The EMBO Journal vol.6 no.5 pp.1251–1257, 1987

Structural variants of human T200 glycoprotein (leukocyte-common antigen)

Stephen J.Ralph, Matthew L.Thomas, Cynthia C.Morton[1] and Ian S.Trowbridge

Department of Cancer Biology, The Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92138, and [1]Department of Genetics, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

Communicated by A.F.Williams

Structural variation in the primary structure of human T200 glycoprotein has been detected. Three cDNA variants have been characterized each of which encode T200 molecules that differ in size as a result of sequence differences in their amino-terminal regions. The largest form of the molecule is distinguished from the smallest by an insert of 161 amino acids, after the first eight amino-terminal residues. The other variant has an insert at the same location of 47 amino acids identical to residues 75–121 in the larger insert. Both extra domains are rich in serine and threonine residues and are likely to display multiple *O*-linked oligosaccharides. These structural variants which probably arise by cell-type-specific alternative splicing provide a molecular basis for the previously observed structural and antigenic heterogeneity of T200 glycoprotein. In addition to the variable amino-terminal region, the external domain of human T200 glycoprotein consists of a second cysteine-rich region of about 400 amino acids, a single transmembrane-spanning region and a large cytoplasmic domain of 707 amino acids shared by all of the structural variants and highly conserved between species. The gene encoding human T200 is located on the long arm of chromosome 1.

Key words: alternative splicing/cDNA sequences/lymphoid cell-specific glycoprotein variants

Introduction

T200 glycoprotein (CD45, McMichael *et al.*, 1987), also known as leukocyte-common antigen (L-CA) (Fabre and Williams, 1977), is a major high mol. wt leukocyte cell surface molecule (Trowbridge *et al.*, 1975; Trowbridge 1978; Sunderland *et al.*, 1979; Omary *et al.*, 1980; Dalchau *et al.*, 1980). It is expressed on all hematopoietic cells except mature red cells and their immediate progenitors. However, it is not found on other differentiated tissues and, as a consequence, can be used as an antigenic marker with which to identify undifferentiated hematopoietic tumors (Battifora and Trowbridge, 1983; Battifora, 1984). It has been well established that the mature glycoprotein exhibits cell-type-specific variation in structure (Trowbridge, 1978; Michaelson *et al.*, 1979; Omary *et al.*, 1980; Newman *et al.*, 1984; Woollett *et al.*, 1985). Initially, it was shown that the glycoprotein found on B cells was larger than that on thymocytes, and subsequent analysis has revealed multiple variants of T200 glycoprotein ranging in M_r from 180 000 to 220 000 present in characteristic arrays on different classes of leukocytes. Further, antigenic variants of the glycoprotein have been defined by monoclonal antibodies in human, mouse and rat (Dalchau and Fabre, 1981; Coffman and Weissman, 1981; Landreth *et al.*, 1983; Spickett *et al.*, 1983). Recently, cDNA clones of rat thymocyte L-CA were obtained from which a partial protein sequence of 1073 amino acids was deduced (Thomas *et al.*, 1985). It was estimated from the size of thymocyte L-CA mRNA that ~100–150 amino acids at the N-terminus of the thymocyte form of the molecule remained to be determined. Subsequently, B cell cDNA clones were obtained which spanned ~ 3 kb of the 3′ end of the molecule. However, no differences were found by limited sequencing and restriction mapping between the rat thymocyte and B cell forms of the glycoprotein. Cell-specific differences in the size of rat L-CA mRNAs from thymocytes and B cells were detected that correlated with the differences in M_r of the glycoproteins displayed on their cell surface (Thomas *et al.*, 1985). Similar observations were later made in the mouse (Shen *et al.*, 1985; Lefrancoise *et al.*, 1986). Further multiple forms of T200 glycoprotein were detected early after synthesis that were unlikely to be accounted for by differences in glycosylation (Lefrancoise *et al.*, 1986). Together, these results suggested that there may be cell-specific variants of T200 glycoprotein that differ in primary structure presumably close to the N-terminus of the molecule.

Here, we report the isolation of cDNA clones of T200 glycoprotein from a variety of human lymphoid cells. The complete primary structure of the molecule has been deduced from the cDNA sequence of these clones and three structural variants have been identified. In the accompanying report (Barclay *et al.*, 1987) related L-CA variants in the rat are described.

Table I. Summary of cDNA libraries and clones

Human cell source	Cell type and reference	Cloning vector	No. of unique recombinants	cDNA clones
Tonsil	–	pcD	$1–1.5 \times 10^5$	pHLC-1, -2
IB4	EB virus transformed lymphocyte (King *et al.*, 1980)	λgt10	1×10^5	λHLC-1
RPMI 4265	B-CML (Imamura *et al.*, 1970)	λgt10	1.3×10^6	λHLC-2
Raji	Burkitt's lymphoma (Pulvertaft, 1965)	λgt11	5×10^5	λHLC-3
Jurkat	T-ALL (Schneider *et al.*, 1977)	λgt10	1.1×10^6	λHLC-4

