Identification of a new isoform of cell–cell adhesion molecule 105 (C-CAM), C-CAM4: a secretory protein with only one Ig domain

Karen EARLEY*, Weiping LUO*, Yuhong QIU*, Nancy L. THOMPSON†, Janice CHOU‡, Douglas C. HIXSON† and Sue-Hwa LIN*§

*Department of Molecular Pathology, Box 89, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, U.S.A.,

†Department of Medical Oncology, Rhode Island Hospital, Brown University, Providence, RI, 02903, U.S.A., and ‡Section of Cell Differentiation, National Institutes of Health, Bethesda, MD 20892, U.S.A.

A series of Southern blot hybridization experiments using probes derived from different regions of the rat liver cell–cell adhesion molecule 105 (C-CAM) cDNA revealed the presence of a 9.6 kb *EcoRI* genomic fragment that seemed to encode a unique C-CAM isoform. An RNase protection study showed that this C-CAM transcript was expressed in placenta, spleen, lung and large intestine. In contrast, the other C-CAM isoforms, C-CAM1 and C-CAM2, are expressed in liver and small intestine. This result also suggests that the new isoform, which we named C-CAM4, was indeed encoded by a new C-CAM gene. A rat placenta cDNA library was then screened and the full-length cDNA coding for C-CAM4 was isolated. The deduced protein contained 142 amino acids and had a calculated molecular mass of 15 kDa.

C-CAM4 was composed of a leader sequence and the first V-like Ig domain typical of C-CAM-family proteins. However, C-CAM4 lacked the C-like Ig domains, the transmembrane domain, and the cytoplasmic domain found in other C-CAM isoforms. Thus, C-CAM4 is different from the other known C-CAMs in that it is a secreted protein. We have previously shown that the first Ig domain of C-CAM1 is crucial for its adhesion function. The V-like Ig domain of C-CAM4 had 92% and 89% sequence identity with the corresponding regions of C-CAM1 and C-CAM2 respectively. Together these results suggest that C-CAM4 may play a role in regulating the function of other C-CAM family proteins.

INTRODUCTION

Cell adhesion molecules (CAMs) play important roles in cell growth and differentiation. To date, many CAMs with diverse structures and functions have been characterized. It has become apparent from these studies that there are multiple isoforms of most CAMs, although the functional significance of these isoforms is not clear. In some cases, soluble isoforms have also been identified, e.g. I-CAM and carcinoembryonic antigen (CEA). The existence of these soluble isoforms, which presumably cannot mediate intercellular adhesion, raises intriguing questions regarding their functions.

Cell–cell adhesion molecule 105 (C-CAM) is a CEA-like adhesion molecule and a member of the Ig supergene family [1]. C-CAM was originally isolated from rat liver plasma membrane [2]. Subsequently, it was found in various epithelia, platelets and granulocytes [3]. During liver development, C-CAM is expressed during differentiation of hepatocytes [4,5]. In hepatocellular carcinoma, there is a consistent decrease in both protein and RNA expression of C-CAM [6,7]. These observations suggest that C-CAM may play a role in regulating hepatocyte growth and differentiation. Rosenberg et al. [8] and Neumaier et al. [9] also demonstrated that mouse and human biliary glycoprotein (BGP), a C-CAM homologue, is down-regulated in colon tumours, suggesting that reduced expression of C-CAM may be associated with epithelial malignancy. Recently we provided further evidence for the role of C-CAM in tumorigenesis by

showing that transfection of C-CAM1 into the prostate cancer cell line PC-3 suppressed the growth of PC-3 cells *in vivo* [10]. Conversely, down-regulation of C-CAM expression by transfection of a vector containing antisense C-CAM cDNA into a non-tumorigenic rat prostate epithelial cell line (NbE) significantly increased the tumorigenicity of this cell line *in vivo* [10]. These results showed clearly that C-CAM1 suppressed the growth of prostate cancer cells.

To elucidate the molecular mechanisms by which C-CAM suppresses tumour growth, it is necessary to characterize various isoforms of C-CAM and to delineate how they interact. We previously characterized three isoforms of C-CAM: C-CAM1 [1], C-CAM2 [11] and C-CAM3, which is an incompletely spliced variant found mostly in the nucleus [12]. In contrast with C-CAM3, both C-CAM1 and C-CAM2 are expressed but it is not known whether they arise from different genes or from different alleles of the same gene [13] or from alternative splicing. Further studies of these two isoforms revealed that they are expressed in the same tissues, e.g. liver and small intestine, at a constant ratio [4]. Because the functional significance of this expression pattern is not yet clear, we characterized the genes for these two isoforms and in the process, we found a previously unidentified isoform of C-CAM, which we named C-CAM4. C-CAM4 is a novel isoform of C-CAM in that it contains only the first Ig domain and lacks the other Ig domains, the transmembrane domain and the cytoplasmic domain of other C-CAM isoforms. Thus C-CAM4 is a soluble isoform of C-CAM. We have previously shown that the

Abbreviations used: CAM, cell adhesion molecule; C-CAM, cell–CAM 105; CEA, carcinoembryonic antigen; BGP, biliary glycoprotein; PSG, pregnancy-specific glycoprotein; SSC, 0.15 M NaCl/0.015 M sodium citrate.

