The Rat Leukocyte Antigen MRC OX-44 Is a Member of a New Family of Cell Surface Proteins Which Appear To Be Involved in Growth Regulation

ALFONSO BELLACOSA, PEDRO A. LAZO,[†] SUSAN E. BEAR, AND PHILIP N. TSICHLIS*

Department of Medical Oncology, Fox Chase Cancer Center,
Philadelphia, Pennsylvania 19111

Received 29 October 1990/Accepted 26 February 1991

Moloney murine leukemia virus (MoMuLV)-induced rat T-cell lymphomas express discrete 1.8-, 2.2-, and 4-kb mRNA transcripts hybridizing under conditions of reduced stringency to a probe derived from a region upstream of the first exon of the *Tpl-1/Ets-1* gene. Screening a cDNA library from one rat T-cell lymphoma with this genomic probe yielded 15 cDNA clones which were derived from 10 different genes. One of these genes, defined by the cDNA clone pRcT7a, was expressed as a 1.8-kb mRNA transcript in spleen and thymus but not in other normal rat tissues. Expression of the gene defined by the pRcT7a cDNA clone in a series of MoMuLV-induced rat T-cell lymphomas showed a perfect correlation with the expression of the rat leukocyte antigen MRC OX-44. Because of this observation, the pRcT7a clone was sequenced and it was shown to identify a gene coding for a 219-amino-acid protein. The homology between pRcT7a and the *Tpl-1* probe used for its detection mapped within the 3' untranslated region of the pRcT7a cDNA clone. The pRcT7a protein, which exhibits four putative transmembrane regions and three putative glycosylation sites, contains a region which is nearly identical in sequence to a peptide derived from the rat leukocyte antigen MRC OX-44. This finding suggested that the pRcT7a cDNA clone defines the gene coding for OX-44. To confirm this finding, a pRcT7a construct in the retrovirus vector pZipNeo was introduced into the OX-44[−] T-cell lymphoma line 2788. Immunostaining with the MRC OX-44 monoclonal antibody followed by flow cytometry revealed that following gene transfer, the 2788 cells became OX-44⁺. Sequence comparisons revealed that pRcT7a/MRC OX-44 is a member of a family of genes which includes the melanoma-specific antigen ME491; the human leukocyte antigen CD37; the protein TAPA-1, which is expressed on the surface of human T cells and appears to be involved in growth regulation; the human gastrointestinal tumor antigen CO-029; and the *Schistosoma mansoni*-associated antigen Sm23.

The use of monoclonal antibodies against hematopoietic cell surface determinants was instrumental for building up our current understanding of the differentiation pathways of these cells (for a review, see reference 14). One of these antibodies, MRC OX-44, has been particularly useful because it defines a specific subpopulation of differentiating thymocytes in the rat (22, 23). This population has the phenotype CD4[−] CD8⁺ OX-44[−]. The presumed next stage in thymocyte differentiation is characterized by CD4⁺ CD8⁺ OX-44[−] cells. The single positive thymocytes appearing after this stage and the mature peripheral T cells are OX-44⁺. This antigen is also expressed on B cells, neutrophils, monocytes, and macrophages (22). Since the human CD37 antigen exhibits a similar tissue distribution and shows similarities in its molecular properties to OX-44, it has been proposed that these two antigens represent species homologs (6).

In this report, we present the cloning and molecular characterization of the rat OX-44 gene and show that (i) although related to the human CD37, the protein it expresses is not the rat homolog of this human antigen and (ii) it codes

for a protein which is a member of a family of membrane proteins most of which are expressed in the late stages of differentiation in a variety of cell types. Some of these proteins appear to be involved in growth regulation, suggesting a potential functional role for this gene in hematopoietic cells.

MATERIALS AND METHODS

Tissues and cell lines. Normal tissues were obtained from 4- to 6-week-old F344 rats. Cell lines were established from primary Moloney murine leukemia virus (MoMuLV)-induced T-cell lymphomas. The immunophenotype of these cell lines was determined by using flow cytometry as previously described (16).

Southern blot analysis of genomic DNA. Genomic DNA isolation and Southern blotting were carried out by using standard procedures (31, 32). Hybridization of the pRcT7a probe to human DNA was carried out under conditions of reduced stringency (40% formamide–6× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] at 37°C). The filters hybridized under these conditions were washed with 2× SSC–0.1% sodium dodecyl sulfate three times at room temperature and twice (30 min each time) at 65°C.

Construction and screening of partial genomic DNA libraries. Genomic DNA cloning was performed as previously

* Corresponding author.

[†] Present address: Unidad de Genética Molecular, Centro Nacional de Biología Celular y Retrovirus, Instituto de Salud "Carlos III", 28220 Majadahonda (Madrid), Spain.

FIG. 1. Expression of the *Tpl-1/Ets-1* probe A in MoMuLV-induced T-cell lymphoma lines. (A) Northern blot analysis of MoMuLV-induced rat T-cell lymphoma lines hybridized to probe A. Poly(A)⁺-selected RNA was electrophoresed in a 1% agarose–2.2 M formaldehyde gel and, following transfer to nylon membranes, was hybridized under conditions of reduced stringency to probe A, derived from a region upstream of the first exon of the *Tpl-1/Ets-1* locus (2). Arrows indicate the positions of the 28S and 18S ribosomal RNA. (B) Map of the rat *Tpl-1/Ets-1* locus depicting the origin of probe A. The arrows indicate the sites of integration and orientation of the integrated proviruses in four primary tumors and cell lines (2). The position and transcriptional orientation of the first *Ets-1* exon is shown. E, *EcoRI*; B, *BamHI*; H, *HindIII*; S, *SacI*. E⁺ is an artificial *EcoRI* site marking the end of a genomic *Tpl-1/Ets-1* lambda clone. The dotted line marks a polymorphic *SacI* site.

described by using standard procedures (31, 32). Briefly, *EcoRI*-digested genomic DNA from the human colon carcinoma cell line CCL221 was fractionated by agarose gel electrophoresis. The DNA fraction from 10 to 20 kb was ligated into the *EcoRI* site of the bacteriophage vector λ KT3, a modification of λ EMBL4N (29a, 30). Following in vitro packaging, the resulting mixture of recombinant phage particles was propagated in *Escherichia coli* LE392. The library was screened with the *BamHI*-*XbaI* fragment of pRcT7a (see Fig. 3), and five identical independent λ clones were isolated.

