The EMBO Journal vol.6 no.5 pp.1259–1264, 1987

Lymphocyte specific heterogeneity in the rat leucocyte common antigen (T200) is due to differences in polypeptide sequences near the NH_2-terminus

A.Neil Barclay, David I.Jackson, Anthony C.Willis[1] and Alan F.Williams

MRC Cellular Immunology Research Unit, Sir William Dunn School of Pathology, and [1]MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, UK

Communicated by A.F.Williams

The leucocyte-common antigen (L-CA, T200 or CD45) consists of a family of heavily glycosylated glycoproteins of apparent M_r 180 000–240 000 which are restricted to lymphoid and myeloid cells. Forms of L-CA which differ in their apparent M_r, antigenicity and glycosylation are expressed on different lymphocyte types. One specific antigenic determinant called MRC OX-22 is of particular interest because it distinguishes two sets of T helper cells that have different functions. From the sequence of different L-CA cDNA clones we now conclude that there is sequence heterogeneity such that at least four forms of L-CA exist with sequences in the range 1118–1250 amino acids. All the sequence variation occurs at a point starting 6 residues from the NH_2-terminus and the last 1112 residues of all forms are identical. Two of the variants can be directly related to the antigenic variation because they include sequence that was determined for a peptide that carries the MRC OX-22 determinant. Analysis of glycopeptides from thymocyte L-CA identified only one non-glycosylated position out of 14 possible N-glycosylation sites and established that all O-glycosylation was within the first 32 amino acids. The extra protein sequence in the longer forms was also suggestive of extensive O-glycosylation.

Key words: alternative splicing/glycoprotein sequence/glycosylation sites/leucocyte common antigen/lymphocyte glycoprotein

Introduction

The leucocyte-common antigen (L-CA, T200, Ly-5 or CD45) consists of a family of glycoproteins that are major components at the surface of all lymphoid and myeloid cells but not on other cell types (Trowbridge *et al.*, 1975; Fabre and Williams, 1977; Standring *et al.*, 1978; Trowbridge, 1978; Michaelson *et al.*, 1979; Dalchau *et al.*, 1980). The L-CA molecule spans the membrane to yield a cytoplasmic domain of 703 amino acids and an extracellular part with 402 or more amino acids (Thomas *et al.*, 1985; Saga *et al.*, 1986). The extracellular part is heavily glycosylated and is visualised in the electron microscope as a rod of length about 28 nm together with a globular domain of diameter 12 nm which contains the cytoplasmic sequence (Woollett *et al.*, 1985a). The function of L-CA is unknown but the cytoplasmic domain may interact with fodrin (Bourgignon *et al.*, 1985) and is known to be phosphorylated at Ser residues (Shackleford and Trowbridge, 1986).

L-CA shows heterogeneity in apparent M_r, antigenicity and glycosylation and different forms are unique to different lymphoid types (Standring *et al.*, 1978; Trowbridge, 1978; Woollett *et al.*, 1985b). On SDS PAGE thymocytes show one main L-CA band of 180 kd apparent M_r whilst T cells show four forms at 180, 190, 200 and 220 kd and B cells one broad band at around 240 kd (Woollett *et al.*, 1985b). These various forms all share at least two sets of antigenic determinants recognised by MAbs whilst other determinants are restricted to the higher M_r forms (Coffman and Weissman, 1981; Dalchau and Fabre, 1981; Spickett *et al.*, 1983). In the rat the specific determinants include one, recognised by the MRC OX-33 MAb, found only on B lymphocyte L-CA and others recognised by MRC OX-22 and OX-32 MAbs that are found on B cells, T cytotoxic cells and about 2/3 of T helper cells, but not on the rest of the helper cells or thymocytes (Woollett *et al.*, 1985b). These latter determinants are of particular interest because the subdivision of T helper cells correlates with important differences in cell function (Spickett *et al.*, 1983; Arthur and Mason, 1986). Also a tryptic peptide that retains the OX-22 antigenic determinant has been purified by MAb affinity chromatography and sequenced (Woollett, 1984; Johnson *et al.*, 1985a).

The basis of the L-CA heterogeneity presents a long-standing puzzle. Is this due to primary sequence differences or to differential glycosylation of the same peptide? It is known that there are differences in the carbohydrate of the various L-CA forms (Brown and Williams, 1982; Childs *et al.*, 1983; Lefrancois *et al.*, 1985) but this seemed unlikely to be the only factor. From the initial cDNA sequencing it was clear that much of the 3′ sequence for L-CA was identical between thymocyte and B cell forms indicating the presence of only one L-CA gene. However the size of mRNA varied in different cell types and this suggested the possibility of alternative forms at the 5′ end (Thomas *et al.*, 1985; Lefrancois *et al.*, 1986). This was also suggested from S1 nuclease mapping data in the mouse (Saga *et al.*, 1986).

We now resolve the heterogeneity question by showing that different cDNA clones of L-CA can have different sequences at the 5′ end. We also report that O-glycosylation is restricted to the area where the extra sequences are found in thymocyte L-CA.

