DIFFERENTIAL USAGE OF THREE EXONS GENERATES AT LEAST FIVE DIFFERENT mRNAs ENCODING HUMAN LEUKOCYTE COMMON ANTIGENS

BY MICHEL STREULI,* LAURIE R. HALL,* YUMIKO SAGA,‡
STUART F. SCHLOSSMAN,* and HARUO SAITO*

From the *Division of Tumor Immunology, Dana-Farber Cancer Institute, and Departments of
Pathology, Medicine, and Biological Chemistry, Harvard Medical School, Boston,
Massachusetts 02115; and ‡Memorial Sloan-Kettering Cancer Center, New York 10021

T lymphocytes consist of heterogeneous subpopulations of cells that can be distinguished by their distinct pattern of cell surface antigen expression. The T4 antigen (CD4), which is thought to be the receptor for the MHC class II antigens, is expressed on a subpopulation of T lymphocytes that provides helper function, whereas the T8 antigen (CD8) is expressed on a subpopulation that provides cytotoxic and suppressor activity (1). A monoclonal antibody, anti-2H4, further subdivides the T4$^+$ T lymphocyte population into T4$^+$, 2H4$^+$ cells which induce suppressor activity, and T4$^+$, 2H4$^-$ cells, with helper function (2). The observation that the induction of suppressor activity is blocked by the anti-2H4 antibody may suggest that the 2H4 antigen is directly involved in the induction process (3). Recent biochemical analysis (4) has demonstrated that the anti-2H4 mAb recognizes a subset of the cell surface glycoproteins known variously as leukocyte common antigens (LCAs),[1] T200, or CD45.

The human LCAs are a family of several structurally related, high-molecular-mass (170–240 kD) proteins that are found in abundance on the surface of lymphocytes and other hematopoietic cells. Some anti-LCA mAbs react with all forms of LCAs, which are resolved into four or more distinct proteins by SDS-PAGE, while other anti-LCA mAbs, including anti-2H4, react only with subsets of these proteins (4–9). It seems, therefore, that LCA molecules have both common and variable epitopes in the extracellular region. B lymphocytes express predominantly the higher-molecular-mass forms, whereas T lymphocytes express mainly lower-molecular-mass forms. This pattern of LCA expression in different lymphocytes also occurs in mice (10–12), rats (13, 14) and in chickens (15).

The primary structures of rat and mouse LCA molecules have been predicted based on the nucleotide sequences of cDNA clones (16, 17). The amino acid sequence suggests that the mouse LCA is an integral membrane protein with a

402-amino-acid extracellular domain, a 22-amino-acid membrane-spanning peptide, and a large, 705-amino-acid cytoplasmic domain.

The precise molecular basis for the generation of the different members of the LCA family is not understood. Differences in the extent of glycosylation have been suggested to account for the various LCA members (18, 19). The possibility that LCAs are encoded by a multigene family has been rendered unlikely by a preliminary analysis of the mouse LCA gene structure (17). Northern blot analysis and S1 mapping experiments have shown the existence of three different size classes of mouse LCA mRNAs (16, 17, 20, 21).

To establish the genetic basis for the generation of the diversity of LCA molecules, it is essential to determine the primary structures of different LCA forms and to correlate the protein structures to the gene structure. In this paper, we present the complete amino acid sequence of a member of the human LCA family deduced from cDNA sequences, and evidence that there are at least five, and potentially eight different LCA mRNAs that are generated by the differential usage of three exons of a single human LCA gene.

Materials and Methods

Molecular Cloning and Analysis. The constructions of the cDNA libraries derived from poly(A)$^+$ RNAs isolated from pooled human tonsils and from the SB cell line, and the genomic DNA library derived from human placental DNA have been described (22, 23). These libraries were screened essentially as described by Benton and Davis (24), with minor modifications. The tonsil cDNA library was hybridized to the nick-translated ^{32}P-labeled (25) 3.2 kb Xba I fragment isolated from the mouse LCA cDNA clone pLY-5–68 (17), in the presence of 4× SSC, 50% (vol/vol) formamide, and 10% (wt/vol) sodium dextran sulfate (26) at 28°C. The filters were washed at 44°C in 0.1× SSC, 0.1% SDS. The SB cDNA library was probed with the cDNA insert of LCA.6, and the human placenta DNA library was screened with probe 1 (see Fig. 1) in the presence of 4× SSC and 10% (wt/vol) sodium dextran sulfate at 65°C. Filters were washed at 65°C in 0.2× SSC, 0.1% SDS. Phage DNAs were prepared and analyzed by restriction mapping using the methods described by Maniatis et al. (27) and by Southern blot analysis (28). DNA fragments were subcloned into the plasmid vector pSP65 (29) before being sequenced according to the method of Maxam and Gilbert (30) using the strategies outlined in Figs. 1 and 6.

Northern Blot Hybridization. Preparation of poly(A)$^+$ RNA from cultured cell lines was according to Maniatis et al. (27). ~2 µg of poly(A)$^+$ RNA was denatured with glyoxal (31), and fractionated by electrophoresis through a 0.8% agarose gel in 10 mM sodium phosphate buffer (pH 6.5), and the RNA was transferred to a nitrocellulose filter by blotting. The filter was hybridized to various nick-translated probes in the presence of 4× SSC, 50% (vol/vol) formamide, and 10% (wt/vol) sodium dextran sulfate at 42°C, and washed at 65°C in 0.2× SSC and 0.1% SDS. After each round of hybridization, the filter was boiled in H$_2$O for 5 min to remove the previously hybridized probe and exposed to x-ray film to confirm that there was no remaining radioactivity bound.

