A novel POU homeodomain gene specifically expressed in cells of the developing mammalian nervous system

Robert G. Collum, Peter E. Fisher, Milton Datta, Scott Mellis, Carol Thiele, Kay Huebner, Carlo M. Croce, Mark A. Israel, Thomas Theil, Tarik Moroy, Ronald DePinho and Frederick W. Alt

The Howard Hughes Medical Institute and the Departments of Biochemistry and Microbiology, College of Physicians and Surgeons of Columbia University, New York, NY, Molecular Genetics Section, Pediatric Branch, National Institutes of Health, Bethesda, MD, Jefferson Cancer Center, Thomas Jefferson University, Philadelphia, PA, Department of Neurological Surgery and Pediatrics, Brain Tumor Research Center, University of California at San Francisco, San Francisco, CA, USA, Institut für Molekularbiologie und Tumorforschung, Philipps Universität, Marburg, Germany and Department of Microbiology, Albert Einstein University, Bronx, NY, USA

Received February 10, 1992; Revised and Accepted July 17, 1992

ABSTRACT

We report the isolation of a novel human POU domain encoding gene named RDC-1. The POU domain of the RDC-1 encoded protein is highly related to the POU domain potentially encoded by the rat brain-3 sequence and to that of the Drosophila I-POU protein; outside of the POU region, RDC-1 is unrelated to any previously characterized protein. The RDC-1 gene is expressed almost exclusively in normal tissues and transformed cells of neural origin. In the developing mouse and human fetus, RDC-1 is expressed in a spatially and temporally restricted pattern that suggests a critical role in the differentiation of neuronal tissues. In addition, RDC-1 is expressed in a unique subset of tumors of the peripheral nervous system including neuroepitheliomas and Ewing's sarcomas but not neuroblastomas. Based on its unique structural characteristics and expression pattern, we discuss potential functions for the RDC-1 protein.

INTRODUCTION

Efforts to elucidate the complex set of interactions governing mammalian development have been aided by studies of development in lower eukaryotes. Genetic analysis of a variety of developmental mutants of Drosophila has resulted in the identification of a family of genes, all containing a homeodomain, that are critical for pattern formation and differentiation in the developing fly (for review see 1). Homeodomains are conserved throughout eukaryotic evolution and the identification of mammalian genes encoding homeodomain proteins has led to speculation that these genes play a similar role in pattern formation and cell lineage determination in mammalian development (2).

MATERIALS AND METHODS

Isolation of genomic and cDNA clones

RDC-1 was isolated from human placental DNA on the basis of hybridization to DNA probes derived from exons 2 and 3 of human N-myc (9). EcoRI-digested genomic DNA from human placenta was used to construct a library in the lambda phage Charon 35A. Hybridizations, were carried out under reduced stringency conditions in 37% formamide-5×SSCPE (0.6 M

* To whom correspondence should be addressed at: Howard Hughes Medical Institute at Children's Hospital, Enders 861, 300 Longwood Avenue, Boston, MA 02115, USA
Figure 1. Identification of the RDC-1 genomic locus. A) Restriction map of 12-kb EcoRI fragment (named pR36.4) isolated from a human genomic library. R = EcoRI, X = XbaI, P = PstI and B = BamHI. B) Southern blot analysis of EcoRI digested genomic DNA (ten micrograms per lane) from human (H), mouse (M), canary (C), Xenopus (X), and Drosophila (D) hybridized with a nick-translated 1.1-kb SmaI fragment (see Materials and Methods). C) Northern blot analysis of RDC-1 transcripts in ten micrograms total RNA derived from the following human tumor samples: colon (lanes 1 and 2), stomach (lanes 3 and 4), esophagus (lane 5), kidney (lanes 6 and 7), myeloma (lane 8), T cell lymphoma (lane 9), neuroblastoma (lane 10), neuroepithelioma (lane 11), and HeLa (lane 12). Two transcripts of approximate size 4.0- and 2.0-kb are detected in neuroepithelioma RNA following hybridization with a nick-translated 0.5-kb EcoRI fragment (see Materials and Methods).

