The cytoplasmic domain of C-CAM is required for C-CAM-mediated adhesion function: studies of a C-CAM transcript containing an unspliced intron

Peter H. CHEUNG,* Ognjen CULIC,* Yuhong QIU,* Karen EARLEY,* Nancy THOMPSON,† Douglas C. HIXSON† and Sue-Hwa LIN*‡

*Department of Molecular Pathology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, U.S.A., and †Department of Medical Oncology, Rhode Island Hospital, Brown University, Providence, RI 02903, U.S.A.

Cell-CAM105 (also named C-CAM) is a cell surface glycoprotein involved in intercellular adhesion of rat hepatocytes. It has four extracellular immunoglobulin (Ig) domains, a transmembrane domain and a cytoplasmic domain and therefore is a member of the Ig supergene family. We have characterized multiple cDNAs of the C-CAM genes in rat intestine. Sequence analyses showed that rat intestine contained not only the previously reported L-form and S-form C-CAMs (renamed C-CAM1 and C-CAM2 respectively) but also a new isoform, C-CAM3. The C-CAM3 transcript codes for a polypeptide with a truncated C-terminus that lacks 65 amino acids from the previously reported C-CAM1 cytoplasmic domain. Unlike C-CAM1, C-CAM3 did not mediate cell adhesion when expressed in insect cells using the baculoviral expression system. Thus the extra 65 amino acids in the cytoplasmic domain of C-CAM1 are important for adhesion phenotype when expressed in insect cells. Although C-CAM1 and C-CAM2 are encoded by different genes, sequence analysis suggests that C-CAM3 is probably derived from alternative splicing of the C-CAM1 gene. To examine this possibility, we have determined the exon organization of the C-CAM1 gene. C-CAM3 differed from C-CAM1 by the presence of a single unspliced intron which contained a stop codon immediately after the regular splice junction. As a result, translation of C-CAM3 terminates at the point where C-CAM1 and C-CAM3 sequences diverge. To investigate the expression of C-CAM1, C-CAM2 and C-CAM3 in different tissues, we used an RNAase-protection assay to simultaneously assess the levels of expression of these transcripts. Using total RNA prepared from various tissues, we showed that expression of C-CAM3 was tissue-specific, and the C-CAM3 transcript accounted for about 25% of the transcripts derived from the C-CAM1 gene. However, further analysis revealed that C-CAM3 transcript was not present in cytosolic RNA, rather it was enriched in nuclear RNA prepared from hepatocytes. Although C-CAM3 cDNA contains the polyadenylation signal and is polyadenylated, these results indicate that C-CAM3 is probably an incomplete spliced product of C-CAM1 gene.

INTRODUCTION

Cell adhesion molecules, which are involved in both cell–cell and cell–matrix interactions, have been implicated in tissue development (Edelman and Crossin, 1991), epithelial cell differentiation (Boyer and Thiery, 1989), carcinogenesis (Thompson et al., 1991) and metastasis (Nicolson, 1989). Cell-CAM105 (also named C-CAM) is a 105 kDa cell-surface glycoprotein involved in intercellular adhesion of hepatocytes (Ocklind and Obrink, 1982) and is postulated to play a role in liver histogenesis (Mowery and Hixson, 1991). We previously reported the cloning and characterization of two cDNAs coding for the rat liver cell-CAM105 (C-CAM1 and C-CAM2) (Culic et al., 1992). Using anti-peptide antibodies specific for the longer form (C-CAM1) and antibodies that react with both isoforms, we have shown that both isoforms (C-CAM1 and C-CAM2) are expressed on the surface of hepatocytes, with the shorter form (C-CAM2) being the predominant form. On SDS/PAGE, the liver C-CAMs have apparent molecular sizes of 110 and 105 kDa for C-CAM1 and C-CAM2 respectively. Anti-C-CAM polyclonal antibodies have been reported to react with proteins with different apparent molecular masses in different tissues (Odin et al., 1988). The biochemical properties of intestinal C-CAM are also significantly different from those of liver C-CAM (Hansson et al., 1989). These observations show that there are tissue-specific differences between the C-CAM molecules.

To determine whether there is tissue-specific expression of various C-CAM genes, we cloned rat intestinal C-CAM cDNAs from a cDNA library and isolated a third isoform of C-CAM. As more isoforms may exist, nomenclature of these molecules should be according to the order of their discovery. Therefore, we have renamed the L-form cell-CAM105 C-CAM1, the S-form cell-CAM105 C-CAM2 and this newly isolated clone C-CAM3. Here, we report the cloning of this new isoform, which is a truncated isoform of C-CAM1. In contrast with C-CAM1, C-CAM3 does not mediate cell adhesion when expressed *in vitro* suggesting an essential requirement of the cytoplasmic domain for adhesion function. In addition, we showed that the transcript of this truncated isoform was generated from retention of an intron from the C-CAM1 gene.

MATERIALS AND METHODS

Isolation and characterization of C-CAM3 cDNAs and DNA sequence analysis

A rat intestinal cDNA library in λ Uni-ZAP (Stratagene Inc., La Jolla, CA, U.S.A.) was screened with a nick-translated *Bam*H1–*Pst*1 fragment excised from C-CAM1 cDNA (Lin and Guidotti, 1989). A total of 10^6 plaques were screened. Some 184 positive clones were obtained, of which 40 were purified and analysed. After *in vivo* excision of phagemids from λ phage, the nucleotide sequences of the double-stranded DNA were de-

Abbreviations used: Ig, immunoglobulin; $C_{12}E_9$, polyoxyethylene 9-lauryl ether; ConA, concanavalin A; CEA, carcinoembryonic antigen.

‡ To whom correspondence should be addressed.

The nucleotide sequences for C-CAM1, C-CAM2, and C-CAM3 have been submitted to the GenBank/EMBL and have accession numbers J04963, Z12019 and M92848 respectively.

termined using the T7 Sequenase kit (United States Biochemical Corp., Cleveland, OH, U.S.A.) according to the manufacturer's instruction. Sequence analysis revealed that we had cloned not only the C-CAM1 and C-CAM2 cDNA but also a new form, which we named C-CAM3.

Cloning of the full-length C-CAM1 and C-CAM3 cDNA into baculoviral expression vectors

The full-length C-CAM1 cDNA was excised from pBS/full plasmid (Lin and Guidotti, 1989) with *Xba*I and *Nsi*I. The *Xba*I–*Nsi*I fragment was then inserted into the *Xba*I and *Pst*I cloning sites of the baculoviral transfer vector pVL1393 (Invitrogen, San Diego, CA, U.S.A.). To insert the full-length C-CAM3 cDNA into the baculoviral vector, a plasmid containing the full-length C-CAM3 was digested with *Eco*R1 and *Xmn*I to produce a 1.63 kb fragment containing the entire coding region. This 1.63 kb fragment was subcloned into the *Eco*R1–*Sma*I site of pVL1392 to generate pVL-C-CAM3.

