

THE PRIMARY STRUCTURE OF THE HUMAN LEUKOCYTE  
ANTIGEN CD37, A SPECIES HOMOLOGUE  
OF THE RAT MRC OX-44 ANTIGEN

BY BRENDAN J. CLASSON,\* ALAN F. WILLIAMS,\* ANTONY C. WILLIS,†  
BRIAN SEED,§ AND IVAN STAMENKOVIC§

*From the \*Medical Research Council Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, United Kingdom; the †MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU; and the §Department of Molecular Biology and the Department of Pathology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114*

The MRC OX-44 leukocyte antigen is a glycoprotein of  $M_r$  40–50 × 10<sup>3</sup> that is of interest because it allows the identification of an intermediate cell type in rat thymopoiesis that has the CD4<sup>+</sup>CD8<sup>+</sup>OX-44<sup>-</sup> phenotype (1, 2). Most CD4<sup>+</sup>CD8<sup>+</sup> thymocytes are also OX-44<sup>-</sup>, whereas other leukocytes are virtually all OX-44<sup>+</sup>. In the mature lymphocyte population, B cells express considerably more antigen than T cells (1).

In other species, no equivalent to OX-44 was obvious, but the human CD37 antigen showed some similarities to OX-44 in tissue distribution and mobility in SDS-PAGE (3). The CD37 antigen was first defined as being abundantly expressed on B lymphocytes, but the antigen is also seen at lower levels on T cells, some thymocytes, neutrophils, monocytes, and macrophages (3, 4). The function of CD37 is unknown, although it has been shown that a CD37 mAb can inhibit the activation of B cells induced by CD20 antibodies plus B cell growth factor (5).

Structural studies were initiated on OX-44 by protein purification and sequencing and on CD37 by cDNA cloning via antigen expression. The results show that these antigens are species homologues and the complete sequence of human CD37 is reported.

### Materials and Methods

*Protein Purification, SDS-PAGE, and Western Blot Analysis.* The rat OX-44 antigen was purified from spleen cell membrane solubilized in deoxycholate by two rounds of mAb affinity chromatography. Purified OX-44 antigen was analyzed by SDS-PAGE followed by silver staining or Western blotting and all techniques were essentially as described (6).

*Protein Sequencing and Amino Acid Analysis.* Purified OX-44 protein was reduced and alkylated (7) and after acetone precipitation was reconstituted in 20% TFA for loading onto a polybrene-treated, precycled glass fibre disc in an ABI 470A gas/liquid/solid-phase protein sequencer (Applied Biosystems, Foster City, CA). Phenylthiohydantoin (PTH) amino acid derivatives

---

I. Stamenkovic is a Special Fellow of the Leukemia Society of America. B. J. Classon is supported by the Leukemia Research Fund, UK. Address correspondence to B. J. Classon, MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, OX1 3RE, United Kingdom.

were identified by HPLC. Amino acid analysis on performic acid oxidized OX-44 was performed (8).

*Preparation of cDNA Library and Recovery of cDNA Clones by Panning.* Poly(A)<sup>+</sup> RNA was prepared from cells isolated by leukopheresis from a patient with chronic lymphocytic leukemia and cDNA was prepared and used to construct an expression library (9).

Anti-CD37 mAbs were obtained from the International Leukocyte Typing Workshop (4). Panning and screening cycles performed by spheroplast fusion were carried out as described (9).

*DNA Sequencing, RNA and DNA Blot Hybridization.* Purified CD37.1 plasmid was digested with restriction enzymes and fragments were subcloned into M13 mp8 and M13mp19. Recombinant M13 clones were sequenced by the dideoxy method using the Sequenase reagent kit (U. S. Biochemicals, Cleveland, OH) with the Sequenase enzyme (U. S. Biochemicals) or T7 DNA polymerase (Pharmacia Fine Chemicals, Uppsala, Sweden). DNA sequences were analyzed using the computer programs of Staden (10). The entire cDNA insert was sequenced in both directions and each nucleotide was determined an average of six times.

DNA and RNA extractions, blotting and hybridization were carried out as described previously (9, 11).