DNA sequence analysis of RDC-1

The plasmids pNE7 and pNE10 were subjected to DNA sequence determination. However, because these cDNA clone lacked approximately 600 bp at the 5' end of the RDC-1 transcript two adjacent Pst fragments of 0.9- and 0.85-kb size were derived from the RDC-1 genomic locus (see Figure 1A) were also sequenced. Restriction endonuclease mapping, primer extension analysis, and S1 nuclease protection experiments demonstrated that the RNA

NaCl, 0.075 M sodium citrate, 0.05 M potassium phosphate, 0.05 M EDTA), 10% dextran sulfate, 1×Denhart's solution, and 100 micrograms salmon sperm DNA per ml at 42°C. The 12-kb RDC-1 locus was subcloned into the EcoRI site of pVcos7, an expression vector containing a Moloney murine leukemia virus long terminal repeat, and transfected into primary rat embryo fibroblasts according to published procedures (10). Double-stranded cDNA was constructed from poly(A) RNA derived from the transfected cells as described (11). The cDNA was made blunt-ended by treatment with the Klenow fragment of E.coli DNA polymerase, ligated to EcoRI linkers, and cloned into the EcoRI site of lambda gt10. This library was hybridized with a 0.55-kb Styl-Xhol fragment derived from the genomic RDC-1 locus to identify a partial (0.5-kb) RDC-1 cDNA clone, denoted 36A. A second cDNA library was constructed from poly(A) RNA derived from the CHP100 neuroepithelioma cell line using a Pharmacia cDNA cloning kit according to the manufacturer's instructions. Approximately 10⁶ recombinant clones were screened with the 0.5-kb EcoRI insert derived from the previously identified 36A cDNA clone. Multiple overlapping cDNA clones were identified were subcloned into Bluescripts SK− and two overlapping clones pNE7 and pNE10, were selected for further analysis.

Figure 2. Comparison of POU domains from RDC-1 to other characterized POU domains. The POU domain is divided into POU-specific boxes A and B separated by a spacer sequence of variable length from a POU-homeodomain. Conserved residues are noted below.
transcript start site and 5' end of the transcript were contained on these two Pst fragments (data not shown). Nested deletions were generated using the ExoIII-Mung Bean nuclease system (Stratagene) and direct subcloning into M13 vectors mp18 and mp19 was used to generate appropriately sized fragment for sequencing. M13 single-stranded DNA or denatured plasmid DNA was sequenced using a Sequenase kit (US Biochemical Corp.).

**RESULTS**

Identification of the RDC-1 genomic locus

DNA probes derived from exons 2 and 3 of N-myc were used to screen a human genomic library under reduced stringency conditions (9). Several recombinant clones were identified in this way. One such clone, denoted RDC-1, contained a 12-kb EcoRI fragment (Figure 1A) and is therefore distinct from previously characterized myc family genes. Additional analysis of the N-myc homologous regions of RDC-1 demonstrated that these regions were short in size but similar enough to N-myc to account for cross-hybridization (data not shown), however, this homology appeared fortuitous as no significant similarities were encoded at the protein level (data not shown).

Southern blot hybridization analyses carried out under normal stringency conditions demonstrated that a DNA probe derived from RDC-1 detects both a 12 kb EcoRI fragment and an additional higher molecular weight fragment in human DNA (Figure 1B, lane H; two distinct fragments were seen using a variety of restriction enzymes, data not shown). Furthermore, hybridizing fragments were also detected in mouse, canary, *xenopus*, and *Drosophila* DNA under normal stringency conditions, indicating a high degree of evolutionary conservation (Figure 1B, lanes M, C, X, D).

**RDC-1 encodes unique transcripts**

The evolutionary conservation of RDC-1 suggested that the gene might encode a factor important for a conserved cellular function.
Therefore, to assay for expression of RDC-1, total cellular RNA from 12 human tumor samples or cell lines (representing various cell types) was examined by Northern blot analysis using an RDC-1 DNA probe. In this analysis, hybridizing transcripts were detected only in RNA of the CHPIOO neuroepithelioma cell line; two transcripts of approximately 4.0- and 2.0-kb were observed (Figure 1C, lane 11). To further characterize the RDC-1 transcripts, a cDNA library constructed from poly A+ RNA from CHPIOO was screened and clones that hybridized to the RDC-1 probe were isolated.

**RDC-11 contains a POU domain**

We determined the nucleotide structure and deduced amino acid codon sequence of the cDNA corresponding to the 4.0 kb RDC-1 transcript. The RDC-1 transcript contains an open reading frame of 978 bp (326 amino acids). The open reading frame is followed by a very large (approximately 2.5-kb) 3' untranslated region; additional analyses indicated that the smaller form of the RDC-1 mRNA differs from the larger only with respect to the amount of 3'-untranslated region present (data not shown). The functional significance of the two forms, if any, is unclear. The open reading frame of the RDC-1 cDNA predicts a protein of approximately 35 kD. A computer-aided search of known protein sequences indicates that RDC-1 contains a POU domain. The POU domain in RDC-1 is found at the very C-terminus of the protein and consists of the two POU-specific boxes A and B separated by a short spacer sequence from a POU-homeodomain (Figure 2).