Results

Sequences of cDNA clones established at least four different forms of rat L-CA

In previous studies rat L-CA cDNA was sequenced to yield a predicted sequence of 1073 amino acids that was incomplete at the NH_2-terminus (Thomas *et al.*, 1985). The sequence has now been extended to yield a predicted size for the smallest processed form of the molecule of 1118 amino acids (Figure 1). The argument that the NH_2-terminus has been identified is indirect because a typical leader sequence for rat L-CA was not found even though 9 independently-isolated cDNA clones extending into the NH_2-terminal region have been sequenced. The amino terminus is designated as a Gln residue in Figure 1 and the cDNA clones also predict the sequence Val Phe Val Thr Gly as part of the possible leader sequence. These residues correspond ex-

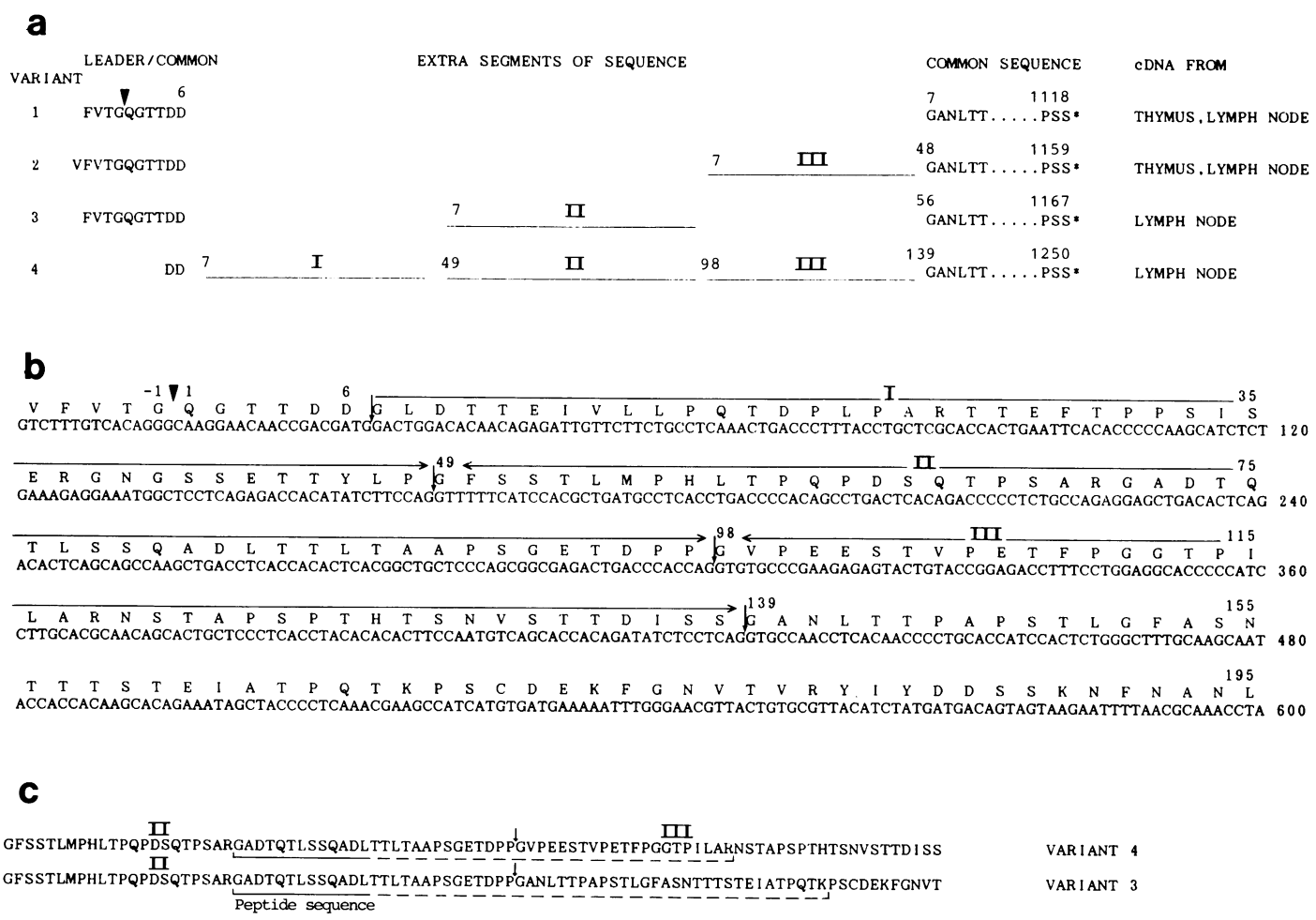

Fig. 1. **(a)** Schematic organisation of variants derived from cDNA clones. Four different types of variants were found containing extra segments I (42 amino acids), II (49 amino acids), III (41 amino acids). The cDNA clones were not full length and the extent of the 5′ end is indicated for each variant. The actual lengths of the clones analysed are given in Materials and methods. **(b)** Sequence at the 5′ end of the longest form of L-CA. The positions of the three extra segments I, II and III are indicated. The figure is a composite utilizing the most 5′ sequence obtained in variant 2 together with the remaining sequence in variant 4. **(c)** The MRC OX-22 peptide sequence in variants 3 and 4. The partial sequence (solid line) of the tryptic peptide purified using the MRC OX-22 antibody is present in sequences of variants 3 and 4. Different peptides would be derived from each variant by tryptic cleavage as shown by the dotted lines. Two peptides were isolated using the OX-22 MAb (see Materials and methods) but their sequences extended only as far as shown with the solid line. Thus it is unclear whether the peptides in both variants are antigenically active.

actly to the end of the predicted leader sequences from mouse and human cDNA clones (Saga *et al.*, 1986; Ralph *et al.*, 1987). The peptide data (see below) also supports the assignment of the amino terminus as shown in Figure 1. The sequence of 1118 amino acids gives an M_r of 127 133 which would yield a weight of about 169 000 for the processed polypeptide given that carbohydrate is 25% by weight of the whole molecule (Brown *et al.*, 1981). This figure corresponds as well as can be expected to various estimates from SDS gels for the apparent M_r of the thymocyte form of L-CA which have varied from 150 000 to 180 000 depending on the percent acrylamide of the gels (Standring *et al.*, 1978; Woollett *et al.*, 1985b). The mol. wt determined for L-CA from gel filtration and sucrose gradients was 172 000 (Fabre and Williams, 1977).

