INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

EVOLUTION OF GLYCOPHOSPHATIDYL INOSITOL ANCHORS IN CARCINOEMBRYONIC ANTIGEN FAMILY MEMBERS: FUNCTIONAL IMPLICATIONS

by

Fakhraddin Naghibalhossaini

A thesis submitted to the Faculty of Graduate Studies and Research

In partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Department of Biochemistry

McGill University

Montreal, Quebec

August, 1999

Table of contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

page

List of Tables

Chapter 1

Chapter 2

Thesis Abstract

Carcinoembryonic antigen (CEA) family members in humans can be divided into two subgroups according to the type of anchorage to the cell membrane: Glycophosphatidyl inositol (GPI)-linked and transmembrane (TM)-linked. The mode of membrane anchorage can profoundly affect the biological functions of CEA family molecules, including the ability to inhibit differentiation. All information so far, suggests that the GPI-anchored CEA family members evolved recently presumably from a more primordial TM-anchored family member such as CEACAM1. The results of this study indicate that very few mutations in the TM domain are required to effect this change. The introduction of a stop codon in the CEACAM1 TM domain at the position corresponding to that of all GPI-linked CEA family members plus two more amino acid changes in this domain resulted in highly efficiently GPI-processed protein. The GPI-anchored mutant CEACAM1 protein is still capable of mediating cell-cell adhesion but, unlike TM-linked CEACAM1, could block myogenic differentiation. Screening of genomic DNA from various mammalian groups by PCR showed convergent evolution of GPI anchorage in the CEA family, in that it evolved twice independently by different packages of mutations during the primate radiation. The same group of GPI-generating mutations as in human CEA family members evolved in the common ancestor of primitive primate groups of tarsiers and New world monkeys, and a novel group of GPI producing mutations evolved later in Cebidae family of New world monkeys. These independent groups of mutations

th[illegible] GPI-

ar[illegible] toring of

Cl[illegible] has been

po[illegible] anchor-

ge[illegible] human

Cl[illegible] protein.

Im[illegible] blocking

ca[illegible] o block

my[illegible] ntribute

to [illegible] perhaps

be[illegible]

that gave rise to the same structural feature imply adaptive evolution of GPI-anchor acquisition in CEA family. This supports the notion that GPI-anchoring of CEA family members is functionally important in vivo and therefore, has been positively selected during evolution. A chimeric cDNA of the novel GPI anchor-generating TM domain linked to the extracellular domains of human CEACAM1-4L was efficiently processed to a GPI-anchored cell surface protein. Importantly, this construct, like CEA/CEACAM6 which have a general blocking capacity on different type of cell differentiation programs, could also block myogenic differentiation completely. Thus a feature which appears to contribute to tumorigenesis in adults is nevertheless advantageous for the species, perhaps because of improved function at another stage in the life of the organism.

Résumé de thèse

Les membres de la famille de molécules de CEA (Antigène Carcino Embryonnaire) se classent en deux sous groupes dépendant du type de lien qu'ils forment avec les membranes, un lien de type Glycophosphatidyl Inositol (GPI) ou un lien de type Transmembranaire (TM). Le mode d'attache de chaque membre de la famille de molécules de CEA influe sur sa fonction biologique, incluant son effet d'inhibition de la différentiation. Les données disponibles jusqu'à date suggèrent que les membres de la famille de CEA, ayant un lien GPI, ont apparu plus récemment et sont probablement dérivés des membres plus primitifs ayant un lien TM tel que le CEACAM1. Les résultats obtenus dans la présente étude indiquent que cette transformation ne requiert que quelques mutations dans le domaine TM. L'introduction d'un codon d'arrêt dans le domaine TM de la molécule CEACAM1 au niveau de la position correspondant à celle de tous les membres de la famille de CEA ayant un lien GPI ainsi qu'un changement sur deux acides aminés de ce domaine ont permis de générer une protéine GPT hautement transformée. La protéine mutante de CEACAM1 ayant un lien GPI ne perd pas sa capacité de faciliter le processus d'adhésion intercellulaire. Cependant, contrairement à la molécule CEACAM1 TM, elle empêche aussi la différentiation myogénique. Le criblage par PCR des ADN génomiques d'une gamme de mammifères démontre qu'il y a eu une évolution convergente du lien GPI parmi les membres de la famille de CEA. Ce dernier a subi deux évolutions indépendantes effectuées par deux séries de mutations au

cours de la radiation des primates. Le groupe de mutations qui génère des molécules GPI chez l'humain a aussi évolué chez les tarsiers, un groupe de primate primitifs et les singes du Nouveau Monde. Plus tard, une nouvelle série de mutations générant des molécules GPI a apparu chez une famille de singes du nouveau monde, les Cebidae . L'évolution indépendante de ces séries de mutations donnant lieu à des changements structuraux semblables implique que la famille de molécules de CEA a subit une évolution adaptatrice au niveau de l'acquisition de lien GPI. Cette observation vient appuyer le fait que le lien GPI des membres de la famille de CEA joue un rôle important *in vivo* et qu'il a été particulièrement sélectionné durant l'évolution. Un cDNA chimérique du nouveau domaine TM ayant généré un lien rattaché aux domaines extracellulaires de la molécule CEACAM1-4L chez l'humain a été transformé de façon très efficace en protéine de surface ayant un lien GPI. Il est important de remarquer que cette nouvelle construction, tout comme celle de CEA/CEACAM6 qui est capable d'empêcher différent types de différentiation cellulaire, est en mesure d'empêcher complètement la différentiation myogénique. Cet effet, qui semble influer sur le processus de la tumurogénèse chez les adultes, apporte malgré tout un avantage pour les espèces. Ceci probablement au niveau de l'amélioration de la fonction qu'il entraîne à d'autres étapes de la vie de l'organisme.

Preface

This thesis includes the contents of original reports submitted for publication and thus, as stipulated in the "Guidelines for Thesis Preparation" from the Faculty of Graduate Studies and Research which follows, the responsibilities of all the authors of the co-authored papers must be cited.

Manuscript and Authorship:

As an alternative to the traditional thesis format, the dissertation can consist of a collection of papers that have a cohesive, unitarycharacter making them a report of a single program of research. The structure for the manuscript-based thesis must conform to thefollowing:

1. Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly-duplicated text (not the reprints) of one or more published papers. These texts must conform to the "Guidelines for Thesis Preparation" with respect to font size, line spacing and margin sizes and must be bound together as an integralpart of the thesis. (Reprints of published papers can be included in the appendices at the end of the thesis.)

2. The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges between the different papers are mandatory.

3. The thesis must conform to all other requirements of the "Guidelines for Thesis Preparation" in addition to the manuscripts.

4. As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis.

5. In general, when co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. This statement should appear in a single section entitled "Contributions of Authors" as a preface to the thesis. The supervisor must attest to the accuracy of this statement at the doctoral oral defence. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest to clearly specify the responsibilities of all the authors of the co-authored papers.

Contributions of Authors

All work described in this thesis is my own except that noted below:

Chapter 2: Co-author C. Ordoñez carried out the assay for integrin function which appears in Figure 7. Robert A. Screaton isolated the N-CAM transfectant (Fig. 7) of L6 myoblasts which was used as a control for this assay.

Chapter 3: Co-authors A. D. Yoder, M. Tobi, C. Neiswanger provided some of the Primate blood or genomic DNA samples used in the study. A. D. Yoder also made intellectual contributions to the work.

Chapter 4: Co-author C. Ordoñez performed the assay for the effect of the CMO novel GPI-encoding protein on integrin function which appears in Figure 5.

Acknowledgments

First, I would like deeply express my gratitude to Dr. Cliff Stanners whom I had the opportunity to learn much in his lab and thank for his support and nice supervision.

I would also like to thank all Cliff's lab members especially, Dr. Chris Ilantzis, luisa Demarte, and Anthony Sharpe for their collaboration and providing a friendly working environment.

I wish to thank Dr. Pelletier's lab members including Majid Ghahramani for all their assistance and Ken McDonald for cell sorting.

I would like to acknowledge the highly valued advice of my advisory committee members, Dr. Jerry Pelletier and Dr. Nahum Sonenberg.

Thanks also to Sarita Benchimol and Maureen Caron for all their valuable helps and to Dr. Petr Draber for willingness to give advice during this work.

Finally, My special thank to my wife, Afsaneh Sazegari, for her patience and support.

Thank you all.

Chapter 1

Introduction to Thesis

Literature Review

1) *Carcinoembryonic Antigen*†

Carcinoembryonic antigen (CEA), was first described by Gold and Freedman as an oncofetal protein, an antigen present in human fetal colonic tissue, and colonic carcinomas but absent in normal adult colon (Gold, 1965). However, CEA was also later discovered in human normal colonic mucosa and serum (Chu et al., 1972; Fritsche and Mach, 1977; Ilantzis et al., 1997; Sundblad et al., 1996). Elevated CEA blood levels are found in a wide variety of epithelial neoplasms including colorectal, pancreas, breast, stomach and lung (Laurence et al., 1972; Shively and Beatty, 1985). Moreover, elevated levels of serum CEA could also be seen in patients with benign liver disease and inflammatory disorders. Nevertheless, CEA has been used as a tumor marker and measurement of its serum concentration is an important parameter in the monitoring of the course of colorectal and pancreatic cancers following tumor resection (Fantini and DeCosse, 1990).

---

† Since the nomenclature used for CEA family members is not uniform, the most recent name has been used for this thesis preparation. Table 1 provide a guide to the various names used within this thesis and in the literature. More complete list of nomenclature could be found at CEA web site @ http://www.cea.uni-Freiburg.de

Table 1. Nomenclature of human CEA family

Recent name	Other names
CEA, CEACAM5	CD66e
CEACAM1	BGP, CD66a
CEACAM4	CGM7
CEACAM6	NCA-50/90, CD66c
CEACAM7	CGM2
CEACAM8	NCA-95, CGM6, CD66b
CEACAM3	CGM1, CD66d

1.1) *Size and structure of the human CEA gene family*

The human CEA gene family consists of 29 gene sequences that are tightly clustered on the long arm of chromosome 19 (Hammarström et al., 1998). The CEA family members code for highly glycosylated proteins. As a subset within the immunoglobulin gene superfamily (Paxton et al., 1987), all CEA members consist of a processed leader sequence and a variable immunoglobulin (Ig V)-like, N-terminal domain of 108-110 amino acids. This is then followed by a variable number (0-6) of immunoglobulin constant-like (Ig C) domains, usually as a pair of 92 or 96 (A domains) and 86 amino acid (B domains) (Fig. 1a) (Thompson et al., 1991). Each IgC-like domain is assumed to be stabilized by a disulfide bridge between two cysteine residues. Because of the absence of cysteine residues in some members, the N-terminal Ig-V like fold is possibly stabilized by a salt bridge between a pair of highly conserved amino acids (Arg, Asp) and/or hydrophobic interactions in these CEA family members (Thompson et al., 1989).

CEA family members can be divided into two groups: the pregnancy specific glycoprotein (PSG) group, which are secreted and the CEACAM group, which are membrane bound cell surface glycoproteins. In this survey, the latter group will be the main subject of discussion. The CEACAM group could also be divided into two subgroups according to the type of anchorage to the cell membrane: glycophosphatidyl inositol (GPI)-linked (CEA, CEACAM6,

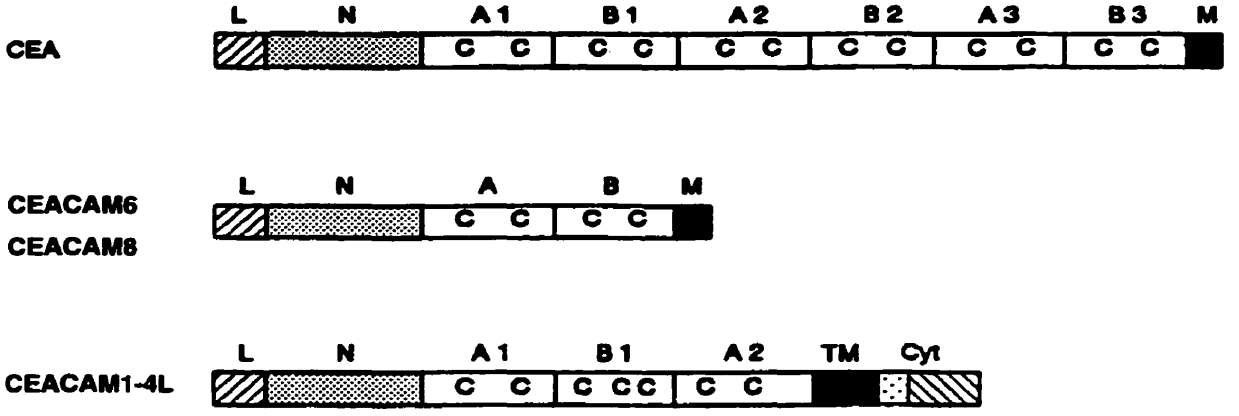
Fig. 1. Human CEA family members

a. Domain structures based on deduced exon amino acid sequence.

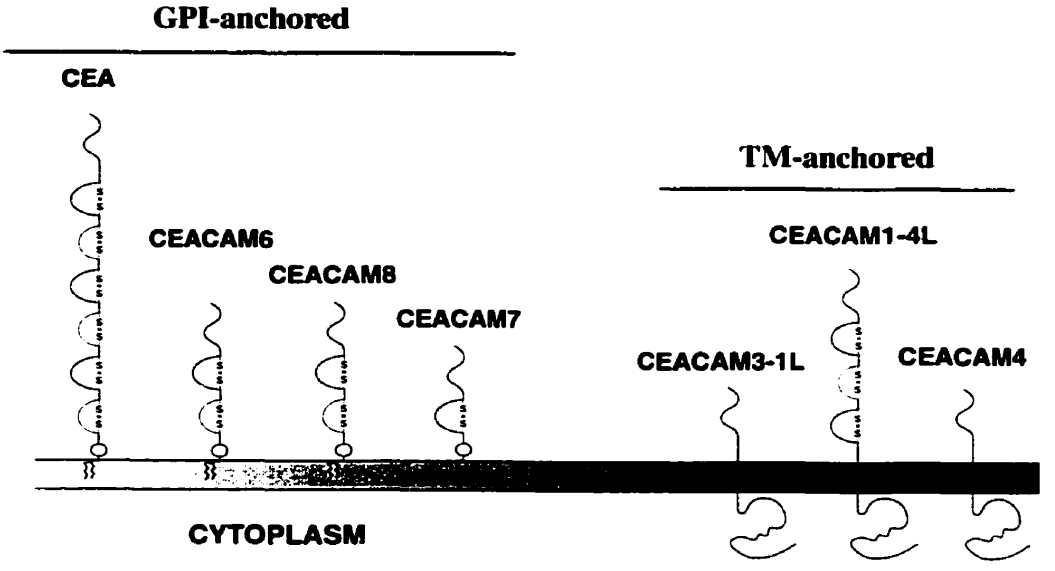
L: Leader peptide, N: N-terminal domain, A1-A3 and B1-B3: IgC like domains, TM: transmembrane domain, M: C-terminal GPI-anchoring signal peptide, Cyt: cytoplasmic domains

b. Type of anchorage to the cell surface

a


CEA
L N A1 B1 A2 B2 A3 B3 M
CEACAM6
CEACAM8
L N A B M
CEACAM1-4L
L N A1 B1 A2 TM Cyt

b


GPI-anchored
CEA
CEACAM6
CEACAM8
CEACAM7
TM-anchored
CEACAM1-4L
CEACAM3-1L
CEACAM4
CYTOPLASM

CEACAM7, and CEACAM8) and TM-linked, including CEACAM3, CEACAM4, and different splice variants of CEACAM1 (Fig. 1b) with either a short (9 amino acids) or long (71 amino acids) cytoplasmic tail (Hammarström et al., 1998).

1.2) *CEA gene family members in other species*

19 murine and 7 rat CEA family members have been characterized to date (Zimmermann, 1998). Because of rapid divergence and independent evolution, it is not possible to determine the human and rodent CEA family counterparts (orthologous genes), by comparing only their nucleotide or amino acid sequences. Rodent CEA family members are either secreted or TM-linked (i.e., CEACAM1-like) at the cell surface. So far, no GPI-linked CEA member has been discovered in rodents. The rodent secreted CEA family members also show a very different domain organization compared to human CEA family members. In this case, three or five IgV-like N-domains are followed by a single IgC-like A domain. The internal N domains are separated by leader-like domains (Thompson et al., 1991). While only one CEACAM1-like gene has been found in human and rat, mice have two CEACAM1 related genes denoted *Ceacam1* and *Ceacam2* (Nedellec et al., 1994). As in the human, alternative splicing of a 53 base pair exon also results in two isoforms with either long (71-73 amino acids) or short (10-12 amino acids) cytoplasmic tails of rodent CEACAM1 related proteins (Edlund et al.,

1993; McCuaig et al., 1993; Najjar et al., 1993). The conservation of this splicing pattern across species possibly has functional significance. Since so far only TM-linked CEACAM1-like CEA family members have been found in rodents, it has been suggested that GPI-linked CEA members appeared after the radiation of rodents in evolution (Stanners et al., 1992).

1.3) *Evolution of the CEA gene family*

The human CEA gene family is located on the long arm of chromosome 19 (19q31.2)(Brandriff et al., 1992). Murine CEA gene family members were mapped on chromosome 7 near the centromere (7A2-A3) (Rettenberger et al., 1995) which is syntenic with the human CEA gene locus on 19q. Both regions have been derived from the same ancestral chromosome region (Stubbs et al., 1996). The relative order of CEA gene family member CEACAM1 and marker genes shows an inversion in mice but overall conservation of the two loci in mouse and human (Zimmermann, 1998). Human and rodent CEA gene family members show a substantial diversification. The most obvious difference between human and rodent CEA family members is the lack of any GPI-linked CEA related protein in rodents. Since only the CEACAM1-related molecules have a conserved domain structure in human and rodents, it has been suggested that all CEA family members evolved from an ancestral CEACAM1-like gene by duplication and exon-shuffling (Zimmermann, 1998). With the exception of the CEACAM1-

related proteins, the other CEA family members in rodents and human show divergent domain organization; therefore, an independent evolution of the family members seems to have occurred in these two mammalian orders. The only domain of common type and position within all human and rodent CEA-related molecules is the N-terminal Ig V-like domain, which is possibly because of its functional importance. Indeed, it has been shown that the N-domain is required for homotypic adhesion of human CEA and CEACAM1 (Teixeira et al., 1994; Zhou et al., 1993a; Zhou et al., 1993b), of rat CEACAM1 (Wikstrom et al., 1996) and as a mouse hepatitis virus (MHV) binding receptor (Dveksler et al., 1993b).

Calculation of the rate of synonymous (Ks) and nonsynonymous nucleotide substitutions (Ka) (see section 3 of this chapter) has lead to the suggestion that the CEA family members are still evolving rapidly (Streydio et al., 1990). The Ks/Ka ratio reflects the functional constraints placed on the protein and the speed at which genes evolve (Li et al., 1985). CEA family members, in a large portion of their sequence, show a high rate of accumulation of the nonsynonymous mutations. The internal Ig C-like A domain seems to be evolving at a particularly high rate, comparable to that of pseudogenes and the mutation frequency of the N-domain exon is twice as high as that of the adjacent intron (Rudert et al., 1989). Comparing the N-domains of human CEA members with rat also suggests that there is little functional constraint on the primary amino acid sequence except for a few key

amino acids that are involved in maintaining the Ig fold (Rudert et al., 1989). It is possible that the selective pressure was not on the primary structure but on the secondary and tertiary structures and on maintaininig the open reading frame. Also the high nucleotide sequence homology of the repeated Ig C-like domains in human CEA and between CEA, CEACAM6 and CEACAM8 (>80%) and the fact that the amino acid alignment is always less than the nucleotide sequence alignment (Beauchemin et al., 1987) imply random changes in the nucleotide sequence without preservation of the actual amino acid sequence (Stanners et al., 1992). Therefore it has been suggested that CEA family members are in a state of transitional evolution and still evolving rapidly.

1.4) *Functions of CEA family members*

Although CEA has for a long time been the most widely used tumor marker in clinical medicine, its normal function(s) and relevance to malignant transformation were totally obscure, until CEA cDNA was cloned (Beauchemin et al., 1987; Oikawa et al., 1987). A number of functions were then ascribed to the CEA family members as follows.

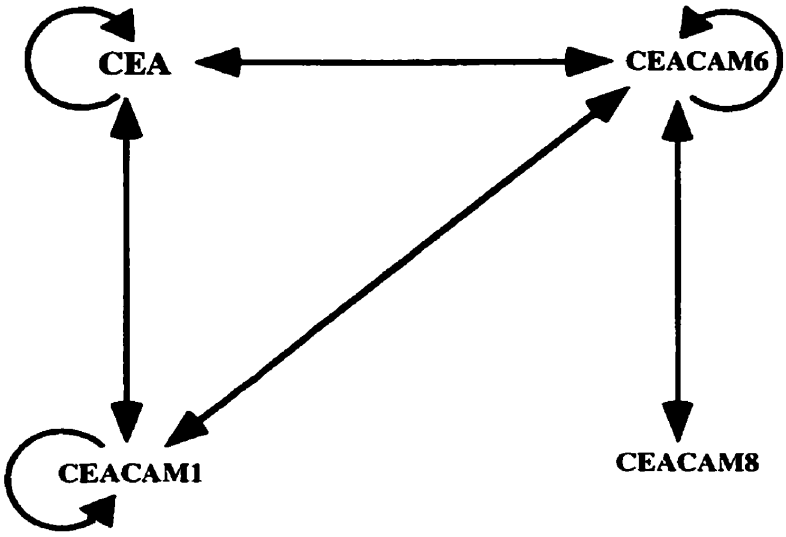
1.4.1) *Adhesion function*

In 1989, using clones of a human coloncarcinoma cell line with high or undetectable expression of CEA and a CHO cell line (LR-73) transfected with CEA cDNA, it was shown for the first time that CEA functioned in vitro, at least, as a homotypic intercellular adhesion molecule (Benchimol et al., 1989). The adhesion function was then extended to other family members, CEACAM6 (Oikawa et al., 1989; Zhou et al., 1990), CEACAM1 (Rojas et al., 1990) and CEACAM8 (Oikawa et al., 1991). Heterotypic interactions between various human CEA family members has also been described (Fig. 2) (Oikawa et al., 1991; Zhou et al., 1990). Interestingly, CEACAM8 is not capable of homotypic adhesion and can only heterotypically interact with CEACAM6 (Oikawa et al., 1991). Cell adhesion functions for the rodent (mouse and rat) homologs of human CEACAM1 have also been reported (Lucka et al., 1995; McCuaig et al., 1992; Olsson et al., 1995; Turbide et al., 1991; Wikstrom et al., 1996).

A. *Mechanism of adhesion of CEA family members*

A double reciprocal anti-parallel binding model has been proposed for CEA-CEA interactions between molecules on the surface of apposite cells (Zhou et al., 1993a). According to this model, the N domain of one CEA molecule interacts with the A3B3 internal domain of another CEA molecule

Fig. 2. Intermolecular adhesive interactions between four human CEA family members. Single headed arrows show homotypic interactions and double headed arrows show heterotypic interactions. Modified from Stanners C. P. 1998.


CEA
CEACAM6
CEACAM1
CEACAM8

and vice versa on apposite cell surfaces. cDNA constructs producing proteins bearing the N or A3B3 domains alone could not mediate homotypic adhesion but could mediate heterotypic adhesion when the transfectants were mixed. The adhesive interaction between CEA molecules could be blocked by bacterially produced fusion peptides containing the N or A3B3 domains but not the A1B1 or A2B2 domains (Zhou et al., 1993a). Experiments using various isoforms and mutants of human TM-linked CEACAM1 transfectants showed that they could only interact by the binding of N-N domains (Teixeira et al., 1994)(Keyston R, Stanners CP, unpublished). Since various splice variants of human CEACAM1, either with short (9 A.As) or long (71 A.As) cytoplasmic tails, are competent in homotypic adhesion (Rojas et al., 1996; Rojas et al., 1990; Stanners et al., 1992; Watt et al., 1994), the length of the cytoplasmic domain does not appear to influence the intracellular adhesion function, although it might be important for other functions like signaling mediated by CEACAM1 (Obrink, 1997).

In rat, as in human CEACAM1, the N-domain seems to be essential for adhesion mediated by CEACAM1 (Wikstrom et al., 1996).

1.4.2) *Role in Carcinogenesis*

Some human and rodent CEA family members show marked changes in expression in various cancers. As mentioned before, CEA and CEACAM6

are upregulated in many different tumors at the protein and/or mRNA level (Cournoyer et al., 1988; Ilantzis et al., 1997; Jothy et al., 1993; Zimmermann et al., 1988). In a cytofluorometric analysis of highly purified single epithelial cell suspensions from freshly excised carcinomas versus adjacent normal tissue, an overexpression of up to 20-fold in the mean cell surface levels of CEA and a greater overexpression (up to 70-fold) of CEACAM6 protein were demonstrated. In addition, the degree of overproduction was negatively correlated with tumor differentiation; the less differentiated and more progressed tumors showed the greatest overproduction (Ilantzis et al., 1997). The question of whether this relationship reflects only the re-expression of a embryonic antigen in cancer cells, or whether a causal relationship exists has been controversial. However, there is increasing evidence which supports the notion that CEA and CEACAM6 might have an instrumental role in tumorigenesis by the inhibition of differentiation (Eidelman et al., 1993; Screaton et al., 1997).

On the other hand, human CEACAM7 (Thompson et al., 1994) and CEACAM1 (Neumaier et al., 1993; Nollau et al., 1997) and its counterparts in rodents (Hixson et al., 1985; Hsieh et al., 1995; Rosenberg et al., 1993) have been reported to be downregulated in some tumors (see below).

CEA and CEACAM6 might promote metastasis through a mechanism that is independent of its homotypic adhesive properties. It has been proposed

that CEA molecules bind to a 80 KD receptor on Kupffer cells by the PELPK pentapeptide located at the hinge region between the N-domain and the first (A1) Ig loop of CEA molecule (amino acids 108-112) and induce cytokines release. Cytokines induce hepatic endothelial cells to express intercellular adhesion molecules and increase adhesion of the tumor cells which results in retension in the liver and hepatic metastasis from colorectal cancer (Edmiston et al., 1997; Gangopadhyay et al., 1996; Gangopadhyay et al., 1998; Gangopadhyay and Thomas, 1996).

A. *CEA and CEACAM6 inhibit cell differentiation*

It seems CEA has a general inhibitory effect on many different types of differentiation programs. Ectopic expression of CEA and CEACAM6, but not CEACAM1, results in a complete block of both the biochemical differentiation (absence of Myo D family transcriptional regulators, and myosin synthesis) and morphological differentiation (absence of fusion into multinucleated myotubes) of rat L6 (Rojas et al., 1996) and mouse C2C12 (Screaton et al., 1997) myoblasts. The block could be relieved by the administration of either N domain or A3B3 domain peptides (Eidelman et al., 1993) showing that the differentiation block depended on CEA-CEA binding.

CEA and CEACAM6 expression can also block the differentiation of retinoic acid-induced neurite extension of mouse P19 embryonal carcinomas

(Malette B, Stanners CP. in preparation) and the adipogenic differentiation of C3H10T1/2 fibroblasts and 3T3L1 adipoblasts (Demarte L, Stanners CP. in preparation). These inhibitory effects of CEA and CEACAM6 were also shown on human colonocyte differentiation in an *in vivo* assay (Illantzis et al. submitted) in which a low proportion of human colonic epithelial test cells are mixed with dissociated fetal rat colonic cells and after *in vitro* reaggregation implanted under the kidney capsule of nude mice and allowed to grow for 7-10 days. CEACAM6 overexpression produced by CEACAM6 transfection of human colorectal carcinoma cell line SW-1222, which express a low level of CEACAM6 caused them to form more aberrant dysplastic crypt-like structures and the transfected cells exhibited CEA and CEACAM6 expression all over their surface a further demonstration of lack of polarity and differentiation. In another study, the human colon carcinoma cell line, Caco2, stably overexpressing CEA and CEACAM6 at the cell surface demonstrated a marked reduction in dome formation compared to control Caco2, and collagen grown Caco-2 transfectants failed to establish a polarized monolayer and cells were found piled up in multilayered sheets, indicative of failure to establish cellular polarization (Illantzis C., Stanners C. P. submitted).

B. *CEA down-regulates E-cadherin*

The cadherin family of intercellular adhesion molecules (Takeichi, 1991) are required for the establishment and maintenance of gap junctions

between cells (Buxton and Magee, 1992; Gumbiner, 1996). E-cadherin downregulation or loss of function has been implicated in the acquisition of metastatic ability in many tumors (Birchmeier and Behrens, 1994; Takeichi, 1993). An inverse correlation between naturally expressed CEA and E-cadherin has been observed in human colorectal carcinoma cells (Yan et al., 1993) and in CEA/CEACAM6 transfected P19 cells (Malette B., in preparation), where high CEA expresser cells showed dramatically lower cell surface expression of endogenous E-cadherin. This observed mutual antagonism between CEA overexpression and cadherin downregulation possibly contributes to the transforming effect of CEA.

1.4.3) *CEACAM1 as a tumor suppressor*

There are numerous reports that human CEACAM1 and its counterparts in mouse and rat are downregulated in tumors relative to normal tissues (Hixson et al., 1985; Rosenberg et al., 1993). It seems also downregulation of CEACAM1 is an early event in the development of human prostate cancer (Kleinerman et al., 1995). Upregulation of CEACAM1 expression was also reported in some tumors such as primary lung tumors (Ohwada et al., 1994) and gastric carcinoma (Kinugasa et al., 1998). It has been shown that overexpression of rat CEACAM1 or mouse Ceacam1 isoforms with the long cytoplasmic tail, in either human prostatic carcinoma cells (Hsieh et al., 1995) or CT51 mouse colonic carcinoma cells (Kunath et al.,

1995), which were negative for endogenous expression of CEACAM1-related proteins, results in reduced growth rates compared to the parental or mock transfected cells. Similarly, in soft agar clonogenic assays these CEACAM1 expressing cells showed significant reduction of colony number (30-80%) and size. Injection of rat CEACAM1 transfected human PC-3 into nude mice (Hsieh et al., 1995) or BALB/C-derived mouse CT51 cells transfected with mouse Ceacam1 cDNA into syngenic mice (Kunath et al., 1995) showed a significant reduction in number and size of sub-cutaneous tumors (70-80%). In another study, benign rat prostate NbE cells transfected with an antisense construct of rat CEACAM1, in contrast to the parental cells, were tumorigenic in nude mice (Hsieh et al., 1995). The tumor suppressor effect was determined to be dependent on the cytoplasmic tail. Whereas both short and long tailed isoforms of mouse Ceacam1 affect the growth rate and the colony formation in soft agar to the same extent, only the long tailed isoform could inhibit tumor development in vivo (Kunath et al., 1995). It has been also demonstrated that rat CEACAM1 with the long cytoplasmic tail can prevent tumor growth of human bladder carcinomas (Kleinerman et al., 1996) and breast carcinomas (Luo et al., 1997).

1.4.4) *Signal transduction mediated by CEA family members*

There is increasing evidence that, in addition to cell adhesion, CEA family members might be involved in cell signaling events. Upon activation

of granulocytes with a variety of agents, such as formylated peptides (e. g. Formyl-Met-Leu-Phe- (fLMP) and phorbol esters (PMA)), cell surface expression of CEA family members, CEACAM1, CEACAM8, and CEACAM6 are upregulated (Ducker and Skubitz, 1992; Kuroki et al., 1992; Kuroki et al., 1995). Monoclonal antibodies against CEA family members (CEACAM1, CEACAM8, CEACAM6, and CEACAM3) augment neutrophil adhesion to human umbilical vein endothelial cells (HUVEC) (Skubitz et al., 1996). This adhesion requires exogenous Ca^{2+} and functional activity of the leukocyte-specific integrins (β2 integrins, CD11/CD18), to result in neutrophil response. Each of the CEA family members expressed on neutrophils are capable of activating neutrophils independently. The data suggest a possible signaling role for CEA family members in the activation and adhesion activity of integrins in human neutrophils.

