cDNA cloning and sequence of MAL, a hydrophobic protein associated with human T-cell differentiation

(subtractive hybridization/ontogeny/T-cell activation/transport proteins)

MIGUEL A. ALONSO* AND SHERMAN M. WEISSMAN

Department of Human Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510

Contributed by Sherman M. Weissman, December 5, 1986

ABSTRACT

We have isolated a human cDNA that is expressed in the intermediate and late stages of T-cell differentiation. The cDNA encodes a highly hydrophobic protein, termed MAL, that lacks a hydrophobic leader peptide sequence and contains four potential transmembrane domains separated by short hydrophilic segments. The predicted configuration of the MAL protein resembles the structure of integral proteins that form pores or channels in the plasma membrane and that are believed to act as transporters of water-soluble molecules and ions across the lipid bilayer. The presence of MAL mRNA in a panel of T-cell lines that express both the T-cell receptor and the T11 antigen suggests that MAL may be involved in membrane signaling in T cells activated via either T11 or T-cell receptor pathways.

Monoclonal antibodies against T cells have identified a number of surface molecules expressed during intrathymic ontogeny (1). This has allowed the definition of discrete stages of T-cell differentiation (2). The earliest identified T-lineage cells express the sheep erythrocyte receptor T11 (stage I). Later, thymocytes express T6, T4, and T8 antigens (stage II). With further maturation, T6 disappears, and thymocytes acquire the T3/T-cell receptor structure and ultimately appear in the periphery as either T4+T8- or T4-T8+ cells (stage III).

Hybridoma technology has defined several surface structures on T cells (3), but other surface molecules have remained elusive. As many as 200 mRNA species are expressed in T cells but absent in B cells. A third of those encode membrane-associated molecules (4). We describe the characterization of a cDNA clone present in mature T cells but not expressed in the earliest stage of T-cell differentiation. This cDNA encodes a 16.7-kDa protein, which we have named MAL. This protein has a predicted secondary structure containing four potential transmembrane domains that resembles the structure of a number of membrane proteins (5).

MATERIALS AND METHODS

Cells were grown in RPMI 1640 supplemented with 10% (vol/vol) fetal bovine serum at 37°C in a 5% CO2/95% air atmosphere. Frozen human tissues were kindly provided by the Department of Surgical Pathology of the Yale University School of Medicine.

Total cytoplasmic RNA from tissue culture cells was prepared by the Nonidet P-40 lysis method (6). Membrane-bound RNA was prepared by mechanical disruption of cells in hypotonic buffer and differential centrifugation (7). When frozen tissues were used, total RNA was isolated by homogenization in 4 M guanidinium thiocyanate, followed by ultracentrifugation through a 5.7 M CsCl cushion (8). Poly-(A)+ RNA was isolated by oligo(dT)-cellulose chromatography (9). High molecular weight genomic DNA was prepared essentially as described by Maniatis et al. (10).

The first strand of cDNA was synthesized by oligo(dT) priming using poly(A)+ RNA from MOLT-4 cells and avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, FL) in the presence of actinomycin D at 100 ng/ml. This cDNA was mixed with a 10-fold mass excess of poly(A)+ RNA from CCRF HSB-2 cells, boiled for 60 sec, and incubated at 68°C in 0.5 M phosphate buffer, pH 6.8/5 mM EDTA/0.1% NaDodSO4 to a R(t) (initial concentration of RNA × time) value of 1500 (11). Unhybridized cDNA was separated from the cDNA-RNA hybrids by chromatography on a hydroxyapatite column using 0.12 M phosphate buffer, pH 6.8/0.1% NaDodSO4 at 60°C. This single-stranded cDNA fraction was then used to construct libraries in the EcoRI site of pBR322 (10) and bacteriophage λgt10 (12). The second strand was synthesized by using RNase H, Erlichia cell DNA polymerase, and T4 DNA ligase (13). cDNA molecules >800 base pairs long were cloned in the unique EcoRI site of λgt10. Subtracted [32P]cDNA probes were generated using a similar protocol except that the cDNA was labeled to specific activities of up to 106 cpm/μg (11). Screening was carried out using 106 cpm per 137-mm filter under standard conditions (10).

High molecular weight genomic DNA was digested with restriction endonucleases and blotted as described by Southern (14). For RNA blots, RNA was denatured in the presence of 50% (vol/vol) formamide and 2.2 M formaldehyde, subjected to electrophoresis on 1.2% agarose/formaldehyde gels, and blotted as described by Thomas (15). Final blot washing conditions were 0.1× SSC/0.1% NaDodSO4 at 50°C. (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0.)

