LETTER TO THE EDITOR

The POU domain: a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2*, and *Caenorhabditis elegans unc-86* gene products

We describe a large, 150- to 160-amino-acid-long region of sequence similarity in the three mammalian proteins Pit-1, Oct-1, and Oct-2 and in the product of the *unc-86* gene of the nematode *Caenorhabditis elegans* called POU (pronounced 'pow'). This domain contains homeobox-related and POU-specific subdomains. The three mammalian proteins have been implicated in transcriptional control, and *unc-86* is a cell lineage and cell differentiation gene. Pit-1 is a rat pituitary-specific transcription factor that regulates expression of the prolactin and growth hormone genes, normally expressed in two different cell types (lactotrophs and somatotrophs) of the anterior pituitary gland, and recognizes the sequence motif ATGNATA$^A/_T$$^A/_T$ (Nelson et al. 1988). Expression of Pit-1 in heterologous cells results in the activation of cotransfected growth hormone and prolactin promoters (Ingraham et al. 1988). Oct-1 and Oct-2 are a pair of differentially expressed human promoter-binding proteins that recognize the same octameric sequence, ATG-CAAAT. This octamer motif is an important *cis*-regulatory element in the promoters of ubiquitously expressed genes (e.g., small nuclear RNA and histone H2B genes) and cell-specific promoters (i.e., immunoglobulin promoters) (for references, see Sturm et al. 1988). Oct-1 is expressed ubiquitously and probably represents the octamer-binding protein variously referred to as NF-A1, OTF-1, NFIII, and OBP100. OTF-1 and NFIII were shown to be identical and to be both a transcription factor and an adenovirus DNA replication factor (O'Neill et al. 1988). In contrast to Oct-1, Oct-2 is a B-cell-specific transcription factor [previously referred to as NF-A2 or OTF-2 (see Clerc et al. 1988)] that can activate immunoglobulin gene transcription in vitro (Scheidereit et al. 1987; LeBowitz et al. 1988). The biochemical activity of the *unc-86* gene product is unknown, but loss of *unc-86* function results in altered neuroblast lineages and failure in the differentiation of at least one class of neuron (Chalfie et al. 1981; Desai et al., 1988).

Figure 1 shows schematically the location of the POU domain and its subdomains, described below, within the four POU proteins, and Figure 2 shows a comparison of the predicted amino acid sequences of Pit-1 (Ingraham et al. 1988), Oct-1 (Sturm et al. 1988, this issue), Oct-2 (Clerc et al. 1988, this issue), and *unc-86* (Finney et al. 1988) in the region of sequence similarity. Table 1 shows the percentage of residues that are identical among all four proteins and in the six different pair-wise combinations. As aligned in Figure 2, the region of similarity extends over 155 (Pit-1), 159 (Oct-1), 162 (Oct-2) and 162

(*unc-86*) amino acid residues. The entire region is referred to as the POU domain, for the Pit-1, Oct-1 and Oct-2, and *unc-86* region of sequence similarity. These four proteins are identical at 34% of the 170 possible positions (58/170) over the entire POU domain. In any pair-wise combination, however, there is a higher level of identity. The Oct-1 and Oct-2 proteins are very related, displaying 87% identity over the entire POU domain. These two proteins also display sequence relationships outside of the POU domain (Clerc et al. 1988; Sturm et al. 1988) and thus are the most highly related pair of POU proteins. Aside from the close Oct-1 and Oct-2 relationship, the four proteins display similar levels of sequence identity, ranging from 42% (Oct-2 : *unc-86*) to 52% (Oct-1 : Pit-1), in the POU domain but no obvious relationship outside of this region.