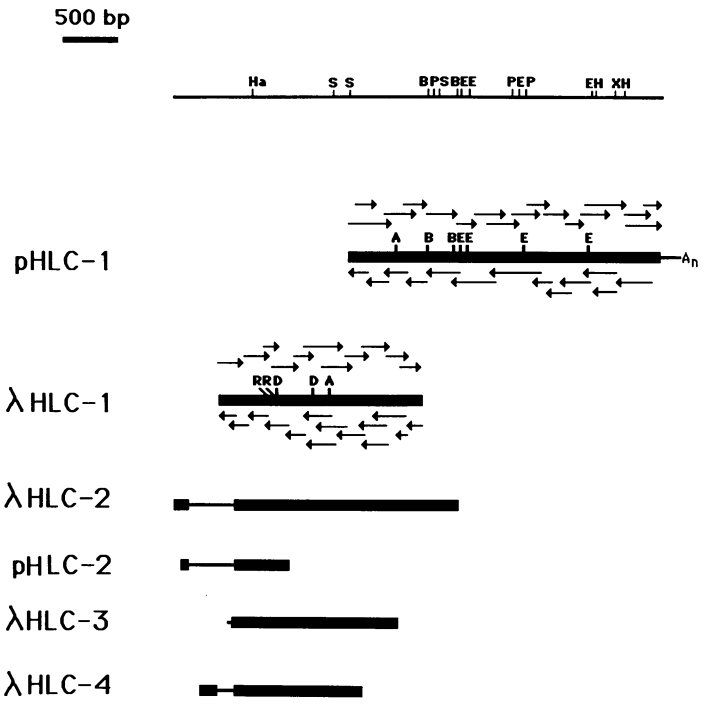

Fig. 1. Analysis of human T200 cDNA clones. The origin and relative positions of the cDNA clones are shown together with a partial restriction enzyme map of the largest T200 form. The complete sequence of pHLC-1 and λHLC-1 was obtained in both directions by the dideoxy method., The arrows indicate individual sequences used to obtain the entire structure. The regions encoding the variant inserts not present at the 5′ end of λHLC-1 are represented by the thin lines. The cellular origin and size of the cDNAs are as follows: pHLC-1 (3.6 kb) and pHLC-2 (1.0 kb) from tonsil, λHLC-1 from IB4 (2.0 kb); λHLC-2 (2.8 kb) from RPMI 4265; λHLC-3 (1.8 kb) from Raji and λHLC-4 (1.8 kb) from Jurkat. The restriction enzyme sites are: A = *Ava*II, B = *Bam*HI, H = *Hin*dIII, Ha = *Hae*II, D = *Dra*I, P = *Pst*I, E = *Eco*RI, R = *Rsa*I, S = *Stu*I, X = *Xba*I.

Results

Isolation of human T200 cDNA clones and complete nucleotide sequence

A cDNA library derived from human tonsil (see Table I) was screened by hybridization using the rat L-CA probe, pLC-1.2, derived from rat thymocyte cDNAs, pLC-1 and pLC-2, by ligation at the unique *Xba*I site (Thomas *et al.*, 1985). This led to the isolation of a plasmid containing a 3.6-kb cDNA insert designated pHLC-1. Limited sequencing by the method of Maxam and Gilbert (1980) established the authenticity of this cDNA and allowed alignment of the clone with the 3′ end of rat L-CA. A 520-bp *Ava*II fragment from the 5′ end of pHLC-1 (see Figure 1) was then used to screen a λgt10 cDNA library derived from the human B lymphoblastoid cell line, IB4 (Table I). A recombinant bacteriophage containing a 2.0-kb insert (λHLC-1) which extended past the 5′ end of pHLC-1 by ~1.1 kb was isolated (Figure 1). These two clones were completely sequenced in both directions (Figure 1). A large open reading frame was found which extended from nucleotide 132 to nucleotide 3575 and spanned the entire coding region of human T200 glycoprotein (Figure 2). The nucleotide sequence and predicted amino acid sequence of T200 encoded by these two clones is shown in Figure 2. There are two ATG triplets (positions 141–143 and 147–149) near the beginning of the open reading frame. It is more likely that the second ATG is the initiation codon based upon the consensus sequence for eukaryotic initiation sites $CC^{A}_{G}CCATG(G)$ (Kozak, 1984). A hydrophobic sequence of 23 amino acids characteristic of a classical leader sequence (Watson, 1984) was present. Based upon the cleavage sites of other signal sequences (Watson, 1984), the most likely N-terminal amino acid of the mature protein is glutamine and thus a mature T200 protein of 1120 amino acids is predicted. There are striking structural similarities between human T200 and the previously reported sequences of rat thymocyte L-CA and mouse thymocyte T200 glycoprotein (Thomas *et al.*, 1985; Saga *et al.*, 1986). There is a hydrophobic stretch of 22 residues (392–413) that are highly conserved in all three species and represent the hydrophobic membrane spanning region. By analogy with rat L-CA (Thomas *et al.*, 1985), human T200 glycoprotein is almost certainly oriented with its carboxy-terminus on the cytoplasmic side of the membrane. The putative cytoplasmic domain of human T200 is 707 amino acids in length and is remarkably similar in structure to that of the rat L-CA and mouse T200. At the amino acid level the conservation of sequence between the three species approaches 90%. As in the other two species, the cytoplasmic domain consists of two subdomains (residues 427–717 and 718–1033) of ~300 amino acids which have an amino acid sequence homology of about 30%. The extracellular domain of human T200 glycoprotein encoded by λHLC-1 consists of a small N-terminal stretch of about 40 amino acids that are rich in serine and threonine residues that are potential sites for *O*-linked glycosylation followed by a relatively cysteine-rich region (16/351 residues) (Figure 2). The size of the extracellular domain of human T200 encoded by λHLC-1 is similar to that of the mouse T200 and rat L-CA isolated from thymocytes (Saga *et al.*, 1986; Barclay *et al.*, 1987). A potential polyadenylation signal ATTAAA was identified at positions 4574–4579 followed by a poly(A) tail ~80 residues downstream. Thus, the 3′ untranslated region of pHLC-1 is ~1.1 kb.

