGSVSEDCNICFWGI
COR	(125-159)	15	LSGHKRVGTTSFG*ADNVAVTSSGDFLKTWDV
COR	(168-201)	8	VEGHSDMITSCENHNHNSQIVTTCKDKKARVSDP
AAC3	(139-153)	138	LASSQSDGTVRVWNE
AAC3	(183-217)	29	LKGHDGSEKISWS*NNDLLASAGTDKVIKWDV
AAC3	(314-347)	96	LYGHTASTYCMEDPTGKYLAAGSADSIYSLWDL

MAK11	(119-133)	118	LVSGGNDDEHRTIYDI
MAK11	(176-190)	23	LLSASEDHKIMLRV
MAK11	(199-232)	8	LKGTARVNDVDLHPTNRIATSVSDHSTRLNWL
MAK11	(350-385)	17	LLGHTNRIKDEKEY*FGTYLVITIGSDGKIVVWDM
MSI1	(145-159)	144	TAGASSDGATYIFDR
MSI1	(216-230)	37	ELSSHNSGQVQVWDY
MSI1	(312-326)	62	LASADSNGRINLWDL
MSI1	(357-371)	11	ATAGQEDGLVKLNDT
MSI1	(380-414)	8	HGGHMLGVNDISWD*DPWLMCSVANDNSVHTWKP
CDC20	(274-288)	273	VTAIALDTALYLWNA
CDC20	(316-330)	8	ISMAREGDNTEINDV
CDC20	(357-371)	7	IATGSRSGEIQINDV
CDC20	(400-414)	9	LASGENDNTVMINDT
CDC20	(445-459)	11	SGGGQIDPKHTHFWS
PWP1	(230-244)	229	AAIGTFPPQIEIWNL
PWP1	(302-316)	38	LASTSADHTVKIWDL
PWP1	(347-361)	11	LLTGGYDSRVALTDV
PWP1	(492-506)	11	MVIGGVNKKVKKINDV
LIG	(72-104)	71	LVFHDAIAISAME.AQGCFGGGGAGGSITIEDT
LIG	(201-230)	96	LSAHSVAAVNDVD...PTIQGLAFDSTEGIDS

Fig. 1. (A) The WD-40 consensus sequence, which represents the consensus for a single repeat, notes those positions at which a single amino acid or group of similar amino acids predominates. n1 and n2 represent stretches of amino acids, variable in both sequence and length, that separate individual members of the WD-40 repeat, and elements A and B within the repeat, respectively. x = any amino acid, ϕ = hydrophobic residues preferred, δ = non-charged amino acid.

(B) Alignment of WD-40 repeats constructed manually. The numbering of the amino acid residues (between parentheses) is according to the references (see text). The number of amino acids constituting n1 is indicated. In the absence of an element recognizable as A, 19 residues (the mean length of the segment otherwise occupied by element A together with n2) have been subtracted. The preferred length for the spacer n1, between individual repeats, is thus around 8 residues. If the length of n2 exceeds 6 amino acids, the extra amino acids are indicated by an asterisk. Some members of the WD40-family are not listed, either because their sequences are not yet available (PRP17), or because they are closely related to members already listed (e.g. the different G β -subtypes and Cblp which, upon closer inspection, might be considered as the *Chlamydomonas* homologue of the 12.3-protein (82% similarity)).

the [D,N]xxxxx[W,F,Y][D,N] pattern in part B. The connecting regions, n1 and n2, contain many charged residues and prolines. The distances between the individual repeat units (n1 in Fig. 1) vary largely, but n1=8 amino acids seems to be favoured, resulting in a total repeat length of about 42-43 amino acids.

In some motifs, considerable deviation from the consensus sequence is observed: some repeats would not have been recognized as such were it not for the fact that the same protein contains other, better conserved repeats with a proper spacing relative to the more degenerate motif. Obviously, even more divergent versions of the motif may have been missed. On a number of occasions, only element B of the repeat could be identified (e.g. PWP1 and CDC20 contain no part A at all), suggesting that part A may be dispensable. In these cases, no obvious homologies between non-A parts immediately upstream of part B were observed. In the databases, no protein was found that contains a complete WD-40 motif (A+B) in the absence of other sequences homologous to the repeat. This might indicate that multiple copies of the motif (or part of the motif) are required to render it functional.

What could be the function of the WD-40 repeat? Its

presence implies some common structural feature(s) for the family members, which may, but need not, imply functional relationships. However, the function(s) of the motif is not known for any of the members of the WD-40 family. The members of the family do not share any obvious functional properties. Furthermore, the subcellular location (if known) of these proteins varies largely: some are found in the nucleus (e.g. CDC4, PRP4 and TUP1), others in the plasma membrane (e.g. Espl, G β -subunits, STE4) or as a cytoskeletal component (coronin). None of the members of the family has been crystallized nor has extensive mutational analysis been done.

