Protein Science (1999), *8*:1557–1561. Cambridge University Press. Printed in the USA.

ACCELERATED COMMUNICATION

TRIADs: A new class of proteins with a novel cysteine-rich signature

BERT A. VAN DER REIJDEN, CLAUDIA A.J. ERPELINCK-VERSCHUEREN, BOB LÖWENBERG, AND JOOP H. JANSEN

Institute of Hematology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

(RECEIVED April 19, 1999; ACCEPTED May 3, 1999)

Abstract

Triad1 was recently identified as a nuclear RING finger protein, which is up-regulated during retinoic acid induced granulocytic differentiation of acute leukemia cells. Here we show that a cysteine-rich domain (C6HC), present in Triad1, is conserved in at least 24 proteins encoded by various eukaryotes. The C6HC consensus pattern C-x(4)-C-x(14–30)-C-x(1–4)-C-x(4)-C-x(2)-C-x(4)-H-x(4)-C defines this structure as the fourth family member of the zinc-binding RING, LIM, and LAP/PHD fingers. Strikingly, in 22 of 24 proteins the C6HC domain is flanked by two RING finger structures. We have termed the novel C6HC motif DRIL (double RING finger linked). The strong conservation of the larger tripartite TRIAD (two RING fingers and DRIL) structure indicates that the three subdomains are functionally linked and identifies a novel class of proteins.

Keywords: LAP/PHD domain; LIM domain; protein–DNA interaction; protein–protein interaction; RING finger

Since the initial discovery of the zinc finger as a DNA-binding protein structure (Miller et al., 1985), several other zinc-binding structures involved in protein–DNA or protein–protein interactions have been identified (Freemont, 1993; Schwabe & Klug, 1994; Aasland et al., 1995; Saha et al., 1995; Saurin et al., 1996; Dawid et al., 1998). The common denominator relating these motifs is the regular spacing of zinc-binding cysteine and histidine residues that form a three-dimensional structure that is stabilized by zinc, thereby providing an interface for interaction. Within this group of zinc-binding motifs, there are three related structures that each bind two zinc atoms. These are the RING (really interesting new gene) (Freemont, 1993) , the LIM (for the proteins in which it was first discovered: Lin11, Isl1, and Mec3) (Schwabe & Klug, 1994), and the LAP (leukemia-associated protein) (Saha et al., 1995) /PHD (plant homeo domain) fingers (Aasland et al., 1995). All these structures use eight cysteine/histidine residues for zinc-binding. In contrast to the classical zinc finger that interacts with both DNA and proteins (Mackay & Crossley, 1998), the LIM and RING fingers are thought to provide interfaces for protein interactions only (Freemont, 1993; Schmeichel & Beckerle, 1994; Schwabe & Klug, 1994; Saurin et al., 1996; Dawid et al., 1998). LIM and RING finger proteins may have either cytoplasmic or nuclear functions and are implicated in various processes such as cell lineage determination, oncogenesis, and embryogenesis (Freemont, 1993; Schwabe & Klug, 1994; Saurin et al., 1996; Dawid et al., 1998). RING fingers are often found in conjunction with other cysteine-rich domains (Saurin et al., 1996). In the RING finger B-box coiled coil (RBCC) proteins, the RING fingers are linked to a specific cysteine-rich zinc-binding motif (B-Box) and a coiled coil domain (Reddy et al., 1992). Interestingly, the three RBCC proteins PML, TIF1, and Rfp were all identified as oncogenic products that are disrupted by chromosomal translocations (Saurin et al., 1996). So far, LAP/PHD proteins appear to be confined to the nucleus. They have been shown to be involved in gene regulation through chromatin interaction or by functioning as bridging factors between transcription factors and the basal transcription machinery (Aasland et al., 1995; Giles et al., 1998). We have identified a novel conserved cysteine-rich motif (DRIL) that constitutes the fourth member of the family of the zinc-binding LAP/PHD, LIM, and RING finger domains.

Reprint request to: Bert A. van der Reijden, Institute of Hematology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; e-mail: vanderreijden@hema.fgg.eur.nl.

