

Cloning and Sequence Analysis of Desmosomal Glycoproteins 2 and 3 (Desmocollins): Cadherin-like Desmosomal Adhesion Molecules with Heterogeneous Cytoplasmic Domains

Jane E. Collins, P. Kevin Legan, Terry P. Kenny, Jennie MacGarvie, Janice L. Holton, and David R. Garrod

Cancer Research Campaign, Medical Oncology Unit, University of Southampton, Southampton General Hospital, Southampton, SO9 4XY United Kingdom

Abstract. Desmosomal glycoproteins 2 and 3 (dg2 and 3) or desmocollins have been implicated in desmosome adhesion. We have obtained a 5.0-kb-long clone for dg3 from a bovine nasal epidermal λ gt11 cDNA library. Sequence analysis of this clone reveals an open reading frame of 2,517 bases encoding a polypeptide of 839 amino acids. The sequence consists of a signal peptide of 28 amino acids, a precursor sequence of 104 amino acids, and a mature protein of 707 amino acids. The latter has the characteristics of a transmembrane glycoprotein with an extracellular domain of 550 amino acids and a cytoplasmic domain of 122 amino acids.

The sequence of a partial clone from the same library shows that dg2 has an alternative COOH terminus that is extended by 54 amino acids. Genomic DNA sequence data show that this arises by splicing out of a 46-bp exon that encodes the COOH-terminal 11 amino acids of dg3 and contains an in-frame stop codon.

The extracellular domain of dg3 shows 39.4% protein sequence identity with bovine N-cadherin and 28.4% identity with the other major desmosomal glycoprotein, dg1, or desmoglein. The cytoplasmic domain of dg3 and the partial cytoplasmic domain of dg2 show 23 and 24% identity with bovine N-cadherin, respectively.

The results support our previous model for the transmembrane organization of dg2 and 3 (Parrish, E. P., J. E. Marston, D. L. Matthey, H. R. Measures, R. Venning, and D. R. Garrod. 1990. *J. Cell Sci.* 96:239-248; Holton, J. L., T. P. Kenny, P. K. Legan, J. E. Collins, J. N. Keen, R. Sharma, and D. R. Garrod. 1990. *J. Cell Sci.* 97:239-246). They suggest that these glycoproteins are specialized for calcium-dependent adhesion in their extracellular domains and, cytoplasmically, for the molecular interactions involved in desmosome plaque formation. Moreover this represents the first example of alternative splicing within the cadherin family of cell adhesion molecules.

DESMOSOMAL glycoproteins 2 and 3 (dg2/3) or desmocollins are major components of desmosomes, intercellular adhesive junctions of epithelia, cardiac muscle, follicular dendritic cells, and meninges. Localization of dg2 and 3 within the desmosome has been demonstrated by immunoelectron microscopy (Miller et al., 1987; Steinberg et al., 1987; Garrod et al., 1990; Matthey et al., 1990). We have also localized the NH₂ termini of bovine dg2 and 3 to the intercellular adhesive material of the desmosome, the desmoglea, by immunoelectron microscopy with specific anti-NH₂-terminal peptide sera (Holton et al., 1990). Inhibition of desmosome assembly in Madin-Darby bovine kidney (MDBK)¹ cells with univalent antibody fragments against dg2/3 has directly implicated these glycoproteins in desmosomal adhesion (Cowin et al., 1984b). Metabolic labeling, immunoprecipitation, and whole cell trypsiniza-

1. *Abbreviations used in this paper:* MDBK, Madin-Darby bovine kidney; N-cad, N-cadherin; PCR, polymerase chain reaction.

tion studies have confirmed that the dg2/3 polypeptides in MDCK cells are transmembrane glycoproteins in which about one-fifth of the relative molecular mass is cytoplasmic. The cytoplasmic domains also show heterogeneity of size and phosphorylation (Parrish et al., 1990).

Immunochemical characterization of the desmosomal glycoproteins using monoclonal antibodies suggests that dg2 and 3 are related to each other and are distinct from the other major desmosomal glycoprotein, dg1 or desmoglein (Cohen et al., 1983). This relationship is supported by one-dimensional peptide mapping and by amino acid analysis (Kapprell et al., 1985). While the distinction between dg2 and 3 and dg1 is generally accepted, it is our experience that polyclonal antisera raised against dg2 or 3 commonly show some cross-reactivity with dg1 (Suhrbier and Garrod, 1986). This suggests the possibility that there are epitopes shared between these glycoproteins. Components immunologically related to dg2 and 3 are found in all desmosome-bearing tissues

examined from mammalian, avian, reptilian, and anuran amphibian species (Cowin and Garrod, 1983; Cowin et al., 1984a; Suhrbier and Garrod, 1986).

Desmosomal adhesion is calcium-dependent (Hennings and Holbrook, 1983; Watt et al., 1984; Mattey and Garrod, 1986; Bologna et al., 1986) and the desmosomal glycoproteins bind calcium ions (Mattey et al., 1987; Steinberg et al., 1987). It is therefore of great interest that we have recently demonstrated significant identity between the NH₂-terminal protein sequence of dg2/3 and the calcium-dependent cell-cell adhesion molecule, N-cadherin (Holton et al., 1990). The deduced amino acid sequence of bovine dgl also shows homology with the extracellular domains of the cadherins. In contrast, the large cytoplasmic domain of dgl shows limited similarity to the cadherin cytoplasmic domain (Koch et al., 1990).

The cadherins are a family of closely related cell-cell adhesion molecules, which includes uvomorulin, A-CAM, cell CAM 120/80, and L-CAM (reviewed by Takeichi, 1988, 1990). Derived amino acid sequences of these proteins reveal a homologous family of transmembrane glycoproteins with a single membrane spanning domain, and a highly conserved cytoplasmic domain (Ringwald et al., 1987; Gallin et al., 1987; Nose et al., 1987; Nagafuchi et al., 1987; Hatta et al., 1988; Miyatani et al., 1989; Shimoyama et al., 1989; Walsh et al., 1990; Liaw et al., 1990).

In this paper we report the isolation of a full-length cDNA clone for dg3 and a partial cDNA clone encoding the cytoplasmic domain of dg2. We show that the entire extracellular domain of dg3 has homology with the cadherin family of cell-cell adhesion molecules and that the cytoplasmic domains of dg2 and 3 are generated by alternative splicing and show only limited similarity to the cytoplasmic domain of the cadherins. In support of our previous biochemical analysis, the cytoplasmic domain of dg2 is longer than that of dg3 and contains a consensus serine phosphorylation site which is absent from the shorter molecule. Our results suggest that dg2 and dg3 are specialized extracellularly for cell-cell adhesion and cytoplasmically for desmosome plaque formation.

Materials and Methods

Library Construction and Screening

Total RNA was isolated from fresh bovine nasal epidermis by the method of MacDonald et al. (1987). Poly(A)⁺ RNA was isolated from total RNA by poly(U) Sepharose 4B (Pharmacia, Milton Keynes, UK) affinity chromatography (Jacobson, 1987). Randomly primed double-stranded cDNA was synthesized from 5 µg of poly(A)⁺ RNA and packaged into λgt11 using commercial kits (Boehringer Mannheim, Lewes, UK; Amersham International, Aylesbury, UK). Plaque lifts of the resulting library were prepared using Hybond-C extra (Amersham International) and screened at a density of 8×10^3 pfu per 90-mm plate using a mixture of two rabbit polyclonal dg2/3 antisera essentially according to Young and Davis (1983) except that antibody binding was detected with ¹²⁵I-protein G (Amersham International). Positive clones were plaque purified by several rounds of rescreeing. Plaque pure clones were tested for reactivity with other antibodies using the spotting method of Green et al. (1988). The library was rescreened at high stringency with ³²P-labeled cDNA inserts from positive clones. Plaque lifts were prepared at a density of 5×10^3 pfu per 90-mm plate using Hybond-N (Amersham International) according to Benton and Davis (1977). High stringency hybridization conditions were as described for Southern blotting in the manufacturer's instructions. Two additional 30-min washes in 0.1% SDS in 0.1× SSC at 65°C were used to reduce background.

Fusion Protein Antibody

Lysogens were made in *Escherichia coli* Y1089, grown as bulk cultures and production of fusion protein induced with IPTG (isopropyl β-D-thiogalactopyranoside). Cells were harvested, lysed by freeze thawing and sonication, and fusion protein purified by electroelution from SDS-gels (Laemmli, 1970) or by passing the lysate over an anti-β-galactosidase affinity column (Promega, Southampton, UK). The fusion protein antiserum was raised by immunizing guinea pigs as previously described (Cowin and Garrod, 1983).

Western Blotting

Western blotting was performed as described by Suhrbier and Garrod (1986) using ¹²⁵I-labeled protein G or ¹²⁵I-labeled anti-mouse immunoglobulins (Amersham International).

Adsorption of Antibody Activity

Guinea pig fusion protein antiserum was diluted 1:10,000 in 1% (wt/vol) bovine serum albumin in 150 mM NaCl, 50 mM Tris-HCl, 0.05% (vol/vol) Tween 20, pH 7.4 (TB) and adsorbed with 30 µg ml⁻¹ SDS-PAGE-purified dg2/3 or 30 µg ml⁻¹ SDS-solubilized β-galactosidase. Reactivity of the adsorbed antiserum was tested on Western blots of whole desmosomes isolated from bovine nasal epidermis (Skerrow and Matoltsy, 1974).