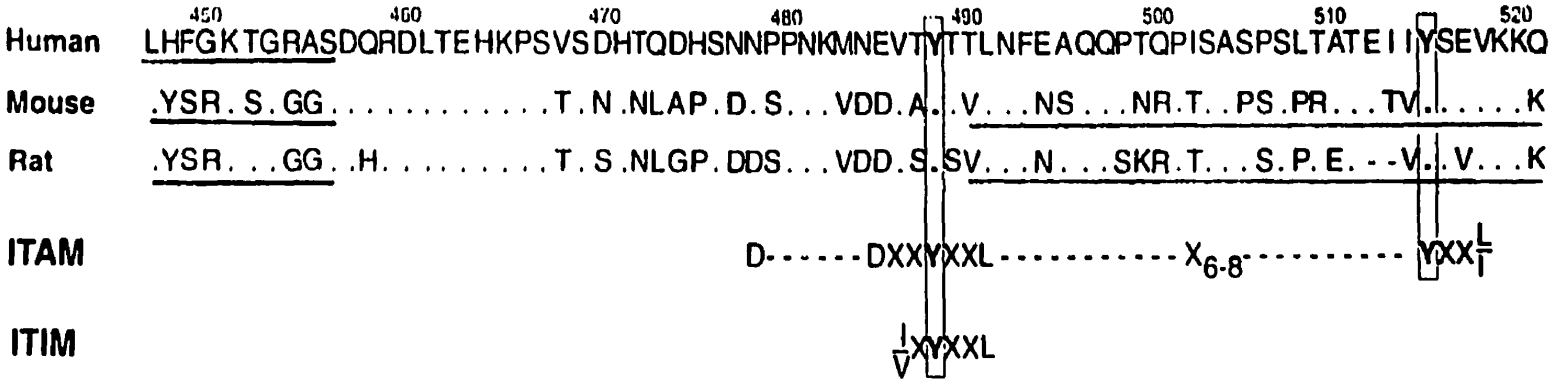
The mechanism of signaling via CEA family members is unclear. A number of possible mechanisms have been described for signaling mediated by CEA family members as follows:

A. *Phosphorylation*

The amino acid sequences of the cytoplasmic domains of mouse, rat and human CEACAM1 are highly homologous (Fig. 3). The cytoplasmic tail

Fig. 3. Amino acid sequences of cytoplasmic tail of human CEACAM1-4L and its counterparts in mouse and rat.

Underlined sequences show calmodulin binding sites (Obrink B. 1997). The Consensus sequences of ITAM and ITIM motifs are shown at the bottom. Dots represent identical amino acids. Dashes in the rat sequence represent gaps. X represents any amino acid. Critical tyrosine residues are boxed.


450
460
470
480
490
500
510
520
Human LHFGKTGRASDQRDLTEHKPSVSDHTQDHSNNPPNKMNEVTYTTLNFEAQQPTQPISASPSLTATEIIYSEVKKQ
Mouse .YSR.S.GG.....T.N.NLAP.D.S...VDD.A..V...NS...NR.T..PS.PR...TV.....K
Rat .YSR...GG..H.....T.S.NLGP.DDS...VDD.S.SV...N...SKR.T...S.P.E.--V..V...K
ITAM D------DXXYXXL----------X6-8----------YXXL/I
ITIM I/VXYXXL

of both short and long splice variants of CEACAM1 in human and rodents contain Ser and Thr residues which are potential phosphorylation sites of protein kinase A (PKA) and C (PKC). The longest splice variant of human CEACAM1 (CEACAM1-4L) has been shown to be phosphorylated mainly on tyrosine, with a lower level on serine, in granulocytes and colon cancer cells (Afar et al., 1992; Skubitz et al., 1992; Skubitz et al., 1993). Treatment of transfected cells or hepatocytes with phorbol esters leads to increased Ser/Thr phosphorylation (Afar et al., 1992; Sippel et al., 1994). The long cytoplasmic domain of CEACAM1 contains two tyrosine residues, absent from the short cytoplasmic domain. The membrane proximal tyrosine residue has been found to be phosphorylated in rat (Margolis et al., 1990) by the insulin receptor tyrosine kinase and in human cells by Src family tyrosine kinases (Brummer et al., 1995; Skubitz et al., 1995). These two tyrosine residues are part of a modified immune-receptor tyrosine activation motif (ITAM) or immunoreceptor tyrosine based inhibition motif (ITIM)-like motif (Obrink, 1997). Mutation of the proximal Tyr in rat CEACAM1 alters bile acid efflux (see next) (Sippel et al., 1994).

In other studies, association of $P^{60\ C\text{-}src}$ with CEACAM1-4L in neutrophils and the colon cancer cell line HT-29 and binding of its SH2 domain to the cytoplasmic tail of CEACAM1-4L were reported (Brummer et al., 1995). A recent study has also demonstrated that the protein tyrosine phosphatase (PTP), SHP-1, binds to the cytoplasmic tail of mouse Ceacam1

through one or both of its SH2 domains (Beauchemin et al., 1997). No evidence has been found that the SHP-1 association to Ceacam1 is required for the growth inhibitory effect of long tailed Ceacam1 (Beauchemin et al., 1997; Luo et al., 1997), and the physiological role of SHP-1 binding to Ceacam1 is not yet clear.

B. ***Calmodulin binding and dimerization***

Both short and long cytoplasmic-tailed CEACAM1 contain a calmodulin binding site close to the cell membrane (Fig.3). Rodent CEACAM also has another calmodulin-binding site at the carboxy-terminal of long cytoplasmic tail (Edlund et al., 1996). Calmodulin, which is a Ca^{2+}-binding regulatory protein (James et al., 1995), binds to rat CEACAM1's cytoplasmic domains in response to increases of intracellular Ca^{2+} concentration (Edlund et al., 1996). Other studies on both intact epithelial cells and purified rat CEACAM1 in solution have demonstrated that both long and short splice variants of rat CEACAM1 are capable of dimerization by cis-interactions (Hunter et al., 1996). Increase of the intracellular Ca^{2+} concentration leads to dissociation of CEACAM1 dimers, possibly through calmodulin binding (Edlund et al., 1996). The functional significance of calmodulin binding and dimerization of CEACAM1 protein is currently unknown.

C. Signalling mediated by GPI-anchored CEA family members

As mentioned previously, GPI-linked CEA and CEACAM6 block various types of differentiation, including L6 myogenic differentiation; however, TM-linked CEACAM1 can not. This result suggests that differentiation blocking capacity could depend upon GPI-membrane anchorage. As will be discussed in the next section, GPI-linked proteins could be involved in a form of transmembrane signaling which is different from that mediated by TM-linked proteins. There is other evidence that suggests the involvement of GPI-anchored CEA members in signal transduction events.

In the human colon carcinoma cell line, Colo 205, CEA in cold nonionic detergent-extracted cells associates specifically with $p^{56\text{-}lck}$ in a complex, but not with $p^{60\text{-}src}$, Src family protein tyrosine kinases that are expressed in these cells (Draber and Skubitz, 1998). TM-linked CEACAM1 was excluded from the complex. Because of its low background of irrelevant signaling events, the rat basophilic leukemia (RBL) cell line was used to show that CEA could in fact mediate signaling (Draber and Skubitz, 1998). These cells express protein tyrosine kinases of the src family, $p^{56\text{-}lck}/p^{53\text{-}lyn}$ and $p^{60\text{-}src}$, but again only $p^{56\text{-}lck}/p^{53\text{-}lyn}$ formed a complex with CEA in cold non-ionic detergent extracts. Signaling was shown by cross-linking CEA with specific monoclonal antibodies and consequent release of β-hexoaminidase and

RBL cells showed no

exoaminidase upon

number of other

which is discussed

d as encoding a

seems both an

asmic tail are

TPase activity

s et al., 1992),

e activity.

ported to be

CAM1. This

gulated by

tyrosine phosphorylation. CEACAM1-4L transfectants of RBL cells showed no protein-tyrosine phosphorylation and no release of β-hexoaminidase upon mAb-mediated cross-linking.

1.4.5) *Other functions of CEA family members*

In addition to what has been mentioned above, a number of other functions have been ascribed to the CEA family members, which is discussed briefly here.

A. *Ecto-ATPase and bile acid transport*

The cDNA encoding rat CEACAM1 was originally cloned as encoding a Ca^{2+}/Mg^{2+}-dependent ecto-ATPase (Lin and Guidotti, 1989). It seems both an intact amino-terminal Ig V-domain and intact long cytoplasmic tail are required for ecto-ATPase activity (Sippel et al., 1996). No ecto-ATPase activity was found for the human homologue of rat CECAM1 (Stanners et al., 1992), and there is also no report that mouse CEACAM has ecto-ATPase activity.

ATP-dependent bile acid (taurocholate) efflux was also reported to be stimulated by the long cytoplasmic splice variant of rat CEACAM1. This activity required phosphorylation of Ser503 and was regulated by

phosphorylation of Tyr488; both residues are located in the cytoplasmic domain (Sippel et al., 1994).

B. ***Bacterial and viral receptors***

CEACAM1 molecules act as receptors for both bacteria and viruses. It has been reported that E.coli and salmonella bind to the carbohydrates on CEA, CEACAM6 and human CEACAM1 (Leusch et al., 1991; Sauter et al., 1993). Recently, binding of Neisseria gonorrhea and Neissria meningitidis to the N-domain of human CEACAM1, CEACAM6, CEACAM3 and CEA via virulence-associated Opa proteins have been reported (Bos et al., 1997; Virji et al., 1996).

Several splice variants of mouse Ceacam1 and Ceacam2 encoded proteins have been shown to act as receptors for mouse hepatitis virus (MHV) (Dveksler et al., 1993a; Nedellec et al., 1994; Ohtsuka et al., 1996). It has been reported that human CEA and CEACAM1 can also function as MHV receptors (Chen et al., 1997). Mutational analysis revealed that the N-terminal IgV-like domain and at least one additional Ig-like domain are required for MHV receptor activity (Dveksler et al., 1995; Wessner et al., 1998).

2) *Glycophosphatidyl inositol(GPI)-linked proteins*

It was in the late 1970s that the covalent modification of proteins by GPI-linkage was recognized as an alternative mechanism for anchoring proteins to cell membranes (Ikezawa et al., 1976; Low and Finean, 1977). These proteins exist in all eukaryotic species including yeast, parasites, insects, and mammals. Over 100 proteins have been found to be GPI-anchored. The functions of the GPI-anchored proteins are diverse, ranging from enzymes, coat proteins and cell-adhesion molecules to receptors (Englund, 1993; Low, 1989).

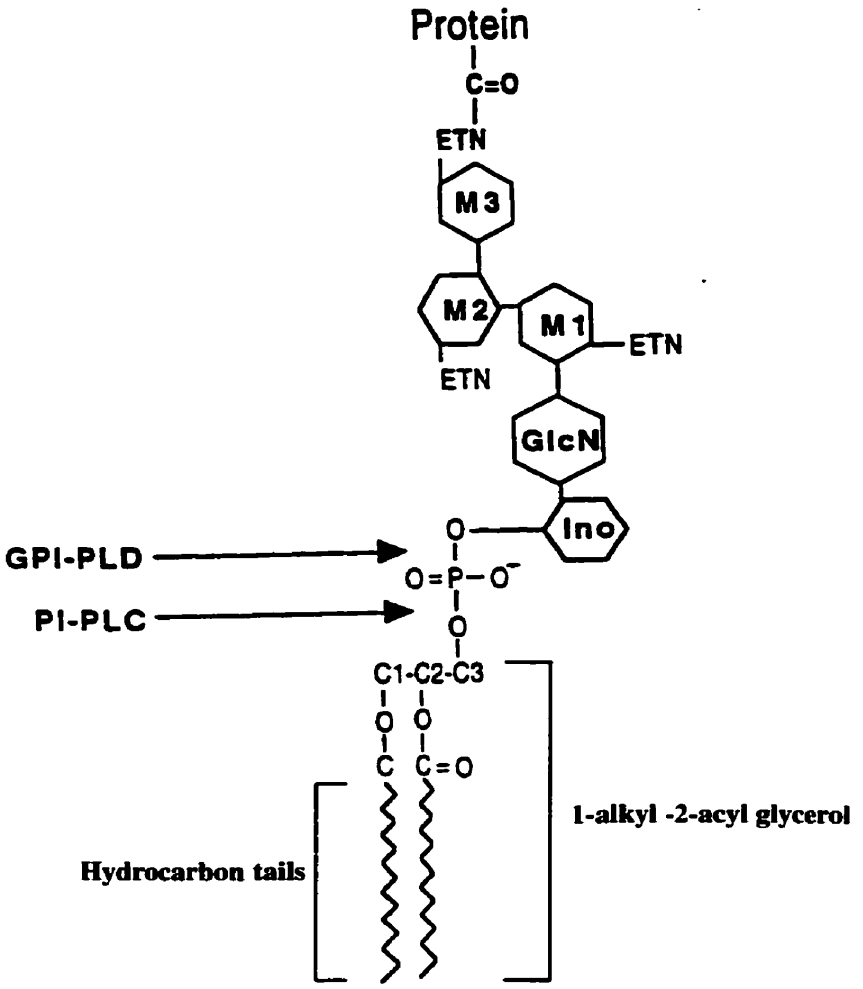
2.1) *GPI structure*

All GPI membrane anchors have an evolutionarily conserved core structure which is illustrated in (Fig. 4). The core structure of the anchor consists of a tetrasaccharide containing three mannoses (M1,2,3) and one unacetylated glucosamine (GlcN) linked to the 6-hydroxyl of phosphatidylinositol (Ino). The terminal mannose (M3) is linked via a phosphodiester bond to ethanolamine (ETN), which in turn is linked to the α-carboxyl group of the carboxy terminal residue of the protein by an amide linkage (Englund, 1993; Stevens, 1995). GPI-anchors can differ due to numerous types of side chain modifications linked to the core structure. Some anchors contain an extra fatty acid esterified to inositol (e.g. human

Fig. 4. A GPI anchor core structure of GPI-linked proteins.

The cleavage sites for two specific Phospholipases have been shown.

ETN= Ethanol amine, M= Mannose, Ino= Inositol


Protein
C=O
ETN
M3
M2
M1
ETN
ETN
GlcN
Ino
GPI-PLD
PI-PLC
O=P–O⁻
C1-C2-C3
C
C=O
Hydrocarbon tails
1-alkyl -2-acyl glycerol

erythrocyte acetylcholinestrase), which usually confers resistance to bacterial PI-PLC (Ferguson, 1992).

2.2) *GPI-anchor biosynthesis*

GPI anchoring is a post-translational modification occurring in the endoplasmic reticulum where preassembled GPI anchor precursors are transferred to proteins bearing a C-terminal GPI signal sequence (see next section).

The mechanism of mammalian GPI biosynthesis has been studied using mutant cells that are defective in different steps of the biosynthetic pathway. Important among these mutants have been GPI-defective murine T-cell lymphomas selected by their inability to express GPI-linked Thy-1 on their surfaces (Puoti et al., 1991). More than fifteen genes are involved in the mammalian GPI-anchor precursor biosynthesis and its transfer to proteins (for review see: (Kinoshita et al., 1997). The first intermediate of GPI synthesis is N-acetylglucosaminylphosphatidyl inositol (GlcNAC-PI), which is made by a transfer of GlcNAC from UDP-GlcNAC to phosphatidyl inositol (PI). The second step is deacetylation of GlcNAC-PI to form glucosaminyl-PI(GlcN-PI). The first two reaction steps of GPI biosynthesis occur on the cytoplasmic side of the ER. Thus the first and second intermediates face the cytoplasmic side. The transfer of the mature GPI-anchor to proteins is believed to occur on the

lumenal side of the ER. The step at which a post-deacetylation intermediate or the mature GPI-anchor precursors flip into the lumen is unclear at the moment.

2.3) *The GPI-anchoring signal*

The precursor of a GPI-anchored protein contains both a N-terminal signal sequence, which directs the polypeptide to endoplasmic reticulum, and also a C-terminal signal peptide which is required for GPI anchor addition. The C-terminal signal is cleaved off at the time of anchor addition. The preformed GPIs that contain an ethanolamine phosphate residue attached to mannose 3 are transferred to the carboxy-terminus of proteins by a GPI-transferase. The GPI-transferase complex is believed to act as a transamidase, to remove the GPI anchoring signal from the C-terminal of proproteins and simultaneously transfer the preformed GPI to what was previously an internal residue (the ω site) (Udenfriend and Kodukula, 1995). Two genes denoted GAA1 and GPI8, which are believed to code for two essential components of putative GPI-transamidase complex in yeast, were cloned (Benghezal et al., 1996; Hamburger et al., 1995). Recently GAA1 and GPI8 homologues in humans were also cloned (Hiroi et al., 1998; Yu et al., 1997). Alignment of the translated C-terminal amino acid sequences of cDNAs encoding different GPI-anchored proteins shows no apparent consensus sequences. Mutational studies revealed, however, that the C-terminal signal

contains a short (15-20 amino acids) hydrophobic sequence, which is preceded by a spacer containing 5-10 hydrophilic amino acids. The cleavage and attachment site to the GPI-anchor is indicated by ω, and the adjacent amino acids by ω+1 and ω+2 respectively. The ω site usually contains small amino acids: G, D, N, A, S or C. The ω+2 site similarly requires small amino acids. The ω+1 site, however, has a less stringent amino acid size requirement. Except for proline and tryptophane, almost all substitutions at the ω+1 site are effective (Udenfriend and Kodukula, 1995).

2.4) *Functions of the GPI membrane anchor*

Although various GPI-linked proteins share a common type of cell surface anchoring, they encompass a diverse family of membrane molecules. They can be involved in such processes as cell adhesion, receptor, catalysis and signaling (Review in (Low, 1989; McConville and Ferguson, 1993). The importance of GPI-anchorage in maintaining cell function and structure is not well understood. GPI-linked proteins with sphingolipids and cholestrol form a complex which is insoluble in nonionic detergents such as Triton X-100 at 4°C (Brown and London, 1997). The GPI-anchor is required for these proteins to associate with detergent-insoluble glycosphingolipid-enriched complexes (DIGs) (Brown and London, 1998). The detergent-resistant membrane microdomains form functional rafts in cell membranes which are

involved in processes such as intracellular membrane sorting and signal transduction (for review see: (Brown and London, 1998).

A. ***Intracellular sorting***

The GPI-anchor has been proposed to act as an apical targeting signal for proteins in polarized epithelial cells like MDCK (Brown and Rose, 1992). However, some other studies suggest that apart from DRM association, the GPI anchor is not required for apical targeting of the proteins (Brown and London, 1998). The apical membrane surface in cells is enriched in sphingolipids (Simons and Wandinger-Ness, 1990). Sphingolipids could self-associate from other membrane lipids by intermolecular hydrogen bond formation. Sphigolipid-cholestrol rafts might provide a site for the clustering and transporting of GPI-anchored proteins and other apically targeted proteins to the apical surface (Simons and Ikonen, 1997).

B. ***Transmembrane signalling***

It has been known for some years that cross-linking with antibodies against Thy-1 and other GPI-linked T-cell surface molecules is mitogenic and induces IL-2 production and calcium mobilization (Gunter et al., 1984; Malek et al., 1986). Cell proliferation and lymphokine production in response to cross-linking of GPI-anchored molecules require their attachment to the

plasma membrane via the GPI anchor, rather than by a transmembrane region (Shenoy-Scaria et al., 1992; Su et al., 1991). Signalling through GPI-anchored molecules in T-cells is also known to depend on the expression of the T-cell receptor (Wegener et al., 1992). It has been shown that GPI-anchored proteins expressed on lymphocytes coimmunoprecipitate with the doubly acylated Src family member tyrosine kinases including lck and fyn after treatment of cells with nonionic detergents (Stefanova et al., 1991; Thomas and Samelson, 1992) (Shenoy-Scaria et al., 1992). Therefore, it has been suggested that clustering of GPI-anchored proteins on the cell surface might result in interaction with signalling proteins thus triggering transmembrane signalling (Simons and Ikonen, 1997).

In nonlymphoid (adipocytes, epithelial and endothelial) cells, GPI-anchored proteins have been localized to small cold Triton-X insoluble, flask-shaped membrane invaginations termed caveolae. Caveolae are rich in glycosphingolipids, cholestrol, and a 22 KD marker protein called Caveolin (Anderson, 1993; Shaul and Anderson, 1998). It has been suggested that caveolae are involved in functions such as moving molecules across the cell by transcytosis, the concentration and internalization of small molecules and ions, a process called **potocytosis**, and signal transduction.

Detection of signalling molecules such as GTP-binding proteins (small and heterotrimeric), calcium mediated signalling molecules, receptor and Src

family non-receptor protein tyrosine kinases concentrated in caveolae (reviewed by Shaul and Anderson, 1998) led to the hypothesis that these glycosphingolipid domains may localize GPI anchored proteins to regions specialized for cellular signalling, which may in fact explain the mechanism by which GPI-anchored proteins activate cells (Simons and Ikonen, 1997).

Various hormones such as insulin, IL-2, NGF, TGF-β and TSH, have been reported to stimulate the release of inositol phosphoglycan (IPG) from cells. Free IPG can mimic some of the effects of these hormones (Anderson, 1994). Hydrolysis of GPI-linked proteins by endogenous enzyme in rafts has been proposed to release IPG from the GPI-proteins. The IPG molecules are then internalized and are proposed to function as second messengers in the cell (Simons and Ikonen, 1997; Varela-Nieto et al., 1996).

Clustering of GPI-anchored proteins in membrane microdomains in vivo has been controversial (Maxfield and Mayor, 1997; Mayor et al., 1994). It has been suggested that, in the absence of the clustering agents such as cross-linking antibodies or detergent treatment, they might be diffused over the plasma membrane. However, recent studies using energy transfer measurements between GPI- or TM-linked isoforms of folate receptor bound to the fluorescent analog of folic acid (Varma and Mayor, 1998) and chemical cross-linking experiments (Friedrichson and Kurzchalia, 1998) support the

existence of small clusters of GPI-linked proteins *in vivo*. This remains controversial.

C. *Other functions of the GPI anchor*

Some other features of GPI-anchored proteins have been proposed as follows:

1. A low turnover rate for GPI-anchored proteins in mammalian cells because of their exclusion from the clathrin-mediated endocytic pathway (McConville et al., 1993).

2. Regulation of their cell surface level via selective release by endogenous phosphatidyl inositol-specific phospholipases (PI-PLC and PI-PLD) (Censullo and Davitz, 1994).

3. Exchange of GPI-anchored proteins between neighboring cells by diffusion (Censullo and Davitz, 1994; Medof et al., 1996).

4. A higher mobility than TM-linked proteins at the cell surface (Edidin et al., 1991; Jacobson et al., 1987). However, recent studies revealed a confined lateral motion of GPI-anchored Thy-1 and N-CAM proteins (Jacobson et al., 1997; Sheets et al., 1997).

3) *Molecular Evolution*

Classically, evolution is defined as a process of change in the diversity and adaptation of populations of living organisms over long periods of time (Mayr, 1978). For centuries the main contributors to knowledge of evolution were biologists working at the level of whole organisms. In recent decades molecular biologists have been able to compare the genes of thousands organisms. They introduced a new way of analyzing organismal evolution and defining evolution, as a process of change in the genetic make up of populations. The molecular basis of evolution relies on the assumption that the heritable differences among organisms results from differences in their DNAs (Wilson, 1985). The process of evolution involves at least two stages: mutation altering the nucleotide sequence within the genetic material, and the selection and fixation of this mutational event into the genomes of an interbreeding population (Wilson, 1985).

3.1) *Mutation*

A basic process in the evolution of DNA sequences is the change in the nucleotide sequence with time. Mutation in the genes and the regulatory regions supplies the raw material for evolution. Mutations can be classified by the type of change caused by the mutational event as **substitutions, deletions, insertions** and **inversions**. Nucleotide substitutions are also divided into

transitions (substitution of a purine for another purine or a pyrimidine for another pyrimidine) and **transversions** (substitutions between a purine and a pyrimidine). Nucleotide substitutions occurring in protein-coding regions can also be characterized by their effect on the protein product. A **synonymous** or **silent** substitution is one that, because of the degeneracy of the genetic code, makes no change to the encoded protein sequence; a **nonsynonymous** substitution results in an amino acid replacement. Some regions of the genome, such as CG dinucleotides and short tandem repeats, are more prone to mutation (hot spots) than others (Freimer and Slatkin, 1996; Ramel, 1997; Schmutte and Jones, 1998). The high frequency of C/T changes seems to be related to methylation of cytosine in the CG dinucleotides in the genome. In eukaryotes, short tandem repeats are often hotspots for deletions and insertions, probably as a result of slipped-strand mispairing. Replication slippage most often results in short (up to 20-30 nucleotides) deletions or insertions (duplication) of DNA segments. Unequal crossing over or DNA transposition usually results in long insertions and deletions in the genome (Li and Graur, 1991c). If mutation occurs at random among the 4 nucleotides, we would expect twice the transversional changes relative to transitional changes. In practice this is not the case, and transitional changes occur more often than expected (Li and Graur, 1991b).

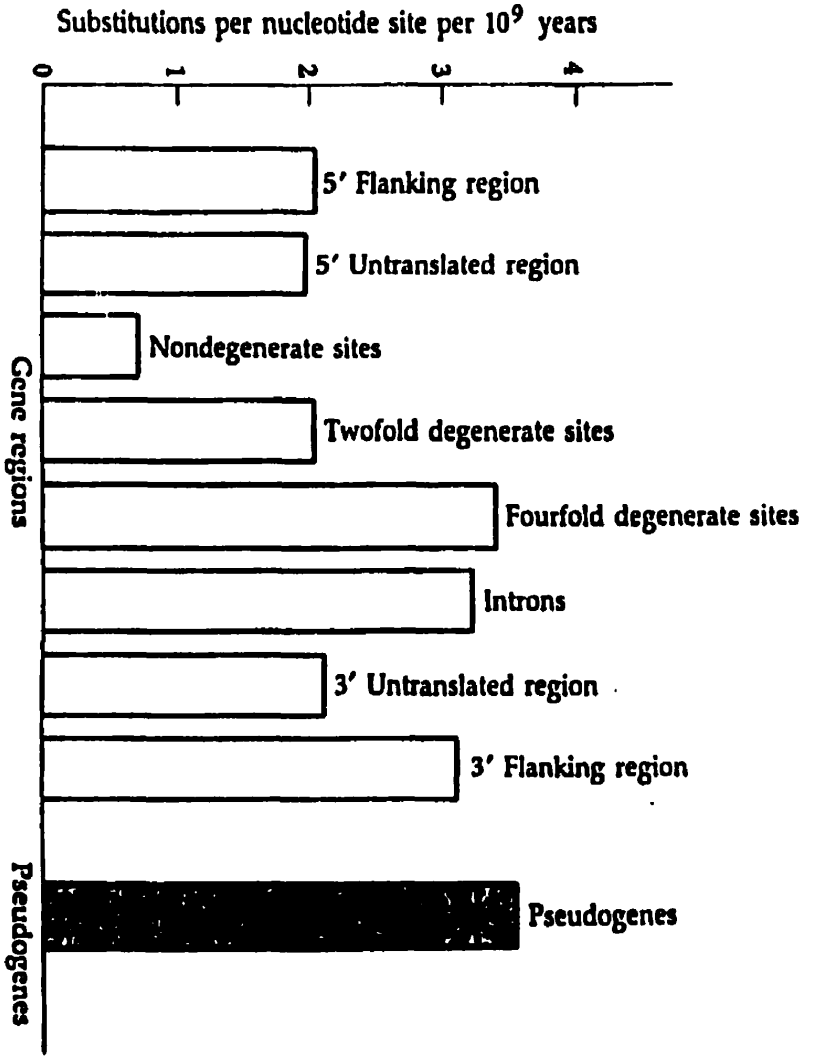
3.2) *Patterns and mechanism of nucleotide substitutions*

If we consider only coding regions of DNA, the majority of the changes are nucleotide substitutions. For the vast majority of the genes, the rate of nucleotide substitution is much higher at the third position of codon than at the first and second positions. This is apparently due to the fact that many nucleotide substitutions at the third position are silent and do not change amino acids. Studies on numerous protein coding genes in different mammalian species shows the average rate of synonymous substitution is five times the average rate of nonsynonymous substitution (Li and Graur, 1991d). The rate of nonsynonymous substitution also varies enormously from gene to gene and by different sites within a gene (Gillespie, 1991; Kimura, 1981), but the rate of synonymous substitution is much more uniform than the nonsynonymous rate.

The rate of substitution also differs between noncoding regions of the genome and between functional genes. Based on the comparison of several genes between humans and rodents (Li and Graur, 1991d), it has been shown that the substitution rate is extremely high in pseudogenes and at fourfold degenerate sites (the third position of codons for which the three possible changes always result in synonymous changes) in genes and slightly lower in introns and 3′-flanking regions (Fig. 5). Different explanations for the cause of

Fig. 5. Average rates of substitution in various regions of genome.

5′ and 3′ flanking regions of genes compared to 5′ and 3′ transcribed untranslated regions, introns, pseudogenes and sites within codons with different levels of degeneracy (the position within codons for which the three possible nucleotide substitutions result in different levels of synonymous changes). From Li W. H. and Graur D. 1991.


Substitutions per nucleotide site per 10^9 years
0
1
2
3
4
Gene regions
5′ Flanking region
5′ Untranslated region
Nondegenerate sites
Twofold degenerate sites
Fourfold degenerate sites
Introns
3′ Untranslated region
3′ Flanking region
Pseudogenes
Pseudogenes

the rate variation of substitution has been explained, which is briefly discussed as follows.

Two factors are involved in the substitution process, the incidence of a mutational event and the fixation process in which a mutant allele eventually spreads through the population to reach a frequency of 100%. For a mutant allele to increase in frequency, factors other than mutation such as selection, random genetic drift, and migration of individuals have important roles. The rate of the fixation process depends on whether the mutation is advantageous, neutral, or deleterious.

3.2.1) *Neutral theory of molecular evolution*

Observations of the high rate of evolutionary change at weakly constrained DNA positions resulted in the introduction of the neutral theory of molecular evolution (Kimura, 1968). According to the neutral theory, the great majority of evolutionary changes at the molecular level and the variability of genes within each species (gene polymorphism) are caused not by Darwinian advantageous selection but by random drift of selectively neutral or nearly neutral mutants. These mutant genes are adaptively neither more or less advantageous than the genes they replace. Neutralists believe advantageous mutations may occur, but they are so rare at the molecular level that their effect as a driving force of molecular evolution can be

neglected (Kimura, 1991). In the neutralists' view, the majority (~90%) of new mutations are disadvantageous and are eliminated by purifying selection and the fate of the rest of the mutations, which are selectively neutral (or nearly neutral), is determined by random genetic drift. The frequency of a neutral allele fluctuates, increasing or decreasing over time. The majority of mutant alleles are lost by chance, but a minority of them eventually become fixed in the population. Five principles of molecular evolution presented by Kimura and Ohta (Kimura and Ota, 1974) are as follows:

1. For each protein, the rate of evolution in terms of amino acid substitution is approximately constant/site/year for various lines, as long as the function and tertiary structure of the molecule remain essentially unaltered.

2. Functionally less important molecules or parts of a molecule evolve (in terms of mutant substitutions) faster than more important ones.

3. Those mutant substitutions that disrupt less the existing structure and function of a molecule (conservative substitutions) occur more frequently in evolution than more disruptive ones.

4. Gene duplication must always precede the emergence of a gene having a new function.

5. Selective elimination of definitely deleterious mutants and random fixation of selectively neutral or very slightly deleterious mutants occur far more frequently in evolution than positive Darwinian selection of definitely advantageous mutants.

Neutralists believe that mutational pressure and random genetic drift have the major roles as causes of evolutionary changes. Some proteins or parts of a protein are functionally more "constrained" than others. Therefore, the rate of mutation in tightly constrained proteins is lower than that in loosely constrained proteins (Li et al., 1985). It seems that this idea is now commonly accepted by molecular biologists in that variable regions of genes among various species are considered unimportant whereas conserved regions of genes are considered functionally important.

3.2.2) *Synthetic theory*

The above view that molecular evolution is the result of accumulation of neutral mutations is in sharp contrast to the neodarwinian selection and adaptive view. According to the synthetic theory, gene substitutions occur as a consequence of selection for advantageous mutations. Selectionists agree with neutralists that most new mutations are deleterious and are quickly removed from the population. However, unlike neutralists, they believe that most of the nondeleterious mutations which contribute to the process of evolution are advantageous and very few are neutral. Selectionists consider environmental conditions to be the major determinants of evolution. They have maintained that there should be a correlation between environmental variability and genetic variability. In the selectionists view, environmental change is the cause of the observed rate variation of protein evolution

(Gillespie, 1991). Slowly evolving proteins such as histones function in a relatively constant environment (intra-nucleus). However, fast evolving proteins, like immunoglobulins, have evolved in response to a rapidly changing environment (pathogens).

Although adaptive evolution could be easily found at the phenotypic level of organisms, molecular adaptation has been difficult to demonstrate. The majority of evolutionary changes could be explained by the neutral theory of molecular evolution. Only a few cases of adaptive evolution have been identified at the molecular level (McDonald and Kreitman, 1991; Messier and Stewart, 1997).

Two major types of sequence tests are currently used to identify adaptive evolution in proteins: sequence convergence, and neutral rate violation (reviewed by (Kreitman and Akashi, 1995). Molecular convergence or parallelism are due to the same changes in proteins having been derived along different evolutionary lineages ✿.g. gut lysozyme evolution in cow and longur monkey, (Messier and Stewart, 1997)✿ Functional convergence through independent substitutions in different lineages are often considered an indication of positive selection or adaptation.

Many molecular evolutionary analyses, particularly those attempting to detect adaptive evolution, rely on distinguishing between synonymous

and nonsynonymous differences in DNA sequences. As discussed before, purifying selection eliminating disadvantageous mutations leading to amino acid substitution is the most prominent form of selection on proteins. Therefore, in most proteins the neutral (synonymous) nucleotide substitution rate (K_S) is much higher than the nonsynonymous rate (K_A). The cases with a higher nonsynonymous rate ($K_A > K_S$) or with an accelerated rate of amino acid replacement are taken as examples of adaptive evolution (Kreitman and Akashi, 1995; Sharp, 1997). The higher rate of nonsynonymous to synonymous substitution rates ($K_A > K_S$) is a stringent criterion. Only in rare cases can positive selection raise the value of K_A for a whole protein to values higher than K_S. An alternative way for showing adaptive changes in a molecule through amino acid substitution rate analyses is to concentrate on a specific domain which we already know is functionally important, thereby excluding other regions which are subject to purifying selection (Sharp, 1997).