Isolation of polyadenylated cellular RNA and Northern blot analysis. Total RNA was isolated from normal rat organs and MoMuLV-induced T-cell lymphoma lines by using the guanidinium isothiocyanate–cesium chloride method (3). Polyadenylated RNA was selected by affinity chromatography in oligo(dT)–cellulose (13). Northern (RNA) blot analysis was carried out by using standard procedures (24). Briefly, 5 to 10 μ g of polyadenylated RNA was electrophoresed in 1% agarose–2.2 M formaldehyde gels. RNA was transferred and cross-linked to nylon membranes (Hybond N, Amersham International) by UV irradiation. Hybridization to the pRcT7a probe was carried out under high-stringency conditions. Hybridization to a probe derived from a region upstream of the first exon of the *Tpl-1/Ets-1* gene (probe A) was performed under reduced stringency (40% formamide–6 $\times$ SSC at 37°C). The filters hybridized under reduced stringency were washed in 2 $\times$ SSC–0.1% sodium dodecyl sulfate three times at room temperature and twice (30 min each time) at 65°C.

cDNA library. Construction and screening. Oligo(dT)-primed cDNA was synthesized from 5 μ g of polyadenylated

RNA, extracted by the T-cell lymphoma line 2772A (16). To obtain double-stranded DNA, we used the method of Gubler and Hoffman (10). Following the addition of *EcoRI* linkers, the double-stranded cDNA was ligated to the *EcoRI* arms of the bacteriophage λ ZapII (Stratagene). Following in vitro packaging, the recombinant phage particles were propagated in *E. coli* XL-1 Blue. The cDNA library was screened with a probe derived from a region upstream of exon I of the *Tpl-1/Ets-1* gene (probe A). Fifteen λ clones were obtained and directly subcloned in pBluescript SK(–) by in vivo excision (28).

DNA sequence analysis. Nested pRcT7a deletions were generated in both directions by using the exonuclease III/mung bean nuclease method (11). Eight plasmid clones with overlapping deletions starting at the 5' end of pRcT7a and nine plasmid clones with overlapping deletions starting at the 3' end were sequenced. The sequencing reactions were carried out on alkali-denatured double-stranded DNA (18) by using the dideoxy chain termination method (25) and employing Sequenase version 2.0 (U.S. Biochemicals) and [α -³⁵S]dATP (NEN Dupont). G+C-rich areas were resolved by sequencing single-stranded DNA and/or by employing dITP. The reactions were analyzed on 4 and 6% polyacrylamide–8.3 M urea sequencing gels. The sequence of two subclones of the human OX-44 genomic clone was determined similarly. In this case, deletions were not generated and only the 5' and 3' ends of two clones (approximately 300 bp each) were sequenced.

Computer comparisons and protein hydrophobicity analysis. Computer comparisons between the rat OX-44, human OX-44, and human CD37 nucleotide sequences were performed with the program Bestfit. The same program was

FIG. 2. Expression of pRcT7a in normal rat tissues and in MoMuLV-induced rat T-cell lymphoma lines. (A) Polyadenylated RNA extracted from normal rat tissues was electrophoresed in formaldehyde-agarose gels, transferred to nylon membranes, and hybridized to the pRcT7a probe. A 1.8-kb transcript was detected in spleen and thymus. No detectable expression was found in testis, brain, or kidney. (B) Polyadenylated RNA extracted from eight MoMuLV-induced T-cell lymphoma lines was electrophoresed in formaldehyde-agarose gels, transferred to nylon membranes, and hybridized to the pRcT7a probe. The expression of the 1.8-kb pRcT7a transcript correlates perfectly with the expression of the rat leukocyte antigen MRC OX-44 (16). Arrows indicate the positions of the 28S and 18S ribosomal RNA.

employed to carry out protein sequence comparisons between the rat OX-44 and related proteins. Multiple sequence alignment of rat OX-44 and related proteins was done by using the output files from Bestfit and editing them with the program Lineup. Hydrophobicity analysis of the rat OX-44 protein was done with the program Peplot. Bestfit, Lineup, and Peplot are part of the Genetics Computer Group-University of Wisconsin Sequence Package (9).

Gene transfer. Immunostaining and flow cytometry. A *Clal-XbaI* DNA fragment containing the entire pRcT7a open reading frame was inserted by blunt-end ligation into the *BamHI* site of the retrovirus vector pZipNeo (4), generating the construct pZip7a. DNA of the pZipNeo vector and the pZip7a retrovirus construct were transfected into the amphotropic retrovirus packaging cell line ψ am (8). The transfected ψ am cells were cocultivated for a week with the OX-44⁺ T-cell lymphoma line 2788, which was thus infected with the pZipNeo or the pZip7a retrovirus. The infected 2788 cells were selected for G418 resistance (active concentration, 400 μ g/ml). The parental and pZipNeo- and pZip7a-infected 2788 cells were immunostained with the MRC OX-44 monoclonal antibody and analyzed by flow cytometry. A total of 10^6 viable cells were washed in ice-cold phosphate-buffered saline with 3% fetal calf serum and incubated with MRC OX-44 for 30 min at 4°C. After the washing, the cells were stained with a fluorescein-conjugated F(ab')₂ fragment of rabbit anti-mouse immunoglobulin G antibody. To determine nonspecific binding, the second antibody alone was used as a negative control. After being washed, the cells were suspended in phosphate-buffered saline with 3% fetal calf serum and propidium iodide to stain nonviable cells. The stained cells were analyzed with a FACStar Plus instrument (Becton-Dickinson and Co., Paramus, N.J.).