Restriction fragments from λMA5 and pMA34 inserts were subcloned into the M13mp8 vector and sequenced (16).

The full-length cDNA insert from λMA5 was subcloned in the appropriate orientation in the EcoRI site of pSP65 (17). Transcription of 2 μg of linearized plasmid was performed with SP6 polymerase in the presence of unlabeled nucleotides and 0.5 mM P]-5'-[7-methylguanosine]-P]-5'-guanosine triphosphate (mGpppG). One-tenth of the reaction mixture was translated in a rabbit reticulocyte lysate system (Promega Biotech, Madison, WI) in the presence of L-[35S]methionine under the conditions suggested by the supplier. The in vitro translation products were subjected to electrophoresis on NaDodSO4/polyacrylamide gels under reducing conditions using 7–15% polyacrylamide gradient gels as described by Laemmli (18).

Abbreviation: mGpppG, P]-5'-[7-methylguanosine]-P]-5'-guanosine triphosphate.

*Present address: Centro de Biologia Molecular, Facultad de Ciencias, Universidad Autonoma, Cantoblanco, 28049-Madrid, Spain.
RESULTS

Table 1 shows the surface marker profile of some of the leukemic T-cell lines used in the present study. In the strategy we applied to isolate cDNA sequences differentially expressed during T-cell ontogeny, sequences shared between MOLT-4 (stage II) and CCRF HSB-2 (stage I) cells were depleted by annealing cDNA synthesized using poly(A)* RNA from MOLT-4 cells as template with poly(A)* RNA from CCRF HSB-2 cells, and subsequently the unhybridized cDNA was separated from the CDNA-RNA hybrids by chromatography on hydroxyapatite columns (11). This single-stranded cDNA was then used to construct a library in the plasmid pBR322. About 3000 recombinants were screened with a probe prepared by subtracting cDNA from HPB-ALL cells, a line of cells derived from a patient with acute T-cell lymphoblastic leukemia (stage II/III) with poly(A)* RNA from CCRF HSB-2 cells. A cDNA probe from HPB-ALL cells was used to focus on stage II/III specific sequences shared by MOLT-4 and HPB-ALL cells, rather than MOLT-4 specific sequences. Five clones hybridized consistently with the probe. The clone pMA34 that carried the largest insert (350 base pairs) was used to test whether the corresponding mRNA is present, in general, in T cells in an advanced stage of differentiation (Fig. 1). A single hybridizing mRNA band of 1.1 kilobases was detected with RNA prepared from the cell lines MOLT-4, HPB-ALL, Jurkat, and an uncharacterized acute T-cell lymphoblastic leukemia (T-ALL), whereas no hybridization was detected with RNA from CCRF HSB-2, CCRF CEM, and a different uncharacterized T-ALL leukemia. Moreover, no expression of pMA34 was evident in three different lines of B-cell origin (JY, G-7, and BL), in the erythroleukemic cell line K-562, in the promyelocytic cell line HL-60, or in HeLa cells. Fig. 1 shows that pMA34 is expressed in human mature T-cell clones (lanes b and c) indicating that pMA34 expression also occurs in normal T lymphocytes and is not restricted to T-cell lines of leukemic origin. The same 1.1-kb RNA species was present at much higher levels in a preparation of membrane-associated RNA (lane a) as compared with total cytoplasmic RNA from MOLT-4 cells (lane d). Fig. 1B also shows that pMA34 cDNA is expressed in thymus (lane h) but not in colon (lane f), adrenal glands (lane g), or liver (lane i).

To isolate a full-length cDNA, we prepared a MOLT-4 cDNA library in λgt10 (12) using the procedure of Gubler and Hoffman (13). Screening of 50,000 recombinants from this unamplified library with nick-translated pMA34 insert gave nine positive clones, of which five of the six analyzed had the same length as the mRNA detected by RNA gel blot analysis. The nucleotide and the deduced amino acid sequences of the cDNA are shown in Fig. 2. A single open reading frame extends from the ATG at nucleotide 1 to the TAA stop codon at base 460, encoding a protein with a predicted molecular mass of 16,700. We have assigned the first methionine codon as the initiator because it is the first in-frame ATG downstream of the stop codon at base −45 and because the sequences flanking this ATG are homologous to the highly conserved sequence CCRCCATGG (where R stands for a purine) that flanks functional initiation sites in eukaryotic mRNAs (24).

Upstream of the 3' end of the MAL cDNA there are no perfect consensus polyadenylation signals (25) though there is an ATAAAA sequence. Although ATAAAA is common to most eukaryotic mRNAs, there are a number of cases in which this sequence is not present (26, 27). The putative polyadenylation signal in the MAL cDNA is adjacent to the sequence TGTCTTAA, which is similar to the consensus sequence YGTGTTYY (where Y stands for a pyrimidine) found between many polyadenylation signals and their poly(A) tails (27).

