Molecular Biology Reports **17**: 123–128, 1993.
© 1993 *Kluwer Academic Publishers. Printed in Belgium.*

Molecular cloning and characterization of the human E-cadherin cDNA

M.J.G. Bussemakers, A.van Bokhoven, S.G.M. Mees, R. Kemler[1] & J.A. Schalken*
Urology Research Laboratory, University Hospital Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands, [1]*Max-Planck-Institut für Immunbiologie, FG Molekulare Embryologie, D-7800 Freiburg, Germany* **Author for correspondence*

Received 20 August 1992; accepted in revised form 22 October 1992

Key words: cDNA cloning, E-cadherin, human, invasion suppressor gene, sequence analysis

Abstract

E-cadherin is a Ca^{2+}-dependent cell adhesion molecule involved in cell-cell interaction. In its normal physiological function it plays an important role in embryonic development and tissue morphogenesis. Recent studies have shown that in cancer development E-cadherin can act as a suppressor of invasion. Indeed, in several kinds of carcinomas allelic loss of the E-cadherin/Uvomorulin locus and decreased E-cadherin expression have been described. The importance of E-cadherin in human cancer development may be substantiated by molecular analysis of the E-cadherin transcript. Therefore, we isolated and characterized the human E-cadherin cDNA. Comparison of the nucleotide and deduced amino acid sequences revealed that the human E-cadherin is highly homologous to the mouse E-cadherin (uvomorulin) and to other members of the cadherin family.

Introduction

Cell adhesion is of fundamental importance in the establishment and maintenance of tissue form and function [1]. Several proteins are known to be involved in cell-cell adhesion like certain integrins, adhesion molecules belonging to the immunoglobulin superfamily, selectins and cadherins [2]. Since adhesion molecules play a role in cell attachment, cellular motility and intercellular communication, uncontrolled expression of these molecules may lead to changes in cellular adhesion and to increased motility, processes that can contribute to the invasive and/or metastatic behaviour of cells. Recent studies revealed a functional relation of E-cadherin (also termed uvomorulin [mouse], L-CAM [chicken], cell-CAM 120/80 [human] and Arc-1 [dog]) with inhibition of invasion and suggest that E-cadherin can act as an invasion suppressor gene product [3–5]. Studies on the expression of E-cadherin in tumors have shown that indeed in invasive tumors the E-cadherin expression is often decreased [6–8]. Also our own studies on prostatic tumors showed that a decreased expression of E-cadherin correlated with tumor progression [9, 10]. Interestingly, restriction fragment length polymorphism (RFLP) analyses of human prostate cancer specimens have shown a frequent loss of heterozygosity of the chromosome 16q [11, 12], including the region where the Uvomorulin/E-cadherin gene is mapped [13]. Also in hepato-cellular carcinomas [14], in breast cancer [15] and in neuroectodermal tu-

mors of the central nervous system [16], frequent deletions involving the same chromosomal region have been reported. If E-cadherin would be the relevant tumor suppressor gene involved, classical tumor suppressor gene theories [17] would predict that the remaining allele is mutated, thus abolishing E-cadherin function. Even more, it should be noted that as in the case of other tumor suppressor genes e.g. p53), dominant negative mutations in E-cadherin may occur. Alternatively, the reduced dosage of the gene due to allelic loss may result in reduced expression of E-cadherin below a critical threshold, which by itself could impair functioning of E-cadherin. Evidence for this possibility comes from studies which showed that a reduction, not elimination, of E-cadherin expression was sufficient to induce invasiveness of kidney carcinoma cells [5].

However, as mentioned before, the aberrant expression of E-cadherin may be caused by genetic changes. Until now, molecular genetical studies were hampered by the fact that the human E-cadherin gene and cDNA have not yet been cloned. Here we report the isolation and characterization of the human E-cadherin cDNA which now enables molecular studies of E-cadherin transcripts.