Adhesion assays

Recombinant baculoviruses carrying either the C-CAM1 or C-CAM3 genes were generated by cotransfection of baculoviral transfer plasmid and wild-type baculoviral DNA using the BaculoGold transfection kit (PharMingen, San Diego, CA, U.S.A.) according to the instructions provided by the manufacturer. *Spodoptera frugiperda* (Sf9) cells were infected with C-CAM1 or C-CAM3 recombinant virus or wild-type virus and cultured in spinner flasks according to the procedures of Summers and Smith (1987). At various time intervals, portions of virus-infected Sf9 cells were removed and treated with 50 units/ml DNAase I (Promega, Madison, WI, U.S.A.). The single cells in these samples were counted with a haemocytometer.

Analysis of C-CAM isoforms expressed in Sf9 cells

To test whether the expressed proteins can bind concanavalin A (ConA), the total lysate of infected Sf9 cells was prepared by solubilizing these cells in 0.5% detergent polyoxyethylene 9-lauryl ether ($C_{12}E_9$) for 15 min at 4 °C. The solubilized sample was centrifuged in an Eppendorf microfuge, and the supernatant was loaded on a ConA–Sepharose 4B gel packed in 1 ml blue pipette tips. The column was first washed with 5 column vol. of washing buffer (0.1% $C_{12}E_9$ in 1 mM $CaCl_2$, 1 mM $MgCl_2$ and 50 mM Tris/HCl, pH 7.4), and then eluted with 0.5 M methyl α-mannoside in 0.1% $C_{12}E_9$. Portions of the column eluates were boiled in SDS sample buffer and analysed by SDS/PAGE (Laemmli, 1970). Western immunoblotting using antibody Ab669 against C-CAM was performed as previously described (Culic et al., 1992) except that alkaline phosphatase-conjugated goat anti-rabbit antibody was used as the second antibody. The coloured stain was developed with Nitroblue Tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate as substrate.

Immunofluorescence

Sf9 cells grown on cover slips were infected with recombinant viruses. After 60 h, the cells were fixed with 4% formaldehyde in PBS for 20 min at room temperature. The cells were then treated with or without 0.1% Triton X-100 in PBS for 10 min. The cells on coverslips were incubated with Ab669 (Culic et al., 1992) at 1:300 dilution in PBS for 45 min at room temperature. The coverslips were washed three times for 5 min each in PBS and further incubated with fluorescein-conjugated goat anti-rabbit secondary antibody for 45 min at room temperature. After the cells had been rinsed with PBS, the coverslips were mounted on slides, examined and photographed using a Nikon fluorescence microscope.

To stain Sf9 cells with antibodies without fixation, Sf9 cells grown on coverslips were infected with recombinant virus. After 48 h, the cells were incubated with Ab669 at 1:300 dilution in insect cell medium containing 10% fetal bovine serum for 45 min at room temperature. The cells were washed three times and further incubated with fluorescein-conjugated goat anti-rabbit secondary antibody (1:250 dilution) for 45 min at room temperature. The antibodies were removed, Trypan Blue in 1:300 dilution was added to the cell and the stained cells were examined on a Nikon fluorescent microscope with two Nikon photo cubes, B2E and G2E.

Isolation and characterization of C-CAM1 genomic clones

A rat genomic library (a *Sau*3A partial digest in λ DASHII) was purchased from Stratagene. The library was screened with the *Eco*R1–*Nsi*I 1.6 kb fragment of the full-length C-CAM1 cDNA labelled with ^{32}P using the Klenow fragment of DNA polymerase I in the presence of random hexanucleotide primers and [α-^{32}P]-dCTP (Sambrook et al., 1989). The phage DNAs of positive clones were characterized by restriction mapping. Restriction fragments from bacteriophage containing the C-CAM1 gene were ligated into the plasmid vectors pBS (Stratagene) or pUC19. The nucleotide sequence of the exons and the adjacent introns was determined using specific oligonucleotide primers.

Plasmid construction and preparation of RNA probe

To construct a C-CAM3-specific probe for RNAase protection studies, a 312 bp *Bgl*II–*Bam*H1 fragment (nt 1178–1490) was excised from the C-CAM3 cDNA and inserted into the *Bam*H1 sites of Bluescript (Stratagene) to produce pSK-IL60(*Bgl*II–*Bam*H1). The orientation of the inserts was determined by double-stranded dideoxy sequencing with T7 DNA polymerase and universal primers using the T7 Sequenase kit (United States Biochemical Corp.) according to the instructions provided by the manufacturer.

To synthesize antisense RNA probes, plasmid pSK-IL60(*Bgl*II–*Bam*H1) was linearized with *Xba*I. Radiolabelled RNA probes were synthesized from the linearized DNA template using T7 RNA polymerase in the presence of [^{32}P]UTP as described by Melton et al. (1984). DNA templates were digested with RNAase-free DNAase I (Promega), and the labelled probes were used directly for RNAase-protection experiments without further purification.

Preparation of RNA from different tissues

RNA was isolated from adult rat tissues by homogenizing frozen tissues in guanidine isothiocyanate. The homogenates were layered over a CsCl cushion and centrifuged as described by Chirgwin et al. (1979). The integrity of the RNA was determined by formaldehyde–agarose gel electrophoresis followed by ethidium bromide staining and u.v. transillumination as described (Sambrook et al., 1989).

Preparation of hepatocyte nuclear and cytoplasmic fractions

Hepatocytes were isolated from a 250 g adult male Sprague–Dawley rat as previously described (Thompson et al., 1993). The resulting hepatocyte cell pellet was washed three times in cold

diethylpyrocarbonate-treated PBS and once in cold PBS containing 10 units/ml RNAase-free DNAase. It was then centrifuged at 500 *g* (2500 rev./min) to give a pellet of approx. 5 ml volume. All the supernatant PBS was removed and the pellet was lysed by resuspension and gentle vortexing in cold Nonidet P40 lysis buffer (0.01 M Tris/HCl, pH 7.9, 0.015 M $MgCl_2$, 0.15 M NaCl, 0.65% Nonidet P40) containing 375 units/ml RNasin RNAase inhibitor. Samples of the cell lysate were monitored microscopically to check for complete cell lysis. Nuclei were pelleted by centrifugation at 1000 *g* (3000 rev./min) for 15 min. The cytoplasmic supernatant was removed into another tube and a sample was examined to ensure that no nuclei were visible in this fraction. The nuclei pellet was reserved for the preparation of nuclear RNA as described below.

Preparation of cytoplasmic RNA

Cytoplasmic RNA was prepared from 2.6 ml of cytoplasmic extract. The extract was added to 5.3 ml of a 1.5 × stock solution of guanidine thiocyanate buffer (Chirgwin et al., 1979), homogenized briefly, and centrifuged over CsCl as described (Chirgwin et al., 1979). The resulting RNA pellet was resuspended in sterile water, extracted with phenol/chloroform (1:1, v/v) and precipitated with ethanol overnight at −20 °C.

Preparation of nuclear RNA

In order to remove contaminating cytoplasmic debris, nuclei were resuspended in solution I (250 mM sucrose, 0.1% Tween 80, 2% citric acid) and centrifuged at 1500 *g* (3000 rev./min), for 10 min, at 4 °C through a sucrose cushion (880 mM sucrose, 0.1% Tween 80, 2% citric acid). All supernatant was removed and the resulting purified nuclear pellet was homogenized in 8 ml of 1 × working guanidine thiocyanate buffer (Chirgwin et al., 1979) and prepared as described above for cytoplasmic RNA preparation. On electrophoresis, a sample of the nuclear RNA preparation was free of visible ribosomal RNA contamination.