Nucleotide sequence accession number. The nucleotide sequence reported in this study has been submitted to the GenBank data base under accession number M57276.

RESULTS

Cloning and expression of pRcT7a. Following the identification and cloning of the *Tpl-1* locus (2), a series of genomic DNA probes from the vicinity of the integrated provirus was analyzed to determine whether the probes detected mRNA transcripts in a variety of cell types. One such probe (probe A), derived from a region upstream of the first exon of the *Tpl-1/Ets-1* gene (Fig. 1B), detected discrete 1.8-, 2.2- and 4-kb mRNA transcripts in several MoMuLV-induced tumor cell lines when hybridized under conditions of reduced stringency (see Materials and Methods) (Fig. 1A). Screening 250,000 independent recombinants in a cDNA library from the tumor cell line 2772A with this probe detected 15 cDNA clones derived from 10 different genes (data not shown). This report describes the characterization of one of these genes, represented by the cDNA clone pRcT7a.

The expression of pRcT7a was examined in a variety of normal rat tissues. The results in Fig. 2A showed that it was expressed exclusively in the spleen and thymus. Analysis of eight MoMuLV-induced rat T-cell lymphoma lines revealed a perfect correlation between the expression of pRcT7a and the leukocyte antigen OX-44 (Fig. 2B).

pRcT7a is a cDNA clone of the OX-44 gene. The pRcT7a clone was trimmed by sequential deletions in both directions. Eight plasmids carrying overlapping deletions starting at the 5' end of the gene and nine plasmids with overlapping deletions starting at the 3' end of the gene were sequenced. This allowed complete sequencing of the pRcT7a clone in

1	<u>EcoRI</u> GAATTCGGCÁCGAGGTATTTTCTCTCTTCACTTCTCTTTTACACAAATAGACATAGÁCTTCTGGGTÁACCAGACTGTGCTCGCCACTTTAAAGATAÁ	100
101	TAAGCAAATTTACCTGAAÁGTACTGAGGÁCACTGCCTTCTAAAGGCACTACTATCÁGAGTGGGÁCATGAGCAGCCTGÁAATTTGCTGAAATATGTTCTGÁ	200
	End MetGlyMetSerSerLeuLysLeuLeuLysTyrValLeu	
201	TTTTCTTTAACTTTCTTTTGGGTCTGTGGCTGTTCATTTTGGGCTTTGGATCCACCTTCTGGTCCAGAATACCTÁTGAATACTCTTCCGTAATC	300
	PhoPhePheAsnPhoLeuPheTrpValCysGlyCysCysIleLeuGlyPheGlyIleHisLeuLeuValGlnAsnThrTyrGlyIleLeuPheArgAsnL	
	Ser Glu Glu	
301	TCCCCTTCTGACACTTGGCAATGTTCTGGTCATTGTGGGCTCCATTATCATGGTAGTTGCTTCTTGGGCTGCATGGGÁTCAATCAAGGAGAATAAGTG	400
	ouProPheLeuThrLeuGlyAsnValLeuValIleValGlySerIleIleMetValValAlaPheLeuGlyCysMetGlySerIleLysGluAsnLysCy	
401	CCTGCTTATGTGTTCTTTGTTCTGCTGCTGCTTATTCTCTGCTGAGGTGACCTTAGGÁCATCTGCTCTTTGTGTATGÁAAAAAGATÁCAACACGTTÁ	500
	sLeuLeuMetSerPhePheValLeuLeuLeuLeuIleLeuLeuAlaGluValThrLeuAlaIleLeuLeuPheValTyrGluLysLysIleAsnThrLeu	
501	GTGGCAGAGGGCTGAATGÁCAGCATCCAÁCATTATCACÁCTGACAACAÁCCAGGATGGCATGGGACTTCATCCAGTÁCACAÁCTGCAÁGTGTTGTGGTG	600
	ValAlaGluGlyLeuAsnAspSerIleGlnHisTyrHisSerAspAsnSerThrArgMetAlaTrpAspPheIleGlnSerGlnLeuGlnCysCysGlyV	
	*	
601	TAAATGGCTÁAAGTGATGÁTCAGTGGTÁCCACATCTTCTGCCCATCÁGGTGCAGATÁGTTCAGGGTÁGTTATAAAAAÁGGGCAGGCGÁGGTTTCACTÁ	700
	aIAsnGlySerSerAspTrpIleSerGlyProProSerSerCysProSerGlyAlaAspValGlnGlyCysTyrLysLysGlyGlnAlaTrpPheHisSe	
	*	
701	CAATTTCTGTATATTGGAÁTCGTTACCAÁCTGCGTATGTGTGATACAGGTGCTGGGÁGTGCTTTTGTCTGACACTGÁACTGCCAGÁTTGACAAAÁCC	800
	rAsnPhoLeuTyrIleGlyIleValThrIleCysValCysValIleGlnValLeuGlyMetSerPheAlaLeuThrLeuAsnCysGlnIleAspLysThr	
	XbaI	
801	AGCCAGGCTTTAGGGCTGTGÁTTTACACCTÁCTTCTGCTTAAAGTATTTATTCTCTCTAGAAAGTCAÁAGCACCCATTCATTAAGGÁGTAAACAATT	900
	SerGlnAlaLeuGlyLeuEnd	
901	ACGTGCCTCACTGGCATTÁTGACTTTCTCTTATTCTTCTTTGACTGGAÁCTCTGTCTTÁTACACATCCÁGGGAAGAGAÁCATTTGTCTÁGGACTTCCCÁ	1000
1001	CATCAAGCAGÁAGACAAACÁTTAACCAÁCTGATAGCAGTÁACCATATCCÁTTAAAGATGÁTGAAACATAÁCTGGGTAATTTTAAAGÁGACACTGTAG	1100
1101	CACTGTAATTTGATATCCAÁCTGCTAGCTGÁAGGTGTGATÁTATCTCAGAÁTACTGGCCTÁCCACATGTAÁTCATTGAGTTTAGCCCCAÁATCCAATAÁ	1200
1201	AAGAGGTCCAÁAGTTTAAATÁTTAGACTCTTÁACAATTGTÁAAATCAATCTCTACCTTCTÁAATCTTGCTGCTGCTCACTCTACAAGTCAÁATAGGAGACÁ	1300
1301	AGCTGAAGGÁAAATTTGGAGTCCCATTCAÁTTAGATAATTGACATGTCCAÁGTTCCCTTTTGAACACAGTCTTGGGAATAGGAATCATACÁGAGATCTTTT	1400
1401	ATCACTGGAÁAAACATTCTGÁTTTCTTTGTTAAATGTÁTTATTGATTTTGTGTACATCAÁCCACACATCCÁTCCATCCAÁTCCATCCATCCAT	1500
1501	CCATCCATAÁAGACTAATÁTTCTCATCÁTACCATCAGÁCAGTTTGCAGCTTTGAATTÁCATTTGCCAAÁGAACAÁCTCÁGAAGAATTAÁTATCTCCAGT	1600
	EcoRI	
1601	GTGTTAGCTGÁATAATCATATCCATGTTGÁTGTTGTTGTTÁTTCTCTATTÁAAATAAGGTTCTGTTAATGTÁCCATTTTAACTGTCTCGTÁGCCGAATTC	1699