The POU domain of RDC-1 is highly related to all previously identified POU proteins, although it is most closely related to those of the *C. elegans* unc-86 gene (84% identical; 6), the brn-3 POU domain isolated from rat brain (100% identical; 17), and the recently described *Drosophila* 1-POU protein (83% identical; 18). The similarity between RDC-1 and brn-3 extends into the spacer region separating the B domain from the homeodomain whereas most of the differences between RDC-1 and 1-POU are found in this spacer region. Outside of the POU domain RDC-1 is completely dissimilar to unc-86 and 1-POU; however, because only the POU domain of brn-3 has been isolated, a complete analysis of the structural relationship between RDC-1 and brn-3 is not possible.

Outside of the POU domain RDC-1 contains an unusually high number of alanine and glycine residues and in this way resembles the POU transcription factor SCIP, isolated from the rat peripheral nervous system (19) and the *Drosophila* POU proteins Cfl-a and pdm-1 (20, 21). The alanine and glycine residues are scattered throughout the proteins and do not form any discrete domains conserved between these three proteins. Furthermore, RDC-1 contains a polyhistidine stretch (residues 15 to 23) similar to stretches found in pdm-1 and a Cfl-a related *Drosophila* clone (21).
RDC-1 is expressed in a tissue- and stage-specific manner during development

Because homeo-box containing proteins are known to play a critical role in pattern formation and development in Drosophila and are thought to play a similar role in mammals, we examined the expression of RDC-1 during human and murine development. To assay for tissue- and stage-specific expression of RDC-1, total RNA was isolated from a variety of human fetal tissues at 15 weeks and 24 weeks gestational age and analyzed by Northern blot hybridization to an RDC-1 probe. This probe identified the 4.0- and 2.0-kb RDC-1 transcripts in 15-week brain, spinal cord, and 3.5- and 2.5-kb in RNA from the brain and spinal cord of day 13.5 mouse embryos (Figure 3B: E13.5, Br and SQ. Expression of RDC-1 in these three neurogenic tissues contrasted sharply with transcripts were readily detected in 6 out of 8 neuroepithelioma samples assayed representing at least 17 different cell lineages (Figures 1C and 4A and data not shown). Two of the samples expressing RDC-1 are from tissues thought to arise from the neural crest: CHP100 and H510 (22, 23).

Several POU homeodomain proteins including the Drosophila proteins I-POU and Cfl-1 are also expressed predominantly in tissues of neural origin. These two proteins appear to be intimately involved in determining the type of neurotransmitter produced by neurons by regulating the expression of the dopacarboxylase gene (encoding an enzyme required by adrenergic neurons; 18). In this regard, we have assayed for RDC-1 expression in neuroepitheliomas and neuroblastomas, two histologically similar tumors of the human peripheral nervous system that respectively produce cholinergic and adrenergic neurotransmitters (23, 24). Total RNA from 8 different neuroepithelioma and 8 different neuroblastoma tumor cell lines was isolated and assayed for expression of RDC-1 by Northern blotting analysis. RDC-1 transcripts were readily detected in 6 of 8 neuroepithelioma samples (Figure 3D, lanes 2–8, 13), but were not detected in any of eight neuroblastomas samples (Figure 3D, lanes 1, 12, 14, 18–22), showing that RDC-1 generally is expressed in the cholinergic tumor lines but not in the adrenergic ones. Ewing’s sarcomas display a number of neural features including the production of cholinergic neurotransmitters (25) and are characterized by an t(11;22) chromosomal translocation (26). Because of these neural features and because an identical t(11;22) translocation is found in neuroepitheliomas (27), RNA from several Ewing’s sarcoma cell lines was tested for expression of RDC-1. Total RNA was prepared from 6 Ewing’s sarcoma samples and assayed by Northern blot hybridization for RDC-1 expression. RDC-1 transcripts were readily detectable in 4 of 6 Ewing’s sarcoma samples tested (Figure 3D, lanes 9–11, 15–17).

RDC-1 maps to human chromosome 13

Although RDC-1 is expressed in a limited set of human tumors, the gene is expressed in two different tumor types that contain a t(11;22)(q23–24;q11–12) chromosomal translocation. To determine if RDC-1 might be directly involved in this translocation, we mapped the chromosomal location of this gene. DNA from twenty five rodent-human hybrids and a mouse and human control were analyzed by Southern blotting with
hybridization to probes derived from the RDC-1 genomic and cDNA clones. Hybrid DNAs which were positive or negative for the RDC-1 homologous fragments were scored and indicated that RDC-1 is located on chromosome 13 (see Materials and Methods). To regionally localize the RDC-1 locus more precisely, two hybrids retaining partial chromosomes 13 (as defined by the presence or absence of 13-linked molecular markers) were screened for hybridization to RDC-1 probes. These data (summarized in Figure 5) demonstrate that the common chromosome region present in the positive hybrids is chromosome region 13q14-q22. A narrower localization to 13q22 has recently been defined using other hybrids carrying partially deleted chromosomes 13 (M. Hansen, W. Cavenee, and K. Huebner, unpublished observations). Therefore RDC-1 is located in chromosome region 13q14-q22 and is not directly involved in the t(11;22) chromosomal translocation.