Results

Complete Amino Acid Sequence of Human LCA Protein Deduced from cDNA Sequences. To isolate human cDNA clones, a λ gt11 cDNA library (22) was screened with the 3.2 kb Xba I–Xba I fragment of the mouse LCA cDNA clone pLy-5-68 (17), which contains most of the protein-encoding sequences. The inserts of 18 independently isolated cDNA clones were subcloned into the plasmid vector pSP65 (29) and restriction maps were constructed. The restriction
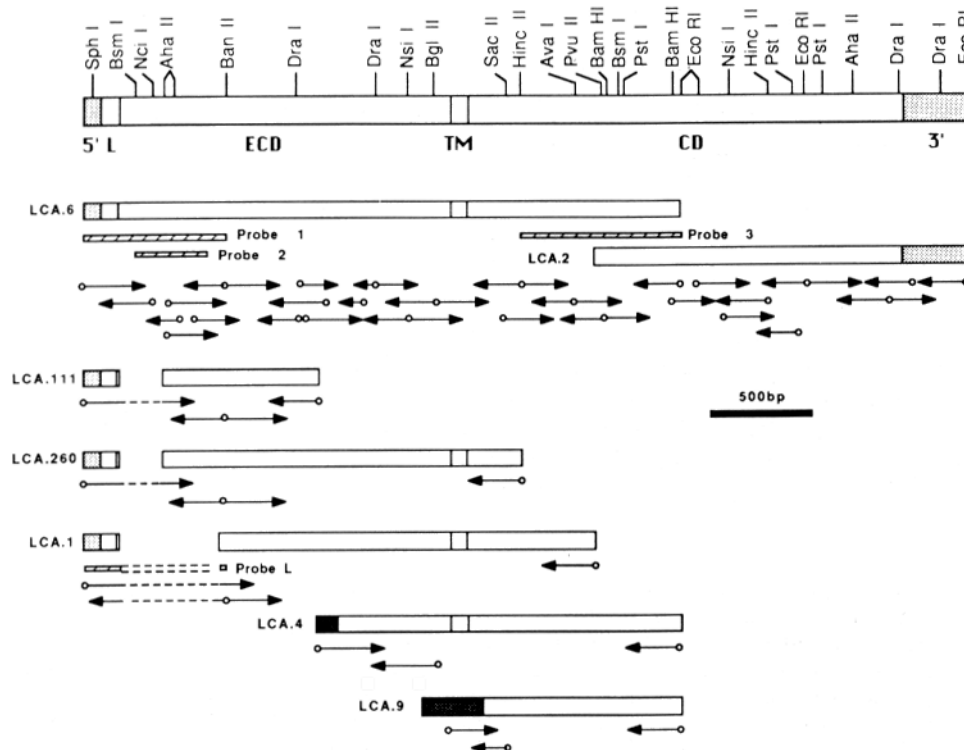
FIGURE 1. Restriction maps and nucleotide sequence strategy of human LCA cDNAs. A summary of the restriction maps of LCA cDNAs as well as the protein structure of LCA.6/2 is shown at the top of the figure. The open boxes marked L (leader peptide), ECD (extracellular domain), TM (transmembrane peptide), and CD (cytoplasmic domain) schematically represent the domains encoded in the open reading frame defined by the two overlapping cDNA clones LCA.6 and LCA.2. The stippled areas, marked 5' and 3', represent the 5'- and 3'-untranslated regions. Various LCA cDNA clones named LCA.6, etc. are shown below and are aligned with the restriction map to indicate the sizes and relative location of each cDNA. The arrows indicate the direction and extent of the nucleotide sequences determined according to Maxam and Gilbert (30). Open circles indicate ^{32}P-labeling at the 5' end. Dashed lines represent gaps. Shaded areas represent sequences that are entirely different from the other LCA cDNA sequences. The hatched bars indicate hybridization probes used for screening of cDNA and genomic DNA libraries, and for Northern blot hybridization.

maps of the relevant cDNA subclones are shown in Fig. 1. By comparing the restriction maps, it was found that two overlapping clones, namely LCA.6 and LCA.2, would together span ~4.3 kb. Therefore, the nucleotide sequences of LCA.6 and LCA.2 were determined according to the method of Maxam and Gilbert (30) using the strategy shown in Fig. 1. The two cDNA clones share the identical sequences in the overlapping region. The combined nucleotide sequence of these two cDNAs, which is 4,315 bp long, is shown in Fig. 2. This cDNA lacks the 3' end of the sequence, since it does not contain a poly(A) stretch or the poly(A) attachment signal sequence (AATAAA), (32). Indeed, the size of mRNA detected by Northern blot analysis is 5.0–5.6 kb (see below). Nevertheless, the combined cDNA sequence (hereafter referred to as LCA.6/2) contained all the information necessary to deduce the amino acid sequence of a human

LCA protein. The reading frame shown in Fig. 2 is the only one that yields an amino acid sequence long enough to encode a LCA protein. As seen with the mouse LCA cDNA sequence, the human cDNA sequence also has two in-frame methionine codons (ATG) separated by three nucleotides near the beginning of the reading frame. We believe that the second ATG is the initiation codon for translation, since the second ATG conforms better to the proposed translation initiation consensus sequence, ACCATGG (33). Assuming this assignment of the initiation codon, the human LCA.6/2 encodes a protein of 1,304 amino acids.

Using the hydropathy analysis of Kyte and Doolittle (34), two stretches of strongly hydrophobic amino acids (indicated in Fig. 2 by underlining) can be identified. The first hydrophobic stretch of 23 amino acids at the beginning of the coding sequence is most likely the signal peptide necessary for the transfer of the protein across the cellular membrane. The calculation according to the algorithm of von Heijne (35) supports this conclusion, and predicts the NH_2 terminus of the mature protein to be the glutamine residue assigned amino acid position 1 in Fig. 2, although this assignment must be considered tentative. The second highly hydrophobic sequence (amino acid positions 553–574) is 22 amino acids long and is most likely the transmembrane peptide. The extracellular domain contains 17 potential N-linked glycosylation sites (N-X-S or N-X-T), which are indicated in Fig. 2 by short underlinings. In summary, the mature protein deduced from these two overlapping cDNA sequences has a 552-amino-acid extracellular domain, a 22-amino-acid transmembrane peptide, and a 707-amino-acid cytoplasmic domain.

Comparison of Human and Mouse LCA Amino Acid Sequences. The human LCA.6/2 amino acid sequence and a complete mouse LCA amino acid sequence (17) were compared to find the degree of conservation between these sequences. The homology between the two sequences is not uniform. The putative signal and transmembrane peptides of the two species are very well conserved (91% homologous), and the cytoplasmic sequences are ~84% homologous. In marked contrast, the extracellular portions of the two species have diverged to an extent that a unique alignment of two sequences is difficult. Another notable feature of the extracellular domain is that the human sequence has an insertion of 161 amino acids near the NH_2-terminus compared with the mouse sequence. The nature of this insertion will be discussed in the following sections. Even if the large insertion is excluded, the homology of the extracellular domains of the two species is only ~39%. In spite of the extensive divergence, there are some conserved features in the extracellular portion. In particular, all of the 16 cysteine residues in the extracellular domain of the human sequence are found in corresponding positions in the mouse sequence (there are two extra cysteine residues in the mouse sequence that have no counterpart in the human sequence). Among the 18 tyrosine residues in the extracellular domain of the human sequence, 13 are conserved in the mouse sequence. Perhaps this conservation of cysteine and tyrosine positions indicates that some higher-order structure, but not the exact primary sequence, is important for the (unknown) function of the extracellular portion of human and mouse LCA proteins.