### Results and Discussion

The OX-44 antigen was purified by two rounds of immunoaffinity chromatography and an aliquot of the eluted material was analyzed by SDS-PAGE. After the first affinity step a broad band of material of  $M_r$   $40 \times 10^3$  was seen with a number of minor contaminants (data not shown). On repassage through the affinity column, the heterogeneous material was again bound and eluted (Fig. 1 A). The OX-44 antigenic material was shown to coincide with the broad band by Western blotting (Fig. 1 B; note the gels in A and B were run with separate antigen preparations), in accord with previous results obtained with radiolabeled OX-44 (1). The material was further assessed by amino acid analysis (Fig. 1 C), which revealed unusually high levels of Leu (13%) and Phe (7%), consistent with the presence of one predominant molecular species. Two separate preparations of reduced and alkylated OX-44 were sequenced, and in each case a single NH<sub>2</sub>-terminal sequence was obtained (Fig. 2 C), with the exception of residue 1, which was not assigned due to the presence of multiple PTH-amino acid derivatives. In a third sequencing run on a separate antigen preparation, a major sequence identical to the previous two runs was obtained and PTH-Gly was unambiguously identified at cycle 1. The NH<sub>2</sub>-terminal sequence of OX-44 was notable for its high content of hydrophobic amino acid residues.

A CD37 clone was isolated using a COS cell expression system (9, 11) and the complete nucleotide sequence of the cDNA insert is shown in Fig. 2 A. The insert of 1124 bp contained a single open reading frame of 732 bp encoding a protein of 244 amino acids with a predicted molecular mass of 27,376 daltons (Fig. 2 B). The insert size of the CD37 cDNA is consistent with the size of the CD37 mRNA (1.2 kb) from various cell types of the B lymphoid and monocytic lineages (Fig. 1 D). Southern blot analysis with DNA from various B cell lines and placenta digested with Eco RI or Hind III showed a pattern consistent with that of a single copy gene (data not shown).

A comparison of the NH<sub>2</sub>-terminal sequences of CD37 and OX-44 clearly shows that the two proteins are homologous, with 14 identities and 4 conservative amino acid substitutions over the first 29 residues (Fig. 2 C). At first inspection, the CD37 NH<sub>2</sub>-terminal sequence (residues 1-26) closely resembled a conventional leader sequence (Fig. 2 B). However, the match with the OX-44 protein sequence which begins only three residues after Met-1 of CD37 (Fig. 2 C) suggests that CD37 lacks a cleaved

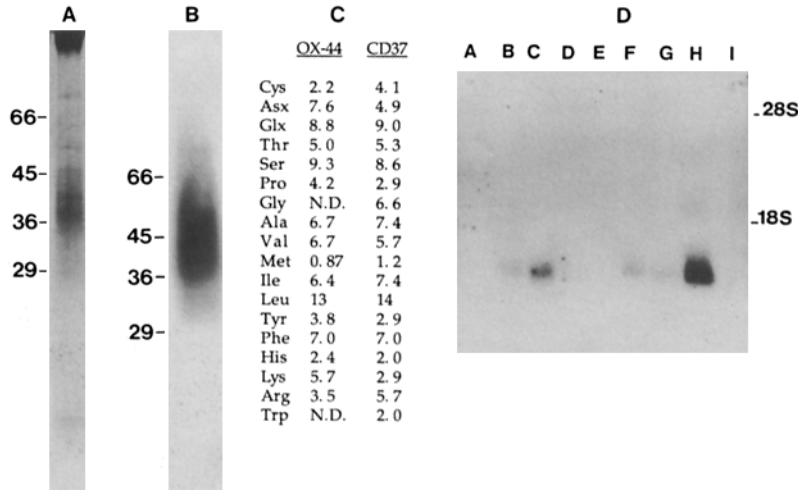


FIGURE 1. Protein and mRNA analysis. (A) 25 pmol (1  $\mu$ g) twice affinity-purified OX-44 electrophoresed on a 10% polyacrylamide gel in SDS under reducing conditions and silver stained. Molecular mass markers (kD) are indicated. (B) Western blot of twice affinity-purified OX-44. Purified OX-44 antigen was subjected to SDS-PAGE and transferred to nitrocellulose membrane. The OX-44 antigen was detected using the OX-44 mAb and <sup>125</sup>I-labeled, affinity-purified rabbit anti-mouse IgG antibody. (C) Amino acid analysis (residues per 100 amino acids) derived from duplicate analyses of 10  $\mu$ g OX-44 after periodic acid oxidation and hydrolysis for 24 h at 110°C with 5.7 N HCl and 0.01 M phenol. Loss of Ser and Thr were not corrected for and Trp was not determined. A value for Gly is not given because this amino acid was present at high levels. Since glycine was used to neutralize the OX-44 preparation after high pH elution this presumably represents glycine contamination remaining after dialysis. The amino acid composition estimated from the translated sequence of the human CD37 cDNA (Fig. 2 B) is shown for comparison. (D) RNA blot analysis of CD37. RNA isolation, electrophoresis and blots were performed as described (11). RNA sources were as follows: (A) HL-60; (B) U937; (C) tonsillar B cells; (D) HUT-102; (E) HPB-ALL; (F) RPMI 8226; (G) CESS; (H) Raji; (I) Nalm 6.