Abbreviations: DRIL, double ring finger linked; LAP, leukemia-associated protein; LIM, Lin11, Isl1, and Mec3; PHD, plant homeo domain; RBCC, RING finger B-box coiled coil; RING, really interesting new gene; TRIAD, two RING fingers and DRIL.

Results

Identification of a novel cysteine-rich motif belonging to the family of LAP/PHD, LIM, and RING finger domains

We have recently identified Triad1 as a nuclear protein that is up-regulated during all-*trans* retinoic acid induced granulocytic

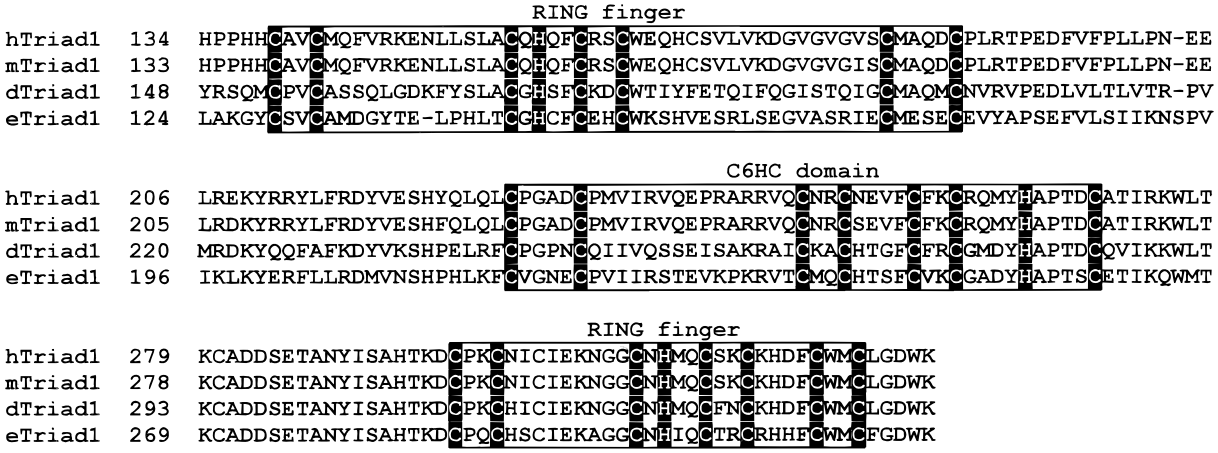

Fig. 1. Triad1 RING finger and C6HC domain conservation in the human (hTriad1), mouse (mTriad1), *D. melanogaster* (dTriad1), and *C. elegans* (eTriad1) orthologs. For clarity only the conserved cysteine/histidine residues related to the C6HC and RING finger domains are indicated. Numbers indicate Triad1 amino acid positions. Human, mouse, *D. melanogaster*, and *C. elegans* Triad1 GenBank accession numbers are respectively, AF099149, AJ130975, AJ01069, and U61944.

differentiation of acute promyelocytic leukemia cells (B.A. van der Reijden et al., in prep.). Triad1 contains two RING fingers, two C-terminal coiled coil domains and an acidic N-terminal region. In addition to the two RING fingers, it contains a cysteine-rich domain (C6HC) that is conserved in the murine, *Drosophila melanogaster*, and *Caenorhabditis elegans* orthologs (Fig. 1). To determine whether this domain is conserved in other proteins, we performed database searches using the human amino acid and pattern sequences and found 20 additional C6HC proteins. These proteins are encoded by various eukaryotes ranging from baker's yeast (*Saccharomyces cerevisiae*) to human. Alignment of the obtained sequences defined the C6HC consensus pattern as C-x(4)-C-x(14–30)-C-x(1–4)-C-x(4)-C-x(2)-C-x(4)-H-x(4)-C (Fig. 2). This pattern is reminiscent of the zinc-binding LAP/PHD, LIM, and RING finger consensus patterns. These motifs are characterized by the presence of eight zinc-coordinating cysteine and histidine residues that bind two zinc atoms (Fig. 2). The spacing between these zinc-coordinating residues is small (1–5 amino acids) with the exception of two loops that vary in length (4–48 amino acids). The C6HC motif differs from these patterns with respect to the histidine location and the absence of length-variation in the C-terminal loop (Fig. 2). The strong resemblance of the C6HC domain to the zinc-binding LAP/PHD, LIM, and RING fingers suggests that C6HC is a novel metal-binding motif possibly involved in protein–protein or protein–DNA interactions (Fig. 2).