An intriguing observation concerns the fact that, in yeast, some of the members of the WD-40 family are genetically associated with members of another family of repeat-containing proteins: the tetratricopeptide (TPR) gene family (for a review see [7]). Based on this finding, it has been proposed that members of the WD-40 family work in pairs with members of the TPR-family [7]. Interestingly, TPR-proteins contain multiple repeats of a 34 amino acid sequence motif of unknown function. As with the members of the WD-40 family, members of the TPR-family are involved in many cellu-

lar functions and are found in a wide variety of subcellular locations. In a single case, a direct physical association between members of the two families has been demonstrated: the TPR-protein, Ssn6, when paired with TUP1, a member of the WD-40 family, functions as a general repressor of transcription in yeast [26,27]. Although the evidence is suggestive, it remains to be established whether TUP1 and Ssn6 bind to each other via their respective repeats, but undoubtedly this hypothesis will soon be tested. Clearly, with the relative ease of genetic analysis and the availability of many mutant strains, the yeast *Saccharomyces cerevisiae* provides an attractive system in which to study the function(s) of the WD-40 repeat.

REFERENCES

- [1] Fong, H.K.W., Hurley, J.B., Hopkins, R.S., Miake-Lye, R., Johnson, M.S., Doolittle, R.F. and Simon, M. (1986) *Proc. Natl. Acad. Sci. USA* 83, 2162-2166.
- [2] Simon, M.I., Strathmann, M.P. and Gautam, N. (1991) *Science* 252, 802-808.
- [3] Whiteway, M., Hougan, L., Dignard, D., Thomas, D.Y., Bell, L., Saari, G.C., Grant, F.J., O'Hara, P. and Mackay, V.L. (1989) *Cell* 56, 467-477.
- [4] Yochem, J. and Byers, B. (1987) *J. Mol. Biol.* 195, 233-245.
- [5] Choi, W.J., Clark, M.W., Chen, J.X. and Jong, A.Y. (1990) *Biochem. Biophys. Res. Commun.* 172, 1324-1330.
- [6] Pringle, J. and Hartwell, L. (1981) *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance* (Strathern, J.N., Jones, E.W. and Brouch, J.R., eds.) 97-142, Cold Spring Harbour Press, NY.
- [7] Goebel, M. and Yanagia, M. (1991) *Trends Biochem. Sci.* 16, 173-1776.
- [8] Sethi, N., Monteagudo, M.C., Koshland, D., Hogan, E. and Burke, D.J. (1991) *Mol. Cell Biol.* 11, 5592-5602.
- [9] Hartley, D.A., Preiss, A. and Artavanis-Tsakonas, S. (1988) *Cell* 55, 785-795.
- [10] Guillemot, F., Billault, A. and Auffray, C. (1989) *Proc. Natl. Acad. Sci. USA* 86, 4594-4598.
- [11] Dalrymple, M.A., Petersen-Bjorn, S., Friesen, J.D. and Beggs, J.D. (1989) *Cell* 58, 811-812.
- [12] Banroques, J. and Abelson, J.N. (1989) *Mol. Cell Biol.* 9, 3710-1719.
- [13] Ruby, S.W. and Abelson, J. (1991) *Trends Gen.* 7, 79-85.
- [14] Zhang, M., Rosenblum-Vos, L.S., Lowry, C.V., Boakye, K.A. and Zitomer, R.S. (1991) *Gene* 97, 153-161.
- [15] Williams, F.E. and Trumbly, R.J. (1990) *Mol. Cell Biol.* 10, 6500-6511.
- [16] Ruggieri, R., Tanaka, K., Nakafuku, M., Kaziro, Y., Toh-e, A. and Matsumoto, K. (1989) *Proc. Natl. Acad. Sci. USA* 86, 8778-8782.
- [17] Hostos, E.L., Bradtke, B., Lottspeich, F., Guggenheim, R. and Gerisch, G. (1991) *EMBO J.* 10, 4097-4104.
- [18] Duronio, R.J., Gordon, J.I. and Boguski, M.S. retrieval code: SWISS:PWPI\$YEAST.
- [19] Schloss, J.A. (1990) *Mol. Gen. Genet.* 221, 443-452.
- [20] Icheo, T. and Wickner, R.B. (1988) *J. Biol. Chem.* 263, 1467-1475.
- [21] Shaw, D.R., Richter, H., Giorda, R., Ohmachi, T. and Ennis, H.L. (1989) *Mol. Gen. Genet.* 218, 453-459.
- [22] Devereux, J., Haerberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* 12, 387-395.
- [23] Gribkov, M., McLachlan, A.D. and Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. USA* 84, 4355-4358.
- [24] Brown, A., Sims, P.F.G., Raeder, U. and Broda, P. (1988) *Gene* 73, 77-85.
- [25] Tien, M. and Tu, C.-P.D. (1987) *Nature* 326, 520-523.
- [26] Williams, F.E., Varanasi, U. and Trumbly, R.J. (1991) *Mol. Cell Biol.* 11, 3307-3316.
- [27] Keleher, C.A., Redd, M.J., Schultz, J., Carlson, M. and Johnson, A.D. (1992) *Cell* 68, 709-719.