LCA cDNA Clones with Different Structures Near the 5' End. Two additional cDNA clones isolated from the tonsil cDNA library (LCA.260 and LCA.1) have

GGAAATTGTT CCTCGTCTGA TAAGACAACA GTGGAGAAAG GACGCATGCT GTTTCTTAGG GACACGGGTG GCTTCCAGAT ATGACC

FIGURE 2. Nucleotide sequence of the LCA.6/2 cDNAs and the deduced LCA amino acid sequence. The numbers shown above the amino acid residue positions. The putative signal and quence designate amino acid residue positions. The putative signal and transmembrane peptides are indicated by underlining. The potential N-linked glycosylation sites in the extracellular domain are shown by short underlining. Amino acids are presented by the standard single-letter code. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00638.

very similar restriction maps to that of LCA.6 (Fig. 1). However, it was noted that two restriction sites (Bsm I and Nci I) that exist in LCA.6 were missing from the other cDNA clones, even though all three cDNA clones share the Sph I site located 5' to these sites. Therefore, the nucleotide sequences of the 5' portions of these cDNA clones were determined and compared with the LCA.6 sequence (Fig. 3). These three cDNAs do indeed share the 5'-untranslated sequence, as well as the sequence encoding the signal peptide and the first eight amino acids of the putative mature proteins. The 5'-untranslated sequences of the three clones are slightly different from each other, probably due to the natural occurrence of sequence polymorphism (the cDNA library was constructed from a pool of several tonsils). More important, however, is the finding that the LCA.260 and LCA.1 sequences have extensive deletions relative to the LCA.6 sequence. The deletions start at the same 5' position, but the 3' endpoints of the deletions are different (Fig. 3). Because the number of bases deleted is a multiple of three (198 bp for LCA.260 and 483 bp for LCA.1), the reading frame is not affected by these deletions. As a result, the proteins encoded by LCA.260 and LCA.1 are 66 and 161 amino acids shorter than the protein encoded by LCA.6/2, respectively. This interpretation assumes that the missing 3' sequences of LCA.260 and LCA.1 are the same as LCA.6/2. Another human LCA cDNA clone (LCA.111), which was isolated from a cDNA library derived from mRNA of a human B cell line SB was also characterized by restriction mapping and partial nucleotide sequence determination (Figs. 1 and 3). The 5' sequence of LCA.111 is the same as that of LCA.260, except for a few differences ascribable to polymorphism. A comparison of these shorter protein sequences with the mouse protein sequence revealed that the mouse cDNA sequence pLy-5-R4 (17) corresponds to the human LCA.1 cDNA sequence. The three distinct structures defined by the clones LCA.6, LCA.260/LCA.111, and LCA.1 demonstrate that there are at least three different classes of human LCA mRNAs.

Two additional cDNA clones isolated from the tonsil library (LCA.4 and LCA.9) have similar restriction maps to that of LCA.6 (Fig. 1). The determination of the partial nucleotide sequences of these cDNA clones confirmed that the 3' sequences are identical to that of LCA.6. However, based on the nucleotide sequence data, these cDNAs seem to have derived from either aberrantly or incompletely spliced LCA mRNAs (data not shown).

Isolation and Characterization of Genomic DNA Encoding Variable Portion of Human LCAs. To define the genetic basis of the diversity of LCA mRNAs, a genomic DNA clone (LCA.204) was isolated from a human placental DNA library (23) using the Eco RI–Ban II fragment of LCA.6 as the hybridization probe (probe 1 shown in Fig. 1). The restriction map of the 13 kb LCA.204 insert is shown in Fig. 4A. Using the combination of restriction mapping and Southern blot hybridizations (28) with the insert of LCA.6 as a probe, we identified and located six exons in this cloned genomic DNA. During the subsequent sequence analysis of these exons, however, it became clear that a small exon of 27 bp existed in the LCA gene. This exon was located by blot hybridization using a synthetic oligonucleotide predicted from the cDNA sequence. The locations of the seven exons, as well as the strategy for determining the nucleotide sequence, are indicated in Fig. 4B, and the nucleotide sequences
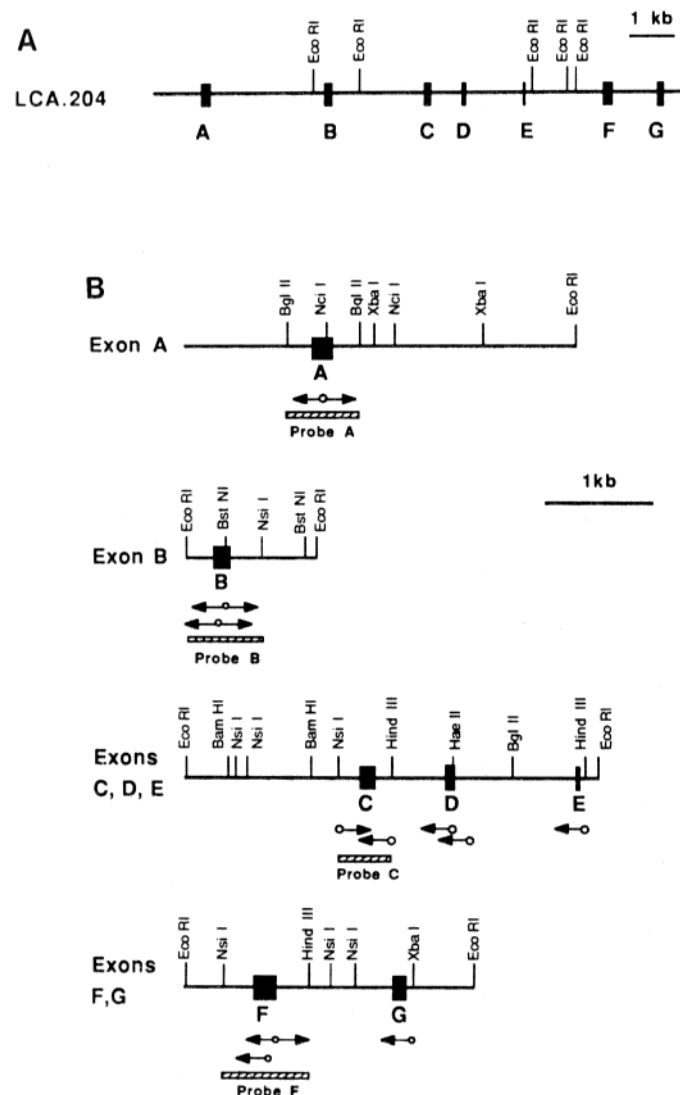
FIGURE 3. Comparison of the different LCA cDNA structures. The 5′ sequences of four cDNA clones, LCA.6, LCA.111, LCA.260, and LCA.1 are compared. The dashes indicate identical nucleotide. The numbers refer to the amino acid positions in the LCA.6/2 sequence.