Structural variants of human T200

Because of the evidence that the structural variants of the mature T200 glycoprotein on different lymphoid cell types may reflect structural differences in primary structure at the N-terminus of the molecule, additional cDNAs from libraries derived from a variety of human lymphoid cells were isolated by screening with the 520-bp *Rsa*I fragment from the 5′ end of the λHLC-1 (see Figure 1). Four additional cDNA clones encoding the amino-terminal region of the molecule were isolated: (i) pHLC-2 was isolated from the original tonsil library, (ii) λHLC-2 from a λgt10 library derived from RPMI 4265, an Epstein–Barr virus (EBV)-transformed lymphoblastoid B cell line, (iii) λHLC-3 from Raji, a Burkitt's lymphoma cell line, and (iv) λHLC-4 from Jurkat, a T cell line (Table I and Figure 1). Sequencing of the 5′ ends of these clones showed that they differed in structure from that of λHLC-1 but that each were related. The cDNA from RPMI 4265 cells, λHLC-2, predicted a protein sequence that was identical to λHLC-1 for the first eight N-terminal amino acids. It then, however, contained an insertion of 483 bp encoding an additonal 161 amino acids (Figure 3). The predicted protein sequence of λHLC-2 returned to that found for λHLC-1 at position nine. The 5′ terminus of the pHLC-2 clone derived from tonsil was identical to that of λHLC-2. A smaller insert was found in the λHLC-4 cDNA derived from Jurkat. The sequence of this smaller insert was identical to nucleotide residues 225–366 lying within the insert of the RPMI 4265 λHLC-2 clone (Figure 3). The λHLC-3 clone from Raji was incomplete. However, the predicted sequence at its N-terminus was identical for four amino acids with

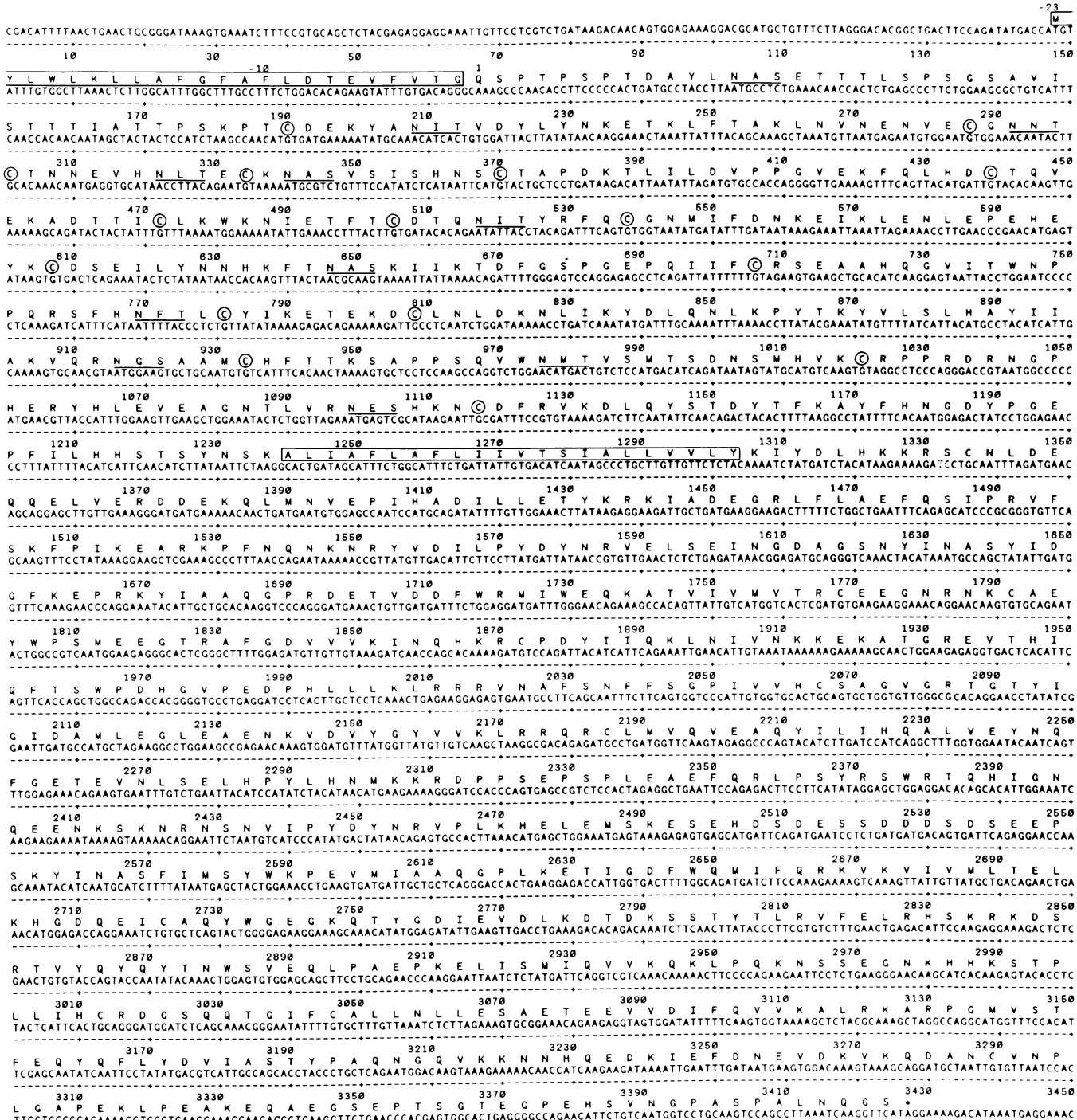

Fig. 2. Complete nucleotide and predicted primary sequence of human T200. The sequence determined from cDNA clones pHLC-1 and λHLC-1 were combined to obtain the complete nucleotide [4597 bp excluding the poly(A) tail] and predicted protein sequence of human T200. The predicted sequence is initiated at methionine corresponding to codon position 147 and terminated at position 3578. The leader sequence and transmembrane regions are boxed. The potential sites of N-linked glycosylation and poly-adenylation signal sequence immediately 5′ to the poly-adenylation tract are underlined and the cysteine residues in the extracellular domain have been circled. The nucleotide numbers are given beneath the sequence and the amino acid number at the end of each line.