Materials and methods

Screening of the cDNA library

A cDNA library constructed according to Huynh et al. [18] (oligo-dT primed, cloned into lambda gt11) using human (non-neoplastic) pancreas tissue, was screened according to Sambrook *et al.* [19] using pV962, a 0.6 kb human E-cadherin cDNA [13] (see Fig. 1), as a probe. Thus, lambda hEc3 was isolated. Using a small subclone derived from the 3′-end of lambda hEc3 the human pancreas cDNA library was rescreened and lambda hEc10 was obtained. Using the 3′-end of this clone we were unable to obtain the complete 3′-end of the human E-cadherin cDNA. Therefore, a cDNA library constructed according to Huynh *et al.* [18] (oligo-dT primed, cloned into lambda gt11) using human (non-neoplastic) liver tissue was screened and lambda hEc14 was isolated.

Nucleotide sequence analysis

DNA fragments were ligated into the polylinker region of M13mp8–19. All of the DNA se-

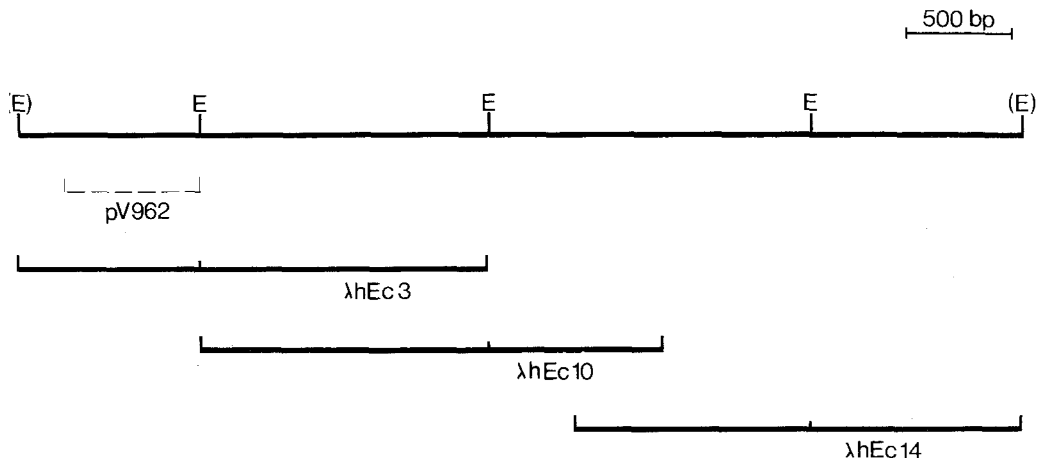

Fig. 1. Schematic presentation of the molecular cloning of the human E-cadherin cDNA. The probe used for the isolation, pV962 [13], is indicated, as are the phage lambda clones that were isolated to obtain the complete cDNA of 4.8 kbp. (E = *Eco* RI) (lambda hEc3 and 10 were isolated from a pancreas cDNA library, lambda hEc14 was isolated from a liver cDNA library).

quences were determined on both strands using the dideoxy sequencing method as described by Sanger *et al.* [20]. The gel readings were recorded and edited using IntelliGenetics computer software (release 5.35). Computer comparison studies were performed using sequences obtained from the EMBL (release 30) and Genbank (release 71) nucleotide sequence databases [21], using CAMMSA computer software.

Results and discussion

Upon screening of the human pancreas cDNA library using pV962 as a probe, we identified lambda hEc3 (see Fig. 1) which upon nucleotide sequence analysis and comparison to the mouse E-cadherin (uvomorulin) [22] could be shown to contain the translation initiation site and 94 nucleotides of the 5′ non-coding region (see Fig. 2). In order to identify the 3′-end of the cDNA, small subclones derived from the 3′-end of lambda hEc3 and lambda hEc10 were used. The nucleotide sequence and deduced amino acid sequence of human E-cadherin is shown in Figure 2. Although no obvious poly-A-addition signal is present (also lacking in the mouse E-cadherin cDNA), the cDNA contained a poly-A-tail confirming that we isolated the full length cDNA. The human E-cadherin cDNA has an open reading frame encoding a protein of 882 amino acids and a 3′ non-coding region of 2035 nucleotides and has a homology with mouse E-cadherin of 78% and with the chicken E-cadherin related sequence, L-CAM [23] of 65%.