FIGURE 4. Restriction map and sequence strategy of a human LCA genomic DNA segment. A. The exon-intron organization in a 13 kb genomic fragment of the human LCA gene is schematically represented. Exons, designated A–G, are indicated by solid boxes. B. The restriction maps of four Eco RI fragments containing LCA exons are shown. The sequencing strategy for the seven exons are indicated. Other symbols are the same as in Fig. 1.

of the seven exons are shown in Fig. 5. Each of the seven exons is flanked at its 5' and 3' ends by sequences that conform to the consensus splice acceptor and donor signal sequences (36). Interestingly, all the splice junctions of the seven exons occur between the first and second bases in a triplet codon.

Because the complete exon-intron organization of the human LCA gene is not yet known, we can not assign definitive exon numbers to each exon. In the following sections, we will tentatively call these exons A–G (Fig. 4). Although we

Exon A

Exon B

Exon C

Exon D

Exon E

Exon F

Exon G

FIGURE 5. Nucleotide sequences of human LCA exons *A–G*. The nucleotide shown. The numbers refer to the amino acid positions in the LCA.6/2 cDNA sequences of each exon together with the 5' and 3' flanking sequences are sequence.

do not know whether there are one or more exons encoding the 5'-untranslated and leader peptide, we will call the putative 5'-exon(s) L.

The comparison of the exon sequences with the cDNA sequences clearly indicates that the deleted sequence in LCA.260 and LCA.111 corresponded precisely to the exon A sequence, while the deletion in LCA.1 corresponds to the sequences of exons A, B, and C. The three different mRNA structures can be described as LABCDE . . . (LCA.6), LBCDE . . . (LCA.260 and LCA.111) and LDE . . . (LCA.1). The comparison of the LCA cDNAs and gene sequences formally proves that the variable LCA mRNA structures are generated by differential usage of exons.

Analysis of LCA mRNA Expression in Human Lymphocyte Cell Lines. To determine if the three LCA cDNA structures could account for all the LCA mRNAs expressed in various T and B lymphocytes, Northern blot analysis (31) was carried out. To make the comparison between several different probes meaningful, one RNA blot filter containing mRNA isolated from three T cell lines (Molt3, HSB2, and Hut78) and four B cell lines (Raji, SB, Daudi, and Namalwa) was recycled for this analysis. Hybridizations using either a cDNA probe corresponding to the 5' variable region (probe 2, Fig. 1) or a probe corresponding to the cytoplasmic region (probe 3, Fig. 1) resulted in essentially the same hybridization pattern, except that the degradation of mRNA is more prominent with the cytoplasmic region probe (Fig. 6,A and B). More important, however, is the observation that the LCA mRNA sizes, which range between 5.0 and 5.6 kb, are variable in different cell lines, and that there appear to be multiple size species of mRNA expressed by a single cell line. For example, Raji cells seem to express at least three different-size mRNAs, while Molt3 cells express at least two.

Because the analysis of cDNAs and gene structure has demonstrated that different-size mRNAs could be produced by differential usage of three exons (A, B, and C), the expression of individual exons in the lymphocyte cell lines was examined by using exon-specific probes. As the leader exon probe, we used the Eco RI–Ban II fragment derived from LCA.1 (probe L, see Fig. 1), which contains 174 bp from the putative leader exon sequence, and 35 bp from exon D. Other probes specific to exons A, B, C, or F were made from the genomic DNA clone (probes A, B, C, and F, See Fig. 4B). The hybridization patterns of the leader (L) exon probe and exon F probe are not significantly different from that of the cytoplasmic region probe (Fig. 6B and C). This result is consistent with the idea that the leader exon, exon F, and the cytoplasmic exons are included in any LCA mRNA. However, when the exon A probe is used, the pattern of hybridization is substantially different. The T cell line Molt3 and the B cell line Raji strongly express exon A, the B cell lines Namalwa and Daudi weakly express exon A–containing mRNA, while the B cell line SB and the T