3.3) *Evolution of multigene families*

The traditional view concerning the creation of gene families includes serial duplication of entire genes and divergence of duplicated (paralogous) genes because of relaxation of the evolutionary pressure on its sequence (Henikoff et al., 1997; Li and Graur, 1991a). Continual duplication and deletion is a feature of all gene clusters (Lewin, 1994a) and acceleration of

amino acid substitution versus synonymous nucleotide substitution is often found to be related to gene duplication (Ohta, 1991). Duplicated genes might evolve in several ways. The copies may retain their original function; this will result in production of a larger quantity of the same RNA or protein products. However, the duplicate genes usually diverge independently, either acquiring new functions or becoming inactive pseudogenes (Li and Graur, 1991a). Whether acceleration of amino acid substitutions in multigene families is the result of positive natural selection or neutral evolution because of relaxation of selective constraints on duplicated, redundant genes, is controversial (Kreitman and Akashi, 1995).

3.3.1) *Exon shuffling*

DNA shuffling is a powerful process for evolution which generates diversity by recombination. Molecular shuffling can efficiently mix sequences from different genes to combine useful mutations from individual genes and therefore accelerate the evolution process. There are two ways of exon shuffling: exon duplication and exon insertion (Li and Graur, 1991a). Exon duplication is one of the most important processes involved in gene elongation. The insertion of an exon from one gene into another, results in production of mosaic or chimeric proteins (Doolittle, 1989).

3.3.2) *Concerted evolution of a gene family*

The members of a multigene family in a species are relatively free to diverge . However, they generally evolve together as a unit (Ohta, 1991). Although the nucleotide sequences of a multigene family can change over time the homogeneity of nucleotide sequence among members is maintained through genetic interaction among its members (Li and Graur, 1991a; Ohta, 1991). This phenomenon is called concerted or horizontal evolution. A variant repeat could be propagated by unequal crossing over and gene conversion to all gene family members. Gene conversion is a unidirectional recombination event in which a small region of one chromosome is replaced with the corresponding sequence of the homologous chromosome. If an unequal exchange takes place between two members of a gene family through the creation of heteroduplex DNA, any mismatches within heteroduplex might be repaired. The result is conversion of one of the nonallelic genes to the sequence of the other (Lewin, 1994b). The divergent member in a gene family might be deleted by unequal crossing over or to be converted to the conserved copy by gene conversion. Therefore, divergence of duplicated genes in a gene family slows down. Under concerted evolution the paralogous members of a family might become more similar to each other within one species than orthologous family members from other, even closely related, species.

In addition to unequal crossing over and gene conversion other processes such as transposition of mobile elements and slippage of DNA polymerase during replication can cause gains or losses of variant genes in a family (Li and Graur, 1991a).

4) *Aims and scope of the present work*

CEA gene family members in humans can be divided into two groups according to the type of anchorage to the cell membrane: GPI-linked and TM-linked. All information so far, including the close nucleotide sequence alignment between CEA members and the fact that to date only TM CEACAM1-like family members have been discovered in rodents, suggests a recent evolution for GPI-linked members. Since the normal function of CEA members *in vivo* is not well understood, the question of whether the acquisition of GPI-anchorage in CEA family members during evolution had any functional advantage seems of interest and was the subject of this study. The observations and experimental findings presented in this thesis for the first time provide detailed information regarding the evolution of GPI-linkage in CEA family members. This could represent a paradigm for such evolution in the much larger and functionally diverse Ig superfamily.

In chapter 2, it has been shown how, by only a few mutations, a GPI-linked CEA family member could evolve from the TM-linked, more primordial CEA family member, CEACAM1. The key mutations in the TM exon of the latter required to shift the mode of membrane linkage from TM to GPI were determined. The results indicate that the introduction of a stop codon in the CEACAM1-4L TM domain, at the corresponding position to the stop codon naturally found in the C-terminal domain of GPI-linked members plus two more amino acid substitutions in this exon could convert TM linked CEACAM1-4L protein into an efficiently GPI-processed protein. This GPI-linked mutant CEACAM1 protein, like normal CEA family members, could still function as an intercellular adhesion molecule *in vitro*. But interestingly, unlike CEACAM1, its expression blocked the myogenic differentiation of rat L6 myoblasts and perturbed the binding of L6 cells to fibronectin in much the same way as CEA and CEACAM6, thus implying the same effect as CEA/CEACAM6 on the function of the $\alpha_5\beta_1$, the integrin mainly responsible for binding to fibronectin.

On the basis of the assumption that the above mentioned stop codon and processing efficiency mutations are indicative of GPI-linkage for CEA family members, primers specific for the TM exons of the CEA family were designed and applied for PCR analysis of genomic DNA of different primate and nonprimate species. Chapter 3 provides the DNA sequence data showing convergent and adaptive evolution of acquisition of GPI-anchorage in CEA

family members. The results indicate that the mode of generating a stop codon found in humans arose in a common ancestor of Tarsisus and New world monkeys, whereas a second novel mechanism of stop codon generation evolved independently later in the Cebidae family of new world monkeys during the primate radiation. The fact that GPI linkage evolved twice independently supports the notion that GPI-linked CEA family members have important functions *in vivo*.

CEA and CEACAM6, in a model system of differentiation, can block the terminal myogenic differentiation of rodent myoblasts, but GPI-linked N-CAM and TM-linked CEACAM1 can not. In chapter 4, the effect of the novel molecular mechanism of GPI-linkage that evolved in the CEA family of new world monkeys on cellular differentiation are reported. Experimental evidence is presented that the novel mechanism for generating GPI-linked sequences in the CEA family that arose during evolution has functional implications in that it produces proteins that block myogenic differentiation possibly through observed modification of integrin-ECM interaction

Chapter 2

A Paradigm for Evolution of GPI Membrane-linkage in the Carcinoembryonic Antigen Family, a Subset of the Ig Superfamily

Abstract

GPI membrane-anchored CEA family members CEA and CEACAM6 are up-regulated in human cancer, whereas TM-anchored family member CEACAM1 is down-regulated. In addition, CEA and CEACAM6 have been shown to inhibit differentiation and increase tumorigenicity of various cell types, whereas CEACAM1 has no effect. The chief structural determinant of these carcinogenic effects of CEA/CEACAM6 is their GPI anchor. GPI-anchored CEA family members evolved in the primate radiation, presumably from a more primordial TM-anchored family member such as CEACAM1, which is found in more primitive species. It was of interest, therefore, to examine the feasibility of derivation of GPI linkage from the CEACAM1 TM domain by mutation, using the known differences between the TM domain of CEACAM1 and the corresponding domains of GPI-linked CEA family members as a guide. The results indicate that very few mutations are required to effect this change. The introduction of a stop codon in the CEACAM1 TM domain at the position corresponding to that of all GPI-linked CEA family members gave a GPI-linked protein that was, however, inefficiently processed. Two other mutations increased the processing efficiency to the high levels of naturally occurring CEA family members. The GPI-linked mutant CEACAM1 demonstrated a dramatic change in function: unlike normal CEACAM1 it now blocked myogenic differentiation. The

results support a paradigm for the evolution of efficient GPI anchorage in this subset of the Ig superfamily.

Introduction

The human Carcinoembryonic antigen (CEA) gene family, a subfamily within the immunoglobulin gene superfamily (Paxton et al., 1987), can be divided into two subgroups according to the type of anchorage to the cell membrane: glycophosphatidyl inositol (GPI)-linked (CEA, CEACAM6, CEACAM8, CEACAM7) and transmembrane (TM)-linked (CEACAM1, CEACAM3, CEACAM4). Members of both subgroups function in vitro, at least, as intercellular homotypic and heterotypic adhesion molecules (Benchimol et al., 1989; Oikawa et al., 1991; Oikawa et al., 1989; Rojas et al., 1990). These two subgroups show opposite molecular properties in several aspects. CEACAM1-mediated intercellular adhesion involves homophilic binding between N terminal domains (Keyston and Stanners, unpublished)(Teixeira et al., 1994) whereas CEA-mediated intercellular adhesion requires double reciprocal heterophilic binding between N terminal and internal domains of apposed antiparallel molecules (Zhou et al., 1993a). Also, unlike GPI-linked members CEA and CEACAM6, which are up-regulated in colorectal carcinomas (Cournoyer et al., 1988; Ilantzis et al., 1997; Jothy et al., 1993), TM-linked members tend to be down-regulated (Neumaier et al., 1993; Nollau et al., 1997). Recent studies suggest that GPI-anchored proteins may be clustered in detergent insoluble membrane microdomains (Friedrichson and Kurzchalia, 1998; Varma and Mayor, 1998) which can

activate signaling pathways that are different from those activated by TM anchored proteins (Casey, 1995; Robinson, 1997; Simons and Ikonen, 1997).

Furthermore, we have shown that the ectopic expression of GPI-linked CEA and CEACAM6 can block the differentiation of rat L6 and mouse C2C12 myoblasts (Eidelman et al., 1993; Screaton et al., 1997), while CEACAM1 cannot (Rojas et al., 1996). In fact, a chimeric cDNA construct consisting of the human CEACAM1-4L external domains linked to the CEA carboxy-terminal domain, which is processed to give a GPI-linked protein product, also blocks myogenic differentiation (Screaton, R.A. and Stanners, C.P., in preparation). In recent work, it has also been demonstrated that CEA and CEACAM6 but not CEACAM1 expression can inhibit retinoic acid induced neurite extension and neurofilament expression by P19 embryonal carcinoma cells (Malette, B. and Stanners, C.P., in preparation), adipogenic differentiation of C3H10T1/2 fibroblasts and 3T3L1 adipoblasts (Demarte, L and Stanners, C.P., in preparation) and polarization and differentiation of colonocytes (Ilantzis and Stanners, 1997) (Ilantzis, C. and Stanners, C.P., submitted) . The mechanism for this pan effect, in L6 myoblasts, P19 cells and colonocytes at least, appears to involve a CEA-mediated perturbation in the function of specific integrins and consequently in the binding to extracellular matrix components, upon which many differentiation programs depend (Ordoñez et al., submitted). It appears, therefore, that CEA has a general inhibitory effect on many different types of differentiation programs and that

and that the mode of membrane linkage is the chief molecular determinant of this effect.

The close nucleotide sequence alignment between GPI-linked CEA family members (higher than 90% in some cases) and the fact that to date only TM CEACAM1-like family members have been discovered in mouse and rats, while both GPI and TM-linked family members have been found in humans, suggests a recent evolution of GPI-linked members from a more primordial TM-linked ancestral member. The CEA gene family members thus represent a naturally occurring model for the evolution of a GPI-anchored proteins in the Ig superfamily. It appears likely that, in the course of evolution, gene and domain duplications of a primordial TM-linked CEA family member gave rise to the rest of the family: a group of 29 CEA related genes and gene-like sequences, including genes coding for GPI linkage, that are clustered on human chromosome 19 (Hammarström et al., 1998).

Mutational studies on GPI-linked proteins, especially Decay Accelerating Factor (DAF) and Placental Alkaline Phosphatase (PLAP), indicates that the signal required for GPI anchor addition resides completely in the C-terminal domain, which is cleaved in the process (Englund, 1993). These signals comprise a short stretch of hydrophobic residues (8-21 amino acids) at the C-terminal end, a hydrophilic spacer region (usually 5-10 residues) and, finally, the cleavage/attachment site to the GPI anchor, denoted

the ω-site. The ω and ω+2 residues are both critical for anchor addition and are usually small amino acid residues.

In this work, we have shown how, by only a few mutations, CEA family members with a GPI membrane linkage could have evolved from the TM-linked, presumably more primordial, CEA family member, CEACAM1. We determined the key differences in the transmembrane exon of the GPI-linked family members that are required to shift the mode of membrane linkage from TM to GPI. We demonstrate that the introduction of a stop codon in the CEACAM1-4L (the longest splice variant of CEACAM1 with a cytoplasmic tail containing 73 amino acids) TM domain, at the corresponding position to the stop codon naturally found in the C-terminal domain of GPI-linked members, could alone confer GPI linkage on CEACAM1-4L, although not efficiently. By substituting another amino acid at the ω+2 site just downstream of the cleavage site of the TM domain and one further substitution in the hydrophilic spacer region, TM-linked CEACAM1-4L could be converted to a completely efficiently processed GPI-linked construct. Importantly, the new CEACAM1-4L GPI-linked construct was also radically altered in its biological function in that, like naturally occuring GPI-linked CEA family members, it could now affect integrin function and block myogenic differentiation.

Matherials And Methods

Mutagenesis - PCR based mutagenesis using a 62-mer mutant antisense and a 27-mer sense primer (Table 1) was applied to introduce a deletion and a substitution to create a stop codon at the same position as the GPI-linked CEA family members in the TM domain of human CEACA1-4L cDNA, cloned in pBluescript SK plasmid. The 296 bp fragment between the unique Pflm1 (nucleotide 1146) and Bsrf1 (nucleotide 1442) sites in wild type CEACAM1-4L cDNA was replaced with the mutant PCR fragment, generating a CEACAM1 mutant, called CC1-t, giving a truncated product. Other CEACAM1-4L mutants were made by inverse PCR-based mutagenesis, as described (Clackson et al., 1992) with minor modifications, using CC1-t-pBluescript SK as template to introduce the appropriate mutations in the TM domain of human CEACAM1-4L cDNA. The oligonucleotides used for mutagenesis (see Table 1) were PAGE purified and used with Pfu DNA polymerase (Stratagene) instead of Taq polymerase for higher fidelity of polymerization. The coding sequences of all mutants were verified by dideoxy sequencing (Pharmacia Biotech T7 sequencing™ kit).

Cell Culture and Transfection - The wild type and mutant cDNAs were inserted into the P91023B expression vector (courtesy of R. Kaufman, Genetics Institute, Boston, MA) for expression in LR-73, a CHO-derived cell line

Table I. Methods of Site-Directed Mutagenesis of CEACAM1-4L

construct	mutation in CC1-4L protein	Primers used for mutagenesis
CC1-t*	A445/stop	Sense: 5′TGGAATCTCCATCCGTTGGTTCTTCAA3′ ▲Pflm1 Bsrf1▼ Antisense: 5′CTTGCCCTGCCGGTCTTCCCGAAATGCA GAAAACATGCCAGGGCTACT**A**-TATCAGAGCAACC3′
CC1-tSG	L419P420/SG	Sense: 5′CTATAATGCT**TC**A**GG**ACAAGAAAATG3′ Antisense‡:5′TTTACGTTCAGCATGATGGGG3′
CC1-tS	L419/S	Sense: 5′CTATAATGCT**TC**ACCACAAGAAAATG3′
CC1-tG	P420/G	Sense: 5′CTATAATGCTCTA**GG**ACAAGAAAATG3′
CC1-tA	L419/A	Sense: 5′CTATAATGCT**GC**ACCACAAGAAAATG3′
CC1-tGT	P420I430/GT	Sense: 5′CTATAATGCTCTA**GG**ACAAGAAAATGGCCTCT CACCTGGGGCCA**C**TGCT3′
CC1-tAT	L419I430/AT	Sense: 5′CTATAATGCT**GC**ACCACAAGAAAATGGCCTCT CACCTGGGGCCA**C**TGCT3′

CEACAM1-4L constructs, sites of their mutations and the oligos which have been used for mutagenesis. The bold, underlined bases represent mismatched nucleotides used to introduce mutations in CC1 cDNA. The dash symbol in the antisense primer used for making CC1-t represents the deleted base corresponding to CC1-4L cDNA. The arrows show the unique restriction sites in CC1-4L cDNA which have been used to replace a 296 b.p. fragment of CC1-4L by the mutant PCR fragment to make the CC1-t construct.

* Other constructs have a stop codon at the same position as CC1-t, in addition to the other specified amino acid substitutions.
‡ This antisense primer was also used for all other inverse PCR based mutagenesis.

(Pollard and Stanners, 1979), or rat L6 myoblasts (Yaffe, 1968). The cells were grown as monolayer cultures in α-MEM (LR-73) or DME (L6) containing 10% fetal bovine serum (FBS) as growth medium (GM; GIBCO BRL, Gaithersburg, MD) supplemented with 100 U/ml penicillin and 100 ug/ml streptomycin (GIBCO BRL) and were incubated at 37°C in a humidified atmosphere with 5% CO2. For transfection, the cells were seeded at 1.5-2 x 10^5 cells/100mm dish and co-transfected 24h later by calcium phosphate coprecipitation with 10ug of P91023B containing full length cDNAs and 1ug pSV2neo plasmid per dish. Geneticin (G-418 sulfate, Gibco BRL)-resistant colonies were pooled after selection with 400ug/ml G-418 in growth medium for 10-14 days. To enrich for stable transfectants, the pooled G-418-resistant colonies were cultured without G-418 for over 1 month and then sorted for high cell surface expression of CEACAM1 protein by FACS, using polyclonal rabbit antihuman CEA antibody. High expressing cell populations were used for further experiments.

FACS Analysis - Transfected cells were analyzed for cell surface expression of proteins by cytofluorometric analysis (FACScan®, Becton Dickinson, Bedford, MA) using polyclonal rabbit anti-CEA as primary antibodies, as previously described (Zhou et al., 1993a). This antibody extensively cross reacts with CEACAM1 and also other CEA family members (Zhou et al., 1993b).

Phosphatidylinositol-Specific Phospholipase C (PI-PLC) Treatment of Intact Cells - Adherent transfected cells were removed from culture dishes with Hank's balanced salt solution lacking Ca^{2+}/Mg^{2+} but containing 0.5 mM EDTA (Hanks/EDTA) for 3min at 37°C, rendered single cell suspensions and the cell concentration determined. Duplicate aliquots of $5x10^5$ cells were washed in cold PBS and resuspended in 150µl of serum free culture medium diluted 1:1 with PBS + 0.2% BSA; one aliquot was treated in suspension at 37°C, in 95% air + 5% CO_2, with 0.2 units PI-PLC from Bacillus Cereus (Boehringer Mannheim) and the other without PI-PLC treatment as a control. After 1hr incubation, the cells were centrifuged and resuspended in 2.5 ml cold PBS and subjected to FACS analysis as described above to measure the residual levels of cell surface CEACAM1 protein. Transfectants of naturally occurring GPI and TM-anchored CEA family members were used as positive and negative controls, respectively.

Differential Extraction of GPI-linked and TM-linked proteins by Triton X-100 - Monolayer cultures of cells were removed with Hanks/EDTA and $2.5x10^6$ LR-73 transfectant cells or $1.25x10^6$ L6 transfectant cells were resuspended in 250µl of ice cold lysis buffer containing 20mM Tris-HCl pH 6.5, 150mM NaCl, 5mM EDTA, and 1% Triton X-100 (TX-100), supplemented with a cocktail of protease inhibitors (1mM phenylmethyl sulfonyl fluoride,

10μg/ml leupeptin and 10μg/ml aprotenin). After mixing, they were incubated for 10min. on ice and centrifuged at 15,000xg at 4°C for 20min. 250μl 2xSDS PAGE sample buffer (Laemmli, 1970) was added to the supernatant and 500μl 1xSDS sample buffer to the pellet. Lysates were boiled for 10min. and 20μl of each supernatant and sediment was analyzed by PAGE for the levels of CEACAM1 proteins by immunoblotting, as described below.

Endoglycosidase H (Endo H) Treatment - For glycosidase treatment, LR(CC1-t) transfectant cells were lysed by TX-100 containing lysis buffer, as described above, and both supernatant and sediment were dissolved in 1x Endo H denaturing buffer provided by the enzyme supplier (New England Biolabs, Beverly, MA), then denatured by boiling at 100°C for 10min. Total protein levels were determined using bicinchoninic acid (BCA) reagent (Pierce, Rockford, IL) and 20ug of total protein was treated with 1250 U EndoH at 37°C for 1hr. For immunoblotting, Endo H treated and untreated samples were mixed with an equal volume of 2xSDS sample buffer and CEACAM1 proteins from 15ug total protein of each supernatant and sediment were resolved by SDS PAGE and immunoblotting, as below.

SDS PAGE and Immunoblotting - Aliquots of supernatant and sediment of TX-100 extracts or Endo H treated samples, prepared as described above, were analyzed by electrophoresis on 7.5% SDS-polyacrylamide gels and

transferred electrophoretically to a 0.45 um nitrocellulose membrane (Protran™, XYMOTECH Biosystem INC.). Membranes were incubated at 4°C overnight with rabbit polyclonal antihuman CEA antibody, diluted 1:3000. After 30 min. washing with buffer containing 0.1% Tween 20, membranes were incubated in 1:4000 diluted HRP-conjugated anti-rabbit antibody, and washed and visualized by ECL reagent according to the manufacturer's instructions (Amersham Life Science, Pittsburgh, PA). Extracts from LR-73 transfectants of naturally occurring GPI or TM-linked CEA family members were used as controls.

Adhesion Assay - Adhesion assays were performed on LR-73 transfectants by visual measurement of the percentage of cells remaining as single cells as a function of time in suspension using a hemocytometer, as previously described (Benchimol et al., 1989).

Myogenic Differentiation Assay - Rat L6 myoblast transfectants were tested for their ability to differentiate as follows. In order to avoid complications in interpretation due to clonal differences unrelated to the effects of the expressed plasmid, pooled populations of transfectants sorted by FACS for relatively high levels of expression, were studied. Cultures of transfectants were seeded at 10^4 cells/cm^2 in 60mm dishes in GM and incubated at 37°C for 3 days, after which the medium was changed to DMEM + 2% horse serum and the incubation continued for a further 7 days. The

cultures were stained with hematoxylin and the extent of differentiation assessed by the percentage of nuclei in single myotube cells with 3 or more nuclei relative to the total number of nuclei. Three independent experiments testing for the ability of various experimental and control transfectants to differentiate were carried out for each of two independent sets of transfectants, with equivalent results.

Assay for Integrin Function - As a measure of integrin function, transfected L6 cells were seeded at $1x10^4$ cells/cm^2 in GM and incubated for a period of 4 days to reach confluency, collected and suspended at a concentration of $4x10^5$ cells/ml in serum free DMEM. 0.25 ml of the cellular suspension was added per well to 24 well plates coated with 10 ug/ml of purified rat plasma fibronectin (0.25 ml/well) (Sigma-Aldrich, Oakville, ON) by overnight incubation at 4°C followed by incubation for a minimum of 8 hours at 4°C with 1% BSA, then washing with serum-free DMEM. Cells were allowed to adhere to the ECM-coated wells for one hour, after which the wells were washed twice with PBS to remove unattached cells. The remaining attached cells were removed with trypsin and their number determined using a particle counter (Coulter Electronics Inc., Hialeah, FL).

RESULTS

To determine the key differences in the transmembrane exon of human CEA family members required to shift the mode of membrane linkage from TM to GPI, we compared the nucleotide and amino acid sequences of the C-terminal hydrophobic exons of both GPI and TM-linked CEA family members (Fig. 1). A significant difference between them is the presence of a stop codon at the same position in all TM exons of GPI-linked members, while CEACAM1-4L has an extended open reading frame of 82 amino acids (TM plus the cytoplasmic domain) after the stop codon.

Effect of introduction of a stop codon into the CEACAM1-4L TM domain - To investigate the role of this stop codon in the mode of membrane linkage of CEA family members, we introduced the same nucleotide sequence differences between CEA and CEACAM1 (a one bp deletion and a one bp substitution) to produce a stop codon at the corresponding position in CEACAM1-4L cDNA by PCR-based site-specific mutagenesis. The mutated CEACAM1-4L cDNA, denoted "CC1-t" for "truncated", was introduced into an expression vector and stably transfected into the CHO cell line, LR-73. After confirming the cell surface expression of a CC-1t protein product by FACS analysis of LR(CC1-t), the GPI-linkage of the protein to the cell membrane of one of the CC1-t expressing clones was demonstrated by sensitivity to PI-PLC. FACS profiles of both PI-PLC-treated and untreated cell populations showed a

Figure 1. Sequence comparison (a. Nucleotides, b. Amino acids) of the transmembrane domain of CEA gene family members. Underlined bases in panel **a** represent stop codons and dashes show a naturally occurring base deletion in comparison with the CEACAM1 gene. Underlined amino acids in panel **b** represent the GPI anchor attachment sites and stars show the position of stop codons. Bold letters show the position of amino acid mutations in various constructed, mutant CC1-4L proteins. The TM domain of mutant GPI-linked CC1-tAT protein and the positions of its mutations (bold letters) are also shown in panel **b** for comparison.

a

```
CEA        CATCTGGAACTTCTCCTGGTCTCTCAGCTGGGGCCACTGTCGGCATCATGATTGGAGTGCTGGTTGGGGTTGCTCTGATA-TAGCAGCCCTGGTGTAGTTTCTTCATTTCAGGAAGACTG
CEACAM6          ...G.G.....TC.........T.....C.........C.............CCA....G.........-...............T....CAT.............
CEACAM8    T.CAA....G.........C.........A.A.........A.................A....CCA....G.........-...T...T.............G.......A........
CEACAM7    T.CAA.C..G...A....AC.............A..G.....A.................A....C....A.G.........-.....T.....
CEACAM1    .TCTACC.CAAGAAAA...C......C.........T..CT.....TG.........AG...CCCT.............GC..T........CA.GT.....G......G.......C.GCAG

CC1-tAT    .TG.ACC.CAAGAAAA...C......C............CT.....TG.........AG...CCCT.............-...T........CA.GT.....G......G.......C.GCAG
```

b

```
CEA          TVSASGTSPGLSAGATVGIMIGVLVGVALI*
CEACAM6      MITV..SA.V...V.....T....AR....*
CEACAM8      DALVQ.S......R...S......AR....*
CEACAM7      YE.VQAS..D....TA.S......A.M...*
CEACAM1      NYN.LPQEN...P..IA..V...VAL....AVAL

CC1-tAT      NYN.APQEN...P..TA..V...VAL....*
```

415

reduction of cell surface levels of both CC1-t and CEA but no reduction of CEACAM1-4L (Fig. 2).

The GPI linkage of the CC1-t protein was also verified by a cold Triton X-100 solubilization assay. In this assay, GPI-linked proteins tend to be insoluble, whereas TM-linked proteins are soluble (Brown and London, 1997; Schroeder et al., 1994). Western blot analysis of cold TX-100 extracts of CEA, CEACAM1-4L and CC1-t transfectants indicated insoluble CEA and CC1-t protein and soluble CEACAM1-4L protein (Fig. 3a). The CC1-t protein, however, showed a major lower MW band in the TX-100 soluble fraction, in addition to the band in the insoluble fraction. This band was presumably due to incomplete glycosylation during intracellular processing, since it was Endo-H sensitive (Fig. 4b), a test for proteins that fail to pass the golgi (Dunphy et al., 1985). Immunoblotting of the supernatant fractions from PI-PLC-treated LR(CC1-t) cells confirmed that only the higher MW, TX-100 insoluble species is released by PI-PLC and therefore GPI-linked and at the cell surface (Fig. 4a). We concluded that the lower MW, TX-100 soluble band is not GPI processed and is sequestered inside the cell. Densitometric analysis of an ECL-developed Western blot showed that about 10% of the total CC1-t protein in the cell is GPI processed. In parallel experiments with LR-73 transfectants producing naturally occurring GPI-linked CEA family members, CEA (Fig. 3a), CEACAM6 (Fig. 3b), or CEACAM8 (data not shown), the lower MW TX-100 soluble unprocessed band could not be detected, although some higher MW,

Figure 2. FACS analysis before and after PI-PLC treatment.

After PI-PLC treatment of LR-73 transfectants, CEA (GPI-linked control) and three CC1-4L mutants, CC1-t, CC1-tSG, CC1-tAT, showed cell surface reduction of expressed protein, but Neo and CC1-4L (TM-linked) controls showed no sensitivity to PI-PLC.

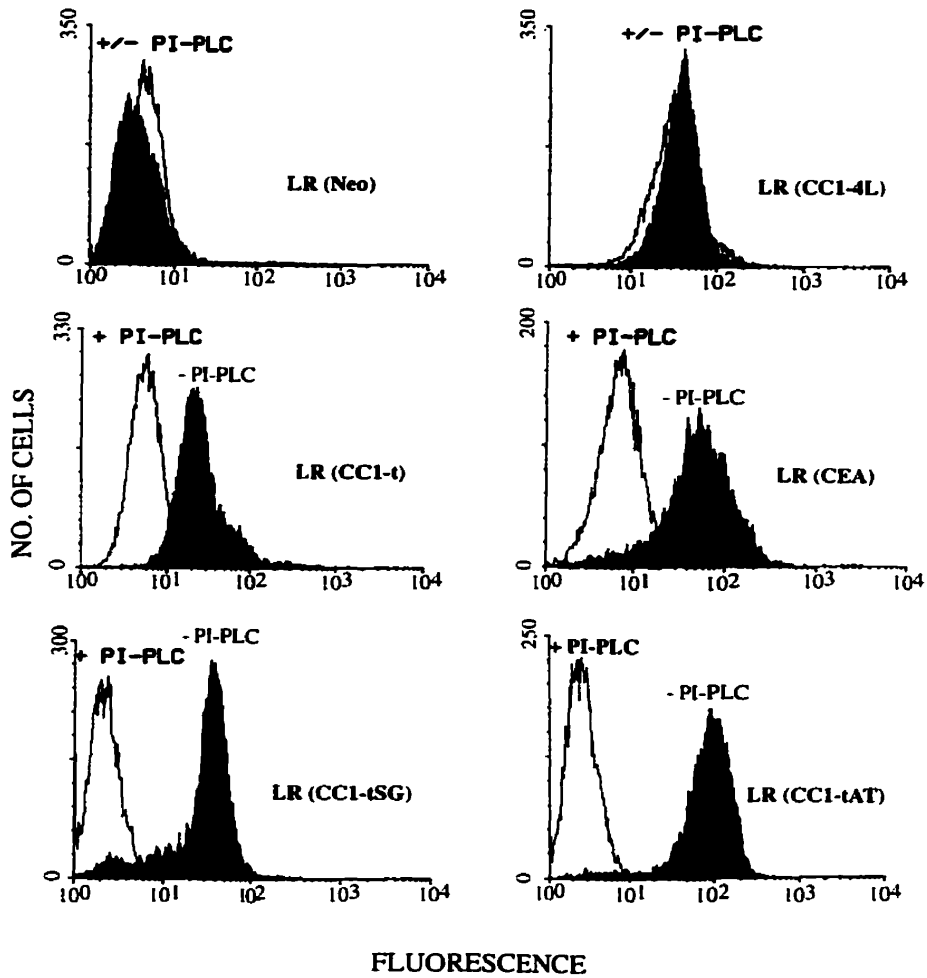

+/- PI-PLC
LR (Neo)
+/- PI-PLC
LR (CC1-4L)
+ PI-PLC
- PI-PLC
LR (CC1-t)
+ PI-PLC
- PI-PLC
LR (CEA)
- PI-PLC
+ PI-PLC
LR (CC1-tSG)
+ PI-PLC
- PI-PLC
LR (CC1-tAT)
NO. OF CELLS
FLUORESCENCE

Figure 3. GPI-processing evaluation of various CC1-4L mutants by cold nonionic detergent solubility assay. Western blot analysis of cold Triton X-100 extracts of LR-73 transfectants was performed. Neo transfectants and naturally occurring CEA family members, CEA, CEACAM6 (GPI-linked), and CC1-4L (TM-linked) were used as controls.

a. CEA and CC1-t proteins were insoluble in TX-100, but CC1-4L protein was soluble. The CC1-t protein, however, showed a major lower MW band in the TX-100 soluble fraction which presumably was due to failure of efficient GPI-processing (see Fig. 4).

b. CC1-tS mutant protein, in contrast to CC1-t (panel a) and CC1-tG, showed a marked improvement in GPI-processing efficiency by the presence of a major 116KD, TX-100 insoluble band and only a minor lower MW, TX-100 soluble band. CEA and CEACAM6 controls, as naturally occurring GPI-linked CEA family members, did not show any soluble, lower MW band.

c. CC1-tGT protein showed a much higher GPI-processing efficiency compared to CC1-tG (panel b).

d. CC1-tAT demonstrated complete GPI-processing as with CEA and CEACAM6, but CC1-tA showed the same level of GPI-processing efficiency as CC1-tS (panel b), since the same lower MW, TX-100 solube band was observed.