RNAase-protection assay

RNA from each sample was hybridized with 2×10^5 c.p.m. of ^{32}P-labelled antisense RNA probe and subjected to an RNAase-protection protocol (Zinn et al., 1983). The protected fragments were analysed on a 4% polyacrylamide/7 M urea gel. The gel was then dried and analysed by autoradiography. The relative amounts of C-CAM isoforms were determined from the gels using a Betascope blot analyser (Betagen Corp., Waltham, MA, U.S.A.).

RESULTS

C-CAM3 cDNA structure

To study whether the sequences of the intestinal C-CAMs were different from liver C-CAMs, we cloned rat intestinal C-CAM cDNAs from a cDNA library. Sequence analyses showed that rat intestine contained not only the two previously reported C-CAM isoforms but also a new isoform, C-CAM3, which differs from the previously isolated C-CAM1 in the C-terminal cytoplasmic domain. The sequence of C-CAM3 is shown in Figure 1.

Of the 2381 bp of C-CAM3 cDNA sequenced, 57 bp were in the 5′-untranslated region, and 1362 bp comprised the open reading frame. As this cDNA is 4.0 kb long, it has a long 3′-untranslated region. Except for a single nucleotide deletion in the 5′-untranslated region, both the 5′-untranslated region and the coding region of this new isoform were identical with the previously published liver C-CAM1 sequence (Culic et al., 1992). However, an in-frame TAG stop codon was found at position 1363–1365. As a result, C-CAM3 has a cytoplasmic domain with only six amino acids, in contrast with C-CAM1, which contains a 71-amino acid cytoplasmic domain. A polyadenylation consensus site (AATAAA) followed by a poly-A tail containing a string of 20 adenosines was identified at the very end of the 3′-untranslated region. The sequence from polyadenylation site to the end of the clone is AATAAAAATTTATATATCACTG$(A)_{20}$. As two clones of this new isoform of different lengths were found, C-CAM3 is probably not an artifact generated during construction of the cDNA library.

Adhesion function of C-CAM1 and C-CAM3

The existence of multiple C-CAM isoforms raises the possibility that these different isoforms may have different functions. To address this question, one can express individual proteins in cells that do not express C-CAM and examine the aggregation phenotype conferred by individual C-CAM isoforms. We have previously shown that expression of C-CAM1 in Sf9 cells caused Sf9 cells to aggregate (Cheung et al., 1993). This result suggests that C-CAM1 can mediate cell adhesion when expressed in insect cells. As C-CAM is one of the few carcinoembryonic antigen (CEA) family members that has a cytoplasmic domain, we took advantage of this newly identified C-CAM3 clone to probe the requirement of C-CAM1 cytoplasmic domain in adhesion. To study the adhesion function of C-CAM3, a portion of cDNA containing the entire coding region for C-CAM3 isoform was inserted into the baculoviral genome via homologous recombination, and C-CAM3 protein was expressed in Sf9 cells by infecting them with the C-CAM3 recombinant baculovirus. Figure 2(c) shows the expression of C-CAM1 and C-CAM3 gene products in Sf9 cells. In Western blot, immunoreactive proteins with apparent molecular masses around 70 kDa were detected in cells infected with C-CAM1 and C-CAM3 recombinant viruses. C-CAM3 protein showed slightly faster mobility than C-CAM1 (Figure 2c). As these proteins could bind to ConA–Sepharose 4B columns and be eluted with methyl α-mannoside (results not shown), they were probably glycosylated. The broad diffuse bands exhibited by these proteins on SDS/PAGE suggest that the glycosylation was heterogeneous and C-CAM3 showed a more diffuse pattern than C-CAM1. As C-CAM1 and C-CAM3 differ in the lengths of their cytoplasmic domains, 71 amino acids and six amino acids for C-CAM1 and C-CAM3 respectively, a molecular mass difference of about 7 kDa is expected. Taking into account the heterogeneity of their glycosylation, the mobility difference observed on SDS/PAGE for these molecules is reasonable. However, the reason for the more diffuse pattern of C-CAM3 than C-CAM1 on SDS/PAGE analysis is not clear. Both Sf9 cells alone and Sf9 cells infected with wild-type virus did not produce any proteins immunoreactive with the anti-C-CAM antibody Ab669 (results not shown).

When Sf9 cells in suspension were infected with C-CAM1 recombinant virus, significant aggregation of Sf9 cells was seen at 48, 72 and 96 h after infection (Figure 2a). Aggregation of infected cells was then determined by measuring the disappearance of single cells. The appearance of cell aggregation correlated well with both the disappearance of single cells and temporal expression of C-CAM1 protein (Figures 2a, 2b and 2c). In contrast, Sf9 cells expressing C-CAM3 did not aggregate despite the fact that both the temporal and quantitative expressions of C-CAM3 were similar to that of C-CAM1 (Figures 2a, 2b and 2c). The phenotypic difference observed for C-CAM1