§ To whom correspondence should be addressed.

The nucleotide sequences for C-CAM4 genomic and cDNA have been submitted to the GenBank/EMBL Data Banks and have accession numbers U23055 and U23056 respectively.

first Ig domain of C-CAM is crucial for its adhesion ability [14]. As C-CAM4 is a soluble isoform with only the first Ig domain, it may regulate the adhesion of other C-CAM isoforms. Here we report the identification and cloning of this novel C-CAM homologue.

EXPERIMENTAL

Genomic Southern blot analysis

Rat genomic DNA was prepared from rat liver by the method of Sambrook et al. [15]. The DNA (20 µg) was digested with *Eco*RI, electrophoresed on a 1.0 % agarose gel, transferred on to a Zeta Probe membrane (BioRad, Hercules, CA, U.S.A.) and hybridized to a ³²P-labelled probe generated by random priming of C-CAM1 cDNA [15].

Isolation and characterization of genomic clones

A rat genomic library (an *Eco*RI partial digest) in λ Charon 4A was purchased from Clontech (Palo Alto, CA, U.S.A.). The library was screened with labelled probe generated by nick-translation of C-CAM1 cDNA as described previously [16]. The positive clones were characterized by restriction mapping. Restriction fragments from a bacteriophage containing C-CAM genes were ligated into M13mp18 or M13mp19 (Pharmacia, Piscataway, NJ, U.S.A.), and their nucleotide sequences were determined with a T7 Sequenase kit (United States Biochemical Corp., Cleveland, OH, U.S.A.). A second rat genomic library (a *Sau*3A partial digest in λ DASHII) was purchased from Stratagene (La Jolla, CA, U.S.A.) and was screened with labelled probe generated by random priming of C-CAM1 cDNA [12]. The 9.6 kb DNA fragment that contained the C-CAM4 gene was subcloned into plasmid pSK (Stratagene) at the *Eco*RI site to create plasmid pSK-9kb.

Plasmid construction and preparation of antisense RNA probe

A 1.4 kb *Eco*RI–*Bam*HI fragment from the 5' end of the genomic clone was inserted into the *Eco*RI–*Bam*HI sites of pBluescript (Stratagene) to produce the plasmid pBS-9kb–*Bam*HI/*Eco*RI. To synthesize antisense RNA probes, plasmid pBS-9kb–*Bam*HI/*Eco*RI was linearized with *Bgl*II, and radiolabelled RNA probes were synthesized from the DNA template by using T3 RNA polymerase and [³²P]UTP as described by Melton et al. [17]. The labelled probe was 311 nt long, and it was used directly in RNase protection experiments without further purification. The RNase protection assay was performed as described previously [4].

Isolation and characterization of C-CAM4 cDNAs from a rat placenta library

A rat placenta cDNA library in λ gt10 was constructed as described by Chen et al. [18]. About 4 × 10⁵ phage plaques were screened with a 0.25 kb *Eco*RI–*Bam*HI fragment derived from the 5' end of the C-CAM1 cDNA [11] and labelled with [α-³²P]dCTP by using random primers and the Klenow fragment of DNA polymerase. After overnight hybridization at 37 °C, the filters were first washed at low stringency (2 × SSC and 0.1 % Sarkosyl for 2 h at 37 °C) and exposed to X-ray film overnight. Because there were so many positive clones (about 1200), the filters were further washed at high stringency (0.1 × SSC and 0.1 % Sarkosyl for 3 h at 60 °C) and exposed to X-ray film again. The intensity of one positive clone did not decrease after the high-stringency wash. This phage was further plaque purified. The cDNA was harvested from plate lysates with a DE-52

column by the method of Helms et al. [19] and subcloned into pBluescript SK (Stratagene). The nucleotide sequence of the double-stranded DNA was determined with a T7 sequenase kit (United States Biochemical Corp.) by the procedure provided by the manufacturer.