From the alignment of the four sequences in Figure 2, two major subdomains of the POU domain can be identified. The 60 amino acid carboxy-terminal subdomain has an overall identity of 37% among the four proteins and is related to the homeo box domain first identified in *Drosophila* homeotic gene products (McGinnis et al. 1984; Scott and Weiner 1984). The relationship between each of the four POU homeo box domains and other homeo boxes is described in detail elsewhere (Clerc et al. 1988; Finney et al. 1988; Ingraham et al. 1988; Ko et al. 1988; Sturm et al. 1988). Briefly, the sequences of these four POU homeo boxes are more closely related to one another than to other homeo domains. The carboxy-terminal third of the homeo domain (labeled WFC) is particularly well matched (71% match over 17 residues). We refer to this subregion as WFC because the tryptophan (W) and phenylalanine (F) residues are extremely well conserved among homeo box domains, whereas the cysteine (C) is unique to the POU class of homeo boxes. This WFC region contains the DNA recognition helix of a putative helix-turn-helix motif that has been described in all homeo domains identified to date (Laughon and Scott 1984; for review, see Scott et al. 1989). Consistent with this sequence similarity being within a DNA recognition domain, the Pit-1 and Oct proteins bind to similar DNA sequences (Nelson et al. 1988; see Ingraham et al. 1988) and mutations in the helix-turn-helix motif of the Oct proteins impair binding to the octamer sequence (Ko et al. 1988; Sturm and Herr 1988).

The POU homeo box is separated by a short (15–27 amino acids) region of sequence dissimilarity from the second major POU subdomain, a 75- to 82-amino-acid region that displays 43% identity among the four pro-
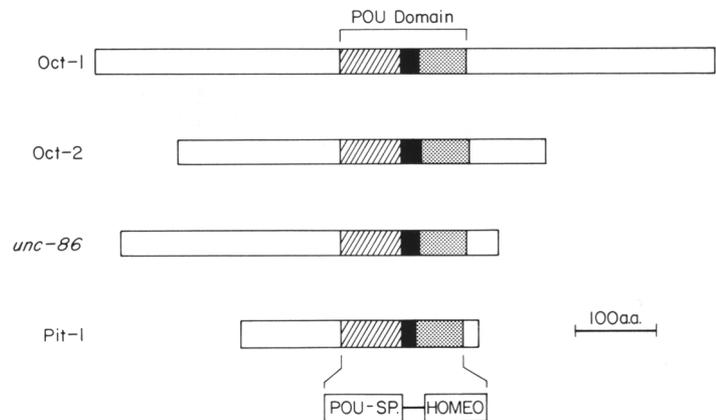
Figure 1. Location of the POU domain in the four POU proteins, Oct-1, Oct-2, *unc-86*, and Pit-1. The open rectangles are proportional representations of the size of the four POU proteins (Oct-1, 743 amino acids (a.a.); Oct-2, 467 a.a.; *unc-86*, 467 a.a.; Pit-1, 291 a.a.) and are aligned with respect to the POU domain. The POU-specific box (POU-SP) is hatched, the POU homeo box (HOMEO) is stippled, and the nonconserved linker region is blackened, as indicated by the expanded and labeled POU domain at the *bottom*.

teins. It is this region that most strongly characterizes the POU proteins as a distinct class. This POU-specific region can be subdivided further into two highly conserved regions (A and B), which overall display 65% and 53% identity among the four proteins. These two subregions are separated by an identical 8-amino-acid sequence in Oct-1 and Oct-2; but these residues differ in Pit-1, and *unc-86* contains an extra three amino acids in this region.

Comparison of the nucleotide sequence of the *oct-1* (Sturm et al. 1988) and *oct-2* (Clerc et al. 1988) genes reveals that the nearly identical POU-specific regions of these two proteins (99% match; see Fig. 2) have been conserved at the amino acid level but not at the nucleotide level, because the nucleotide sequence at silent positions has diverged greatly. This comparison suggests that the POU-specific region has been conserved through functional constraints on the protein products. This conclusion is substantiated further by the high degree of sequence conservation in proteins present in organisms as diverse as nematodes and humans.

It is not immediately evident why the POU family of proteins share such a large region of sequence identity.