FIG. 3. Nucleotide and deduced amino acid sequence of the pRcT7a clone. The proposed ATG codon at nucleotide position 162 follows an in-frame stop codon (marked "End"), and it is included within a favorable context for initiation of translation (Kozak's consensus is indicated by broken box) (15). The 29-amino-acid boxed sequence in the amino-terminal portion of the protein is nearly identical to the published sequence of a peptide derived from the purified MRC OX-44 protein. The two sequences differ toward their carboxy-terminal portion, where direct peptide sequencing detected a serine and two glutamic acid residues in positions where cDNA sequencing detected cysteine residues. However, the correct identification of these three amino acid residues in the direct sequencing of the MRC OX-44 peptide has been questioned (6) (see Results). The positions of three putative N-glycosylation sites are indicated by asterisks. The multiply repeated tetranucleotide unit (ATCC) in the 3' untranslated region is underlined.

both directions. Inspection of this sequence, presented in Fig. 3, revealed that the pRcT7a clone represents a gene coding for a 219-amino-acid hydrophobic protein. A 29-amino-acid segment in the amino terminal portion of the deduced protein was almost identical in sequence to the published sequence of a tryptic peptide derived from the purified rat MRC OX-44 protein (6). The only difference between the sequence deduced from the analysis of the pRcT7a clone and the peptide sequence was in three amino acids whose identity in the peptide sequence has been disputed. This is because these amino acids would compromise the presumed structure and membrane association of the OX-44 protein and because the equivalent positions in the human CD37 protein are occupied by dissimilar amino

acids (6). The sequence of the pRcT7a clone was also compared to the sequence of the genomic DNA probe which originally detected it. The comparison, which is not shown here, revealed that the common sequence between the two clones consisted of a multiply repeated tetranucleotide (ATCC) unit, present in the 3' untranslated region of the pRcT7a gene (Fig. 3).

To confirm the identity of the pRcT7a clone, a pRcT7a retrovirus construct (pZip7a) was generated by using the vector pZipNeo (4) (Fig. 4). Subsequently, the OX-44⁺ T-cell lymphoma line 2788 was infected with the pZip7a or the pZipNeo retrovirus as described in Materials and Methods. The parental and pZipNeo- and pZip7a-infected 2788 cells were immunostained with the monoclonal antibody

FIG. 4. Structure of the pZip7a retroviral construct. An 860-bp *ClaI*-*XbaI* fragment containing the entire pRcT7a coding frame (depicted in black) was inserted by blunt-end ligation into the *Bam*HI site of the retrovirus vector pZipNeo. The *ClaI* site is present in the polylinker region of pRcT7a; the wavy line immediately to the right of the *ClaI* site represents polylinker sequences. LTR, Long terminal repeat; 3'SS, splice acceptor site; Neo, G418 resistance gene; SVori and pBRori, simian virus 40 (SV40) and pBR322 origins of replication.

MRC OX-44 and analyzed by flow cytometry. The results (Fig. 5) confirmed that pRcT7a codes for a protein recognized by the monoclonal antibody MRC OX-44.

The OX-44 rat leukocyte antigen is not the rat homolog of

the human CD37. Because of similarities between the structures of the OX-44 and CD37 proteins as well as similarities of the sequence of an isolated OX-44 peptide to a portion of the deduced amino acid sequence of CD37, it has been

FIG. 5. Fluorescence-activated cell sorter analysis of the MRC OX-44 immunostained parental (panel 1), pZipNeo-infected (panel 2), and pZip7a-infected (panel 3) 2788 T-cell lymphoma lines. Panels 1', 2', and 3' show the flow cytometric profiles of the same cells stained with the second antibody only. The OX-44⁻ 2788 parental cell line becomes OX-44⁺ upon infection with the pZip7a retrovirus construct.