RESULTS

Identification of a new C-CAM gene by Southern blotting

In our previous studies on the structure of the C-CAM1 gene, we found that its coding region was composed of nine exons and was distributed over 17.5 kb of genomic DNA [12]. This genomic DNA can be cleaved with *Eco*RI into two fragments: a 13.5 kb fragment containing exons 1–7 and a 4 kb fragment containing the last two exons. Genomic Southern blot analysis with a probe generated from the full-length C-CAM1 cDNA showed that in addition to the 13.5 kb and 4 kb *Eco*RI fragments, an additional 9.6 kb *Eco*RI fragment also hybridized with the probe (Figure 1A). To identify which part of C-CAM1 hybridized to this 9.6 kb genomic fragment, the same blot was stripped of the probe and rehybridized separately with probes generated from different regions of the C-CAM1 cDNA. As shown in Figures 1(B), 1(C) and 1(D), the 9.6 kb genomic fragment only hybridized with probes generated from the 5' portion of C-CAM1 cDNA, suggesting that this fragment was similar to the first Ig domain, but not to other domains of C-CAM1. This result also suggests that there is another C-CAM-like gene.

Isolation of genomic DNA coding for the new C-CAM gene

To further characterize this new C-CAM-like gene, we screened two rat genomic libraries with a probe generated from the 5' portion of the C-CAM1 cDNA and obtained several clones. Restriction mapping showed that there were two types of clones; one group had restriction maps identical with that of the C-CAM1 gene, while the other group contained the 9.6 kb *Eco*RI fragment and the restriction maps were different from those of C-CAM1. The restriction map of this new gene (named C-CAM4) is shown in Figure 2 (upper panel). Consistent with the genomic

Figure 1 Genomic Southern blot

High-molecular-mass DNA (20 µg) isolated from rat liver was digested with *Eco*RI at 37 °C overnight, subjected to Southern blot analysis and hybridized with probes generated from the full-length C-CAM1 cDNA (A), the 5' *Eco*RI–*Bam*HI, 250 bp fragment (B), the middle *Nco*I–*Pst*II, 650 bp fragment (C), and the 3' *Pst*II–*Nsi*I, 500 bp fragment (D) portions of C-CAM1. The vertical lines indicate the positions of the markers whose sizes are given on the left.

Figure 2 Structure of the C-CAM4 gene

(Upper panel) The restriction map of the C-CAM4 genomic clone. R, *EcoRI*; B, *BamHI*; H, *HindIII*. The restriction fragment that hybridized with the 5' C-CAM1 (*EcoRI*–*BamHI*, 250 bp fragment) is marked with an asterisk. The positions of the first and second exons found in this genomic clone are shown in the expanded region. The restriction fragment that contains the third exon is marked with two asterisks. The region used in the preparation of the antisense RNA probe is also shown. (Lower panel) The nucleotide sequence of the 2 kb genomic DNA fragment that contains the first two exons of C-CAM4. The exon sequence is in capital letters and the intron sequence is in lower-case letters. The translated amino acid sequence of the two exons is also shown.

C-CAM1	65-	CCTCACTTTT	AACCTACTGG	AGCCCTCTCA	CCACTGCCCA	
C-CAM2						
C-CAM4						
C-CAM1	105-	AGTCACCGTA	GACGCTGTGC	CACCCAACGT	TGTTGAGGAG	AAGAGCGTTC
C-CAM2						.GC.....
C-CAM4						.AC.....
C-CAM1	155-	TTTACTCGC	TCACAATCTG	CCGCAGGAGT	TCCAAGTCTT	TTACTGGTAC
C-CAM2						
C-CAM4			.G.....			
C-CAM1	205-	AAGGGAACCA	CTTTGAATCC	AGACAGTGAA	ATTGCACGAT	ATATAAGATC
C-CAM2			.T.....	.AC.GG.TT	.A.....	
C-CAM4			.G.....	.CC.AA.CT	.G.....	
C-CAM1	255-	CGATAATATG	AGTAAACAG	GACCTGCATA	CAGCGGCAGA	GAGACAATAT
C-CAM2		.AG.....	.AC.....	.A.....		.T.....
C-CAM4		.GA.....	.GT.....	.CA.C.....	.G.....	.A.....
C-CAM1	305-	ACAGCAATGG	ATCC			
C-CAM2						
C-CAM4						

Figure 3 Sequence comparison of C-CAM1, C-CAM2 and C-CAM4 in the 5' region chosen for the preparation of the antisense RNA probe

Southern blot analysis, this 9.6 kb DNA fragment did not hybridize with probes generated from the middle or 3' portions of C-CAM1, suggesting that it does not contain these portions of the sequence (data not shown). We further characterized the organization of this gene by restriction mapping and by sequencing the exons and the adjacent intron regions. The portion that hybridized to the 5' C-CAM1 probe was located within a 1.4 kb *EcoRI*–*Bam*HI fragment (Figure 2, upper panel). Within this 1.4 kb fragment, two exons were identified; their locations and sizes are shown in Figure 2 (upper panel). Of the 9.6 kb of genomic DNA, we sequenced about 2 kb, including the two exons (Figure 2, lower panel). It was similar to the C-CAM1 gene; the first exon encoded the 5' untranslated region and two-thirds of the signal peptide and the second exon encoded the last third of the signal peptide and the entire first Ig domain.