In other homeo-box-containing proteins that have no additional large conserved domain, the homeo box domain appears to be sufficient for DNA sequence recognition (Desplan et al. 1988; Hoey et al. 1988). This result would suggest that the POU-specific region is not necessary for DNA binding. Nevertheless, the close relationship and distinct nature of the POU proteins in the region of the putative homeo box DNA recognition helix (WFC) may indicate that these proteins form a different type of DNA-binding protein in which the homeo box is insufficient for DNA binding. Consistent with this interpretation, deletion of the Oct-1 POU A box debilitates Oct-1 DNA-binding activity in a gel retardation assay (Sturm and Herr 1988). The prd box, which is also associated with a particular prd-type homeo domain in certain *Drosophila* genes (Baumgartner et al. 1987), may represent a similar type of structure as the POU domain.

The four POU-related proteins were characterized initially either biochemically (Pit-1, Oct-1, and Oct-2) or genetically (*unc-86*) and not by sequence similarity. There will probably be other homologous proteins that contain POU-related domains. At present, it is not known whether the POU-specific and POU homeo box

Table 1. *Sequence similarity among the POU domains of the Oct-1, Oct-2, Pit-1, and* unc-86 *proteins*

Sequence comparison (number of positions)	POU domain 159–170 (%) (no.)	POU subdomains				
		POU-specific 75–83 (%) (no.)	A 26 (%)	B 34 (%)	homeo 60 (%)	WFC 17 (%)
Oct-1:Oct-2	87 (162)	99 (75)	96	100	88	100
Oct-1:Pit-1	52 (164)	61 (80)	73	74	57	82
Oct-1:*unc-86*	45 (166)	52 (82)	73	59	47	76
Oct-2:Pit-1	50 (167)	61 (80)	73	74	55	82
Oct-2:*unc-86*	42 (169)	52 (82)	73	59	45	76
Pit-1:*unc-86*	44 (159)	53 (83)	65	62	45	76
POU family	34 (170)	43 (83)	65	53	37	71

The percentage of identical amino acids over the total number of residues for each region is shown for each pair-wise combination and for all four POU proteins together. The sequence alignment used and the boundaries of the different subdomains are as shown in Fig. 2. The entire POU domain extends from the amino terminus of the POU-specific box to the carboxyl terminus of the POU homeo box. The total number of positions in each region is shown at the top of each column, except for where the number is variable, in which case the number of positions for each comparison is shown in parentheses.
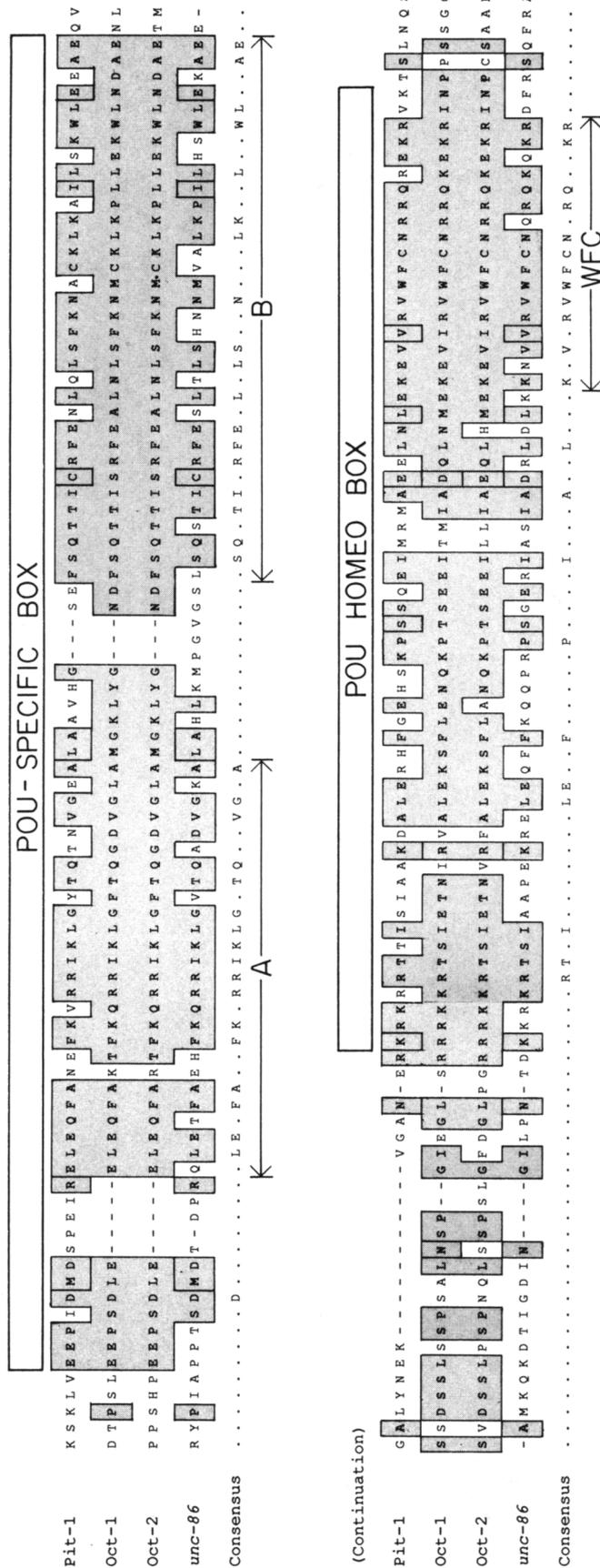
POU- SPECIFIC BOX

