Protein Science (1997), 6:249–253. Cambridge University Press. Printed in the USA.

FOR THE RECORD

SAM as a protein interaction domain involved in developmental regulation

JÖRG SCHULTZ,[1] CHRISTOPHER P. PONTING,[2] KAY HOFMANN,[3] AND PEER BORK[1]

[1]EMBL, Heidelberg and Max-Delbrück-Center for Molecular Medicine, Berlin-Buch, Germany
[2]University of Oxford, Fibrinolysis Research Unit, The Old Observatory, South Parks Road, Oxford, OX1 3RH, United Kingdom
[3]Swiss Institute for Experimental Cancer Research (ISREC), Lausanne, Switzerland

(RECEIVED August 28, 1996; ACCEPTED October 30, 1996)

Abstract: More than 60 previously undetected SAM domain–containing proteins have been identified using profile searching methods. Among these are over 40 EPH-related receptor tyrosine kinases (RPTK), Drosophila bicaudal-C, a p53 from *Loligo forbesi*, and diacylglycerol-kinase isoform δ. This extended dataset suggests that SAM is an evolutionary conserved protein binding domain that is involved in the regulation of numerous developmental processes among diverse eukaryotes. A conserved tyrosine in the SAM sequences of the EPH related RPTKs is likely to mediate cell–cell initiated signal transduction via the binding of SH2 containing proteins to phosphotyrosine.

Keywords: alignment; EPH receptor protein tyrosine kinases; homology search; protein binding; SAM

The SAM (sterile alpha motif) domain was first described (Ponting, 1995) as a module that is present in single copies in a small set of yeast sexual differentiation and Drosophila polyhomeotic proteins. As we are interested in the modular architecture of protein kinases we have used the previously identified (Ponting, 1995) SAM domain of the yeast kinase Ste11p to search current sequence databases using Blastp (Altschul et al., 1990). DD36938, a *Dictyostelium discoideum* ankyrin repeat-containing protein, showed significant similarity to the Ste11p SAM sequence ($p = 3 \times 10^{-4}$); this sequence was used as the query in a subsequent reciprocal Blastp search. Significantly matching ($p < 0.01$) SAM domains were then used to construct a multiple alignment as the initiating procedure of complementary and iterative profile searches (Gribskov et al., 1987; Birney et al., 1996). Subsequently, more than 60 new SAM domains were identified (those annotated with asterisks in Table 1; for details of methods see the legend to Fig. 1), leading to considerable improvement of the alignment (Fig. 1), the accuracy of secondary structure prediction, and the prediction of domain function.

Reprint requests to: Peer Bork, EMBL, Meyerhofstr.1, 69117 Heidelberg, Germany; e-mail: Bork@EMBL-Heidelberg.de.

As SAM domains are found in such diverse organisms as fungi, protozoa, and animals, they are likely to have evolved prior to the divergence of eukaryotes. Literature data on the functions of SAM-containing proteins reveal a surprisingly common theme. These proteins' functions, when known (Table 1), are all compatible with their participation in developmental regulation in single- and multicellular eukaryotes.

For a variety of the SAM-containing proteins, developmental regulation has been directly shown by experiments. Novel SAM domains were identified in the C-terminus of all known EPH-related receptor protein tyrosine kinases (RPTKs); methods used previously (Ponting, 1995) were unable to detect these domains as SAMs. The EPH subfamily is currently the largest group of RPTKs (van der Geer et al., 1994; Tuzi & Gullick, 1994) whose restricted patterns of expression have implicated them in a variety of processes in early development, including regulation of forebrain patterning in zebrafish (Xu et al., 1996). In the adult, several EPH-like RPTKs are thought to mediate different functions (Valenzuela et al., 1995), and some have been implicated in carcinogenesis (Kiyokawa et al., 1994).

Fungal SAM-containing proteins also appear to play roles in development. *Schizosaccharomyces pombe* Byr2p, Ste4p and *Saccharomyces cerevisiae* Ste11p participate in mating pheromone response pathways (Neiman et al., 1993) leading to sexual differentiation from haploid to diploid cells (Bardwell et al., 1994). The *S. cerevisiae* Bem1p-binding proteins Boi1p and Boi2p (also called Bob1p and Beb1p) function in the maintenance of cell polarity essential for bud formation (Bender et al., 1996; Matsui et al., 1996). Each of these proteins contains other domains commonly found among signaling molecules (Fig. 2). Levels of the SAM-containing protein-tyrosine kinase DPYK1 in *Dictyostelium discoideum* have been shown to increase as the protozoan initiates its signal-mediated developmental cycle (Tan & Spudich, 1990).