The first and second C-CAM4 exons were translated (Figure 2, lower panel) and when compared with those of C-CAM1 and C-CAM2 were 92% and 89% identical respectively, at the protein level (data not shown).

Detection of C-CAM4 message by RNase protection

As the product of this gene has not yet been identified, it is not clear whether this gene is expressed and, if so, in which tissues. The 440 bp C-CAM4 sequence of the two exons had only 16 and 23 nucleotide substitutions when compared with the corresponding regions of C-CAM1 and C-CAM2 respectively. The high degree of identity suggested that it would be possible to distinguish between the isoforms by Northern hybridization but the C-CAM4 message could be isolated from those of C-CAM1 and C-CAM2 by using an RNase protection assay. The region chosen for the RNase protection probe is shown in Figure 2 (upper panel). The probe synthesized contained nucleotides 1106 (*Bgl*II site) to 1384 (*Bam*HI site) of the 9.6 kb genomic clone (Figure 2, upper panel), the most divergent part of C-CAM4, and part of the plasmid vector. The predicted coding sequence of this probe is equivalent to nucleotides 65–318 of the C-CAM1 and C-CAM2 cDNAs with several nucleotide substitutions (Figure 3). Detection of a 254 nt transcript would indicate the presence of this new C-CAM isoform in a tissue. Because of the great similarity of the three isoforms, both C-CAM1 and C-CAM2 messages should be partially protected, and fragments of about

200 nt were expected. However, the C-CAM1 and C-CAM2 messages could not be clearly distinguished using this probe.

The results of three RNase protection experiments are shown in Figure 4. Although there were slight variations in C-CAM isoform expression among the individual experiments, a consistent pattern of C-CAM4 expression in different tissues was discernible. In the first experiment, C-CAM4 was mainly expressed in lung and spleen (Figure 4, upper left panel). However, a very faint signal was also detected in RNA prepared from intestine (Figure 4, upper left panel) and RNA was prepared from different parts of the intestine for a further RNase protection study. As shown in Figure 4 (upper right panel), C-CAM4 message was not detected in RNA prepared from either the whole or parts of the small intestine. However, a small amount of C-CAM4 message was detected in placenta RNA (Figure 4, upper right panel). In the third experiment, RNA from small and large intestine was tested and we found that the small amount of C-CAM4 message detected in the first experiment was due to large intestine RNA (Figure 4, lower panel). Taken together, the RNase protection studies showed that C-CAM4 was expressed in spleen, placenta, lung and large intestine (Figure 4). The relative amounts of C-CAM4 were determined from the gel shown in Figure 4 (lower panel) with a Betascope blot analyser (Betagen Corp., Waltham, MA, U.S.A.). Large intestine, spleen and lung had signals 11-, 8- and 6-fold greater than background levels respectively, whereas the remaining tissues had levels 2-fold or less above the background reading (Table 1). Interestingly, both small intestine and liver expressed large amounts of C-CAM1 and C-CAM2 messages, as shown by the strong ~ 200 bp signals, but no C-CAM4 message was detected. On the other hand, large intestine and placenta expressed only C-CAM4, and there was no detectable C-CAM1 or C-CAM2 message. All three messages were detected in spleen and lung. This result indicates that C-CAM4 is expressed in normal tissues and that its tissue distribution is different from those of C-CAM1 and C-CAM2. During our previous cDNA cloning of C-CAM isoforms from rat liver and small intestine we did not detect C-CAM4 and the present finding confirms that C-CAM4 was not expressed in these two tissues. This result indicates that there are multiple genes in the C-CAM family and this genomic clone may code for a yet unidentified C-CAM isoform.

cDNA cloning of C-CAM4

Although the sequence of the genomic clone contained two exons, it was not clear whether C-CAM4 contained domains other than those found in C-CAM1 and C-CAM2. Since C-CAM4 message was detected in RNA prepared from placenta, we decided to clone the full-length C-CAM4 cDNA from that tissue. A rat placenta cDNA library in λ gt10 [18] was screened with a probe derived from the 5' end (*EcoRI*–*Bam*HI fragment, 250 bp) of the C-CAM1 cDNA. Under low-stringency hybridization conditions, about 1200 positive clones were detected. This abundance may have been due to the presence of a large amount of message for pregnancy-specific glycoproteins (PSGs) which are also members of the CEA family [18]. After a high-stringency wash, the intensity of most signals decreased significantly and only one remained unchanged. Sequence analysis revealed that this clone was identical with the first and second exons predicted from the C-CAM4 genomic clone, suggesting that this cDNA indeed coded for C-CAM4. The cDNA sequence and the predicted domain structure of C-CAM4 are shown in Figure 5. The entire cDNA clone was 611 bp long; it had a 82 bp 5' untranslated region, a 426 bp open reading frame and a 103 bp 3' untranslated region. An in-frame TAA stop codon was found at position