POU HOMEO BOX

Figure 2. Sequence comparison of the POU domains in the mammalian gene products Pit-1, Oct-1 and Oct-2, and the *C. elegans unc-86* gene product. The predicted sequences are as described by Ingraham et al. (1988) for Pit-1, Sturm et al. (1988) for Oct-1, Clerc et al. (1988) for Oct-2, and Finney et al. (1988) for *unc*-86. The sequences are shown from left to right in the amino- to carboxy-terminal direction and continue from the top portion into the lower portion of the diagram. Amino acids that are present in two or more proteins at a single position are boxed and shaded. A consensus sequence showing the amino acids in those positions that are identical in all four proteins is shown below the set of four sequences. In this letter the POU domain is defined as the region extending from the amino terminus of the POU-specific box to the carboxyl terminus of the POU homeo box. The different POU subdomains discussed in the text are delineated above and below the sequences.

subregions can be functionally dissociated. A computer-aided search of known protein and DNA sequences revealed no sequences with a high similarity to the POU-specific box.

The POU proteins themselves may well be related functionally, as well as structurally. The Pit-1 and Oct-2 proteins are involved in cell-specific transcriptional regulation; the unc-86 protein controls cell-specific developmental fates. It seems likely that the unc-86 protein is a transcriptional regulator, and it is plausible that the Pit-1 and Oct-2 proteins specify the fates of pituitary and lymphoid cells, respectively. The Oct-1 protein is also a transcriptional regulator, but this POU protein has been found in all cell types analyzed and is apparently involved in the transcription of generally expressed genes.

Further studies of this new family of proteins should establish whether the POU domain functions only in DNA sequence recognition or also in direct interactions with other transcriptional factors. Such studies should also reveal whether, like the unc-86 protein, the Pit and Oct proteins control cell fates during development.

Acknowledgments

Winship Herr, Richard A. Sturm

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA

Roger G. Clerc,[1] Lynn M. Corcoran,[3] David Baltimore,[2,3] Phillip A. Sharp[1,2]

[1]Center for Cancer Research and [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA; [3]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142 USA

Holly A. Ingraham, Michael G. Rosenfeld

Howard Hughes Medical Institute, Eukaryotic Regulatory Biology Program, Center for Molecular Genetics, University of California, San Diego, School of Medicine, San Diego, California 92093 USA

Michael Finney,[4,5] Gary Ruvkun,[4] H. Robert Horvitz[5]

[4]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114 USA; [5]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

References

Baumgartner, S., D. Bopp, M. Burri, and M. Noll. 1987. Structure of two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during *Drosophila* embryogenesis. *Genes Dev.* **1**: 1247–1267.