C-x_2-	C-$x_{(17-19)}$-**H**-x_2-	C-x_2-	C-x_2-C-$x_{(16-20)}$-	**H**-x_2-C	LIM pattern
C-x_2-	C-$x_{(9-39)}$-	C-$x_{(1-3)}$-**H**-$x_{(2-3)}$-C-x_2-C-$x_{(4-48)}$-	C-x_2-	C	RING finger pattern
C-$x_{(1,2)}$-	C-$x_{(9-21)}$-	C-$x_{(2-4)}$-C-$x_{(4-5)}$-**H**-x_2-C-$x_{(12-46)}$-	C-x_2-	C	LAP/PHD pattern
C-x_4-	C-$x_{(14-30)}$-	C-$x_{(1-4)}$-C-x_4-	C-x_2-C-x_4-	**H**-x_4-C	DRIL pattern

Fig. 2. The novel C6HC pattern (DRIL) consists of seven regularly spaced cysteine and one histidine residues and closely resembles the LAP/PHD, LIM, and RING finger patterns. Zinc-binding ligands are C = cysteine, H = histidine (bold), whereas x can be any residue. In virtually all RING fingers the residues upstream of C5 and downstream of C6 are hydrophobic (Freemont, 1993). C8 in the LIM domain consensus can also be an aspartic acid or histidine residue (Dawid et al., 1998).

C6HC motif conservation identifies a novel class of proteins

Remarkably, alignment of the 24 C6HC containing proteins revealed that the novel domain is flanked by two RING finger structures in 22 of the 24 proteins (Fig. 3). Given this association we propose the name DRIL (double RING finger linked) for the C6HC domain. The striking conservation of the two RING fingers and DRIL (TRIAD) motifs defines a novel conserved tripartite motif. The preservation of both the order and the distance of these three domains suggests that the spatial organization has functional significance. Downstream of the C-terminal RING finger an additional conserved cysteine and histidine residue is present (Fig. 3). Notably, many TRIAD RING fingers differ from the RING finger consensus pattern (Saurin et al., 1996). In the majority (18 of 22) of the N-terminal RING fingers, the spacing between C7–C8 is four instead of two amino acids, whereas in the C-terminal RING fingers the spacing between C5–C6 varies between 1–4 amino acids instead of two amino acids. In the latter finger the histidine residue is less conserved (14 of 22), whereas downstream of C6 a preference (12 of 22) for a basic as opposed to the consensus hydrophobic residues is found (Figs. 2, 3). In addition, several other conserved noncysteine/histidine residues are present, particularly in the C-terminal RING finger. While some of these residues are unique to the TRIAD structure, other residues can also be found in several RING fingers of non-TRIAD proteins indicating that they are not unique to the TRIAD structure but rather may identify a subclass of RING fingers in general.

To identify TRIAD domain conservation, an identity and similarity matrix was calculated (Fig. 4) (Thompson et al., 1994). Cross-species conservation of nonligand binding amino acids was found for at least four distinct groups of TRIAD domains. The very high similarity (61–100%) within these groups indicates that TRIAD domains are well preserved during evolution, and suggests that they have important biological functions.