Figure 4 Expression of C-CAM4 in various rat tissues

(Upper left panel) and (upper right panel) The protected RNA fragments were separated in 4% acrylamide/urea gels as described previously [4] and exposed to X-ray film. (Lower panel) The dried gel was exposed to a phosphorimager cassette and the signal was detected with a phosphorimager (Molecular Dynamics, Sunnyvale, CA, U.S.A.). sm., small; lg., large.

427–429. This stop codon terminates translation immediately after the second exon. As a result, C-CAM4 cDNA does not code for a membrane-anchored protein. Instead, it codes for a soluble protein containing only one Ig domain. As shown in Figure 6, comparison of the C-CAM4, C-CAM1 and C-CAM2 sequences showed that this Ig domain was very similar to the V-like Ig domains of C-CAM1 and C-CAM2, with only 12 and 16 differences respectively, in 142 amino acids.

In a previous study, we found that C-CAM3, a truncated form of C-CAM1, was due to intron retention during RNA splicing [12]. Further analysis revealed that the C-CAM3 transcript was not present in cytosolic RNA but was enriched in nuclear RNA.

This observation suggested that C-CAM3 was an incompletely spliced product of C-CAM1 gene. It was possible that the early termination of C-CAM4 at the end of the second exon/intron junction was also due to incomplete splicing. When the C-CAM4 cDNA sequence was compared with its genomic sequence, it was clear that the two sequences diverged immediately after the second exon (data not shown). This finding suggested that the C-CAM4 transcript was not due to intron retention. However, the sequence of a third exon of C-CAM4 was not found in the 2.0 kb of genomic C-CAM4 DNA sequenced (Figure 2, lower panel). To test the possibility that this exon was present in the rest (about 7 kb) of the genomic DNA, an oligonucleotide (oligo 45,

Table 1 Expression of C-CAM4 in various rat tissues

Source of RNA	Relative intensity
tRNA (background control)	1
Kidney	1.7
Brain	1.7
Spleen	8.2
Lung	6.3
Heart	1.8
Small intestine	2.2
Large intestine	10.8
Liver	2.2
Testis	2.0

5'-CACTTGCATTTAATTCTCCTTTGA-3') from the 3' untranslated region of C-CAM4 was synthesized, and it hybridized to 3.5 kb and 1.4 kb DNA fragments of the genomic clone digested with *Hind*III and *Bam*HI respectively (data not shown). This result indicated that there was a third C-CAM4 exon in the 9.6 kb genomic fragment and that it was in the 1.4 kb *Bam*HI fragment, as indicated in Figure 2 (upper panel).

DISCUSSION

We have identified a new isoform of the C-CAM family by both genomic and cDNA cloning. This new isoform, C-CAM4, is a

CAM1	1-	MELASARLLR	GQIPWRGLLL	TASLLTYWSP	LTTAQVTVDA
CAM2	1-	MELASARLLR	GQIPWRGLLL	TASLLTYWSP	LTTAQVTVDA
CAM4	1-	MELASARLLR	GQIPWRGLLL	TASLLTYWSP	LTTAQVTVDA
CAM1	41-	VPPNVVEE <u>K</u> S	VLLL <u>A</u> HN <u>L</u> PQ	EFQVFYWK <u>G</u>	TT <u>L</u> NPDSEIA
CAM2	41-	VPPNVVEE <u>S</u> S	VLLL <u>T</u> HN <u>L</u> PQ	EFQVFYWK <u>Y</u>	TT <u>T</u> GLNSEIA
CAM4	41-	VPPNVVEE <u>N</u> S	VLLL <u>A</u> HN <u>M</u> PQ	EFQVFYWK <u>G</u>	TT <u>P</u> NLDSEIA
CAM1	81-	RYIRSDN <u>M</u> SK	TGPAYSGR <u>E</u> T	IYSNGSLFFQ	NV <u>N</u> KTD <u>E</u> RAY
CAM2	81-	RYIRSDN <u>T</u> SQ	TEPAYSGR <u>V</u> T	IYSNGSLFFQ	NV <u>N</u> KTD <u>E</u> GPY
CAM4	81-	RYIRSDN <u>V</u> HQ	TGPAYSGR <u>E</u> T	IYSNGSLFFQ	NV <u>T</u> KKD <u>E</u> GAY
CAM1	121-	TL <u>S</u> VE <u>D</u> QQFN	PIQTSVQFRV	YP	
CAM2	121-	TL <u>S</u> VI <u>D</u> KQFN	PIQTSVQFRV	YP	
CAM4	121-	TL <u>T</u> VI <u>D</u> QQFN	PIQTSVQFRV	YP	