In Figure 2, various important functional domains of E-cadherin are indicated. In Figure 3, the structure of the E-cadherin protein is shown. E-cadherin is synthesized as a precursor polypeptide which upon posttranslational modifications including proteolytic cleavage and glycosylation gives rise to the mature protein [24] (the first amino acid of the mature protein is indicated with ◆ in Fig. 2, the precursor segment is indicated with the thin line in Fig. 3; putative glycosylation sites are indicated by ●●● in Fig. 2). Furthermore, in Figure 3 the homology between human and mouse E-cadherin is shown. It is clear that the mature protein is more highly conserved than the precursor segment. When the extracellular domain of E-cadherin is considered, three internally repeated domains of 112 amino acids which are involved in the Ca^{2+}-binding can be discriminated [22]. The small region near the proteolytic cleavage site is well conserved and may be important for the recognition and the correct cleavage of E-cadherin, since incorrect or no cleavage of the precursor polypeptide abolishes the functioning of E-cadherin [24]. Also the three internally repeated domains in the extracellular part of E-cadherin are highly conserved (compared to chicken L-CAM the three internally repeated domains have a homology of 75, 57, and 60% respectively). Comparison of the Ca^{2+}-binding domains with mouse sequences showed that the few changes that occurred in these domains during evolution are conservative amino acid changes. The transmembrane domain of the human E-cadherin is identical to the mouse E-cadherin (to chicken L-CAM 96%) and also the cytoplasmic domain is almost identical (whereby mainly conservative amino acid changes occurred / compared to chicken L-CAM the homology is 89%). The cytoplasmic domain has been shown to be of importance for the regulation of the cell-cell binding function of the extracellular domain of the molecule, through interactions with cytoskeletal components by catenins [25, 26].

Comparison of the human E-cadherin to the other well characterized members of the cadherin family, showed a homology to human P-cadherin [27] of 61% at the DNA-level and of 76% at the protein level when compared to the mature protein; with human N-cadherin [28] a homology of 58% at the DNA-level and 67% at (mature) protein level was found.