Southern and Northern Blotting

Genomic DNA was isolated from cultured MDBK cells according to Little (1987). 10 µg aliquots of genomic DNA were digested to completion with EcoRV, EcoRI, BglII, or BamHI, run on a 0.7% agarose gel, transferred to Hybond-N (Southern et al., 1975) and UV cross-linked. Northern blots were prepared by running 1 µg aliquots of glyoxal denatured bovine nasal epidermal poly(A)⁺ RNA on 1% agarose gels (McMaster and Charnichael, 1977) before transfer to Hybond-N and UV cross-linking. Hybridization to ³²P-labeled cDNA probes and washing were performed at high stringency according to the manufacturer's instructions.

cDNA Probe Purification and Labeling

Restriction endonuclease fragments from the cDNA inserts of positive clones were separated by agarose gel electrophoresis and extracted from the gel using GeneClean (Stratagene Scientific Ltd., Luton, UK). Purified fragments were random primer labeled with ³²P-CTP (Amersham Multi-Prime, Amersham International) to a sp act of $1.2-4.0 \times 10^8$ dpm µg⁻¹ and unincorporated label removed using Nensorb columns (Dupont Ltd., Stevenage, UK).

Amplification of Specific DNA Fragments by Polymerase Chain Reaction (PCR)

PCR was performed essentially according to Saiki et al. (1985). Oligodeoxynucleotides were synthesized on an automated solid-phase synthesizer (model 381A; Applied Biosystems Ltd.) by β cyanoethyl phosphoramidite chemistry. For RNA templates, primers, sense S1 (5'-CTGGAATTCGTTAAGGGAGTGACGGA-3') and antisense A1 (5'-CCTGAATTCATAGCTATAGGAGC-3') complementary to regions ~80 bp upstream and downstream of the CN35 46-bp insertion sequence were used with 5' EcoRI restriction sites incorporated to facilitate subcloning. First strand cDNA was synthesized from 1.25 µg total bovine epidermal RNA using 3' oligonucleotide primer A1 and 5 units of AMV reverse transcriptase (Boehringer Mannheim, Lewes, UK) at 42°C for 1 h. The enzyme was inactivated by 5-min incubation at 95°C and the reaction diluted fivefold with 50 mM KCl, 10 mM Tris-HCl pH 8.3, 1.5 mM Mg Cl₂, and 0.01% gelatin. Primer S1 was added with 2.5 U *Taq* DNA polymerase (Promega, Southampton, UK), followed by 30 cycles of 94°C, 1 min; 60°C, 1 min; and 72°C, 1 min. Sequencing of 10 templates per fragment was performed.

For genomic DNA templates primers S1 and A1 were used in addition to sense primer S2 (5'-CCTGAATTCATTAGGACACACT-3') and antisense primer A2 (5'-GCCAATTCAGTGTGCTCTAATGG-3') also incorporating 5' EcoRI restriction sites to facilitate subcloning into sequencing vectors. S2 and A2 are complementary to the CN35 insertion sequence. Amplification was performed as described for RNA with the reverse transcriptase step omitted, using the primers in all three possible combinations.

DNA Sequencing

Inserts of 2 kb or less from positive clones and PCR reaction products were subcloned into M13mp18 or mpl9 in both orientations for sequencing. Inserts >2 kb were first cloned into pUC18, restriction mapped, and overlapping restriction fragments subcloned into M13mp18 or mpl9 in both orientations. To obtain complete sequence from each subclone Exonuclease III deletions (Henikof 1984; Pharmacia) were prepared from pZ523 (5 Prime-3 Prime, Cramlington) column-purified M13RF DNA. Sequencing was by the dideoxynucleotide chain termination method (Sanger et al., 1977) using Sequenase V2.0 (United States Biochemical Corp., Cambridge Bioscience, Cambridge, UK).

Protein and DNA Sequence Analysis

Sequence analysis was carried out using Microgenie software. Release 13 of the Swiss-prot protein sequence databank and release 21 of the EMBL databank library were searched using FASTA, a modification of FASTP (Lipman and Pearson, 1985). The databases and supporting programs have been installed in the Southampton University IBM3090-150VF with the aid of the Southampton University Computing Project Fund.

Results

Screening

The λ gt11 DNA expression library prepared from bovine nasal epidermal poly(A)⁺ RNA was screened with a mixture of two rabbit polyclonal antisera against dg2 and 3. 29 positive clones were identified from $\sim 1.8 \times 10^5$ pfu's screened. The fusion proteins of two of these clones, CN4 and CN29, reacted with the monoclonal antibody, 52-3D, which specifically recognizes the cytoplasmic domains of dg2 and 3 (Parrish et al., 1990).

cDNA inserts from CN4 and CN29 were sized at 0.59 and 2.14 kb by agarose gel electrophoresis, restriction mapped (Fig. 1), and shown to cross-hybridize with each other. The complete insert of CN4 and the 0.9-kb HincII-EcoRI fragment of CN29, which does not contain any CN4 sequences

(Fig. 1), were found to hybridize with a single band of 5.8-kb on Northern blots of bovine nasal epidermal poly(A)⁺ RNA (Fig. 2 a). Single bands were obtained with CN4 and CN29 on Southern blots of MDBK genomic DNA (Fig. 2 b).

The β -galactosidase fusion protein from a lysogen of CN4 was used to raise a guinea pig antiserum. The antiserum reacted specifically with dg2 and 3 on Western blots of whole desmosomes from bovine nasal epidermis (Fig. 2 c). Binding to dg2 and 3 could be blocked by preincubation of the antiserum with gel purified dg2 and 3, but not with β -galactosidase.

From the above data we concluded that CN4 and CN29 were partial clones encoding at least parts of the cytoplasmic domains of dg2 and/or 3.

To obtain further clones, the cDNA insert from CN4 was used to rescreen a further 1.2×10^5 pfu of the λ gt11 library. 17 more positive clones were identified. Clone inserts were sized by EcoRI digestion of miniprep DNA and the 5.0-kb insert of clone CN35 was selected for further study. The insert from CN35 was subcloned into pUC18 and a detailed restriction map prepared. The map of the 3' half of CN35 was found to be identical with the map of CN29 (Fig. 1).

Sequencing of cDNA Clones

Overlapping restriction fragments of the CN35 insert were subcloned into M13mp18 or mpl9 and sequenced (Fig. 1). Complete sequence on both strands was obtained from the 5' end of CN35 to the PstI site at base 2,953. This includes the entire coding region of the clone. Sequence from the PstI site at base 2,953 to the 3' end of CN35 was obtained largely on one strand and is identical to the corresponding 3' sequence of CN29 (see below).

Sequence analysis showed that clone CN35 contains a single open reading frame of 2,517 bases from base 319 to

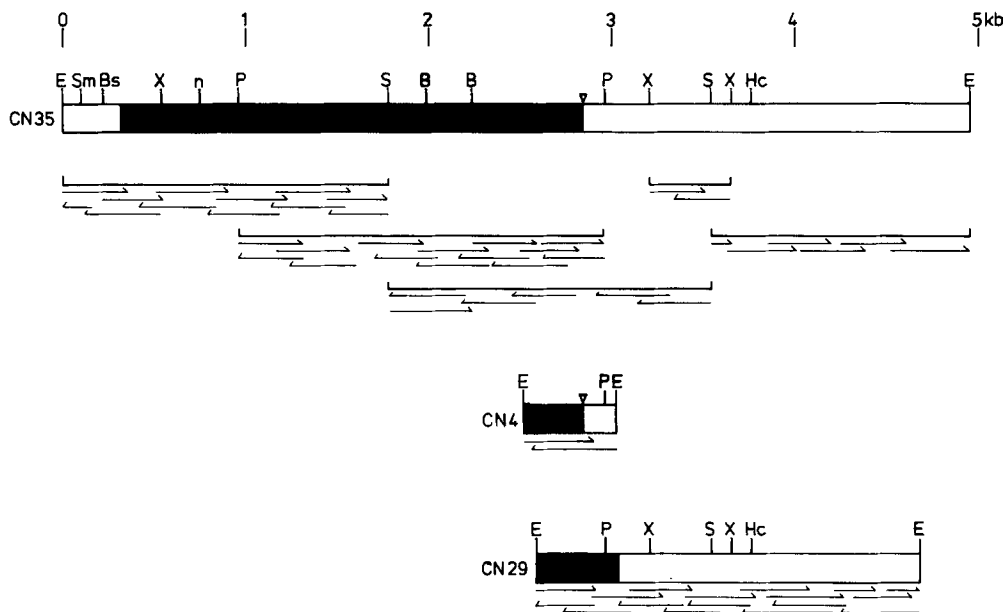


Figure 1. Restriction map and sequencing strategy for dg2 and 3 cDNA clones. Restriction maps for clones CN35, CN29, and CN4 are shown. The open reading frame is shown by black boxes, non-coding sequence by open boxes. Overlapping restriction fragments subcloned from CN35 for sequencing are shown by thin lines. The length and direction of sequence obtained from exonuclease III deletions of each clone is shown by arrows. The position of a 46-bp sequence found in CN4 and CN35 but not in CN29 is shown (∇). The 46bp sequence introduces an in-frame stop codon and shifts the downstream sequence out of the reading frame. The absence of the sequence in CN29 extends the open read-

ing frame by 195bp. The position of the N-terminus of mature dg3 is marked on CN35 (n). E = EcoRI, Sm = SmaI, Bs = Bst EII, X = XbaI, P = PstI, S = SstI, B = BglII, and Hc = HincII.