A

R-Ox44 261 CTCTGGTCCAGAACTACTTATGGAATACTCTTCGGTAATCTCCCGTTCCT 310
H-Ox44 1 CTGCTGATGTCACAACTCTTCGGAGTGTCTTCCATAACCTCCCGTTCCT 50
H-Cd37 169 ATGCTCATGCGAAGACCAGCTTCGTGTCTCTTGTGGCGTTGGGCTTCGT 218

R-Ox44 311 GACACT.....TGGCAATGTTCTGGTCAATTGGGGCTCCATTATCA 351
H-Ox44 51 CAGGCT.....GGGCAATGTTGTTGTCATGTTGGGGCTTATTATCA 91
H-Cd37 219 GCCTCTGCAGATCTGTGTCCTAAGTCTTGGCCATCTCAGGAATCTTACCA 268

R-Ox44 352 TGGTAGTTGGCTCTTGGGCTGCATGGGAATCAATCAAGGAGAAATAGTGC 401
H-Ox44 92 TGGTAGTTGGCTCTTGGGCTGCATGGGCTATCAGGAGAAATAGTGC 141
H-Cd37 269 TGGGCTGCGCCCTCTGGGTTGTGTGGGGCCCTCAGGAGCTTCGGCTGC 318

R-Ox44 402 CTGGCTATGTGCG 413
H-Ox44 142 CTGCTTATGTGCG 153
H-Cd37 319 CTCTTGGGCGTCG 330

B

R-Ox44 34 LLVQNTYGI¹LF²RNL³PF⁴LT⁵L...GNV⁶LV⁷IG⁸SI⁹IM¹⁰V¹¹AF¹²LC¹³MG¹⁴SI¹⁵KE¹⁶NK¹⁷CL¹⁸LS 84
H-Ox44 1 LLIH¹NN²GV³LF⁴PH⁵NL⁶SL⁷LT...GNV⁸LV⁹IG¹⁰SI¹¹IM¹²V¹³AF¹⁴LC¹⁵MG¹⁶SI¹⁷KE¹⁸NK¹⁹CL²⁰LS 51
H-Cd37 36 ILID¹TS²VS³VF⁴GL⁵AF⁶VP⁷IQ⁸WS⁹KV¹⁰LA¹¹IS¹²GI¹³PT¹⁴MG¹⁵IALL¹⁶CG¹⁷CA¹⁸KL¹⁹RC²⁰LL²¹GL 89

FIG. 6. Sequence comparison of portions of the rat OX-44, human OX-44, and human CD37 genes. (A) Nucleotide sequence comparison of a coding segment from a human OX-44 genomic clone with the corresponding portion of the rat OX-44 and human CD37. (B) Amino acid sequence comparison of the homologous regions of the rat OX-44 and the human OX-44 and CD37 proteins.

proposed that the two proteins represent the homologs of a single gene in two species (6). To test this hypothesis, the two major *Eco*RI bands that were detected under conditions of reduced stringency by Southern blot analysis of human genomic DNA and hybridization to the pRcT7a probe were cloned (data not shown). Sequence analysis of coding portions of the human OX-44 genomic clones allowed us to perform a direct sequence comparison between human OX-44 and CD37 genes. The comparison, shown in Fig. 6, revealed that the two genes are indeed different.

Structural features of OX-44. Comparison of OX-44 to other, related proteins. The deduced sequence of the MRC OX-44 membrane-associated protein suggests that it has four hydrophobic regions (Fig. 7A), possibly corresponding to transmembrane domains, and three putative glycosylation sites (Fig. 3). On the basis of these data, we propose that the association of the MRC OX-44 protein with the cellular membrane can be described best by the model shown in Fig. 7B.

OX-44 is a member of a family of proteins which are known to be expressed in a variety of cell types in a differentiation-specific manner (Table 1). These include the melanoma-specific antigen ME491 (12); the human leukocyte antigen CD37 (6, 7); the protein TAPA-1, which appears to be involved in growth regulation (21); the human tumor-associated antigen CO-029 (29); and the *Schistosoma mansoni*-associated antigen Sm23 (33). Figure 8 shows a comparison of the amino acid sequences of these proteins. The comparison revealed that some portions of these proteins are highly conserved among members of the family while other portions are variable. Examination of the conserved and variable domains with reference to the protein structure and membrane association predicted by the model in Fig. 7B

TABLE 1. Tissue distribution of the expression of the OX-44 family of proteins

Protein	Tissue distribution	References
OX-44	B cells, monocytes, macrophages, neutrophils, single (CD4 or CD8) positive thymocytes, peripheral T cells	22, 23
CD37	B cells, monocytes, macrophages, neutrophils, single (CD4 or CD8) positive thymocytes, peripheral T cells	6, 19, 26
TAPA-1	Hematolymphoid, neuroectodermal, and mesenchymal tumor cell lines; distribution in normal cells not reported	21
CO-029	Gastric, colon, rectal, and pancreatic carcinomas	29
ME491	Dysplastic nevi, radial growth phase primary melanomas, hematopoietic cells, tissue macrophages	1, 12
Sm23	Preadult and adult <i>S. mansoni</i>	33

revealed that the conserved regions are localized primarily within the transmembrane domains. This suggests that the conserved regions may be important for the structural integrity and the interactions of these proteins with other components of the cellular membrane, while the variable portions may determine the specific function of each individual protein.