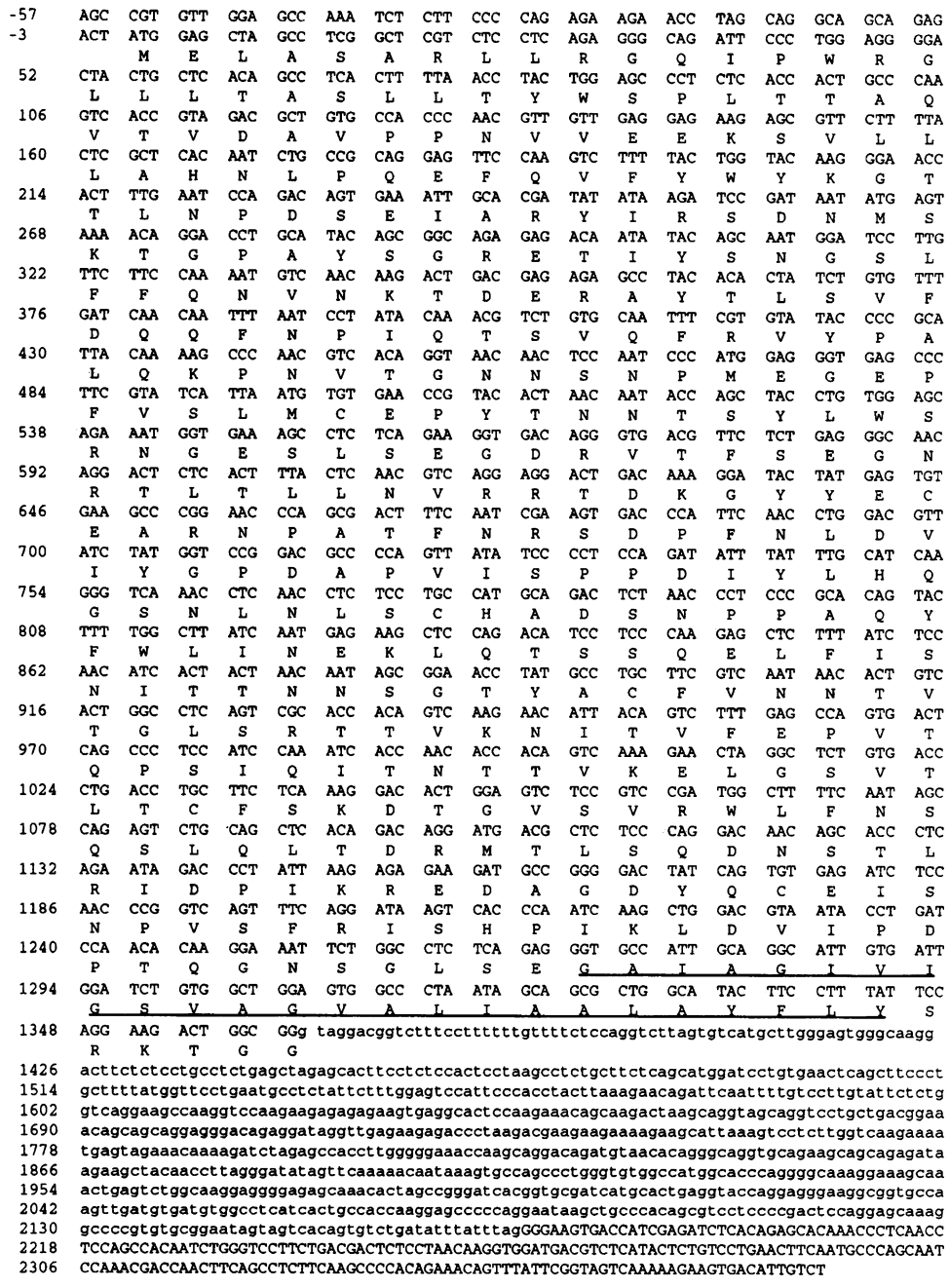


Figure 1 Nucleotide sequence (numbered lines) and deduced amino acid sequence (unnumbered lines) of C-CAM3

Capital nucleotide letters indicate that the cDNA sequence was derived from exon, and lower-case letters indicate that the sequence was derived from intron.

and C-CAM3 was not due to different amounts of protein expressed, as C-CAM3 never supported cell adhesion even at a high level of expression. On the other hand, C-CAM1 consistently gave the adhesion phenotype as early as 48 h after infection when only small amounts of C-CAM1 had started to be expressed. It is possible that failure of C-CAM3 to show the aggregation phenotype may be due to the fact that it was not targeted to the cell surface. This possibility was addressed by the following two approaches. In the first approach, cells were fixed with 4% formaldehyde followed with immunofluorescence staining. Strong immunofluorescence staining was seen in C-CAM3 recombinant virus-infected cells with or without Triton X-100 solubilization (results not shown). In a control experiment, we found that an anti-peptide antibody (anti-C1) (Lin et al., 1991), which recognizes the cytoplasmic domain of C-CAM1, could stain C-CAM1-expressing cells only after Triton X-100 permeabilization. This observation suggests that fixation with formaldehyde did not make cells permeable to antibodies. Therefore staining of C-CAM3 by Ab669 without Triton X-100 solubilization was consistent with the expression of C-CAM3 on the cell surface. In the second approach, the C-CAM3 recombinant virus-infected cells were first incubated with antibodies (Ab669) followed by incubation with fluorescein-conjugated goat anti-rabbit antibodies without fixation. The

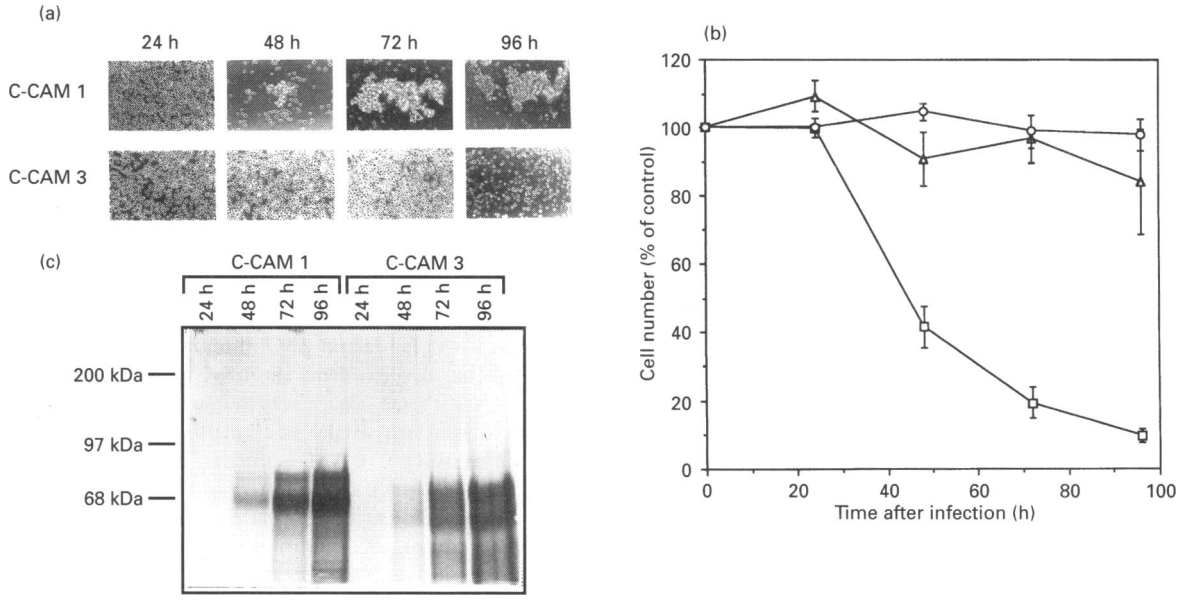


Figure 2 Aggregation of cells in suspension

Sf9 cells cultured in suspension were infected with wild-type virus (△), C-CAM1 recombinant virus (□) or C-CAM3 recombinant virus (○). At various time intervals, samples were removed for microscopic examination (**a**), cell number determination (**b**) and Western immunoblot analysis with Ab669 (**c**). In (**b**), the number of cells infected at time 0 was used as 100%, and the results are means ± S.E.M. of three separate experiments.

viability of the cells was assessed by addition of 0.3% Trypan Blue. Significant amounts of cells, which were not stained by Trypan Blue, were positive with Ab669 staining. This result further supports the surface localization of C-CAM3. As the extracellular protein sequences of C-CAM1 and C-CAM3 were identical, this result indicates that the phenotypic differences observed between these two alternatively spliced isoforms were probably due to the difference in their cytoplasmic domains.

C-CAM3 is derived from unusual splicing of C-CAM1

In previous studies, we cloned C-CAM1 and C-CAM2 cDNAs from rat liver cDNA libraries. As there are several nucleotide differences between these two isoforms scattered throughout the sequences, it is likely that C-CAM1 and C-CAM2 are derived from different genes rather than from alternative splicing. In contrast, the coding sequence of C-CAM3 is identical with that of C-CAM1 except for the shorter cytoplasmic domain in C-CAM3. This finding suggests that C-CAM3 is probably derived from alternative splicing of the C-CAM1 gene product. Thus both different genes and alternative splicing contribute to the diversity of the C-CAM family.