FIGURE 6. Northern blot analysis of T and B lymphocyte mRNAs with LCA probes. Northern blot analysis of poly(A)+ RNA isolated from the T cell lines Molt3, HSB-2, and Hut78, and the B cell lines Raji, SB, Daudi, and Namalwa was done as described by Thomas (31). 2 µg of each RNA was used per lane. Ribosomal RNA was run in parallel as a size marker. Hybridization probes were: A, the cDNA probe 2 (see Fig. 1); B, the cDNA probe 3 (Fig. 1); C, the exon-specific probes, L, A, B, C, and F depicted in Fig. 5. One Northern blot filter was reused for all seven hybridizations. After each hybridization the probes were removed by boiling the filter in water for 5 min. A and B: Autoradiography was without an intensifying screen for 5 d. C: Autoradiography was with an intensifying screen for 2–5 d.
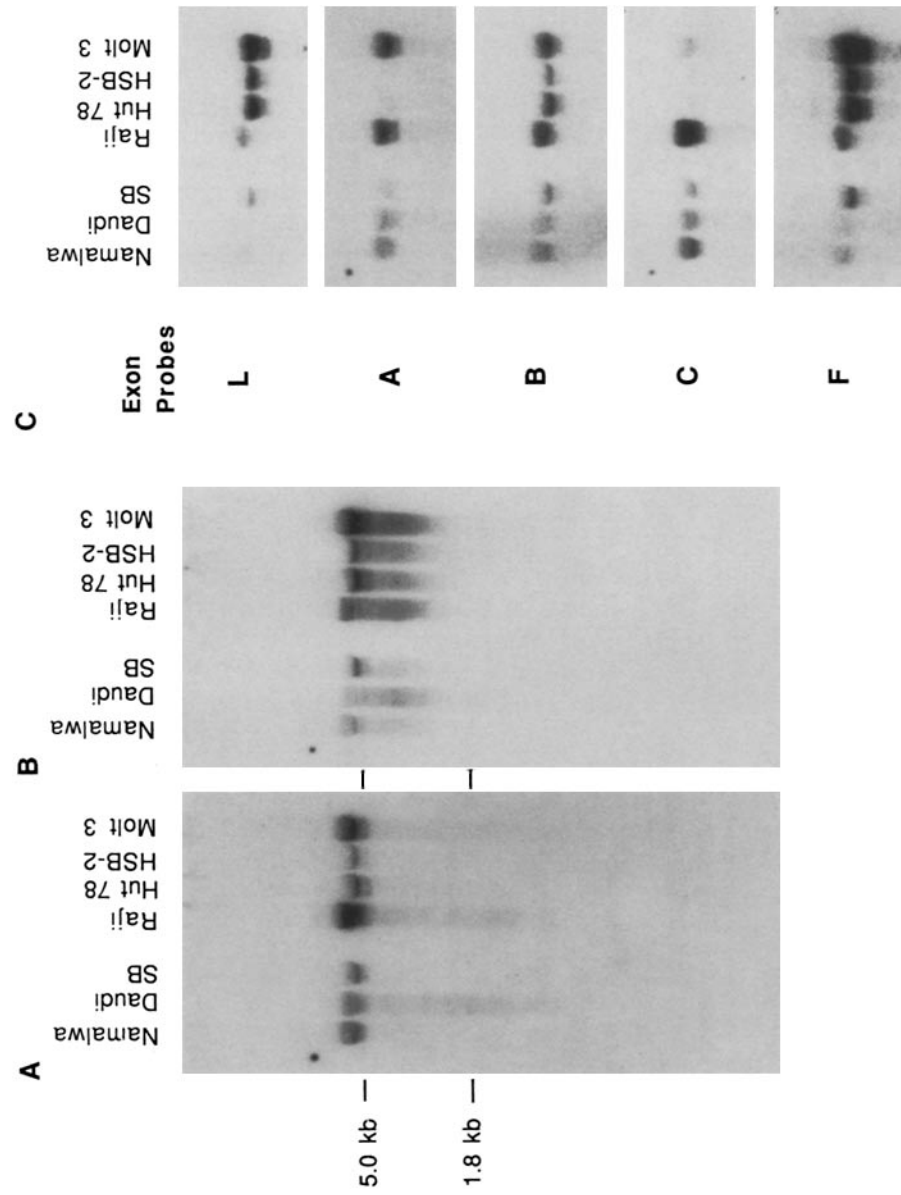
cell line HSB2 and Hut78 do not express this exon (Fig. 6C). Exon B is expressed in all of the lymphocyte cell lines tested (Fig. 6C). Exon C is expressed very strongly in Raji cells, only weakly in the other three B cells, and is barely expressed in the three T cell lines.

The interpretation of these results is not straightforward, since each cell line may produce multiple forms of LCA mRNA. Nevertheless, some tentative conclusions can be deduced. The most abundant mRNA expressed in the B cell line Raji probably contains all three exons and has the structure corresponding to the cDNA clone LCA.6 (LABCDE. . .). This conclusion is consistent with Raji cells having the larger-size mRNA as the predominant species (Fig. 6). Daudi and Namalwa cells seem to have a pattern similar to the Raji cells, except that the overall levels of expression are much lower. On the other hand, the most abundant LCA mRNA detected in the B cell line SB contains exons B and C, but not A. This mRNA species is consistent with the structure of a cDNA clone, LCA.111 (LBCDE. . .) isolated from the same cell line. The predominant mRNA species of the T cell lines HSB2 and Hut78 contain only exon B and not exons A or C (Fig. 6C). Therefore, it is necessary to assume the existence of a mRNA with the structure of LBDE. . . . The other T cell line, Molt3, expresses exon A and B equally well, but the expression of exon C is very weak (Fig. 6C). This can be interpreted as the expression of mRNA having the LABDE . . . structure, or alternatively it is possible that the cell expresses similar amounts of two different mRNAs or LADE . . . and LBDE . . . structures. Whatever the case may be, one also has to assume yet another structure of LCA mRNA (LABDE . . . or LADE . . .). The existence of a LCA mRNA species without any of the three exons, A, B, and C, is predicted from the cDNA clone LCA.1, but it is difficult to assess the expression of this mRNA species by Northern blot analysis. However, the lowest-molecular-mass bands visible in the three T cell lines (Molt3, HSB2, and Hut78) using either the leader exon or exon F probes (Fig. 6C) may correspond to this LCA mRNA species.

In conclusion, Northern analysis supports the existence of three different forms of human LCA mRNA predicted from the isolated cDNA clones. Furthermore, the existence of at least two more species of mRNA was demonstrated. Fig. 7 summarizes the structures of human LCA mRNAs.