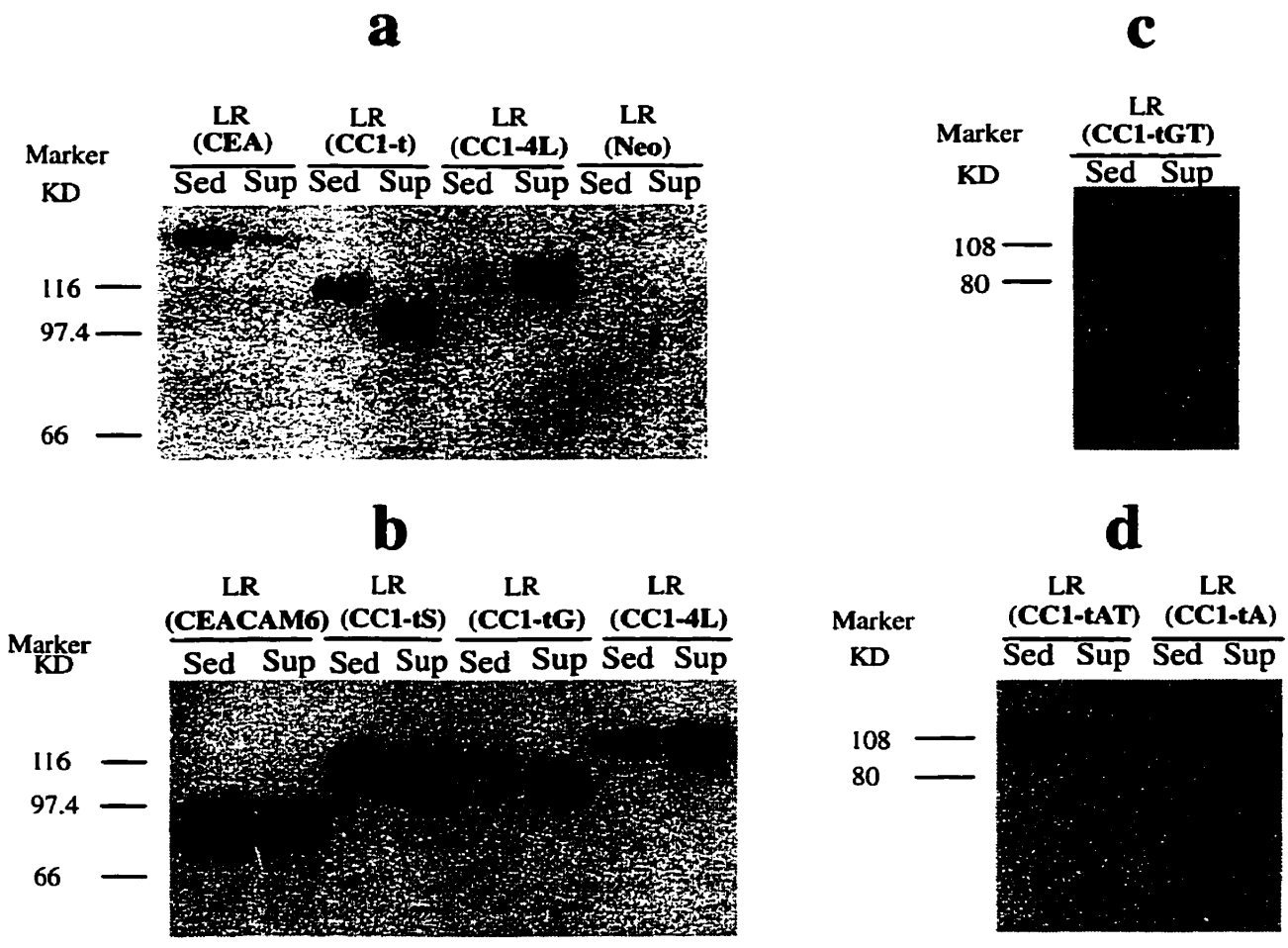
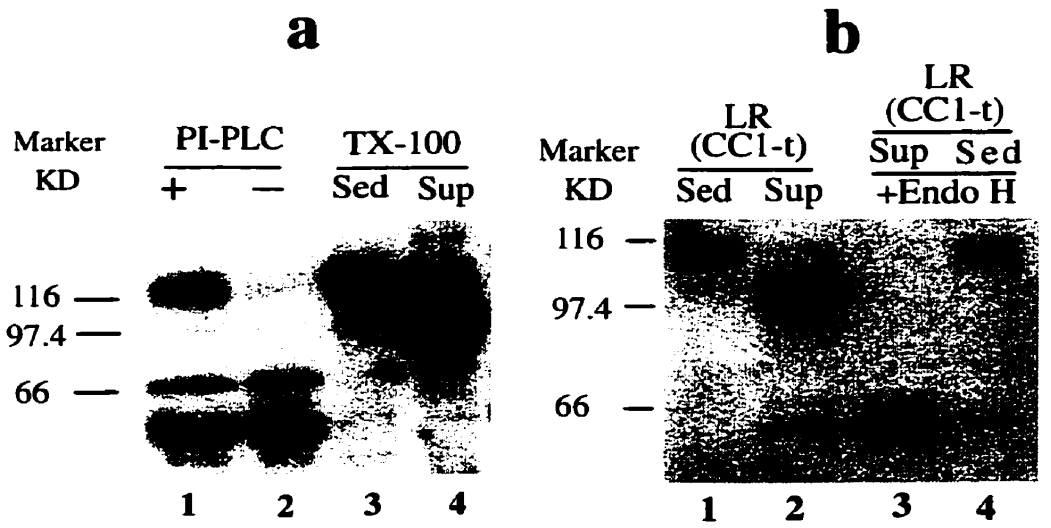

a
LR (CEA)
LR (CC1-t)
LR (CC1-4L)
LR (Neo)
Marker KD
Sed Sup Sed Sup Sed Sup Sed Sup
116
97.4
66
b
LR (CEACAM6)
LR (CC1-tS)
LR (CC1-tG)
LR (CC1-4L)
Marker KD
Sed Sup Sed Sup Sed Sup Sed Sup
116
97.4
66
c
LR (CC1-tGT)
Marker KD
Sed Sup
108
80
d
LR (CC1-tAT)
LR (CC1-tA)
Marker KD
Sed Sup Sed Sup
108
80

Figure 4. Tests for showing intracellular retention of the lower MW, TX-100 soluble fraction of CC1-t protein. a. Immunoblot of the supernatant reaction medium of LR(CC1-t) transfectants after 1 hr incubation with PI-PLC enzyme shows only one band (lane 1) which corresponds to the CC1-t, TX-100 insolubl, higher MW band (lane 3) that is released from cell surface. Two other low MW bands which could be seen in both PI-PLC treated and untreatedcells (lane 1 and 2) but not in TX-100 extracts (lane 3 and 4) are presumably due to other cross-reacting components in the reaction medium (see Materials & Methods). **b.** Immunoblot of both TX-100 soluble, and insoluble fractions with (lane 3 and 4) and without (1 and 2) treatment with Endo H shows that only the lower MW, TX-100 soluble fraction (lane 3) is sensetive to the enzyme and thus is retained inside cells.


a
Marker
KD
PI-PLC
+
—
TX-100
Sed
Sup
116
97.4
66
1
2
3
4
b
Marker
KD
LR
(CC1-t)
Sed
Sup
LR
(CC1-t)
Sup
Sed
+Endo H
116
97.4
66
1
2
3
4

fully processed soluble protein was seen. These results show that the introduction of a stop codon in the CC1-4L TM domain at the corresponding position to the stop codon naturally found in the C-terminal domain of GPI-linked CEA family members is sufficient to confer GPI linkage to CC1-4L, although the efficiency of processing is low.

Mutations giving increased GPI processing efficiency of CC1-t - We predicted that the low efficiency of GPI processing of CC1-t might be ascribed to a weak cleavage and anchor attachment signal at the amino terminal end of its TM domain. The precise TM cleavage site (ω-site) for the CC1-t protein is unknown but, on the basis of previous studies on GPI-linked proteins (Englund, 1993; Micanovic et al., 1990), the two best possible candidates are either N417 or A418; the ω+2 residues, which are also critical for efficient GPI processing, would then be L419 or P420, respectively. However, both of the latter residues have been found to be weak as GPI processing signals (Gerber et al., 1992; Kodukula et al., 1993). We therefore substituted the two candidate ω+2 residues in CC1-t protein for the residues in CEA at these positions (S419 and G420). PI-PLC treatment of stable transfectants of this mutant, denoted LR(CC1-tSG), showed that the protein reaching the cell surface was GPI-linked (Fig. 2). The efficiency of GPI-processing of the mutant CC1-tSG protein by the Triton X-100 solublization assay gave 77% insolubility in TX-100 (Table 2), suggesting a much improved processing efficiency for GPI-linkage.

Table II. GPI-processing efficiency of CC1-4L mutants

CONSTRUCT	CELL LINE	
	LR-73	L6
CC1-t	8%	10%
CC1-tSG	77	ND
CC1-tS	80	77
CC1-tG	5	ND
CC1-tA	80	78
CC1-tGT	56	ND
CC1-tAT	100	ND

GPI-processing efficiency of CC1-4L mutants transfected into the Chinese hamster cell line LR-73 and rat L6 myoblasts, measured by scanning ECL-developed immunoblots. ND: Not Determined

To determine which of the two above substitutions in CC1-t protein was effective for higher GPI-processing, CC1-tS and CC1-tG with single L419S, and P420G substitutions, respectively, were constructed. GPI linkage of the two mutant proteins was shown by PI-PLC treatment of stable transfectants of LR-73 cells, as before (data not shown). The cold TX-100 solubility assay indicated 80% insoluble CC1-tS but only 5% insoluble CC1-tG (Fig. 3B and Table 2). Since mutating L419 to S but not P420 to G could markedly improve GPI processing of CC1-t proteins, we conclude that N417 and not A418 is most likely the ω site for GPI addition in CC1-t.

In spite of the quite dramatic improvement in processing activity in CC1-tS over CC1-t, the efficiency does not match the 100% efficiency of naturally occurring GPI-linked members of the CEA family. The Ser residue in the ω+2 position of CC1-tS, from studies with other GPI-linked proteins (Kodukula et al., 1993; Moran and Caras, 1994), would not be considered optimal for efficient GPI processing. We substituted S419 for Ala to produce CC1-tA, a more favorable residue (Kodukula et al., 1993; Moran and Caras, 1994). CC1-tA cDNA was stably transfected in LR-73 and GPI-linkage of its protein on the cell surface was confirmed by PI-PLC sensitivity as before (data not shown). The GPI-processing efficiency of CC1-tA protein, by the cold TX-100 solubility assay, however, was not improved over that of CC1-tS (Fig. 3d and Table 2).

The failure to obtain 100% GPI-processing activity with mutations at the ω+2 site, led us to consider differences in amino acid sequence further downstream. There is a conserved Thr in the TM domain of GPI-linked CEA, CEACAM6, CEACAM8 and in virtually all primate GPI-linked C-terminal hydrophobic exons (Naghibalhossaini & Stanners, Chapter 3) at a position corresponding to Ile430 in CEACAM1-4L protein (Fig. 1) that, in CC1-t, tends to shorten the hydrophilic region after the ω+2 residue required for efficient processing (**Furukawa et al., 1997**). In order to investigate the role of this residue, an inefficiently processed CEACAM1 mutant with the same residues at the and ω+2 sites as CEA, CC1-tG, was used. CC1-tGT, with Ile430 replaced by Thr, was therefore constructed from CC1-tG. The immunoblot of TX-100 extracts of LR(CC1-tGT) showed a marked improvement of GPI-processing relative to the CC1-tG protein (56% of CC1-tGT protein versus 5% of CC1-tG) (Fig. 3c and Table 2). These results indicate that the evolutionarily conserved Thr430 residue in fact has an important role in GPI-processing of CEA family members.

Finally, mutant CC1-tAT, with both advantageous mutations, L419A and I430T, was constructed and tested for processing efficiency. This mutant protein was well expressed on the surface of LR-73 cells and sensitive to PI-PLC (Fig. 2). The GPI-processing efficiency for CC1-tAT was essentially 100%,

since there was no detectable lower MW cold TX-100 soluble band, as with the naturally occurring CEA family members (Fig. 3d and Table 2).

We conclude that CEACAM1-4L can be converted into an efficiently processed GPI-linked protein by relatively few mutations in its TM domain; the introduction of a stop codon, a mutation at the ω+2 site and a mutation extending the length of the hydrophilic spacer region suffice.

GPI processing of mutant CEACAM1-4L proteins is cell line independent - To test whether or not the efficiency of GPI processing of the CEACAM1-4L mutants is cell line dependent, we transfected CC1-t, CC1-tS, and CC1-tA cDNA into rat L6 myoblasts. The GPI-processing was verified by PI-PLC sensitivity (data not shown) and the efficiency measured by immunoblotting of cold TX-100 extracts of L6 transfectants, as before (Table 2). The latter showed GPI-processing efficiencies of approximately 10, 77, and 78% for CC1-t, CC1-tS and CC1-tA proteins in L6 cells, respectively. Thus, there was no significant difference of GPI-processing efficiency of the CEACAM1-4L mutants in rat L6 versus Chinese hamster LR-73 cells, even though the GPI-processed mutant proteins (and CEA) were consistently lower in MW in L6 than LR-73, presumably because of less glycosylation (data not shown).

GPI-linked CC1-tAT functions as an adhesion molecule - In order to test whether the mode of membrane linkage of CEACAM1-4L changes its

function as an intercellular adhesion molecule (Rojas et al., 1990), LR(CEACAM1-4L) and LR(CC1-tAT) transfectant cells were subjected to an aggregation assay in suspension (Benchimol et al., 1989). The results, shown in Fig. 5, indicate that CC1-tAT functions well as a homotypic intercellular adhesion molecule. Thus, changing the mode of membrane linkage does not delete the intercellular adhesion function of CEACAM1-4L.

GPI-linked CC1-tAT blocks myogenic differentiation and perturbs binding to fibronectin - GPI-linked CEA and CEACAM6 inhibit the differentiation of many different cell types, whereas TM-linked CEACAM1 does not (see Introduction). In order to test whether changing the mode of membrane linkage of CEACAM1 to GPI anchorage now conferred on CEACAM1 the property of blocking myogenic differentiation, CC1-tAT was expressed in rat L6 myoblasts and the transfectants tested for their ability to fuse into myotubes. Unlike L6(CC1-4L) transfectants and like L6(CEA) transfectants, L6(CC1-tAT) transfectants were completely unable to fuse and differentiate into myotubes (Fig. 6).

The mechanism for the inhibition of differentiation by CEA and CEACAM6 has been shown to result from perturbation in the function of certain integrins, $\alpha_5\beta_1$, in the case of L6 myoblasts, resulting in a change in the binding to fibronectin, a key component of the extracellular matrix (Ordoñez et al., submitted). L6 myoblasts expressing CC1-tAT, obtained from confluent

Figure 5. Homotipic adhesion mediated by mutant CEACAM1 protein: LR-73 cells transfected by Neo and CC1-4L as controls and mutant, GPI-linked CC1-tAT were subjected to the homotypic adhesion assay in suspension. The percentage of single cells was measured as a function of time in suspension. The mean expression levels of CC1-4L and CC1-tAT by FACS analysis were 280 and 98, respectively.

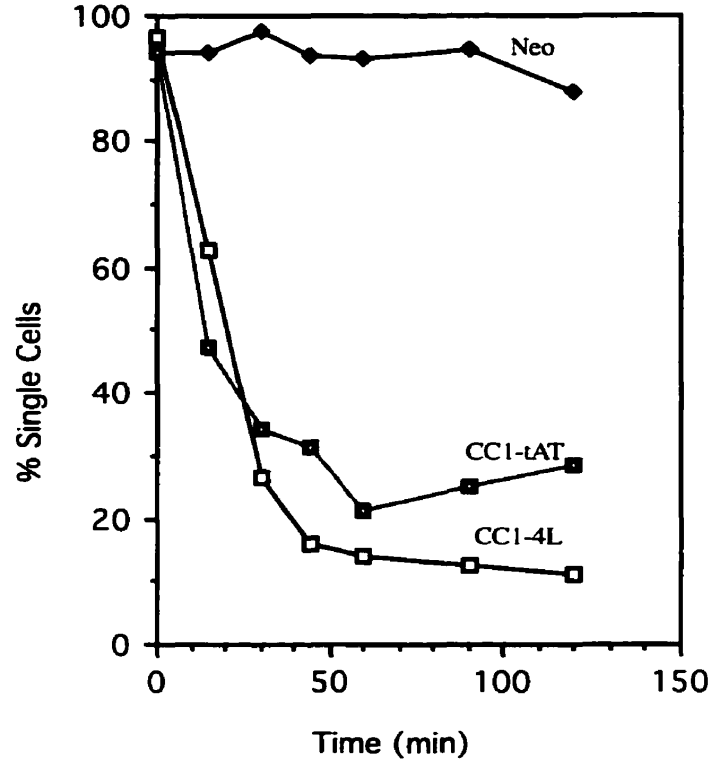
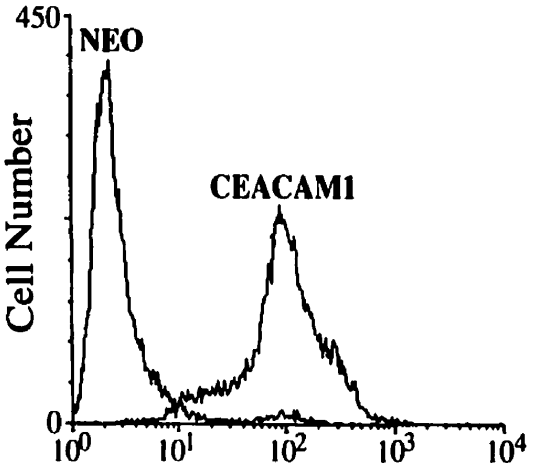

100
80
60
40
20
0
% Single Cells
Neo
CC1-tAT
CC1-4L
0
50
100
150
Time (min)

Figure 6. Photomicrographs of hematoxylin stained cultures of various L6 transfectants incubated in DM for 7 days. FACS profiles show the relative cell surface expression level of proteins.

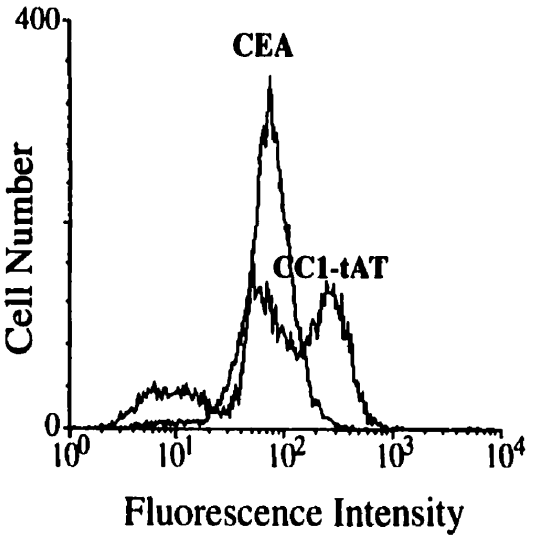
L6(NEO)

L6(CEACAM1)



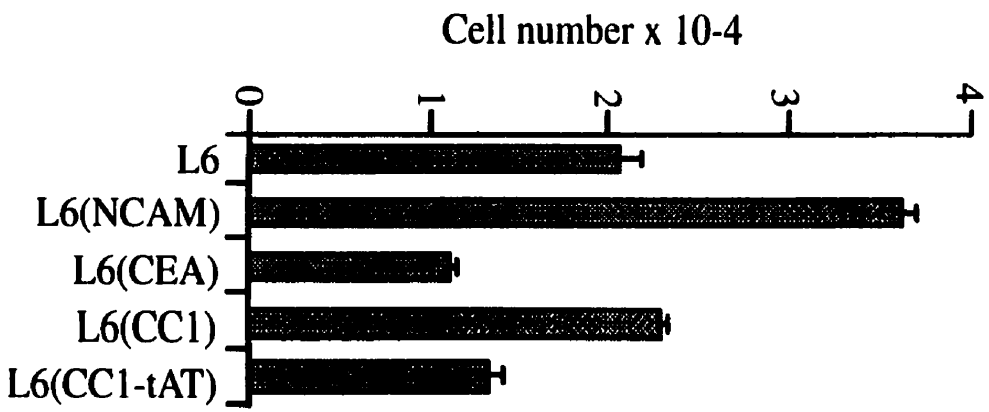
L6(CEA)

L6(CC1-tAT)



cultures in which differentiation would normally occur, bound less well to purified fibronectin than parental L6 myoblasts (Fig. 7), much as L6(CEA) transfectants, whereas the binding of L6(CC1-4L) and L6(NCAM) transfectants was unaffected.

Figure 7. GPI anchored CEA and CC1-tAT expression on the surface of L6 myoblast modifies cell adhesion to fibronectin. Transfected L6 myoblasts were allowed to bind fibronectin-coated tissue culture surfaces for 1 hour at 37°C. L6(CEA) and L6(CC1-tAT) bound to fibronectin significantly less than L6 parental, L6 (CC1-4L) and L6(NCAM) controls. These experiments were repeated three times with similar results.


Cell number x 10-4
0
1
2
3
4
L6
L6(NCAM)
L6(CEA)
L6(CC1)
L6(CC1-tAT)

DISCUSSION

In this work we have investigated the molecular requirements for GPI-linkage of CEA family members in the context of evolution. GPI-linked members, CEA and CEACAM6, differ radically in function from the TM-linked member, CEACAM1, in that the former can block cell differentiation and contribute to tumorigenesis (**Eidelman et al., 1993; Screaton et al., 1997**) while the latter cannot (**Rojas et al., 1996**). Consistent with these findings, CEA and CEACAM6 are upregulated in human cancer, while CEACAM1 is downregulated, possibly acting as a tumor suppressor, as shown for the murine equivalent of CEACAM1 (**Hsieh et al., 1995; Kunath et al., 1995**). Since only TM-linked CEACAM1-like genes have been discovered in rodents, we have suggested a recent evolution of GPI-linked CEA family members from a primordial CEACAM1-like gene (**Stanners et al., 1995; Stanners et al., 1992**). In fact, in an evolutionary study, using a rapid PCR-based assay on genomic DNA, we have identified GPI-linked CEA members only in primates from all mammalian orders tested (Naghibalhossaini et al., Chapter 3). In view of the rather dramatic change in function implicated by the acquisition of the GPI anchor, we decided to examine the ease of derivation of GPI linkage from the TM domain of CEACAM1 by determining which of the sequence differences observed between CEACAM1 and the GPI-linked CEA family members are required to shift the mode of membrane anchorage. We have shown here that very few mutations are required to convert TM-linked

CEACAM1 into an efficiently processed GPI-linked molecule: the introduction of a stop codon and two upstream substitutions in the TM domain of CEACAM1 suffice.

There are actually three TM-linked genes in the human CEA family: CEACAM1, CEACAM3 and CEACAM4. CEACAM1 was chosen as a starting point for our mutational study because of its closer relatedness in nucleotide sequence and domain organization to the human GPI-linked family members (Hammarström et al., 1998).

Measurement of GPI-processing efficiency - Previous work by others on the determination of molecular signals required for GPI processing has been carried out by transient expression of mutant GPI-linked proteins or chimeras, such as placental alkaline phosphatase (Kodukula et al., 1993), acetylcholinesterase (Bucht and Hjalmarsson, 1996) or decay accelerating factor (Moran and Caras, 1991), using enzyme activity or expression at the cell surface as a measure of the efficiency of GPI-processing. This approach depends on a constant transfection efficiency. In the present work, we used two properties of mutant CEACAM1 proteins to quantitate GPI processing efficiency. Due to the linkage of GPI-linked proteins to saturated acyl chains, these proteins associate with sphingolipids and cholesterol in cold detergent-resistant membrane domains (DRM) in the cell membrane and become insoluble (Brown and London, 1997). Also, CEA family members are highly

glycosylated on asparagine residues, constituting up to 50% of their molecular weight; failure to cleave the hydrophobic C terminus and attach a GPI anchor results in retention in the ER and incomplete glycosylation, resulting in a faster gel mobility detectable by immunoblotting.

The GPI-processing efficiency can therefore be assessed by the presence of cold detergent soluble bands of lower than normal molecular weight. The justification for this assumption derives from both precedent and experiment. Thus, uncleaved GPI anchoring signals have been reported to function as intracellular retention signals for GPI-unprocessed proteins (Moran and Caras, 1992). Also, a mutant chimeric protein of rat α-globin (α-GL) fused in-frame with the C-terminal domain of placental alkaline phosphatase (α-GL-PLAP), unlike wild-type α-GL-PLAP, failed to be GPI-processed and associated with the ER-resident chaperon, Bip/GRP78, resulting in retention within the ER and rapid degradation (Oda et al., 1996). In the case of CEACAM1 constructs that were inefficiently GPI-processed, a high MW TX-100 insoluble band was shown to be released from the cell surface by digestion with PI-PLC, demonstrating this form to be the mature, properly localized GPI-linked protein. In the TX-100 soluble fraction, two bands were obtained, one of full size and another of lower MW. The latter was shown to be selectively sensitive to Endoglycosidase H (Endo-H), an enzyme that cleaves the high-mannose oligosaccharides added in the ER that have not been processed by β-N-acetylglucosamine transferase I and α-mannosidase II

in the medial golgi cisternae (Dunphy et al., 1985). No unprocessed CEACAM1 protein was found in the extracellular medium, at least after an incubation of 1 hr, as described for GPI-uncleavable hGH-DAF28 fusion protein (Moran and Caras, 1992). The larger MW TX-100 insoluble band was therefore used as a measurement of CEACAM1 GPI-processing efficiency. Since the rate of degradation of the GPI-unprocessed, Endo-H sensitive, mutant protein was not assessed, our measures of the percentage of GPI-processed protein based on steady state levels are likely to represent overestimates of the true processing efficiencies.

Requirement for a truncated TM-domain - To determine the changes in the CEACAM1 TM domain required to change its mode of membrane linkage, we made a series of site-directed mutations. The CEACAM1-4L TM domain was first truncated by the insertion of a nonsense mutation at the position corresponding to A445. This position corresponds to the stop codon in all human GPI-linked CEA family members (see Fig. 1). The introduction of a stop codon by a one bp deletion and a substitution at this position alone could confer GPI-linkage to the mutant CC1-t protein, although inefficiently (~10% processing efficiency). This indicates that a latent GPI-anchoring signal must exist in the CEACAM1-4L TM domain that becomes functional upon removal of the cytoplasmic tail. Precedents for this observation have been reported; for example, deletion of the pentapeptide cytoplasmic tail of the T-cell receptor (TCR) β-chain resulted in its surface expression as a GPI-

anchored polypeptide (Bell et al., 1994). The identical position and sequence of the stop codons in all human GPI-linked CEA family members plus the high sequence homology between them suggests they were all derived from the same gene after evolution of the first TM exon giving GPI linkage [probably attached to an CEACAM6-like gene (Naghibalhossaini et al., Chapter 3)]. The discovery in New World monkeys of a second stop codon generating mechanism further downstream of that seen in humans and most primates (Naghibalhossaini et al., Chapter 3) indicates that other sites of introduction of stop codons in the TM domain can be equally effective in generating GPI linkage.

Reqirements for efficient GPI-processing – Studies on other GPI-linked proteins (Udenfriend and Kodukula, 1995) suggest that the site for cleavage and addition of a GPI anchor is restricted to the 6 small amino acids: D,N,G,A,S and C; the ω+2 residue, which is also important as a GPI-processing signal, is limited to only A,G and S. N417 was decided to be the most likely ω site in CC1-t based on the fact that mutations L419S or L419A at the ω+2 site for this residue both markedly increased the efficiency of GPI processing. These mutations, however, resulted in processing at 80% efficiency, not 100% as seen in naturally occuring CEA GPI-linked CEA family members. Other differences between the latter and CEACAM1 TM domains were therefore

sought that could account for the residual improvement in processing efficiency.

The GPI processing signal peptide includes a hydrophilic spacer sequence between the ω-site and the C-terminal hydrophobic region. We speculated that the incomplete GPI-processing of CC1-tS and also the highly inefficient processing of the CC1-tG proprotein, which actually has the same residues at the positions corresponding to the ω and ω+2 sites in the CEA TM domain, might be due to a defective hydrophilic spacer signal. The length and sequence of the spacer differs among various GPI-linked proteins, the length alone varying from 4 to 17 residues. Mutational studies on the spacer region of 5′-nucleotidase showed that, of 6-14 residues, a length of 8 residues was optimally compatible for GPI-modification (Furukawa et al., 1997). If N417 is the ω-site in CC1-t TM, the spacer would be 9 residues (A418-S426) in length. All GPI-linked CEA family members of human (Fig. 1) and other primates (Naghibalhossaini et al., Chapter 3) have a T residue at position 430 (except for human CEACAM7 which has an A at this site preceded by a T residue), whereas all TM-linked family members have a hydrophobic Ile residue at this site. It is possible that this Thr, which is at the border of the hydrophilic spacer and the C-terminal hydrophobic domain, is required for efficient GPI-processing, possibly by increasing and optimizing the length of the hydrophilic spacer by 4 residues. Substitution of I430 for T in CC1-tG and CC1-tA, giving CC1-tGT and CC1-tAT mutants respectively, resulted in a much

higher processing efficiency (56% for CC1-tGT and 100% for CC1-tAT protein), bringing the efficiency for CC1-tAT, at least, up to that of naturally occurring GPI-linked CEA family members.

Testing of some of the above CEACAM1-4L mutants also showed the same level of GPI-processing efficiency in rat L6 myoblasts as in the Chinese hamster cell line, LR-73, indicating that the GPI-processing is cell line independent, at least in two mammalian systems.

As all of the naturally occurring GPI-linked CEA family members tested (CEA, CEACAM6 and CEACAM8) showed complete GPI-processing, it is interesting that, although the evolution of GPI-linked CEA family members appears to have happened relatively recently in the primate radiation, all of them accumulated enough mutations to be GPI-processed with high efficiency. This change of mode of cell anchorage from TM to GPI-linked did not abolish the intercellular adhesive function of CC1-4L, an expected result considering the fact that both GPI-linked and TM-linked CEA family members have been found to mediate homotypic and heterotypic intercellular adhesion (Rojas et al., 1996). GPI-linked CEA and CEACAM6, however, show radical differences from CEACAM1 in their effects on several important cellular properties in that they block the differentiation of a number of different cell types and can disrupt the tissue architecture and polarization of human colonocytes (see Introduction). Interestingly, at least one of these differences in function was found for CC1-tAT: this efficiently processed GPI-linked derivative of CC1-4L completely blocked the myogenic

differentiation of rat L6 myoblasts. For GPI-linked CEA and CEACAM6, the block of myogenic differentiation is accompanied by many other effects on the cell phenotype, all apparently consequences of CEA/CEACAM6-mediated alterations of the function of specific integrins (Ordonez, Malette & Stanners, in preparation). CC1-tAT expression was found to perturb the binding of L6 cells to fibronectin in much the same fashion as CEA and CEACAM6, thus implying the same effect as CEA/CEACAM6 on the function of the $\alpha_5\beta_1$, the integrin mainly responsible for binding to fibronectin. The question as to whether the "package" of functional consequences that arise due to the perturbation of this integrin could provide a rationale for the evolution of GPI linkage is currently under investigation.

Since GPI-linked proteins are found in very primitive organisms, such as protozoa and yeast, where in some cases most of the cell surface proteins are GPI-anchored (McConville and Ferguson, 1993), it seems assured that the machinery for producing the GPI anchors themselves was already in place when the generation of GPI-linked members of the CEA family occurred during evolution. In fact, free GPI-related structures have been isolated from bacteria as well as from eukaryotic cells (Ferguson and Williams, 1988; Gaulton and Pratt, 1994) . This basic GPI structure, probably having an independent function, therefore most likely predated the evolution of GPI-anchored proteins. Recruitment of protein GPI-anchoring machinery in so wide a variety of eukaryotic organisms suggests a strong selective advantage for the

acquisition of this form of anchoring during evolution, with functional significance that is only now being explored.

On the basis of above results, we propose the following model for evolution of GPI-linked CEA family members: 1. TM linked CEACAM1 represents the more primordial conserved member of the CEA family, since it is present in both rodents and primates. 2. Truncation of the TM domain of CEACAM1 protein by the introduction of a stop codon, followed by a few more amino acid substitutions in the CEACAM1 TM exon, resulted in the evolution of an efficiently processed primitive GPI-linked gene. 3. All other GPI-linked CEA family members were derived from this gene by gene duplication and exon shuffling relatively recently in the evolutionary time scale. This scheme provides a paradigm for evolution of GPI anchors in the Ig superfamily: the introduction of a stop codon by mutation in a relatively wide region of a TM domain produces an inefficiently processed GPI-linked protein isoform that can nevertheless allow "testing" for an advantageous change in function; if the latter accrues, the advantage will favor the acquisition of further mutations giving efficient processing and high cell surface levels of expression.

Acknowledgments - This work was supported by grants from the Medical Research Council of Canada and the National Cancer Institute of Canada. FN is a recipient of a studentship from the Ministry of Health and Medical Education, Government of Islamic Republic of Iran.

The work presented in the previous chapter examines the possibility of derivation of GPI-linked CEA family members from more primordial conserved TM-linked CEA family members during evolution. The results show the truncation of the TM domain of CEACAM1 protein by the introduction of a stop codon, followed by a few more amino acid substitutions in the CEACAM1 TM exon possibly resulted in the evolution of the first efficiently processed primitive GPI-linked CEA family member.

Since the normal function of CEA family members *in vivo* is not well understood, a study of the appearance during evolution of a structural feature which has a dramatic effect on function, such as GPI membrane linkage, could provide important insights into CEA family function *in vivo*. On the assumption that GPI-linked CEA family members evolved from TM-linked CEACAM1-like CEA family members and that a stop codon in TM exon of CEA family members is necessary and indicative of GPI-anchorage of CEA family members as shown in chapter 2, we established a rapid PCR-based screening of genomic DNA. The PCR analysis of TM exons of the CEA family members of various species presented in the next chapter was designed to directly determine where in evolution the functionally significant change of membrane anchorage in CEA family occurred.