In conclusion, E-cadherin is well conserved between the different species and also the sequence homology among the members of the cadherin family is high. By determining the nucleotide sequence of the human E-cadherin cDNA we are now able to molecularly study the E-cadherin transcript (mutation analysis) in human cancer development.

```
                                                 50
GCTTGCGGAAGTCAGTTCAGACTCCAGCCCGCTCCAGCCCGGCCCGACCCGACCGCACCCGGCGCCTGCCCTCGCTCGGCGTCCCCGGCCAGCC

    100                                                         150
ATG GGC CCT TGG AGC CGC AGC CTC TCG GCG CTG CTG CTG CTG CTG CAG GTC TCC TCT TGG CTC TGC CAG GAG CCG GAG
M   G   P   W   S   R   S   L   S   A   L   L   L   L   L   Q   V   S   S   W   L   C   Q   E   P   E

                                    200                                                             250
CCC TGC CAC CCT GGC TTT GAC GCC GAG AGC TAC ACG TTC ACG GTG CCC CGG CGC CAC CTG GAG AGA GGC CGC GTC CTG
P   C   H   P   G   F   D   A   E   S   Y   T   F   T   V   P   R   R   H   L   E   R   G   R   V   L

                                                            300
GGC AGA GTG AAT TTT GAA GAT TGC ACC GGT CGA CAA AGG ACA GCC TAT TTT TCC CTC GAC ACC CGA TTC AAA GTG GGC
G   R   V   N   F   E   D   C   T   G   R   Q   R   T   A   Y   F   S   L   D   T   R   F   K   V   G

                        350                                                                 400
ACA GAT GGT GTG ATT ACA GTC AAA AGG CCT CTA CGG TTT CAT AAC CCA CAG ATC CAT TTC TTG GTC TAC GCC TGG GAC
T   D   G   V   I   T   V   K   R   P   L   R   F   H   N   P   Q   I   H   F   L   V   Y   A   W   D

                                                    450
TCC ACC TAC AGA AAG TTT TCC ACC AAA GTC ACG CTG AAT ACA GTG GGG CAC CAC CAC CGC CCC CCG CCC CAT CAG GCC
S   T   Y   R   K   F   S   T   K   V   T   L   N   T   V   G   H   H   H   R   P   P   P   H   Q   A

                500                                                                 550
TCC GTT TCT GGA ATC CAA GCA GAA TTG CTC ACA TTT CCC AAC TCC TCT CCT GGC CTC AGA AGA CAG AAG AGA GAC TGG
S   V   S   G   I   Q   A   E   L   L   T   F   P   N   S   S   P   G   L   R   R   Q   K   R   D   W
                                                                                                ◆
                                            600
GTT ATT CCT CCC ATC AGC TGC CCA GAA AAT GAA AAA GGC CCA TTT CCT AAA AAC CTG GTT CAG ATC AAA TCC AAC AAA
V   I   P   P   I   S   C   P   E   N   E   K   G   P   F   P   K   N   L   V   Q   I   K   S   N   K

        650                                                                 700
GAC AAA GAA GGC AAG GTT TTC TAC AGC ATC ACT GGC CAA GGA GCT GAC ACA CCC CCT GTT GGT GTC TTT ATT ATT GAA
D   K   E   G   K   V   F   Y   S   I   T   G   Q   G   A   D   T   P   P   V   G   V   F   I   I   E

                                    750
AGA GAA ACA GGA TGG CTG AAG GTG ACA GAG CCT CTG GAT AGA GAA CGC ATT GCC ACA TAC ACT CTC TTC TCT CAC GCT
R   E   T   G   W   L   K   V   T   E   P   L   D   R   E   R   I   A   T   Y   T   L   F   S   H   A

 800                                                            850
GTG TCA TCC AAC GGG AAT GCA GTT GAG GAT CCA ATG GAG ATT TTG ATC ACG GTA ACC GAT CAG AAT GAC AAC AAG CCC
V   S   S   N   G   N   A   V   E   D   P   M   E   I   L   I   T   V   T   D   Q   N   D   N   K   P

                            900                                                                 950
GAA TTC ACC CAG GAG GTC TTT AAG GGG TCT GTC ATG GAA GGT GCT CTT CCA GGA ACC TCT GTG ATG GAG GTC ACA GCC
E   F   T   Q   E   V   F   K   G   S   V   M   E   G   A   L   P   G   T   S   V   M   E   V   T   A

                                                        1000
ACA GAC GCG GAC GAT GAT GTG AAC ACC TAC AAT GCC GCC ATC GCT TAC ACC ATC CTC AGC CAA GAT CCT GAG CTC CCT
T   D   A   D   D   D   V   N   T   Y   N   A   A   I   A   Y   T   I   L   S   Q   D   P   E   L   P