Of the *Drosophila melanogaster* proteins, Bicaudal-C (Mahone et al., 1995) is required during oogenesis for the migration of follicle cells and for anterior-posterior patterning in oocytes (Schupbach & Wieschaus, 1991). Anterior-posterior patterning is also regulated by the *Polycomb* group (PcG) of genes, which encode transcriptional repressors. Among the PcG are two SAM-encoding

Table 1. *SAM domain-containing proteins*

Name	Accession	Species[a]	Function	Remarks[b]
Developmental regulation experimentally shown				
EPH-family		animals	receptor kinases	more than 40 distinct entries in EMBL protein binding *
Byr2p	P28829	*S. pombe*	conjugation and sporulation	protein binding
Ste4p	P36622	*S. pombe*	sexual differentiation	protein binding
Ste11p	P23567	*S. cerevisiae*	sexual differentiation	protein binding
Bep1p (= Boi2p)	P39969	*S. cerevisiae*	cell polarization	
Bob1p (= Boi1p)	P38041	*S. cerevisiae*	cell polarization	
Bicaudal-C	U15928	*D. melanogaster*	required during oogenesis	*
Scm	U49739	*D. melanogaster*	anterior-posterior patterning	*
ph	P39769	*D. melanogaster*	anterior-posterior patterning	
Rae-28	S73882	*M. musculus*	early stages of development	
DPYK1	U32174	*D. discoideum*	developmental cycle	*
Involved in signal transduction				
SLP-76h	U20158	*H. sapiens*	T cell antigen receptor mediated signal transduction	*
SLP-76m	U20159	*M. musculus*	T cell antigen receptor mediated signal transduction	*
Mg11	U15635	*M. musculus*	interferon-γ induced	
Lip1_a	U22815	*H. sapiens*	binds to protein tyrosine phosphatase	protein binding *
R01H10.8	Z31590	*C. elegans*	contains PDZ domain	
C33B4.3	Z48367	*C. elegans*	contains PDZ domain	
KIAA0229	D86982	*H. sapiens*	contains PTB domain	*
DGKδ	D73409	*H. sapiens*	diacylglycerol kinase	*
F42A9.1	U61952	*C. elegans*	diacylglycerol kinase	*
Sqp53	U43595	*L. forbesi*	p53	*
No information				
D52	U03288	*D. melanogaster*	arrests cell cycle in S. pombe	*
D69	U03277	*D. melanogaster*	arrests cell cycle in S. pombe	*
GOK	U52426	*H. sapiens*	*	
T21H8.1	Z78546	*C. elegans*	*	
M7.4	Z68337	*C. elegans*	*	
R13F6.6	U00046	*C. elegans*	*	
F59F5.6	Z50794	*C. elegans*	*	
F53B1.2	U40953	*C. elegans*	*	
F13B10.1	Z49936	*C. elegans*	*	
C05D10.4	U13645	*C. elegans*		
C46H3.2	U41271	*C. elegans*	*	
DD36938	U36938	*D. discoideum*	contains N-terminus of tyrosine kinase	*
51CP	L24444	*H. sapiens*	DNA repair protein	*
Pmk1	U53872	*S. pombe*	wrongly annotated as MEK kinase homolog	*

[a]Species: *S. cerevisiae*, *Saccharomyces cerevisiae*; *S. pombe*, *Schizosaccharomyces pombe*; *D. discoideum*, *Dictyostelium discoideum*; *D. melanogaster*, *Drosophila melanogaster*; *M. musculus*, *Mus musculus*; *H. sapiens*, *Homo sapiens*; *C. elegans*, *Caenorhabditis elegans*; *L. forbesi*, *Loligo forbesi*.
[b]Proteins that have been shown experimentally to bind proteins are marked as "protein binding"; previously unidentified SAM domains are marked with *.

genes, *Sex comb on midleg* (*Scm*) (Bornemann et al., 1996) and *polyhomeotic* (*ph*) (DeCamillis et al., 1992). A comparable function is known for Rae-28, a transcriptional repressor that participates in morphogenesis during the early stage of mouse development (Nomura et al., 1994). Two further *Drosophila melanogaster* SAM proteins, D52 and D69, were cloned due to their ability to arrest the cell cycle when expressed in *S. pombe* (B. Edgar, unpublished, EMBL accession code U03288/U03277). These proteins have only their SAM domains in common, suggesting that the SAM domain is responsible for their ability to induce cell-cycle arrest in the assay used.