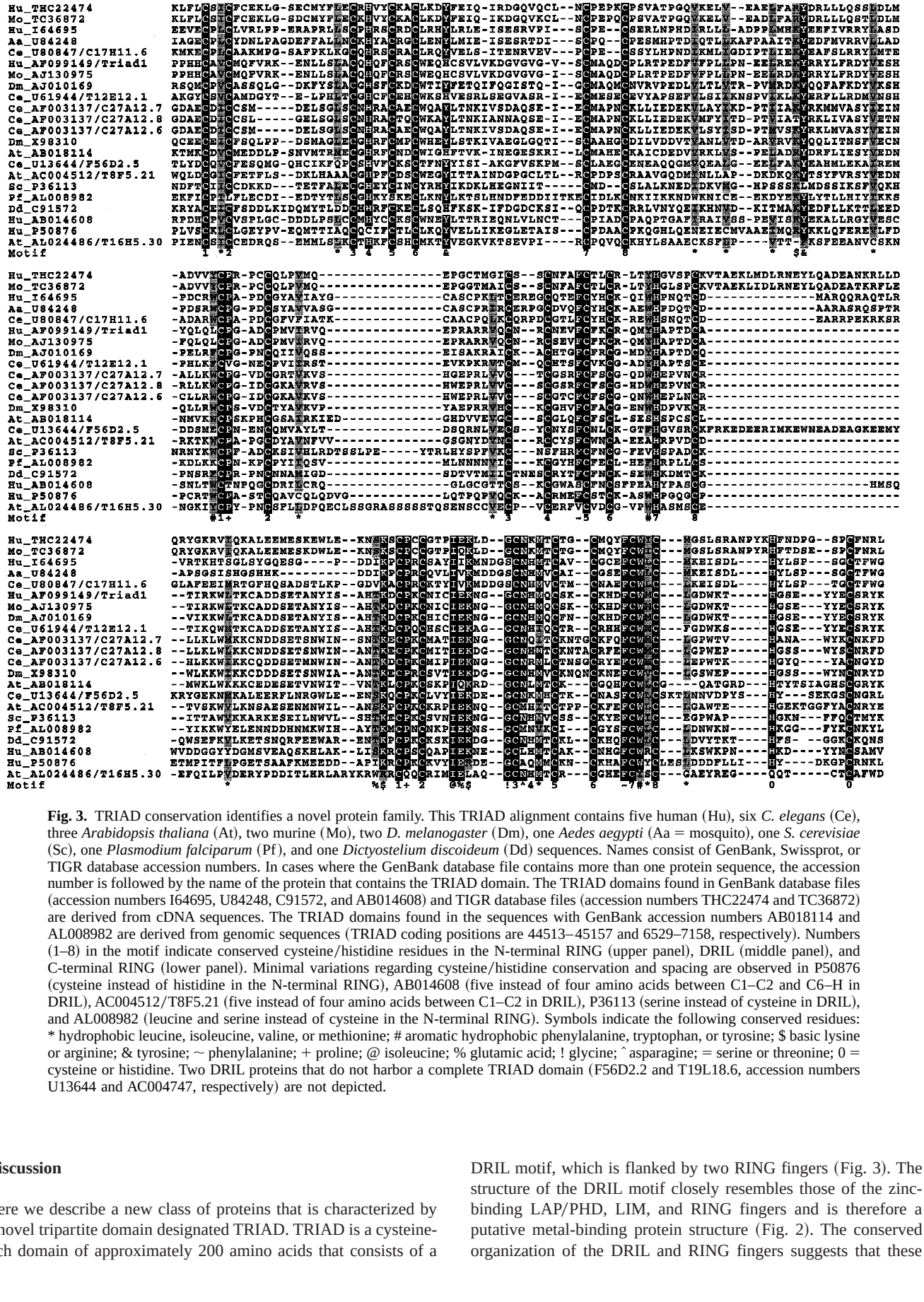

Fig. 3. TRIAD conservation identifies a novel protein family. This TRIAD alignment contains five human (Hu), six *C. elegans* (Ce), three *Arabidopsis thaliana* (At), two murine (Mo), two *D. melanogaster* (Dm), one *Aedes aegypti* (Aa = mosquito), one *S. cerevisiae* (Sc), one *Plasmodium falciparum* (Pf), and one *Dictyostelium discoideum* (Dd) sequences. Names consist of GenBank, Swissprot, or TIGR database accession numbers. In cases where the GenBank database file contains more than one protein sequence, the accession number is followed by the name of the protein that contains the TRIAD domain. The TRIAD domains found in GenBank database files (accession numbers I64695, U84248, C91572, and AB014608) and TIGR database files (accession numbers THC22474 and TC36872) are derived from cDNA sequences. The TRIAD domains found in the sequences with GenBank accession numbers AB018114 and AL008982 are derived from genomic sequences (TRIAD coding positions are 44513–45157 and 6529–7158, respectively). Numbers (1–8) in the motif indicate conserved cysteine/histidine residues in the N-terminal RING (upper panel), DRIL (middle panel), and C-terminal RING (lower panel). Minimal variations regarding cysteine/histidine conservation and spacing are observed in P50876 (cysteine instead of histidine in the N-terminal RING), AB014608 (five instead of four amino acids between C1–C2 and C6–H in DRIL), AC004512/T8F5.21 (five instead of four amino acids between C1–C2 in DRIL), P36113 (serine instead of cysteine in DRIL), and AL008982 (leucine and serine instead of cysteine in the N-terminal RING). Symbols indicate the following conserved residues: * hydrophobic leucine, isoleucine, valine, or methionine; # aromatic hydrophobic phenylalanine, tryptophan, or tyrosine; $ basic lysine or arginine; & tyrosine; ~ phenylalanine; + proline; @ isoleucine; % glutamic acid; ! glycine; ^ asparagine; = serine or threonine; 0 = cysteine or histidine. Two DRIL proteins that do not harbor a complete TRIAD domain (F56D2.2 and T19L18.6, accession numbers U13644 and AC004747, respectively) are not depicted.

Discussion

Here we describe a new class of proteins that is characterized by a novel tripartite domain designated TRIAD. TRIAD is a cysteine-rich domain of approximately 200 amino acids that consists of a DRIL motif, which is flanked by two RING fingers (Fig. 3). The structure of the DRIL motif closely resembles those of the zinc-binding LAP/PHD, LIM, and RING fingers and is therefore a putative metal-binding protein structure (Fig. 2). The conserved organization of the DRIL and RING fingers suggests that these