                    1050                                                                1100
GAC AAA AAT ATG TTC ACC ATT AAC AGG AAC ACA GGA GTC ATC AGT GTG GTC ACC ACT GGG CTG GAC CGA GAG AGT TTC
D   K   N   M   F   T   I   N   R   N   T   G   V   I   S   V   V   T   T   G   L   D   R   E   S   F

                                                1150
CCT ACG TAT ACC CTG GTG GTT CAA GCT GCT GAC CTT CAA GGT GAG GGG TTA AGC ACA ACA GCA ACA GCT GTG ATC ACA
P   T   Y   T   L   V   V   Q   A   A   D   L   Q   G   E   G   L   S   T   T   A   T   A   V   I   T

            1200                                                                1250
GTC ACT GAC ACC AAC GAT AAT CCT CCG ATC TTC AAT CCC ACC ACG TAC AAG GGT CAG GTG CCT GAG AAC GAG GCT AAC
V   T   D   T   N   D   N   P   P   I   F   N   P   T   T   Y   K   G   Q   V   P   E   N   E   A   N
                                            ●●●●●●●●●●●
                                            1300
GTC GTA ATC ACC ACA CTG AAA GTG ACT GAT GCT GAT GCC CCC AAT ACC CCA GCG TGG GAG GCT GTA TAC ACC ATA TTG
V   V   I   T   T   L   K   V   T   D   A   D   A   P   N   T   P   A   W   E   A   V   Y   T   I   L

        1350                                                        1400
AAT GAT GAT GGT GGA CAA TTT GTC GTC ACC ACA AAT CCA GTG AAC AAC GAT GGC ATT TTG AAA ACA GCA AAG GGC TTG
N   D   D   G   G   Q   F   V   V   T   T   N   P   V   N   N   D   G   I   L   K   T   A   K   G   L

                                    1450                                                        1498
GAT TTT GAG GCC AAG CAG CAG TAC ATT CTA CAC GTA GCA GTG ACG AAT GTG GTA CCT TTT GAG GTC TCT CTC ACC ACC
D   F   E   A   K   Q   Q   Y   I   L   H   V   A   V   T   N   V   V   P   F   E   V   S   L   T   T

                                                                1550
TCC ACA GCC ACC GTC ACC GTG GAT GTG CTG GAT GTG AAT GAA GCC CCC ATC TTT GTG CCT CCT GAA AAG AGA GTG GAA
S   T   A   T   V   T   V   D   V   L   D   V   N   E   A   P   I   F   V   P   P   E   K   R   V   E

                            1600                                                            1650
GTG TCC GAG GAC TTT GGC GTG GGC CAG GAA ATC ACA TCC TAC ACT GCC CAG GAG CCA GAC ACA TTT ATG GAA CAG AAA
V   S   E   D   F   G   V   G   Q   E   I   T   S   Y   T   A   Q   E   P   D   T   F   M   E   Q   K

                                                        1700
ATA ACA TAT CGG ATT TGG AGA GAC ACT GCC AAC TGG CTG GAG ATT AAT CCG GAC ACT GGT GCC ATT TCC ACT CGG GCT
I   T   Y   R   I   W   R   D   T   A   N   W   L   E   I   N   P   D   T   G   A   I   S   T   R   A

                    1750                                                            1800
GAG CTG GAC AGG GAG GAT TTT GAG CAC GTG AAG AAC AGC ACG TAC ACA GCC CTA ATC ATA GCT ACA GAC AAT GGT TCT
E   L   D   R   E   D   F   E   H   V   K   N   S   T   Y   T   A   L   I   I   A   T   D   N   G   S
                                            ●●●●●●●●●●●
                                                1850
CCA GTT GCT ACT GGA ACA GGG ACA CTT CTG CTG ATC CTG TCT GAT GTG AAT GAC AAC GCC CCC ATA CCA GAA CCT CGA
P   V   A   T   G   T   G   T   L   L   L   I   L   S   D   V   N   D   N   A   P   I   P   E   P   R

            1900                                                                1950
ACT ATA TTC TTC TGT GAG AGG AAT CCA AAG CCT CAG GTC ATA AAC ATC ATT GAT GCA GAC CTT CCT CCC AAT ACA TCT
T   I   F   F   C   E   R   N   P   K   P   Q   V   I   N   I   I   D   A   D   L   P   P   N   T   S
                                                                                        ●●●●●●●●●●●
                                        2000
CCC TTC ACA GCA GAA CTA ACA CAC GGG GCG AGT GCC AAC TGG ACC ATT CAG TAC AAC GAC CCA ACC CAA GAA TCT ATC
P   F   T   A   E   L   T   H   G   A   S   A   N   W   T   I   Q   Y   N   D   P   T   Q   E   S   I
                                                ●●●●●●●●●●●
```