To characterize the alternative splicing patterns of C-CAM1 and C-CAM3, we determined the exon organization of C-CAM1 gene by restriction mapping and sequencing the exons and the adjacent intron regions using oligonucleotide primers derived from the cDNA sequence. The coding region of the C-CAM1 cDNA was separated into nine exons and distributed within 17.5 kb of genomic DNA. The locations and sizes of these nine exons are shown in Figure 3(a) and Table 1. All the exon/intron junction sequences, which included eight 5′ splice junctions and several 3′ splice junctions (Table 1), followed the GT-AT rule (Breathnach and Chambon, 1981). There is an interesting correlation between the arrangement of the exons in the C-CAM1 gene and the functions of the C-CAM1 protein domains (Figure 3c). The first exon encodes the 5′-untranslated region and about two-thirds of the signal peptide. The second exon encodes the last third of the signal peptide and the first Ig domain. Exons 3, 4 and 5 code for the second, third and fourth Ig domains respectively (Figure 3c and Table 1). Exon 6 encodes the sequence following the fourth Ig domain, the transmembrane region, and the beginning of the cytoplasmic domain. The rest of the C-CAM1 cytoplasmic domain is distributed among exons 7, 8 and 9. Both exons 7 and 8 are short: 53 bp and 32 bp respectively. Exon 9 encodes the last 37 amino acids of the C-CAM1 cytoplasmic domain and the 3′-untranslated sequence. As only the coding sequence and part of the 3′-untranslated region are known (Lin and Guidotti, 1989), the exact size of exon 9 cannot be defined. However, clearly the entire coding region of C-CAM1 is distributed within nine exons.

When we compared the cDNA sequence of C-CAM3 with the C-CAM1 genomic sequence, we found that the C-CAM3 splicing pattern differed from that of C-CAM1 in that an 813 nt intron between exons 6 and 7 was not removed (Figure 3b). This unspliced mRNA contained a stop codon immediately after the regular splice junction. As a result, translation of C-CAM3 terminates at the position where the C-CAM1 and C-CAM3 sequences diverge. After the 'C-CAM3 intron', the C-CAM3 cDNA sequence continued with exons 7, 8 and 9 as found in C-CAM1. Therefore the unspliced intron was the only difference between the C-CAM1 and C-CAM3 cDNA sequences.

RNAase-protection analyses

To identify tissues that express C-CAM3 transcript, we performed RNAase-protection assays using probes that could specifically and simultaneously distinguish the C-CAM3 transcript from those of C-CAM1 and C-CAM2 (Figure 4). A ^{32}P-labelled antisense RNA probe specific for C-CAM3 was synthesized from pSK-IL60(*Bgl*II–*Bam*HI). This 409 nt antisense

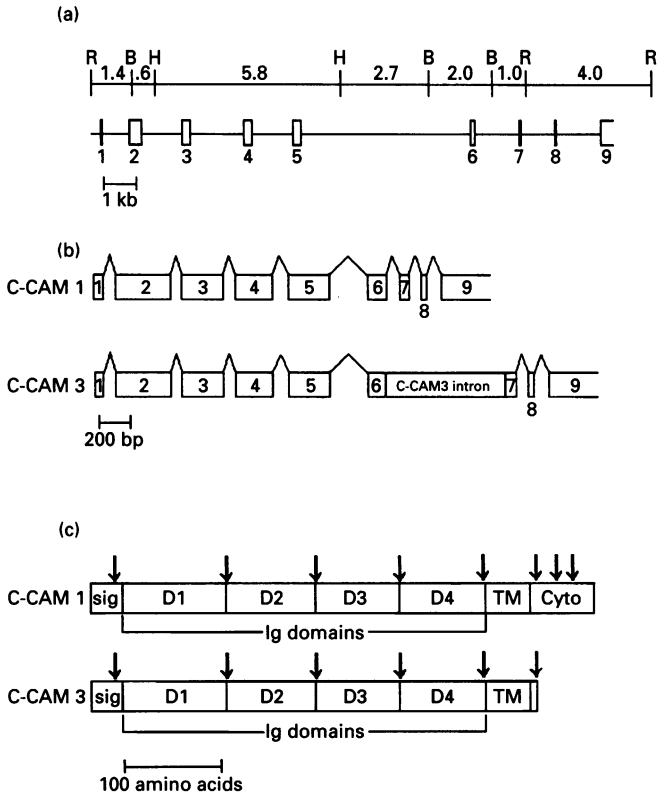


Figure 3 Structure of C-CAM1 gene and alternative splicing patterns of C-CAM1 and C-CAM3 RNA

(**a**) The restriction map and exon arrangement of the C-CAM1 genomic clone. R, *Eco*RI; B, *Bam*HI; H, *Hin*dIII. (**b**) The splicing of C-CAM1 and C-CAM3, as deduced from comparison of genomic and cDNA sequences. (**c**) Positions of intron–exon boundaries relative to the domains of C-CAM1 and C-CAM3. Arrows indicate intron–exon boundaries. Sig, signal peptide sequence; D1, first Ig domain; D2, second Ig domain; D3, third Ig domain; D4, fourth Ig domain; TM, transmembrane region; Cyto, cytoplasmic domain.

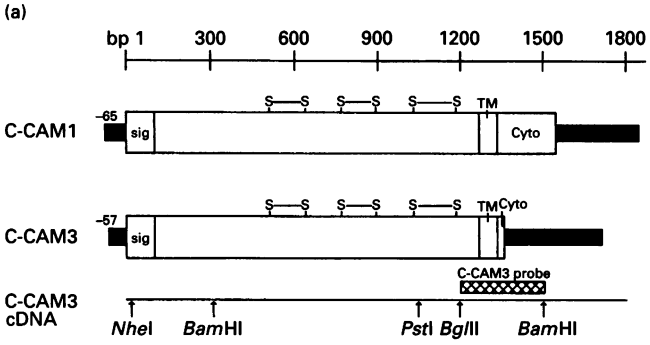
probe contained nt 1178–1490 of the C-CAM3 and a segment from the vector. Because 184 nt of this probe was complementary to the previously isolated isoforms (Figure 4b), it could protect 184 nt fragments from the C-CAM1 and C-CAM2 transcripts in addition to a 312 nt fragment of the C-CAM3 transcript. Figure 5 shows that strong signals corresponding to the C-CAM3 message were detected throughout the small intestine, including the duodenum, ileum and jejunum. Moderate signals were also detected in adult liver, the liver of a 5-day-old rat and lung. Only weak signals were detected in kidney, spleen, placenta and an established transplantable hepatocellular carcinoma (T52THC). Testis did not give a detectable signal. These results indicate that C-CAM3 was expressed predominantly in epithelial tissues, as are the other two isoforms (Figure 5) (Cheung et al., 1993). The same RNA probe also protected a 270 nt fragment with an expression pattern similar to that of C-CAM3 (Figure 5) which may be another C-CAM isoform. The band of about 214 nt (Figure 5) was probably due to non-specific priming because it also appeared in the control sample (tRNA).