The functions of each of the remaining SAM-containing proteins, when known, involve signal transduction. Since regulation of developmental processes is frequently mediated by signalling pathways (for a review see Karin (1992)), this is compatible with the common theme of an involvement of SAM domains in development. SLP-76, a SAM- and SH2-containing protein, was originally isolated by virtue of its association with the adaptor protein Grb2, and is likely to participate in intracellular signal transduction mediated by the T cell antigen receptor (Jackman et al., 1995). LIP1 binds the membrane-distal D2 protein tyrosine phosphatase (PTP) domain of LAR and is proposed to participate in focal adhesion disassembly (Serra-Pages et al., 1995) which would influence a diverse range of cellular processes including cell migration and differentiation (Hynes & Lander, 1992). Several SAM-containing proteins also possess other domains commonly found in

```
GGCEK6A_1       PDFTAFTSVEDWLSAVKMS--QYRDNFLSAGFTSLQLVAQMTSEDLLR--IGVTLAGHQKKILNSIQSMRV  {876-942}
KSEK_MOU        PEFSAVVSVGDWLQAIKMD--RYKDNFTAAGYTTLEAVVHMSQDDLAR--IGITAITHQNKILSSVQAMRT  {908-974}
KEK5_CHICK      PDYTSFNTVDEWLDAIKMS--QYKESFASAGFTTFDIVSQMTVEDILR--VGVTLAGHQKKILNSIQVMRA  {919-985}
MMMDK5_1        PDYTTFTTVGDWLDAIKMG--RYKESFVGAGFASFDLVAQMTAEDLLR--IGVTLAGHQKKILCSIQDMRL  {917-983}
XLERTK_1        PDYTTFPTVSDWLEAIKMG--QYQENFLSAGFTSFHLVAQMTAEDLLR--IGVTLAGHQKKLLNSVQDMRL  {898-964}
MM06834_1       PHYSAFGSVVEWLRAIKMG--RYEESFAAAGFGSFEMVSQISAEDLLR--IGVTLAGHQKKILASVQHMKW  {904-970}
XLEPHRTK_1      PHYSSFSSVSEWLHAIKMG--RYEDGFRNAGFTTFSRVQNISTEDLLR--MGVTLAGHQKKILSSLQQILP  {245-311}
MMKIN1_1        PDFTAFCSVGEWLQAIKME--RYKDNFTAAGYNSLESVARMTIDDVMS--LGITLVGHQKKIMSSIQTMRA  {920-986}
GG23783_1       PDFPSLSNAHEWLDAIKMG--RYKENFDQAGLITFDVISRMTLEDLQR--IGITLVGHQKKILNSIQLMKV  {926-992}
HSEHK1_1        LGSGAYRSVGEWLEAIKMG--RYTEIFMENGYSSMDAVAQVTLEDLRR--LGVTLVGHQKKIMNSLQEMKV  {962-1028}
MMECK_1         SEGVPFRTVSEWLESIKMQ--QYTEHFMVAGYTAIEKVVQMSNEDIKR--IGVRLPGHQKRIAYSLLGLKD  {902-968}
HSRTKEPH_1      SDGIPYRTVSEWLESIRMK--RYILHFHSAGLDTMECVLELTAEDLTQ--MGITLPGHQKRILCSIQGFKD  {625-691}
MMMEK4_1        VDIATFHTTGDWLNGMRTA--HCKEIFTGVEYSSCDTIAKISTDDMKK--VGVTVVGPQKKIISTIKALET  {908-974}
Byr2p           EYYT-SKEVAEWLKSIGLE--KYIEQFSQNNIEG-RHLNHLTLPLLKD--LGIENTAKGKQFLKQRDYLRE  {2-66}
Ste4p           WNWN-NEAVCNWIEQLGFP---HKEAFEDYHILG-KDIDLLSSNDLRD--MGIESVGHRIDILSAIQSMKK  {9-72}
Ste11p          EKTNDLPFVQLFLEEIGCT--QYLDSFIQCNLVTEEEIKYLDKDILIA--LGVNKIGDRLKILRKSKSFQR  {17-83}
Beb1p           EFWS-PEEITAYFIMEGYD-VQSASRFQKHKISG-KILLELELVHLKE--LDINSFGTRFEIFKEIEKIKE  {264-329}
Bob1p           KSWS-PEEVTDYFSLVGFDQ-STCNKFKEHQVSG-KILLELELEHLKE--LEINSFGIRFQIFKEIRNIKS  {226-291}
Bicauda1-C      MQLAKHKDIQTLLTSLGLE--HYIKIFVLNEIDL-EVFTTLTEENLME--LGIAAFGARKKLLTAIHTLLA  {800-865}
Scm             IDWT-IEEVIQYIESNDNSLAVHGDLFRKHEIDG-KALLLLNSEMMMKY-MGLKLGPALKICNLVNKVNGR  {803-871}
ph              SSWS-VDDVSNFIRELPGCQ-DYVDDFIQQEIDG-QALLLLKEKHLVNA-MGMK-LGPALKIVAKVESIKE  {1511-1567}
Rae-28          SQWS-VEEVYEFIASLQGCQ-EIAEEFRSQEIDG-QALLLLKEEHLMSA-MNIK-LGPALKICAKINVLKE  {946-1011}
DPYK1           KNWA-PNDVAIWLESFNYG--QYRKNFRDNNISG-RHLEGITHAMLKND-LGIEPYGHREDIINRLNRMIQ  {906-971}
SLP-76h         LGWD-PDSLADYFKKLNYK--DCEKAVKKYHIDG-ARFLNLTENDIQKF-PKLRVPILSKLSQEINKNEER  {13-79}