	Hu_THC22474	Mo_TC36872	Hu_I64695	Aa_U842486	Ce_U80847/C17H11.6	Hu_AF099149/Triad1	Mo_AJ130975	Dm_AJ010169	Ce_U61944/T12E12.1	Ce_AF003137/C27A12.7	Ce_AF003137/C27A12.8	Ce_AF003137/C27A12.6	Dm_X98310	At_AB018114	Ce_U13644/F56D2.5	At_AC004512/T8F5.21	Sc_P36113	Pf_AL008982	Dd_C91572	Hu_AB014608	Hu_P50876	At_AL024486/T16H5.30
Hu_THC22474	--	**96**	33	32	36	34	34	36	33	36	33	36	35	35	47	31	27	31	40	32	29	26
Mo_TC36872	**92**	--	32	31	35	34	34	36	34	35	33	36	35	36	47	31	26	30	40	31	28	27
Hu_I64695	24	23	--	**67**	**64**	33	32	33	32	34	33	34	33	33	26	32	31	31	39	31	31	27
Aa_U84248	23	22	**61**	--	**61**	31	31	34	32	32	31	34	33	30	29	32	28	28	38	31	33	28
Ce_U80847/C17H11.6	26	26	**54**	**50**	--	33	33	34	33	35	35	35	33	32	30	32	32	30	36	31	33	27
Hu_AF099149/Triad1	23	22	22	22	22	--	**100**	**66**	**65**	*49*	*50*	*50*	*51*	41	33	47	43	38	33	42	34	30
Mo_AJ130975	22	21	22	22	22	**98**	--	**66**	**65**	*49*	*50*	*50*	*51*	41	32	47	43	38	33	42	34	30
Dm_AJ010169	25	25	24	24	22	**58**	**59**	--	**66**	*51*	*52*	*51*	*52*	43	33	46	43	43	38	38	31	31
Ce_U61944/T12E12.1	22	21	20	21	22	**53**	**54**	**56**	--	*50*	*51*	*49*	*48*	41	30	44	42	39	36	37	30	34
Ce_AF003137/C27A12.7	23	22	24	23	24	36	37	38	37	--	**88**	**88**	**63**	51	29	46	44	42	41	36	30	31
Ce_AF003137/C27A12.8	23	23	24	23	25	39	40	41	38	**80**	--	**85**	**66**	51	31	47	46	40	40	36	31	33
Ce_AF003137/C27A12.6	23	23	25	24	24	39	40	42	38	**79**	**77**	--	**63**	50	31	48	44	41	40	35	30	31
Dm_X98310	25	24	27	26	25	39	39	44	34	**51**	**57**	**53**	--	49	32	49	44	40	39	37	28	32
At_AB018114	27	28	24	21	23	34	35	34	32	36	39	37	37	--	29	43	36	36	43	34	28	28
Ce_U13644/F56D2.5	35	34	19	20	24	24	23	22	21	22	22	22	23	22	--	31	25	30	34	27	32	23
At_AC004512/T8F5.21	24	23	23	23	23	36	36	35	30	35	36	37	38	34	24	--	43	43	40	35	29	31
Sc_P36113	21	20	22	21	23	30	30	31	28	33	34	34	34	29	19	33	--	37	33	32	27	31
Pf_AL008982	23	22	21	20	20	29	29	32	30	31	30	32	30	26	24	32	28	--	38	35	30	29
Dd_C91572	28	27	27	27	27	25	26	30	25	29	28	28	31	32	26	30	25	27	--	32	36	29
Hu_AB014608	22	22	23	24	23	31	31	30	28	28	28	28	28	25	21	30	23	27	25	--	29	28
Hu_P50876	21	19	24	24	23	25	25	25	21	21	24	22	24	20	23	23	20	22	29	23	--	28
At_AL024486/T16H5.30	18	18	18	19	19	21	21	20	23	20	23	21	21	20	15	20	20	17	22	19	19	--