Fig. 2. Nucleotide sequence and deduced amino acid sequence (in one-letter-code) of human E-cadherin. ◆ the first amino acid of the mature E-cadherin protein; ●●● consensus sequence for N-linked glycosylation; ▬▬▬ the transmembrane domain; *** stopcodon; the extracellular repeated domains are underlined; the Ca^{2+}-binding domains are indicated by a double line (sequence data have been deposited with the EMBL/Genbank Data Libraries under the accession no. Z13009).

```
2050                                                           2100
ATT TTG AAG CCA AAG ATG GCC TTA GAG GTG GGT GAC TAC AAA ATC AAT CTC AAG CTC ATG GAT AAC CAG AAT AAA GAC
I   L   K   P   K   M   A   L   E   V   G   D   Y   K   I   N   L   K   L   M   D   N   Q   N   K   D

                                2150                                                                2200
CAA GTG ACC ACC TTA GAG GTC AGC GTG TGT GAC TGT GAA GGG GCC GCC GGC GTC TGT AGG AAG GCA CAG CCT GTC GAA
Q   V   T   T   L   E   V   S   V   C   D   C   E   G   A   A   G   V   C   R   K   A   Q   P   V   E

                                                                2250
GCA GGA TTG CAA ATT CCT GCC ATT CTG GGG ATT CTT GGA GGA ATT CTT GCT TTG CTA ATT CTG ATT CTG CTG CTC TTG
A   G   L   Q   I   P   A   I   L   G   I   L   G   G   I   L   A   L   L   I   L   I   L   L   L   L

                        2300                                                                2350
CTG TTT CTT CGG AGG AGA GCG GTG GTC AAA GAG CCC TTA CTG CCC CCA GAG GAT GAC ACC CGG GAC AAC GTT TAT TAC
L   F   L   R   R   R   A   V   V   K   E   P   L   L   P   P   E   D   D   T   R   D   N   V   Y   Y

                                                        2400
TAT GAT GAA GAA GGA GGC GGA GAA GAG GAC CAG GAC TTT GAC TTG AGC CAG CTG CAC AGG GGC CTG GAC GCT CGG CCT
Y   D   E   E   G   G   G   E   E   D   Q   D   F   D   L   S   Q   L   H   R   G   L   D   A   R   P

                2450                                                            2500
GAA GTG ACT CGT AAC GAC GTT GCA CCA ACC CTC ATG AGT GTC CCC CGG TAT CTT CCC CGC CCT GCC AAT CCC GAT GAA
E   V   T   R   N   D   V   A   P   T   L   M   S   V   P   R   Y   L   P   R   P   A   N   P   D   E

                                                2550
ATT GGA AAT TTT ATT GAT GAA AAT CTG AAA GCG GCT GAT ACT GAC CCC ACA GCC CCG CCT TAT GAT TCT CTG CTC GTG
I   G   N   F   I   D   E   N   L   K   A   A   D   T   D   P   T   A   P   P   Y   D   S   L   L   V

        2600                                                            2650
TTT GAC TAT GAA GGA AGC GGT TCC GAA GCT GCT AGT CTG AGC TCC CTG AAC TCC TCA GAG TCA GAC AAA GAC CAG GAC
F   D   Y   E   G   S   G   S   E   A   A   S   L   S   S   L   N   S   S   E   S   D   K   D   Q   D

                                    2700
TAT GAC TAC TTG AAC GAA TGG GGC AAT CGC TTC AAG AAG CTG GCT GAC ATG TAC GGA GGC GGC GAG GAC GAC TAG GGG
Y   D   Y   L   N   E   W   G   N   R   F   K   K   L   A   D   M   Y   G   G   G   E   D   D   ***

2750                                          2800                                                2849
ACTCGAGAGAGGCGGGCCCCAGACCCATGTGCTGGGAAATGCAGAAATCACGTTGCTGGTGGTTTTCAGCTCCCTTCCCTTGAGATGAGTTTCTGGGGAAAA

                                            2900                                                2950