Amongst cDNA clones isolated from thymus and lymph node libraries three other sequences in addition to the minimal one have been identified as shown in Figure 1. In all cases the shortest L-CA sequence is interrupted 6 residues after the NH_2-terminus with the addition of extra sequence segments of 41, 49 or 132 amino acids. In the longest form both the 41 and 49 residue segments are present plus another 42 amino acid segment on the NH_2-terminal side of these. After the extra segments the sequence resumes to give what appears to be identical sequence in all forms.

In cDNA libraries from lymph nodes which contain mature T and B lymphocytes, all 4 forms of cDNA have been found, whereas in thymocyte libraries only the two smallest forms have thus far been detected. The smallest form presumably accounts for the major band of 180 000 M_r seen in the thymus while the second sequence with the extra 41 amino acids (variant 2) is probably the 190 000 apparent M_r form which is clearly seen in T cells but only as a minor band in thymocytes. The extra 41 amino acids would give an extra protein M_r of 4122 and carbohydrate could easily double the apparent M_r contributed by this segment in analysis by SDS PAGE.

The two largest forms (variants 3 and 4) are found only in lymph node libraries and these seem likely to code for some of the larger T or B cell L-CA forms. This view is supported by the fact that these two sequences include the sequence determined for a peptide that retained the antigenic activity detected by the MRC OX-22 MAb (Figure 1c and Materials and methods). The MRC OX-22 MAb gives virtually no labelling of thymocytes but

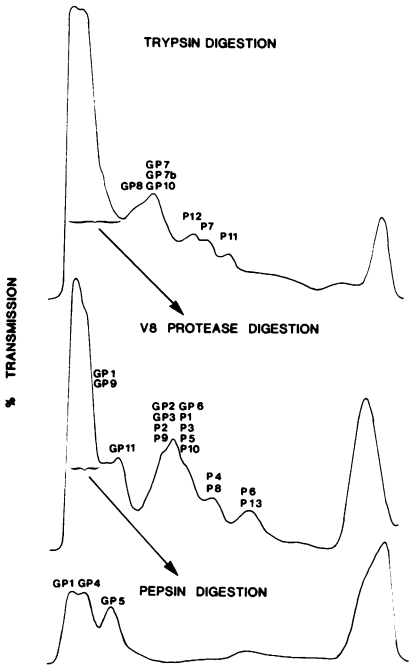

Fig. 2. Gel filtration of peptides from the large tryptic fragment of thymocyte L-CA. The positions of the peptides and glycopeptides analysed and sequenced (see Table I and Figure 3) are indicated above their approximate elution positions. In all cases except for the peaks obtained after pepsin digestion the peaks from gel filtration were pooled and separated further by h.p.l.c. prior to amino acid analysis and sequencing. In many cases the glycopeptides were recovered as multiple peaks after h.p.l.c. but they were usually derived from one peak on gel filtration. Column effluent was 0.1 M NH_4HCO_3 and the columns were monitored for absorbance at 206 nm.

reacts with B cells and most T cells. In immunochemical studies the antibody reacted only with the B cell form of L-CA and subfractions of the 190, 200 and 220 kd T cell forms. It could be that both variants 2 and 3 give rise to proteins that coincide in the 190 000 M_r L-CA band. In both variants the extra sequences are about the same size and the possibility of more than one form within the 190 kd band is consistent with the fact that the OX-22 MAb does not react with all of this band (Woollett *et al.*, 1985b).

The extra segments of sequence in variants 2–4 were unusual in containing about 33% Thr or Ser residues and about 14% proline (Figure 1b). The sequences are of the type that might be expected to be O-glycosylated (McMullen and Fujikawa, 1986).

Peptides isolated from the tryptic fragment of thymocyte L-CA

When L-CA is digested with trypsin in the absence of detergent a fragment of about 100 000 M_r is quickly produced (Woollett *et al.*, 1985a). This is fairly stable to further digestion except that the 100 000 M_r chain is cleaved to yield two fragments of 50 000 apparent M_r that are held together by disulphide bonding. The tryptic fragment seemed likely to retain the hydrophobic sequence since in electron microscopy 28 nm rods were seen and these were found to be associated at one end to form rosette-like structures as seen with the intact molecule (Woollett *et al.*, 1985a).

The thymocyte L-CA fragment was reduced and alkylated and then succinylated after which tryptic cleavage of the whole chain was undertaken. Peptides were fractionated by gel filtration (Figure 2) and then by h.p.l.c. (not shown). Large M_r material from the tryptic cleavage was further digested with V-8 protease under conditions for cleavage after Glu residues and residual large material was again digested with pepsin as shown in Figure 2. At all stages peptides were purified by h.p.l.c. except after the pepsin cleavage. The peptides that were isolated and characterised are shown in Figure 3 with compositions for glycopeptides

Table I. Amino acid and amino sugar composition of glycopeptides used for sequencing