the carboxy-terminus of the cDNA clones containing the large inserts encoded by λHLC-2 and pHLC-2 isolated from RPMI 4265 and tonsil. Restriction mapping of the four clones did not reveal any differences at the 3′ ends of the clones and the fragments obtained were consistent with the nucleotide sequence derived from λHLC-1 and pHLC-1. It is likely, therefore, that

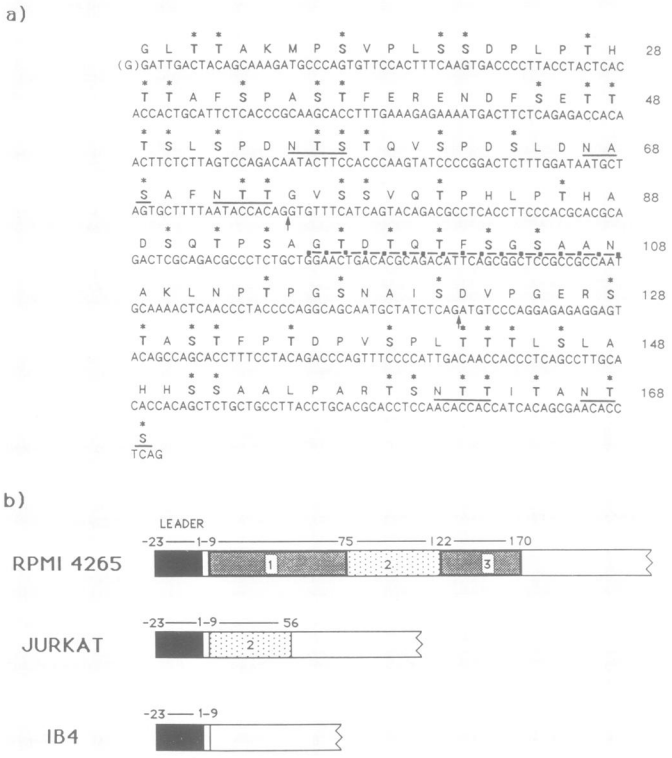

Fig. 3. The nucleotide and predicted primary sequence of the T200 variant inserts. (**a**) shows the nucleotide and predicted amino acid sequence of the 161 amino acid insert encoded within λHLC-2. The hydroxy amino acids are shown in bold type and marked by asterisks. The potential sites for N-linked glycosylation are underlined. The amino acid sequence (residues 96–108) homologous to that of the tryptic peptide known to bear the restricted L-CA antigenic determinant recognized by MAb OX-22 in the rat (Johnson *et al.*, 1985) is indicated by the dashed underline. (**b**) is a schematic representation of the 5′ ends of the three variants of human T200 that have been identified by DNA sequencing. Each consists of identical leader sequences (solid) and shared coding sequences (open). The insert of 161 amino acids in the T200 cDNA from RPMI 4265 cells (residues 9–169) are shown as segments 1–3. The insert of 47 amino acids in the T200 cDNA from Jurkat cells is identical to that encoded by segment 2 of RPMI 4265 cells indicated by arrows in (a).

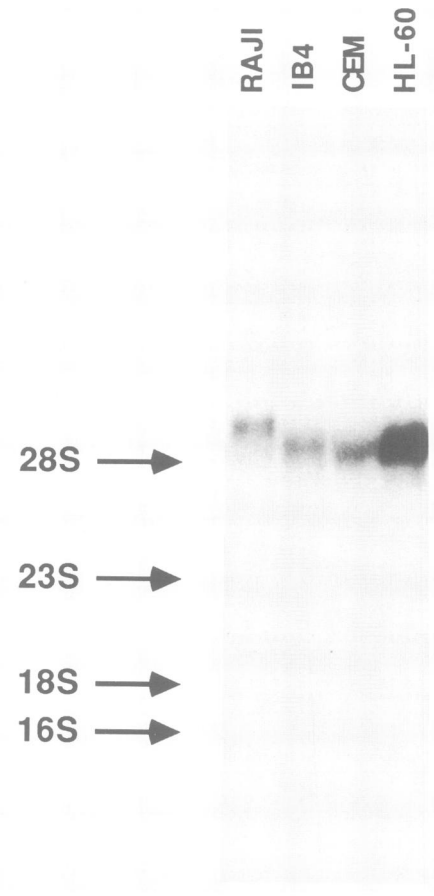

Fig. 4. Northern blot analysis of human T200 mRNA. Poly(A)$^+$ mRNA (5 μg each) from Raji, IB4, CEM and HL 60 cells were fractionated on a 0.7% agrarose formaldehyde denaturing gel and blotted onto Zeta-probe. The filter was hybridized with a nick-translated probe of the pHLC-1 and λHLC-1 cDNAs. Eukaryotic and prokaryotic ribosomal RNAs were used as markers.

the structural variation is limited to the 5′ regions of the clones sequenced. The nucleotide sequence and predicted amino acid sequence of the larger insert are shown in Figure 3. The sequence is extremely rich in serine and threonine residues (59 of 161 residues), proline (17 residues) and small aliphatic amino acids suggesting that this region of the molecule has little ordered secondary structure and is likely to provide multiple attachment sites for *O*-linked sugars.

Northern blot analysis of human T200 mRNA

It has been shown in the rat (Thomas *et al.*, 1985) and the mouse (Shen *et al.*, 1985; Lefrancoise *et al.*, 1986) that cell-type specific differences in the size of T200 mRNA can be detected. To confirm that similar differences exist in the human, Northern blot analysis of T200 mRNA from four human lymphoid cell lines is shown in Figure 4. T200 mRNA isolated from Raji cells was clearly larger than those isolated from IB4, CCRF-CEM and HL-60 cells. This is consistent with the limited Raji cDNA sequence suggesting that the cDNA isolated from Raji cells contains a large insert. The size of T200 mRNA from IB4 cells is consistent with the cDNA sequence data and suggests these cells express a low M_r form of T200 glycoprotein. The T200 mRNA isolated from Jurkat cells could not be distinguished from that of IB4 or CCRF-CEM cells (data not shown).