On the basis of quantitative β-emission scanning by a Betascope blot analyser (Betagen), the steady-state message level of C-CAM3 (the 312 nt signal) was about 10% of that of the other two isoforms (the 184 nt signal), indicating that C-CAM3 comprised a small portion of the total C-CAM isoform messages (results not shown). As the relative steady-state message level of C-CAM1 is about half that of C-CAM2 (Cheung et al., 1993), the ratio of RNA for the three C-CAM isoforms (C-CAM1, C-CAM2 and C-CAM3) is probably 3:6:1. As C-CAM1 and C-CAM2 are obviously from different genes whereas C-CAM1 and C-CAM3 are from the same gene, this result indicates that C-CAM3 constitutes about 25% of C-CAM1 gene product.

As total RNA preparation from tissues may be contaminated by small amounts of nuclear RNA, we could not exclude the possibility that the 10% C-CAM3 message detected may have been derived from nuclear RNA. To address this we undertook the task of evaluating the abundance of C-CAM3 message in cytosolic and nuclear RNA. In order to avoid RNA degradation, it is common to use isolated cells rather than whole tissue for cytosolic RNA preparation. As C-CAM3 message was also

Table 1 Exon and domain organization of C-CAM1

Exon number	Corresponding protein domain	Corresponding position in cDNA	Length	Exon	Intron
1	5′-untranslated region + 2/3 signal sequence	−65 to 64	129	...CTG CTC ACA G Leu Leu Thr A	gtaaggagac....tccctttag
2	1/3 signal sequence + 1st Ig domain	65 to 425	361	CC TCA CTT....GTA TAC CC la Ser Leu Val Tyr Pr	gtaagtattt....ND
3	2nd Ig domain	426 to 703	278	C GCA TTA.....GTT ATC T o Ala Leu Val Ile T	gtgagtaaac....ND
4	3rd Ig domain	704 to 958	255	AT GGT CCG....GTC TTT G yr Gly Pro Val Phe G	gtaagtaggc....ND
5	4th Ig domain	959 to 1234	276	AG CCA GTG....GTA ATA C lu Pro Val Val Ile P	gtaagtgctc....ND
6	Transmembrane region	1235 to 1361	127	CT GAT CCA....ACT GGC GG ro Asp Pro Thr Gly Gl	gtaggacggt....atttatttag
7	Cytoplasmic domain	1362 to 1414	53	G GGA AGT.....AGC CAC A y Gly Ser Ser His A	gtaagtaaag....gctcccctag
8	Cytoplasmic domain	1415 to 1446	32	AT CTG GGT....CCT AAC AAG sn Leu Gly Pro Asn Lys	gtaagcgtgc....atacttgcag
9	Cytoplasmic domain	1447-ND	ND	GTG GAT GAC.... Val Asp Asp	



(b)

```
C-CAM1 1178- AGATCTCCAA CCCGGTCAGT TTCAGGATAA GTCACCCAAT CAAGCTGGAC
C-CAM2 1178- AGATCTCCAA CCCGGTCAGT TTCAGGATAA GTCACCCAAT CAAGCTGGAC
C-CAM3 1178- AGATCTCCAA CCCGGTCAGT TTCAGGATAA GTCACCCAAT CAAGCTGGAC

C-CAM1 1228- GTAATACCTG ATCCAACACA AGGAAATTCT GGCCTCTCAG AGGGTGCCAT
C-CAM2 1228- GTAATACCTG ATCCAACACA AGGAAATTCT GGCCTCTCAG AGGGTGCCAT
C-CAM3 1228- GTAATACCTG ATCCAACACA AGGAAATTCT GGCCTCTCAG AGGGTGCCAT

C-CAM1 1278- TGCAGGCATT GTGATTGGAT CTGTGGCTGG AGTGGCCCTA ATAGCAGCGC
C-CAM2 1278- TGCAGGCATT GTGATTGGAT CTGTGGCTGG AGTGGCCCTA ATAGCAGCGC
C-CAM3 1278- TGCAGGCATT GTGATTGGAT CTGTGGCTGG AGTGGCCCTA ATAGCAGCGC

C-CAM1 1328- TGGCATACTT CCTTTATTCC AGGAAGACTG GCGGGGGAAG TGACCATCGA
C-CAM2 1328- TGGCATACTT CCTTTATTCC AGGAAGACTG GCGG
C-CAM3 1328- TGGCATACTT CCTTTATTCC AGGAAGACTG GCGGGTAGGA CGGTCTTTCC

C-CAM1 1378- GATCTCACAG AGCACAAACC CTCAACCTCC AGCCACAATC TGGGTCCTTC
C-CAM2 1362-                                        ATC TGGGTCCTTC
C-CAM3 1378- TTTTTTGTTT TCTCCAGGTC TTAGTGTCAT GGTTGGGAGT GGGCAAGGAC

C-CAM1 1428- TGACGACTCT CCTAACAAGG TGGATGACGT CTCATACTCT GTCCTGAACT
C-CAM2 1375- TGACGACTCT CCTAACAAGG TGGATGACGT CTCATACTCT GTCCTGAACT
C-CAM3 1428- TTCTCTCCTG CCTCTGAGCT AGAGCACTTC CTCTCCACTC CTAAGCCTCT

C-CAM1 1478- TCAATGCCCA GC
C-CAM2 1425- TCAATGCCCA GC
C-CAM3 1478- GCTTCTCAGC AT
```

Figure 4 (a) Schematic diagram of C-CAM1 and C-CAM3 and the region used in the preparation of the antisense RNA probe and (b) sequence comparison of C-CAM1, C-CAM2 and C-CAM3 in the C-terminal cytoplasmic region chosen for the preparation of the antisense RNA probe

In (**a**) the nucleotides are numbered like those of the previously isolated C-CAM clones (Culic et al., 1992). Untranslated regions are illustrated as filled boxes. The coding regions are illustrated as open boxes.

detected in rat liver (Figure 5) and the procedure of isolating hepatocytes is established, we chose to prepare cytosolic and nuclear RNA from hepatocytes. The presence of C-CAM3 message in these two RNA preparations was further quantified by RNAase-protection assay. As shown in Figure 6 and Table 2, the C-CAM3 transcript accounts for 8–10% of total C-CAM messages in total RNA samples from liver tissue and isolated hepatocytes. In contrast, C-CAM3 transcript accounts for 0.2% and about 79% in RNA samples from cytosol and nuclei respectively. Although slight degradation of cytosolic RNA reduced the signal intensities of all C-CAM transcripts, their relative ratios should not be affected. Therefore the low percentage of C-CAM3 in cytosol and the high percentage of this