Chapter 3

Evolution of GPI-linked Ig Superfamily Members in CEA Subfamily Indicates Adaptive Function of GPI Membrane Anchors

Abstract

Glycophosphatidyl inositol (GPI) anchors, linking glycoprotein molecules to the external cell surface, would be expected to lead to different inter-molecular coupling than transmembrane (TM) anchors that can directly link exo-signals with internal cellular machinery by their cytoplasmic domains. We show here that, for the human carcinoembryonic antigen (CEA) family, a subfamily of the immunoglobulin superfamily (IgSF), GPI anchorage demonstrates convergent evolution, in that it evolved twice independently by different packages of mutations, presumably from more primitive TM anchors, during the primate radiation. This finding suggests that GPI anchors have adaptive function conferring selective advantage. For CEA family members, conversion of the mode of membrane linkage from TM to GPI confers radical changes in the cellular properties they mediate: from tumor suppression or neutrality towards disordered cellular and tissue architecture and inhibition of differentiation, thereby contributing to adult tumorigenesis. The present results suggest that these features must nonetheless be advantageous under certain situations during the life of the organism.

Introduction

Cell surface linkage of external proteins by GPI anchors is common in animal cells (Englund, 1993; McConville and Ferguson, 1993). The human CEA family consists of 29 highly similar gene-like sequences, closely clustered on the long arm of chromosome 19, and includes members comprised of NH2-terminal V-like Ig domains followed by a variable number of C2 Ig domains and anchored to the external cell surface by both GPI and TM anchors. GPI-anchored family members CEA and CEACAM6 demonstrate radically different functions from TM-linked member CEACAM1 in that they inhibit the differentiation of many cell types (Eidelman et al., 1993) and disrupt cell polarity and tissue architecture of colonic epithelial cells, whereas CEACAM1 has no such effect (Ilantzis and Stanners, 1997). In cell lines, these effects of CEA and CEACAM6 lead to an increase in tumorigenicity (Screaton et al., 1997). Mouse and rat CEACAM1, on the other hand, appear to act as tumor suppressors (Kleinerman et al., 1996; Kunath et al., 1995; Luo et al., 1997). In fact, approximately 50% of human cancers show deregulated over-expression of CEA and/or CEACAM6, while CEACAM1 is usually down-regulated (Cournoyer et al., 1988; Neumaier et al., 1993; Nollau et al., 1997); tumor differentiation status is negatively correlated with the cell surface expression levels of CEA and CEACAM6 (Ilantzis et al., 1997).

The structural feature that determines these radically different effects on the cell phenotype is the CEA-derived GPI anchor, as CEACAM1 can acquire the myogenic differentiation-inhibitory properties of CEA by swapping its TM anchor for the GPI anchor of CEA and vice versa (Screaton R. & Stanners C. P. in preparation). Since rodents are known to possess only TM-linked CEA family members (Zimmermann, 1998), whereas humans have both (Obrink, 1997), this, along with other features of the CEA family, have led us to suggest that GPI anchors arose relatively recently in evolution by replication and mutation of a primordial TM exon (Naghibalhossaini et al. Chapter 2). These considerations therefore raise an important question: can a structural feature that enhances tumorigenicity by perturbing cell and tissue architecture be advantageous and therefore maintained in evolution? In this study we address this question and present molecular evidence for positive selection of GPI-anchored CEA family members in the course of evolution.

Species and genomic DNA preparation - The species included in this study are listed by their common name, scientific name, and tissue source: American opossum, *Didelphys virginiana*, O.K cells, ATCC; rabbit, *Oryctolagus cuniculus*, blood; rat, *Rattus norvegicus*, L6 myoblast; mouse, *Mus musculus*, C2C12 myoblast; tree shrew, *Tree shrew glis*, purified genomic DNA provided by C. A. Porter, Wayne State University, Detroit, Michigan; flying lemur, *Cynocephalus volans*, liver, Field Museum of Natural History, Chicago, Illinois; bat, *Tadarida brasiliensis*, Tb1lu, ATCC; dog, *Canis familiaris*, MDCK; cat, *Felis catus*, blood; sheep, *Ovis aries*, blood; goat, *Capra hircus*, blood; cow, *Bos taurus*, blood; horse, *Equus cabalus*, blood; human, *Homo sapiens*, blood; chimpanzee, *Pan troglodytes*, blood, Yerks Regional Primate Centre, Georgia; rhesus macaque, *Macaca mulatta*, blood, Tulane Regional Primate Centre (TRPC), Florida; mangaby, *Cercocebus atys*, blood, TRPC; cotton top taumarin, *Saguinus oedipus*, blood, New England Primate Centre (NEPC), Mass.; common marmoset, *Callithrix jacchus*, blood, NEPC; wooly monkey, *Lagothrix lagothrica*; spider monkey, *Ateles geoffroyi*; black howler monkey, *Alouatta caraya*; red howler monkey, *Alouatta seniculus*; red uakari, *Cacajao rubicundus*; titi monkey, *Callicebus molloch*, blood, California Regional Primate Research Centre; squirrel monkey, *Saimiri sciurus*, blood, NEPC; tarsier, *Tarsisus syrichta*, liver, Duke University Primate Centre (DUPC); *Propithecus tattersalli*, liver, DUPC.

Genomic DNA was extracted from ACD-anticoagulated blood, cells or frozen tissues as described (Sambrook et al., 1989).

PCR and RT-PCR analysis - Sense (S) 1, antisense (AS) 1 and antisense 2 PCR primers (Fig. 1) were based on the human CEA gene TM sequence. 0.5 ug of genomic DNA was amplified for 25-30 cycles in 100ul reaction volume using 0.5 units Taq polymerase (BIO/CAN Scientific Inc. Mississauga, ON) using 42°C annealing and 72°C polymerization temperatures for 45 Sec. To overcome reaction failure because of possible 3′-end mismatch of primers with template DNA and for higher fidelity of polymerization for samples for which no positive PCR bands using Taq polymerase were obtained, a mix of 0.5 U Taq+0.1 U Pfu (Stratagene) polymerases was used. Three sequences of callicebus monkey cDNAs with the CEA stop package (RT 1-3, Fig. 1) were obtained by doing RT-PCR on total RNA extracted by the single-step method (Chomczynski and Sacchi, 1987) using guanidium isothiocyanate from isolated blood leukocytes (Boyum, 1974). To synthesize cDNA, 2 ug of RNA was incubated with 100 pmol Pd(N)6 random hexamer (Pharmacia Biotech Inc.) and 200 U of M-MLV reverse transcriptase (GIBCO BRL, Burlington, ON) in a total reaction volume of 20 ul for 1 hr at 38°C. The RT-reaction product was used for PCR amplification as above using primer AS-2 and a sense (5′GTTITTCTACTTGTICACAATCTGCC3′) primer residing in N-terminal IgV-like domain of human CEA family members. To obtain upstream flanking sequence of the Ceb stop TM domain of Callicebus monkey, a S-2

primer (Fig. 1) (5′GTACCAGGTAGTTCTCCT3′) was designed based on a consensus sequence of human and Callicebus CEA family members. Using S-2/AS-2 primers for PCR under low stringency conditions, i.e., 37°C annealing temperature for 2 min, a longer PCR band of Callicebus CEA TM domain with the Ceb stop package was obtained (Fig. 1).

DNA electrophoresis and Southern blot - PCR products were seperated by electrophoresis through a 2% agarose gel containing 0.5 ug/ml ethidium bromide in TBE buffer and photographed under UV illumination prior to southern blotting. DNA was transferred to a nylon membrane (Genescreen plus, Boston, Mass) in 10x SSC for 16hr and Southern hybridization for detection of CEA family members was carried out in 50% formamide containing hybridization buffer at 42°C overnight. The 46 mer 5′-end, P^{32}-labeled oligo corresponding to the sequence residing between the S-1 primer and the human CEA gene stop codon (Fig. 1) was used as a CEA TM specific probe. The membrane was washed once by 2x SSC for 10 min at room temperature and once in 1xSSC, 1% SDS for 20 min at 37°C and exposed to X-ray film.

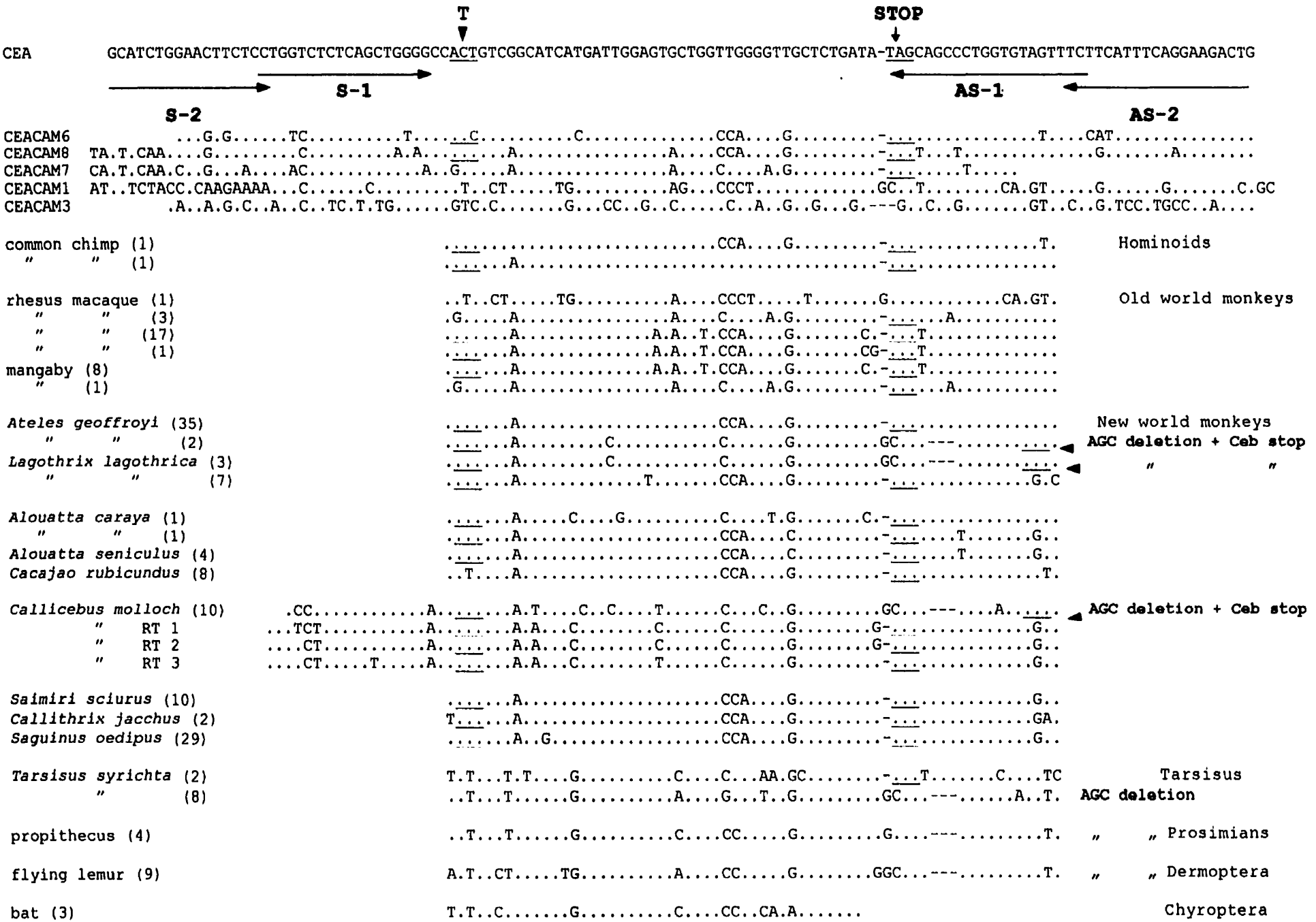
Cloning and sequencing - Positive PCR bands from 3-5 pooled PCR reactions were cloned (TA cloning kit, Invitrogen) and the indicated number of independent clones (Fig. 1) were sequenced for each species. Sequencing in

both directions was accomplished using a dideoxy sequencing kit (Pharmacia Biotech T7 sequensing™ kit).

Results and Discussion

The known nucleotide sequences of the TM domains of human CEA family members (Fig. 1) indicate the common presence at the same position of a stop codon in all GPI-anchored members by a one bp deletion and a substitution, relative to the sequences of TM-linked members. Interestingly, these occurred in a region consisting of 3 trinucleotide (AGC) imperfect repeats, which might be expected to be a mutational hotspot due to slippage of replication enzymes (Freimer and Slatkin, 1996; Ramel, 1997) and which, in fact, show deletion of one of the repeats as well as other mutations (see below). A stop codon and a few upstream mutations to increase the efficiency of processing have been shown to be sufficient for efficient GPI anchorage (Naghibalhossaini F., & Stanners C. P. Ch. 2). On the assumption that a stop codon in the TM domain of CEA family members is indicative of GPI-linkage of family members (Ch. 2), the TM exon sequences were used to design two sets of primers for PCR reactions with genomic DNA, one of which could amplify only sequences containing the human CEA-like stop codon (AS-1 in Fig. 1). The latter had a mismatch at its 3′ end for human CEACAM1 and failed to amplify cloned human CEACAM1 cDNA using Taq polymerase but

Figure 1. Nucleotide sequence alignment of PCR amplified TM domains of CEA family members of human and other species. The sequences were compared to human GPI-anchored CEA. Identical nucleotides are indicated as dots. Dashes show naturally occurring deletions. Stop codons are underlined. The conserved Thr codon in GPI-linked members, shown to be important in GPI processing, is underlined in green color. Positions of 2 sets of sense (S) and antisense (AS) PCR primers are shown by horizontal arrows. Numbers of sequenced independent clones of PCR products for each species have been indicated in brackets. 3 cloned RT-PCR amplified bands of callicebus monkey cDNA (see methods) are shown as RT 1-3. A positive PCR band for the bat CEA family member was obtained only by using a combination of S-1/AS-1 primers. The sequence is two nucleotides shorter than the expected human control. This was possibly because of pairing of the AS-1 primer at its 3′ terminus to the TAG sequence which precedes the sequence corresponding to the position of the CEA stop of those CEA members with AGC deletions in their TM domains.


T
STOP
CEA GCATCTGGAACTTCTCCTGGTCTCTCAGCTGGGGCCACTGTCGGCATCATGATTGGAGTGCTGGTTGGGGTTGCTCTGATA-TAGCAGCCCTGGTGTAGTTTCTTCATTTCAGGAAGACTG
S-2
S-1
AS-1
AS-2
CEACAM6 ...G.G.....TC.....T.....C.....C.....CCA....G.....-...T....CAT.....
CEACAM8 TA.T.CAA....G.....C.....A.A.....A.....A....CCA....G.....-...T...T.....G.....A.....
CEACAM7 CA.T.CAA.C..G...A....AC.....A..G.....A.....A....C....A.G.....-.....T.....
CEACAM1 AT..TCTACC.CAAGAAAA...C.....C.....T..CT.....TG.....AG...CCCT.....GC..T.....CA.GT.....G.....G.....C.GC
CEACAM3 .A..A.G.C..A..C..TC.T.TG.....GTC.C.....G...CC..G..C.....C..A..G..G...G.---G..C..G.....GT..C..G.TCC.TGCC..A....
common chimp (1) .....CCA....G.....-.....T. Hominoids
" " (1) .....A.....-.....
rhesus macaque (1) ..T..CT.....TG.....A....CCCT.....T.....G.....CA.GT. Old world monkeys
" " (3) .G.....A.....A....C....A.G.....-.....A.....
" " (17) .....A.....A.A..T.CCA....G.....C.-...T.....
" " (1) .....A.....A.A..T.CCA....G.....CG-...T.....
mangaby (8) .....A.....A.A..T.CCA....G.....C.-...T.....
" (1) .G.....A.....A....C....A.G.....-.....A.....
Ateles geoffroyi (35) .....A.....CCA....G.....-..... New world monkeys
" " (2) .....A.....C.....C.....G.....GC...---.....
Lagothrix lagothrica (3) .....A.....C.....C.....G.....GC...---..... AGC deletion + Ceb stop
" " (7) .....A.....T.....CCA....G.....-.....G.C " "
Alouatta caraya (1) .....A.....C....G.....C....T.G.....C.-.....
" " (1) .....A.....CCA....C.....-.....T.....G..
Alouatta seniculus (4) .....A.....CCA....C.....-.....T.....G..
Cacajao rubicundus (8) ..T....A.....CCA....G.....-.....T.
Callicebus molloch (10) .CC.....A.....A.T....C..C....T.....C...C..G.....GC...---....A..... AGC deletion + Ceb stop
" RT 1 ...TCT.....A.....A.A...C.....C.....C.....G.....G-.....G..
" RT 2 ....CT.....A.....A.A...C.....C.....C.....G.....G-.....G..
" RT 3 ....CT.....T.....A.....A.A...C.....T.....C.....G.....-.....G..
Saimiri sciurus (10) .....A.....CCA....G.....-.....G..
Callithrix jacchus (2) T.....A.....CCA....G.....-.....GA.
Saguinus oedipus (29) .....A..G.....CCA....G.....-.....G..
Tarsisus syrichta (2) T.T...T.T....G.....C....C...AA.GC.....-...T.....C....TC Tarsisus
" (8) ..T...T.....G.....A....G...T..G.....GC...---.....A..T. AGC deletion
propithecus (4) ..T...T.....G.....C....CC.....G.....G....---.....T. " " Prosimians
flying lemur (9) A.T..CT.....TG.....A....CC.....G.....GGC...---.....T. " " Dermoptera
bat (3) T.T..C.....G.....C....CC..CA.A..... Chyroptera

succeeded using repair-competent Pfu polymerase (Fig. 2) (Lundberg et al., 1991).

PCR bands of the expected size were obtained only for primates and related species (Fig. 3a). Southern blot analysis of the PCR products using a human CEA gene TM specific probe, confirmed that they were CEA-related amplified sequences (Fig. 3b). The positive PCR bands were cloned and sequenced to confirm their validity and to compare their nucleotide and predicted amino acid sequences. All other species were negative with both sets of primers with both Taq and Taq/Pfu polymerase mixes, including tree shrew, opossum, rabbit, dog, cat, sheep, goat, cow, horse, mouse and rat. The latter rodent species are known to possess homologs of the human CEACAM1 gene and were negative because of known mismatches with the primers used. The failure to detect specific PCR products in many of these species therefore does not exclude the existence of more distantly related CEA-like homologs.

We obtained PCR products of the expected size for three closely related mammalian orders: Microchiroptera (bat, *Tadarida brasiliensis*), Dermoptera (colugo or flying lemur, *Cynocephalus volans*) and various groups of primates (Hominoids, Old world and New world monkeys, Tarsiers, and Prosimians). The nucleotide sequences of many independent clones of each of these TM domains from several independent PCR reactions indicate that the package of mutations giving GPI anchorage in the present-day CEA family, which we shall term the "CEA package", arose in a common ancestor of

Fig. 2. DNA agarose gel of a PCR experiment using S-1 and stop codon based AS-1 (Fig. 1) primers, showing specificity of the AS-1 primer for amplification of human CEA-like GPI-linked family members. Using Taq polymerase, AS-1 primer amplifies mutant CC1-tAT cDNA (lane 2) which contains the human CEA-like stop codon (Chapter 2) but because of one nucleotide mismatch at its 3′-end for template DNA, it does not amplify TM-linked human CEACAM1 cDNA (lane 3). However, using either Pfu polymerase alone (lane 5) or mix of Taq+Pfu (lane4) which can remove the 3′-end mismatch of primer to the template DNA, the CEACAM1 TM domain can be amplified. Arrowhead points out positive PCR band. The lower molecular weight band as investigated by Southern blot analysis (data not shown) is artifact.

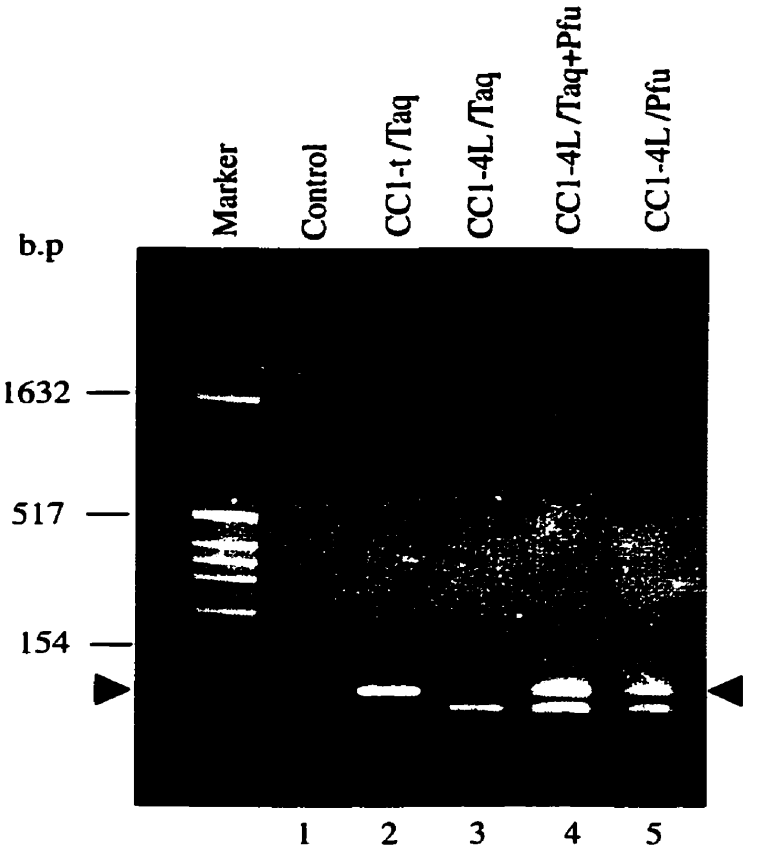

Marker
Control
CC1-t /Taq
CC1-4L /Taq
CC1-4L /Taq+Pfu
CC1-4L /Pfu
b.p
1632
517
154
1
2
3
4
5

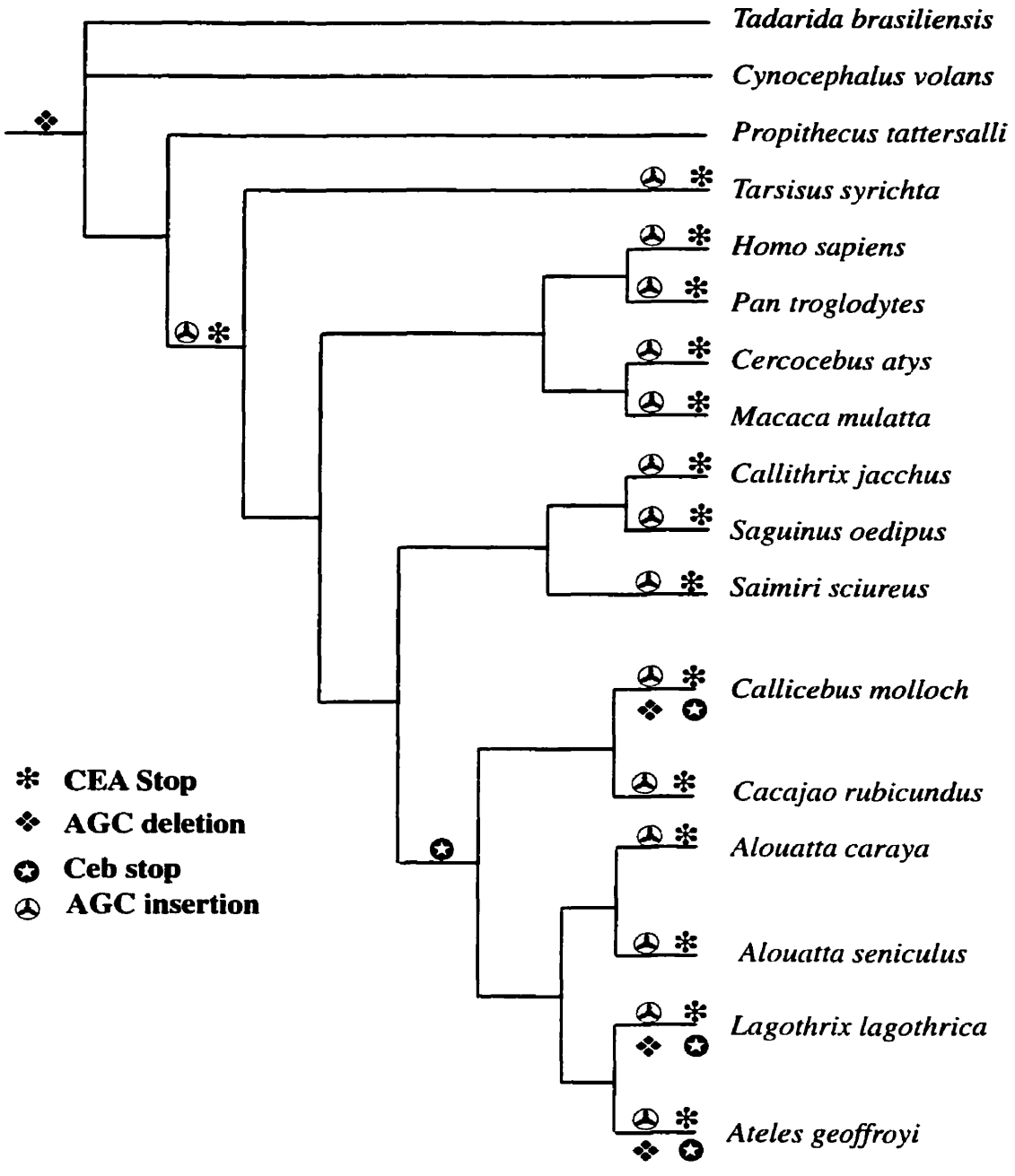
Fig. 3. a. Ethidium bromide stained DNA agarose gel of a typical PCR amplification of genomic DNA of various mammalian species screening for CEA family-like TM domains, using S-1/AS-2 (Fig. 1) primers. Arrowhead points out the position of the size of PCR bands expected for human.
b. Southern blot analysis of the PCR amplified DNA using human CEA TM specific probe (see text) indicating CEA positive PCR amplified bands.

a
B.P
1632
517
220
Marker
Control
Human
Bat
Flying lemur
Macaque
Callithrix
Chimpanzee
Saimiri
Callicebus
b
Human
Bat
Flying lemur
Macaque
Callithrix
Chimpanzee
Saimiri
Callicebus

tarsiers and anthropoids (Haplorrhini) (Figs. 1&4). Surprisingly, a second GPI mutational package arose, the "Ceb package", presumably starting first actually prior to the primate radiation in a common ancestor of Dermoptera, Propithecus and Tarsisus with a deletion of the third AGC repeat and followed later in the Cebidae radiation in New world monkeys with a stop codon just downstream of the AGC triplet deletion and other significant mutations in TM domain (see below) (Figs. 1&4); the failure to detect the Ceb package in the alouatta species of the Cebidae family is probably due to the relatively few clones that were sequenced for these species. The predicted C-terminal domain for the Ceb package is 4 amino acids longer than the CEA package. A cDNA construct consisting of the human CEACAM1-4L external domain linked to the Callicebus Molloch C-terminal domain with the Ceb package was efficiently processed to give GPI membrane linkage in two rodent cell transfectants and blocked myogenic differentiation (Naghibalhossaini F. & Stanners C. P. Ch. 4).

By comparing the amino acid sequences of various family members, several groups have pointed out the seeming lack of constraint on amino acid substitutions in the external domains of CEA (Rudert et al., 1989; Stanners et al., 1992) and PSG (Streydio et al., 1990) family members during evolution: the overall rates of synonymous mutations not giving rise to amino acid changes are similar to the rates of non-synonymous mutations giving amino acid substitutions. For most genes, the rates of non-synonymous mutations are

Figure 4. Phylogenetic relationship (Porter et al., 1997) and evolution of two different modes of stop codon (GPI-producing) mutations in CEA gene family members of various primates and nonprimate species.


Tadarida brasiliensis
Cynocephalus volans
Propithecus tattersalli
Tarsisus syrichta
Homo sapiens
Pan troglodytes
Cercocebus atys
Macaca mulatta
Callithrix jacchus
Saguinus oedipus
Saimiri sciureus
Callicebus molloch
Cacajao rubicundus
Alouatta caraya
Alouatta seniculus
Lagothrix lagothrica
Ateles geoffroyi
CEA Stop
AGC deletion
Ceb stop
AGC insertion

much lower than the synonymous rates (Kreitman and Akashi, 1995), which is consistent with the notion that many amino acid substitutions would be expected to be deleterious and therefore not retained. The TM exon of the CEA family and the closely related PSG family (Streydio et al., 1990) appears to represent an exception to the other exons in that here there is evidence of sequence conservation and adaptive amino acid substitution (Streydio et al., 1990). If we assume that the Ceb package (GPI anchored) evolved from the more ancient AGC deletion (still TM anchored) (Fig. 4), which seems likely since the Ceb package includes the same triplet deletion and excludes the finding of the triplet deletion alone as well in all species studied (Fig. 1 and 4), we find by comparison mutations giving not only the new stop codon but several others giving molecularly adaptive amino acid substitutions (Fig. 5). Thus, our recent work has shown that the introduction of a stop codon into the TM-linked domain of splice variant CEACAM1-4L can give GPI linkage but with an efficiency of only 10%; further mutations, that replace the normally hydrophobic amino acids in the TM-anchored domain of CEACAM1 downstream of the new GPI cleavage site with hydrophilic residues, are required for efficient GPI membrane linkage (Furukawa et al., 1997; Udenfriend and Kodukula, 1995). Such mutations, I688T, G690S, V692T and I693T, representing changes in 4 of 6 contiguous amino acid residues, extend the hydrophilic spacer, while retaining a short hydrophobic stretch of amino acids at the most C-terminus, that is necessary for efficient GPI processing and arose

Figure 5. Alignment of deduced amino acid sequence of TM domains with the Ceb stop (B) and TM domains with the AGC deletion but without Ceb stop (D) with TM domain of human GPI-anchored (A) and TM-anchored (C) CEA family members. Sequences are compared to the human CEA sequence. Dots represent identical residues and the star symbol indicates a stop codon. The site of cleavage/attachment to the GPI-anchor (ω site) in CEA and CEACAM6 has been shown by vertical arrows. Four conserved hydrophobic residues in TM-linked members which have been substituted with four hydrophilic residues in 3 cebidae species' TM domain, which improve the efficiency of GPI-processing (see text) are shown as bold underlined letters.

A

673 ▼

```
CEA          TVSASGTSPGLSAGATVGIMIGVLVGVALI*
CEACAM6      MITV.!SA.V...V.....T....AR....*
CEACAM8      DALVQ.S......R...S......AR....*
CEACAM7      YE.VQAS..D....TA.S......A.M...*
```

B

```
Lagothrix                        .S..T...A.....AALV*
Ateles                           .S..T...A.....AALV*
Callicebus              A....T..S.TT.L.A.....AALM*
```

C

```
CEACAM1      NYN.LPQEN...P..IA..V...VAL....AVALA
CEACAM3      FHVYQENA...PV..VA..VT........VAALVC
```

D

```
Tarsisus                         I...V....G.....AALVY
Propithecus                      I...V....A.....VALVY
Cynocephalus                     IA..V..V.A....MAALVY
```

in most GPI-anchored CEA family members with both the CEA and Ceb packages of mutations.

Intuitively, the probability that different stop codons along with the above 4 substitutions all affecting efficient GPI linkage arose randomly without significant selection by adaptive advantage twice during evolution seems remote. In a comparison of the entire TM domains of *Callicebus molloch* (Ceb package) with *Tarsisus syrichta* (triplet deletion only), the calculated nonsynonymous amino acid substitution rate according to the method of Nei and Gojoborii (Nei and Gojobori, 1986) was 0.27. Assuming a divergence time of 57 Mya (Porter et al., 1997), the amino acid substitution rate is 2.4×10^{-9} per site per year, which is higher than the rate at the highly mutable Ig kappa C region (Gillespie, 1991) and .7x the average nonsynonymous mutation rates in primates (Ohta, 1997). The mutation rate of the CEA family TM domain is therefore very high and this tends to be restricted to subdomains that involve GPI anchorage.

So far only a few cases of adaptive evolution have been explained at the molecular level (Kreitman and Akashi, 1995). Since adaptive substitutions are likely restricted to the protein domains, it would be difficult to find adaptive changes without concentrating on the domains which are functionally important. Change of TM domain and mode of cell surface

attachment in CEA family members represents another paradigm for adaptive evolution of proteins at the molecular level.

Lipid-like GPI structures allow more lateral movement in the membrane than TM anchors, as they are not linked to the cell cytoskeleton through cytoplasmic domains, and would be expected to change radically the repertoire of cell surface elements with which CEA family members could interact. These novel interactions apparently confer a cell and tissue state maintaining fluidity and lack of architecture and inhibiting cellular specialization (Benchimol et al., 1989; Ilantzis and Stanners, 1997); apoptosis is also inhibited (Ordonez C. & Stanners C. P. in preparation). Unlike many oncogenes, this is not achieved by stimulating cellular proliferation. We speculate that such a cell and tissue state, although it exacts the price of increased predilection for cancer in adult life, is advantageous during development where delay of specialization and structure by the explicit deployment of GPI-anchored CEA family members could offer improved temporal and spatial control of morphogenesis.

