LETTER TO THE EDITOR

The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence

We compared the sequence-specific DNA binding properties of the proteins encoded by the murine homolog of the *ets*-1 proto-oncogene and two *ets*-related genes, murine *PU.1* and *Drosophila E74*. The protein products of these genes share sequence similarity within a region of ~85 amino acids that we propose to call the ETS-domain. This amino acid sequence similarity, combined with the observation that these proteins bind similar DNA sequences, provides the basis for our proposal that the ETS-domain defines a new family of eukaryotic DNA-binding proteins.

The proto-oncogene *ets*-1 was discovered first as one of two cellular sequences transduced by the avian retrovirus E26. Chicken *ets*-1 and *myb* gene sequences are fused to *gag* sequences to form the tripartite E26 oncogene (Leprince et al. 1983; Nunn et al. 1983). The *ets* portion of this oncogene is required for the induction of erythroblastosis in infected chickens (Nunn and Hunter 1989) and also appears to affect the myeloid transformation process (Golay et al. 1988).

Over the last five years, a variety of genes have been described whose protein products have predicted amino acid sequences that show strong similarity with that of the avian *ets*-1 proto-oncogene. These include human *ets*-1 (Watson et al. 1988), murine *ets*-1 (Gunther et al. 1990), *ets*-2 (murine, human: Watson et al. 1988; avian: Boulukos et al. 1988), human *erg* (Reddy et al. 1987), human *elk*-1 and *elk*-2 (Rao et al. 1989), *Drosophila E74* (Burtis et al. 1990), *Drosophila ets*-2 (Pribyl et al. 1988), and murine *PU.1* (Klemsz et al. 1990). Representatives of each of the protein products of these genes are aligned in Figure 1 according to the ~85-amino-acid region of strong conservation. We designate this region the ETS-domain. Five of the seven proteins listed have their ETS-domains near the carboxyl terminus. Twenty-two amino acid positions are identical among all members. At 12 positions, only one of the seven sequences is dissimilar from the others. The ETS-domains of Ets-1 and Ets-2 are the most similar, whereas those of PU.1 and E74 are the most divergent from Ets-1. Outside of the ETS-domain, most of the sequences diverge significantly. An exception is Ets-2, which is ~50% identical to Ets-1 (Watson et al. 1988). The human *erg* protein also contains some, more limited, sequence similarity with Ets-1 and Ets-2 proteins, outside of the ETS-domain (Reddy et al. 1987).

A role for this conserved protein domain has been demonstrated recently by our discovery of the sequence-specific DNA binding activities of three ETS-domain proteins: murine Ets-1, murine PU.1, and *Drosophila* E74A (Gunther et al. 1990; Klemsz et al. 1990; Urness and Thummel 1990). The murine *ets*-1 homolog was isolated by screening a cDNA expression library with a Moloney murine sarcoma virus (MSV) LTR promoter sequence (Gunther et al. 1990). In a similar manner, a murine *PU.1* cDNA was isolated by screening an expression library with a regulatory sequence upstream from the MHC class II I-Aβ gene. The *PU.1* protein product also binds to a purine-rich sequence within the SV40 enhancer (Klemsz et al. 1990). The *Drosophila ets*-related gene *E74* was isolated by virtue of its location within an early ecdysone-inducible puff. This complex gene encodes two related proteins, designated E74A and E74B. These proteins have unique sequences at their amino termini joined to an identical carboxy-terminal region that contains the E74 ETS-domain (Burtis et al. 1990).

The sequence-specific DNA binding of PU.1, E74A, and Ets-1 suggests that their shared ETS-domains may mediate this common activity. Indeed, fragments of the PU.1 and Ets-1 proteins that retain sequence-specific DNA binding activity contain the conserved ETS-domain (Gunther et al. 1990; Klemsz et al. 1990). The configuration of conserved amino acids within the ETS-domain, however, has no apparent similarity to any well-characterized DNA-binding motifs (Johnson and McKnight 1989; Mitchell and Tjian 1989). There are no appropriately spaced cysteines or histidines to stabilize a zinc finger structure. Furthermore, the conserved α-helical permissive regions are not spaced in a recognizable helix–turn–helix configuration. Potential α-helices also do not show the amphipathy speculated to be crucial in the helix–loop–helix model (Murre et al. 1989). The proto-oncogene *ets*-1 encodes three tryptophans with 17- to 18-amino-acid spacing (Fig. 1) similar to the tryptophan repeats observed in *myb* and *myb*-related protein products (Anton and Frampton 1988). These tryptophans are highly conserved among the ETS-domain proteins, PU.1 being the only exception, with a substitution of a tyrosine for one of the tryptophans. Although basic amino acids are concentrated in the carboxyl half of the ETS-domain, they cannot be aligned readily with either the basic amino acids of the leucine zipper/basic region family of proteins (Landschulz et al. 1989) or with the basic domain of the helix–loop–helix family of proteins (Murre et al. 1989). Therefore, we conclude that the ETS-domain encodes a new structural motif for binding DNA and that proteins containing this conserved sequence, the ETS-domain proteins, constitute a new family of eukaryotic DNA-binding proteins.

