Protein Science (1995), *4*:1928–1930. Cambridge University Press. Printed in the USA.

FOR THE RECORD

SAM: A novel motif in yeast sterile and *Drosophila* polyhomeotic proteins

CHRISTOPHER P. PONTING

Fibrinolysis Research Unit, University of Oxford, The Old Observatory, Oxford OX1 3RH, United Kingdom

(RECEIVED May 5, 1995; ACCEPTED June 16, 1995)

Abstract: Single copies of an ≈65–70 residue domain are shown to be present in the sequences of 14 eukaryotic proteins, including yeast byr2, STE11, ste4, and STE50, which are essential participants in sexual differentiation. This domain, named SAM (sterile alpha motif), appears to participate in other developmental processes because it is also present in *Drosophila polyhomeotic* gene product and related homologues, which are thought to regulate determination of segmental specification in early embryogenesis. Its appearance in byr2 and STE11, which are MEK kinases, and in proteins containing pleckstrin homology, *src* homology 3, and discs-large homologous region domains, suggests possible participation in signal transduction pathways.

Keywords: *Drosophila* development; homology; signal transduction; yeast sterile genes

Identification, by sequence analysis, of a homologous domain family, can prompt detailed experimental investigation leading to elucidation of domain and molecular functions. For example, identifications of large families of *src* homology (SH2 and SH3) and pleckstrin homology (PH) domains have facilitated a greater understanding of the regulatory nature of signal transduction pathways (reviewed in Cohen et al., 1995; Pawson, 1995). More recently, further domain families present in signal transduction proteins, including the discs-large homologous region (DHR, or GLGF; recently renamed 'PDZ') (Ponting & Phillips, 1995) and phosphotyrosine interaction domain (PID) (Bork & Margolis, 1995) have been documented. Here it is reported that a novel domain is common to yeast proteins that are essential for sexual responses induced by mating pheromones and to animal proteins that are essential during embryo morphogenesis.

During an investigation of DHR-containing protein sequences, a region of a *Caenorhabditis elegans* putative protein sequence (R01H10.8) was found initially to be similar to *Schizosaccharomyces pombe* ste4 and mouse Mg11 sequences, and eventually to be similar to a total of 13 sequences (Fig. 1). These similarities, particularly conservation of hydrophobic residues throughout the alignment, indicate that these sequences encode homologous domains that have a common evolutionary ancestor. Unlike most other intracellular domains, the residue limits of SAM sequences are well defined, because SAMs in byr2 (residues 1–66) and C33B4.3 (residues 1,045–1,110) begin and end with N-terminal and C-terminal residues, respectively (Fig. 2).

Four proteins, byr2, STE11, ste4, and STE50, which contain this domain, are essential participants in sexual differentiation in yeasts: mutations in their corresponding genes induce sterility (Rhodes et al., 1990; Okazaki et al., 1991; Wang et al., 1991; Ramezani Rad et al., 1992). Consequent to this, and to the all (α-)helical predicted secondary structures of these sequences, this domain has been named SAM, an acronym for sterile alpha motif. *S. pombe* byr 2 (also called ste8) and *Saccharomyces cerevisiae* STE11 are orthologous MEK kinases that participate in the MAP kinase cascades as part of the Ras1 and pheromone response pathways (reviewed in Neiman, 1993; Herskowitz, 1995). The N-terminal noncatalytic regions of STE11 and byr2 contain binding sites for STE5 and Ras1, respectively (Choi et al., 1994; Masuda et al., 1995), as well as negatively regulating kinase activity (Cairns et al., 1992; Stevenson et al., 1992). These protein-protein interactions suggest possible functional roles for the byr2 and STE11 N-terminal SAMs. Other MEK kinases, tobacco NPK1 and mouse MEKK, do not appear to possess a SAM-related sequence.

SAMs also occur within three proteins that share a similar domain composition and that are most similar within their SAM sequences. These are: *Drosophila melanogaster polyhomeotic* gene product (ph) (Dura et al., 1987), mouse RAE-28 (Nomura et al., 1994), and *Drosophila* tumor suppressor gene *lethal(3)malignant brain tumour* product (l(3)mbt) (J. Wismar et al., unpubl.). *Drosophila* ph is thought to be among the Polycomb group of genes that encode chromatin proteins; these maintain the process of spatial regulation during determination of segmental identity in early embryogenesis (Jürgens, 1985; Dura et al., 1987). RAE-28 appears to be a ph counterpart in mouse, and a human ph homologue is partly encoded by an expressed sequence tag (EST, Genbank code T09455), implying that this developmental control gene is present across the animal kingdom.