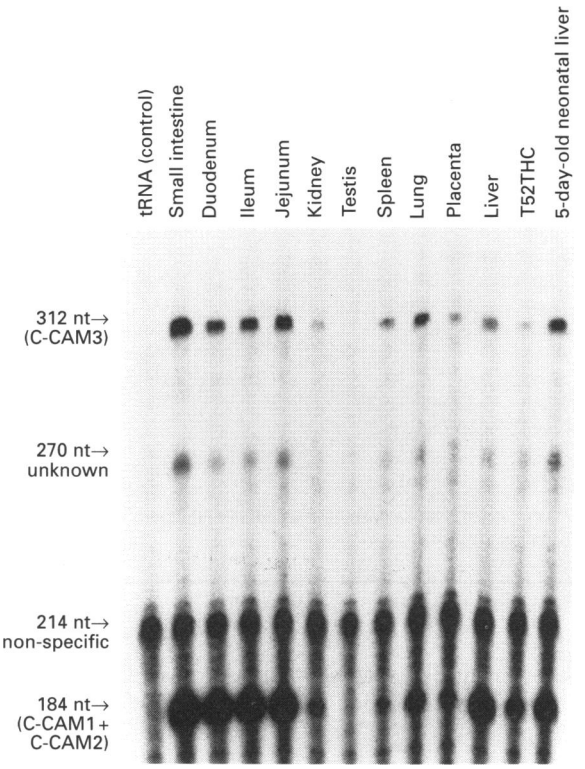


Figure 5 Expression of C-CAM3 in total RNA (10 μg) from various rat tissues as detected by RNAase protection

T52THC is a transplantable hepatocellular carcinoma cell line.

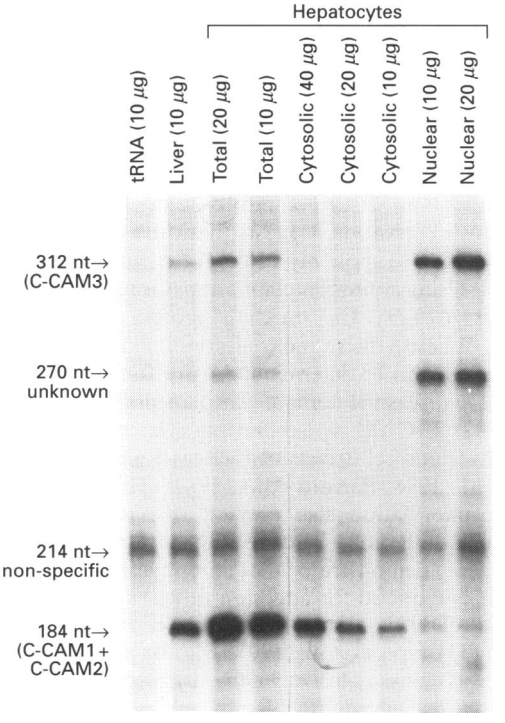


Figure 6 Expression of C-CAM3 in total, cytosolic and nuclear RNAs isolated from rat hepatocytes

Table 2 Relative abundance of C-CAM3 transcript in various RNA preparations

	C-CAM3*	(C-CAM1 + C-CAM2)*	C-CAM3 (% of total C-CAM)
Liver total RNA			
10 μg	523	4442	10.5
Hepatocyte total RNA			
20 μg	1379	16006	7.9
10 μg	1116	11247	9.0
Hepatocyte cytosolic RNA			
40 μg	12	5447	0.2
20 μg	0	2634	0
10 μg	0	1311	0
Hepatocyte nuclear RNA			
20 μg	4877	1261	79.5
10 μg	2508	723	77.6

*Numbers represent radioactivities determined with a Betascope blot analyser during 120 min periods.

message in nuclei suggest that C-CAM3 is probably an incompletely spliced product of C-CAM1 gene.

DISCUSSION

This paper reports the cloning, expression, genomic organization and functional study of a new C-CAM isoform, C-CAM3. We have shown that the expression of C-CAM3 containing only six instead of 71 amino acids intracellularly failed to show adhesion activity, suggesting that the cytoplasmic domain of C-CAM plays a significant role in adhesion function. This is in contrast with CEA and non-specific cross-reacting antigen, the cytoplasmic domains of which are not necessary for adhesion function (Benchimol et al., 1989; Oikawa et al., 1989, 1991). An important question arising from this study concerns the mechanism by which the C-CAM1 cytoplasmic domain regulates C-CAM1 adhesion function. The cytoplasmic domain of C-CAM1 could interact with cytosolic or cytoskeletal factors that modulate the intercellular adhesion mediated by the C-CAM1 extracellular domain. It was reported that calmodulin could bind to C-CAM polypeptides immobilized on nitrocellulose filters (Blikstad et al., 1992). This observation suggests that calmodulin could interact with the C-CAM1, possibly through the cytoplasmic domain of C-CAM1. Using immunoprecipitation, Lim et al. (1990) observed that a 65/67 kDa Ca^{2+}-binding protein is associated with C-CAM. Whether these molecules modulate C-CAM adhesion *in vivo* remains to be determined. A previous study showed that C-CAM1 cytoplasmic domain was phosphorylated *in vivo* (Culic et al., 1992). C-CAM is also phosphorylated by insulin receptor both in liver and cell-free system (Rees-Jones and Taylor, 1985; Accili et al., 1986; Perrotti et al., 1987). Therefore it is possible that phosphorylation also plays a role in C-CAM-mediated intercellular adhesion. Whatever the mechanism, results presented here demonstrate that the cytoplasmic domain of C-CAM1 is required for adhesion function.

Several cDNAs arising from intron retention have been reported (Thompson et al., 1991; Mattox et al., 1992). In most cases, it was not conclusively demonstrated whether the cDNA in question was indeed mRNA or unspliced heteronuclear RNA. Also the protein corresponding to the intron-retained cDNA was not identified. One of the well-studied examples of 'intron retention' during splicing is the transformer-2 (*tra-2*) gene of *Drosophila*. The *tra-2* gene regulates sex determination in *Drosophila*. In this gene, a combination of alternative promoter usage and alternative splicing generate four different mRNAs, the products of which play different functional roles in sex determination (Amrein et al., 1988, 1990; Mattox et al., 1990; Mattox and Baker, 1991). One male germ line-specific *tra-2* mRNA includes a single unspliced intron called M1 (Amrein et al., 1990; Mattox et al., 1990; Mattox and Baker, 1991). It was shown that repression of M1 splicing in the male germ line is regulated by specific polypeptide products of the *tra-2* gene itself. In the case of C-CAM, using an RNAase-protection assay, we were able to show that C-CAM3 was enriched in nuclear RNA but not in cytoplasmic RNA prepared from hepatocytes. In contrast, C-CAM1 and C-CAM2 messages were enriched in cytosolic rather than nuclear RNA. This result suggests that C-CAM1 and C-CAM2 are indeed mRNAs whereas C-CAM3 is probably an incomplete spliced mRNA despite the fact that it does contain polyadenylation signal and is polyadenylated. Consistent with this observation, we have previously demonstrated the expression of both C-CAM1 and C-CAM2 proteins in rat liver using anti-peptide antibodies (Culic et al., 1992; Thompson et al., 1993).