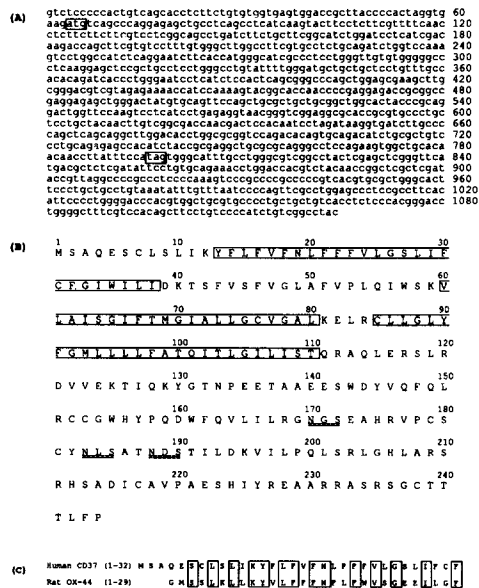


FIGURE 2. (A) The complete nucleotide sequence of the CD37 cDNA insert. The start codon at position 64 and the stop codon at position 798 are boxed. (B) The complete protein sequence of CD37. Postulated transmembrane sequences are boxed and N-linked glycosylation sites are underlined. (C) NH<sub>2</sub>-terminal sequence homology between CD37 and OX-44. Identical amino acid residues are boxed. The OX-44 sequence was derived from three independent sequencing runs. These sequence data have been submitted to the EMBL/GenBank Data Libraries under accession number X14046.

leader sequence and is therefore orientated in the membrane such that the NH<sub>2</sub> terminus is located inside the cell. Given that CD37 and OX-44 are species homologues, the latter would also be expected to contain an uncleaved NH<sub>2</sub>-terminal signal sequence and have a Met at position 1 but this is not the case (Fig. 2 C). It is possible that OX-44 has lost one or more NH<sub>2</sub>-terminal residues by proteolysis during purification of the antigen from spleen extracts.

A notable feature of the CD37 protein sequence is the high content of hydrophobic residues in the NH<sub>2</sub>-terminal half of the molecule (Fig. 3). It seems likely that this part of the protein will cross the membrane more than once and three putative transmembrane sequences have been assigned (Fig. 2 B). These are tentatively suggested since the fairly uniform hydrophobicity makes precise demarcation difficult. The positions of polar and charged amino acids may mark turns in the sequence that are outside the lipid bilayer but this is unclear since polar residues are sometimes found in putative transmembrane sequences of molecules that cross the membrane a number of times (12) or those that form multimolecular complexes in the bilayer (13). The significance of the two Glu residues assigned at positions 24 and 25 in the OX-44 sequence is unclear, given their absence from equivalent positions in the CD37 sequence. These residues may conflict with the transmembrane assignments shown in Fig. 2 B and the confirmation of these Glu residues at the cDNA level will be needed before any firm conclusions are drawn.

The COOH-terminal half of CD37 has a high content of polar and charged amino acids and is concluded to be extracellular since it contains all the *N*-linked glycosylation sites. Enzymatic deglycosylation studies have suggested that CD37 contains two or three such structures (3), which is consistent with the three possible *N*-linked sites shown in the CD37 sequence (Fig. 2 B). Both CD37 and OX-44 migrate as extremely

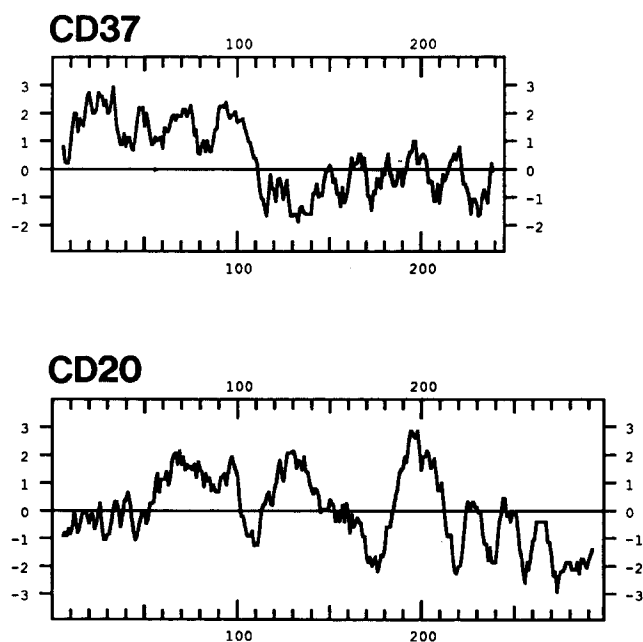


FIGURE 3. Hydropathicity profile for CD37 and CD20. The algorithm of Kyte and Doolittle was used with the DNA Strider computer software program (15).

heterogeneous bands in SDS-PAGE and this has been established to be due to N-linked carbohydrate for CD37 (reference 3). The carbohydrate moieties of CD37 may be unusually large since a maximum of three N-linked sites must account for  $\sim 20 \times 10^3$  of the average apparent  $M_r$  of  $45 \times 10^3$  on SDS-PAGE. In the case of Thy-1, three N-linked carbohydrate structures increase the apparent  $M_r$  of the protein by  $\sim 12 \times 10^3$ , as assessed by SDS-PAGE (8).