A computer search through the GenBank† and Protein Identification Resource‡ databases found no significant over-

![Fig. 1.](image)

Table 1. Surface phenotypic profile of T-lineage leukemic cells

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Surface phenotype*</th>
<th>T-cell receptor†</th>
<th>MAL RNA</th>
<th>Stage of differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCRF HSB-2</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>I</td>
</tr>
<tr>
<td>CCRF CEM</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>I/II</td>
</tr>
<tr>
<td>MOLT-4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>II/III</td>
</tr>
<tr>
<td>HPB-ALL</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>II/III</td>
</tr>
<tr>
<td>Jurkat</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

The presence (+) or absence (−) of each antigen on the surface of each type of cell is shown.

*Data taken from Reinherz et al. (19) and Greaves et al. (20).

†Data taken from Collins et al. (21), Royer et al. (22), and Sangster et al. (23).


all homology between the MAL nucleotide and amino acid sequences and any known DNA or protein sequences.

Genomic DNA from a variety of human cells expressing or not expressing MAL mRNA and from human placenta was digested with either EcoRI or BamHI, electrophoresed, and subjected to hybridization with the isolated insert of the full-length cDNA clone λMAS. No evidence of DNA rearrangements was apparent (unpublished results). There is no EcoRI site in the full-length cDNA but there are two bands in the EcoRI genomic DNA blots and one in the BamHI blots, so that the gene is present as a single copy in the human genome and contains at least one intron.

There are no N-glycosylation sites (Asn-Xaa-Ser/Thr) in the deduced amino acid sequence (Fig. 2), and there are two serines in the COOH terminus of the molecule in a configuration that resembles those favored by cAMP- and cGMP-dependent protein kinases (28). A hydrophathy plot (29) of the derived amino acid sequence of the MAL protein is presented in Fig. 3. This plot predicts four hydrophobic regions of 19–25 amino acids in length corresponding to peaks I–IV, interspersed with hydrophilic domains that contain a predicted β-turn secondary structure (30). Although peak b predicts a turn, which is due to proline-23 in the first hydrophobic segment, its effect in the α-helix structure could be minimized because of the tendency of transmembrane segments to span the lipid bilayer as α-helices (29). The alternation of hydrophobic domains of 19–25 amino acids in length with hydrophilic segments has been seen in transmembrane proteins spanning the membrane several times (5, 6, 31).

The predicted MAL protein lacks an NH2-terminal hydrophobic signal sequence characteristic of many membrane proteins (32). Most of the transmembrane proteins that lack a cleavable signal peptide are considered to have their NH2 terminus on the cytoplasmic face of the membrane (5, 33), although there are examples in which the NH2 terminus is located in the extracytoplasmic space (34).

Fig. 2. Nucleotide sequence of the MAL cDNA and deduced amino acid sequence of the MAL protein. The in-frame stop codons of the open reading frame are indicated by asterisks. Numbers shown above the amino acid sequence designate amino acid residue positions. Numbers beneath the nucleotide sequence show nucleotide positions. The nucleotides were numbered from the position of the presumed initiator methionine codon. The location of the putative polyadenylation signal (ATTAAAA) is underlined.

Fig. 3. Hydrophathy plots and predicted turns plots of the deduced amino acid sequence of the MAL protein. (A) β-Turns plot of the MAL amino acid sequence. The curve was generated by analysis of the sequence according to the method of Chou and Fasman (30) using a window of four amino acid residues. Peak clusters corresponding to potential turns are named a, b, c, d, e, and f. (B) Hydrophathy plot of the MAL protein. The plot was generated using the algorithm of Kyte and Doolittle (29). The curve is the average of a hydrophobicity index for each residue over a window of 20 residues. Positive values indicate hydrophobic regions, and negative values represent hydrophilic segments. The four potential transmembrane domains are named I–IV.
of amphiphilic transmembrane domains in the MAL protein suggests that it may line an aqueous channel (36).