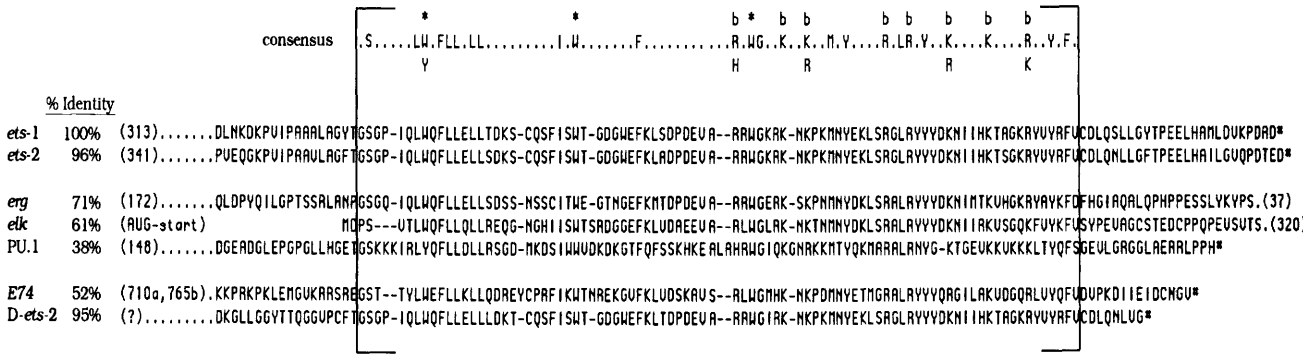

Figure 1. Alignment of conserved amino acids within the ETS-domain. The amino acid sequences of the *ets*-related proteins are aligned within their ETS-domains: murine Ets-1 (Gunther et al. 1990), human Ets-2 (Boulukos et al. 1988; Watson et al. 1988), human Erg (Reddy et al. 1987), human Elk (Rao et al. 1989), murine PU.1 (Klemsz et al. 1990), *Drosophila* E74 (Burtis et al. 1990), and *Drosophila* Ets-2 (Pribyl et al. 1988). The ETS-domain, shown in brackets, consists of 85 amino acids in Ets-1. Listed at the top is the consensus sequence which consists of the 22-amino-acid identities and some positions with conservative amino acid substitutions. Positions that have conserved basic amino acids (b) and the conserved tryptophan residues (*) are indicated above the consensus. Listed to the *left* of each sequence is its name, the number of residues to its putative start codon, and its percent identity with the ETS-domain of the murine Ets-1 protein. (The human, murine and avian Ets-1 homologs are 100% identical within the ETS-domain.) Stop codons are shown at the end of each sequence, or the number of residues to the stop codon for Erg and Elk. We note that *Drosophila* Ets-2 does not appear to be an Ets-2 homolog, as proposed originally, because its sequence similarity with Ets-2 does not extend outside of the ETS-domain (Pribyl et al. 1988).

A comparison of the DNA sequences bound by Ets-1, PU.1, and E74A suggests that a common binding specificity may be shared among ETS-domain proteins. As shown in Figure 2, each of the ETS-domain binding sites contain a centrally located, purine-rich sequence, 5′-GGAA-3′. We speculate that recognition of this sequence may be a conserved feature of all ETS-domain proteins because the DNA-binding sites of Ets-1 and the ETS-domain proteins most divergent from Ets-1, PU.1, and E74, contain this GGAA motif.