Residues	GP1	GP2	GP3	GP4	GP5	GP6	GP7	GP7b	GP8	GP9	GP10	GP11
Cys C	.3,0	1.0,1	-,0	2.0,2	1.3,1	1.0,1	.3,0	.4,0	1.0,1	1.7,2	1.0,1	1.0,1
Asx D,N	4.0,4	.9,1	2.0,1	6.4,4	6.0,5	1.9,1	3.1,2	2.5,1	3.7,4	2.6,2	4.2,4	2.6,2
Thr T	7.5,9	1.5,2	1.1,1	4.8,2	3.9,3	1.0,1	3.1,3	1.6,1	1.7,2	3.1,3	.9,1	1.8,2
Ser S	3.1,3	.8,1	.6,0	3.7,3	1.0,0	1.0,1	.4,0	.6,0	.4,0	3.7,4	2.3,3	2.4,3
Glx E,Q	2.2,2	1.8,2	.3,0	1.7,1	1.6,1	1.2,1	.8,0	2.5,0	1.9,2	2.4,2	1.8,2	1.6,1
Pro P	2.5,2	1.5,2	-,0	2.4,1	3.0,3	-,0	-,0	-,0	-,0	2.0,2	1.5,2	-,0
Gly G	2.9,3	.1,0	1.2,1	2.3,1	1.7,1	.2,0	1.3,1	1.1,0	.6,0	1.7,1	2.2,2	2.2,2
Ala A	2.9,3	.9,1	.2,0	1.1,0	.5,0	1.2,1	1.2,1	.9,0	1.6,2	2.4,2	1.6,2	.5,0
Val V	.3,0	.1,0	1.9,2	1.7,1	1.2,1	.4,0	1.4,1	1.4,1	1.5,2	2.3,2	.9,1	3.0,4
Met M	-,0	-,0	-,0	.1,0	.1,0	-,0	-,0	-,0	-,0	1.0,1	-,0	.1,0
Ile I	.3,0	.7,1	.2,0	1.0,2	1.0,1	1.0,1	.3,0	.4,0	.5,1	1.0,1	1.5,2	.4,0
Leu L	2.2,2	-,0	.2,0	2.9,2	1.6,1	1.1,1	4.1,3	3.0,2	2.2,3	3.3,3	.2,0	1.2,1
Tyr Y	-,0	.1,0	.3,0	-,0	-,0	-,0	.4,0	.4,0	1.6,2	1.4,2	.9,1	.6,0
Phe F	1.0,1	-,0	1.0,1	.6,0	1.0,1	.3,0	1.0,0	.4,0	-,0	-,0	-,0	1.9,2
His H	.3,0	-,0	-,0	.2,0	1.0,1	-,0	1.8,2	1.0,1	-,0	2.7,3	-,0	.6,0
Lys K	.5,0	1.1,1	1.3,1	1.7,2	.3,0	.3,0	1.0,1	1.0,1	.9,1	2.1,2	.1,0	2.2,3
Arg R	-,0	-,0	.8,1	-,0	-.0	-,0	1.1,1	1.0,1	.9,1	-.0	.8,1	-,0
GlucNac	7.8	.1	1.9	10.5	5.2	.8	2.3	1.7	2.2	4.0	1.0	3.0
GalNac	5.5	.5		.2								
Total nmol	23	30	29	25	25	17	6	1.9	12.5	10.5	13.5	5.6

For each glycopeptide the observed compositions in residues per mole of peptide are given together with the compositions predicted from the cDNA sequence in Figure 3. The compositions were not corrected for degradation of Thr and Ser. GP4 and GP5 were analysed directedly from gel filtration; other peptides were analysed after h.p.l.c. In some cases the same glycopeptide was obtained in more than one peak and the total nmol peptide is the sum of these. It will be noted that the yield of peptides from the second half of the sequence was lower than for peptides from the NH_2-terminal part. This might be due to loss of the second part, because of the presence of the hydrophobic region, occurring during dialysis after the reduction, alkylation and succinylation steps which splits the disulphide bond between the 2 × 50 000 apparent M_r pieces.

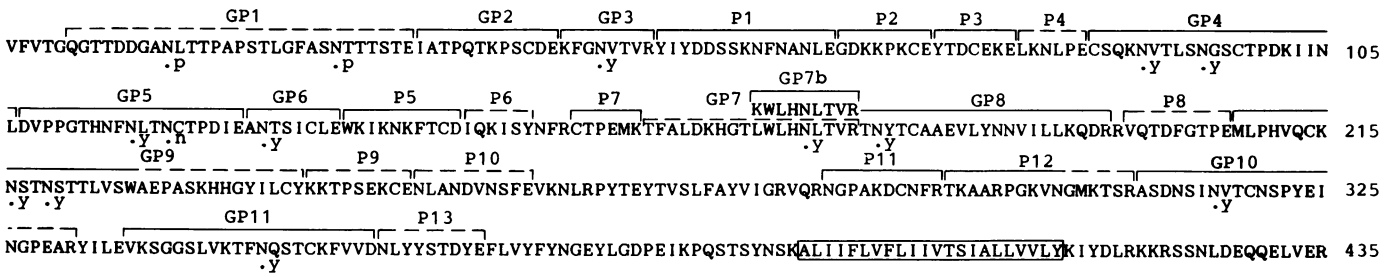

Fig. 3. Positions of peptides isolated from the tryptic fragment of thymocyte L-CA. The residues that were unequivocally determined by sequencing are indicated by a solid bar over the sequence deduced from the cDNA sequence. Dashed lines indicate that an isolated peptide fits well with the indicated sequence on the basis of amino acid composition. The elution positions of all the peptides are shown in Figure 2 and compositions of the glycopeptides are given in Table I. The possible N-linked glycosylation sites are indicated by a spot followed by a 'y' if it is known to be glycosylated from the sequence or 'p' if it is probably glycosylated because that glycopeptide is known to contain glucosamine and in one case 'n' as sequencing showed that this site was not glycosylated. The transmembrane sequence is boxed.

in Table I. The positions of peptides isolated in relation to the amino sequence predicted from the cDNA sequence are shown in Figure 3.

GP1 is designated as the NH_2-terminal peptide since none of the proteases used would be expected to cleave a Gly Gln linkage and the composition of GP1 in Table I argues strongly against the presence of any residues other than those indicated in Figure 3. No result was obtained when GP1 was sequenced and this would be expected if it were the NH_2-terminal peptide with blocking being due either to the fact that the Gln residue was cyclised to a pyroglutamic acid group or because the NH_2 group was succinylated. A second preparation of GP1 was isolated without succinylation but sequencing was still unsuccessful. This strongly implies the presence of pyroglutamic acid but attempts to remove this with pyroglutamyl amino peptidase (Podell and Abraham, 1978) also resulted in material that gave no sequence. It could be that glycosylation of GP1 (Table I) interfered with the enzyme.