% identity

% similarity

Fig. 4. Percentages of TRIAD similarity and identity. Names are explained in Figure 3. TRIAD orthologs/paralogs are in bold and boxed (identity ≥50%). Related TRIAD family members (≥48% similarity) are in italics and boxed (dashed line). The similarity (≥40%) of lower eukaryotes (*S. cerevisiae*, *D. discoideum*, and *P. falciparum*) to several different conserved TRIAD proteins (dotted boxes) further illustrates the evolutionary preservation of TRIAD.

domains operate together as one functional unit that may mediate complex protein or DNA interactions. Both TRIAD RING fingers differ from the classical RING finger consensus pattern with respect to cysteine spacing and nonligand binding amino acid conservation (Fig. 3). These characteristic differences may result in structural variations that are essential to TRIAD functioning.

To date, the function of TRIAD containing proteins remains unknown. However, within the set of 22 TRIAD domains reported here, at least four groups with strong cross-species conservation are present, suggesting that they have an important function. More distant TRIAD homologies were found for parkin, a protein mutated in juvenile parkinsonism (Kitada et al., 1998), and RBCK, a novel protein kinase C interacting protein (Tokunaga et al., 1998) (data not shown). Neither protein completely matches the TRIAD consensus nor do they share certain cysteine/histidine or other conserved residues, formally disqualifying them as TRIAD proteins. Like TRIAD proteins, the exact functions of parkin and RBCK are unknown.

The TRIAD proteins described here vary in size from 292 to more than 1,754 amino acids (data not shown). In the smallest member (P50876), the TRIAD domain spans almost the entire length of the protein. If TRIAD represents an interface for multiple protein interactions, the principal role of the smaller TRIAD proteins may be to integrate the function of other proteins, whereas the larger ones may have additional roles.

Protein structure analysis predicts that two-thirds of the TRIAD proteins are nuclear (data not shown). Interestingly, all these nuclear TRIAD proteins also contain predicted coiled coil regions (Lupas, 1997), whereas all but two cytoplasmic proteins do not (data not shown). Apparently, the presence of coiled coil domains is essential to nuclear TRIAD protein functioning, perhaps providing additional surfaces for protein interaction. Since nuclear proteins may be involved in various processes that are pertinent to cell-cycle regulation, cellular differentiation, gene expression, and oncogenesis, it will be of interest to determine in which nuclear processes TRIAD proteins are implicated.