DISCUSSION

The rat leukocyte antigen MRC OX-44 is a membrane protein expressed in B cells, monocytes, macrophages, and T cells in the late stages of differentiation (22). This marker has been particularly useful in studies of thymic cell differentiation because it has permitted the identification of a T-cell population at an intermediate stage of development (22, 23). The cloning and characterization of MRC OX-44 presented in this manuscript have shown that it is a member of a family of membrane proteins which are expressed in the late stages of differentiation in a variety of cell types (Table 1). OX-44 and the other members of this family contain a hydrophobic amino-terminal domain of 100 to 130 amino acids. In this domain, the predominant hydrophobic regions are separated by two short stretches of polar or charged amino acids, suggesting that these proteins may contain three consecutive transmembrane domains. According to this hypothesis, the polar and charged amino acids mark turns that lie outside the lipid bilayer. Additionally, the polar and charged amino acids may be involved in the formation of macromolecular complexes between these and other membrane proteins. The hydrophobic region is followed by a 60- to 70-amino-acid hydrophilic domain (131 amino acids long in CD37) that in three members of the family (OX-44, CD37 and ME491) contains three putative glycosylation sites, while in the other three members (TAPA-1, CO-029, and Sm23) it contains only one. In all the members of the family, the carboxy-terminal region contains an additional hydrophobic domain followed by a hydrophilic tail. On the basis of this data, we have presented a model according to which these proteins contain a carboxy-terminal cytoplasmic tail with an amino-terminal and carboxy-terminal domain embedded into the membrane and separated by an extracytoplasmic glycosylated region. It is interesting that the most conserved portions of these proteins lie within the putative

FIG. 7. Hydrophobicity analysis and proposed model for the structure and membrane association of the rat MRC OX-44 protein. The dotted line in panel A is based on the Goldman algorithm, and the full line is based on the Kyte-Doolittle algorithm. The numbers correspond to amino acid residues. (A) The hydrophobicity plot shows four hydrophobic regions, possibly corresponding to transmembrane domains. (B) Model for the structure and membrane association of the MRC OX-44 leukocyte antigen. The model predicts four transmembrane regions, three cytoplasmic domains, and two extracellular domains. The main extracellular domain contains three putative N-glycosylation sites. The dotted lines represent the attached carbohydrate moieties.

transmembrane regions, suggesting that these regions may be critical for the structural integrity of these proteins and their interactions with other cellular elements. The hydrophilic extracellular portion may be variable because of its involvement in binding to diverse ligands.

The expression of members of this family of proteins in the late stages of differentiation in a variety of cell types raises the question of whether they may be involved in the transmission of differentiation signals to the respective cells. Whether this is true should be a subject of further study. However, several pieces of evidence have suggested that at least some of the members of this family are involved in the regulation of cell proliferation. Thus, monoclonal antibodies against CD37 inhibit the activation of B cells induced by anti-CD20 antibodies and B-cell growth factor but enhance the mitogenic effect of anti-immunoglobulin antibodies on the same cells (17). Similarly, antibodies to TAPA-1 exert an antiproliferative effect on a variety of hematopoietic tumor cell lines in culture. TAPA-1 appears to be noncovalently linked to other membrane proteins, thereby forming macromolecular complexes (21). Monoclonal antibodies against one of these proteins (Leu13) inhibit the mitogenic effect of anti-CD3 antibodies on T cells (5). Finally, ME491 is expressed at high levels in dysplastic nevi and radial growth phase primary melanomas which have limited growth potential. The same protein is expressed at very low levels or not at all in advanced, rapidly growing melanoma cells, suggest-

ing that its expression may have an antiproliferative effect (1). These data suggest that OX-44, like other members of this family of proteins, may be involved in the control of the proliferation of differentiated functional hematopoietic cells.

The inclusion of a *Schistosoma* protein (Sm23) in this family is particularly interesting. Schistosomes have a complex life cycle which includes stages that evolve only within specific mammalian tissues (20). The logical extension of our current knowledge on the function of these proteins is that Sm23 may be involved in interactions with natural ligands present in these tissues. Such interactions may be critical for the growth and development of the parasite.

It is interesting that we cloned OX-44 by using a probe that detected a repeated sequence present in the 3' untranslated region of the mRNA. This repeat is present in low copy numbers in genomic DNA (data not shown), and it is apparently a structural feature of the mRNAs transcribed from a limited number of genes, one of which is OX-44. This raises the possibility that a gene containing this repeat in its 3' untranslated region may be present upstream from *Ets-1*. The potential functional role of this repeat is currently unknown. Since it is known, however, that the structure of the 3' untranslated region may affect the stability of a given message (27), we suggest that this repeat may play a role in affecting the stability of the OX-44 message. Because of this, it may contribute to the tissue-specific expression of this gene.