Figure 6 Sequence similarity between C-CAM proteins

The amino acid sequences of the first Ig domains of C-CAM1, C-CAM2 and C-CAM4 are aligned. Amino acids that are not identical among the three isoforms are underlined and in bold.

unique C-CAM-family member in that it is a soluble protein. Although the PSG-family genes, which are mainly expressed in placenta, also code for soluble proteins, and the message for C-CAM4 was found in placenta, sequence comparison showed that C-CAM4 bore more similarities to the C-CAMs than to PSG

Figure 5 cDNA and protein sequence of C-CAM4

(Upper panel) Domain structure of C-CAM4. TM, transmembrane; cyto, cytoplasmic. (Lower panel) Nucleotide sequence and deduced amino acid sequence of C-CAM4. The possible leader peptide region is underlined and its amino acids are in italics. The arrow head marks the boundary between exon 1 and exon 2 and the arrow marks the boundary between exon 2 and exon 3.

(data not shown). Thus, C-CAM4 belongs to the C-CAM family rather than to the PSG family.

The C-CAM/BGP, PSG, CEA and related genes make up a large family of genes with structural identity [20]. However, their functions are diverse. CEA is a marker for colonic cancer [21,22]. Clinically, serum levels of CEA are used to monitor post-operative tumour recurrence [23]. The correlation between CEA level and recurrence suggests that CEA may have a role in the progression of cancer. CEA is a membrane phospholipid-anchored protein [24–27]. The protein sequence derived from cDNA also shows that, although CEA does not have a cytoplasmic domain, it does contain a transmembrane domain [28–30]. These results suggest that CEA may exist in both membrane-bound and soluble forms. It is conceivable that the elevated serum levels of CEA detected in certain malignant conditions reflect increases in the soluble form of CEA. CEA was recently shown to be an intercellular adhesion molecule that mediates the aggregation of cultured human colon adenocarcinoma cells and the adhesive properties of CEA were postulated to contribute to the development of colonic carcinoma [31]. Furthermore, using rat L6 myoblasts stably transfected with CEA as a model system of differentiation, Eidelman et al. [32] showed that fusion of L6 myoblasts into myotubes was abrogated by the ectopic expression of CEA, suggesting that expression of CEA can block terminal differentiation of myoblasts. Although this evidence suggests that membrane-anchored CEA has a causative role in tumorigenesis, the function of soluble CEA in this process and how the process is regulated are still not clear.

PSG, a major secretory product of human placenta, has been used clinically to confirm pregnancy and to predict some pregnancy-related complications [33,34]. Furthermore, PSG is also found in the sera of most patients with hydatidiform mole, invasive mole or choriocarcinoma [33,34]. These observations suggest that PSG is associated with both normal and pathological tissue growth. Although the high level of PSG produced by the placenta indicates its importance during pregnancy, its physiological role is still unknown.

C-CAM/BGP is the major transmembrane member of the CEA family. Several studies have shown that, in contrast to CEA, the level of C-CAM/BGP decreases in malignancy, including liver carcinoma [6], colonic cancer [7,8] and prostate cancer [35]. In normal adult human prostate, C-CAM/BGP is predominantly localized in the basal cell layer along the cell membrane. An overall decreased staining of C-CAM/BGP was seen in benign prostatic hyperplasia and prostatic intra-epithelial neoplasia, and C-CAM/BGP was not detected in prostate carcinomas [35]. These observations suggest that expression of C-CAM/BGP may decrease early in hyperplastic/neoplastic transformation. This, in turn, suggests that C-CAM/BGP may be a tumour suppressor. This possibility was supported by our recent finding [10] that expression of C-CAM1 in the human prostate cancer cell line PC-3 suppressed the tumorigenicity of the cells. In addition, down-regulation of C-CAM in a non-tumorigenic cell line, NbE, by transfection of the cells with antisense C-CAM1 cDNA increased the tumorigenicity of the transfected cells. The successful use of C-CAM1-expressing recombinant adenovirus to suppress tumour growth *in vivo* [36] further supports the notion that C-CAM1, unlike CEA, may suppress tumours. C-CAM/BGP and CEA therefore have opposite roles in tumour development, and the CEA-family proteins do not have a unified functional role.