Three independent forms of experimental evidence are consistent with a role for the GGAA sequence in ETS-domain DNA binding. First, methylation interference analysis identified the two guanine residues within the GGAA sequence as close contact points for Ets-1 and PU.1 protein binding (Gunther et al. 1990; Klemsz et al. 1990). Second, the binding of PU.1 to an oligonucleotide with the sequence GGAA was not competed with a mutant oligonucleotide bearing the sequence CCAA (Klemsz et al. 1990). Similarly, site-directed mutagenesis of the MSV LTR promoter (GCGGAA changed to CGCCAA) eliminated Ets-1 binding (Gunther et al. 1990). Third, E74A was used to select oligonucleotides of random sequence to statistically define a consensus sequence for binding (Urness and Thummel 1990). This analysis demonstrated that the GGAA sequence is essential for E74A binding and must be preceded by either an A or C. A precise assessment of the significance of the flanking nucleotides will require further biochemical characterization of these and other ETS-domain proteins.

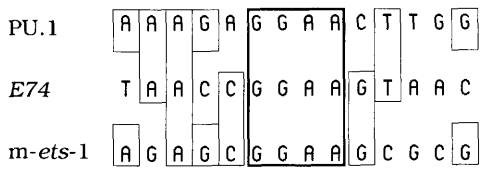

Figure 2. Comparison of the DNA sequences in the binding sites for PU.1 (SV40 enhancer, position 304 to 330; Klemsz et al. 1990), Ets-1 (MSV LTR; Gunther et al. 1990) and the E74 protein (within an *E74* intron; Urness and Thummel, 1990). The DNA sequences are aligned by the common GGAA motif.

Our studies of PU.1, E74A, and Ets-1 suggest that ETS-domain proteins function as transcriptional regulators. PU.1 *trans*-activates a reporter gene placed downstream of PU.1 binding sites in transient expression assays in tissue culture cells (Klemsz et al. 1990). Likewise, disruption of the Ets-1 binding site in the MSV LTR reduces the transcriptional activity of this viral promoter (Gunther et al. 1990). Although there is no established function for the DNA binding activity of E74A, the *E74* gene is postulated to play a central role in the ecdysone regulatory hierarchy during *Drosophila* metamorphosis (Burtis et al. 1990; Urness and Thummel 1990).

The tissue distribution of the known mammalian ETS-domain proteins suggests that they could regulate the expression of tissue-specific genes. For example, PU.1 is expressed in macrophages and B-cells (Klemsz et al. 1990); Ets-1 is most abundant in T lymphocytes (Bhat et al. 1987, 1989), and *elk* gene expression is restricted to the lung and testis (Rao et al. 1989). On the other hand, Ets-2 is more broadly distributed such that, in T lymphocytes, *ets*-1 and *ets*-2 are both expressed (Bhat et al. 1987, 1990). Similarly, in *Drosophila*, the *E74A* and *E74B* mRNAs have overlapping temporal and spatial patterns of expression (Burtis et al. 1990; Thummel et al. 1990). In these cases, ETS-domain proteins may compete for occupancy of a common set of binding sites. Further studies of the known ETS-domain proteins and the iso-

lation of new members of this family should provide insights into how this novel DNA-binding domain is regulated.

Acknowledgment

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Felix D. Karim, Lisa D. Urness, Carl S. Thummel

Howard Hughes Medical Institute, Department of Human Genetics, University of Utah Medical Center, Salt Lake City, Utah 84132 USA

Michael J. Klemsz, Scott R. McKercher, Antonio Celada, Charles Van Beveren, Richard A. Maki

Cancer Research Center, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, California 92037 USA

Cathy V. Gunther, Julie A. Nye, Barbara J. Graves

Department of Cellular, Viral, and Molecular Biology, University of Utah Medical Center, Salt Lake City, Utah 84132 USA

References

Anton, I.A. and J. Frampton. 1988. Tryptophans in *myb* proteins. *Nature* **336:** 719.

Bhat, N.K., R.J. Fisher, S. Fujiwara, R. Ascione, and T.S. Papas. 1987. Temporal and tissue-specific expression of mouse *ets* genes. *Proc. Natl. Acad. Sci.* **84:** 3161–3165.

Bhat, N.K., K.L. Komschlies, S. Fujiwara, R.J. Fisher, B.J. Mathieson, T.A. Gregorio, H.A. Young, J.W. Kasik, K. Ozato, and T.S. Papas. 1989. Expression of *ets* genes in mouse thymocyte subsets T cells. *J. Immunol.* **142:** 672–678.