SLP-76m         LAWN-SDNLADYFRKLNYR--DCEKAVKKYHIDG-ARFLNLTENDIQKF-PKLRMPLLSKLSQDINKNEER  {13-79}
Mg11            RTWE-PEDVCSFLENRGFREKKVLDIFRDNKIAG-SFLPFLDEDRLED--LGVSSLEERKKMIECIQQLSQ  {44-110}
Lip1_a_1        AQWD-GPTVVVWLELWVGMPAWYVAACRANVKSG-AIMSALSDTEIQRE-IGISNPLHRLKLRLAIQEIMS  {876-943}
Lip1_a_2        GDMNHEWIGNEWLPSLGLP--QYRSYFMECLVDA-RMLDHLTKKDLRGQ-LKMVDSFHRNSFQCGIMCLRR  {960-1026}
Lip1_a_3        LVWS-NDRVIRWILSIGLK--EYANNLIESGVHG--ALLALDETFDFSA-LALLLQIPTQNTQARAVLERE  {1065-1130}
R01H10.8        EQWK-GKEIARWIEGLGDQMNPYLGMIRDNIKSG-KQLEALDDDSLKK--IGISALGARKTIFQAVSLLLY  {49-115}
C33B4.3         DVWS-VDDVIGWLSSLHLS--EYTPAFRSQRING-RCLRQCDRSRFTQ--LGVTRIAHRQIIESALRGLLQ  {1046-1110}
KIAA0229_1      SRTL-EQSVGEWLESIGLQ--QYESKLLLNGFDDVHFLGSNVMEEQDLRDIGISDPQHRRKLLQAARSLPK  {740-807}
KIAA0229_2      YDGNSPPSVPSWLDSLGLQ--DYVHSFLSSGYSSIDTVKNLWELELVN-VLKVQLLGHRKRIIASLADRPY  {813-880}
DGKδ            HLWG-TEEVAAWLEHLSLC--EYKDIFTRHDIRG-SELLHLERRDLKD--LGVTKVGHMKRILCGIKELSR  {1100-1162}
F42A9.1         PYWT-SEEVCAWLSSIGMS--EYGSTFRKNDIQG-SELMHLERSDIMD--IGITKIGHVKRLQSAIVDLRA  {1170-1234}
Sqp53           EPTE--NTIAQWLTKLGLQ--AYIDNFQQKGLHNMFQLDEFTLEDLQS--MRIG-TGHRNKIWKSLLDYRR  {453-516}
D52             YVVNHAANVEQILMHMGLE--NYVTNFEEAHIDL-VELASLERADLVK--IGLNTDEDCNRIMDVLHTL--  {154-217}
D69             RNVG-MSGIGLWLKSLRLH--KYIELFKNMTYEE---MLLITEDFLQS--VGVT-KGASHKLALCIDKLKE  {22-83}
GOK             YNWT-VDEVVQWLITYVELP-QYEETFRKLQLSGHAMPRLAVTNTTMTG-TVLKMTDRSHRQKLQLKALDT  {129-195}
T21H8.1_1       VDWR-SEQLADWIAEIGYP--QYMNEVSRHVRSG-RHFLNMSMNEYEGV-LNIKNPVHRKRVAILLRRIEE  {728-793}
T21H8.1_2       NKWD-VHQTLRWLDDIGLP--QYKDVFAENVVDG-PLLLSLTANDAVE--MKVVNAHHYATLARSIQFLKK  {800-864}
T21H8.1_3       VRWT-HSATCEWLRKIDLA--EFTQNLLFAGVPG-ALMIYEPSFTAESL-AEILQMPPHKTLLRRHLTSHF  {891-956}
M7.4            WDINYFTDPSMVLAQLGCS--EYMTQLRDQEIDM-HAFLLLDEQNLKD--IGVSTIGARKKIHHAILSSFS  {119-185}
R13F6.6         SHWV--ETVLGRVANDKKFLDRVNAAVFRNRITG-ARLLGMTQNDLEH--LGVHKVGSRIELMKMIRKLAD  {49-115}
F59F5.6_1       ALWN-GPTVVAWLELWVGMPAWYVAACRANVKSG-AIMSALSDQEIQKE-IGISNPLHRLKLRLAIQEMVS  {880-948}
F59F5.6_2       GDMNHEYIGNDWLPCLGLA--QYRSAFMECLLDA-RMLEHLSKRDLRTH-LRMVDTFHRTSLQYGIMCLKK  {965-1032}
F59F5.6_3       LVWS-NERVQRWVEEIGLG--VFSRNLVDSGIHG--ALIALDETFDASA-FAYALQIGSQDVPNRQLLEKK  {1055-1120}
F53B1.2         NEWR-CEDVGNWLKKIGMA--KYADLIAMKHKVDGKCLLALTDTDLKDP-PVSINCLGDIKKILFAIEFLS  {12-79}
F13B10.1_1      PGWT-CADVQYWVKKIGFE--EYVEKFAKQMVDG-DLLLQLTENDLKHD-VGMISGLHRKRFLRELQTLKV  {580-645}
F13B10.1_2      SSVD-ESNLDNFLMGLSPELSVYTYQMLTNGVNR-SLLSSLTDEMMQNA-CGITNPIHRLKLTQAFETAKH  {650-717}
C05D10.4        KEWS-LDDVLLWLQSAQMD--DVAGLLIGYDLRG-EDLLQWNDQTLAQ--LGVSNPEIRKKLLDDLSKIIE  {221-285}
C46H3.2_1       SGVPHAEVLANWLDSINMS--NYLAVFLKQGYDL-QTIARCTPADLLS--LGIKNPDHRKKLMSDIHSWKI  {114-180}
C46H3.2_2       SVVP-NNSLREWLHAIALA--EYIPMFESQRYTSVSDVLDLQWEDFED--IGVKRLGHLKRFGLTIKKLKD  {185-251}
DD36938         ARIKKYKDLFDWLQKHGFE--QYKDAFLKEEMFL-DELGEMSEDILNK--MGITSTGTRLRILKETSNLAN  {237-303}
51CP            SGLG-EAGMSAWLRAIGLE--RYEEGLVHNGWDDLEFLSDITEEDLEE--AGVQDPAHKRLLLDTLQLSK-  {60-124}
Pmk1            QDWS-LNSVLQFLKLYKFN-KEWEDVFIKSRIEMDLFINLADQSKAEE--FAFKNKLSKESAIQLSSCIRK  {29-95}

consensus       .ths..psltpWLptltht..pYhpthhptthss.phl.thstpcl.p..hGlp..ttppclhptlpphtt
PHD             LLLLLL.HHHHHHHHHhhL.hHHHHHHHHhh...hhHHHHHhHHHHHHh..L....L.HHHHHHHHHhLL
```