AAAAGAGACTGGTTAGTGATGCAGTTAGTATAGCTTTATACTCTCTCCACTTTATAGCTCTAATAAGTTTGTGTTAGAAAAGTTTCGACTTATTTCTTAAAGC

                                         3000                                               3050
TTTTTTTTTTTTCCCATCACTCTTTACATGGTGGTGATGTCCAAAAGATACCCAAATTTTAATATTCCAGAAGAACAACTTTAGCATCAGAAGGTTCACCCAG

                                     3100                                               3150
CACCTTGCAGATTTTCTTAAGGAATTTTGTCTCACTTTTAAAAAGAAGGGGAGAAGTCAGCTACTCTAGTTCTGTTGTTTTGTGTATATAATTTTTTAAAAAA

                                    3200                                               3250
AATTTGTGTGCTTCTGCTCATTACTACACTGGTGTGTCCCTCTGCCTTTTTTTTTTTTTTTTAAGACAGGGTCTCATTCTATCGGCCAGGCTGGAGTGCAGTGG

                                3300                                               3350
TGCAATCACAGCTCACTGCAGCCTTGTCCTCCCAGGCTCAAGCTATCCTTGCACCTCAGCCTCCCAAGTAGCTGGGACCACAGGCATGCACCACTACGCATGA

                            3400                                               3450
CTAATTTTTTAAATATTTGAGACGGGGTCTCCCTGTGTTACCCAGGCTGGTCTCAAACTCCTGGGCTCAAGTGATCCTCCCATCTTGGCCTCCCAGAGTATTG

                        3500                                               3550
GGATTACAGACATGAGCCACTGCACCTGCCCAGCTCCCCAACTCCCTGCCATTTTTTAAGAGACAGTTTCGCTCCATCGCCCAGGCCTGGGATGCAGTGATGT

                    3600                                               3650
GATCATAGCTCACTGTAACCTCAAACTCTGGGGCTCAAGCAGTTCTCCCACCAGCCTCCTTTTTATTTTTTTGTACAGATGGGGTCTTGCTATGTTGCCCAAG

                3700                                               3750
CTGGTCTTAAACTCCTGGCCTCAAGCAATCCTTCTGCCTTGGCCCCCCAAAGTGCTGGGATTGTGGGCATGAGCTGCTGTGCCCAGCCTCCATGTTTTAATAT

            3800                                               3850
CAACTCTCACTCCTGAATTCAGTTGCTTTGCCCAAGATAGGAGTTCTCTGATGCAGAAATTATTGGGCTCTTTTAGGGTAAGAAGTTTGTGTCTTTGTCTGGC

        3900                                               3950
CACATCTTGACTAGGTATTGTCTACTCTGAAGACCTTTAATGGCTTCCCTCTTTCATCTCCTGAGTATGTAACTTGCAATGGGCAGCTATCCAGTGACTTGTT

        4000                                               4050
CTGAGTAAGTGTGTTCATTAATGTTTATTTAGCTCTGAAGCAAGAGTGATATACTCCAGGACTTAGAATAGTGCCTAAAGTGCTGCAGCCAAAGACAGAGCGG

    4100                                               4150
AACTATGAAAAGTGGGCTTGGAGATGGCAGGAGAGCTTGTCATTGAGCCTGGCAATTTAGCAAACTGATGCTGAGGATGATTGAGGTGGGTCTACCTCATCTC

    4200                                               4250
TGAAAATTCTGGAAGGAATGGAGGAGTCTCAACATGTGTTTCTGACACAAGATCCGTGGTTTGTACTCAAAGCCCAGAATCCCCAAGTGCCTGCTTTTGATGA

  4300                                               4350
TGTCTACAGAAAATGCTGGCTGAGCTGAACACATTTGCCCAATTCCAGGTGTGCACAGAAAACCGAGAATATTCAAAATTCCAAATTTTTTCTTAGGAGCAAG

 4400                                               4450                                              4497
AAGAAAATGTGGCCCTAAAGGGGGTTAGTTGAGGGGTAGGGGGTAGTGAGGATCTTGATTTGGATCTCTTTTTATTTAAATGTGAATTTCAACTTTTGACAAT

                                                4550                                              4600
CAAAGAAAAGACTTTTGTTGAAATAGCTTTACTGTTTCTCAAGTGTTTTGGAGAAAAAAATCAACCCTGCAATCACTTTTTGGAATTGTCTTGATTTTTCGGC

                                            4650                                              4700
AGTTCAAGCTATATCGAATATAGTTCTGTGTAGAGAATGTCACTGTAGTTTTGAGTGTATACATGTGTGGGTGCTGATAATTGTGTATTTTCTTTGGGGGTGG

                                          4750
AAAAGGAAAACAATTCAAGCTGAGAAAAGTATTCTCAAAGATGCATTTTTATAAATTTTATTAAACAATTTTGTT(A)n
```