Discussion

Although an accumulated body of evidence has made it plausible that the different forms of the LCA molecules are encoded by a single gene, the exact mechanism of the generation of diversity was not known. In this paper, it was demonstrated that a single human LCA gene can generate multiple forms of mRNAs that encode proteins of different structures. The isolation of the LCA cDNA clones and Northern blot analysis demonstrate the existence of at least five distinct LCA mRNAs. Furthermore, it was shown that these different mRNAs result from differential usage of three exons by alternative splicing. Each exon can be either included or excluded from a LCA mRNA independently of the other two exons, and without disturbing the coding frame. Therefore, it is possible that the number of different forms of the LCA mRNAs and proteins is as large as eight (2^3). This calculation, of course, assumes that only exons A,
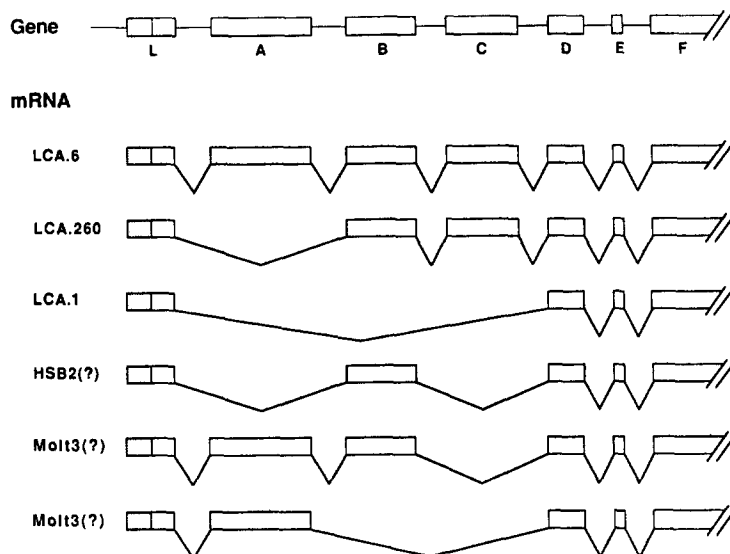
FIGURE 7. Summary of the differential exon usage of the human LCA gene. The LCA gene and the possible LCA mRNA products are schematically shown with boxes representing exons. The mRNA structures LCA.6, LCA.260, and LCA.1 are derived from cDNA cloning, and the existence of HSB2(?) and one or the other of Molt3(?) mRNA structures are demonstrated by Northern blot data (for details, see text). Not drawn to scale.

B, and C are used differentially and that the LCA gene does not contain exons for which we have not found any corresponding cDNAs. So far there is no convincing evidence that any other exon, particularly exon D, might be used differentially. Therefore, at the moment, the best estimate of the diversity of the human LCA mRNAs and proteins is between five and eight.

The generation of multiple structures of mRNAs and proteins by differential splicing is not uncommon among eukaryotic genes (for a review, see reference 37). For example, the troponin T gene can produce at least 10 different mRNAs and protein structures (38). Other examples, among integral membrane proteins, are the N-CAM (39) and Lyt-2 (40) genes, which produce two alternative forms of mRNAs. The alternative splicing of these genes encodes proteins with differences in the cytoplasmic domains. The LCA gene is, however, unusual in that the diversity occurs in the extracellular domain, that the degree of the diversity is large, and that the pattern of exon usage is not stochastic.

Because of the similar sizes of the three variable exons, together with the very large sizes of the LCA mRNAs and proteins, it will be very difficult to distinguish some mRNAs or their products from each other solely based on their electrophoretic mobilities. For example, mRNAs or proteins of the structure LABDE . . ., LBCDE . . ., and LACDE . . . will most likely be indistinguishable. This is probably a reason that the true complexity of the LCA system is often overlooked. Of course, it is possible that a regulatory mechanism of LCA mRNA splicing somehow precludes one or more of the potential combinations of these three exons. The evaluation of the complexity of the LCA gene system in a more quantitative manner will require an exhaustive series of experimentation using S1 nuclease protection, primer extension, and blot hybridization with various

probes, and perhaps more cDNA cloning. The knowledge of the general structure of the LCA mRNAs, as well as the structure of variable exons, that are reported in this paper, will make this type of approach feasible. Another useful tool to study the complex expression of various members of the LCA family at the protein level will be mAbs directed against epitopes expressed by the individual variable exons. Some mAbs with these properties are already available (e.g., anti-2H4 [4]). However, even such mAbs may not be able to distinguish two or more different LCA proteins that share the exon to which they are specific. For example, anti–exon B antibodies would react to both LABDE . . . and LBCDE . . . structures, which may not be distinguishable by SDS-PAGE analysis. For this reason, antibodies specific to the joints of two segments will be particularly powerful since they will have a very limited specificity.

The complete amino acid sequence of one member of the human LCA family has been determined based on the nucleotide sequence of LCA cDNA clones, and the amino acid sequence of two other LCA family members is inferred based on incomplete cDNA structures. The differences in the sizes of the primary structures (1,304, 1,238, and 1,143 amino acids) as well as the variation in the number of potential N-linked glycosylation sites in the extracellular domain (17, 14, and 11 sites, respectively) are consistent with the observed range in molecular masses ($\sim$170–240 kD) of the human LCA members. Furthermore, the variability of the amino acid sequences in the NH_2-proximal region of these three LCA members is compatible with studies, using mAbs that have suggested that the LCAs have common and variable epitopes (4, 6, 12). The human LCA amino acid sequences have no significant homology with any known protein sequence (except to the rat and mouse LCAs) in the database of the Protein Identification Resource. There is, however, an internal sequence homology of amino acid numbers 588–878 and 879–1,194 located in the cytoplasmic domain. A similar homology was previously shown for the rat LCA sequence (16).

The rate of amino acid change in the various domains of the LCA is uneven. While the signal peptide, transmembrane peptide, and cytoplasmic domain are well conserved (91, 91, and 84% amino acid homology between mouse and human), the extracellular domain is poorly conserved ($\sim$39% homology). However, the conservation of cysteine residues in the extracellular domain suggests that higher-order structure may be maintained. Because there is no sequence information of the rodent counterparts of the human LCA exons A, B, and C, the possibility that these exons are more homologous remains. Nevertheless, the degree of the conservation of the LCA extracellular domain sequences appears to be much less than most known membrane proteins. Another example of a membrane glycoprotein that has uneven rates of amino acid changes is the T8 (Lyt-2) molecule, which has $\sim$42% homology between the human and mouse extracellular V-like domains (41, 42). In this case, it has been suggested that the high degree of sequence divergence may reflect coevolution of the T8 protein with the MHC class I antigens, which are believed to be the ligands of T8. The elucidation of the possible ligand(s) and the significance of the uneven evolution of the LCA domains require further study.

A unique feature of LCA proteins that distinguishes them from most known membrane glycoproteins is that both the extracellular domain (either 552, 486,

or 391 amino acids) and the cytoplasmic domain (707 amino acids) are very large. Among the few membrane proteins that have an extracellular and a cytoplasmic domain comparable in size to the LCA domains are the receptors for epidermal growth factor (EGF) (43) and platelet-derived growth factor (44), the product of the protooncogene c-*fms* (45), and the oncogene product neu (46, 47). Although the LCA proteins have no primary sequence homology to these growth factor receptors and epidermal growth factor receptor–related proteins, the similarity in overall architecture, as well as the finding that LCAs are phosphoproteins (48, 49) might suggest that LCAs also function as cell-surface receptors. If the LCAs are indeed receptors, the fact that potentially five to eight distinct extracellular forms of the LCA molecule exist might suggest that there could be several different ligands.