DISCUSSION

We have isolated a novel human gene, RDC-1, that encodes a POU domain protein. Although the POU domain encoded by RDC-1 shows striking similarity to three well-characterized transcriptional activators—Pit-1, Oct-1, and Oct-2 (3, 4, 5)—its POU-encoded region is most closely related to those of the unc-86 gene, the recently described Drosophila I-POU protein, and the brn-3 nucleotide sequence. The unc-86 gene encodes a C.elegans protein of unknown function (6) while the brn-3 POU sequence was derived from the rat genome by PCR (17) and has not yet been identified as part of a distinct gene. The I-POU protein is a putative dominant, negative transcriptional inhibitor that blocks activation by Cfl-a, a positive POU-domain regulator of neuron-specific genes (18).

The high degree of similarity between the POU domains of RDC-1 and unc-86, despite the large evolutionary distance separating the species from which these genes originate, may reflect a highly conserved functional role played by these two genes. Unc-86 is intimately involved in cell lineage determination and differentiation of the nervous system of C.elegans. Homozygous unc-86 loss of function mutants have specific nervous system defects including the absence of certain types of neurons and the failure to differentiate other neuronal cell types. Because RDC-1 is expressed specifically in embryonic tissues that give rise to neural structures, it also may play such a role in cell lineage determination and differentiation in humans.

We have isolated and determined the nucleotide sequence of the RDC-1 hybridizing Drosophila gene and found that it was identical to I-POU (M. Data, R. Collum, and F. Alt, unpublished). However, the homology resided completely within the POU domain, and outside of this region RDC-1 and I-POU are completely dissimilar. The negative regulatory activity of I-POU is ascribed to the lack of two critical basic amino acid residues within the POU domain (18). These two residues are retained within the RDC-1 POU domain. Therefore, it seems very unlikely that RDC-1, despite the fact that I-POU appears to be the closest Drosophila relative, would have an I-POU like negative regulatory property.

N-terminal to the POU domain the RDC-1 protein is characterized by glycine—alanine rich regions and a polyhistidine stretch. In this way, RDC-1 resembles a number of POU proteins from other species that appear to play a role in nervous system development. These proteins include pdm-1, Cfl-a, and a Cfl-a related clone from Drosophila and SCIP from the rat (19, 20, 21). The potential function of these regions in POU domain proteins particularly in relation to neural development remains speculative.

Although some POU domain encoding genes are expressed ubiquitously (5), most POU domain genes have a more limited pattern of expression (3, 17). RDC-1 resembles the latter group of genes in that its expression is highly restricted during development and in tumor cells. In this regard, the RDC-1 and brn-3 sequences have highly related POU domains and are both expressed primarily in neural tissues. However, within neural tissues, the expression patterns of these two sequences are quite distinct. Levels of brn-3 RNA increase during embryonic development to reach their highest levels at embryonic day 18 (in the rat) with continuing expression into adulthood (17); while RDC-1 expression, on the other hand, peaks early during embryogenesis (day E13.5) and is undetectable by 14 days after birth (Figure 3B).

RDC-1 is expressed in the developing brain, spinal cord, and eye during a limited period of active growth and differentiation. Although the precise cell types that express RDC-1 within these tissues have not been established, the expression patterns of RDC-1 both in vivo and in tumor cell lines strongly implicate this gene in the development of certain cells of neural crest origin. In this regard, the common expression of RDC-1 in neuroepitheliomas and Ewing's sarcomas supports the notion that these two tumors share a common developmental lineage and/or function (25). Although RDC-1 is expressed specifically in cells and tissues with neuronal characteristics, it is not expressed in all neuronal or neuronal-like cells tested. Indeed, the expression of RDC-1 in peripheral neuroepitheliomas versus the morphologically similar neuroblastomas should provide an important marker for further distinguishing them. Furthermore, the expression of RDC-1 correlates with the type of neurotransmitter produced by these tumors: neuroepitheliomas are cholinergic and neuroblastomas are adrenergic (23). Notably, the Drosophila I-POU and Cfl-a POU domains proteins also are expressed predominantly in tissues of neural origin where they may be intimately involved in regulating the adrenergic neuropeptide system (18). Given the restricted expression of RDC-1 in cholinergic versus adrenergic neural tumors and with the strong homology to I-POU, it seems possible that RDC-1 also may play a role in the regulation of specific gene expression within a subset of neuronal lineages.

REFERENCES