Chalfie M., H.R. Horvitz, and J.E. Sulston. 1981. Mutations that lead to reiterations in the cell lineages of *C. elegans. Cell* **24**: 59–69.

Clerc, R.G., L.M. Corcoran, H.H. LeBowitz, D. Baltimore, and P.A. Sharp. 1988. The B-cell-specific Oct-2 protein contains POU-box and homeo-box-type domains. *Genes Dev.* **2**: 1570–1581.

Desai, C., G. Garriga, S.L. McIntire, and H.R. Horvitz. 1988. A genetic pathway for the development of the *Caenorhabditis elegans* HSN motor neurons. *Nature* (in press).

Desplan, C., J. Theis, and P.H. O'Farrell. 1988. The sequence specificity of homeodomain-DNA interaction. *Cell* **54**: 1081–1090.

Finney, M., G. Ruvkun, and H.R. Horvitz. 1988. The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein containing a homeo domain and extended sequence similarity to mammalian transcription factors. *Cell* (in press).

Hoey, T., R. Warrior, J. Manak, and M. Levine. 1988. DNA binding activities of the *Drosophila* even-skipped protein are *melanogaster* mediated by its homeo domain and influenced by protein context. *Mol. Cell. Biol.* **8**: 4598–4607.

Ingraham, H.A., R. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C.R. Lin, D.M. Simmons, L. Swanson, and M.G. Rosenfeld. 1988. A tissue specific transcription factor containing a homeo domain specifies a pituitary phenotype. *Cell* **55**: 519–529.

Ko, H.-S., P. Fast, W. McBride, and L.M. Staudt. 1988. A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeobox domain. *Cell* **55**: 135–144.

Laughon, A. and M.P. Scott. 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA binding proteins. *Nature* **310**: 25–31.

LeBowitz, J.H., T. Kobayashi, L. Staudt, D. Baltimore, and P.A. Sharp. 1988. Octamer-binding proteins from B or HeLa cells stimulate transcription of the immunoglobulin heavy-chain promoter in vitro. *Genes Dev.* **2**: 1227–1237.

McGinnis, W., M.S. Levine, E. Hafen, A. Kuroiwa, and W.J. Gehring. 1984. A conserved DNA sequence in homeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature* **308**: 428–433.

Nelson, C., V.R. Albert, H.P. Elsholtz, L.I.-W. Lu, and M.G. Rosenfeld. 1988. Activation of cell-specific expression of rat growth hormone and prolactin genes by a common transcription factor. *Science* **239**: 1400–1405.

O'Neill, E.A., C. Fletcher, C.R. Burrow, N. Heintz, R.G. Roeder, and T.J. Kelly. 1988. Transcription factor OTF-1 is functionally identical to the DNA replication factor NFIII. *Science* **241**: 1210–1213.

Scheidereit, C., A. Heguy, and R.G. Roeder. 1987. Identification and purification of a human lymphoid-specific octamer binding protein (OTF-2) that activates transcription of an immunoglobulin promoter in vitro. *Cell* **51**: 783–793.

Scott, M.P. and A.J. Weiner. 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proc. Natl. Acad. Sci.* **81**: 4115–4119.

Scott, M., J.W. Tamkun, and G.W. Hartzell. 1989. The structure and function of the homeo domain. *BBA Rev. Cancer* (in press).

Sturm, R.A., G. Das, and W. Herr. 1988. The ubiquitous octamer binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev.* **2**: 1582–1599.

Sturm, R.A. and W. Herr. 1988. The POU domain is a bipartite DNA binding structure. *Nature* (in press).

CSH
PRESS

Genes
& Development

The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products.

W Herr, R A Sturm, R G Clerc, et al.

References	This article cites 15 articles, 8 of which can be accessed free at: **http://genesdev.cshlp.org/content/2/12a/1513.full.html#ref-list-1**
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**