The expression of the different LCA forms appears to correlate with the function of cells. For example, T4$^+$ cells that induce suppressor activity express LCAs with the 2H4 epitope, while T4$^+$ cells that have helper activity do not express the 2H4 epitope. The precise localization of the 2H4 epitope is not yet known, although it is within the variable portion encoded by the exons A, B, and C of the LCA gene (our unpublished results). The molecular basis of this structure-function relationship might be analogous to that of T4 molecules. The biological function of the T4 molecules is, among other things, to facilitate the interaction of T4$^+$ cells with antigen-presenting cells, which bear MHC class II molecules. Therefore, it is plausible that the function of LCAs is to facilitate the interactions of lymphocytes among themselves and with other cells of the immune system. Furthermore, different LCA forms might specify the way(s) the cells bearing those forms interact with other cells.

Since the completion of this work, papers by Ralph et al. (50) and Barclay et al. (51) appeared that drew similar conclusions based on cDNA sequences concerning the sequences of human LCAs, and the basis of heterogeneity in human and rat LCAs.

Summary

Leukocyte common antigens (LCAs, also known as T200 and CD 45) are integral membrane proteins expressed exclusively on hematopoietic cells. These molecules exhibit varying molecular masses and epitopes when expressed in different cell types. To determine the genetic bases for the generation of this diversity, three classes of human LCA cDNA clones that are different near their 5' ends have been isolated. These differences arose as a result of differential usage of three exons as determined from an analysis of a genomic DNA clone. Furthermore, Northern blot analysis with LCA exon–specific probes demonstrates the existence of at least two more LCA mRNA forms that are generated by differential splicing. A comparison of the human and mouse LCA protein sequences revealed a marked difference only in the extracellular domain.

References

1. Reinherz, E. L., and S. F. Schlossman. 1980. The differentiation and function of human T lymphocytes. *Cell.* 19:821.
2. Morimoto, C., N. L. Letvin, J. A. Distaso, W. R. Aldrich, and S. F. Schlossman. 1985. The isolation and characterization of the human suppressor inducer T cell subset. *J. Immunol.* 134:1508.
3. Takeuchi, T., C. E. Rudd, S. F. Schlossman, and C. Morimoto. 1987. Induction of suppression following autologous mixed lymphocyte reaction: role of a novel 2H4 antigen. *Eur. J. Immunol.* 17:97.
4. Rudd, C. E., C. Morimoto, L. L. Wong, and S. F. Schlossman. 1987. The structure of the 2H4 antigen: A subset of the LCA/T200 family of antigens involved in immune suppression. *In* Leukocyte Typing III. A. J. McMichael, editor. Oxford University Press, Oxford, United Kingdom. 242.
5. Omary, M. B., I. S. Trowbridge, and H. A. Battifora. 1980. Human homologue of murine T200 glycoprotein. *J. Exp. Med.* 152:842.
6. Dalchau, R., and J. W. Fabre. 1981. Identification with a monoclonal antibody of a predominantly B lymphocyte-specific determinant of the human leukocyte common antigen: evidence for structural and possible functional diversity of the human leukocyte common molecule. *J. Exp. Med.* 153:753.
7. Berger, A. E., J. E. Davis, and P. Cresswell. 1981. A human leukocyte antigen identified by a monoclonal antibody. *Hum. Immunol.* 3:231.
8. Ledbetter, J. A., L. M. Rose, C. E. Spooner, P. G. Beatty, P. J. Martin, and E. A. Clark. 1985. Antibodies to common leukocyte antigen p220 influence human T cell proliferation by modifying IL-2 receptor expression. *J. Immunol.* 135:1819.
9. Nieminen, P., and E. Saksela. 1986. NK-9, a distinct sialylated antigen of the T200 family. *Eur. J. Immunol.* 16:513.
10. Tung, J.-S., M. P. Scheid, M. A. Pierotti, U. Hammerling, and E. A. Boyse. 1981. Structural features and selective expression of three Ly-5$^+$ cell-surface molecules. *Immunogenetics.* 14:101.
11. Sarmiento, M., M. R. Loken, I. S. Trowbridge, R. L. Coffman, and F. W. Fitch. 1982. High molecular weight lymphocyte surface proteins are structurally related and are expressed on different cell populations at different times during lymphocyte maturation and differentiation. *J. Immunol.* 128:1676.
12. Lefrancois, L., and M. J. Bevan. 1985. Novel antigenic determinants of the T200 glycoprotein expressed preferentially by activated cytotoxic T lymphocytes. *J. Immunol.* 135:374.
13. Woollett, G. R., A. N. Barclay, M. Puklavec, and A. F. Williams. 1985. Molecular and antigenic heterogeneity of the rat leukocyte-common antigen from thymocytes and T and B lymphocytes. *Eur. J. Immunol.* 15:168.
14. Spickett, G. P., M. R. Brandon, D. W. Mason, A. F. Williams, and G. R. Woollett. 1983. MRC OX-22, a monoclonal antibody that labels a new subset of T lymphocytes and reacts with the high molecular weight form of the leukocyte-common antigen. *J. Exp. Med.* 158:795.
15. Houssaint, E., S. Tobin, J. Cihak, and U. Losch. 1987. A chicken leukocyte common antigen: biochemical characterization and ontogenetic study. *Eur. J. Immunol.* 17:287.
16. Thomas, M. L., A. N. Barclay, J. Gagnon, and A. F. Williams. 1985. Evidence from cDNA clones that the rat leukocyte-common antigen (T200) spans the lipid bilayer and contains a cytoplasmic domain of 80,000 M_r. *Cell.* 41:83.
17. Saga, Y., J.-S. Tung, F.-W. Shen, and E. A. Boyse. 1986. Sequences of Ly-5 cDNA: Isoform-related diversity of Ly-5 mRNA. *Proc. Natl. Acad. Sci. USA.* 83:6940.
18. Morishima, Y., S. Ogata, N. H. Collins, B. Dupont, and K. O. Lloyd. 1982. Carbohydrate differences in human high molecular weight antigens of B- and T-cell lines. *Immunogenetics.* 15:529.