The structure of CD37 is unusual, and among lymphoid antigens, shows some similarities only to the CD20 molecule of human B cells (11, 14). CD20 also lacks a conventional NH<sub>2</sub>-terminal signal sequence, and as shown in Fig. 3, has three hydrophobic segments which are more clearly defined by interrupting polar and charged residues than those in CD37 (Fig. 3). However, it seems that the COOH-terminal region of CD20 may be intracellular (14) and if so, the similarity between CD37 and CD20 reduces to the likelihood that both molecules have multiple transmembrane sequences. There is no indication of an evolutionary relationship between CD37 and CD20 as assessed by sequence similarity.

In the rat, CD37 (OX-44) is particularly interesting as a marker that allows the identification of an early cell in thymopoiesis (1, 2). An immediate question is whether CD37 antibodies will be useful in identifying an equivalent cell in the human thymus.

### Summary

Comparison of NH<sub>2</sub>-terminal protein sequence from the rat OX-44 antigen with the sequence of the human CD37 antigen deduced from a cDNA clone shows that these antigens are species homologues. The CD37 sequence is 244 amino acids in length and lacks a conventional leader sequence. The molecule is likely to have an NH<sub>2</sub>-terminal cytoplasmic domain followed by three transmembrane sequences that lie within the first 110 amino acids. The rest of the molecule is hydrophilic and contains three sites for N-linked glycosylation.

We thank Dr. A. Neil Barclay for helpful discussions, Helen Wilmore for technical assistance, Mr. Tony Gascoyne for amino acid analysis, and Stan Buckingham for photography.

*Received for publication 14 November 1988 and in revised form 18 January 1989.*

### References

1. 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.
2. 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.
3. 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.
4. Ling, N. R., I. C. M. MacLennan, and D. Y. Mason. 1987. B-cell and plasma cell antigens: new and previously defined clusters. In *Leukocyte Typing III*. A. McMichael, et al., editors. Oxford University Press, Oxford. 302-335.
5. 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. In *Leukocyte Typing III*. A. McMichael, et al., editors. Oxford University Press, Oxford. 339-340.

6. Johnson, P., J. Gagnon, A. N. Barclay, and A. F. Williams. 1985. Purification, chain separation and sequence of the MRC OX-8 antigen, a marker of rat cytotoxic T lymphocytes. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:2539.
7. Classon, B. J., J. Tsagaratos, I. F. C. McKenzie, and I. D. Walker. 1986. Partial primary structure of the T4 antigens of mouse and sheep: Assignment of intrachain disulphide bonds. *Proc. Natl. Acad. Sci. USA.* 83:4499.
8. Campbell, D. G., J. Gagnon, K. B. M. Reid, and A. F. Williams. 1981. Rat brain Thy-1 glycoprotein. The amino acid sequence, disulphide bonds and an unusual hydrophobic region. *Biochem. J.* 195:15.
9. Seed, B., and A. Aruffo. 1987. Molecular cloning of the human CD2 antigen, the T cell erythrocyte receptor by rapid immunoselection procedure. *Proc. Natl. Acad. Sci. USA.* 84:3365.
10. Staden, R. 1986. The current status and portability of our sequence handling software. *Nucleic Acids Res.* 14:217.
11. Stamenkovic, I., and B. Seed. 1988. Analysis of two cDNA clones encoding the B lymphocyte antigen CD20 (B1, Bp35), a type III integral membrane protein. *J. Exp. Med.* 167:1975.
12. Liao, M. -J., and H. G. Khorana. 1984. Removal of the carboxyl-terminal peptide does not affect refolding or function of bacteriorhodopsin as a light-dependent proton pump. *J. Biol. Chem.* 259:4194.
13. Clevers, H., B. Alarcon, T. Wileman, and C. Terhorst. 1988. The T cell receptor/CD3 complex: A dynamic protein ensemble. *Annu. Rev. Immunol.* 6:629.
14. Einfield, D. A., J. P. Brown, M. A. Valentine, E. A. Clark, and J. A. Ledbetter. 1988. Molecular cloning of the human B cell CD20 receptor predicts a hydrophobic protein with multiple transmembrane domains. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:711.
15. Marck, C. 1988. "DNA Strider": a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* 16:1829.