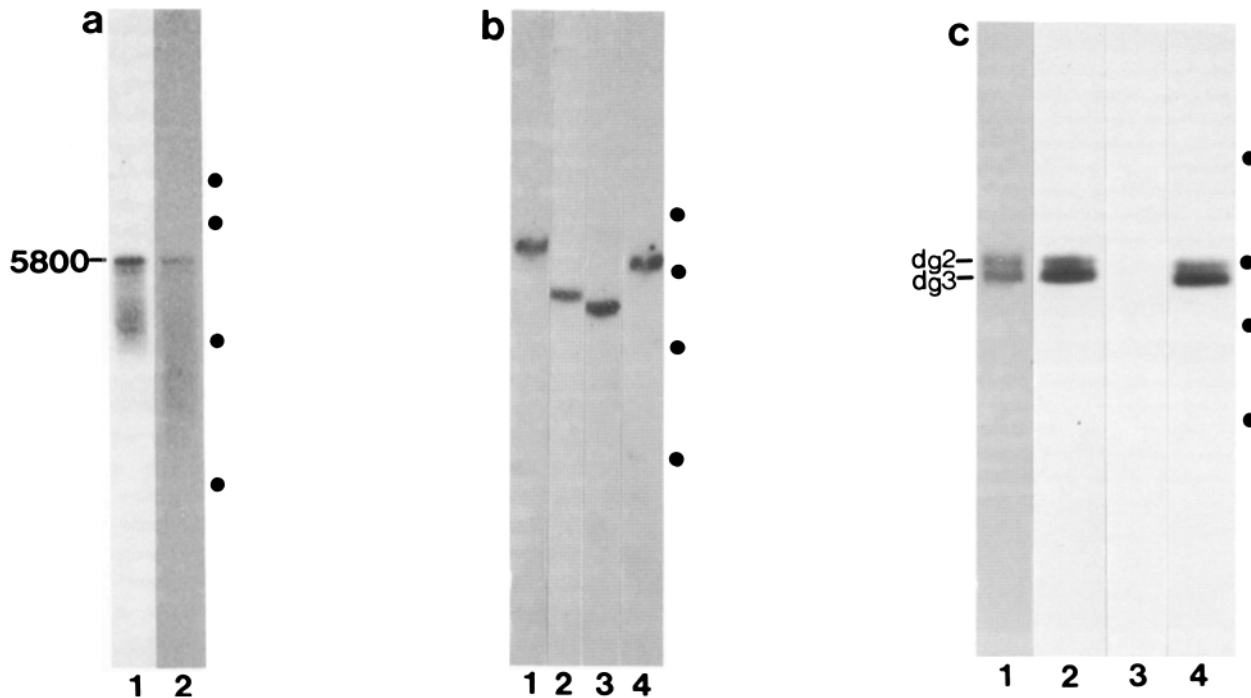


Figure 2. (a) Northern blot analysis. Poly (A)⁺ RNA from bovine nasal epidermis (1 μ g per lane) was probed with the 900-bp HincII-EcoRI fragment of CN29, exclusive of CN4 sequence (lane 1) or with the complete insert of CN4 (lane 2). Both probes hybridized to a single band of 5.8 kb. Markers are (Gibco Ltd., Paisley, UK) 9.5, 7.5, 4.5, and 2.4 kb. (b) Southern blot analysis. MDBK genomic DNA digested with EcoRV (lane 1), EcoRI (lane 2), BglII (lane 3), and BamHI (lane 4) was hybridized to the complete cDNA insert of CN29. The presence of single bands in these lanes suggests the presence of a single gene. An identical pattern was obtained using CN4 as probe. Markers, HindIII digest of λ cI857 DNA, are 23.1, 9.4, 6.5, and 4.3 kb. (c) Antibody reactivities on Western blots of whole desmosomes from bovine nasal epidermis. Lane 1, mouse monoclonal antibody 52-3D reacting with dg2 and 3. This antibody recognized the fusion proteins encoded by CN4 and CN29. Lane 2, fusion protein antiserum (1/10,000) reacting with dg2 and 3. Lane 3, fusion protein antiserum previously adsorbed with SDS-polyacrylamide gel purified dg2/3 showing specific inhibition of dg2/3 reactivity. Lane 4, fusion protein antiserum adsorbed with SDS-solubilized β -galactosidase showing that specific dg2/3 reactivity remains. Molecular weight standards (Sigma Chemical Co., Poole, UK) 205,000, 116,000, 97,000, and 67,000.

2,835, encoding a polypeptide of 839 amino acids (Fig. 3). A match with the NH₂-terminal sequence of dg2/3 determined by protein sequencing (Holton et al., 1990) is found at residue 133–156 of the polypeptide (Fig. 3). In addition the NH₂ terminus of a V8 protease fragment of dg2 obtained as described by Holton et al. (1990) also showed a sequence match between residue 252 and 265 (Fig. 3). The 3' noncoding region of CN35 stretches for 2,100 bases and contains four polyadenylation signals but a poly(A) tail was not found. Comparing this clone with the 5.8-kb mRNA indicates that up to 0.8 kb of sequence is missing from the noncoding regions.

The ATG at position 319 is the first Met codon in the sequence and is thought to be the initiation codon even though the surrounding sequence conforms poorly to the consensus initiation sequence of Kozak (1987). The first 28 amino acids of the open reading frame are predominantly hydrophobic and may represent a signal peptide (Fig. 3). Amino acids 22–29 conform to the signal peptide cleavage site proposed by van Heijne (1983). The signal peptide is followed by a 104 amino acid precursor sequence. Identifying the NH₂-terminal sequence of mature dg2/3 at position 133 of the CN35 polypeptide indicates that the mature protein consists of 707 residues, with a deduced molecular weight of 79,081. This

Figure 3. cDNA and derived amino acid sequence for dg3 and the cytoplasmic domain of dg2. The complete DNA and derived amino acid sequence of clone CN35 is shown. The 5' and 3' ends of clone CN4 and CN29 are marked beneath the sequence. The 46-bp sequence found only in CN35 and CN4 is boxed. In-frame stop codons are marked by asterisks. The putative signal peptide and transmembrane domain are doubly underlined. Sequences matching the NH₂-terminal sequence of mature dg2 and 3 (Holton et al., 1990) and of a V8 protease fragment of dg2 are thinly underlined. Potential N-linked glycosylation sites are marked by a solid triangle. The deduced amino acid sequence for the cytoplasmic domain of dg2, from clone CN29, is identical with that of dg3 (CN35, CN4) from residues 738–828. The 46-bp sequence in CN35 and CN4 introduces 11 amino acids of coding sequence unique to dg3 followed by an in-frame stop codon and a reading frame shift which renders the downstream sequence noncoding. The reading frame of dg2 (CN29) lacks this sequence and continues for a further 195 bases (base 2,849–3,043 of the CN35 sequence) encoding 65 amino acids unique to dg2. PCR primers S1, S2, A1, and A2 used in analysis of the alternative splicing of this 46-bp sequence are arrowed. A consensus serine phosphorylation site in dg2 is marked by solid circles. The 3' noncoding sequence of CN35 is 2,100 bp long and contains four polyadenylation signals (*underlined*). The first of these signals surrounds the TAA stop codon that terminates the reading frame encoding the cytoplasmic tail of dg2. These sequence data are available from EMBL/GenBank/DBJ under accession numbers X56966/7/8.