The full-length cDNA insert from λMAS was subcloned into pSP65, a plasmid vector that contains the bacteriophage SP6 promoter (37). pSP6-MA5 DNA was linearized in the MAL-coding sequence with BstEII at a site 37 nucleotides downstream of the stop codon, with HindIII downstream of the 3' end of the cDNA, or with HindIII in the vector polylinker. Synthetic m'GpppG-capped mRNA was transcribed by SP6 polymerase in the presence of m'GpppG (37) giving truncated (BstEII and HindIII-linearized plasmids) or full-length (HindIII-linearized plasmid) run-off transcripts (Fig. 5). Translation of the full-length RNA in a rabbit reticulocyte lysate system gave at least four discrete bands of apparent molecular mass of 20, 26, 32, and 40 kDa (Fig. 5) that were not present in the control RNA (lane a vs. b). The position of these proteins and the presence of additional weak bands in the upper part of the gel suggest that they represent oligomers either of the 20-kDa protein or of a protein (=14 kDa) masked in the autoradiogram by the globin excess in the reticulocyte lysate. The same pattern was obtained when template RNA prepared from HindIII-linearized plasmid was used (lane d). Since the HindII transcript does not contain enough information to encode proteins over 20 kDa, this result rules out the possibility that the multiple bands were due to aberrant translation of the full-length RNA. When RNA synthesized from BstEII-linearized pSP65-MA5 was used as template for in vitro translation reactions, a single band of the predicted size was observed (lane c). This suggests that the COOH-terminal half of the molecule is needed for both the aggregation and anomalous mobility of the MAL protein in NaDodSO4/polyacrylamide gels. The presence of canine microsomal membranes (38) in the in vitro translation mixture did not result in any alteration in the mobility of the proteins synthesized.

**DISCUSSION**

Proteins involved in the transport of water-soluble molecules and ions across cellular membranes are believed to span the lipid bilayer several times (4, 5, 31). Such proteins often have hydrophilic residues confined to one face of the helix. Polar faces of adjacent helices, from the same or different subunits, could form a pore or a channel through the membrane (31, 36). In the case of ion-channel proteins, at least one of the transmembrane domains is strongly amphipatic and contains charged residues (31, 36). In addition, most of the sequenced membrane-transport proteins lack an NH2-terminal signal peptide, and some of them have both NH2 and COOH termini positioned in the cytoplasm (5, 33). The following secondary structure predictions support the idea that MAL could be involved in transport across the membrane: (i) MAL has four potential transmembrane domains. (ii) At least one of the presumed α-helices spanning the membrane is strongly amphipatic and contains charged residues. (iii) MAL lacks a NH2-terminal signal peptide. (iv) The formation of oligomers in the in vitro translation reaction indicates that the MAL protein has a strong tendency to self-aggregate and suggests that the putative channel may be formed by a complex of MAL molecules, although in the cellular membrane MAL may be associated with different protein subunits.

To our knowledge, no membrane proteins traversing the membrane multiple times have been described that are specifically associated with T cells, although the existence of K+ (39) and Ca2+ (40) channels in peripheral T lymphocytes is well established. T-cell activation by mitogens (41), antigens (42), or monoclonal antibodies against the T3/T-cell receptor protein complex (43) or the T11 glycoprotein (44) results in an increase in cytoplasmic free Ca2+. This has led a number of groups to postulate the existence of at least a Ca2+ channel linked to the T11 structure and/or T3/T-cell receptor complex (40, 44). Based on the strong labeling of the 20-kDa nonglycosylated ε chain of the T3 complex with photoactivatable hydrophobic reagents, it was speculated that the ε subunit could be the putative Ca2+ channel (40). However, molecular cloning of the cDNA encoding the ε chain of the T3 complex has shown that the deduced amino acid sequence for this protein predicts a structure with one membrane-spanning domain, similar to other single-spanning membrane proteins (45). The transmembrane arrangement of the MAL protein and its presence in leukemic T-cell lines.
expressing T11 and the T3/T-cell receptor protein complex (HPB-ALL and Jurkat; Table 1) and in normal mature T-cell clones make MAL a candidate for involvement in membrane signaling in T cells activated either via T11 or T-cell receptor pathways.

The major obstacles to generating monoclonal antibodies against T-cell-specific cell-surface molecules reside both in the greater immunogenicity and abundance of other surface antigens. The secondary-structure predictions for the deduced amino acid sequence of MAL postulate the existence of two highly charged segments. Synthetic peptides covering these regions coupled to appropriate carrier could act as immunogens to raise monoclonal antibodies. The availability of such antibodies could be used to study the function of MAL protein.

We are grateful to Dr. V. Marchesi and D. Engelman for their helpful suggestions, to Dr. E. Engleman for his gift of human T-cell clones, and to P. Rogan for his assistance in the computer analysis of the MAL sequence. We also acknowledge Jeff Shapiro and Yoon-ock Kim for their expert technical assistance and Ann M. Mullvey for assistance in the preparation of this manuscript. M.A.A. was the recipient of a European Molecular Biology Organization fellowship. This work was supported by a grant from the National Institutes of Health (S.M.W.).