With trypsin digestion cleavage would be expected only after Arg residues since the Lys groups were succinylated but two chymotryptic-like cleavages also occurred after Tyr to produce GP9 and P6. With V-8 protease, cleavage after Glu is expected but in two cases the sequence was cut after Asp residues to produce P5 and GP11. Such cleavages are not unexpected with the prolonged digestions that were used.

Because of succinylation, tryptic cleavage at Lys should not occur and P7 was the only peptide detected that ended with Lys. This was noted before (Thomas *et al.*, 1985) and it was suggested that the nicking of the 100 000 M_r chain to give the fragments of 2 × 50 000 apparent M_r gave the cleavage that resulted in P7. This interpretation is supported by current data because sequencing of GP7 (whose composition is shown in Table I) gave no sequence at all except for that of GP8 which was present as a minor contaminant along with GP7. If cleavage occurred between P7 and GP7 prior to succinylation then the NH_2-terminus of GP7 would be blocked by succinylation. GP7b was isolated as a low-yield peptide and this was sequenced to give the variant sequence shown above GP7 in Figure 3. This clearly identifies a polymorphism in the protein sequence and it would be expected that a peptide prior to GP7b should exist to provide the rest of the variant GP7 sequence and that this other peptide should also have been been blocked by succinylation. A candidate peptide was isolated at the level of 4 nmol and this gave a clean composition including Thr (1), Phe (1), Pro (1.1), Leu (.95), Asp (1.1), Lys (.9), Arg (1.1) Gly (.35). Sequencing of this peptide gave no results and GP7 and GP1 were the only other clean peptides for which sequencing was not successful. It is thus possible that the above peptide constitutes the rest of a polymorphic variant in the GP7 region.

N-linked and O-linked glycosylation sites

Amino acid analysis of peptides was done under protein hydrolysis conditions but reasonable recoveries of both glucosamine and galactosamine were still obtained. Table I shows analyses for all the glycopeptides with glucosamine being taken to indicate N-glycosylation and galactosamine to mark O-linked sugars. Only GP1 and GP2 had convincing levels of galactosamine with the presence of this amino sugar in GP4 being due to contamination from GP1 (the gel filtration peaks were pooled to ensure purity of GP1 rather than GP4, Figure 2). Sites for O-glycosylation in GP1 could not be determined as this peptide gave no sequence but as reported before (Thomas *et al.*, 1985) sequencing of GP2 indicated that Thr 3 from this peptide carried an O-linked sugar. It also seems likely that Thr 6 in the OX-22 peptide sequence is O-glycosylated because this Thr was not detected in the sequencer whilst Thr 4 and Ser 8 and 9 were detected in good yield.

All peptides with possible N-linked glycosylation sites were sequenced and glycosylation was proven in all but three cases by the complete absence of Asn at the expected residue in the sequencer run. In GP5 this occurred for the first site but not for the second which gave Asn in the same yield as adjacent residues. At this position the sequence is Asn Cys Thr Pro and the presence of the Pro seems likely to have prevented glycosylation. It is well known that the sequence Asn Pro Thr debars glycosylation (Bause and Hettkamp, 1979) and an Asn Thr Thr Pro sequence in IgG has also been reported as being unglycosylated (Frangione *et al.*, 1980). In GP1 there are two possible N-linked sites but glycosylation was not proven by sequencing. However the amount of glucosamine found in this peptide suggests that both sites are glycosylated. The glycosylation site in GP7 was proven by sequencing of GP7b and by the composition of GP7 (Table I).

Peptides from the larger variants of L-CA

Small numbers of peptides were isolated from L-CA that contained the OX-22 determinant to confirm the identity of the sequence at the protein as well as the cDNA level in regions after the NH_2-terminal segment. When L-CA is purified from spleen with the MRC OX-22 MAb all material obtained is of the higher M_r forms and most will be from B cells. This material was purified and degraded with trypsin to give the large tryptic fragment as described above for thymus L-CA. In the case of spleen L-CA the internal split occurs to give material of about 66 000 (NH_2-terminal) and 50 000 apparent M_r rather than the 2 ×