19. Childs, R. A., R. Dalchau, P. Scudder, E. F. Hounsell, J. W. Fabre, and T. Feizi. 1983. Evidence for the occurrence of O-glycosidically linked oligosaccharides of poly-N-acetyllactosamine type on the human leucocyte common antigen. *Biochem. Biophys. Res. Commun.* 110:424.

20. Lefrancois, L., M. L. Thomas, M. J. Bevan, and I. S. Trowbridge. 1986. Different classes of T lymphocytes have different mRNAs for the leukocyte-common antigen, T200. *J. Exp. Med.* 163:1337.

21. Raschke, W. C. 1987. Cloned murine T200 (Ly-5) cDNA reveals multiple transcripts within B- and T-lymphocyte lineages. *Proc. Natl. Acad. Sci. USA.* 84:161.

22. Weis, J. J., D. T. Fearon, L. B. Klickstein, W. W. Wong, S. A. Richards, A.deB. Kops, J. A. Smith, and J. H. Weis. 1986. Identification of a partial cDNA clone for the C3d/Epstein-Barr virus receptor of human B lymphocytes: homology with the receptor for fragments C3b and C4b of the third and fourth components of complement. *Proc. Natl. Acad. Sci. USA.* 83:5639.

23. Lawn, R. M., E. F. Fritsch, R. C. Parker, G. Blake, and T. Maniatis. 1978. The isolation and characterization of linked delta- and beta-globin genes from a cloned library of human DNA. *Cell.* 15:1157.

24. Benton, W. D., and R. W. Davis. 1977. Screening lambda gt recombinant clones by hybridization to single plaques in situ. *Science (Wash. DC).* 196:180.

25. Rigby, P. W., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polyermase I. *J. Mol. Biol.* 113:237.

26. Wahl, G. M., M. Stern, and G. R. Stark. 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization using dextran sulphate. *Proc. Natl. Acad. Sci. USA.* 76:3683.

27. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 1–545.

28. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503.

29. Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12:7035.

30. Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499.

31. Thomas, P. S. 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. *Methods Enzymol.* 100:255.

32. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding region sequence in eukaryotic messenger RNA. *Nature (Lond.).* 263:211.

33. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell.* 44:283.

34. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105.

35. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683.

36. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459.

37. Leff, S. E., M. G. Rosenfeld, and R. M. Evans. 1986. Complex transcriptional units: Diversity in gene expression by alternative RNA processing. *Ann. Rev. Biochem.* 55:1091.

38. Breitbart, R. E., H. T. Nguyen, R. M. Medford, A. T. Destree, V. Mahdavi, and B. Nadal-Ginard. 1985. Intricate combinational patterns of exon splicing generate multiple regulated troponin T isoforms from a single gene. *Cell.* 41:67.

39. Hemperly, J. J., B. A. Murray, G. M. Edelman, and B. A. Cunningham. 1986.

Sequence of a cDNA clone encoding the polysialic acid–rich and cytoplasmic domains of the neural cell adhesion molecule N-CAM. *Proc. Natl. Acad. Sci. USA.* 83:3037.

40. Zamoyska, R., A. C. Vollmer, K. C. Sizer, C. W. Liaw, and J. R. Parnes. 1985. Two Lyt-2 polypeptides arise from a single gene by alternative splicing patterns of mRNA. *Cell.* 43:153.

41. Nakauchi, H., G. P. Nolan, C. Hsu, H. S. Huang, P. Kavathas, and L. A. Herzenberg. 1985. Molecular cloning of Lyt-2, a membrane glycoprotein marking a subset of mouse T lymphocytes: Molecular homology to its human counterpart, Leu-2/T8, and to immunoglobulin variable regions. *Proc. Natl. Acad. Sci. USA.* 82:5126.

42. Littman, D. R. 1987. The structure of the CD4 and CD8 genes. *Ann. Rev. Immunol.* 5:561.

43. Ullrich, A., L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield, and P. H. Seeburg. 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature (Lond.).* 309:418.

44. Yarden, Y., J. A. Escobedo, W.-J. Kuang, T. L. Yang-Feng, T. O. Daniel, P. M. Tremble, E. Y. Chen, M. E. Ando, R. N. Harkins, U. Francke, V. A. Fried, A. Ullrich, and L. T. Williams. 1986. Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. *Nature (Lond.).* 323:226.

45. Coussens, L., C. van Beveren, D. Smith, E. Chen, R. L. Mitchell, C. M. Isacke, I. M. Verma, and A. Ullrich. 1986. Structural alteration of viral homologue of receptor proto-oncogene *fms* at the carboxyl terminus. *Nature (Lond.).* 320:277.

46. Bargmann, C. I., M.-C. Hung, and R. A. Weinberg. 1986. The *neu* oncogene encodes an epidermal growth factor receptor-related protein. *Nature (Lond.).* 319:226.

47. Yamamoto, T., S. Ikawa, T. Akiyama, K. Semba, N. Nomura, N. Miyajima, T. Saito, and K. Toyoshima. 1986. Similarity of protein encoded by the human c-*erb*-B-2 gene to epidermal growth factor receptor. *Nature (Lond.).* 319:230.

48. Omary, M. B., and I. S. Trowbridge. 1980. Disposition of T200 glycoprotein in the plasma membrane of a murine lymphoma cell line. *J. Immunol.* 255:1662.

49. Shackelford, D. A., and I. S. Trowbridge. 1986. Identification of lymphocyte integral membrane proteins as substrates for protein kinase C: phosphorylation of the interleukin-2 receptor, class I HLA antigens, and T200 glycoprotein. *J. Biol. Chem.* 261:8334.

50. Ralph, S. J., M. L. Thomas, C. C. Morton, and I. S. Trowbridge. 1987. Structural variants of human T200 glycoprotein (leukocyte-common antigen). *EMBO (Eur. Mol. Biol. Organ.) J.* 6:1251.

51. Barclay, A. N., D. I. Jackson, A. C. Willis, and A. F. Williams. 1987. Lymphocyte specific heterogeneity in the rat leucocyte common antigen (T200) is due to differences in polypeptide sequences near the NH₂-terminus. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:1259.