Fig. 1. Multiple alignment of SAM domains. A Blastp (Altschul et al., 1990) search using the Ste11p SAM as the query sequence suggested a previously undetected SAM domain in a *D. discoideum* protein, DD36938 ($p = 3 \times 10^{-4}$, BLOSUM62). A reciprocal Blastp search with DD36938 led to further significant ($p < 0.01$, PAM220) hits, that were used to construct a multiple alignment using ClustalW (Thompson et al., 1994). This alignment was the starting point for iterative profile searches using SWise (Birney et al., 1996). Newly identified SAM domains were included in subsequent alignments that were used for successive iterations. With profiles derived from the alignment shown, all SAM domains scored higher (>3874) than the adjudged first false positive that scored 3764. A subset of EPH-family sequences were chosen as those with an evolutionary distance >0.9 calculated by the ClustalW neighbor-joining method using the BLOSUM30 matrix. (Identifier and Accession code: GGCEK6A_1, Z19110; KSEK_MOUSE, Q03137; KEK5_CHICK, P28693; MMMDK5_1, Z49086; XLERTK_1, L43620; MM06834_1, U06834; XLEPHRTK_1, L43622; MMKIN1_1, X79082; GG23783_1, U23783; HSEHK1_1, X95425; MMECK_1, X78339; HSRTKEPH_1, Z27409; MMMEK4_1, M68513.) Reciprocal complementary Blastp searches (Bork & Gibson, 1996) were performed with each newly identified domain. For example using Blastp (PAM220 matrix) with the C46H3.2 SAM as the query, EPH sequences and the previously described SAM of C33B4.3 were identified with probabilities of aligning by chance between 10^{-6} and 10^{-1}. The last column shows domain boundaries. The consensus line immediately beneath the alignment indicates amino acid residues conserved in 70% of all sequences: h, hydrophobic; l, aliphatic; p, polar; c, charged; s, small; t, turnlike (polar or tiny). PHD (Rost et al., 1994) secondary structure prediction: H/h, helix; L/dot, loop with >80%/70% expected accuracy. These methods were not able to positively identify previously described (Ponting, 1995) SAM domains in *S. cerevisiae* Ste50p and *D. melanogaster* lethal(3)malignant brain tumor (l(3)mbt). However, there are good reasons to believe that these sequences represent SAMs. Ste50p is a likely orthologue of *S. pombe* Ste4p, since apart from a putative leucine zipper motif, these proteins possess the same domain organisation; in addition, *l(3)mbt* shows a domain organization similar to *Scm*. That these sequences were not detected immediately suggests that the family of SAM domain-containing proteins is even more widespread than shown here.