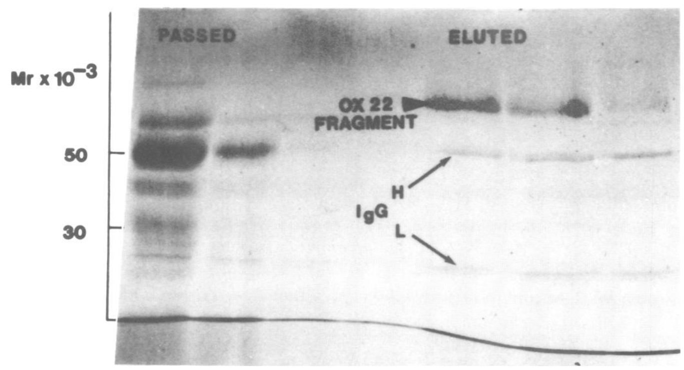

Fig. 4. Purification of the NH_2-terminal part of the large tryptic fragment from spleen L-CA using MRC OX-22 antibody. Purified OX-22$^+$ spleen L-CA was digested with trypsin to obtain the large tryptic fragment, then reduced and alkylated in SDS and subjected to an antibody affinity column after re-addition of deoxycholate. The fractions passing through the MRC OX-22 column and fractions eluted at pH 11.5 were analysed by SDS PAGE stained with Coomassie blue. Bands from contaminating IgG that was eluted from the column are indicated.

```
             LEADER ▼ COMMON  ↓←――――――――――――― I ――――――――――――――→↓
RAT     VFVTGQGTT--DDGLDTTEIVLLPQ-TDPLPARTTEFTPPSISERGNGSSETTYLP
HUMAN   VFVTGQSPTPSPTGLTTAKMPSVPLSSDPLPTHTTAFSPASTFERENDFSETTTSL

                     ↓←――――――――――――― II ―――――――――――――――――
---------------------GFSSTLMPHLTPQPDSQTPSARGADTQTLSSQADLTTLTAAP
SPDNTSTQVSPDSLDNASAFNTTGVSSVQTPHLPTHADSQTPSA-GTDTQTFSGSAANAKLNPTP

――――→↓←――――――――――――――― III ――――――――――――――→↓  COMMON
SGETDPPGVPEESTVPETFPGG-------TPILARNSTAPSPTHTSNVSTTDISSGANLTTPAPS
-GSNAISDVPGERSTASTFPTDPVSPLTTTLSLAHHSSAALPARTSNTTITANTSDAYLNASETT
```

Fig. 5. Alignment of the NH_2-terminal region of the longest forms of L-CA from rat and human. Identities are underlined and dashes indicate gaps introduced to maximise homology. The positions of the extra segments in rat L-CA are indicated using the nomenclature from Figure 1. Data for human L-CA are from Ralph *et al.* (1987).

50 000 fragment as for thymocyte L-CA. It was planned to isolate peptides close to the NH_2-terminal region and to do this the 66 000 piece was first purified by a second round of OX-22 MAb affinity chromatography. The OX-22$^+$ tryptic fragment was reduced and alkylated and dissociated into the 66 000 and 50 000 M_r pieces in SDS, then deoxycholate was added back to form mixed micelles with free SDS. The material was then passed through the OX-22 column to yield the high M_r band as shown in Figure 4. The use of this material unequivocally establishes that peptides isolated will be from a high M_r L-CA form. This material was digested with trypsin and peptides were isolated by gel filtration and h.p.l.c. Two peptides, whose composition (see Materials and methods) indicated derivation from sequence near the NH_2-terminus, gave the unambiguous sequences Phe Gly (Asn) Val Thr Val Arg and Tyr Ile Tyr Asp Asp Ser Ser Lys which correspond to sequences within peptides GP3 and P1 in Figure 4. These sequences confirm that the larger M_r form of L-CA has the same sequence as the smallest form in the NH_2-terminal region close to the position where the variant forms are found.

Discussion

The finding of extra sequence segments near the NH_2-terminus of L-CA provides a molecular basis for the heterogeneity in apparent M_r and antigenicity of the various forms of L-CA. Similar data have been obtained for human L-CA (Ralph *et al.*, 1987) and comparison of the human and rat sequences in Figure 5 shows considerable conservation of sequence in the region of variation.

In the rat there are at least 5 forms of L-CA seen on SDS gels and the heterogeneity must be greater than this since antibodies that are specific for the higher M_r forms bind only a subfraction of the bands that they react with (Woollett *et al.*, 1985b). cDNA variants I and II may respectively account for the major 180 000 M_r band and minor 190 000 M_r band of thymocytes but it seems unlikely that variants three and four could account for all the larger L-CA forms of T and B cells. In the case of human L-CA one of the sequences is longer than any of the rat sequences found so far (Figure 5) (Ralph *et al.*, 1987) and this could consititute another variant in both species. Further variation could also come from differences in the carbohydrates.

The most likely method for the generation of the various sequences of L-CA is by alternative splicing amongst exons of the primary RNA transcript. The complete exon structure of rat L-CA is not known but the segments I, II and III (Figure 1a) are known to be encoded within individual exons (D.I.Jackson, unpublished). A number of examples are now known whereby different forms of a protein are generated by alternative RNA splicing (e.g. Nawa *et al.*, 1984; Kornblitt *et al.*, 1985; Zamoyska *et al.*, 1985; Murray *et al.*, 1986). One example which may be similar to that of L-CA is in the regulatory muscle protein troponin T where selection from 5 small exons coding for 4–7 amino acids near the NH_2-terminus gives a minimum of 10 different mRNAs. The expression of these forms is controlled in development and in tissue expression (Breitbart *et al.*, 1985). The generation of the various forms of L-CA seems a particularly interesting system for further study because of the regulation of expression that occurs during differentiation of a closely related set of cells.

Peptide analysis showed that all but one of the potential N-linked glycosylation sites in the external part of thymocyte L-CA were likely to be glycosylated. The exception had the sequence Asn Cys Thr Pro in which the Pro seems likely to have prevented glycosylation. O-Linked sugars were present only in the NH_2-terminal region of thymocyte L-CA and are probably also abundant in the extra variant segments. Thus the external part of the L-CA contains two distinct regions. The membrane proximal region which consists of about 350 amino acids and is likely to exist in tightly folded domains stabilised by disulphide bonds with carbohydrate only of the N-linked type. Attached to this at the amino terminus are segments of varying size that are notable for high levels of Thr, Ser and Pro and the lack of Cys. This region is likely to have an extended structure and to be heavily O-glycosylated. It seems likely that the NH_2-terminal region expresses the carbohydrates that carry the determinants specific for soybean lectin which binds to B cell but not to thymocyte or to most T cell L-CA (Brown and Williams, 1982).

Protein sequencing identified a polymorphism in one peptide. Rat L-CA is known to express at least two polymorphic antigenic determinants (now called RT7-1 and RT7-2) (Carter and Sunderland, 1980; Newton *et al.*, 1986). The L-CA used for protein sequencing was from an open breeding colony of Sprague–Dawley rats and this is known to include both polymorphic determinants of L-CA (A.F.Williams, unpublished). Thus the protein polymorphism may account for the RT7 antigenic determinants.