Acknowledgements. We thank Dr. Calvin A. Porter, Dr. Laurence Haney and Institutions listed in the method section for provision of specimens used in this study. This work was supported by grants from the NCI(C) and MRC of Canada to CPS. FN is a recipient of a studentship from the Ministry of Health and Medical Education, Government of Islamic Republic of Iran.

The following chapter describes the direct testing of the GPI-processing of the novel stop codon generating mechanism in CEA family members described in chapter 3. Since it has been shown that GPI-linked CEA family members CEA and CEACAM6 can block myogenic differentiation, but GPI-linked N-CAM and TM-linked CEACAM1 cannot, we studied the effect of ectopic expression of a CEACAM1 construct, in which the TM domain was exchanged with the novel GPI-generating sequence, on differentiation of rat L6 myoblasts.

Chapter 4

Tumorigenic Properties of a CEA Family GPI Anchor Found in New World Monkeys

Abstract

Glycophosphatidyl-inositol (GPI) anchorage of human carcinoembryonic antigen (CEA) family cell surface glycoproteins has been shown to confer tumorigenic properties relative to transmembrane (TM) anchorage, which is neutral or tumor suppressive. Furthermore, recent evidence supports the view that functional specificity resides in GPI anchors determined by different carboxy-terminal exons, since the NCAM GPI anchor, unlike the CEA GPI anchor, cannot confer such properties. The GPI-determining exons found in the human CEA family evolved in the primate radiation, being totally absent in rodents. A second CEA family exon predicted to generate GPI linkage by a different means evolved later in the Cebidae radiation of New World monkeys. It is important to determine whether the second GPI anchor conveys similar functional properties to the first, since independent evolution would then imply selective advantage of the tumorigenic properties of CEA. We show here that this novel exon, when linked to the external domains of human CEACAM1, is efficiently processed to give GPI anchorage. The novel GPI-determining domain, like the human CEA GPI domain, confers the ability to completely block the myogenic differentiation of rat L6 myoblasts; also, like CEA, this inhibition is due to perturbation of the functional activity of an integrin responsible for cellular binding to fibronectin. These results thus imply a paradox: that tumorigenic properties can confer adaptive function and positive selection.

Introduction

The two major modes of anchorage of cell surface proteins to external cell membranes, transmembrane (TM) and glycophosphatidyl inositol (GPI)-anchorage, would be expected to confer quite different properties on the proteins that they bind: TM-linkage allows direct coupling with cytoplasmic elements including components of the cytoskeleton whereas GPI-linkage leads to direct coupling only with other membrane components but allows greater lateral mobility in the plane of the membrane (**Brown and London, 1998; Edidin et al., 1991**). The different biological consequences of such linkages is well illustrated in the carcinoembryonic antigen (CEA) family of cell surface glycoproteins. This family represents a subset within the immunoglobulin gene superfamily (**Paxton et al., 1987**) and, in humans, consists of 29 gene-like sequences (including the related subfamily of pregnancy specific glycoproteins), closely clustered on chromosome 19, coding for proteins consisting of an amino-terminal V-like Ig domain followed by a variable number of C2-like Ig domains and terminated either by a hydrophobic processed domain giving GPI anchorage or by a TM domain and cytoplasmic tail (**Hammarström et al., 1998**).

The GPI and TM-linked members in the human CEA family exhibit quite different biological functions. Thus, although all members tested can mediate intercellular adhesion (**Stanners and Fuks, 1998**), GPI-linked

members CEA and CEACAM6 (formerly NCA) have been shown to be upregulated in many different types of cancers (Ilantzis et al., 1997; Shively and Beatty, 1985), whereas TM-linked CEACAM1 (formerly BGP) is usually downregulated (Neumaier et al., 1993; Nollau et al., 1997) and rodent CEACAM1 has been demonstrated to act like a tumor suppressor (Hsieh et al., 1995; Kunath et al., 1995; Luo et al., 1997). Consistent with these observations, ectopic expression or unregulated over-expression of CEA and CEACAM6 were shown to inhibit the myogenic differentiation of rat L6 myoblasts (Eidelman et al., 1993) and mouse C2C12 myoblasts (Screaton et al., 1997) and to disrupt cell polarity and tissue architecture of human colonic Caco-2 epithelial cells (Ilantzis and Stanners, submitted), whereas CEACAM1 had no such effects (Rojas et al., 1996). These effects lead to an increased tumorigenicity of L6 myoblasts (Screaton et al., 1997) and Caco2 colonocytes (Ilantzis etal., submitted) over-expressing CEA. In addition, CEA and CEACAM6 but not CEACAM1 inhibited anoikis of L6 and Caco2 cells, i.e., apoptosis due to deprivation of anchorage (Ordoñez at al., submitted). These effects of GPI-linked human CEA family members have been shown to be due to perturbation of the function of specific integrins, which is $\alpha_5\beta_1$ in the case of L6 myoblasts and Caco2 cells (Ordoñez et al., submitted).

The myogenic differentiation block has been demonstrated to depend critically on the mode of membrane anchorage, since exchanging the GPI and TM anchors of CEA and CEACAM1 reverses their effects; i.e., CEACAM1 can

be converted to a myogenic differentiation-blocking molecule by attaching its extracellular domains to the CEA GPI anchor and CEA attached to the TM domain of CEACAM1 loses its differentiation blocking activity (Screaton et al., submitted). Furthermore, GPI-linked NCAM125 did not block myogenic differentiation itself but could also be converted to a blocking molecule by switching its carboxy terminal GPI anchor-determining domain for the carboxy terminal domain of CEA (Screaton at al., submitted). This result implies specificity for biological function in the GPI anchor itself. Our present model for the structural requirements of CEA family members for the myogenic differentiation block is therefore as follows: self adhesive extracellular domains which confer clustering required for signaling, attached to the CEA-derived GPI anchor which confers the specificity of signaling and function (Screaton et al., submitted).

The CEA families of rodents and humans have been investigated in detail and show only TM-anchored members in rodents, and both GPI and TM anchored members in humans. It has been suggested, therefore, that GPI-linked CEA family members evolved from an ancestral CEACAM1-like gene after the radiation of rodents (Stanners et al., 1992). We have previously shown how, by only a few mutations, an efficiently processed GPI-anchored CEA family member could evolve from the TM-linked CEACAM1 gene (Naghibalhossaini et al., submitted). Thus, by introduction of a stop codon in the TM domain of splice variant CEACAM1-4L, at the corresponding position

to the GPI-linked CEA family members, and replacement of two other amino acid residues in the TM domain, CEACAM1 was converted into an efficiently processed GPI-linked protein (Naghibalhossaini et al., submitted). An evolutionary study of the nucleotide sequences of cloned genomic and cDNA carboxy-terminal domains of CEA family members demonstrated the emergence of the package of mutations giving CEA-like GPI linkage in common ancestor of Tarsiers and Anthropoids in the primate radiation (Naghibalhossaini et al., submitted). Interestingly, a second different package of mutations giving GPI linkage also appeared in the Cebidae arm of the New World monkey radiation, implying convergent evolution and adaptive function conferring selective advantage of GPI anchorage (Naghibalhossaini et al., submitted). The amino acid translation of this novel carboxy terminal exon indicates that the sequence contains the signals necessary for efficient GPI-anchorage. To test experimentally whether this is indeed the case, we made a chimeric cDNA of the novel carboxy terminal domain from one of the New World monkeys, *Callicebus Molloch* (CMO), linked to the extracellular domains of human CEACAM1-4L, which is denoted the "CC1-CMO" construct. We show here that the CC1-CMO construct determines a protein that is efficiently processed to give cell surface expression with GPI anchorage. Importantly, from the standpoint of exploring the function(s) conferred by GPI linkage giving convergent evolution, this construct, like CEA, could also block myogenic differentiation completely.

Materials and Methods

Plasmid constructs

CC1-CMO cDNA was constructed in two steps: (1) A 24mer sense primer (5'CAAGAAAATGCCCTCTCAGCTGGG3'), which includes a 9 nucleotide tail corresponding to nucleotides 1333-1341 of the 5' end of the human CEACAM1 TM exon, and a 21mer antisense primer (5'CAGTCTTCCTGAAATGAAGAA3') complementary to the indicated CEA cDNA sequence downstream of its stop codon (Fig. 1), were used to amplify a 79 bp cloned CEA-related genomic TM exon of CMO monkey which has the novel stop codon generating mutational package of Cebidae (Fig. 1) (Naghibalhossaini et al., submitted). The PCR product was gel purified using the QIAEX (QIAGEN, Chatsworth, CA) DNA gel extraction kit. (2) The 109 bp PCR product was used as an antisense megaprimer along with a 24mer CEACAM1 sense primer corresponding to nucleoides 1445-1468 of the human CEACAM1 cDNA (5'GCAGGGCAAGCGACCAGCGTGATC3') for inverse PCR-based mutagenesis, as described (Clackson et al., 1992), using CEACAM1-4L-pBluescript SK as a template to make the CC1-CMO cDNA chimera consisting of human CEACAM1-4L extracellular domains linked to CMO monkey TM exon. We used Pfu polymerase (Stratagene, La Jolla, CA) instead of Taq polymerase in all PCR reactions for higher fidelity of amplification and to prevent nontemplate-dependent addition of an extra A residue at the 3'-end of PCR products. The coding sequence was verified by dideoxy sequencing

(Pharmacia Biotech T7 sequencing™ kit). The sequence at the CEACAM1/CMO junction is shown in Fig. 1.

Cell Cultures and Transfection

LR-73, a CHO-derived cell line (Pollard and Stanners, 1979) and rat L6 myoblasts (Yaffe, 1968) were co-transfected with wild type and chimeric cDNA inserted into the P91023B expression vector (courtesy of R. Kaufman, Genetics Institute, Boston, MA) and with pSV2Neo by the calcium phosphate precipitation, as previously described (Naghibalhossaini et al., submitted). Geneticin (G-418 sulfate, Gibco BRL)-resistant total transfectant populations were selected with 400ug/ml G418 for 10-14 days. The pooled G418-resistant colonies were sorted for high cell surface expression of CEACAM1 proteins by FACS using polyclonal rabbit antihuman CEA antibody, as described previously (Zhou et al., 1993a). Pooled populations expressing relatively high cell surface levels of CEACAM1 proteins were used in experiments to avoid difficulties in interpretation due to the possibility of clonal variation in cellular properties unrelated to the effects of the test molecules. G418 was removed from growth media 24 hr before the application of the various assays.

GPI-Processing Assay

The GPI-anchorage of mutant chimeric proteins to the cell surface was verified both by phosphatidylinositol-specific phospholipase C (PI-PLC) treatment and cold nonionic detergent extraction of LR-73 transfectants, as described previously (Naghibalhossaini et al., submitted). Briefly, duplicate aliquots of $5X10^5$ cells, were suspended in serum free culture medium diluted with PBS + 0.2% BSA; one aliquot was treated with 0.2 units PI-PLC from Bacillus cereus (Boehringer Mannheim) and the other without treatment as a control. After 1 hr incubation the residual levels of cell surface CEACAM1 protein were measured by FACS analysis as described previously (Zhou et al., 1993a).

For nonionic detergent extraction, $2.5X10^6$ LR-73 transfectants cells were suspended in 250ul of ice cold lysis buffer containing 1% Triton X-100 at pH 6.5. After 10 min. incubation on ice, they were centrifuged at 15,000g at 4^oC for 20 min. 250ul 2XSDS PAGE sample buffer (Laemmli, 1970) was added to the supernatant and 500ul 1XSDS sample buffer to the pellet. Lysates were boiled for 10 min. and 20ul of each supernatant and sediment were analyzed by electrophoresis on 7.5% SDS-polyacrylamide gels and detected for the CEACAM1 proteins by immunoblotting as described previously (Chapter 2).

Myogenic Differentiation and Fusion Assay

To promote fusion and differentiation, rat L6 myoblasts were seeded at 10^4 cells/cm^2 in 60-mm tissue culture plastic Petri dishes in 4 ml growth medium (D-MEM plus 10% FBS), and cultured without changing the medium for 3 days. When cells were nearly confluent, the medium was replaced with 4 ml differentiation medium (D-MEM plus 2% horse serum), and the cells cultured for an additional 7 days. The ability of cells to fuse was assessed by assessing the number of nuclei in cells with >3 nuclei and dividing by the total number of nuclei, as described previously (Eidelman et al., 1993).

Fibronectin Binding Assay

Transfected L6 cells were seeded at $1x10^4$ cells/cm^2 in growth medium and incubated for 4 days. Confluent cells were collected and suspended at a concentration of $4x10^5$ cells/ml in serum free DMEM. A 24 well plates coated with 10 ug/ml of purified fibronectin (Sigma-Aldrich, Oakville, ON) overnight at 4°C followed by incubation with 1% BSA for a minimum of 8 hours at 4°C to block fibronectin uncoated surface, then washed with serum-free DMEM. 250 ul/well of the Cellular suspension, was allowed to adhere to the fibronectin-coated wells for one hour, after which the wells were washed twice with PBS to remove unattached cells. The remaining attached cells were

removed with trypsin and their number determined using a particle counter (Coulter Electronic Inc. Hialeach, FL).

Results

In an evolutionary study, we previously reported molecular evidence suggesting that GPI-anchored CEA family members evolved from more primitive TM-linked CEA members twice independently (Naghibalhossaini et al., submitted). Three species in the Cebidae radiation of New World monkeys showed a CEA family member with a novel carboxy terminal exon containing a different type of stop codon-generating mutational mechanism, consisting mainly of a triplet deletion followed by a new stop codon downstream of the stop codon commonly found in all members of the human CEA family. Visual inspection of their TM sequence shows they have the necessary signals to encode for GPI-cell surface attachment of proteins (Englund, 1993). We suggested from this finding that the acquisition of GPI cell surface anchorage in CEA family members in the course of evolution could indicate adaptive advantage giving positive selection (Naghibalhossaini et al., submitted).

The novel stop codon-containing TM exon of the CMO CEA family member encodes a GPI-anchored protein.

To verify experimentally whether the novel stop codon generating mutational mechanism leads to the expression of a cell surface GPI-linked protein, a chimeric cDNA consisting of 25 amino acids from the carboxy terminal exon of the CMO monkey linked to the extracellular domains of

human TM-anchored CEACAM1-4L was constructed. This chimeric cDNA, denoted "CC1-CMO", which was expected to utilize the upstream Ala or Ser residue of the CMO TM domain as the ω site for cleavage and addition of the GPI anchor (Fig. 1b), was stably transfected into the CHO-derived cell line, LR-73. The cell surface localization of the expressed CC1-CMO chimeric protein was confirmed by FACS analysis of whole transfectant cells; the GPI-cell surface anchorage of the protein was demonstrated by its complete sensitivity to PI-PLC digestion (Fig. 2).

The GPI-linkage and efficiency of GPI processing of the CC1-CMO protein was also verified by a cold nonionic detergent solubilization assay. GPI-linked proteins form a complex in the cell membrane with sphingolipids and cholestrol, which is insoluble in nonionic detergents such as Triton X-100 at 4°C (Brown and London, 1997; Simons and Ikonen, 1997). TM-linked proteins, on the other hand, are soluble under the same conditions. Immunoblot analysis of soluble and insoluble fractions after cold Triton X100 extraction of LR-73 transfectant cells showed that CEACAM1 (denoted CC1-4L) was soluble and CEA was mainly insoluble, as expected; under the same conditions, CC1-CMO was mainly insoluble, indicating GPI anchorage (Fig. 3). We previously showed that incomplete GPI-processing of CEACAM1 proteins with mutant TM domains could be distinguished by the presence of a less glycosylated lower MW, TX-100 soluble band (Naghibalhossaini et al., submitted). Since no lower MW, TX-100 soluble band could be seen in the

Figure 1. Sequence comparison (a. nucleotides, b. amino acids) of the transmembrane domain of CEA gene family members of humans and Callicebus molloch (CMO) monkey. Underlined bases in panel **a** and star signs in panel **b**, represent stop codon positions and dashes show naturally occurring base deletions. Horizontal arrows in panel **a** show the PCR primers that were used to amplify the CMO TM exon in making the CC1-CMO construct. The tilted 5′-end of the sense primer represents a 9 nucleotide tail which matches the CEACAM1 cDNA at the corresponding position (see Materials and Methods). The amino acids indicated by ω in panel **b** represent possible points of cleavage for addition of the GPI anchor.

a. Nucleotide Sequences

```
CEA          GCATCTGGAACTTCTCCTGGTCTCTCAGCTGGGGCCACTGTCGGCATCATGATTGGAGTGCTGGTTGGGGTTGCTCTGATA-TAGCAGCCCTGGTGTAGTTTCTTCATTTCAGGAAGACTG

CMO-CEACAM                                   .CC...........A........A.T...C..C....T......C...C..G.........GC...---....A.......

                                                  Sense                                                                      Antisense

CC1-4L   ATAAT..TCTACC.CAAGAAAA...C......C.........T..CT.....TG..........AG...CCCT.............GC..T........CA.GT.....G......G.......C.GC
CC1-CMO  ATAAT..TCTACC.CAAGAAAA..CC...........A........A.T...C..C....T......C...C..G.........GC...---....A.........................
```

b. Amino Acid Sequences

```
CEA          TVSASGTSPGLSAGATVGIMIGVLVGVALI*
CC1-4L       NYN.LPQEN...P..IA..V...VAL....AVALA.....
CMO-CEACAM            A....T..S.TT.L.A.....AALM*
CC1-CMO      NYN.LPQENA....T..S.TT.L.A.....AALM*
                      ↑   ↑
                      ω   ω
```

Figure 2. FACS analysis of PI-PLC sensitivity of CC1-CMO protein on cell surfaces. CC1-CMO, CEACAM1-4L (CC1-4L) and control Neo LR-73 transfectants were treated with PIPL-C and cell surface levels before (black profiles) and after (white profiles) treatment assessed by FACS analysis. The CC1-CMO construct showed a reduction of cell surface protein, but Neo and CC1-4L (TM-linked) controls showed no sensitivity to treatment.

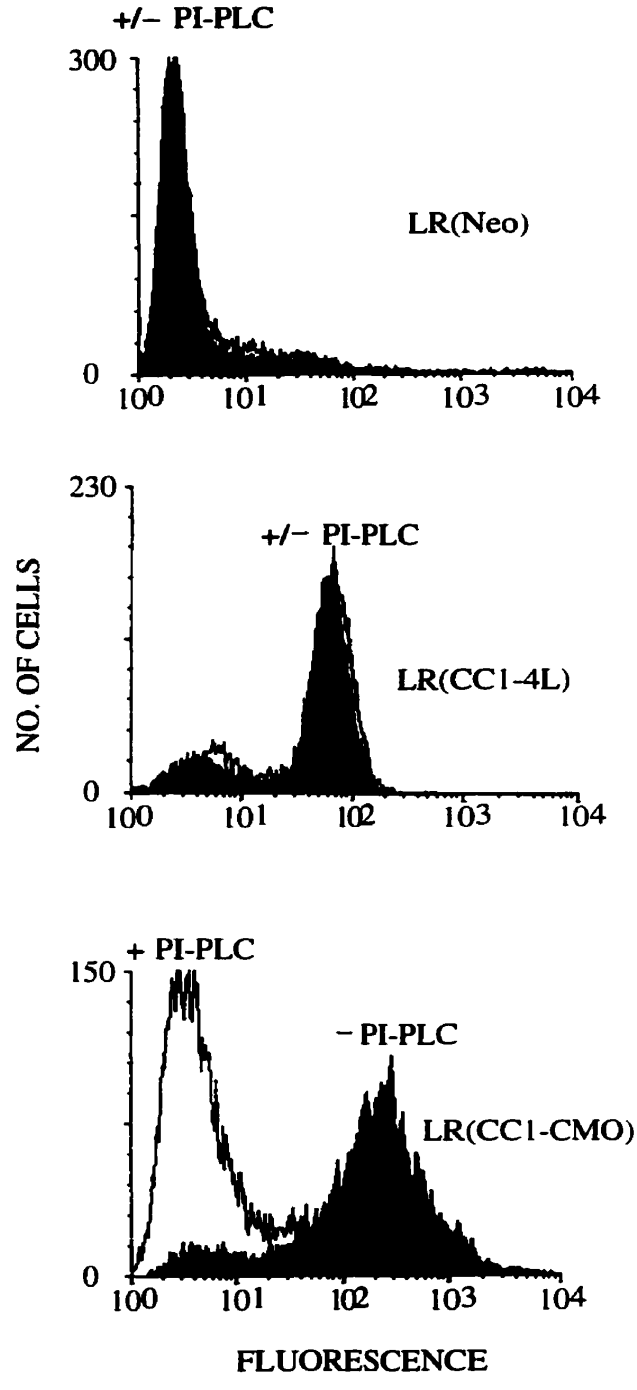
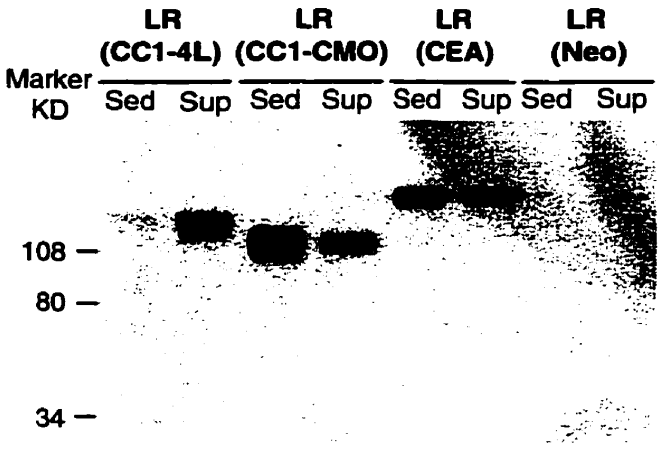

+/− PI-PLC
300
0
LR(Neo)
100
101
102
103
104
230
+/− PI-PLC
LR(CC1-4L)
NO. OF CELLS
+ PI-PLC
150
− PI-PLC
LR(CC1-CMO)
FLUORESCENCE

Figure 3. GPI-processing efficiency of the CC1-CMO protein assessed by cold nonionic detergent solubility. Western blot analysis of cold Triton X-100 extracts of CC1-CMO, CC1-4L, CEA and control Neo LR-73 transfectants. CC1-CMO protein was mainly insoluble in cold TX-100 as was CEA, but CC1-4L protein was mainly soluble. No soluble band of lower MW, indicative of inefficient processing, could be detected for CC1-CMO protein.

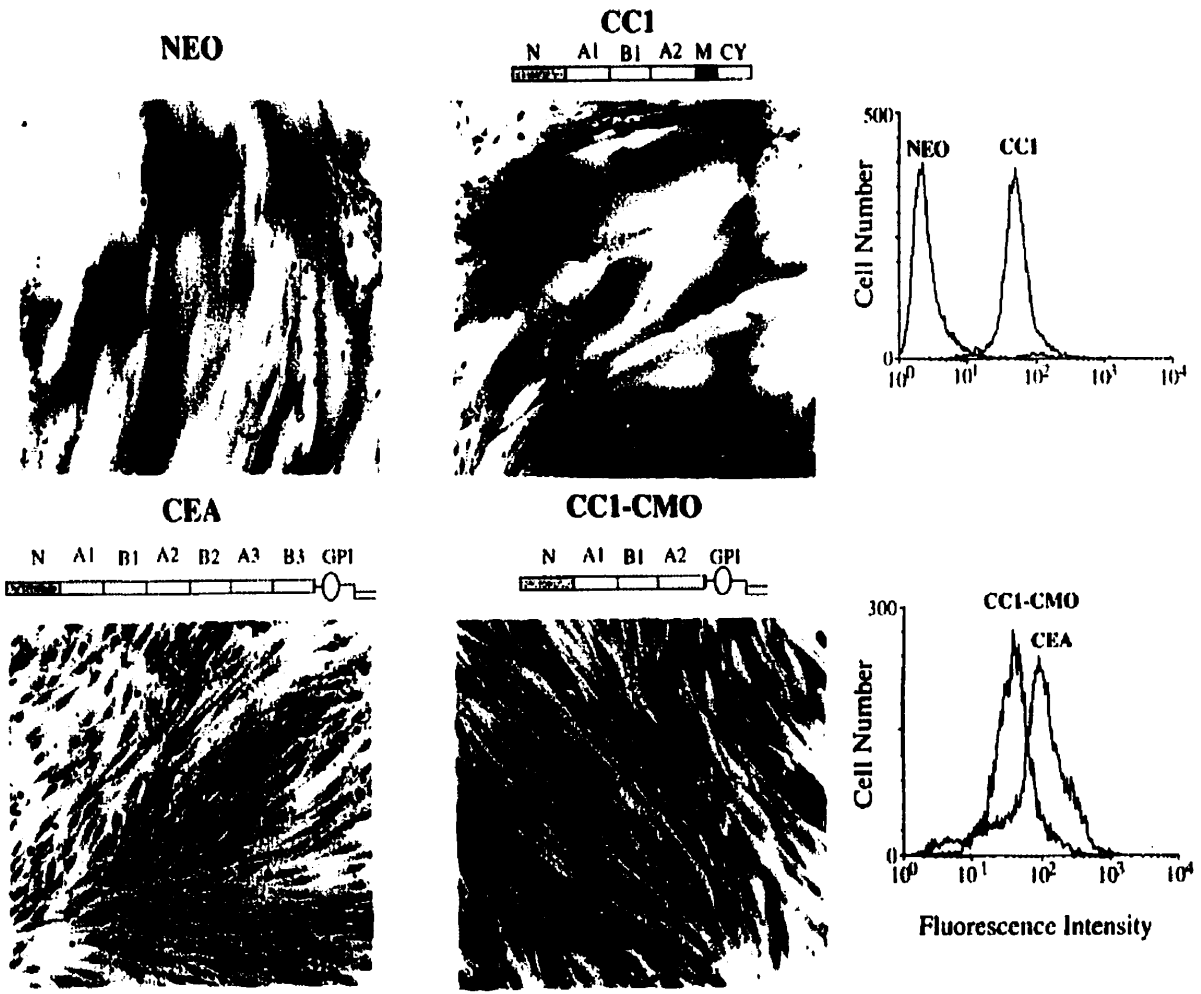

LR (CC1-4L)
LR (CC1-CMO)
LR (CEA)
LR (Neo)
Marker KD
Sed
Sup
Sed
Sup
Sed
Sup
Sed
Sup
108
80
34

immunoblot of the CC1-CMO chimeric protein (Fig. 3), we conclude that this chimeric protein, as with the naturally occurring GPI-linked human CEA family members, is efficiently GPI-processed.

The novel GPI-anchored protein blocks myogenic differentiation

Having demonstrated that the CC1-CMO chimeric protein was in fact GPI-anchored, it remained to explore the functional implications of the acquisition of this novel GPI anchor. To this end, CC1-CMO cDNA was transfected into rat L6 myoblasts and, after selection of a pooled population of clones stably expressing relatively high cell surface levels of CC1-CMO protein, their ability to differentiate and fuse into multinuclear myotubes was assessed. Parallel pooled stable transfectant populations expressing CEA and CEACAM1 were also assessed for their differentiation capacity as positive and negative controls, respectively. The results indicate that, while L6(Neo) (vector only control) and L6(CEACAM1) total transfectant populations started to fuse on day 3 and showed more than 70-80% fusion after 7 days in differentiation medium, L6 (CC1-CMO) total transfectants, like CEA transfectants, were totally incapable of fusing and forming myotubes, even after 9 days in differentiation medium (Fig. 4). The complete block in myogenic differentiation by CC1-CMO was confirmed by the absence of staining with anti-myosin antibody (data not shown). The results were reproducible and were repeated for at least two independently isolated pooled total transfectant populations. These results demonstrate that the novel CEA

Figure 4. Effect of CC1-CMO protein on myogenic differentiation of L6 myoblasts. Photomicrographs of hematoxylin-stained cultures of various rat L6 myoblast transfectants incubated in differentiation medium for 7 days. FACS profiles show the relative cell surface expression levels of the indicated proteins.

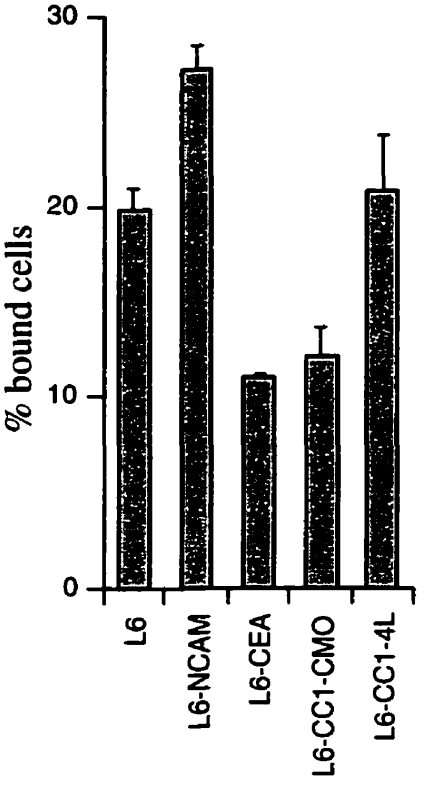

NEO
CC1
N A1 B1 A2 M CY
500
NEO
CC1
Cell Number
0
10^0 10^1 10^2 10^3 10^4
CEA
N A1 B1 A2 B2 A3 B3 GPI
CC1-CMO
N A1 B1 A2 GPI
300
CC1-CMO
CEA
Cell Number
0
10^0 10^1 10^2 10^3 10^4
Fluorescence Intensity

family GPI anchor found in CMO monkeys can, like the human CEA-derived GPI anchor, confer the property of abrogation of myogenic differentiation.

The novel GPI-anchored protein inhibits binding to fibronectin

As outlined in the Introduction, the molecular basis for the inhibition of differentiation and distortion of tissue architecture by CEA expression has been shown to be due to a perturbation in the function of specific integrins which, in the case of L6 myoblasts from confluent cultures, results in an inhibition of binding to the extra-cellular matrix and to a major component of the extra-cellular matrix, fibronectin. To test whether such a perturbation could also underlie the myogenic differentiation block mediated by the CC1-CMO protein, binding of L6 CC1-CMO transfectants relative to control transfectants to fibronectin was assessed. The results of several independent experiments (Fig. 5) show that CC1-CMO L6 transfectants demonstrate a similar inhibition of binding to fibronectin relative to parental L6 cells as CEA transfectants, whereas GPI-linked NCAM and CEACAM1 transfectants show no such effect. Thus, the inhibition of myogenic differentiation by the CC1-CMO protein is likely to be due to specific integrin perturbation.

Figure 5. **The novel GPI-anchored protein inhibits binding to fibronectin.** L6 transfected myoblasts expressing CEA or a fusion protein that contains the CMO specific anchor and the extracellular domain of the CC1-4L glycoprotein bound to fibronectin less than L6 parental, L6 transfected myoblasts expressing GPI-linked NCAM or L6 myoblasts expressing the whole transmembrane CC1-4L glycoprotein.


30
20
10
0
% bound cells
L6
L6-NCAM
L6-CEA
L6-CC1-CMO
L6-CC1-4L

Discussion

There is increasing evidence suggesting that the GPI mode of cell surface attachment of proteins confers functional properties on cells which differ from those of TM-linked proteins (for review see Brown, 1998 #221) and, as outlined in the Introduction, our previous results with GPI and TM-anchored CEA family members supports this contention.

In an evolutionary study, a novel package of mutations in a carboxy terminal exon of the CEA family of the Cebidae radiation of New World monkeys was discovered which was predicted to give GPI anchorage (Naghibalhossaini et al., submitted). This mutational package differs from that used by human GPI-linked CEA family members; the fact that GPI anchorage seems to have evolved twice independently implies convergent and parallel evolution for a structural feature (i.e., the GPI anchor) that confers functions giving adaptive advantage and positive selection (Zhang and Kumar, 1997). Since this finding represents an opportunity to gain insight into the normal functions of the GPI-linked members of the CEA family and considering the evidence that GPI anchors derived from the expression of different carboxy terminal exons can determine specificity of function (see Introduction), it is important to investigate whether the novel GPI anchor confers functions similar to those of the common human CEA GPI anchor.

In this work, we first confirmed experimentally, by sensitivity to PIPL-C digestion and insolubility in cold non-ionic detergent solutions, that the CEA

carboxy terminal exon of the CMO New World monkey in fact specifies the GPI mode of cell surface attachment of proteins. In addition, in agreement with an analysis of the predicted amino acid sequence coded for by this exon, the carboxy-terminal sequence was efficiently processed. Efficient processing requires mutations and amino acid substitutions between the cleavage site and stop codon (Naghibalhossaini et al., submitted); the presence of these mutations in both types of carboxy terminal exons giving GPI anchorage therefore represents further evidence for positive selection for this mode of membrane anchorage.