GCCAGTGGCTGC 12
GTTTGGGGTGT... 111
GAGAAAGCAGT... 210
GGTTACCCTGCC... 309
CCACTTCTGAG... 408
M A V A S A A P G S I F W K Q L L F S L L V L I L F C D A C 30
CAGAAAATTTCT... 507
Q K I S L Q V P S H L R A E A L V G K V N L K E C L Q S A S L I L S 63
TCAAGTGACTCT... 606
S S D P D F R I L E D S I Y T T H D L V L S S G K S F S I L L S 96
GACAGTCAGGGAC... 705
D S Q G Q G Q K E I E I L E A G G K K V P K R H M K D A V L R R 129
ACGAAGAGCGGAT... 804
T K R R W A P I P C S L M E N S L G P F P Q H V Q Q V Q S D A Q 162
AACTACACCATCT... 903
N Y T I F Y S I S G P G V D K E P F N L F F I E K D T G D I F C T 195
AGAAGCATAGACC... 1002
R S I D R E Q Y Q E F P I Y A Y A T T A D G Y A P E Y P L P L V F 228
AAAGTTGAAGATA... 1101
K V E D N D N A P Y F E N K L T V P E N C R T G T S V G K 261
GTGACTGCAATAG... 1200
V T A I D L D E P D T L H T R L K Y K I L Q Q I P N N P R H F T V 294
CATCCAGACAGGT... 1299
H P D C T G T T T P L L D R E K C D T Y K L I M E V R D M G G G G C 327
CAACCTTTGGTTA... 1398
Q P F G L F N T G T I T I S L E D E N D N A P Y F T E T S Y T V E 360
GTAGAAGAAAACA... 1497
V E E N R I D V E I L R M A V H D H D L P N T P H S R A V Y Q I L 393
CAGGAAAAGAAA... 1596
Q G N E N G T F K I S T A D P N T N E A V L C V V K P L N Y E V N R 426
CAAGTTGTTTTGA... 1695
Q V V L Q I G V L N E A Q F A K A V N S K T T T T M C T T V V T V 459
AAAGTTAAGACCA... 1794
K V K D H G E P E C P P V K V I Q S E D C L P T G T E L L G Y 492
AAAGCAGTGGACC... 1893
K A V D P E R G T G E G L R Y K K I Q D E D N W F E I N E Y T G D 525
TTGAAAAGTAAA... 1992
L K T V K V R E L D R E S T F V K N N Q Y N V S V I A F D A D G R S C 558
ACTGGAACGTTAG... 2091
T G T L V V F L E D K N D H P P Q I K Q E E L T I C R H D K D Y V 591
GTCCTCGAACCTA... 2190
V L E P T D Q D G P D N G P P F Q F I L D N S A S K L W T V E T R 624
GATGGTAAACTGC... 2289
D G K L R A I L R G R Q I L R I K D R H G A C S A T 657
CACATATTACAGT... 2388
H I L P V R V C D C T I P S E C R M P S K L S R E A A L A N V F L 690
GGAAAATGGGCA... 2487
G K W A I L A M V L G S V L L L C I L F T C F C V T V K K T V K G 723
TGTTTTCCAGAAG... 2586
C F P L E D V A Q Q N L I V S I N T E G P G E E V M D A N I R L P T Q 756
ACATCCAACGTT... 2685
T S N V C D T S I S I S G T L G G Q G V K T Q Q S F E M V K G G Y T 789
TTGGATGCCAAC... 2784
L D A N K G G G H Q T L E S V K G V T D T G R Y T Y S W H A F T C 822
S2 A2
CAACCTCGGCTT... 2883
Q P R L G E E S I R G H T L V K N *** K V Y L C G Q A D E E H C 839
S1 A1
GCTCTGCGAAGC... 2982
L C E D Y V R S Y S Y S V G K G S V G C C S D R Q E E E G L
TGACTTCTAGAT... 3081
D F L D H L E P K F I L A K T C V K K ***
ATATTGAATGTA... 3180
C T G A A T A A C T T A G C A T C A G C A T C G T A A A G T T T G T T A T G G A G G T A A C T C C A G T T G T A T A G A T A A G G C T C C C T A A A T T C T C T T G T 3180
TGCGATAACTT... 3279
G C T T T C C A C C A C T T C T G T T G A A A A C T A C G T T A C T T G G T A T G C T G G A A A G T A T G G A G G A T T A G G A T T G G A G A T T A A A G A A A A C A C A 3378
TTGTTATTTAA... 3477
T T G T A T T T A A A T G A A A T G A A A T T T T C T C A G A A T C T T G G G G A C A A T A T G C T T T A G A A G T A T T G C T T T G G T G C T A C A G A A T T G C T T C T T A 3477
CCTGAATGGCT... 3576
C C T G A A T G G C T T G T G A A G A A T T T C T A T G C A T T C T G A T G C A A T T T G A T T G G T G A A T A A G A G C T A C T T T C A T T G T G T T T C A C C T G T G 3576
GAGGTGCTCC... 3675
G A G G T G C T C C C A G C T T G A T C C T G G G A T C T G T T G A T G T G G A T G T G T A C A G G T T C C A C A G T T A G A A T T G T T G G A A G A T T G G A C C A G A A G A G 3675
GCTCACAGAA... 3774
G C T C A C A G A A T T T G A G T C A T T G A G A T T T T A A A A A A A A G C A T T C C A T G A A A G A A T T G C A G C C T A A T G C C C A T G G G T T G A C T G A A A C A A A A T C 3774
GTTTTGTTGCC... 3873
G T T T T G T T G C C C A C A T G A T G A T T A T T A C T T T T A A A A T G T A A A T A C C T T G G C A A G G A C A T G T T C T C T A G T T T G A C A C C A T C C C A C C A T C C T C 3873
AAGCATTTGAG... 3972
A A G C A T T T G A G T A A C T A G A C C A G T A G G C A T T A G G C T T A C A A C C C T G A T C G C C C T T G A A T G T G C T A A C A C T G A A T G A C T A C C A T T C C A A T G G A T 3972
CTCAGGCTAC... 4071
C T C A G G C T A C A T T G G T A A T C A C T T C A G A A A T G T T A A A T A A G A A A A G T G T C T G G C A T A T A T G T T A A A A A C T G A A A T G T T A T T A A C T A A A T C 4071
TTACCAAAC... 4170
T T A C C A A C A T T T G T G A G T A G C T G G A T A A A G A A T G A A A C A T G A G T T A A T G A G C T A T T T T A A T A C A T A G C A T T T A A T A A G A T A T T C A T T T T A 4170
TTACTTTAGA... 4269
T T A C T T T A G A A A A A C C A A T A T T T A A A T A T T T T A A T T T T T G C C A A A T G C T A T A G A T T C A T A A A C A T T T G C A G T A T A T A C A A G G A C 4269
ATTTTACTCA... 4368
A T T T T A C T C A G A A A A A A C T T T T T C A C T T T T A G C T A C T T A C C T C A A A A A A A A T G A G T G A C T A A T T T C T T A A A T T T G T G A A T T A T T T C C 4368
TTCTTAGTC... 4467
T T C T T A G T C A C A G G A T T A A A A G A T T T C T T G A A T G A A A T T G A A T A G T A G C A G A A A A A A A G T A G C T T A G T A A A T A C A G T T T T A T T T T A 4467
AAAGTGCCTC... 4566
A A A G T G C C C T G A T T A T T T T A T G G C A T T G T C A G T A A A A T T A A T T T G G A G A G T A A T G C A G C T T C T G A A T A T C T G T A T C T G T A T C T G T A 4566
TTCTGACAA... 4665
T T C T G A C A A T G A T G C T A G T T T G A A A T T T A T A A T A T T A A A A T T T T G G A A A T A T A T C A T A A T A A T G T T C T A A G T A T A T T T T G A A C A T A G G T G C A G A A 4665
CN29
TGTTTGTGAG... 4764
G A A C A T A A G G T A C A A A T A A A G G C C A A A G G C T G A T T C T C A A A C C A A C A G A G A G A A G A C T T A A A A C T A T G C C A G A A T T A T T A G A A A G T G T A G G A A A T 4764
GAATTTGATA... 4935
G A A T T T G A T A A A G G A A T T G A C A T T T A A G A T G T T C A A A A T C A T T A A A T T G T T A A A T T G T A T C C A T 4935

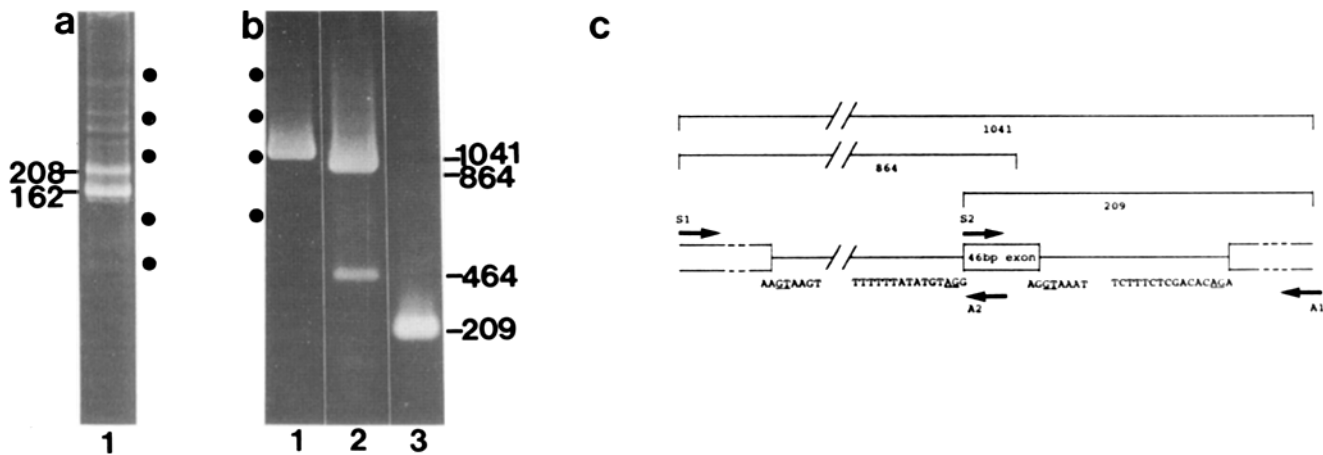


Figure 4. (a) PCR amplification to verify the existence of alternative splicing in the COOH-terminal domains of dg2 and dg3. Lane 1, reverse transcriptase-initiated PCR using primers S1 and A1 flanking the CN35 insertion sequence yielded two major bands of 208 and 162 bp. Sequencing of these showed the 208 fragment to have 46 bp of additional sequence corresponding exactly to the insertion sequence in CN35. Markers, HaeIII digest of pBR322, are 587, 434, 267, 124, and 64 bp. (b) PCR amplification of 0.25 μ g of whole genomic MDBK DNA. Fragments generated by S1 and A1 (lane 1), S1 and A2 (lane 2), and S2 and A1 (lane 3). Markers, EcoRI/HindIII digest of λ I857 DNA, are 2,027, 1,330, 983, and 564 bp. (c) Organization of amplified genomic DNA fragments. Sequencing of the products shown in b revealed that the 864-bp product in lane 2 overlapped with the 209-bp product in lane 3. Together they corresponded to the 1,041-bp product in lane 1. The 464-bp fragment in lane 2 was a nonspecific amplification product. The sequence contained 751- and 83-bp introns 5' and 3' to the 46-bp insertion, with consensus sequences at the splice junctions. The most 5' and 3' bases of the introns are underlined.

is significantly smaller than the relative molecular weights for dg2 and dg3, 115,000 and 107,000, respectively, determined by SDS-PAGE (Parrish et al., 1990). Such a molecular weight discrepancy is commonly found for transmembrane glycoproteins (e.g., L-CAM, Gallin et al., 1987).