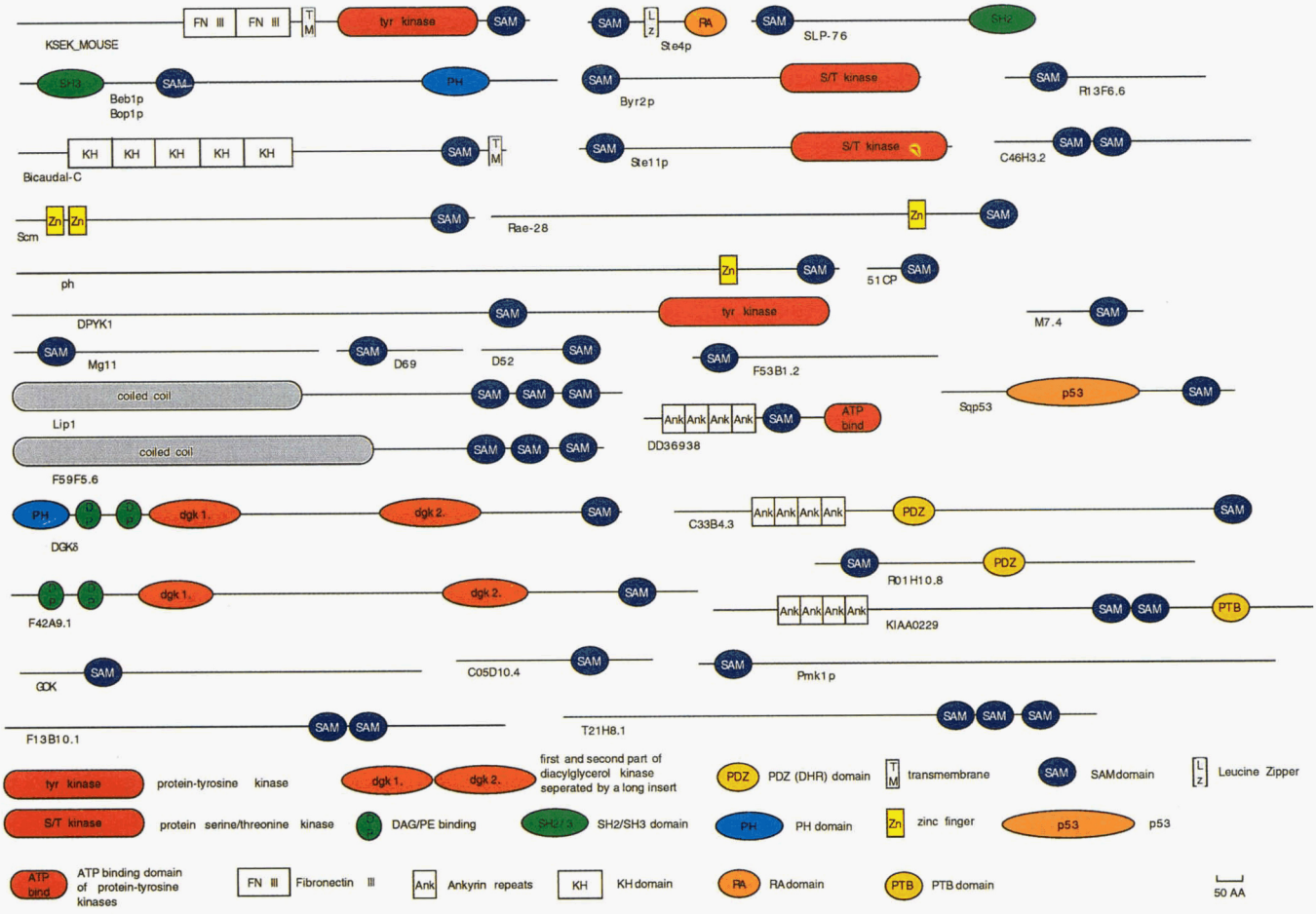

Fig. 2. SAM as a module in diverse proteins (drawn to scale)— As all EPH-family members are identical in their modular construction, just one is shown. Domains shown are: Ank, Ankyrin repeats; C4, C4 zinc finger; DAG, DAG/PE binding; PDZ, PDZ (DHR) domain (Ponting & Phillips, 1995); PTB Phospho-Tyrosine-Binding domain; RA Ras associating domain (Ponting & Benjamin, 1996); FN, fibronectin type-III; KH, KH motif; LZ, Leucine-Zipper; PH, PH domain; SH2, SH2 domain; SH3, SH3 domain; TM, transmembrane.

signal transduction molecules (Fig. 2): single PDZ (or "DHR") domains (Ponting & Phillips, 1995) are present in *C. elegans* hypothetical proteins R01H10.8 and C33B4.3, and diacylglycerol kinase (DAGK) domains are present in DAGK δ isoforms (Sakane et al., 1996).

An unusual occurrence of the SAM domain is as part of the C-terminal extension of the tumor suppressor protein p53 in the Cephalopod *Loligo forbesi* (P. Winge, S. Friend, and J.T. Fleming, unpublished, EMBL accession code U43595). The functional significance of this extension is unknown and no other currently known p53 contains a SAM domain.