Bhat, H.K., C.B. Thompson, T. Lindsten, C.H. June, S. Fujiwara, S. Koizum, R.J. Fisher, and T.S. Papas. 1990. Reciprocal expression of human *ETS1* and *ETS2* genes during T-cell activation: Regulatory role for the protooncogene *ETS1*. *Proc. Natl. Acad. Sci.* **87:** 3723–3727.

Boulukos, K.E., P. Pognonec, A. Begue, F. Galibert, J.C. Gesquiere, D. Stehelin, and J. Ghysdael. 1988. Identification in chickens of an evolutionarily conserved cellular *ets*-2 gene (c-*ets*-2) encoding nuclear proteins related to the products of the c-*ets* proto-oncogene. *EMBO J.* **7:** 697–705.

Burtis, K.C., C.S. Thummel, C.W. Jones, F.D. Karim, and D.S. Hogness. 1990. The *Drosophila* 74EF early puff contains *E74*, a complex ecdysone-inducible gene that encodes two *ets*-related proteins. *Cell* **61:** 85–99.

Golay, J., M. Introna, and T. Graf. 1988. A single point mutation in the v-*ets* oncogene affects both erythroid and myelomonocytic cell differentiation. *Cell* **55:** 1147–1158.

Gunther, C.V., J.A. Nye, R.S. Bryner, and B.J. Graves. 1990. Sequence-specific DNA binding of the proto-oncoprotein *ets*-1 defines a transcriptional activator sequence within the long terminal repeat of the Moloney murine sarcoma virus. *Genes Dev.* **4:** 667–679.

Johnson, P.F. and S.L. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58:** 799–839.

Klemsz, M.J., S.R. McKercher, A. Celada, C. Van Beveren, and R.A. Maki. 1990. The macrophage and B cell specific transcription factor PU.1 is related to the *ets* oncogene. *Cell* **61:** 113–124.

Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1989. The DNA-binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243:** 1681–1699.

Leprince, D., A. Gegonne, J. Coll, C. de Taisne, A. Schneeberger, C. Lagrou, and D. Stehelin. 1983. A putative second cell-derived oncogene of the avian leukemia retrovirus E26. *Nature* **306:** 395–397.

Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245:** 371–378.

Murre, C., P.S. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. *Cell* **56:** 777–783.

Nunn, M.F. and T. Hunter. 1989. The *ets* sequence is required for induction of erythroblastosis in chickens by avian retrovirus E26. *J. Virol.* **63:** 398–402.

Nunn, M.F., P.H. Seeburg, C. Moscovici, and P.H. Duesberg. 1983. Tripartite structure of the avian erythroblastosis virus E26 transforming gene. *Nature* **306:** 391–395.

Pribyl, L.J., D.K. Watson, M.J. McWilliams, R. Ascione, and T.S. Papas. 1988. The *Drosophila ets*-2 gene: Molecular structure, chromosomal localization and developmental expression. *Dev. Biol.* **127:** 45–53.

Rao, V.N., K. Huebner, M. Isobe, A. Ar-Rushdi, C.M. Croce, and E.S.P. Reddy. 1989. *elk*, tissue-specific *ets*-related genes on chromosomes X and 14 near translocation breakpoints. *Science* **244:** 66–70.

Reddy, E.S.P., V.N. Rao, and T.S. Papas. 1987. The *erg* gene: A human gene related to the *ets* oncogene. *Proc. Natl. Acad. Sci.* **84:** 6131–6135.

Thummel, C.S., K.C. Burtis, and D.S. Hogness. 1990. Spatial and temporal patterns of E74 transcription during *Drosophila* development. *Cell* **61:** 101–111.

Urness, L.D. and C.S Thummel. 1990. Molecular interactions within the ecdysone regulatory hierarchy: DNA binding properties of the *Drosophila* ecdysone-inducible E74A protein. *Cell* (in press).

Watson, D.K., M.J. McWilliams, P. Lapis, J.A. Lautenberger, C.W. Schweinfest, and T.S. Papas. 1988. Mammalian *ets*-1 and *ets*-2 genes encode highly conserved proteins. *Proc. Natl. Acad. Sci.* **85:** 7862–7866.

The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence.

F D Karim, L D Urness, C S Thummel, et al.

Genes Dev. 1990, **4:**
Access the most recent version at doi:10.1101/gad.4.9.1451