Chromosomal mapping of human T200

In situ hybridization was performed using the 2-kb λHLC-1 cDNA subcloned into pUC18 to identify the location of the human T200 gene. A total of 152 metaphase spreads were analyzed and of 571 grains, 26 (4.6%) were found localized on chromosome 1 at the region of q31-32 (Figure 5). This was the only chromosomal region which gave a grain count significantly above background.

Discussion

Three variants of human T200 glycoprotein that differ in primary structure in their N-terminal regions have been identified. As determined by cDNA sequencing, the largest form of the molecule from the B cell line, RPMI 4265, differs from the smallest found in another human B cell line, IB4, by an insert of 161 additional amino acids after residue 8. The third variant isolated from the human T cell line Jurkat has a 47 amino acid insert at the same position with a sequence identical to residues 75–121 of the larger insert (Figure 3).

The existence of multiple forms of T200 mRNA which differ in their coding region provides a likely molecular basis for the different variants of the mature T200 glycoprotein found on the

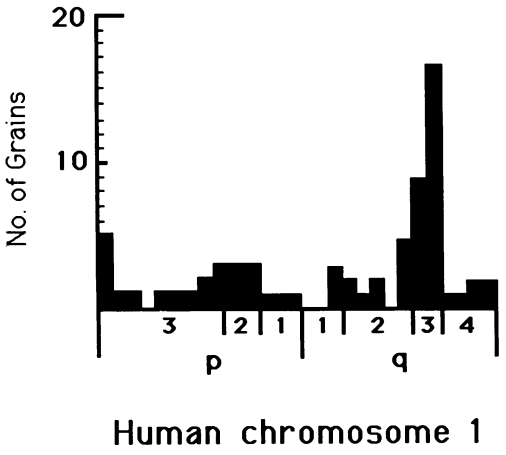

Fig. 5. Assignment of human T200 gene to chromosome 1. *In situ* hybridization of the λHLC-1 cDNA probe to the entire complement of human chromosomes was performed as described in Materials and methods. The figure shows the silver grain distribution for chromosome 1 with significant hybridization localized to the q31-32 region. The grain distribution for other chromosomes was at background levels.

cell surface. At least four different variants of human T200 glycoprotein ranging in size from 180 — 220 K can be distinguished by SDS−polyacrylamide gel electrophoresis. A reasonable hypothesis is that the structural variants encoded by the different cDNAs correspond to these discrete molecular species. The largest human T200 cDNA variant probably encodes the predominant B cell form of the molecule whereas the smallest variant encodes the basic thymocyte form of the glycoprotein. Consistent with this idea, T200 mRNA from the Burkitt's lymphoma cell line, Raji, which expresses the high M_r form of the glycoprotein (Deane *et al.*, 1984) was significantly larger than that isolated from IB4 cells and the T cell lines CCRF-CEM and Jurkat. Studies using antibodies specific for the insert sequences should clarify these relationships.

Rat and mouse T200 cDNAs encoding similar variants have been isolated (Barclay *et al.*, 1987; Thomas *et al.*, 1987). However, the largest T200 variants found in these species lack a contiguous stretch of 26 amino acids present in the largest human insert (see Barclay *et al.*, 1987). These variants may therefore be related but distinct from the largest human form of the molecule. The smallest human T200 variant identified by cDNA sequencing is homologous to that of a T200 (Ly-5) cDNA isolated from mouse thymocytes (Saga *et al.*, 1986) and the smallest L-CA cDNA isolated from rat thymocytes (Barclay *et al.*, 1987). The fact that the human cDNA was isolated from the EBV-transformed B cell line, IB4, does not preclude that it is the basic thymocyte form of the molecule. The pattern of T200 glycoprotein species on human B cell tumor cell lines is variable and some express only the low M_r form of the glycoprotein (Morishima *et al.*, 1982; Deane *et al.*, 1984). It will be of interest to determine at which stage of B cell development expression of the high mol. wt form of T200 glycoprotein occurs. It is not known how many different variants of T200 glycoprotein are expressed in the human. At least one additional form of the molecule has been detected in the rat (Barclay *et al.*, 1987) and it is possible more variants exist in both species.

The amino acid sequence of the human T200 variants deduced from the cDNA nucleotide sequences provides insight into the molecular basis of the antigenic heterogeneity of T200 glycoprotein. In the human, mouse and the rat, monoclonal antibodies have been identified which react only with forms of T200 glycoprotein expressed on B cells, cytotoxic T cells and a fraction of T helper cells (Dalchau and Fabre, 1981; Kincade *et al.*, 1981; Ledbetter *et al.*, 1985; Woollett *et al.*, 1985). These antibodies appear to react preferentially with the high mol. wt forms of T200 glycoprotein. One of these antibodies with restricted specificity, OX-22, has been shown to react with a peptide of known sequence derived from the N-terminal region of rat L-CA (Johnson *et al.*, 1985; Barclay *et al.*, 1987). The homologous sequence in human T200, GTDTQFSGSAAN (Figure 3) is located in the large B cell insert as well as the smaller insert found in the T200 cDNA from Jurkat cells. It is likely, therefore, that monoclonal antibodies which react with restricted subpopulations of human T200 of which F10-89-4 is the prototype (Dalchau and Fabre, 1981) recognize an antigenic determinant located in the same region of the molecule. It might be predicted that antibodies against antigenic determinants found only in the largest human T200 glycoprotein variant would have an even more restricted cellular distribution. Monoclonal antibodies have been identified in the mouse and rat which appear to react selectively with B cells (Coffman, 1983; Woollett *et al.*, 1985). A more precise analysis of functional human lymphocyte subsets may now be possible with antibodies or nucleic acid probes specific for each of the variant forms of T200.