	CY1	TM1	EX1	TM2				
	1				75			
R-Ox44	MGMSLLKL	IKYVLFFNF	LFWVCGCCIL	GFGIHLVQN	T..YGILFRN	LPFLTL....	...GNVLVIV	GSITM
H-CD37	MSAQESCLSL	IKYFLFVFN	FFFVLGSLIF	CFGIWILIDK	TSFVSFVGL.	...AFVPLQI	W...SKVLAIS	GIFTM
Co-029	MAGVSAC	IKYSMTFNF	LFWLGCILIL	ALAIWVRVSN	DSQAIFGSED	VGSSS....	YVAVDILIAV	GAITM
Tapal	MGVEGCTKC	IKYLLFVNF	VFWLAGGVIL	GVALWLRH..	DPQTNLLYLE	LGDKPAPNTF	YVGIIYILIAV	CAVMM
Me491	MAVEGGMKC	VKFLLYVLL	AFCACAVGLI	AVGVAQLVL	.SQTIIQGAT	PGSL.....	...PVVIIAV	GVFLF
Sm23	MATLGTGMRC	IKSCVFVNL	ICLLCSLVLI	GAGAYVEVKF	.SQY...GDN	LHKVWQAA..	...PIAIIIV	GVIIIL
Consensus	Mnaveg..kc	IKYllfvnf	lfwlcg..il	g.glw..v.n	.sq.i.lgln	lgsl.....	...p.vliav	G.iim
	CY2	TM3						
	76							150
R-Ox44	VVAFILGCGS	IKENKLLMS	FFVLILLILL	AEVTLAILLF	VYEKKINTLV	AEGLNDSIQH	Y....HSDNS	TRMAW
H-CD37	GIALILGCGA	LKELRCLLGL	YFGMLLLFA	TQITLGILIS	TQRAQLERSL	RDVVEKTIQK	YGTNPE.ETA	AESWS
Co-029	ILGFLGCGA	IKESRMLLL	FFIGLILL	LQVATGILGA	VFKSKSDRIV	NETLYENTKL	LSATGESEKQ	FQEI
Tapal	FVGFLGCGA	IKESQCLLT	FFTCVLILFA	CEVAAGTNGF	VNKDQIAKDV	KQFYDQALQ.	.AVVDDANN	AKAVV
Me491	LVAFLGCGA	CKENYCLMIT	FAIFLSLIML	VEVAANTAGY	VFRDKVMSEF	NNNFQOMEN	Y...P.KNNH	TASIL
Sm23	IVSFLGCGA	IKENVMCLYM	YAFFLVLLI	AELAAALVAV	VYKDRIDSEI	DAL...MTGA	L.DKPTKEIT	EFMNL
Consensus	ivafllGCGa	ikEnrCllg.	ffiflll.11	aevaa..llgf	v.kdkid.ev	ne.l.q.q.q.	y...pe.en.	...a.
				EX2				225
R-Ox44	DFIQSLOCC	G.....VN	GSSDWISGPP	SS.....				...CP
H-CD37	DYVQFLRCC	GWHYPQDFQ	VLILRGNGSE	AHRVPCSCYN	LSATNDSTIL	DKVILPQLSR	LGHARSRS	ADICA
Co-029	IVFQEEFKCC	G.....LVN	GAADWGNFQ	HYELCAC..	L	DK.....	...QRPCQS	YN...
Tapal	KTFHETLDC	GS.....					...ST	LTALTTSVLK
Me491	DRMQADFCK	G.....	..AANYTDW..	...EKIPS..	M	SK.....	...NRVPS	CCINV
Sm23	..IQSSEFKCC	G.....AKG	PDDYRGNV..				...PASC	KE...
Consensus	d..q.q.kCC	G.....	..aad.gng..				...r...s	..n.c.
				TM4		CY3		
	226							294
R-Ox44	SGADV.....	...QGCKYK	KGOAWFHSNF	LYIGIVTICV	CVIQVLGMSF	ALTNCQIDK	TSQALGL	
H-CD37	VPAE.....	SHIYREGCAQ	GLQKWLHNNL	ISIVGICLGV	GLELGFMTL	SIFLCRNLDH	VYNRLARYR	
Co-029	G	KQVYKETCIS	FIKDFLAKNL	IIVIGISFGL	AVIEILGLVF	SMVLKQIGN	K	
Tapal	SGSNII...	SNLFKEDCHQ	KIDDLFSGKL	YLIGIAAIVV	AVIMIFEMIL	SNVLCCGIRN	SSVY	
Me491	TVGCGINFNE	KAIHKEGVE	KIGGWLRKNV	LVVAAAALGI	AFVEVLGIVF	ACCLVKSIRS	GVEVM	
Sm23	E	NLTYTEGVS	VFGAFLKRNL	VIVACVAFGV	CFFQLLSIVI	ACCLGROIKE	YENV	
Consensus		..iykegcV.	kl..wlhkn1	.i....a.gv	avie.lgmvf	.c.lccq1.n	..nvl..YR	

FIG. 8. Sequence alignment of the rat OX-44 protein, the human CD37, CO-029, TAPA-1, and ME491 proteins, and the *S. mansoni* Sm23 protein. The bottom line presents the consensus sequence. Only the amino acids which were identical in all the proteins are boxed. The positions of the presumed cytoplasmic (CY), extracellular (EX), and transmembrane (TM) domains are indicated. The extracellular domain 2 (EX2) of CD37 is longer than that of all the other members of the family.

ACKNOWLEDGMENTS

We thank A. Joshua Wand and Robert A. Beckman for reviewing the manuscript and Pat Bateman for secretarial assistance.

This work was supported by Public Health Service grants CA-38047 (P.N.T.), CA-06927, and RR-05539 from the National Institutes of Health and by an appropriation from the Commonwealth of Pennsylvania to the Fox Chase Cancer Center. A.B. is a Fellow of the Lawrence Greenwald Foundation for Leukemia and Lymphoma Research.