Furthermore, caution is also needed in assigning similar functional roles to members of the same C-CAM subfamily. Although we have demonstrated that C-CAM1 is a tumour suppressor in the progression of prostate cancer, the function of

C-CAM2 is not clear. Because the C-CAM3 transcript is an incompletely spliced product of C-CAM1, there is probably no C-CAM3 protein product in tissues [12]. At present, the functional role of C-CAM4, a soluble protein with a different tissue distribution from the other C-CAM proteins, can only be inferred from its structural characteristics. We have previously shown that of the four extracellular Ig domains of C-CAM1, only the first, which is a V-like Ig structure, is critical for C-CAM1-mediated intercellular adhesion [14]. It was therefore very interesting to find that the sequence of C-CAM4, which has only one Ig domain, was very similar to that of the first Ig domains of C-CAM1 and C-CAM2. The high degree of identity suggested that C-CAM4 may interfere with binding (homophilic or heterophilic) between the membrane-associated C-CAM1 and C-CAM2 molecules.

Using a polyclonal antibody against C-CAM, Odin et al. [3] localized the expression of C-CAM to the apical surface of epithelial cells and also to granulocytes, macrophages and platelets. Because the polyclonal antibodies could not distinguish different isoforms, the expression and distribution of C-CAM isoforms in these tissues are not known. Studies of the expression of C-CAM isoforms have so far focused on liver [11,13] and intestine ([12]; X. Zang, O. Culic, K. Earley, N. Thompson, D. C. Hixson and S.-H. Lin, unpublished work). The expression of C-CAM1, C-CAM2 and C-CAM4 in granulocytes, macrophages and platelets has not been characterized. In the present study, we found that C-CAM4 message was detected in large intestine, spleen and lung. Judging from the amount of RNA detected in the RNase protection assays, the numbers of C-CAM4 molecules in these tissues are not high (Figure 4). It is possible that C-CAM4 may be expressed in granulocytes, macrophages and platelets where it would have different functions from those of C-CAM1 and C-CAM2, which are transmembrane proteins. Study of the expression and localization of these three isoforms in tissues will require development of isoform-specific antibodies that can distinguish the individual C-CAM isoforms. Such antibodies will also be useful for demonstrating the functional role of C-CAM4 *in vivo*. With modifications of the established protocols for generating monoclonal antibodies against peptide antigens (T.-C. Liang, W. Luo, J.-T. Hsieh and S.-H. Lin, unpublished work), it is now possible to design and generate monoclonal antibodies that can distinguish similar protein sequences. These isoform-specific antibodies will be useful for studying the functional roles of individual isoforms.

We previously cloned the C-CAM1 and C-CAM2 cDNAs, which differ in both the extracellular and cytoplasmic domains [11]. Although a few substitutions were found, it was not possible to determine whether C-CAM1 and C-CAM2 were the products of different genes or of different alleles of one gene. Based on the following observations, it is likely that C-CAM1 and C-CAM2, which are 85% identical in their first Ig domain, are from different alleles of one gene. First, we have screened two genomic libraries for C-CAM1 and C-CAM2 genes and found only the genomic clone for C-CAM1. Secondly, in the genomic Southern blot, we detected only one gene for C-CAM1 (and C-CAM2) (Figure 1). Thirdly, C-CAM1 and C-CAM2 were expressed at different levels in different rat strains [13]. Fourthly, it was reported that the many mouse C-CAM isoforms are generated from different alleles through alternative splicing [37]. On the other hand, in the present study, we found that the C-CAM4 transcript, although very similar to both C-CAM1 and C-CAM2 (92% and 89% identical respectively), was encoded by a gene distinct from C-CAM1/C-CAM2.

In summary, the discovery of a new gene coding for a unique C-CAM-family protein (C-CAM4) in rat provides new insights

into the diversity of C-CAM proteins and the molecular basis for them. This information is essential for further studies of the functions of C-CAM proteins in both normal and pathological conditions.

We thank Professor Guido Guidotti, Harvard University, Cambridge, MA, U.S.A., in whose laboratory the initial work was performed, for his support and advice during the course of this study. This work was supported by grants GM43189 (to S.-H.L.) and CA42715 (to D.C.H.) from the National Institutes of Health.