Reprint requests to: Christopher P. Ponting, Fibrinolysis Research Unit, University of Oxford, The Old Observatory, South Parks Road, Oxford OX1 3RH, United Kingdom; e-mail: chris@biop.ox.ac.uk.

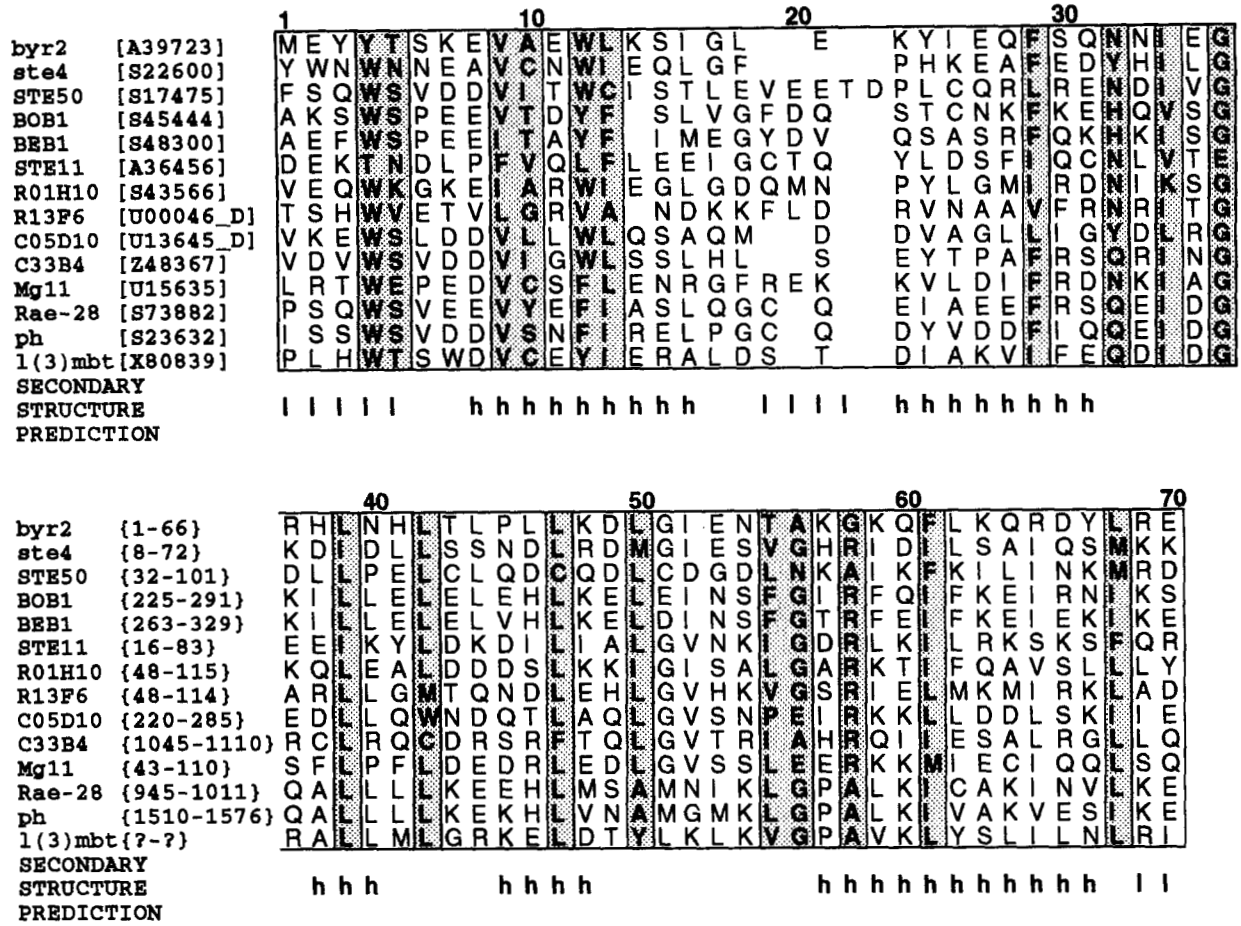

Fig. 1. Multiple alignment of 14 SAM sequences, displayed using Alscript (Barton, 1993a). An initial database search with the *C. elegans* R01H10.8 sequence, using BLAST (Altschul et al., 1994) at the NCBI yielded two candidate homologues, *S. pombe* ste4 and mouse Mg11. These three sequences showed pairwise sequence similarities (*Z*-scores >8.5σ, probabilities of matching by chance, $P < 1 \times 10^{-4}$, $P < 8 \times 10^{-7}$, $P < 6 \times 10^{-7}$) indicative of homology. Subsequent sequence alignment using AMPS (Barton, 1990) yielded a profile and a pattern that were scanned against PIR (v44), SWISSPROT, and PATCHX databases, using local similarity and pattern-matching algorithms (Barton & Sternberg, 1989; Barton, 1993b). Sequences were added to the alignment if they scored significantly higher than both the adjudged levels of noise in both scanning procedures. A further three iterations of this procedure, as well as identification of SAMs in *C. elegans* C33B4.3 and *Drosophila* l(3)mbt using BLAST (Altschul et al., 1994) yielded a total of 13 SAMs. A final SAM from the byr2 homologue, *S. cerevisiae* STE11, was appended as it scored at levels equivalent to the highest "noise" score. Sequences from the *ets* gene family (e.g., PIR accession code TVHUE2) scored among the highest noise scores, yet considerable dissimilarity was noted between SAM and ets consensus sequences at their C-termini; therefore, no positive identification of ets sequences as containing SAMs was possible. In these procedures, the BLOSUM62 matrix was used throughout. The highest probability that either of two alignment "blocks" (positions 1–13 and 29–70) was arrived at by chance was calculated using MACAW (Schuler et al., 1991) as $P < 3 \times 10^{-12}$. The average of pairwise *Z*-scores and percentage identities among the 14 SAMs was 5.7σ and 22%, respectively (excluding two pairs of sequences where percentage identities were above 60%). Secondary structures, predicted using the program PHD at an expected accuracy of >82% (Rost & Sander, 1993), are indicated by **h** (helices) and **l** (loops). Accession codes and residue limits are given in parentheses. The ph sequence shown corresponds to the proximal unit of the two independent repeats. Positions where the chemical character of residues is conserved in >75% are shown in bold and shaded.