Materials and methods

Isolation of L-CA cDNA clones from thymocyte and lymph node cDNA libraries

RNA was prepared from AO rat thymocytes or intact lymph nodes in guanidine isothiocyanate (Chirgwin *et al.*, 1979; Thomas *et al.*, 1985) and poly A$^+$ RNA was selected using oligo dT cellulose (Maniatis *et al.*, 1982). cDNA was syn-

thesised by the RNase H method of Gubler and Hoffman (1983) using a kit from Amersham International, UK. A seventeen base antisense oligonucleotide matching nucleotides 1114–1130 in pLC-1 (transmembrane region) (Thomas *et al.*, 1985) was phosphorylated with T4 polynucleotide kinase (Maniatis *et al.*, 1982) and added to the first strand cDNA synthesis at 5 ng/ml to derive cDNA clones with 5′ L-CA sequence. Lymph node cDNA of >1 kb was selected by sucrose gradient (10–30% gradient w/v) centrifugation. cDNA was blunt end ligated into pAT153/PVUII/8 or the derivative pATX and used to transform *Escherichia coli* MC1061 (Thomas *et al.*, 1985). cDNA libraries were amplified about 50-fold. Thymocyte cDNA libraries of complexity 50 000 and 250 000 and two lymph node libraries each of complexity 120 000 were screened using a cDNA probe from the 5′ end of pLC-1 to the *Eco*RI site (Thomas *et al.*, 1985) and a 17 long oligonucleotide (corresponding to residues 594–610 in Figure 1). The following cDNA clones coding variants 1–4 (Figure 1) were studied. Thymocyte cDNA clones: variant 1; pLC 29 (4.5 kb), pLC 40 (1.3 kb), variant 2; pLC 39 (1.4 kb), lymph node cDNA clones; variant 1; pLC 42 (1.2 kb), pLC 64 (1.2 kb), variant 2; pLC 70 (1.3 kb), variant 3; pLC 44 (2.5 kb), variant 4; pLC 41 (2.8 kb), pLC 65 (2.8 kb). The cDNA clones were fragmented with restriction enzymes or by sonication and subcloned into M13mp8. Clones from sequencing by the dideoxy method were picked randomly or screened on nitrocellulose with appropriate cDNA probes (Messing, 1983; Biggin *et al.*, 1983; Staden, 1986). The 5′ ends of each clone were sequenced in both directions and the rest of the clone checked by limited sequencing and restriction map analysis.

Preparation of peptides from thymocyte L-CA

The tryptic fragment of L-CA was purified from rat thymocytes as in Woollett *et al.* (1985b) except that for the final purification after the trypsin digestion a second MRC OX-1 affinity column step was used. The amount of L-CA was estimated by amino acid analysis and 55 nmol was reduced, alkylated and succinylated (Campbell *et al.*, 1981). The material was digested at 37°C in 0.1 M NH_4HCO_3 for 24 h with a total of 80 μg trypsin added in two aliquots. Phenylmethylsulphonyl fluoride was added to 2 mM and peptides were fractionated on a Biorad Biogel P-30 column (1 cm × 150 cm) in 0.1 M NH_4HCO_3 and all retarded fractions were subjected to reverse-phase h.p.l.c. as in Johnson *et al.* (1985b). The unretarded fractions were lyophilised, resuspended in 0.1 M NH_4HCO_3 and digested for 24 h at 37°C with 80 μg V-8 protease added in two aliquots. The fractionation was repeated as after the trypsin digestion and the unretarded material was lyophilised and suspended in 0.5 M acetic acid for digestion at 37°C for 24 h with 70 μg pepsin added in two aliquots. The material was then neutralised with NH_4HCO_3 and centrifuged to remove precipitate prior to gel filtration on Biogel P30 in 0.1 M NH_4HCO_3.

Peptides were hydrolysed for amino acid analysis on an LKB 4400 amino acid analyser (Campbell *et al.*, 1981) or subjected to sequencing on an Applied Biosystem 470A gas phase sequencer using the 02CPTH program or on a Beckman model 890C spinning cup sequencer.

Peptides from L-CA that express OX-22 determinants

OX-22 positive L-CA was purified from spleen membranes solubilised in deoxycholate using an OX-22 MAb affinity column and the large tryptic fragment was produced from this as described for thymocyte L-CA. About 23 nmol of material was solubilised in 0.5 ml 1% SDS and reduced and alkylated after which 2 ml 0.5% Na deoxycholate was added (Johnson *et al.*, 1985a). This material was passed down a 5 ml OX-22 MAb (10 mg/ml) Sepharose 4BCL column and bound material was eluted with 0.05 M diethylamine HCl pH 11.5 plus 0.5% deoxycholate and deoxycholate removed by dialysis against 0.1 M NH_4HCO_3. The material was then trypsin digested and peptides purified by gel filtration and h.p.l.c. as for thymocyte peptides. Two peptides that were recovered in amounts 1.4 nmol and 2.6 nmol respectively were chosen for sequencing and their compositions in residues per mol peptide were: Asp (1.1), Thr (.9), Ser (.4), Glu (.5), Gly (1.4), Ala (.5), Val (2.2), Phe (1), Lys (.5), Arg (.8), glucosamine (1.5) and Asp (2.3), Thr (.1), Ser (2.1) Glu (.3), Gly (.3), Ala (.3), Ile (1), Tyr (1.3), Lys (1.3).

To isolate the OX-22 positive tryptic peptide, OX-22$^+$ spleen L-CA was reduced and alkylated and fully digested with trypsin. Affinity chromatography was carried out and the eluted material was purified on h.p.l.c. to give two peaks in amounts about 2.0 and 0.5 nmol. Both were sequenced (Woollett, 1984; Johnson *et al.*, 1985a) and gave the same partial sequence: Gly Ala Asp Thr Gln–Leu Ser Ser Glu Ala Asp Leu. Position 6 gave no residue and residue 5 is a Gln (Woollett, 1984) not Gly as given in Johnson *et al.* (1985a) due to a typographical error.