Although literature data support roles for SAM domains in developmental regulation, in the majority of cases the molecular functions of SAMs remain unknown. Indeed, the original report of SAM domains (Ponting, 1995) provided no clues as to their function. However, for some of the novel SAM domains identified here, molecular details have been revealed experimentally, strongly suggesting that SAM domains are protein-binding modules. A variety of SAMs have been shown to bind specifically to other SAMs or to SH2 domains.

Deletion experiments with the *S. pombe* proteins Ste4p and Byr2p have shown that they interact via their N-terminal SAM-containing regions (Ste4p: 1-160; Byr2p: 1-392) (Barr et al., 1996). Furthermore, the Byr2p SAM appears to mediate this interaction since substitution of Asn with Ile at residue 28, located at the end of the predicted second α-helix of the Byr2p SAM (Fig. 1), abolishes the interaction with Ste4p (Barr et al., 1996). In addition, SAM-containing portions of Ste4p and its *S. cerevisiae* orthologue, Ste50p, form homodimers (Barr et al., 1996), indicating that SAMs form homotypic and heterotypic homodimers.

The cytoplasmic domain of the Eph-related tyrosine kinase ELK binds to the SH2 domains of Grb10 and Grb2 (Stein et al., 1996). This interaction is dependent on self-phosphorylation of ELK and, at least for Grb10, the presence of tyrosine 929 (Stein et al., 1996). This tyrosine is conserved in the majority of SAM sequences (Fig. 1), suggesting that participation of these domains in signaling pathways involves interaction of SH2 domain-containing proteins with phosphotyrosine. For EPH-like RPTKs, this mechanism is likely to regulate the phenotype of RPTK-expressing cells during bidirectional signaling (Holland et al., 1996).

The region of LIP1 that binds the transmembrane protein tyrosine phosphatase LAR has been determined as amino acids 854–1132 (Serra-Pages et al., 1995). This region is coincident with the threefold repeat of SAM domains (amino acids 875–1113), indicating that this interaction is mediated by at least one of these three domains.

The Ste5p-binding region of Ste11p (residues 1-351) includes its SAM domain (Marcus et al., 1994), as does the Src-like adapter protein-binding region of Eck (Pandey et al., 1995).

We conclude that SAM domains can be described as protein–protein interaction domains, which are able to bind other SAM domains or, via phosphorylation of a conserved tyrosine, SH2 domains. They occur in proteins involved in both signaling and developmental regulation. Since SAM domains occur in *S. cerevisiae*, *C. elegans*, and *Drosophila melanogaster* proteins, good model systems are available for mutational studies of their functions. In particular, the presence of this domain in the large and well-documented family of EPH-related RPTKs (more than 40 distinct entries in the EMBL database) may allow further insight into members of this important family that have been previously implicated in neural and brain development (Xu et al., 1996), carcinogenesis (Iwase et al., 1993) and are overexpressed in various tumors (Kiyokawa et al., 1994; Hirai et al., 1987).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Bardwell L, Cook JG, Inouye CJ, Thorner J. 1994. Signal propagation and regulation in the mating pheromone response pathway of the yeast *Saccharomyces cerevisiae*. *Dev Biol 166*:363–379.

Barr MM, Tu H, van Aelst L, Wigler M. 1996. Identification of Ste4 as a potential regulator of Byr2 in the sexual response pathway of *Schizosaccharomyces pombe*. *Mol Cell Biol 16*:5597–5603.

Bender L, Lo HS, Lee H, Kokojan V, Petersen J, Bender A. 1996. Associations among PH and SH3 domain-containing proteins and Rho-type GTPases in yeast. *J Cell Biol 133*:879–894.

Birney E, Thompson JD, Gibson TJ. 1996. PairWise and SearchWise: Finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Res 24*:2730–2739.

Bork P, Gibson T. 1996. Applying motif and profile searches. *Meth Enzymol 266*:162–184.

Bornemann D, Miller E, Simon J. 1996. The *Drosophila Polycomb* group gene *Sex comb on midleg* (*Scm*) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development 122*:1621–1630.

DeCamillis M, Cheng NS, Pierre D, Brock HW. 1992. The *polyhomeotic* gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with *Polycomb*. *Genes Dev 6*:223–232.

van der Geer P, Hunter T, Lindberg RA. 1994. Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu Rev Cell Biol 10*:251–337.

Gribskov M, McLachlan AD, Eisenberg D. 1987. Profile analysis: Detection of distantly related proteins. *Proc Natl Acad Sci USA 84*:4355–4358.

Hirai H, Maru Y, Hagiwara K, Nishida J, Takaku F. 1987. A novel putative tyrosine kinase receptor encoded by the eph gene. *Science 238*:1717–1720.

Holland SJ, Gale NW, Mbamalu G,Yancopoulos GD, Henkemeyer M, Pawson T. 1996. Bidirectional signalling through the EPH-family receptor Nuk and its transmembrane ligands. *Nature 383*:722–725.