Consistent with the known transmembrane nature of dg2/3 (Parrish et al., 1990) hydropathy plots (Hopp and Woods, 1981) identified a single region of 35 amino acids showing characteristics of a transmembrane sequence (Fig. 3). The extracellular domain of the mature CN35 polypeptide consists of 550 amino acids, containing three potential N-linked glycosylation sites (Fig. 3). The cytoplasmic domain of CN35 consists of 122 amino acids.

Cytoplasmic Domain Heterogeneity of dg2 and dg3

The cDNA insert from CN29 was sequenced on both strands (Fig. 1) and was found to lie within the cytoplasmic and 3' noncoding regions of CN35. CN29 lacks a 46-bp insertion found in CN35 (Fig. 3). This insertion in CN35 encodes a unique COOH terminus of 11 amino acids and contains an in-frame stop codon. The absence of this 46-bp insertion extends the open reading frame of CN29 to encode a COOH terminus of 65 amino acids not found in CN35. CN29 thus encodes a cytoplasmic domain 54 amino acids longer (6,126 molecular weight) than that encoded by CN35 and also contains a consensus serine phosphorylation site not found in the CN35 sequence (Fig. 3). The sequence of CN4 is identical to bases 2,497–3,016 of CN35, and includes the 46-bp insertion (Fig. 3).

Parrish et al. (1990) proposed a model for the cytoplasmic organization of dg2 and 3 in MDCK cells in which the cytoplasmic domain of dg2 is 4,000–5,000 M larger than that of dg3. Since the relative molecular weights of the bovine proteins are identical to those in MDCK cells the results of sequence analysis correspond well with this model, suggesting that CN29 encodes the cytoplasmic domain of dg2 whilst CN35 encodes the complete sequence of dg3.

Cytoplasmic Heterogeneity of dg2 and 3 Is Generated by Alternative Splicing

Southern blots suggested the presence of a single gene (Fig. 2) raising the possibility that the presence or absence of the 46-bp sequence might be due to alternative mRNA splicing.

To study this possibility amplification of DNA by the PCR was performed with cDNA synthesized from total RNA of bovine nasal epidermis using two oligonucleotide primers 80 bp 5' and 3' of the 46-bp sequence (S1 and A1 in Fig. 3). Products of 208 and 162 bp were obtained (Fig. 4 a). The identities of these were confirmed by sequencing and they were shown to correspond to the appropriate regions of CN35 and CN29, with the 208-bp product containing the 46-bp sequence. This suggests the presence of mRNAs coding for the two differing sequences. That only a single band of 5.8 kb was found on Northern blots is explained by the inability of the technique to resolve a difference of 46 bases.

To confirm the existence of the 46-bp sequence as a discrete exon, PCR was carried out on genomic DNA from MDBK cells using additional primers from within CN35 (S2 and A2 in Fig. 3). Using three different combinations of the primers, three fragments were generated of 1,041, 864, and 209 bp (Fig. 4 b). Sequence analysis of these fragments revealed the predicted exon boundaries and consensus donor and acceptor sites (Shapiro and Senepathy, 1987) showing that the 46-bp sequence was present as a single exon (Fig. 4 c). An additional fragment of 464 bp was a nonspecific amplification product. These results confirm that the differing structure of clones CN35 and CN29 are produced by alternative splicing of this exon.

Sequence Homologies with Cadherins and dgl

Extracellularly dg3 shows 39.4% identity with bovine N-cadherin but cytoplasmically only 23%. Table I shows percent identities between the extracellular domains of dg3, representative cadherins, and dgl. The extracellular domains of

Table 1. Percentage Amino Acid Identities of the Extracellular Domains of Representative Cadherins, dg3 and dg1

N-cad (bovine)		N-cad (chick)		P-cad (mouse)		E-cad (mouse)		L-CAM (chick)		dg3 (bovine)		dg1 (bovine)	
88.5		44.3											
44.9		47.5		53.4									
47.2		48.0		51.6		58.4							
47.4		39.3		36.3		35.3		34.4					
39.4		28.5		28.9		27.5		28.5		28.4			
30.0													

Percentage identity between extracellular domain sequences can be found by referring to the figure at the intersection of rows and columns corresponding to any pair of sequences. Bovine N-cadherin, Liaw et al. (1990); chick N-cadherin, Hatta et al. (1988); mouse P-cadherin, Nose et al. (1987); mouse E-cadherin, Nagafuchi et al. (1987); chick L-CAM, Gallin et al. (1987); bovine dg1, Koch et al. (1990).

dg3 and the cadherins share more identity than they do with dg1. Members of the cadherin family show greater similarity to each other than to either of the desmosomal glycoproteins.

More detailed comparison of the extracellular domains of dg3, N-cadherin (N-cad) and dg1 reveals several features of

interest. (a) The extracellular domains of dg3 and N-cad are very similar in size (dg3: 62,155 mol wt, 550 residues;

N-cad: 60,923 mol wt, 555 residues), whereas that of dg1 is significantly smaller (55,068 mol wt, 499 residues) (Fig. 5).

(b) For ease of comparison with the cadherins, all three ex-

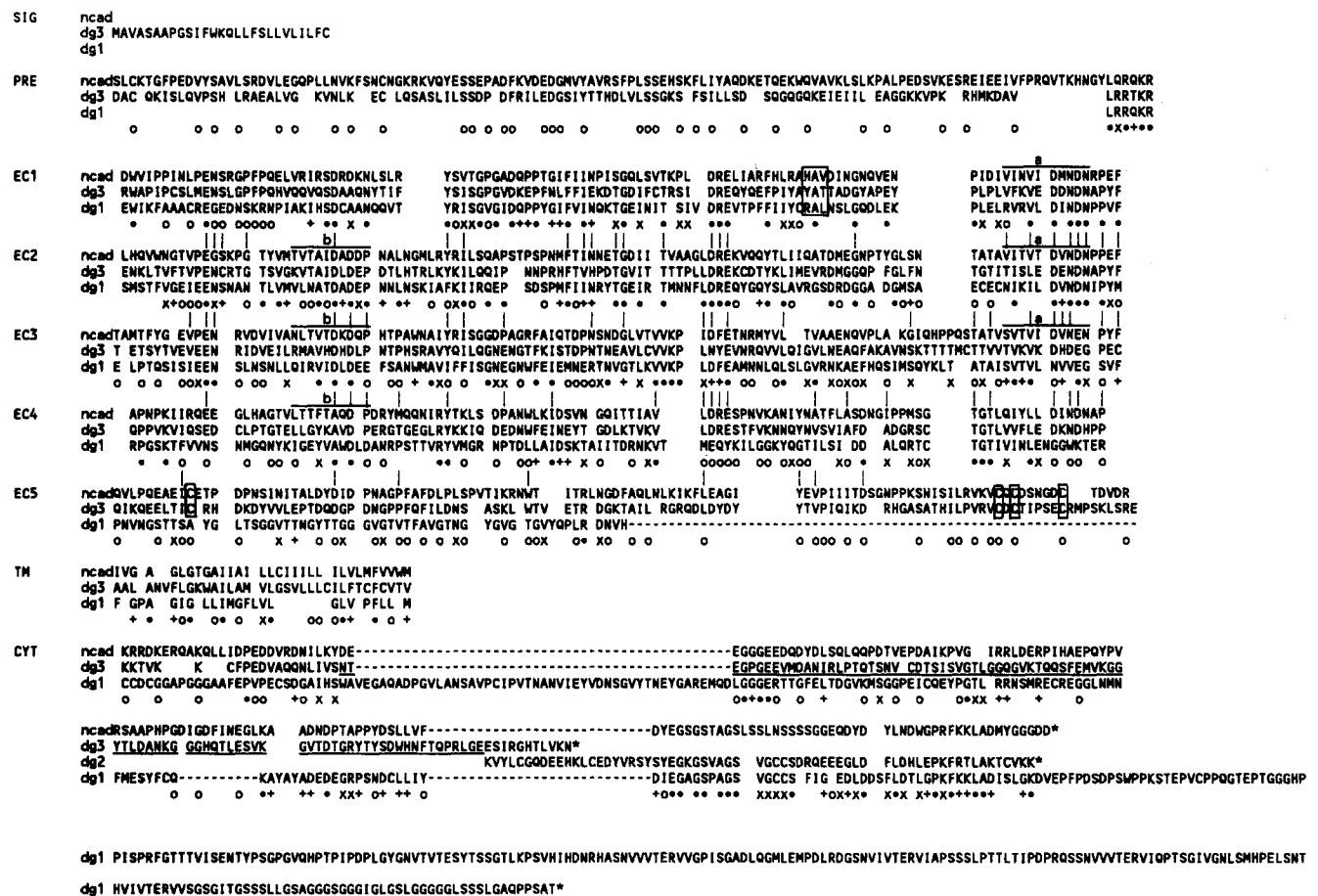


Figure 5. Comparison of amino acid sequences of bovine dg3 and dg2 cytoplasmic domain with bovine N-cadherin and bovine dg1. Sequences were divided into domains and aligned according to Hatta et al. (1988). SIG, signal peptide; PRE, precursor region; ECI-5, extracellular region divided into five homologous repeats; TM, transmembrane region; CYT, cytoplasmic region. Residues identical in dg2/3, N-cadherin and dg1 are marked by a solid circle; in N-cadherin and dg2/3 by an open circle; in dg2/3 and dg1 by an X; and in dg1 and N-cadherin by a plus sign. Residues repeated in ECI-5 are marked by vertical lines. The HAV sequence of N-cadherin and the corresponding regions in dg2/3 and dg1 are boxed. Four conserved cysteines in EC5 are also boxed. Putative Ca²⁺ binding sites are overlined and labeled a or b after Ringwald et al. (1987). Sequence of the cytoplasmic domain of dg3 that is known to be identical in dg2 is underlined. The 11 amino acid COOH terminus unique to dg3, and 65 amino acid COOH terminus unique to dg2 are shown individually. Large gaps in EC5 and the cytoplasmic domain introduced to optimize the alignment are dashed. Sequences were taken from Liaw et al. (1990) and Koch et al. (1990).