The investigation of the biological functions conferred by the novel GPI anchor revealed that expression of the CC1-CMO chimeric construct in rat L6 myoblasts completely blocked myogenic differentiation and fusion, exactly as the common human GPI anchor. In the latter case, the myogenic differentiation block is due to a perturbation in the function of the $\alpha_5\beta_1$ integrin, as shown by the effects of monoclonal antibodies against this integrin and against its chief ligand, fibronectin (Ordoñez et al., submitted). The CC1-CMO construct had the same effect on cellular binding to fibronectin as CEA, thus implying the same perturbation of $\alpha_5\beta_1$ function as that produced by CEA.

In all, the CC1-CMO construct had the same biological functions as the CC1-CEA construct studied previously (Screaton et al., submitted): the

inhibition of cellular differentiation due to integrin perturbation. Other functions of CEA, such as the inhibition of anoikis, remain to be investigated. The tissue specific pattern of expression of the new GPI-determining exon in New World monkeys, which might provide clues regarding *in vivo* function, has not yet been determined The positive selection of a molecular feature conferring tumorigenic properties twice during evolution represents a paradox. Speculations concerning possible utility of such a function have been given elsewhere (Naghibalhossaini et al., submitted).

Chapter 5

Thesis Conclusions and General Discussion

Discussion and future direction

As reported here and elsewhere for the CEA family, the mode of anchorage to the cell surface has a dramatic effect on the functional properties of the molecule. GPI-linked CEA and CEACAM6 show differences with TM-linked CEACAM1 in several aspects due to the difference of the type of anchorage to the cell surface (Rojas et al., 1996). Although, several functions have been ascribed for CEA gene family members (Obrink, 1997; Stanners, 1998; Stanners and Fuks, 1998), there is no clear picture of their function *in vivo*. Most of the studies have been carried out on CEA family members of humans and rodents. Little is known about the extent of CEA family members in other species. The presence of antigens in some primates which cross react with antihuman CEA or CEACAM6 antibodies has been already reported (Engvall et al., 1976; Haagensen et al., 1982; Tobi et al., 1994). A biological system can only be fully understood when its evolutionary history has been elucidated. The work presented here provides the first detailed study of CEA family members in species other than rodents and humans. A study of the appearance of GPI-anchorage in the CEA family during evolution could provide important insights into its function *in vivo*. The purpose of this thesis, therefore, was to study the pattern of acquisition of GPI-anchors in the course of evolution.

Since only TM-linked CEA family members have been found in rodents, but humans have both GPI and TM-linked CEA family members, it has been suggested that acquisition of GPI-membrane anchorage by CEA family members is a recent evolutionary event (**Stanners et al., 1992**). The first part of the work (chapter 2), verifies the feasibility of derivation of the GPI-linked CEA family members from the more primordial, conserved TM-linked member, CEACAM1, in the course of evolution. Previous studies on other GPI-linked proteins has revealed that a signal at the C-terminus of proteins is required for GPI-anchor addition. However, there is no clear consensus sequence for GPI-processing of proteins (**Englund, 1993**). The key mutations in the CEACAM1-4L transmembrane domain, which are required to shift the mode of anchorage from TM to GPI-linked were determined by site specific mutagenesis. Introduction of a stop codon in the CEACAM1-4L TM domain, at the corresponding position to the stop codon naturally found in the C-terminal domain of GPI-linked CEA members resulted in activation of a latent GPI-processing signal. The mutant, truncated CEACAM1, was GPI-processed, although not efficiently, as verified by PI-PLC treatment and cold triton X-100 extraction of the transfected cells. By substituting one amino acid just down stream of the cleavage site of the TM domain, in addition to the stop codon, the GPI-processing efficiency of mutant CEACAM1 protein increased significantly. But the GPI-processing still was not complete. By substituting one more amino acid residue, an Ile by a conserved Thr of GPI-linked CEA family members, in the middle of the TM domain, TM-linked

CEACAM1 was converted to an efficiently GPI-processed protein, comparable to the naturally occurring GPI-linked CEA members. Therefore, these results demonstrated how easily, by only a few changes in the TM domain, GPI-linked CEA family members could have evolved from a more primordial TM-linked member. These results also showed the essential role of the introduction of a stop codon in the TM domain of CEACAM1 for activation of the latent GPI-processing signal and the role of other amino acid substitutions for increasing the GPI-processing efficiency of the protein. The change of mode of cell surface attachment from TM to GPI did not abolish the intercellular adhesive property of CEACAM1-4L molecule but, interestingly like CEA/CEACAM6 conferred myogenic differentiation blocking capacity to it and modified cell adhesion to fibronectin, a component of the ECM. It has been argued that positive natural selection is needed for acquisition a gene family with a new function (Ohta, 1991). We therefore suggest that, in the course of evolution of CEA family, the mutations which possibly happened in the CEACAM1-4L tail resulting in truncation gave rise to a GPI-anchored version of the protein in low yield that could be tested by the cell for an advantageous change in function. Since the protein was functionally selected, very few additional mutations could have resulted in its expression at the cell surface in high yield.

To understand the pattern of the evolution of GPI-linked CEA gene family members, we established a rapid PCR based screening. On the basis of

the assumption that the above mentioned stop codon is indicative of GPI-linkage of CEA family members, we designed primers specific for the TM exons of the CEA family and applied them to PCR screening of genomic DNA form various mammalian species. Because of the high homology between CEA family members, this stop codon was the only key marker used to distinguish GPI-linked from TM-linked CEA family members by this technique. Sequencing of the positive PCR bands from various species revealed two different modes of generating a stop codon in CEA family members. The same mode of generating a stop codon as humans was found in various group of primates such as Hominoids, Old world monkeys, New world monkeys, and Tarsisus. A second novel mechanism of stop codon generation in the CEA family was found in three new world monkey species. A chimeric construct of the TM domain, containing this novel stop codon-generating sequence linked to the CEACAM1-4L extracellular domain, was GPI-processed efficiently as shown in Chapter 4. Flying lemur and two other primitive primates (Tarsisus and Propitecus) showed a homologous TM domain with half of the mutations of the novel GPI generating TM sequence ; a trinucleotide deletion, but no stop codon. We assume this sequence encodes a TM-linked CEA family member in these primitive species. This TM-linked CEA family member of primitive species was assumed to be the ancestral gene for the novel GPI-anchored CEA family members of new world monkeys. Therefore, it appears that the GPI-anchored genes evolved twice from TM-linked genes independently in CEA family. The same mode of

generating a stop codon found in human GPI-linked CEA family members evolved from a TM-linked CEACAM1-like gene before radiation of the primitive primate group, called Tarsiers, whereas a second novel mechanism of stop codon generation evolved independently from another TM-linked CEA member later in Cebidae family of new world monkeys.

Diversification of the type of cell surface attachment following gene duplication was possibly the mechanism for evolution of some other GPI-anchored proteins from TM-linked proteins in Ig superfamily. Duplicated genes encode GPI- and TM-linked forms of decay-accelerating factor (DAF) proteins in mouse (Spicer et al., 1995). A similar mechanism possibly resulted in the evolution of murine GPI-anchored Class I transplantation antigen Qa-2. Replacement of a single amino acid Asp in TM domain of Qa-2 by the corresponding Val residue of TM-anchored members, converted it to a transmembrane protein (Waneck et al., 1988). Fcγ receptor III molecules in humans are encoded by two alternative membrane-anchored forms by two separate genes, the GPI-linked form (FcγRIIIB) for which expression is restricted to primate neutrophils and a highly homologous and more widely distributed TM-linked form (FcγRIIIA). It has been shown that only a single amino acid substitution in upstream TM domain of FcγRIIIB plus a shorter cytoplasmic tail than FcγRIIIA (4 A.A of IIIB versus 25 A.A residues in IIIA) is sufficient for efficient GPI-processing (Kurosaki and Ravetch, 1989). mRNA

splicing is another mechanism of generation of protein isoforms with different modes of membrane attachment in the Ig superfamily, such as LFA-3 (Wallich et al., 1998), N-CAM (Cunningham et al., 1987) and guinea pig DAF protein (Nonaka et al., 1995).

It has been proposed that selective constraints are relaxed after gene duplication (Lewin, 1994a). A duplicated gene is relatively free to evolve to gain a new function. In this context it is plausible to think that, after duplication of the primordial TM-linked CEA gene, one copy retained the original function and the other copy accumulated enough mutations in its transmembrane domain to be converted to a GPI-linked CEA member.

There are two major views regarding the possible mechanism of amino acid replacements in proteins in the course of evolution (Kreitman and Akashi, 1995). Either the amino acid substitutions have had selective advantage by which they improve the functional condition of the molecule or they have been replaced by chance (random genetic drift), being selectively neutral. Evolution of the novel GPI-linked CEA family member in new world monkeys from a TM-linked CEA gene is unlikely to have happened by chance. Conversion of a TM-linked to a GPI-linked protein, needs highly ordered amino acid replacements. Only certain amino acids in a specific regions at the C-terminal domain of proteins are compatible with GPI-processing (Englund, 1993; Udenfriend and Kodukula, 1995). Also, a major

change in protein structure such as conversion of the mode of anchorage from TM to GPI and its functional consequences could not be considered as a neutral event. It has been proposed that the rate of evolution in primates is slower than that of other mammalian groups (Gibbons, 1995; Li et al., 1996), since the majority of amino acid replacements in proteins are deleterious and subject to purifying selection. Thus, an accelerated directional amino acid replacement (discussed in Chapter 3) which gave rise to the novel GPI-anchored CEA family members, suggests that the TM domain of this gene was subjected to non-neutral mechanisms of change. It has been suggested that CEA gene family members are in a state of transitional evolution and still evolving rapidly (Stanners et al., 1992; Streydio et al., 1990). It seems that the selective pressure is on the secondary and tertiary structure (Ig fold) of the CEA family members, but not on the primary structure (Rudert et al., 1989). After evolution of the first GPI-linked CEA gene, the family expanded by gene duplications (Hammarström et al., 1998). Contrary to rodents, which lack any GPI-anchored CEA members, most of the genes in the CEACAM subgroup of the human CEA family are GPI-linked. The TM domain of the four human GPI-linked CEA family members (CEA, CEACAM6, CEACAM8, and CEACAM7) also diverged in some degree from each other. However, the necessary molecular signals for efficient GPI-processing have been conserved. The significance of this conservation will be more evident when considering the complexity of the GPI-processing signal. Most amino acid replacements at the C-terminal of a GPI-linked protein would affect GPI-processing efficiency

and would result in loss of or decreased expression at the cell surface or even conversion to a TM-linked protein (Waneck et al., 1988).

We do not know the functional advantage of acquisition of GPI-anchorage in CEA family. The importance of GPI-anchorage in maintaining cell function and structure is only beginning to be understood. Certain functional advantages which this type of attachment to the cell surface confers to the protein have been discussed, such as increased lateral mobility, sorting to the apical surface in polarized cells, accessibility to regulation by cell surface or extracellular proteolytic or lipolytic enzymes and involvement in different signal transduction pathways (see Chapter 1). Several studies suggest that CEA family members in human hematopoietic cells are possibly involved in transmembrane signalling. Monoclonal antibodies (mAb) against CEACAM8, a GPI-anchored CEA family member, induce an increase in cytoplasmic Ca^{2+} and an oxidative burst in neutrophils (Lund-Johansen et al., 1993). Protein kinase activity was found in Immunoprecipitates of CEACAM1-4L, CEACAM8 and CEACAM6 in neutrophils (Skubitz et al., 1995) and Src family kinases, lyn and hck, were the associated tyrosine kinases in the CEA family member immunoprecipitates in neutrophils. In another study, Src kinases lyn and hck were also found associated with CEA, but not CEACAM1, in cold nonionic detergent extracts of rat basophilic leukemia (RBL) cells (Draber and Skubitz, 1998). Src family kinases have also been

found to form complexes with GPI-anchored proteins in other cells, as described in the Introduction.

CEA and CEACAM6 for a long time have been considered to be oncofetal antigens and tumor markers. An instrumental role has been also proposed for CEA/CEACAM6 in promoting carcinogenesis (Ilantzis et al., 1997). Therefore, it might seem that GPI-anchor acquisition during evolution by CEA members was a disadvantageous event which could increase cancer risk. However, if this is the case, they would have been eliminated by purifying selection. The convergent and independent evolution of the GPI-anchored CEA family members through different genetic mechanisms indicates that the acquisition of GPI-anchor by CEA family members represents a favored character in higher mammalian species. These data support the notion that the GPI-anchored CEA family members have important functions *in vivo*.

GPI-linked CEA and CEACAM6 block myogenic differentiation, whereas TM-linked CEACAM1 does not (Rojas et al., 1996). GPI-linked human N-CAM-125 is also unable to block and actually accelerates myogenic differentiation (Dickson et al., 1990). However, GPI-linked N-CAM and TM-linked CEACAM1, could be converted into differentiation-blocking molecules by swapping their GPI-anchors or TM, respectively, with the CEA GPI anchor (Screaton R. and Stanners C.P. in preparation). The functional implications of

a novel mechanism for generating GPI-linked sequences in the CEA family were tested for blocking myogenic differentiation of L6 myoblasts in Chapter 4. Ectopic expression of the CEACAM1 construct containing the novel GPI-generating sequence from Callicebus molloch in rat L6 myoblasts, like GPI-linked CEA and CEACAM6, completely blocked myogenic differentiation and inhibited binding to the ECM component, Fibronectin. GPI linked CEA and CEACAM6 are overexpressed during embryonic development of the gastrointestinal tract and in colonic tumors. A possible role for GPI-linked CEA family members in inhibiting differentiation during embryonic development of the colon in situations where cell populations need to be held together in an undifferentiated state as morphogenesis proceeds and/or in the maintenance of tissue architecture in adult colonic tissue has been suggested (Benchimol et al., 1989; Stanners, 1998). Three GPI-linked CEA family members of *Callicebus moloch*, with the same stop codon generating mechanism as in humans, (not the novel GPI-anchoring mechanism), were cloned by RT-PCR on total RNA extracted from blood. It is intriguing to speculate that CEA family members with the novel GPI-linkage mechanism like human CEA, is a tissue specific expressed gene which replaces human CEA in colonic epithelium of this species of monkey. Human CEA, which has the most complex domain organization in CEA family members, possibly evolved later in higher primates by gene and exon duplications. Cloning of the complete cDNA of the Callicebus monkey novel CEA family member and comparison of its sequence and deduced domain organization to human CEA

gene family members could provide insight regarding their function *in vivo*. To address this issue unfortunately there was no colonic tissue from the above monkey available.

It has also been proposed that the overproduction of CEA leads to a disruption of normal tissue architecture and resumption of a multilayered embryonic configuration in adult colonic epithelium (**Benchimol et al., 1989**). A deletion mutant of the N domain of CEA, missing the last 75 amino acids of the N domain which is incapable of mediating intercellular adhesion, does not block myogenic differentiation (**Eidelman et al., 1993**). In addition, N and A3B3 but not A1B1 or A2B2 domain peptides were capable of releasing the differentiation block (Eidelman 93). These are precisely the same peptides that inhibited intercellular adhesion mediated by binding between the N and A3B3 domains of CEA. Therefore, the determined structural requirements for myogenic differentiation block are self-binding external domains coupled to the CEA-specified GPI anchor.

Regarding the possible mechanism of the differentiation block mediated by GPI-linked CEA family members, there is accumulating evidence that the inhibitory effects of CEA and CEACAM6 on different types of differentiation programs are due to perturbations in integrin/ECM interactions (Ordonez C. & Stanners C. P., in preparation). It has been shown that relative to the parental L6 cells, the binding to L6 ECM of transfectants

L6(CEA) and L6(CEACAM6) that were blocked in differentiation was significantly inhibited. However, the binding of L6 (ΔNCEA) a deletion mutant of the N domain defective in adhesion, L6(NCAM) a GPI-linked Ig superfamily member, and L6(CEACAM1-4L) was unaffected or increased (Ordonez C & Stanners C. P., in preparation. in preparation). Further experiments have shown that the presence of CEA or CEACAM6 decreases binding to the purified ECM components, fibronectin, and that the reduced binding of L6(CEA) to the ECM can be promptly reversed by treatment with N or A3B3 but not A2B2 domain peptides (Ordonez C. & Stanners C. P., in preparation). There is extensive evidence implicating a perturbation of the function of particular integrins by the over-expression of CEA and CEACAM6. (Ordonez C. & Stanners C. P. in preparation). CC1-tAT and CC1-CMO also blocked myogenic differentiation of L6 myoblasts much like CEA and CEACAM6, possibly by the observed inhibition of binding to the ECM component, fibronectin.

A hypothetical model has been proposed that overproduction of CEA leads to lower cell surface expression of E-cadherin, leading in turn to a progressive loss of cell polarization, followed by circumferential CEA localization and, consequently, compromised integrin function, giving reduced ECM binding and, finally, inhibited differentiation of colonocytes as they migrate up the crypt axis. The differentiation block leaves cells with division potential and susceptibility to further oncogenic activation and the

acquisition of tumorigenicity, as demonstrated by the ability of CEA to cooperate with Myc and Bcl-2 in cellular transformation (Screaton et al., 1997).

This work has established that differentiation blocking is an inherent feature of CEA family member GPI anchors which co-evolved with two different groups of GPI-anchoring mutational mechanism during evolution. In this study, we have shown that conversion of TM-linked CEACAM1 to a GPI-linked protein by fusing its extracellular domain to the TM domain of the novel CEA family member of callicebus monkey or directly by introducing a stop codon and substituting only two amino acids in CEACAM1 TM domain, confers myogenic differentiation blocking capacity on CEACAM1. As discussed before, TM-linked CEACAM1 and another GPI-anchored Ig superfamily member, N-CAM125 did not block myogenic differentiation but the substitution of their TM domains with the CEA specified-GPI gave construct that did block L6 myogenic differentiation (Screaton et al. in preparation). It would therefore be of special interest to determine the molecular signals which specify differentiation blocking capacity of the GPI anchor in the CEA family members' TM domain. This could be inferred and experimentally determined by comparing and altering the TM domain sequence of human and other species (e.g. Callicebus monkey) which specify GPI linkage.

Although the GPI anchor core structure is evolutionarily conserved (see Ch. 1) but, there might be differences in the side chains or linked lipid structures of various GPI anchors (McConville and Ferguson, 1993). It has been proposed that the basis for the difference between the myogenic differentiation-promoting effect of NCAM and the inhibitory effect of CEA is because of their non-overlapping localization in different membrane microdomains due to their different GPI-anchors (Screaton et al. in preparation). Structural analysis of the human CEA or Callicebus monkey novel CEA family member GPI anchor in comparison with the NCAM GPI anchor could directly test this issue and provide important information concerning the relationship between GPI anchor structure and function.

It has been shown that the GPI-linked form of proteins could be secreted or shed from the cell surface spontaneously (Censullo and Davitz, 1994), One possibility is that the release of the soluble form of GPI-linked CEA-like proteins into the medium and their interaction with ECM or other cell surface components could result in blocking myogenic differentiation. This would reduce the importance of GPI anchors in this process, except for facilitating the protein to be more readily released into the medium. Although this possibility in our view is low, it could be tested by transfection of L6 cells with an anchor minus, secretory form of CEACAM1 and CEA protein or alternatively by exogeneously adding of PI-PLC cleaved GPI-anchored CEA-like proteins from another cultured cell population surface to

cultured parental L6 cells without transfecting them. This alternative simple experiment as mentioned below could also be tested for one of the models which is discussed regarding the way which GPI-anchored proteins initiate cell signalling.

As discussed in the first Chapter, different models have been proposed to explain signal induction by GPI-anchored proteins. One of these models suggests that the endogenous lipolytic enzymes like PI-PLC cleave the GPI-anchor from protein, yielding a bioactve second messenger from the GPI anchor which is internalized and triggers intracellular signal transduction (Malek et al., 1994; Varela-Nieto et al., 1996). This could be also addressed by exogeneously adding of PI-PLC cleaved GPI-linked CEA-like proteins from the surface of another cultured cell populations to the cultured parental L6 or other cells lines.

General findings and major original contributions to knowledge in the field

1. Only a few mutational changes are required to convert a TM-linked CEA family member into a GPI-linked one.

2. The evolution of GPI-linked CEA family members happened twice independently. The mode of generating a stop codon found in humans evolved in a common ancestor of the primitive primate groups, Tarsiers and anthropoids. The second novel mechanism of stop codon generation evolved independently in Cebidae family of new world monkeys during the primate radiation. Also, mutations, predicted (on the basis of the results of chapter 2) to give highly efficient GPI processing, independently evolved.

The independent evolution that gave rise to the same structural feature implies useful function(s) giving adaptive advantage and positive selection of the acquisition of GPI linkage in the CEA family. These results support the notion that GPI-linked CEA family members have important functions *in vivo*.

3. The two different mechanisms discovered for generating GPI-linked sequences in the CEA family during evolution both lead to proteins that inhibit cellular differentiation and modifies cell-ECM interaction.

Bibliography

Refrences

Afar, D.E., C.P. Stanners, and J.C. Bell. 1992. Tyrosine phosphorylation of biliary glycoprotein, a cell adhesion molecule related to carcinoembryonic antigen. *Biochimica et Biophysica Acta.* 1134:46-52.

Anderson, R.G. 1993. Caveolae: where incoming and outgoing messengers meet. *Proceedings of the National Academy of Sciences of the United States of America.* 90:10909-10913.

Anderson, R.G. 1994. Functional specialization of the glycosylphosphatidylinositol membrane anchor. *Seminars in Immunology.* 6:89-95.

Beauchemin, N., S. Benchimol, D. Cournoyer, A. Fuks, and C.P. Stanners. 1987. Isolation and characterization of full-length functional cDNA clones for human carcinoembryonic antigen. *Molecular & Cellular Biology.* 7:3221-3230.

Beauchemin, N., T. Kunath, J. Robitaille, B. Chow, C. Turbide, E. Daniels, and A. Veillette. 1997. Association of biliary glycoprotein with protein tyrosine phosphatase SHP-1 in malignant colon epithelial cells. *Oncogene.* 14:783-790.

Bell, L.M., K.R. Solomon, J.P. Gold, and K.N. Tan. 1994. Cytoplasmic tail deletion of T cell receptor (TCR) beta-chain results in its surface expression as

glycosylphosphatidylinositol-anchored polypeptide on mature T cells in the absence of TCR-alpha. *Journal of Biological Chemistry*. 269:22758-22763.

Benchimol, S., A. Fuks, S. Jothy, N. Beauchemin, K. Shirota, and C.P. Stanners. 1989. Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell*. 57:327-334.

Benghezal, M., A. Benachour, S. Rusconi, M. Aebi, and A. Conzelmann. 1996. Yeast Gpi8p is essential for GPI anchor attachment onto proteins. *EMBO Journal*. 15:6575-6583.

Birchmeier, W., and J. Behrens. 1994. Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. *Biochimica et Biophysica Acta*. 1198:11-26.

Bos, M.P., F. Grunert, and R.J. Belland. 1997. Differential recognition of members of the carcinoembryonic antigen family by Opa variants of Neisseria gonorrhoeae. *Infection & Immunity*. 65:2353-2361.

Boyum, A. 1974. Separation of blood leucocytes, granulocytes and lymphocytes. *Tissue Antigens*. 4:269-274.

Brandriff, B.F., L.A. Gordon, K.T. Tynan, A.S. Olsen, H.W. Mohrenweiser, A. Fertitta, A.V. Carrano, and B.J. Trask. 1992. Order and genomic distances

among members of the carcinoembryonic antigen (CEA) gene family determined by fluorescence in situ hybridization. *Genomics*. 12:773-779.

Brown, D.A., and E. London. 1997. Structure of detergent-resistant membrane domains: does phase separation occur in biological membranes? *Biochemical & Biophysical Research Communications*. 240:1-7.

Brown, D.A., and E. London. 1998. Functions of lipid rafts in biological membranes (review). *Annual Review of Cell & Developmental Biology*. 14:111-136.

Brown, D.A., and J.K. Rose. 1992. Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell*. 68:533-544.

Brummer, J., M. Neumaier, C. Gopfert, and C. Wagener. 1995. Association of pp60c-src with biliary glycoprotein (CD66a), an adhesion molecule of the carcinoembryonic antigen family downregulated in colorectal carcinomas. *Oncogene*. 11:1649-1655.

Bucht, G., and K. Hjalmarsson. 1996. Residues in Torpedo californica acetylcholinesterase necessary for processing to a glycosyl phosphatidylinositol-anchored form. *Biochimica et Biophysica Acta*. 1292:223-232.

Buxton, R.S., and A.I. Magee. 1992. Structure and interactions of desmosomal and other cadherins. *Seminars in Cell Biology*. 3:157-167.

Casey, P.J. 1995. Protein lipidation in cell signaling. *Science*. 268:221-225.

Censullo, P., and M.A. Davitz. 1994. How GPI-anchored proteins turnover: or where do they go after arrival at the plasma membrane. *Seminars in Immunology*. 6:81-88.

Chen, D.S., M. Asanaka, F.S. Chen, J.E. Shively, and M.M. Lai. 1997. Human carcinoembryonic antigen and biliary glycoprotein can serve as mouse hepatitis virus receptors. *Journal of Virology*. 71:1688-1691.

Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*. 162:156-159.

Chu, T.M., G. Reynoso, and H.J. Hansen. 1972. Demonstration of carcinoembryonic antigen in normal human plasma. *Nature*. 238:152-153.

Clackson, T., D. Güssow, and P.T. Jones. 1992. General Applications of PCR to Gene Cloning and Manipulation. *In* PCR a Practical Approach. M.J. Mcpherson, P. Quirke, and G.R. Taylor, editors. IRL Press at Oxford University Press, New York. 187-214.

Cournoyer, D., N. Beauchemin, D. Boucher, S. Benchimol, A. Fuks, and C.P. Stanners. 1988. Transcription of genes of the carcinoembryonic antigen family in malignant and nonmalignant human tissues. *Cancer Research*. 48:3153-3157.

Cunningham, B.A., J.J. Hemperly, B.A. Murray, E.A. Prediger, R. Brackenbury, and G.M. Edelman. 1987. Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science*. 236:799-806.

Dickson, G., D. Peck, S.E. Moore, C.H. Barton, and F.S. Walsh. 1990. Enhanced myogenesis in NCAM-transfected mouse myoblasts. *Nature*. 344:348-351.

Doolittle, R.F. 1989. Similar amino acid sequences revisited. *Trends in Biochemical Sciences*. 14:244-245.

Draber, P., and K.M. Skubitz. 1998. Signal transduction mediated by the CEA family. *In* Cell Adhesion and Communication Mediated by the CEA Family. Vol. 5. C.P. Stanners, editor. Hardwood Academic Publishers, Amsterdam. 121-140.

Ducker, T.P., and K.M. Skubitz. 1992. Subcellular localization of CD66, CD67, and NCA in human neutrophils. *Journal of Leukocyte Biology*. 52:11-16.

Dunphy, W.G., R. Brands, and J.E. Rothman. 1985. Attachment of terminal N-acetylglucosamine to asparagine-linked oligosaccharides occurs in central cisternae of the Golgi stack. *Cell*. 40:463-472.

Dveksler, G.S., A.A. Basile, C.B. Cardellichio, and K.V. Holmes. 1995. Mouse hepatitis virus receptor activities of an MHVR/mph chimera and MHVR mutants lacking N-linked glycosylation of the N-terminal domain. *Journal of Virology*. 69:543-546.

Dveksler, G.S., C.W. Dieffenbach, C.B. Cardellichio, K. McCuaig, M.N. Pensiero, G.S. Jiang, N. Beauchemin, and K.V. Holmes. 1993a. Several members of the mouse carcinoembryonic antigen-related glycoprotein family are functional receptors for the coronavirus mouse hepatitis virus-A59. *Journal of Virology*. 67:1-8.

Dveksler, G.S., M.N. Pensiero, C.W. Dieffenbach, C.B. Cardellichio, A.A. Basile, P.E. Elia, and K.V. Holmes. 1993b. Mouse hepatitis virus strain A59 and blocking antireceptor monoclonal antibody bind to the N-terminal domain of cellular receptor. *Proceedings of the National Academy of Sciences of the United States of America*. 90:1716-1720.

Edidin, M., S.C. Kuo, and M.P. Sheetz. 1991. Lateral movements of membrane glycoproteins restricted by dynamic cytoplasmic barriers. *Science*. 254:1379-1382.

Edlund, M., I. Blikstad, and B. Obrink. 1996. Calmodulin binds to specific sequences in the cytoplasmic domain of C-CAM and down-regulates C-CAM self-association. *Journal of Biological Chemistry*. 271:1393-1399.

Edlund, M., H. Gaardsvoll, E. Bock, and B. Obrink. 1993. Different isoforms and stock-specific variants of the cell adhesion molecule C-CAM (cell-CAM 105) in rat liver. *European Journal of Biochemistry*. 213:1109-1116.

Edmiston, K.H., A. Gangopadhyay, Y. Shoji, A.P. Nachman, P. Thomas, and J.M. Jessup. 1997. In vivo induction of murine cytokine production by carcinoembryonic antigen. *Cancer Research*. 57:4432-4436.

Eidelman, F.J., A. Fuks, L. DeMarte, M. Taheri, and C.P. Stanners. 1993. Human carcinoembryonic antigen, an intercellular adhesion molecule, blocks fusion and differentiation of rat myoblasts. *Journal of Cell Biology*. 123:467-475.

Englund, P.T. 1993. The structure and biosynthesis of glycosyl phosphatidylinositol protein anchors. *Annual Review of Biochemistry*. 62:121-138.

Engvall, E., M. Vuento, and E. Ruoslahti. 1976. A monkey antigen crossreacting with carcinoembryonic antigen, CEA. *British Journal of Cancer*. 34:341-345.

Fantini, G.A., and J.J. DeCosse. 1990. Surveillance strategies after resection of carcinoma of the colon and rectum. *Surgery, Gynecology & Obstetrics*. 171:267-273.

Ferguson, M.A. 1992. Site of palmitoylation of a phospholipase C-resistant glycosylphosphatidylinositol membrane anchor. *Biochemical Journal*. 284:297-300.

Ferguson, M.A., and A.F. Williams. 1988. Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. *Annual Review of Biochemistry*. 57:285-320.

Freimer, N.B., and M. Slatkin. 1996. Microsatellites: evolution and mutational processes. *Ciba Foundation Symposium*. 197:51-67; discussion 67-72.

Friedrichson, T., and T.V. Kurzchalia. 1998. Microdomains Of Gpi-Anchored Proteins In Living Cells Revealed By Crosslinking. *Nature*. 394:802-805.

Fritsche, R., and J.P. Mach. 1977. Isolation and characterization of carcinoembryonic antigen (CEA) extracted from normal human colon mucosa. *Immunochemistry*. 14:119-127.

Furukawa, Y., K. Tsukamoto, and H. Ikezawa. 1997. Mutational analysis of the C-terminal signal peptide of bovine liver 5'-nucleotidase for GPI anchoring: a study on the significance of the hydrophilic spacer region. *Biochimica et Biophysica Acta*. 1328:185-196.

Gangopadhyay, A., O. Bajenova, T.M. Kelly, and P. Thomas. 1996. Carcinoembryonic antigen induces cytokine expression in Kuppfer cells: implications for hepatic metastasis from colorectal cancer. *Cancer Research*. 56:4805-4810.

Gangopadhyay, A., D.A. Lazure, and P. Thomas. 1998. Adhesion of colorectal carcinoma cells to the endothelium is mediated by cytokines from CEA stimulated Kupffer cells. *Clinical & Experimental Metastasis*. 16:703-712.

Gangopadhyay, A., and P. Thomas. 1996. Processing of carcinoembryonic antigen by Kupffer cells: recognition of a penta-peptide sequence. *Archives of Biochemistry & Biophysics*. 334:151-157.

Gaulton, G.N., and J.C. Pratt. 1994. Glycosylated phosphatidylinositol molecules as second messengers. *Seminars in Immunology*. 6:97-104.

Gerber, L.D., K. Kodukula, and S. Udenfriend. 1992. Phosphatidylinositol glycan (PI-G) anchored membrane proteins. Amino acid requirements adjacent to the site of cleavage and PI-G attachment in the COOH-terminal signal peptide. *Journal of Biological Chemistry*. 267:12168-12173.

Gibbons, A. 1995. When it comes to evolution, humans are in the slow class [news]. *Science*. 267:1907-1908.

Gillespie, J.H. 1991. Protein evolution. *In* The causes of molecular evolution. Vol. 1. Oxford university press, New York. 3-63.

Gold, S.O.F.a.P. 1965. Specific carcinoembryonic antigensof the human digestive system. *Journal Experimental Medicine.* 122:467-481.

Gumbiner, B.M. 1996. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell.* 84:345-357.

Gunter, K.C., T.R. Malek, and E.M. Shevach. 1984. T cell-activating properties of an anti-Thy-1 monoclonal antibody. Possible analogy to OKT3/Leu-4. *Journal of Experimental Medicine.* 159:716-730.

Haagensen, D.E., Jr., R.S. Metzgar, B. Swenson, W.G. Dilley, C.E. Cox, S. Davis, J. Murdoch, N. Zamcheck, and S.A. Wells, Jr. 1982. Carcinoembryonic antigen in nonhuman primates. *Journal of the National Cancer Institute.* 69:1073-1076.

Hamburger, D., M. Egerton, and H. Riezman. 1995. Yeast Gaa1p is required for attachment of a completed GPI anchor onto proteins. *Journal of Cell Biology.* 129:629-639.