Although cDNA sequencing has established that there are multiple protein variants of T200, previous work suggested that different oligosaccharides are displayed on the B and T cell forms of the molecule (Morishima *et al.*, 1982; Brown and Williams, 1982). Only the human B cell form of the glycoprotein shows strong reactivity with anti-i and anti-I antibodies which recognize antigenic determinants on *O*-linked linear or branched oligosaccharides, respectively (Childs and Feizi, 1981; Childs *et al.*, 1983). It is probable that some of the multiple serine and threonine residues found in the high M_r variants of T200 glycoprotein serve as attachment sites for *O*-linked oligosaccharide chains comprised of repeating N-acetyl lactosamine units with which the anti-i and anti-I antibodies react. The calculated M_r of the three human T200 variant polypeptides are 128 234, 134 876 and 144 590, respectively. Previous estimates for the mol. wts of the different forms of mouse T200 after removal of N-linked sugars or blocking of N-linked glycosylation are between 150 and 170 K for the T cell forms and 190 K for the B cell form (Tung *et al.*, 1984; Banga *et al.*, 1984; Lefrancoise *et al.*, 1985). It is likely, therefore, that the higher mol. wt variants of T200 glycoprotein contain additional oligosaccharides that contribute to their M_r determined by SDS−polyacrylamide gel electrophoresis.

The molecular mechanism by which the variants of T200 glycoprotein are generated has not been firmly established. However, the present evidence implicates alternative mRNA splicing. First, overlapping recombinant lambda clones encompassing 80 kb of the mouse genome have been isolated and their analysis suggests that there is only one gene encoding T200 glycoprotein that encompasses at least 60 kb (Saga *et al.*, 1986; Thomas *et al.*, 1987). Thus, it is unlikely the variant T200 molecules represent the products of different genes. Furthermore, homologous nucleotide sequences span both the 5′ and 3′ ends of the inserts of the two larger human T200 cDNAs. At the junction of the 5′ end of the larger insert is the sequence CCACTGGA which is identical at six of eight positions with the sequence CCACAGGT found at the 5′ end of smaller insert. The sequence, CTCAGATG, is found at the junction of the 3′ ends of both inserts. These repeat motifs are related to the consensus sequence

($^{C}_{A}$AG/GT) for exon/exon splice junctions (Mount, 1982; Rogers, 1985) suggesting that the multiple variants of human T200 glycoprotein arise by cell-type-specific alternative splicing. If so, this would be analogous to the fibronectin variants generated by alternative splicing involving both exon skipping and exon subdivision (Hynes, 1985; Paul *et al.*, 1986). It is possible that either or both of these mechanisms also operate to generate the multiple variants of T200 glycoprotein.

A comparison of the cytoplasmic domains of human rat and mouse T200 glyco-proteins reveals a remarkable structural conservation that approaches 90% homology at the amino acid level (Thomas *et al.*, 1985; Saga *et al.*, 1986; Thomas *et al.*, 1987). This degree of conservation over the entire cytoplasmic domain of more than 700 amino acids is consistent with the possibility that this domain interacts with structural elements at the cytoplasmic face of the cell membrane. It has been suggested that T200 is specifically associated with a cytoskeletal component fodrin (Bourguignon *et al.*, 1985). It is noteworthy that with the exception of receptors such as those for epidermal growth factor and insulin that have large intracytoplasmic domains (542 and 402 respectively) with intrinsic tyrosine-kinase activity (Ullrich *et al.*, 1984, 1985; Ebina *et al.*, 1985), the only other well-characterized membrane protein with a relatively large cytoplasmic domain (362 amino acids) is the 160 K M_r-form of the chicken neural cell-adhesion molecule N-CAM (Hemperly *et al.*, 1986). This polysialic acid-rich glycoprotein is believed to mediate cell−cell interactions with embryonic and adult neural tissues. T200 glycoprotein may play a similar role within the lymphoid system with its variable extracellular domain modulating interactions with other cells or extracellular matrixes.

The gene encoding human T200 has been mapped by *in situ* hybridization to the long arm of chromosome 1 in the region q31-32. It is known from segregation studies of the Ly-5 allotypes that the T200 gene is also on chromosome 1 in the mouse (Scheid *et al.*, 1982). Regions of the short arm of human chromosome 1 show homology to mouse chromsomes 3 and 4 (Sawyer and Hozier, 1986). However, the locus for peptidase C which maps to the long arm of human chromosome 1 is found on mouse chromosome 1 (Lalley and McKusick, 1985). It is noteworthy that the genes encoding C4 binding protein, factor H, and the complement receptors CR-1 and CR-2 map to the same region of human chromosome 1 (Klickstein *et al.*, 1985; Weis *et al.*, 1987).

Materials and methods

The human tonsil and Raji cDNA libraries were obtained from Dr R.Wetsel and Dr G.Nemerow respectively, at Scripps Clinic and Research Foundation, La Jolla, CA. The RPMI-4265 and Jurkat cDNA libraries were purchased from Clontech Laboratories, Palo Alto, CA. The IB4 cDNA library was obtained from Dr S.Speck, Dana-Farber Cancer Institute, Boston, MA. The properties of the libraries and the origin of the cells from which they were derived are given in Table I. The nick-translation kit used to radiolabel probes was from BRL Life Technologies, Inc. Gaithersburg, MD. Other reagents and enzymes were from commercial sources. Radioisotopes were purchased from New England Nuclear.