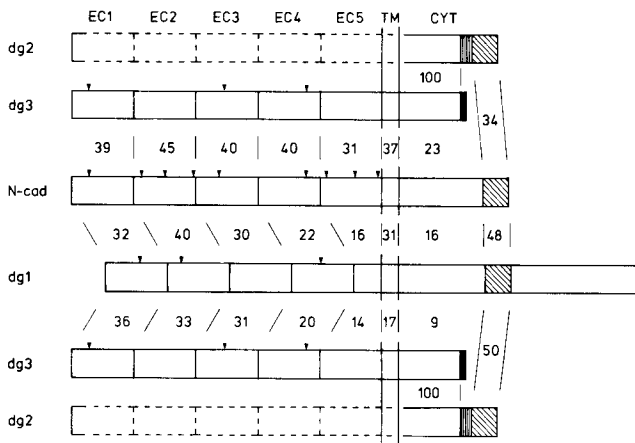


Figure 6. Representation of similarities between the sequences of bovine dg3, N-cadherin, and dg1. Each sequence is represented by open boxes drawn to scale. The NH₂ terminus of dg2 has been shown to be similar to that of dg3 by protein sequencing and antibody reactivity (Holton et al., 1990). The part of the dg2 sequence between the NH₂ terminus and the 5' end of CN29 is unknown and this part of the molecule is therefore indicated by broken lines. The 11 amino acid COOH terminus of dg3 is shown by the solid box. The 65 amino acid COOH terminus of dg2 is represented by the vertically and diagonally hatched boxes together. The cadherin-like region within the COOH terminus of dg2 and homologous regions in N-cadherin and dg1 are shown by the diagonally hatched boxes. Percentage identities are shown for each region. Potential N-linked glycosylation sites are marked with arrowheads.

tracellular domains have been divided into five homologous regions according to Hatta et al. (1988). Such a comparison reveals that the membrane proximal repeat (EC5) of dg1 is significantly smaller than the corresponding region of the other molecules (Figs. 5 and 6). (c) The greatest similarities are between the EC2 domains of dg3 and N-cad and the EC1 domains of dg3 and dg1 (Fig. 6). Within the dg3 molecule the greatest identity is between EC2 and EC4 (30.4%) and the least is between EC1 and EC3 (14.4%). (d) The HAV adhesion sequence of N-cad (Blaschuk et al., 1990; Nose et al., 1990) is represented by YAT in dg3 and RAL in dg1 (boxed in Fig. 5). (e) Several putative calcium binding sites (Ringwald et al., 1987) are conserved between the three molecules (Fig. 5, *overlined*). (f) Four cysteine residues are conserved between dg3 and N-cad in the membrane proximal region (EC5), but not in dg1 (Fig. 5). (g) dg3 has three consensus sites for N-linked glycosylation, two of which correspond with sites in N-cad, which has nine sites in total. None of the three putative glycosylation sites in dg1 correspond with those in dg3 or N-cad and one of them lies within the first putative calcium binding site in EC2 (Fig. 5).

Overall the cytoplasmic domain of dg3 and the partial cytoplasmic sequence for dg2 diverge significantly from the highly conserved cytoplasmic domains of cadherins. However, comparison of the sequences of dg3, dg2, N-cad, and dg1 reveals several features of interest. (a) The cytoplasmic domain of dg3 is considerably shorter than that of N-cad (dg3: 13,247 mol wt, 122 residues; N-cad: 17,690 mol wt, 160 residues). If the cytoplasmic, membrane-proximal sequence of dg2 is the same as dg3, the cytoplasmic domain of dg2 would be slightly larger than that of N-cad (dg2: 19,373 mol wt, 176 residues). The cytoplasmic domain of dg1 is very much larger than the other polypeptides (44,763

mol wt, 435 residues). (b) The dg3 sequence has a short, charged region immediately inside the membrane that may represent a stop signal for protein translocation into the endoplasmic reticulum (Sabatini et al., 1982) and which resembles the same region in N-cad. This contrasts with the cysteine-rich membrane-proximal region of dg1. (c) The cytoplasmic domain of N-cad has no cysteine residues. There are two in dg3, an additional five in the COOH-terminal tail of dg2 and 12 in the cytoplasmic domain of dg1. (d) The most highly conserved region lies within the COOH-terminal 44 amino acids of dg2. It shows 50% identity with dg1 and 34% identity with N-cad. Within this region the 10 amino acid motif YEGXGSXAGS is conserved throughout the cadherins and dg1 (Fig. 5). In dg2 this sequence is immediately preceded by a consensus serine phosphorylation site and followed by the sequence VGCCS which is identical in dg1. (e) dg3, dg2, and N-cadherin share a common motif of six amino acids, EGXGEE, conserved throughout the cadherins. This motif is also found in the COOH-terminal region of α -tubulin (Sullivan, 1988). (f) Both dg3 and N-cad lack an insertion of 47 amino acids found in dg1. dg1 has a large, unique COOH-terminal region of 236 amino acids.

Other points of interest are that the precursor sequence of dg3 shows 27% similarity to the precursor sequence of N-cadherin. The six amino acids immediately preceding the mature NH₂ termini of these proteins are highly conserved in all cadherins. No significant homologies with other sequences were found in searches of the databanks with the complete sequence of dg3 and the cytoplasmic sequence of dg2.

Discussion

Desmosomes are unique punctate membrane-junctional domains. The intercellular adhesive material (desmoglea) is highly ordered, showing the presence of a midline joined to the cell plasma membranes by cross-bridges (Rayns et al., 1969). Cytoplasmically there is an electron-dense plaque close to the membrane. This serves as a site for association of intermediate filaments (IFs). However, the IFs do not attach to the plaque directly but via filaments which extend into the cytoplasm from the inner face of the plaque (Kelly, 1966). Transmembrane desmosomal glycoproteins therefore require extracellular domains that participate in cell-cell adhesion and cytoplasmic domains that participate in plaque formation involving linkage, directly or indirectly, to IFs.

The results of our present study confirm that dg3 is a member of the cadherin family of Ca²⁺-dependent cell-cell adhesion molecules, supporting our previous evidence (Holton et al., 1990). The deduced amino acid sequence of dg3 is in excellent accord with previous cell biological and biochemical observations. Thus the resemblance to cadherins extracellularly supports the view that dg3 has a primary role in calcium-dependent cell-cell adhesion in desmosomal junctions (Cowin et al., 1984b; Gorbsky and Steinberg, 1981; Matthey et al., 1987; Garrod et al., 1990; Holton et al., 1990). The overall structure of an asymmetrically distributed glycoprotein with a single transmembrane domain supports the model of Parrish et al. (1990).

Divergence in sequence homology from the cadherins in the dg3 cytoplasmic domain is significant because this part of the molecule is the most highly conserved region among other members of the family. We suggest that the structure

of the dg3 cytoplasmic domain is specialized for participation in plaque formation.

Our present data do not reveal the entire structure of the larger of these glycoproteins, dg2. However, we demonstrate a second COOH-terminal structure coded for by an alternatively spliced message derived from the same gene as dg3. This structure is that of a longer cytoplasmic region containing a consensus serine phosphorylation site. These features are consistent with the structure predicted for the COOH-terminal of dg2 by Parrish et al. (1990).

While alternative splicing is well documented among the Ig-like cell-cell adhesion molecules (Seed, 1987; Dustin et al., 1987; Lai et al., 1987; Walsh and Dickson, 1989; Cunningham and Edelman, 1990) this is the first reported case of alternative splicing in a cadherin-like adhesion molecule. Among the Ig-like adhesion molecules, the example most similar to that described here is found in the myelin-associated glycoprotein where splicing in of a 45-bp exon containing an in-frame stop codon gives rise to a protein isoform with a shorter cytoplasmic domain and a unique COOH terminus (Lai et al., 1987). The function of alternative splicing in adhesion molecules is not fully understood. We suggest that the alternative splicing of dg2 and 3 cytoplasmic domains has some specific role in relation to desmosomal plaque formation.

We have also shown that the N-terminus of dg2 is cadherin-like and resembles that of dg3 (Holton et al., 1990). It may be that dg2 is identical to dg3 except for this heterogeneity in the cytoplasmic domain, although further alternative splicing may generate additional heterogeneity extracellularly. Up to four isoforms of dg2 and 3 have been reported in various tissues and cell types (Cohen et al., 1983; Suhrbier and Garrod, 1986; Parrish et al., 1990). It may be that the other isoforms are also generated by alternative splicing although we have not found evidence for this here.