REFERENCES

- Lin, S.-H., Culic, O., Flanagan, D. and Hixson, D. C. (1991) *Biochem. J.* **278**, 155–161
- 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
- Cheung, P. H., Thompson, N. L., Earley, K., Culic, O., Hixson, D. and Lin, S.-H. (1993) *J. Biol. Chem.* **268**, 6139–6146
- Thompson, N. L., Panzica, M. A., Hull, G., Lin, S.-H., Curran, T. R., Ruppuso, P. A., Baum, O., Reutter, W. and Hixson, D. (1993) *Cell Growth Differ.* **4**, 257–268
- Hixson, D. C. and McEntire, K. D. (1989) *Cancer Res.* **49**, 6788–6794
- Thompson, N. L., Lin, S.-H., Panzica, M. A. and Hixson, D. C. (1994) *Pathobiology* **62**, 209–220
- Rosenberg, M., Nedellec, P., Jothy, S., Fleiszer, D., Turbide, C. and Beauchemin, N. (1993) *Cancer Res.* **53**, 4938–4945
- Neumaier, M., Paululat, S., Chan, A., Matthes, P. and Wagener, C. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10744–10748
- Hsieh, J.-T., Luo, W., Song, W., Wang, Y., Kleinerman, D. I., Van, N. T. and Lin, S.-H. (1995) *Cancer Res.* **55**, 190–197
- Culic, O., Huang, Q.-H., Flanagan, D., Hixson, D. and Lin, S.-H. (1992) *Biochem. J.* **285**, 47–53
- Cheung, P. H., Culic, O., Qiu, Y., Earley, K., Thompson, N., Hixson, D. C. and Lin, S.-H. (1993) *Biochem. J.* **295**, 427–435
- Edlund, M., Gaardsvoll, H., Bock, E. and Obrink, B. (1993) *Eur. J. Biochem.* **213**, 1109–1116
- Cheung, P. H., Luo, W., Qiu, Y., Zhang, X., Earley, K., Millirons, P. and Lin, S.-H. (1993) *J. Biol. Chem.* **268**, 24303–24310
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Lin, S. H. and Guidotti, G. (1989) *J. Biol. Chem.* **264**, 14408–14414
- Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984) *Nucleic Acids Res.* **12**, 7035–7056
- Chen, H., Plouzek, C. A., Liu, J.-L., Chen, C.-L. and Chou, J. Y. (1992) *DNA Cell Biol.* **11**, 139–148
- Helms, C., Graham, M. Y., Dutchik, J. E. and Olson, M. V. (1985) *DNA* **4**, 39–49
- Thompson, J. A., Grunert, F. and Zimmermann, W. (1991) *J. Clin. Lab. Anal.* **5**, 344–366
- Terry, W. D., Henkart, P. A., Colligan, J. E. and Todd, C. W. (1974) *Transplant. Rev.* **20**, 100–129
- Shuster, J., Thompson, D. M. P., Fuks, A. and Gold, P. (1980) *Prog. Exp. Tumor Res.* **25**, 89–139
- Shively, J. E. and Beatty, J. D. (1985) *CRC Crit. Rev. Oncol. Hematol.* **2**, 355–399
- Jean, F., Malapert, P., Rougon, G. and Barbet, J. (1988) *Biochem. Biophys. Res. Commun.* **155**, 794–800
- Kuroki, M., Arakawa, F., Yamamoto, H., Shimura, H., Ikehara, Y. and Matsuoka, Y. (1988) *Cancer Lett.* **43**, 151–157
- Takami, N., Misumi, Y., Kuroki, M., Matsuoka, Y. and Ikehara, Y. (1988) *J. Biol. Chem.* **263**, 12716–12720
- Hefta, S. A., Hefta, L. J. F., Lee, T. D., Paxton, R. J. and Shively, J. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4648–4652
- Zimmerman, W., Ortlieb, R., Friedrich, R. and von Kleist, S. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2960–2964
- Oikawa, S., Nakazato, H. and Kozaki, G. (1987) *Biochem. Biophys. Res. Commun.* **142**, 511–518
- Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A. and Stanners, C. P. (1987) *Mol. Cell. Biol.* **7**, 3221–3230
- Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shiota, K. and Stanners, C. P. (1989) *Cell* **57**, 327–334
- Eidelman, F. J., Fuks, A., DeMarte, L., Taheri, M. and Stanners, C. P. (1993) *J. Cell Biol.* **123**, 467–475
- Tatarinov, Y. S. and Masyukevich, V. N. (1970) *Byull. Eksp. Biol. Med.* **69**, 66–68
- Lin, T.-M., Halbert, S. P. and Spellacy, W. N. (1974) *J. Clin. Invest.* **54**, 576–582
- Kleinerman, D. I., Troncoso, P., Lin, S.-H., Pisters, L. L., Sherwood, E. R., Brooks, T., von Eschenbach, A. C. and Hsieh, J.-T. (1995) *Cancer Res.* **55**, 1215–1220
- Kleinerman, D. I., Zhang, W.-W., Lin, S.-H., Van, N. T., von Eschenbach, A. C. and Hsieh, J.-T. (1995) *Cancer Res.* **55**, 2831–2836
- Dveksler, G. S., Dieffenbach, C. W., Cardellicchio, C. B., McCuaig, K., Pensiero, M. N., Jiang, G.-S., Beauchemin, N. and Holmes, K. V. (1993) *J. Virol.* **67**, 1–8