Hynes RO, Lander AD. 1992. Contact and adhesive specificities in the associations, migrations, and targeting of cells and axons. *Cell 68*:303–322.

Iwase T, Tanaka M, Suzuki M, Naito Y, Sugimura H, Kino I. 1993. Identification of protein-tyrosine kinase genes preferentially expressed in embryo stomach and gastric cancer. *Biochem Biophys Res Commun 194*:698–705.

Jackman JK, Motto DG, Sun Q, Tanemoto M, Turck CW, Peltz GA, Koretzky GA, Findell PR. 1995. Molecular cloning of SLP-76, a 76-kDa tyrosine phosphoprotein associated with Grb2 in T cells. *J Biol Chem 270*:7029–7032.

Karin M. 1992. Signal transduction from cell surface to nucleus in development and disease. *FASEB J 6*:2581–2590.

Kiyokawa E, Takai S, Tanaka M, Iwase T, Suzuki M, Xiang YY, Naito Y, Yamada K, Sugimura H, Kino I. 1994. Overexpression of *ERK*, an *EPH* family receptor protein tyrosine kinase, in various human tumors. *Cancer Res 54*:3645–3650.

Mahone M, Saffman EE, Lasko PF. 1995. Localized *Bicaudal-C* RNA encodes a protein containing a KH domain, the RNA binding motif of FMR1. *EMBO J 14*:2043–2055.

Marcus S, Polverino A, Barr M, Wigler M. 1994. Complexes between STE5 and components of the pheromone-responsive mitogen-activated protein kinase module. *Proc Natl Acad Sci USA 91*:7762–7766.

Matsui Y, Matsui R, Akada R, Toh-e A. 1996. Yeast src homology region 3 domain-binding proteins involved in bud formation. *J Cell Biol 133*:865–878.

Neiman AM, Stevenson BJ, Xu HP, Sprague GFJ, Herskowitz I, Wigler M, Marcus S. 1993. Functional homology of protein kinases required for sexual differentiation in *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* suggests a conserved signal transduction module in eukaryotic organisms. *Mol Biol Cell 4*:107–120.

Nomura M, Takihara Y, Shimada K. 1994. Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: One of the early inducible clones encodes a novel protein sharing several highly homologous regions with a *Drosophila Polyhomeotic* protein. *Differentiation 57*:39–50.

Pandey A, Duan H, Dixit VM. 1995. Characterization of a novel Src-like adapter protein that associates with the Eck receptor tyrosine kinase. *J Biol Chem 270*:19201–19204.

Ponting CP. 1995. SAM: A novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. *Protein Sci 4*:1928–1930.

Ponting CP, Benjamin DR. 1996. A novel family of ras binding domains. *Trends Biochem Sci* (in press).

Ponting CP, Phillips C. 1995. DHR domains in syntrophins, neuronal NO synthases and other intracellular proteins. *Trends Biochem Sci 20*:102-103.

Rost B, Sander C, Schneider R. 1994. PHD—An automatic mail server for protein secondary structure prediction. *Comput Appl Biosci 10*:53–60.

Sakane F, Imai S, Wada I, Kanoh H. 1996. Molecular cloning of a novel diacylglycerol kinase isozyme with a pleckstrin homology domain and a C-terminal tail similar to those of the EPH family of protein-tyrosine kinases. *J Biol Chem 271*:8394–8401.

Schupbach T, Wieschaus E. 1991. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics 129*:1119–1136.

Serra-Pages C, Kedersha NL, Fazikas L, Medley Q, Debant A, Streuli M. 1995. The LAR transmembrane protein tyrosine phosphatase and a coiled-coil LAR-interacting protein co-localize at focal adhesions. *EMBO J 14*:2827–2838.

Stein E, Cerretti DP, Daniel TO. 1996. Ligand activation of ELK receptor tyrosine kinase promotes its association with Grb10 and Grb2 in vascular endothelial cells. *J Biol Chem 277*:23588–23593.

Tan JL, Spudich JA. 1990. Developmentally regulated protein-tyrosine kinase genes in *Dictyostelium discoideum*. *Mol Cell Biol 10*:3578–3583.

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.

Tuzi NL, Gullick WJ. 1994. *eph*, the largest known family of putative growth factor receptors. *Br J Cancer 69*:417–421.

Valenzuela DM, Rojas E, Griffiths JA, Compton DL, Gisser M, Ip NY, Goldfarb M, Yancopoulos GD. 1995. Identification of full-length and truncated forms of Ehk-3, a novel member of the Eph receptor tyrosine kinase family. *Oncogene 10*:1573–1580.

Xu Q, Alldus G, Macdonald R, Wilkinson DG, Holder N. 1996. Function of the Eph-related kinase *rtk1* in patterning of the zebrafish forebrain. *Nature 381*:391–322.