Desmosomal adhesion is homotypic in the sense that it occurs between two half desmosomes contributed by adjacent cells. Cadherin-mediated adhesion is homotypic at the molecular level (reviewed by Takeichi, 1990). Adhesive binding in the cadherins is believed to involve the amino acid triplet HAV (Fig. 5) (Blaschuk et al., 1990; Nose et al., 1990). The corresponding sequence in dg3 is YAT and in dg1 is RAL. These may be candidates for adhesion binding sites in the desmosomal glycoproteins. Even single amino acid changes in the region adjacent to the HAV site of cadherins affects binding specificity (Nose et al., 1990). Since the sequences of the desmosomal glycoproteins are quite dissimilar in this region, it is possible that they participate in specific homotypic binding. If the structure of the extracellular domain of dg2 proves to be identical to that of dg3, adhesive interaction between these two molecules might be anticipated unless specificity is also influenced by differences in the cytoplasmic domains. Dg3 contains neither an HAV nor an RAL sequence and dg1 contains neither HAV nor YAT.

The extracellular domain of dg3 is approximately the same size as those of the cadherins and retains some of their structural features. The cysteines in the membrane proximal region are conserved suggesting that they have an important structural role. It is interesting that dg1 differs significantly in the membrane proximal region in that about half of the EC5 region is absent. Furthermore, the remaining part of this region in dg1 consists of three repeats of a hydrophobic

sequence (Koch et al., 1990), lacks cysteine and is quite unlike dg3 and the cadherins. It is possible that this region has a different function in dg1, possibly in membrane interactions or lateral molecular interactions.

All three desmosomal glycoproteins have been shown to bind Ca^{2+} (Mattey et al., 1987; Steinberg et al., 1987) and putative Ca^{2+} binding sites identified in uvomorulin (Ringwald et al., 1987) are conserved in both dg1 and dg3. The role of Ca^{2+} in cadherin mediated adhesion is unclear, but certainly of central importance. Ca^{2+} binding may be required for dg3 to adopt an active conformation (Mattey et al., 1987, 1990), or Ca^{2+} may be directly involved in adhesion (Steinberg, 1958; Pethica, 1961).

Three consensus N-linked glycosylation sites are found in dg3, compared to nine in N-cadherin. Two of the three correspond to sites in N-cadherin in EC1 and EC4. A reduction in the relative molecular mass of dg2 and 3 by $\sim 10,000$ D in tunicamycin treated MDCK cells suggests three to four N-linked glycan side chains in dg2 and 3 (Penn et al., 1987). This is in good agreement with the three possible glycosylation sites of bovine dg3. The role of glycosylation in desmosome assembly is unclear. Inhibition of N-linked glycosylation with tunicamycin does not inhibit desmosome formation (Overton, 1982; Mattey et al., 1987). In embryonic cells the attachment of carbohydrate appears to protect the glycoproteins from proteolysis (Overton, 1982).

The cytoplasmic domain of dg3 is considerably shorter than that of N-cadherin or dg1 and shows little similarity to either. The motif EGXGEE conserved in dg3, dg2, and the cadherins was also found in the COOH terminus of α -tubulin. A specific role for the region of α -tubulin in vivo is not known but it may have a role in microtubule assembly (Sullivan, 1988). The cytoplasmic domain of dg2 is terminated by a 65 amino acid tail which shows homology to both N-cadherin and dg1. The tail is characterized by the motifs YEGXGSAGS, conserved in the cadherins and dg1, and VGCCS, found only in dg2 and dg1.

Cadherins and dg2 and 3 frequently occur in the same cell types but occupy different membrane locations. Although the presence of small amounts of cadherins has been reported in desmosomes in one situation (Jones, 1988), cadherins are normally located in nonjunctional lateral membrane or in the actin-associated zonula adherens junctions (Boller et al., 1985). These distinct locations are likely to be determined by association with distinct classes of cytoplasmic and cytoskeletal proteins. The very high conservation of the cadherin cytoplasmic domain ($>55\%$ identity) suggests a common role for this domain in the different cadherin subclasses. Recent work has shown that this may be the anchorage of cadherins to actin filaments via interactions with at least three accessory proteins, the catenins (Ozawa et al., 1989, 1990).

The cytoplasmic domain sequences of dg2 and 3 have overall only 24 and 23% identity with the cytoplasmic domain of bovine N-cadherin. This individuality of structure presumably reflects specialization for the molecular interactions involved in desmosome plaque formation, association with the unique desmosomal proteins (desmosomal proteins 1 and 2 [desmoplakins] desmocollin, desmoyokin, and B6B, reviewed by Schwartz et al., 1990) and linkage (probably indirect) to intermediate filaments. Sequence homologies between the desmosomal glycoproteins and the cadherins may suggest shared structural features or binding specificities.

Desmosomal protein 3 (plakoglobin) may interact with the cytoplasmic domains of both classes of glycoproteins since it occurs in both desmosomes and zonula adherens junctions (Docherty et al., 1984; Cowin et al., 1986; Gorbisky et al., 1985; Miller et al., 1987).

The presence of a consensus serine phosphorylation site in the cytoplasmic tail of dg2 suggests that its interactions may be modulated by phosphorylation. dg2 becomes serine phosphorylated during desmosome assembly in MDCK cells but dg3 does not (Parrish et al., 1990). It is possible that in the MDCK glycoproteins, as in the bovine proteins described here, dg2 possesses a phosphorylation site that is absent from dg3 and that this accounts for the observed difference in phosphorylation.

The desmosomal glycoproteins are clearly members of the cadherin family of cell adhesion molecules but diverge significantly from the other cadherins in sequence homology. This demonstrates that the cadherin family is more diverse than previously believed.

We thank Dr. D. A. Rees and his colleagues at the National Institutes for Medical Research and Drs. J. G. Williams, and S. D. Gregory for advice in the early stages of this project. We thank Dr. Elaine Parrish for valuable discussion and also Bridget Warland for typing the manuscript.

This work was supported by the Cancer Research Campaign and the Agricultural and Food Research Council. Antisera to dg2 and 3 were raised by Dr. Derek Matthey. Oligonucleotide primers were synthesized by Dr. Mark Pickett of the Department of Microbiology, University of Southampton. NH₂-terminal protein sequence from the V8 protease fragment of dg2 was obtained by Dr. Kathryn Lilley of Leicester University.

Received for publication 4 December 1990 and in revised form 14 January 1991.

References

Benton, W. D., and R. W. Davis. 1977. Screening λ gt11 recombinant clones by hybridisation to single plaques in situ. *Science (Wash. DC)*. 196:180-182.

Blaschuk, O. W., R. Sullivan, S. David, and Y. Pouliot. 1990. Identification of a cadherin cell adhesion recognition sequence. *Dev. Biol.* 139:227-229.

Boller, K., D. Vestweber, and R. Kemler. 1985. Cell adhesion molecule uvomorulin is localized in the intermediate junctions of adult intestinal epithelial cells. *J. Cell Biol.* 100:327-332.

Bologna, M., R. Allen, and R. Dulbecco. 1986. Organization of cyokeratin bundles by desmosome in rat mammary cells. *J. Cell Biol.* 102:560-567.

Cohen, S. M., G. Gorbisky, and M. S. Steinberg. 1983. Immunohistochemical characterisation of related families of glycoproteins in desmosomes. *J. Biol. Chem.* 258:2621-2627.

Cowin, P., and D. R. Garrod. 1983. Antibodies to epithelial desmosomes show wide tissue and species cross reactivity. *Nature (Lond.)*. 302:148-150.

Cowin, P., D. L. Matthey, and D. R. Garrod. 1984a. Distribution of desmosomal components in the tissues of vertebrates, studied by fluorescent antibody staining. *J. Cell Sci.* 66:119-132.

Cowin, P., D. L. Matthey, and D. R. Garrod. 1984b. Identification of desmosomal surface components (desmocollins) and inhibition of desmosome formation by specific FAB'. *J. Cell Sci.* 70:41-60.

Cowin, P., H.-P. Kapprell, W. W. Franke, J. Tamkun, and R. O. Hynes. 1986. Plakoglobin: a protein common to different kinds of intercellular adhering junctions. *Cell*. 46:1063-1073.

Cunningham, B. A., and G. M. Edelman. 1990. Structure, expression and cell surface modulation of cell adhesion molecules. In *Morphoregulatory Molecules*. G. M. Edelman, B. A. Cunningham, and J. -P. Thiery, editors. John Wiley & Sons, New York. 9-40.

Docherty, R. J., J. G. Edwards, D. R. Garrod, and D. L. Matthey. 1984. Chick embryonic pigmented retina is one of the group of epithelioid tissues that lack cyokeratins and desmosomes and have intermediate filaments composed of vimentin. *J. Cell Sci.* 71:61-74.

Dustin, M. L., P. Selvaraj, R. J. Mattaliano, and T. A. Springer. 1987. Anchoring mechanisms for LFA-3 cell adhesion glycoprotein at membrane surface. *Nature (Lond.)*. 329:846-848.

Gallin, W. J., B. C. Sorkin, G. M. Edelman, and B. A. Cunningham. 1987. Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule L-CAM. *Proc. Natl. Acad. Sci. USA*. 84:2808-2812.

Garrod, D. R., E. P. Parrish, D. L. Matthey, J. E. Marston, H. R. Measures,

and M. J. Vilela. 1990. Desmosomes. In *Morphoregulatory Molecules*. G. M. Edelman, B. A. Cunningham, and J. -P. Thiery, editors. John Wiley & Sons, Chichester. 315-339.

Gorbisky, G., and M. S. Steinberg. 1981. Isolation of the intercellular glycoproteins of desmosomes. *J. Cell Biol.* 90:243-248.

Gorbisky, G., S. M. Cohen, H. Shida, G. J. Guidice, and M. S. Steinberg. 1985. Isolation of the non-glycosylated proteins of desmosomes and immunolocalisation of a third plaque protein: desmoplakin III. *Proc. Natl. Acad. Sci. USA*. 82:810-814.