REFERENCES

- Atkinson, B., C. S. Ernst, B. F. D. Ghrist, A. H. Ross, W. H. Clark, M. Herlyn, D. Herlyn, G. Maul, Z. Stepwski, and H. Koprowski. 1985. Monoclonal antibody to a highly glycosylated protein reacts in fixed tissue with melanoma and other tumors. *Hybridoma* 4:243-255.
- Bear, S. E., A. Bellacosa, P. A. Lazo, N. A. Jenkins, N. G. Copeland, C. Hanson, G. Levan, and P. N. Tsichlis. 1989. Provirus insertion in *Tpl-1*, an *Ets-1*-related oncogene, is associated with tumor progression in Moloney murine leukemia virus-induced rat thymic lymphomas. *Proc. Natl. Acad. Sci. USA* 86:7495-7499.
- Berger, S. L. 1987. Preparation and characterization of RNA: overview. *Methods Enzymol.* 152:215-227.
- Cepko, C. L., B. E. Roberts, and R. C. Mulligan. 1984. Construction and applications of a highly transmissible murine retrovirus shuttle vector. *Cell* 37:1053-1062.
- Chen, Y. X., K. Welte, D. H. Gebhard, and R. L. Evans. 1984. Induction of T cell aggregation by antibody to a 16 kd human leucocyte surface antigen. *J. Immunol.* 133:2496-2501.
- Classon, B. J., A. F. Williams, A. C. Willis, B. Seed, and I. Stamenkovic. 1989. The primary structure of the human leucocyte antigen CD37, a species homolog of the rat MRC OX-44 antigen. *J. Exp. Med.* 169:1497-1502.
- Classon, B. J., A. F. Williams, A. C. Willis, B. Seed, and I. Stamenkovic. 1990. The primary structure of the human leucocyte antigen CD37, a species homolog of the rat MRC OX-44 antigen. *Correction. J. Exp. Med.* 172:1007.
- Cone, R. D., and R. C. Mulligan. 1984. High-efficiency gene transfer into mammalian cells: generation of helper-free recombinant retrovirus with broad mammalian host range. *Proc. Natl. Acad. Sci. USA* 81:6349-6353.
- Devereux, J., P. Haeblerli, and O. Smithies. 1984. A comprehensive set of sequence analysis program for the VAX. *Nucleic Acids Res.* 12:387-395.
- Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28:351-359.
- Hotta, H., A. H. Ross, K. Huebner, M. Isobe, S. Wendeborn, M. V. Chao, R. P. Ricciardi, Y. Tsujimoto, C. M. Croce, and H. Koprowski. 1988. Molecular cloning and characterization of an antigen associated with early stages of melanoma tumor progression. *Cancer Res.* 48:2955-2962.
- Jacobson, A. 1987. Purification and fractionation of poly(A)⁺ RNA. *Methods Enzymol.* 152:254-261.
- Knapp, W., B. Dörken, W. R. Gilks, E. P. Rieber, R. E. Schmidt, H. Stein, and A. E. G. K. von dem Borne (ed.). 1989. Leucocyte typing IV. Oxford University Press, Oxford.
- Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15:8125-8148.
- Lazo, P. A., A. J. P. Klein-Szanto, and P. N. Tsichlis. 1990. T-cell lymphoma lines derived from rat thymomas induced by Moloney murine leukemia virus: phenotypic diversity and its implications. *J. Virol.* 64:3948-3959.
- Ledbetter, J. A., G. Shu, and E. A. Clark. 1987. Monoclonal

- antibodies to a new gp40-45 (CD37) B-cell associated cluster group modulate B-cell proliferation, p. 339-340. In A. McMichael, P. C. L. Beverley, S. Cebbold, M. J. Crumpton, W. Gilks, F. W. Gotch, N. Hogg, M. Horton, N. Ling, I. C. M. MacLennan, D. Y. Mason, C. Milstein, D. Spiegelhalter, and H. Waldmann (ed.), *Leucocyte typing III*. Oxford University Press, Oxford.
18. Lim, H. M., and J. J. Pene. 1988. Optimal conditions for supercoil DNA sequencing with the *Escherichia coli* DNA polymerase I large fragment. *Gene Anal. Tech.* **5**:32-39.
 19. Ling, N. R., I. C. M. MacLennan, and D. Y. Mason. 1987. B-cell and plasma cell antigens: new and previously defined clusters, p. 302-335. In A. McMichael, P. C. L. Beverley, S. Cebbold, M. J. Crumpton, W. Gilks, F. W. Gotch, N. Hogg, M. Horton, N. Ling, I. C. M. MacLennan, D. Y. Mason, C. Milstein, D. Spiegelhalter, and H. Waldmann (ed.), *Leucocyte typing III*. Oxford University Press, Oxford.
 20. Mahmoud, A. A. F. 1990. Trematodes (Schistosomiasis) and other flukes, p. 2145-2151. In G. L. Mandell, R. G. Douglas, Jr., and J. E. Bennett (ed.), *Principles and practice of infectious diseases*. Churchill Livingstone, New York.
 21. Oren, R., S. Takahashi, C. Doss, R. Levy, and S. Levy. 1990. TAPA-1, the target of an antiproliferative antibody, defines a new family of transmembrane proteins. *Mol. Cell. Biol.* **10**:4007-4015.
 22. Paterson, D. J., J. R. Green, W. A. Jefferies, M. Puklavec, and A. F. Williams. 1987. The MRC OX-44 antigen marks a functionally relevant subset among rat thymocytes. *J. Exp. Med.* **165**:1-13.
 23. Paterson, D. J., and A. F. Williams. 1987. An intermediate cell in thymocyte differentiation that expresses CD8 but not CD4 antigen. *J. Exp. Med.* **166**:1603-1608.
 24. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 25. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
 26. Schwartz-Albiez, R., B. Dorken, W. Hofmann, and G. Moldenhauer. 1988. The B cell-associated CD37 antigen (gp 40-52): structure and subcellular expression of an extensively glycosylated glycoprotein. *J. Immunol.* **140**:905-914.
 27. Shaw, G., and R. Kamen. 1986. A conserved Au sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**:659-667.
 28. Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse. 1988. λ ZAP: a bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16**:7583-7600.
 29. Szala, S., Y. Kasai, Z. Stepiewski, U. Rodeck, H. Koprowski, and A. J. Linnenbach. 1990. Molecular cloning of cDNA for the human tumor-associated antigen CO-029 and identification of related transmembrane antigens. *Proc. Natl. Acad. Sci. USA* **87**:6833-6837.
 - 29a. Tartof, K. D. Unpublished data.
 30. Tartof, K. D., and C. A. Hobbs. 1988. New cloning vectors and techniques for easy and rapid restriction mapping. *Gene* **67**:169-182.
 31. Tschlis, P. N., P. G. Strauss, and L. F. Hu. 1983. A common region for proviral DNA integration in MoMuLV induced rat thymic lymphomas. *Nature (London)* **302**:445-449.
 32. Tschlis, P. N., P. G. Strauss, and M. A. Lohse. 1985. Concerted DNA rearrangements in Moloney murine leukemia virus-induced thymomas: a potential synergistic relationship in oncogenesis. *J. Virol.* **56**:258-267.
 33. Wright, M. D., K. J. Henkle, and G. F. Mitchell. 1990. An immunogenic M_r 23,000 integral membrane protein of *Schistosoma mansoni* worms that closely resembles a human tumor-associated antigen. *J. Immunol.* **144**:3195-3200.