The remaining seven SAM-containing proteins possess no well-documented functions. *S. cerevisiae* BOB1 and BEB1 are proteins that bind an SH3-containing molecule, BEM1, and each contains SH3 and PH domains, implying their participation in signal transduction pathways. *C. elegans* putative proteins R01H10.8 and C33B4.3 each possess a single DHR domain, which again suggests their participation in signalling. Mouse Mg11 possesses an N-terminal SAM, followed by a C-terminal tail homologous to N-terminal regions of two putative proteins of unknown function: *C. elegans* ZK177.8 and *Bacillus subtilis ipa-93d* gene product (accession codes U21321 and S39748).

The functions of SAMs remain to be determined although their small size (65–70 residues) precludes any catalytic role. As outlined for byr2 and STE11, they may mediate a protein-binding function. In contrast, it is possible that they may possess DNA-binding functions, via recognition helices. This is suggested by the high content, in SAMs, of short helices separated by turn-forming glycines (Fig. 1), which are features common

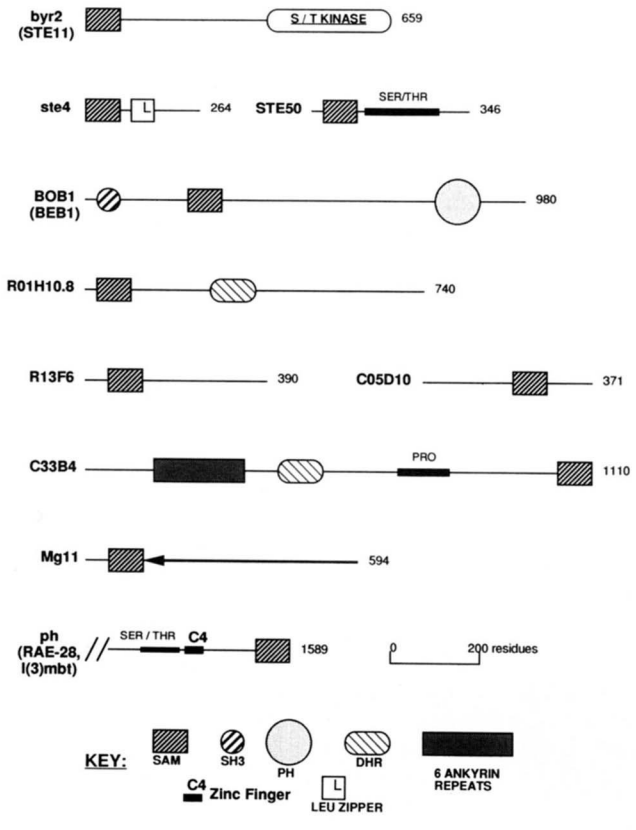

Fig. 2. Domain organization of SAM-containing proteins. Close homologues are given in parentheses. Domains shown are: SH3, PH (Cohen et al., 1995; Pawson, 1995, and references therein), DHR (Ponting & Phillips, 1995), ankyrin repeats (Bork, 1993), C4 zinc finger (Evans & Hollenberg, 1988), and leucine zipper (Landschulz et al., 1988). Regions rich in particular amino acids are shown by thick lines. The C-terminal region of Mg11 (shown by an arrow) is homologous to *C. elegans* putative protein ZK177.8 and *B. subtilis* putative *ipa-93d* gene product. Total numbers of amino acids are shown at their C-terminal ends.

to other DNA-binding sequences such as helix-turn-helix and helix-loop-helix motifs. Whatever their molecular role, their presence in such a variety of eukaryotic proteins indicates a general function in cell differentiation in organisms as divergent as yeast and vertebrates.