Fig. 2. Continued.

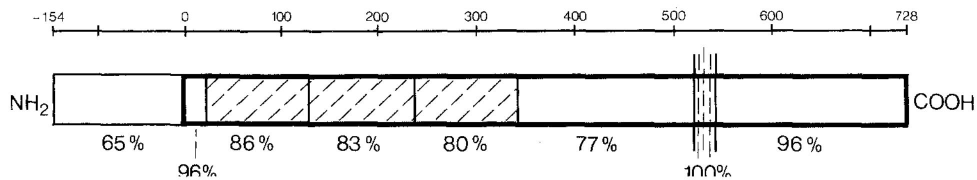

Fig. 3. Schematic presentation of the structure of human E-cadherin. The thin line indicates the precursor segment; the fat line indicates the mature protein; the cross-hatched boxes indicate the extracellular repeated domains; the transmembrane domain is vertically striped. The percentage homology (amino acids) of the different regions of the human E-cadherin with the mouse E-cadherin [22] is also indicated.

Acknowledgements

These studies were supported by the Dutch Cancer Society (NUKC 92-81) (M.J.G.B., A.v.B.) and the Royal Dutch Academy of Arts and Sciences (J.A.S.).

References

1. Edelman GM (1988) Biochemistry 27: 3533–3543
2. Edelman GM & Crossin KL (1991) Annu. Rev. Biochem. 60: 155–190
3. Frixen UH, Behrens J, Sachs M, Eberle G, Voss B, Warda A, Löchner D & Birchmeier W (1991) J. Cell Biol. 113: 173–185
4. Behrens J, Mareel MM, Van Roy FM & Birchmeier W (1989) J. Cell Biol. 108: 2435–2447
5. Vleminckx K, Vakaet Jr. L, Mareel MM, Fiers W & Van Roy FM (1991) Cell 66: 107–119
6. Shimoyama Y & Hirohashi S (1991) Cancer Letters 57: 131–135
7. Hashimoto M, Niwa O, Nitta Y, Takeichi M & Yokoro KAF (1989) Jpn. J. Cancer Res. 80: 459–463
8. Schipper JH, Frixen UH, Behrens J, Unger A, Jahnke K & Birchmeier W (1991) Cancer Res. 51: 6328–6337
9. Bussemakers MJG, Van Moorselaar RJA, Giroldi LA, Ichikawa T, Isaacs JT, Takeichi M, Debruyne FMJ & Schalken JA (1992) Cancer Res. 52: 2916–2922
10. Umbas R, Schalken JA, Aalders TW, Carter BS, Karthaus HFM, Schaafsma HE, Debruyne FMJ & Isaacs WB (1992) Cancer Res. 52: In press
11. Carter BS, Ewing CM, Ward WS, Treiger BF, Aalders TW, Schalken JA, Epstein JI & Isaacs WB (1991) Proc. Natl. Acad. Sci. USA 87: 8751–8755
12. Bergerheim US, Kunimi K, Collins VP & Ekman P (1991) Genes Chromosome Cancer 3: 215–220
13. Mansouri A, Spurr N, Goodfellow PN & Kemler R (1988) Differentiation 38: 67–71
14. Tsuda H, Zhang W, Shimosato Y, Yokota J, Terada M, Sugimura T, Miyamura T & Hirohashi S (1990) Proc. Natl. Acad. Sci. USA 87: 6791–6794
15. Sato T, Tanigami A, Yamakawa K, Akiyama F, Kasumi F, Sakamoto G & Nakamura Y (1990) Cancer Res. 50: 7184–7189
16. Thomas GA & Raffel C (1991) Cancer Res. 51: 639–643
17. Knudson AG (1971) Proc. Natl. Acad. Sci. USA 68: 820–823
18. Huynh TU, Young RA & Davis RW (1985) DNA Cloning Techniques: A practical approach. IRL Press, Oxford
19. Sambrook J, Fritsch EF & Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
20. Sanger F, Coulson AR, Barrell BG, Smith AJH & Roe BA (1980) J. Mol. Biol. 143: 161–178
21. Devereux J, Haeberli P & Smithies O (1984) Nucleic Acids Res. 12: 387–395
22. Ringwald M, Schuh R, Vestweber D, Eistetter H, Lottspeich F, Engel J, Dölz R, Jähnig F, Epplen J, Mayer S, Müller C & Kemler R (1987) EMBO J. 6: 3647–3653
23. Gallin WJ, Sorkin BC, Edelman GM & Cunningham BA (1987) Proc. Natl. Acad. Sci. USA 84: 2808–2812
24. Ozawa M & Kemler R (1990) J. Cell Biol. 111: 1645–1650
25. Nagafuchi A & Takeichi M (1988) EMBO J. 7: 3679–3684
26. Ozawa M, Ringwald M & Kemler R. (1990) Proc. Natl. Acad. Sci. USA 87: 4246–4250
27. Nose A, Nagafuchi A & Takeichi M (1987) EMBO J. 6: 3655–3661
28. Walsh FS, Barton CH, Putt W, Moore SE, Kelsell D, Spurr N & Goodfellow PN (1990) J. Cell. Biol. 55: 805–812