Acknowledgements

We are grateful to Mark Smith and Helen Wilmore for excellent technical help, Tony Gascoyne for amino acid analysis and to Ms Caroline Griffin for preparing the manuscript.

References

Arthur,R.P. and Mason,D.W. (1986) *J. Exp. Med.*, **163**, 774–786.
Bause,E. and Hettkamp,H. (1979) *FEBS Lett.*, **108**, 341–344.
Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
Bourguignon,L.Y.W., Suchard,S.J., Nagpal,M.L. and Glenney,J.R. Jr (1985) *J. Cell Biol.*, **101**, 477–487.
Breitbart,R.E., Nguyen,H.T., Medford,R.M., Destree,A.T., Mahdavi,V. and Nadal-Ginard,B. (1985) *Cell*, **41**, 67–82.
Brown,W.R.A., Barclay,A.N., Sunderland,C.A. and Williams,A.F. (1981) *Nature*, **289**, 456–460.
Brown,W.R.A. and Williams,A.F. (1982) *Immunology*, **46**, 713–726.
Campbell,D.G., Gagnon,J., Reid,K.B.M. and Williams,A.F. (1981) *Biochem. J.*, **195**, 15–30.
Carter,P.B. and Sunderland,C.A. (1980) *Immunogenetics*, **10**, 583–593.
Childs,R.A., Dalchau,R., Scudder,P., Hounsell,E.F., Fabre,J.W. and Feizi,T. (1983) *Biochem. Biophys. Res. Commun.*, **110**, 424–431.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.
Coffman,R.L. and Weissman,I.L. (1981) *Nature*, **289**, 681–683.
Dalchau,R. and Fabre,J.W. (1981) *J. Exp. Med.*, **153**, 753–765.
Dalchau,R., Kirkley,J. and Fabre,J.W. (1980) *Eur. J. Immunol.*, **10**, 737–744.
Fabre,J.W. and Williams,A.F. (1977) *Transplantation*, **23**, 349–359.
Frangione,B., Rosenwasser,E., Prelli,F. and Franklin,E.C. (1980) *Biochemistry*, **19**, 4304–4308.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Johnson,P., Williams,A.F. and Woollett,G.R. (1985a) In Springer,T.A. (ed.), *Hybridoma Technology in the Biosciences and Medicine*. Plenum Press, New York and London, pp. 163–175.
Johnson,P., Gagnon,J., Barclay,A.N. and Williams,A.F. (1985b) *EMBO J.*, **4**, 2539–2545.
Kornblitt,A.R., Umezawa,K., Vibe-Pedersen,K. and Baralle,F.E. (1985) *EMBO J.*, **4**, 1755–1759.
Lefrancois,L., Puddington,L., Machamer,C.E. and Bevan,M.J. (1985) *J. Exp. Med.*, **162**, 1275–1293.
Lefrancois,L., Thomas,M.L., Bevan,M.J. and Trowbridge,I.S. (1986) *J. Exp. Med.*, **163**, 1337–1342.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (eds) (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
McMullen,B.A. and Fujikawa,K. (1985) *J. Biol. Chem.*, **260**, 5328–5341.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Michaelson,J., Scheid,M. and Boyse,E.A. (1979) *Immunogenetics*, **9**, 193–197.
Murray,B.A., Owens,G.C., Prediger,E.A., Crossin,K.L., Cunningham,B.A. and Edelman,G.M. (1986) *J. Cell Biol.*, **103**, 1431–1439.
Nawa,H., Kotani,H. and Nakanishi,S. (1984) *Nature*, **312**, 729–734.
Newton,M.R., Wood,K.J. and Fabre,J.W. (1986) *J. Immunogenetics*, **13**, 41–50.
Podell,D.N. and Abraham,G.N. (1978) *Biochem. Biophys. Res. Commun.*, **81**, 176–185.
Ralph,S.J., Thomas,M.L., Morton,C.C. and Trowbridge,I.S. (1987) *EMBO J.*, **6**, 1251–1257.
Saga,Y., Tung,J.-S., Shen,F.-W. and Boyse,E.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6940–6944.
Shackelford,D.A. and Trowbridge,I.S. (1986) *J. Biol. Chem.*, **261**, 8334–8341.
Spickett,G.P., Brandon,M.R., Mason,D.W., Williams,A.F. and Woollett,G.R. (1983) *J. Exp. Med.*, **158**, 795–810.
Staden,R. (1986) *Nucleic Acids Res.*, **14**, 217–231.
Standring,R., McMaster,W.R., Sunderland,C.A. and Williams,A.F. (1978) *Eur. J. Immunol.*, **8**, 832–839.
Thomas,M.L., Barclay,A.N., Gagnon,J. and Williams,A.F. (1985) *Cell*, **41**, 83–93.
Trowbridge,I.S. (1978) *J. Exp. Med.*, **148**, 313–323.
Trowbridge,I.S., Ralph,P. and Bevan,M.J. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 157–161.
Woollett,G.R. (1984) D.Phil. Thesis, Oxford University.
Woollett,G.R., Williams,A.F. and Shotton,D.M. (1985a) *EMBO J.*, **4**, 2827–2830.
Woollett,G.R., Barclay,A.N., Puklavec,M. and Williams,A.F. (1985b) *Eur. J. Immunol.*, **15**, 168–173.
Zamoyska,R., Vollmer,A.C., Sizer,K.C., Liaw,C.W. and Parnes,J.R. (1985) *Cell*, **43**, 153–163.

Received on December 29, 1986; revised on February 27, 1987