Materials and methods

Detection of C6HC/DRIL proteins

C6HC/DRIL proteins were identified by screening the nonredundant, the expressed sequence tag, and the Swissprot GenBank databases (version 2.0.6) at the National Center for Biotechnology Information (NCBI) (http://ncbi.nlm.nih.gov), and the tentative human and mouse consensus databases at The Institute for Genomic Research (TIGR) (http://www.tigr.org), with the human Triad1 DRIL protein sequence using the TBLASTN and BLASTP algorithms (Altschul et al., 1997). The pattern C-x(4)-C-x(4,25)-C-x(1,4)-C-x(4)-C-x(2)-C-x(4,25)-H-x(4)-C was used to screen the ISREC prosite database by the pattern search algorithm (Hofmann et al., 1999).

Alignment of TRIAD domains and matrix calculations

Protein sequence alignments and identity/similarity matrix calculations were performed using the CLUSTAL W (version 1.74)

algorithm (Thompson et al., 1994) at IBCP (http://www.ibcp.fr), using default parameters, except for the gap-penalty (8 instead of 10), and were adjusted by hand. TRIAD gene expression was confirmed by screening the expressed sequence tag database at NCBI with putative TRIAD protein sequences (that were derived from genomic DNA sequences) using the TBLASTN algorithm (Altschul et al., 1997). All proteins were shown to match to expressed sequences, except for AF003137/C27A12.7, AC024486/T16H5.30, and AB018114.

Acknowledgments

We acknowledge Rachel H. Giles for critical reading of the manuscript and Paul Freemont for helpful discussions.

References

Aasland R, Gibson TJ, Stewart AF. 1995. The PHD finger: Implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci 20*:56–59.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:33389–33402.

Dawid IB, Breen JJ, Toyama R. 1998. LIM domains: Multiple roles as adapters and functional modifiers in protein interactions. *Trends Genet 14*:156–162.

Freemont PS. 1993. The RING finger. A novel protein sequence motif related to the zinc finger. *Ann NY Acad Sci 684*:174–192.

Giles RH, Peters DJ, Breuning MH. 1998. Conjunction dysfunction: CBP/p300 in human disease. *Trends Genet 14*:178–183.

Hofmann K, Bucher P, Falquet L, Bairoch A. 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res 27*:215–219.

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N. 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature 392*:605–608.

Lupas A. 1997. Predicting coiled coil regions in proteins. *Curr Opin Struct Biol 7*:388–393.

Mackay JP, Crossley M. 1998. Zinc fingers are sticking together. *Trends Biochem Sci 23*:1–4.

Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J 4*:1609–1614.

Reddy BA, Etkin LD, Freemont PS. 1992. A novel zinc finger coiled coil domain in a family of nuclear proteins. *Trends Biochem Sci 17*:344–345.

Saha V, Chaplin T, Gregorini A, Ayton P, Young BD. 1995. The Leukemia-Associated Protein (LAP) domain, a cysteine-rich motif, is present in a wide range of proteins, including MLL, AF10, and MLLT6 proteins. *Proc Natl Acad Sci USA 92*:9737–9741.

Saurin AJ, Borden KL, Boddy MN, Freemont PS. 1996. Does this have a familiar RING? *Trends Biochem Sci 21*:208–214.

Schmeichel KL, Beckerle MC. 1994. The LIM domain is a modular protein-binding interface. *Cell 79*:211–219.

Schwabe JW, Klug A. 1994. Zinc mining for protein domains. *Nat Struct Biol 1*:345–349.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.

Tokunaga C, Kuroda S, Tatematsu K, Nakagawa N, Ono Y, Kikkawa U. 1998. Molecular cloning and characterization of a novel protein kinase C-interacting protein with structural motifs related to RBCC family proteins. *Biochem Biophys Res Commun 244*:353–359.