Although the C-CAM1 gene contains at least eight introns, only the C-CAM3 intron is retained in the polyadenylated C-CAM3 cDNA clone, indicating that removal of the C-CAM3 intron is not an efficient process. The established consensus sequences at the 3′ splice site consist of an extensive polypyrimidine tract, a pyrimidine residue immediately preceding the splice junction AG dinucleotide and a branchpoint consensus sequence, YNYURAY, within 18–40 nucleotides upstream of the polypyrimidine tract (Mount, 1982; Reed and Maniatis, 1985). A branchpoint sequence, GTCTGAT, 15 nucleotides upstream of the 3′ AG dinucleotide could be found upon close examination of the sequence within the unspliced C-CAM3 intron (Figure 1). As the most 5′ residue of that sequence is a G and not a pyrimidine residue, the branchpoint sequence is similar to but not identical with the consensus. Moreover, there was no long stretch of polypyrimidine residues between the branchpoint sequence and the 3′ AG dinucleotide. Instead a short polypyrimidine tract with two purine substitutions, ATTTATT, was found in this region. As RNA splicing is affected by the length of the polypyrimidine tract and the branchpoint sequence (Reed, 1989; Smith et al., 1989), the shortness of the polypyrimidine tract with purine substitutions and the 'imperfect' branchpoint sequence in C-CAM3 intron could account for the inefficiency of removing this intron.

C-CAM was first purified by monitoring the ability of the protein to neutralize the inhibition of hepatocyte aggregation by antibodies against liver plasma membranes (Ocklind and Obrink, 1982). The sequences as predicted from cDNAs indicate that these proteins are members of the Ig superfamily. Their sequences are very similar to those of the CEA family proteins (Aurivillius et al., 1990; Lin and Guidotti, 1989; Lin et al., 1991). CEA and its related proteins were recently shown to mediate intercellular adhesion when expressed in mammalian cells (Benchimol et al., 1989; Oikawa et al., 1989, 1991; Rojas et al., 1990). The sequence similarity between CEA and C-CAM family members supports the original observation that C-CAM molecules mediate cell adhesions. The phenotype exhibited by expression of C-CAM1 in Sf9 cells is also consistent with the notion that C-CAM proteins are involved in cell adhesion functions (Cheung et al., 1993). However, in hepatocytes, C-CAM is mainly localized at the apical surface, which is formed by the out-pocketing of membrane from two adjacent cells. Detailed localization of C-CAM by electron microscopy also revealed that C-CAM is

distributed in both canalicular and pericanalicular domains (Mowery and Hixson, 1991). These localizations suggest that C-CAM may not mediate intercellular adhesion *in vivo*; rather, it may play a role in the formation and maintenance of the specialized membrane structure of the apical surface. This study presents evidence to support the possible involvement of cytoplasmic domain in C-CAM adhesion function. Detailed characterizations of molecules interacting with C-CAM both intracellularly and extracellularly should provide clues to the *in vivo* functions of these molecules.

This work was supported by grant GM43189 (to S.-H.L.), a core grant CA16672 from the National Institutes of Health to University of Texas M.D. Anderson Cancer Center, and by grant CA42715 from the National Institutes of Health (to D.C.H.).

REFERENCES

Accili, D., Perrotti, N., Rees-Jones, R. and Taylor, S. I. (1986) Endocrinology **119**, 1274–1280
Amrein, H., Gorman, M. and Nothiger, R. (1988) Cell **55**, 1025–1035
Amrein, H., Maniatis, T. and Nothiger, R. (1990) EMBO J. **9**, 3619–3629
Aurivillius, M., Hansen, O. C., Lazrek, M. B. S., Bock, E. and Obrink, B. (1990) FEBS Lett. **264**, 267–269
Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K. and Stanners, C. P. (1989) Cell **57**, 327–334
Blikstad, I., Wikstrom, T., Aurivillius, M. and Obrink, B. (1992) FEBS Lett. **302**, 26–30
Boyer, B. and Thiery, J. P. (1989) J. Membr. Biol. **112**, 97–108
Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349–383
Cheung, P. H., Thompson, N. L., Earley, K., Culic, O., Hixson, D. and Lin, S. H. (1993) J. Biol. Chem. **268**, 6139–6146
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
Culic, O., Huang, Q. H., Flanagan, D., Hixson, D. C. and Lin, S. H. (1992) Biochem. J. **285**, 47–53

Received 15 January 1993/19 May 1993; accepted 24 May 1993

Edelman, G. M. and Crossin, K. L. (1991) Annu. Rev. Biochem. **60**, 155–190
Hansson, M., Blikstad, I. and Obrink, B. (1989) Exp. Cell Res. **181**, 63–74
Laemmli, U. K. (1970) Nature (London) **227**, 680–685
Lim, Y.-P., Dai, J. H., Josic, D., Callanan, H., Reutter, W. and Hixson, D. C. (1990) J. Cell Biol. **111**, 355a
Lin, S.-H. and Guidotti, G. (1989) J. Biol. Chem. **264**, 14408–14414
Lin, S.-H., Culic, O., Flanagan, D. and Hixson, D. C. (1991) Biochem. J. **278**, 155–161
Mattox, W. and Baker, B. S. (1991) Genes Dev. **5**, 786–796
Mattox, W., Palmer, M. J. and Baker, B. S. (1990) Genes Dev. **4**, 789–805
Mattox, W., Ryner, L. and Baker, B. S. (1992) J. Biol. Chem. **267**, 19023–19026
Melton, D. A., Krieg, P. A., Rebagaliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984) Nucleic Acids Res. **12**, 7035–7056
Mount, S. M. (1982) Nucleic Acids Res. **10**, 459–472
Mowery, J. and Hixson, D. C. (1991) Hepatology **13**, 47–56
Nicolson, G. L. (1989) Curr. Opin. Cell Biol. **1**, 1009–1019
Ocklind, C. and Obrink, B. (1982) J. Biol. Chem. **257**, 6788–6795
Odin, P., Asplund, M., Busch, C. and Obrink, B. (1988) J. Histochem. Cytochem. **36**, 729–739
Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G. and Nakazato, H. (1989) Biochem. Biophys. Res. Commun. **164**, 39–45
Oikawa, S., Inuzuka, C., Kuroki, M., Arakawa, F., Matsuoka, Y., Kosaki, G. and Nakazato, H. (1991) J. Biol. Chem. **266**, 7995–8001
Perrotti, N., Accili, D., Marcus-Samuels, B., Rees-Jones, R. W. and Taylor, S. I. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 3137–3140
Reed, R. (1989) Genes Dev. **3**, 2113–2123
Reed, R. and Maniatis, T. (1985) Cell **41**, 95–105
Rees-Jones, R. W. and Taylor, S. I. (1985) J. Biol. Chem. **260**, 4461–4467
Rojas, M., Fuks, A. and Stanners, C. P. (1990) Cell Growth Differ. **1**, 527–533
Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Smith, C. W., Porro, E. B., Patton, J. G. and Nadal-Ginard, B. (1989) Nature (London) **342**, 243–247
Summers, M. D. and Smith, G. E. (1987) Tex. Agric. Exp. Stn. Bull. 1555
Thompson, J. A., Grunert, F. and Zimmermann, W. (1991) J. Clin. Lab. Anal. **5**, 344–366
Thompson, N. L., Panzica, M. A., Hull, G., Lin, S. H., Curran, T. R., Gruppuso, P. A., Baum, O., Reutter, W. and Hixson, D. C. (1993) Cell Growth Differ. **4**, 257–268
Zinn, K., Dimaio, D. and Maniatis, T. (1983) Cell **34**, 865–879