Acknowledgments: This work was supported by the Medical Research Council (UK) and the Oxford Centre for Molecular Sciences. I thank R.B. Russell for many valuable discussions and G.J. Barton for construction of sequence databases.

References

Altschul SF, Boguski MS, Gish W, Wootton JC. 1994. Issues in searching molecular sequence databases. *Nature Genet 6*:119–129.

Barton GJ. 1990. Protein multiple sequence alignment and flexible pattern matching. *Methods Enzymol 183*:403–428.

Barton GJ. 1993a. ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng 6*:37–40.

Barton GJ. 1993b. An efficient algorithm to locate all locally optimal alignments between two sequences allowing for gaps. *Comput Appl Biosci 9*:729–734.

Barton GJ, Sternberg MJE. 1989. Flexible protein sequence patterns. A sensitive method to detect weak structural similarities. *J Mol Biol 212*:389–402.

Bork P. 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? *Proteins Struct Funct Genet 17*:363–374.

Bork P, Margolis B. 1995. A phosphotyrosine interaction domain. *Cell 80*: 693–694.

Cairns BR, Ramer SW, Kornberg RD. 1992. Order of action of components in the yeast pheromone response pathway revealed with a dominant allele of the STE11 kinase and the multiple phosphorylation of the STE7 kinase. *Genes & Dev 6*:1305–1318.

Choi KY, Satterberg B, Lyons DM, Elion EA. 1994. Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in *S. cerevisiae*. *Cell 78*:499–512.

Cohen GB, Ren R, Baltimore D. 1995. Modular binding domains in signal transduction proteins. *Cell 80*:237–248.

Dura JM, Randsholt NB, Deatrick J, Erk I, Santamaria P, Freeman JD, Freeman SJ, Weddell D, Brock HW. 1987. A complex genetic locus, *polyhomeotic*, is required for segmental specification and epidermal development in *D. melanogaster*. *Cell 51*:829–839.

Evans RM, Hollenberg SM. 1988. Zinc fingers: Gilt by association. *Cell 52*:1–3.

Herskowitz I. 1995. MAP kinase pathways in yeast: For mating and more. *Cell 80*:187–197.

Jürgens G. 1985. A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature 316*:153–155.

Landschulz WH, Johnson PF, McKnight SJ. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science 240*:1759–1764.

Masuda T, Kariya K, Shinkai M, Okada T, Kataoka T. 1995. Protein kinase Byr2 is a target of Ras1 in the fission yeast *Schizosaccharomyces pombe*. *J Biol Chem 270*:1979–1982.

Neiman AM, Stevenson BJ, Xu HP, Sprague GF Jr, 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 highly homologous regions with a *Drosophila polyhomeotic* protein. *Differentiation 57*:39–50.

Okazaki N, Okazaki K, Tanaka K, Okayama H. 1991. The ste4$^+$ gene, essential for sexual differentiation of *Schizosaccharomyces pombe*, encodes a protein with a leucine zipper motif. *Nucleic Acids Res 19*:7043–7047.

Pawson T. 1995. Protein modules and signalling networks. *Nature 373*: 573–579.

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

Ramezani Rad M, Xu G, Hollenberg CP. 1992. STE50, a novel gene required for activation of conjugation at an early step in mating in *Saccharomyces cerevisiae*. *Mol Gen Genet 236*:145–154.

Rhodes N, Connell L, Errede B. 1990. STE11 is a protein kinase required for cell-type-specific transcription and signal transduction in yeast. *Genes & Dev 4*:1862–1874.

Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol 232*:584–599.

Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins Struct Funct Genet 9*:180–190.

Stevenson BJ, Rhodes N, Errede B, Sprague GF Jr. 1992. Constitutive mutants of the protein kinase STE11 activate the yeast pheromone response pathway in the absence of the G protein. *Genes & Dev 6*:1293–1304.

Wang Y, Xu HP, Riggs M, Rodgers L, Wigler M. 1991. *byr2*, a *Schizosaccharomyces pombe* gene encoding a protein kinase capable of partial suppression of the *ras1* mutant phenotype. *Mol Cell Biol 11*:3554–3563.