Isolation of cDNA clones

Between 1 and 4 × 10^5 transformants from each library were screened with nick translated probe of specific radioactivity between 1 and 2 × 10^8 c.p.m./μg according to the procedures outlined by Maniatis *et al.* (1982) except that filters were soaked for 5 min at room temperature in 50 mM Tris, pH 8, 1 M NaCl, 0.1% SDS before prehybridization and hybridization. The prehybridization and hybridization conditions were those of Gatti *et al.* (1984). The first human T200 cDNA clone was obtained by screening a tonsil library (Table I) under conditions of low stringency (0.3 M NaCl, 20 mM sodium citrate, pH 7.0, 0.1% SDS at 52°C) using the rat cDNA probe, pLC-1.2. The library was then rescreened using the small insert from this clone to obtain pHLC-1. The 5′ *Ava*I fragment of pHLC-1 (see Figure 1) was used as a probe to obtain λHLC-1 and subsequently the 5′ *Rsa*I (500 bp) fragment of this clone containing the leader sequence was used to obtain all the other cDNA clones.

DNA sequencing strategy

Overlapping deletions of each cDNA was obtained according to the procedure of Dale *et al.* (1985) as modified in the procedure manual for the Cyclone system of rapid deletion subcloning (International Biotechnologies, Inc., New Haven, CT). The overlapping cDNA deletions were sequenced using the dideoxy procedure of Sanger *et al.* (1977) and the DNA sequences were compiled and analyzed using the programs of Staden (1982, 1984) and Devereux *et al.* (1984).

Northern blotting

Total RNA was isolated from cells using the guanidinium isothiocyanate lysis method of Chirgwin *et al.* (1979). Poly(A)$^+$ RNA was selected by a single passage over oligo(dT)−cellulose (Maniatis *et al.*, 1982). RNA was electrophoresed at 30 mA for 24 h on a 0.7% agarose gel containing formaldehyde and blotted onto Zetaprobe (Biorad Laboratories, Richmond, CA) using 20 mM sodium phosphate buffer (pH 6.5) for transfer. Prehybridization and hybridization were by the method of Gatti *et al.* (1984) except that the Northern was prewashed at 50°C in 7.5 mM sodium citrate, pH 7, 75 mM NaCl, 0.1% SDS from 1 h before prehybridizing at 42°C overnight. The probe used was a mixture of pHLC-1 (1.5 × 10^7 c.p.m.) and λHLC-1 (1.5 × 10^7 c.p.m.) in 20 ml hybridization buffer. The filter was hybridized at 42°C overnight before washing in 1.5 mM sodium citrate, pH 7, 15 mM NaCl, 0.1% SDS at 52°C and exposure to Kodak X-omat AR-5 film.

Chromosomal mapping

The λHLC-1 cDNA was subcloned into the *Eco*RI site of pUC18 and this recombinant plasmid was used for chromosomal *in situ* hybridization as previously described (Morton *et al.*, 1984).

Acknowledgements

We thank Drs S.Speck and J.Strominger, R.Wetsel and B.Tack and G.Nemerow for generously providing cDNA libraries. We also thank Dr J.Weis for sharing data prior to publication and for advice. We thank Aileen Chain, Pamela Reynolds, Ashleigh Head and Albert Smith for technical assistance and Ami Koide for preparation of the manuscript. This work was supported by grant CA 17733 from the National Cancer Institute. M.L.T. is a Special Fellow of the Leukemia Society of America. S.J.R. was supported by an overseas training fellowship of the New Zealand Cancer Society. C.C.M. was supported by a postdoctoral fellowship, CA-07511, of the National Cancer Institute.

References

Banga,J.P., Guarnotta,G., Harte,A., Pryce,G., Campbell,M.A., Quartey Papafio,R., Lydyard,P.M. and Roitt,I.M. (1984) *Scand. J. Immunol.*, **19**, 11−21.

Barclay,A.N., Jackson,D.I., Willis,A.C. and Williams,A.F. (1987) *J. Eur. Mol. Biol. Org.*, in press.

Battifora,H. (1984) *Seminars in Diagnostic Pathology*, **1**, 251−271.

Battifora,H. and Trowbridge,I.S. (1983) *Cancer*, **51**, 816−821.

Bourguignon,L.Y.W., Suchard,S.J., Nagpal,M.L. and Glenny,J.R. (1985) *J. Cell Biol.*, **101**, 477−487.

Brown,W.R.A. and Williams,A.F. (1982) *Immunology*, **46**, 713−726.

Childs,R.A. and Feizi,T. (1981) *Biochem. Biophys, Res. Commun.*, **102**, 1158−1164.

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. (1983) *Immunol. Rev.*, **69**, 5−23.

Coffman,R.L. and Weissman,I.L. (1981) *Nature*, **289**, 681−683.

Dalchau,R., Kirkley,J. and Fabre,J.W. (1980) *Eur. J. Immunol.*, **10**, 737−744.

Dalchau,R. and Fabre,J.W. (1981) *J. Exp. Med.*, **153**, 753−765.

Dale,R.M.K., McClure,B.A. and Houchins,J.P. (1985) *Plasmid*, **13**, 31−40.

Deane,D.L., Cohen,B.B., Morton,J.E. and Steel,C.M. (1984) *Int. J. Cancer*, **34**, 459−462.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387−395.

Ebina,Y., Ellis,L., Jarnagin,K., Edery,M., Graf,L., Clauser,E., Ou,J.-H., Masiarz,F., Kan,J.W., Goldfine,I.D., Roth,R.A. and Rutter,W.J. (1985) *Cell*, **40**, 747−758.

Fabre,J.W. and Williams,A.F. (1977) *Transplantation*, **23**, 349−359.

Gatti,R.A., Concannon,P. and Salser,W. (1984) *Biotechniques*, **2**, 148−155.