Hammarström, S., A. Olsen, S. Teglund, and V. Baranov. 1998. The Nature and Expression of the Human CEA Family: Basic and Clinical Perspectives. *In*

Cell Adhesion and Communication Mediated by the CEA Family. Vol. 5. C.P. Stanners, editor. Hardwood Academic Publishers, Amsterdam. 1-30.

Henikoff, S., E.A. Greene, S. Pietrokovski, P. Bork, T.K. Attwood, and L. Hood. 1997. Gene families: the taxonomy of protein paralogs and chimeras. *Science*. 278:609-614.

Hiroi, Y., I. Komuro, R. Chen, T. Hosoda, T. Mizuno, S. Kudoh, S.P. Georgescu, M.E. Medof, and Y. Yazaki. 1998. Molecular cloning of human homolog of yeast GAA1 which is required for attachment of glycosylphosphatidylinositols to proteins. *FEBS Letters*. 421:252-258.

Hixson, D.C., K.D. McEntire, and B. Obrink. 1985. Alterations in the expression of a hepatocyte cell adhesion molecule by transplantable rat hepatocellular carcinomas. *Cancer Research*. 45:3742-3749.

Hsieh, J.T., W. Luo, W. Song, Y. Wang, D.I. Kleinerman, N.T. Van, and S.H. Lin. 1995. Tumor suppressive role of an androgen-regulated epithelial cell adhesion molecule (C-CAM) in prostate carcinoma cell revealed by sense and antisense approaches. *Cancer Research*. 55:190-197.

Hunter, I., H. Sawa, M. Edlund, and B. Obrink. 1996. Evidence for regulated dimerization of cell-cell adhesion molecule (C-CAM) in epithelial cells. *Biochemical Journal*. 320:847-853.

Ikezawa, H., M. Yamanegi, R. Taguchi, T. Miyashita, and T. Ohyabu. 1976. Studies on phosphatidylinositol phosphodiesterase (phospholipase C type) of Bacillus cereus. I. purification, properties and phosphatase-releasing activity. *Biochimica et Biophysica Acta*. 450:154-164.

Ilantzis, C., S. Jothy, L.C. Alpert, P. Draber, and C.P. Stanners. 1997. Cell-surface levels of human carcinoembryonic antigen are inversely correlated with colonocyte differentiation in colon carcinogenesis. *Laboratory Investigation*. 76:703-716.

Ilantzis, C., and C.P. Stanners. 1997. A colonic tissue architecture assay applied to human colon carcinoma cells. *In Vitro Cellular & Developmental Biology. Animal*. 33:50-61.

Jacobson, K., A. Ishihara, and R. Inman. 1987. Lateral diffusion of proteins in membranes. *Annual Review of Physiology*. 49:163-175.

Jacobson, K.A., S.E. Moore, B. Yang, P. Doherty, G.W. Gordon, and F.S. Walsh. 1997. Cellular determinants of the lateral mobility of neural cell adhesion molecules. *Biochimica et Biophysica Acta*. 1330:138-144.

James, P., T. Vorherr, and E. Carafoli. 1995. Calmodulin-binding domains: just two faced or multi-faceted? *Trends in Biochemical Sciences*. 20:38-42.

Jothy, S., S.Y. Yuan, and K. Shirota. 1993. Transcription of carcinoembryonic antigen in normal colon and colon carcinoma. In situ hybridization study and implication for a new in vivo functional model. *American Journal of Pathology*. 143:250-257.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature*. 217:624-626.

Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 78:454-458.

Kimura, M. 1991. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proceedings of the National Academy of Sciences of the United States of America*. 88:5969-5973.

Kimura, M., and T. Ota. 1974. On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 71:2848-2852.

Kinoshita, T., K. Ohishi, and J. Takeda. 1997. GPI-anchor synthesis in mammalian cells: genes, their products, and a deficiency. *Journal of Biochemistry*. 122:251-257.

Kinugasa, T., M. Kuroki, H. Takeo, Y. Matsuo, K. Ohshima, Y. Yamashita, T. Shirakusa, and Y. Matsuoka. 1998. Expression of four CEA family antigens (CEA, NCA, BGP and CGM2) in normal and cancerous gastric epithelial cells: up-regulation of BGP and CGM2 in carcinomas. *International Journal of Cancer*. 76:148-153.

Kleinerman, D.I., C.P. Dinney, W.W. Zhang, S.H. Lin, N.T. Van, and J.T. Hsieh. 1996. Suppression of human bladder cancer growth by increased expression of C-CAM1 gene in an orthotopic model. *Cancer Research*. 56:3431-3435.

Kleinerman, D.I., P. Troncoso, S.H. Lin, L.L. Pisters, E.R. Sherwood, T. Brooks, A.C. von Eschenbach, and J.T. Hsieh. 1995. Consistent expression of an epithelial cell adhesion molecule (C-CAM) during human prostate development and loss of expression in prostate cancer: implication as a tumor suppressor. *Cancer Research*. 55:1215-1220.

Kodukula, K., L.D. Gerber, R. Amthauer, L. Brink, and S. Udenfriend. 1993. Biosynthesis of glycosylphosphatidylinositol (GPI)-anchored membrane proteins in intact cells: specific amino acid requirements adjacent to the site of cleavage and GPI attachment. *Journal of Cell Biology*. 120:657-664.

Kreitman, M., and H. Akashi. 1995. Molecular evidence for natural selection (review). *Annual Review of Ecology & Systematics*. 26:403-422.

Kunath, T., C. Ordonez-Garcia, C. Turbide, and N. Beauchemin. 1995. Inhibition of colonic tumor cell growth by biliary glycoprotein. *Oncogene*. 11:2375-2382.

Kuroki, M., Y. Matsuo, T. Kinugasa, and Y. Matsuoka. 1992. Augmented expression and release of nonspecific cross-reacting antigens (NCAs), members of the CEA family, by human neutrophils during cell activation. *Journal of Leukocyte Biology*. 52:551-557.

Kuroki, M., T. Yamanaka, Y. Matsuo, S. Oikawa, H. Nakazato, and Y. Matsuoka. 1995. Immunochemical analysis of carcinoembryonic antigen (CEA)-related antigens differentially localized in intracellular granules of human neutrophils. *Immunological Investigations*. 24:829-843.

Kurosaki, T., and J.V. Ravetch. 1989. A single amino acid in the glycosyl phosphatidylinositol attachment domain determines the membrane topology of Fc gamma RIII [published erratum appears in Nature 1990 Jan 25;343(6256):390]. *Nature*. 342:805-807.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227:680-685.

Laurence, D.J., U. Stevens, R. Bettelheim, D. Darcy, C. Leese, C. Turberville, P. Alexander, E.W. Johns, and A.M. Neville. 1972. Role of plasma carcinoembryonic antigen in diagnosis of gastrointestinal, mammary, and bronchial carcinoma. *British Medical Journal*. 3:605-609.

Leusch, H.G., Z. Drzeniek, S.A. Hefta, Z. Markos-Pusztai, and C. Wagener. 1991. The putative role of members of the CEA-gene family (CEA, NCA an BGP) as ligands for the bacterial colonization of different human epithelial tissues. *Zentralblatt fur Bakteriologie*. 275:118-122.

Lewin, B. 1994a. Gene numbers: replication and redundancy. *In* genes V. Vol. 1. Oxford university press, New York. 701-731.

Lewin, B. 1994b. Recombination of DNA. *In* genes V. Vol. 1. Oxford university press, New York. 967-997.

Li, W.H., D.L. Ellsworth, J. Krushkal, B.H. Chang, and D. Hewett-Emmett. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Molecular Phylogenetics & Evolution*. 5:182-187.

Li, w.H., and D. Graur. 1991a. Evolution by gene duplication and exon shuffling. *In* Fundamentals of molecular evolution. Vol. 1. Sinauer associates, sunderland. 136-171.

Li, w.H., and D. Graur. 1991b. Evolutionary change in nucleotide sequences. *In* Fundamentals of molecular evolution. Vol. 1. Sinauer associates, sunderland. 42-66.

Li, w.H., and D. Graur. 1991c. Gene structure and mutation. *In* Fundamentals of molecular evolution. Vol. 1. Sinauer associates, sunderland. 3-19.

Li, w.H., and D. Graur. 1991d. Rates and patterns of nucleotide substitution. *In* Fundamentals of molecular evolution. Vol. 1. Sinauer associates, sunderland. 67-98.

Li, W.H., C.I. Wu, and C.C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology & Evolution*. 2:150-174.

Lin, S.H., and G. Guidotti. 1989. Cloning and expression of a cDNA coding for a rat liver plasma membrane ecto-ATPase. The primary structure of the ecto-ATPase is similar to that of the human biliary glycoprotein I. *Journal of Biological Chemistry*. 264:14408-14414.

Low, M.G. 1989. The glycosyl-phosphatidylinositol anchor of membrane proteins. *Biochimica et Biophysica Acta*. 988:427-454.

Low, M.G., and J.B. Finean. 1977. Release of alkaline phosphatase from membranes by a phosphatidylinositol-specific phospholipase C. *Biochemical Journal*. 167:281-284.

Lucka, L., I. Cichocka, K. Baumler, K. Bechler, and W. Reutter. 1995. A short isoform of carcinoembryonic-antigen-related rat liver cell-cell adhesion molecule (C-CAM/gp110) mediates intercellular adhesion. Sequencing and recombinant functional analysis. *European Journal of Biochemistry*. 234:527-535.

Lund-Johansen, F., J. Olweus, F.W. Symington, A. Arli, J.S. Thompson, R. Vilella, K. Skubitz, and V. Horejsi. 1993. Activation of human monocytes and granulocytes by monoclonal antibodies to glycosylphosphatidylinositol-anchored antigens. *European Journal of Immunology*. 23:2782-2791.

Lundberg, K.S., D.D. Shoemaker, M.W. Adams, J.M. Short, J.A. Sorge, and E.J. Mathur. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene*. 108:1-6.

Luo, W., C.G. Wood, K. Earley, M.C. Hung, and S.H. Lin. 1997. Suppression of tumorigenicity of breast cancer cells by an epithelial cell adhesion molecule

(C-CAM1): the adhesion and growth suppression are mediated by different domains. *Oncogene*. 14:1697-1704.

Malek, T.R., T.J. Fleming, and E.K. Codias. 1994. Regulation of T lymphocyte function by glycosylphosphatidylinositol (GPI)-anchored proteins. *Seminars in Immunology*. 6:105-113.

Malek, T.R., G. Ortega, C. Chan, R.A. Kroczek, and E.M. Shevach. 1986. Role of Ly-6 in lymphocyte activation. II. Induction of T cell activation by monoclonal anti-Ly-6 antibodies. *Journal of Experimental Medicine*. 164:709-722.

Margolis, R.N., M.J. Schell, S.I. Taylor, and A.L. Hubbard. 1990. Hepatocyte plasma membrane ECTO-ATPase (pp120/HA4) is a substrate for tyrosine kinase activity of the insulin receptor. *Biochemical & Biophysical Research Communications*. 166:562-566.

Maxfield, F.R., and S. Mayor. 1997. Cell surface dynamics of GPI-anchored proteins. *Advances in Experimental Medicine & Biology*. 419:355-364.

Mayor, S., K.G. Rothberg, and F.R. Maxfield. 1994. Sequestration of GPI-anchored proteins in caveolae triggered by cross-linking. *Science*. 264:1948-1951.

Mayr, E. 1978. Evolution. *Scientific American*. 239:46-55.

McConville, M.J., and M.A. Ferguson. 1993. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. *Biochemical Journal*. 294:305-324.

McCuaig, K., M. Rosenberg, P. Nedellec, C. Turbide, and N. Beauchemin. 1993. Expression of the Bgp gene and characterization of mouse colon biliary glycoprotein isoforms. *Gene*. 127:173-183.

McCuaig, K., C. Turbide, and N. Beauchemin. 1992. mmCGM1a: a mouse carcinoembryonic antigen gene family member, generated by alternative splicing, functions as an adhesion molecule. *Cell Growth & Differentiation*. 3:165-174.

McDonald, J.H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in Drosophila [see comments]. *Nature*. 351:652-654.

Medof, M.E., S. Nagarajan, and M.L. Tykocinski. 1996. Cell-surface engineering with GPI-anchored proteins. *FASEB Journal*. 10:574-586.

Messier, W., and C.B. Stewart. 1997. Episodic adaptive evolution of primate lysozymes. *Nature*. 385:151-154.

Micanovic, R., K. Kodukula, L.D. Gerber, and S. Udenfriend. 1990. Selectivity at the cleavage/attachment site of phosphatidylinositol-glycan anchored membrane proteins is enzymatically determined. *Proceedings of the National Academy of Sciences of the United States of America.* 87:7939-7943.

Moran, P., and I.W. Caras. 1991. Fusion of sequence elements from non-anchored proteins to generate a fully functional signal for glycophosphatidylinositol membrane anchor attachment. *Journal of Cell Biology.* 115:1595-1600.

Moran, P., and I.W. Caras. 1992. Proteins containing an uncleaved signal for glycophosphatidylinositol membrane anchor attachment are retained in a post-ER compartment. *Journal of Cell Biology.* 119:763-772.

Moran, P., and I.W. Caras. 1994. Requirements for glycosylphosphatidylinositol attachment are similar but not identical in mammalian cells and parasitic protozoa. *Journal of Cell Biology.* 125:333-343.

Najjar, S.M., D. Accili, N. Philippe, J. Jernberg, R. Margolis, and S.I. Taylor. 1993. pp120/ecto-ATPase, an endogenous substrate of the insulin receptor tyrosine kinase, is expressed as two variably spliced isoforms. *Journal of Biological Chemistry.* 268:1201-1206.

Nedellec, P., G.S. Dveksler, E. Daniels, C. Turbide, B. Chow, A.A. Basile, K.V. Holmes, and N. Beauchemin. 1994. Bgp2, a new member of the carcinoembryonic antigen-related gene family, encodes an alternative receptor for mouse hepatitis viruses. *Journal of Virology*. 68:4525-4537.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology & Evolution*. 3:418-426.

Neumaier, M., S. Paululat, A. Chan, P. Matthaes, and C. Wagener. 1993. Biliary glycoprotein, a potential human cell adhesion molecule, is down-regulated in colorectal carcinomas. *Proceedings of the National Academy of Sciences of the United States of America*. 90:10744-10748.

Nollau, P., H. Scheller, M. Kona-Horstmann, S. Rohde, F. Hagenmuller, C. Wagener, and M. Neumaier. 1997. Expression of CD66a (human C-CAM) and other members of the carcinoembryonic antigen gene family of adhesion molecules in human colorectal adenomas. *Cancer Research*. 57:2354-2357.

Nonaka, M., T. Miwa, N. Okada, and H. Okada. 1995. Multiple isoforms of guinea pig decay-accelerating factor (DAF) generated by alternative splicing. *Journal of Immunology*. 155:3037-3048.

Obrink, B. 1997. CEA adhesion molecules: multifunctional proteins with signal-regulatory properties. *Current Opinion in Cell Biology*. 9:616-626.

Oda, K., I. Wada, N. Takami, T. Fujiwara, Y. Misumi, and Y. Ikehara. 1996. Bip/GRP78 but not calnexin associates with a precursor of glycosylphosphatidylinositol-anchored protein. *Biochemical Journal*. 316:623-630.

Ohta, T. 1991. Multigene families and the evolution of complexity. *Journal of Molecular Evolution*. 33:34-41.

Ohta, T. 1997. Role of random genetic drift in the evolution of interactive systems. *Journal of Molecular Evolution*. 44:S9-14.

Ohtsuka, N., Y.K. Yamada, and F. Taguchi. 1996. Difference in virus-binding activity of two distinct receptor proteins for mouse hepatitis virus. *Journal of General Virology*. 77:1683-1692.

Ohwada, A., H. Takahashi, I. Nagaoka, and S. Kira. 1994. Biliary glycoprotein mRNA expression is increased in primary lung cancer, especially in squamous cell carcinoma. *American Journal of Respiratory Cell & Molecular Biology*. 11:214-220.

Oikawa, S., C. Inuzuka, M. Kuroki, F. Arakawa, Y. Matsuoka, G. Kosaki, and H. Nakazato. 1991. A specific heterotypic cell adhesion activity between

members of carcinoembryonic antigen family, W272 and NCA, is mediated by N-domains. *Journal of Biological Chemistry*. 266:7995-8001.

Oikawa, S., C. Inuzuka, M. Kuroki, Y. Matsuoka, G. Kosaki, and H. Nakazato. 1989. Cell adhesion activity of non-specific cross-reacting antigen (NCA) and carcinoembryonic antigen (CEA) expressed on CHO cell surface: homophilic and heterophilic adhesion. *Biochemical & Biophysical Research Communications*. 164:39-45.

Oikawa, S., H. Nakazato, and G. Kosaki. 1987. Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence. *Biochemical & Biophysical Research Communications*. 142:511-518.

Olsson, H., K. Wikstrom, G. Kjellstrom, and B. Obrink. 1995. Cell adhesion activity of the short cytoplasmic domain isoform of C-CAM (C-CAM2) in CHO cells. *FEBS Letters*. 365:51-56.

Paxton, R.J., G. Mooser, H. Pande, T.D. Lee, and J.E. Shively. 1987. Sequence analysis of carcinoembryonic antigen: identification of glycosylation sites and homology with the immunoglobulin supergene family. *Proceedings of the National Academy of Sciences of the United States of America*. 84:920-924.

Pollard, J.W., and C.P. Stanners. 1979. Characterization of cell lines showing growth control isolated from both the wild type and a leucyl-tRNA synthetase mutant of Chinese hamster ovary cells. *Journal of Cellular Physiology*. 98:571-585.

Porter, C.A., S.L. Page, J. Czelusniak, H. Schneider, M.P.C. Schneither, I. Sampio, and M. Goodman. 1997. Phylogeny and evolution of selected primates as determined by sequences of the e-globin locus and 5′ flanking regions. *International journal of primatology*. 18:261-295.

Puoti, A., C. Desponds, C. Fankhauser, and A. Conzelmann. 1991. Characterization of glycophospholipid intermediate in the biosynthesis of glycophosphatidylinositol anchors accumulating in the Thy-1-negative lymphoma line SIA-b. *Journal of Biological Chemistry*. 266:21051-21059.

Ramel, C. 1997. Mini- and microsatellites. *Environmental Health Perspectives*. 105:781-789.

Rettenberger, G., W. Zimmermann, C. Klett, U. Zechner, and H. Hameister. 1995. Mapping of murine YACs containing the genes Cea2 and Cea4 after B1-PCR amplification and FISH-analysis. *Chromosome Research*. 3:473-478.

Robinson, P.J. 1997. Signal transduction via GPI-anchored membrane proteins. *Advances in Experimental Medicine & Biology*. 419:365-370.

Rojas, M., L. DeMarte, R.A. Screaton, and C.P. Stanners. 1996. Radical differences in functions of closely related members of the human carcinoembryonic antigen gene family. *Cell Growth & Differentiation*. 7:655-662.

Rojas, M., A. Fuks, and C.P. Stanners. 1990. Biliary glycoprotein, a member of the immunoglobulin supergene family, functions in vitro as a Ca2(+)-dependent intercellular adhesion molecule. *Cell Growth & Differentiation*. 1:527-533.

Rosenberg, M., P. Nedellec, S. Jothy, D. Fleiszer, C. Turbide, and N. Beauchemin. 1993. The expression of mouse biliary glycoprotein, a carcinoembryonic antigen-related gene, is down-regulated in malignant mouse tissues. *Cancer Research*. 53:4938-4945.

Rudert, F., W. Zimmermann, and J.A. Thompson. 1989. Intra- and interspecies analyses of the carcinoembryonic antigen (CEA) gene family reveal independent evolution in primates and rodents. *Journal of Molecular Evolution*. 29:126-134.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Analysis and Cloning of Eukaryotic Genomic DNA. Cold Spring Harbor Laboratory Press, New York. 9.2-9.62 pp.

Sauter, S.L., S.M. Rutherfurd, C. Wagener, J.E. Shively, and S.A. Hefta. 1993. Identification of the specific oligosaccharide sites recognized by type 1 fimbriae from Escherichia coli on nonspecific cross-reacting antigen, a CD66 cluster granulocyte glycoprotein. *Journal of Biological Chemistry*. 268:15510-15516.

Schmutte, C., and P.A. Jones. 1998. Involvement of DNA methylation in human carcinogenesis. *Biological Chemistry*. 379:377-388.

Schroeder, R., E. London, and D. Brown. 1994. Interactions between saturated acyl chains confer detergent resistance on lipids and glycosylphosphatidylinositol (GPI)-anchored proteins: GPI-anchored proteins in liposomes and cells show similar behavior. *Proceedings of the National Academy of Sciences of the United States of America*. 91:12130-12134.

Screaton, R.A., L.Z. Penn, and C.P. Stanners. 1997. Carcinoembryonic antigen, a human tumor marker, cooperates with Myc and Bcl-2 in cellular transformation. *Journal of Cell Biology*. 137:939-952.

Sharp, P.M. 1997. In search of molecular darwinism [news; comment]. *Nature*. 385:111-112.

Shaul, P.W., and R.G. Anderson. 1998. Role of plasmalemmal caveolae in signal transduction. *American Journal of Physiology*. 275:L843-851.

Sheets, E.D., G.M. Lee, R. Simson, and K. Jacobson. 1997. Transient confinement of a glycosylphosphatidylinositol-anchored protein in the plasma membrane. *Biochemistry*. 36:12449-12458.

Shenoy-Scaria, A.M., J. Kwong, T. Fujita, M.W. Olszowy, A.S. Shaw, and D.M. Lublin. 1992. Signal transduction through decay-accelerating factor. Interaction of glycosyl-phosphatidylinositol anchor and protein tyrosine kinases p56lck and p59fyn 1. *Journal of Immunology*. 149:3535-3541.

Shively, J.E., and J.D. Beatty. 1985. CEA-related antigens: molecular biology and clinical significance. *Critical Reviews in Oncology-Hematology*. 2:355-399.

Simons, K., and E. Ikonen. 1997. Functional rafts in cell membranes. *Nature*. 387:569-572.

Simons, K., and A. Wandinger-Ness. 1990. Polarized sorting in epithelia. *Cell*. 62:207-210.

Sippel, C.J., R.J. Fallon, and D.H. Perlmutter. 1994. Bile acid efflux mediated by the rat liver canalicular bile acid transport/ecto-ATPase protein requires serine 503 phosphorylation and is regulated by tyrosine 488 phosphorylation. *Journal of Biological Chemistry*. 269:19539-19545.

Sippel, C.J., T. Shen, and D.H. Perlmutter. 1996. Site-directed mutagenesis within an ectoplasmic ATPase consensus sequence abrogates the cell aggregating properties of the rat liver canalicular bile acid transporter/ecto-ATPase/cell CAM 105 and carcinoembryonic antigen. *Journal of Biological Chemistry*. 271:33095-33104.

Skubitz, K.M., K.D. Campbell, K. Ahmed, and A.P. Skubitz. 1995. CD66 family members are associated with tyrosine kinase activity in human neutrophils. *Journal of Immunology*. 155:5382-5390.

Skubitz, K.M., K.D. Campbell, and A.P. Skubitz. 1996. CD66a, CD66b, CD66c, and CD66d each independently stimulate neutrophils. *Journal of Leukocyte Biology*. 60:106-117.

Skubitz, K.M., T.P. Ducker, and S.A. Goueli. 1992. CD66 monoclonal antibodies recognize a phosphotyrosine-containing protein bearing a carcinoembryonic antigen cross-reacting antigen on the surface of human neutrophils. *Journal of Immunology*. 148:852-860.

Skubitz, K.M., T.P. Ducker, A.P. Skubitz, and S.A. Goueli. 1993. Antiserum to carcinoembryonic antigen recognizes a phosphotyrosine-containing protein in human colon cancer cell lines. *FEBS Letters*. 318:200-204.

Spicer, A.P., M.F. Seldin, and S.J. Gendler. 1995. Molecular cloning and chromosomal localization of the mouse decay-accelerating factor genes. Duplicated genes encode glycosylphosphatidylinositol-anchored and transmembrane forms. *Journal of Immunology*. 155:3079-3091.

Stanners, C.P. 1998. Contributions of the human CEA family to malignant transformation. *In* Cell Adhesion and Communication Mediated by the CEA Family. Vol. 5. C.P. Stanners, editor. Hardwood Academic Publishers, Amsterdam. 141-154.

Stanners, C.P., L. DeMarte, M. Rojas, P. Gold, and A. Fuks. 1995. Opposite functions for two classes of genes of the human carcinoembryonic antigen family. *Tumour Biology*. 16:23-31.

Stanners, C.P., and A. Fuks. 1998. Properties of Adhesion Mediated by the Human CEA Family. *In* Cell Adhesion and Communication Mediated by the CEA Family. Vol. 5. S.C. P., editor. Hardwood Academic Publishers, Amsterdam. 57-71.

Stanners, C.P., M. Rojas, H. Zhou, A. Fuks, and N. Beauchemin. 1992. The CEA family: a system in transitional evolution? *International Journal of Biological Markers*. 7:137-142.

Stefanova, I., V. Horejsi, I.J. Ansotegui, W. Knapp, and H. Stockinger. 1991. GPI-anchored cell-surface molecules complexed to protein tyrosine kinases. *Science*. 254:1016-1019.

Stevens, V.L. 1995. Biosynthesis of glycosylphosphatidylinositol membrane anchors. *Biochemical Journal*. 310:361-370.

Streydio, C., S. Swillens, M. Georges, C. Szpirer, and G. Vassart. 1990. Structure, evolution and chromosomal localization of the human pregnancy-specific beta 1-glycoprotein gene family. *Genomics*. 7:661-662.

Stubbs, L., E.A. Carver, M.E. Shannon, J. Kim, J. Geisler, E.E. Generoso, B.G. Stanford, W.C. Dunn, H. Mohrenweiser, W. Zimmermann, S.M. Watt, and L.K. Ashworth. 1996. Detailed comparative map of human chromosome 19q and related regions of the mouse genome. *Genomics*. 35:499-508.

Su, B., G.L. Waneck, R.A. Flavell, and A.L. Bothwell. 1991. The glycosyl phosphatidylinositol anchor is critical for Ly-6A/E-mediated T cell activation. *Journal of Cell Biology*. 112:377-384.

Sundblad, A.S., E.M. Pellicer, and L. Ricci. 1996. Carcinoembryonic antigen expression in stages I and II breast cancer: its relationship with clinicopathologic factors. *Human Pathology*. 27:297-301.

Takeichi, M. 1991. Cadherin cell adhesion receptors as a morphogenetic regulator. *Science*. 251:1451-1455.

Takeichi, M. 1993. Cadherins in cancer: implications for invasion and metastasis. *Current Opinion in Cell Biology*. 5:806-811.

Teixeira, A.M., J. Fawcett, D.L. Simmons, and S.M. Watt. 1994. The N-domain of the biliary glycoprotein (BGP) adhesion molecule mediates homotypic binding: domain interactions and epitope analysis of BGPc. *Blood*. 84:211-219.

Thomas, P.M., and L.E. Samelson. 1992. The glycophosphatidylinositol-anchored Thy-1 molecule interacts with the p60fyn protein tyrosine kinase in T cells. *Journal of Biological Chemistry*. 267:12317-12322.

Thompson, J., W. Zimmermann, P. Nollau, M. Neumaier, J. Weber-Arden, H. Schrewe, I. Craig, and T. Willcocks. 1994. CGM2, a member of the carcinoembryonic antigen gene family is down-regulated in colorectal carcinomas. *Journal of Biological Chemistry*. 269:32924-32931.

Thompson, J.A., F. Grunert, and W. Zimmermann. 1991. Carcinoembryonic antigen gene family: molecular biology and clinical perspectives. *Journal of Clinical Laboratory Analysis*. 5:344-366.

Thompson, J.A., E.M. Mauch, F.S. Chen, Y. Hinoda, H. Schrewe, B. Berling, S. Barnert, S. von Kleist, J.E. Shively, and W. Zimmermann. 1989. Analysis of the size of the carcinoembryonic antigen (CEA) gene family: isolation and sequencing of N-terminal domain exons. *Biochemical & Biophysical Research Communications*. 158:996-1004.

Tobi, M., M. Memon, K. Kithier, and N. Clapp. 1994. A putative CEA moiety is shared by the cotton-top tamarin (Saguinus oedipus) and humans. *Cancer Letters*. 77:7-13.

Turbide, C., M. Rojas, C.P. Stanners, and N. Beauchemin. 1991. A mouse carcinoembryonic antigen gene family member is a calcium-dependent cell adhesion molecule. *Journal of Biological Chemistry*. 266:309-315.

Udenfriend, S., and K. Kodukula. 1995. Prediction of omega site in nascent precursor of glycosylphosphatidylinositol protein. *Methods in Enzymology*. 250:571-582.

Varela-Nieto, I., Y. Leon, and H.N. Caro. 1996. Cell signalling by inositol phosphoglycans from different species. *Comparative Biochemistry & Physiology. Part B, Biochemistry & Molecular Biology*. 115:223-241.

Varma, R., and S. Mayor. 1998. Gpi-Anchored Proteins Are Organized In Submicron Domains At the Cell Surface. *Nature*. 394:798-801.

Virji, M., K. Makepeace, D.J. Ferguson, and S.M. Watt. 1996. Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic neisseriae. *Molecular Microbiology*. 22:941-950.

Wallich, R., C. Brenner, Y. Brand, M. Roux, M. Reister, and S. Meuer. 1998. Gene structure, promoter characterization, and basis for alternative mRNA splicing of the human CD58 gene. *Journal of Immunology*. 160:2862-2871.

Waneck, G.L., M.E. Stein, and R.A. Flavell. 1988. Conversion of a PI-anchored protein to an integral membrane protein by a single amino acid mutation. *Science*. 241:697-699.

Watt, S.M., J. Fawcett, S.J. Murdoch, A.M. Teixeira, S.E. Gschmeissner, N.M. Hajibagheri, and D.L. Simmons. 1994. CD66 identifies the biliary glycoprotein (BGP) adhesion molecule: cloning, expression, and adhesion functions of the BGPc splice variant. *Blood*. 84:200-210.

Wegener, A.M., F. Letourneur, A. Hoeveler, T. Brocker, F. Luton, and B. Malissen. 1992. The T cell receptor/CD3 complex is composed of at least two autonomous transduction modules. *Cell*. 68:83-95.

Wessner, D.R., P.C. Shick, J.H. Lu, C.B. Cardellichio, S.E. Gagneten, N. Beauchemin, K.V. Holmes, and G.S. Dveksler. 1998. Mutational analysis of the virus and monoclonal antibody binding sites in MHVR, the cellular

receptor of the murine coronavirus mouse hepatitis virus strain A59. *Journal of Virology*. 72:1941-1948.

Wikstrom, K., G. Kjellstrom, and B. Obrink. 1996. Homophilic intercellular adhesion mediated by C-CAM is due to a domain 1-domain 1 reciprocal binding. *Experimental Cell Research*. 227:360-366.

Wilson, A.C. 1985. The molecular basis of evolution. *Scientific American*. 253:164-173.

Yaffe, D. 1968. Retention of differentiation potentialities during prolonged cultivation of myogenic cells. *Proceedings of the National Academy of Sciences of the United States of America*. 61:477-483.

Yan, Z., A. Robinson-Saddler, S. Winawer, and E. Friedman. 1993. Colon carcinoma cells blocked in polarization exhibit increased expression of carcinoembryonic antigen. *Cell Growth & Differentiation*. 4:785-792.

Yu, J., S. Nagarajan, J.J. Knez, S. Udenfriend, R. Chen, and M.E. Medof. 1997. The affected gene underlying the class K glycosylphosphatidylinositol (GPI) surface protein defect codes for the GPI transamidase. *Proceedings of the National Academy of Sciences of the United States of America*. 94:12580-12585.

Zhang, J., and S. Kumar. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology & Evolution*. 14:527-536.

Zhou, H., A. Fuks, G. Alcaraz, T.J. Bolling, and C.P. Stanners. 1993a. Homophilic adhesion between Ig superfamily carcinoembryonic antigen molecules involves double reciprocal bonds. *Journal of Cell Biology*. 122:951-960.

Zhou, H., A. Fuks, and C.P. Stanners. 1990. Specificity of intercellular adhesion mediated by various members of the immunoglobulin supergene family. *Cell Growth & Differentiation*. 1:209-215.

Zhou, H., C.P. Stanners, and A. Fuks. 1993b. Specificity of anti-carcinoembryonic antigen monoclonal antibodies and their effects on CEA-mediated adhesion. *Cancer Research*. 53:3817-3822.

Zimmermann, W. 1998. The nature and expression of the rodent CEA families: evolutionary considerations. *In* Cell Adhesion and Communication Mediated by the CEA Family. Vol. 5. C.P. Stanners, editor. Hardwood Academic Publishers, Amsterdam. 31-55.

Zimmermann, W., B. Weber, B. Ortlieb, F. Rudert, W. Schempp, H.H. Fiebig, J.E. Shively, S. von Kleist, and J.A. Thompson. 1988. Chromosomal localization of the carcinoembryonic antigen gene family and differential expression in various tumors. *Cancer Research*. 48:2550-2554.