Hemperly,J.J., Murray,B.A., Edelman,G.M. and Cunningham,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3037–3041.
Hynes,R. (1985) *Annu. Rev. Cell Biol.*, **1**, 67–90.
Imamura,T., Huang,C.C., Minowada,J., Takahashi,M. and Moore,G.E. (1970) *J. Natl. Cancer Inst.*, **44**, 845–853.
Johnson,P., Williams,A.F. and Woollett,G.R. (1985) In Springer,T.A. (ed.), *Hybridoma Technology in the Biosciences and Medicine*. Plenum Press, New York, pp. 163–175.
Kincade,P.W., Lee,G., Watanabe,T., Sun,L. and Scheid,M.P. (1981) *J. Immunol.*, **127**, 2262–2268.
King,W., Thomas-Powell,A.L., Raab-Traub,N., Hawke,M. and Kieff,E. (1980) *J. Virol.*, **36**, 506–518.
Klickstein,L.D., Wong,W.W., Smith,J.A., Morton,C., Fearon,D.T. and Weis,J.H. (1985) *Complement*, **2**, 44 (abstract).
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Lalley,P.A. and McKusick,V.A. (1985) *Cytogenet. Cell Genet.*, **40**, 536–566.
Landreth,K.S., Kincade,P.W., Lee,G. and Medlock,E.S. (1983) *J. Immunol.*, **131**, 572–580.
Ledbetter,J.A., Rose,L.M., Spooner,C.E., Beatty,P.G., Martin,P.J. and Clark,E.A. (1985) *J. Immunol.*, **135**, 1819–1825.
Lefrancoise,L., Puddington,L., Machamer,C.E. and Bevan,M.J. (1985) *J. Exp. Med.*, **162**, 1275–1293.
Lefrancoise,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. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
McMichael *et al.* (1987) *Leukocyte Typing. III:. White Cell Differentiation Antigens*. Oxford University Press, Oxford, in press.
Michaelson,J., Scheid,M. and Boyse,E.A. (1979) *Immunogenetics*, **9**, 193–197.
Morishima,Y., Ogata,S.-I., Collins,E.F., Dupont,B. and Lloyd,K.O. (1982) *Immunogenetics*, **15**, 529–535.
Morton,C.C., Kirsch,I.R., Taub,R.A., Orkin,S.H. and Brown,J.A. (1984) *Am. J. Hum. Genet,*, **36**, 576–585.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 458–472.
Newman,W., Targan,S.R. and Fast,L.D. (1984) *Mol. Immunol.*, **21**, 1113–1121.
Omary,M.B., Trowbridge,I.S. and Battifora,H.A. (1980) *J. Exp. Med.*, **152**, 842–852.
Paul,J.I., Schwarzbauer,J.E., Tamkun,J.W. and Hynes,R.O. (1986) *J. Biol. Chem.*, **261**, 12258–12265.
Pulvertaft,R.J.V. (1965) *J. Clin. Path.*, **18**, 261–273.
Rogers,J.H. (1985) *Int. Rev. Cytol.*, **93**, 187–279.
Saga,Y., Tung,J.-S., Shen,F.-W. and Boyse,E.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6940–6944.
Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Sawyer,J.R. and Hozier,J.C. (1986) *Science*, **232**, 1632–1635.
Scheid,M.P., Landreth,K.S., Tung,J.-S. and Kincade,P.W. (1982) *Immunol. Rev.*, **69**, 141–159.
Schneider,U., Schwenk,H.-U. and Bornkamm,G. (1977) *Int. J. Cancer*, **19**, 621–626.
Shen,F.-W., Saga,Y., Litman,G., Freeman,G., Tung,J.-S., Cantor,H. and Boyse,E.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7360–7363.
Spickett,G.P., Brandon,M.R., Mason,D.W., Williams,A.F. and Woollett,G.R. (1983) *J. Exp. Med.*, **157**, 795–810.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 4731–4751.
Staden,R. (1984) *Nucleic Acids Res.*, **12**, 521–538.
Sunderland,C.A., McMaster,W.R. and Williams,A.F. (1979) *Eur. J. Immunol.*, **9**, 155–159.
Thomas,M.L., Barclay,A.N., Gagnon,J. and Williams,A.F. (1985) *Cell*, **41**, 83–93.
Thomas,M.L., Shackelford,D., Ralph,S. and Trowbridge,I.S. (1987) *J. Receptor Research*, in press.
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.
Tung,J.-S., Deere,M.C. and Boyse,E.A. (1984) *Immunogenetics*, **19**, 149–154.
Ullrich,A., Coussens,L., Hayflick,J.S., Dull,T.J., Gray,A., Tam,A.W., Lee,J., Yarden,Y., Liberman,T.A., Schlessinger,J., Downard,J., Mayes,E.L.V., Whittle,N., Waterfield,M.D. and Seeburg,P.H. (1984) *Nature*, **309**, 418–425.
Ullrich,A., Bell,J.R., Chen,E.Y., Herrera,R., Petruzzelli,L.M., Dull,T.J., Gray,A., Coussens,L., Liao,Y;-C., Tsubokawa,M., Mason,A., Seeburg,P.H., Grunfeld,C., Rosen,O.M. and Ramachandran,J. (1985) *Nature*, **313**, 756–761.
Watson,M.E.E. (1984) *Nucleic Acids Res.*, **12**, 5145–5164.
Weis,J.H., Morton,C.C., Bruns,G.A.P., Weis,J.J., Klickstein,L.L., Wong,W.W. and Fearson,D.T. (1987) *J. Immunol.*, **138**, 312–315.
Woollett,G.R., Barclay,A.N., Puklavec,M. and Williams,A.F. (1985) *Eur. J. Immunol.*, **15**, 168–173.

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

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00062.