Green, K. J., R. D. Goldman, and R. L. Chisholm. 1988. Isolation of cDNAs encoding desmosomal plaque proteins: evidence that bovine desmoplakins I and II are derived from two mRNAs and a single gene. *Proc. Natl. Acad. Sci. USA*. 85:2613-2617.

Hatta, K., A. Nose, A. Nagafuchi, and M. Takeichi. 1988. Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion molecule: its identity in the cadherin gene family. *J. Cell Biol.* 106:873-881.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene*. 28:351-359.

Hennings, H., and K. A. Holbrook. 1983. Calcium regulation of cell-cell contact and differentiation of epidermal cells in culture. *Exp. Cell Res.* 143:127-142.

Holton, J. L., T. P. Kenny, P. K. Legan, J. E. Collins, J. N. Keen, R. Sharma, and D. R. Garrod. 1990. Desmosomal glycoprotein 2 and 3 (desmocollins) show N-terminal similarity to calcium-dependent cell-cell adhesion molecules. *J. Cell Sci.* 97:239-246.

Hopp, T. P., and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA*. 78:3824-3828.

Jacobson, A. 1987. Purification and fractionation of poly (A)⁺ RNA. *Methods Enzymol.* 152:254-261.

Jones, J. C. R. 1988. Characterisation of a 125K glycoprotein associated with bovine epithelial desmosomes. *J. Cell Sci.* 89:207-216.

Kapprell, H.-P., P. Cowin, W. W. Franke, H. Postingl, and H. J. Opferkuch. 1985. Biochemical characterisation of desmosomal proteins isolated from bovine muzzle epidermis: amino acid and carbohydrate composition. *Eur. J. Cell Biol.* 36:217-229.

Kelly, D. E. 1966. Fine structure of desmosomes, hemi-desmosomes and an adepidermal globular layer in developing newt epidermis. *J. Cell Biol.* 28:51-72.

Koch, P. J., M. J. Walsh, M. Schmeltz, M. D. Goldschmidt, R. Zimbelmann, and W. W. Franke. 1990. Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules. *Eur. J. Cell Biol.* 53:1-12.

Kozak, M. 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* 196:947-950.

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

Lai, C., M. A. Brow, K.-A. Nave, A. B. Noronha, R. H. Quarles, F. E. Bloom, R. J. Miller, and J. G. Sutcliffe. Two forms of 1B236/myelin-associated glycoprotein, a cell adhesion molecule for post-natal neural development, are produced by alternative splicing. *Proc. Natl. Acad. Sci. USA*. 84:4337-4341.

Liaw, C. W., C. Cannon, M. D. Power, P. K. Kibonaka, and L. L. Rubin. 1990. Identification and cloning of two species of cadherins in bovine endothelial cells. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:2701-2708.

Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science (Wash. DC)*. 227:1435-1441.

Little, P. F. R. 1987. Choice and use of cosmid vectors. In *DNA Cloning*. Vol. III. A Practical Approach. D. M. Glover, editor. IRL Press Ltd., Oxford. 19-42.

MacDonald, R. J., G. H. Swift, A. E. Przybyla, and J. M. Chirgwin. 1987. Isolation of RNA using guanidium salts. *Methods Enzymol.* 152:219-227.

McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA*. 74:4835-4838.

Matthey, D. L., and D. R. Garrod. 1986. Calcium-induced desmosome formation in cultured kidney epithelial cells. *J. Cell Sci.* 85:95-111.

Matthey, D. L., A. Suhrbier, E. Parrish, and D. R. Garrod. 1987. Recognition, calcium and the control of desmosome formation. *Ciba Found. Symp.* 125:49-65.

Matthey, D. L., G. Burdge, and D. R. Garrod. 1990. Development of desmosomal adhesion between MDCK cells following calcium switching. *J. Cell Sci.* 97:607-614.

Miller, K., D. Matthey, H. Measures, C. Hopkins, and D. Garrod. 1987. Localisation of the protein and glycoprotein components of bovine nasal epithelial desmosomes by immunoelectron microscopy. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:885-889.

Miyatani, S., K. Shimamura, M. Hatta, A. Nagafuchi, A. Nose, M. Matsunaga, K. Hatta, and M. Takeichi. 1989. Neural cadherin: role in selective cell-cell adhesion. *Science (Wash. DC)*. 245:631-635.

Nagafuchi, A., Y. Shirayoshi, K. Okazaki, K. Yasuda, and M. Takeichi. 1987. Transformation of cell adhesion properties by exogenously introduced E-cadherin cDNA. *Nature (Lond.)*. 329:341-343.

Nose, A., A. Nagafuchi, and M. Takeichi. 1987. Isolation of placental cadherin cDNA: identification of a novel gene family of cell-cell adhesion molecules.

- EMBO (Eur. Mol. Biol. Organ.) J.* 6:3655-3661.
- Nose, A., K. Tsuji, and M. Takeichi. 1990. Localization of specificity determining sites in cadherin cell adhesion molecules. *Cell*. 61:147-155.
- Overton, J. 1982. Inhibition of desmosome formation with tunicamycin and with lectin in corneal cell aggregates. *Dev. Biol.* 92:66-72.
- Ozawa, M., H. Baribault, and R. Kemler. 1989. The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:1711-1717.
- Ozawa, M., M. Ringwald, and R. Kemler. 1990. Uvomorulin-catenin complex formation is regulated by a specific domain in the cytoplasmic region of the cell adhesion molecule. *Proc. Natl. Acad. Sci. USA*. 87:4246-4250.
- Parrish, E. P., J. E. Marston, D. L. Matthey, H. R. Measures, R. Venning, and D. R. Garrod. 1990. Size heterogeneity, phosphorylation and transmembrane organisation of desmosomal glycoproteins 2 and 3 (desmocollins) in MDCK cells. *J. Cell Sci.* 96:239-248.
- Penn, E. J., C. Hobson, D. A. Rees, and A. I. Magee. 1987. Structure and assembly of desmosome junctions: biosynthesis, processing, and glycoprotein components in cultured epithelial cells. *J. Cell Biol.* 105:57-68.
- Pethica, B. A. 1961. The physical chemistry of cell adhesion. *Exp. Cell Res. (Suppl.)*. 8:123-40.
- Rayns, D. G., F. O. Simpson, and J. M. Ledingham. 1969. Ultrastructure of desmosomes in mammalian intercalated disc; appearances after lanthanum treatment. *J. Cell Biol.* 42:322-326.
- Ringwald, M., R. Schuh, D. Vestweber, H. Eistetter, F. Lottspeich, J. Engel, R. Dolz, F. Jahng, J. Epplen, S. Mayer, C. Muller, and R. Kemler. 1987. The structure of cell adhesion molecule uvomorulin. Insights into the molecular mechanism of Ca²⁺-dependent cell adhesion. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3647-3653.
- Sabatini, D. D., G. Kreibich, T. Morimoto, and M. Adesnik. 1982. Mechanisms for the incorporation of proteins in membranes and organelles. *J. Cell Biol.* 92:1-22.
- Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science (Wash. DC)*. 230:1350-1354.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463-5467.
- Schwarz, M. A., K. Owaribe, J. Kartenbeck, and W. W. Franke. 1990. Desmosomes and hemidesmosomes: constitutive molecular components. *Annu. Rev. Cell Biol.* 6:461-491.
- Seed, B. 1987. An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor CD2. *Nature (Lond.)*. 329:840-842.
- Shapiro, M. V., and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15:7155-7174.
- Shimoyama, Y., T. Yoshida, M. Terada, Y. Shimozato, O. Abe, and S. Hirohashi. 1989. Molecular cloning of a human Ca²⁺-dependent cell-cell adhesion molecule homologous to mouse placental cadherin: its low expression in human placental tissues. *J. Cell Biol.* 109:1787-1794.
- Skerrow, C. J., and A. G. Matoltsy. 1974. Isolation of epidermal desmosomes. *J. Cell Biol.* 63:515-523.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
- Steinberg, M. S. 1958. On the chemical bonds between animal cells: a mechanism for type-specific association. *Am. Nat.* 92:65-82.
- Steinberg, M. S., H. Shida, G. J. Guidice, N. H. Patel, and O. Blaschuk. 1987. On the molecular organisation, diversity and functions of desmosomal proteins. *Ciba Found. Symp.* 125:3-25.
- Suhrbier, A., and D. Garrod. 1986. An investigation of the molecular components of desmosomes in epithelial cells of five vertebrates. *J. Cell Sci.* 81:223-242.
- Sullivan, K. F. 1988. Structure and utilisation of tubulin isotypes. *Annu. Rev. Cell Biol.* 4:687-716.
- Takeichi, M. 1988. The cadherins: cell-cell adhesion molecules controlling animal morphogenesis. *Development*. 102:639-655.
- Takeichi, M. 1990. Cadherins: a molecular family important in selective cell-cell adhesion. *Annu. Rev. Biochem.* 59:237-252.
- Van Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* 133:17-21.
- Walsh, F. S., and G. Dickson. 1989. Generation of multiple N-CAM polypeptides from a single gene. *BioEssays*. 11:83-88.
- Walsh, F. S., C. H. Barton, W. Putt, S. E. Moore, D. Kelsell, N. Spurr, and P. N. Goodfellow. 1990. N-Cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin gene. *J. Neurochemistry*. 55:805-812.
- Watt, F. M., D. L. Matthey, and D. R. Garrod. 1984. Calcium-induced reorganization of desmosomal components in cultured human keratinocytes. *J. Cell Biol.* 99:2211-2215.
- Young, R. A., and R. W. Davis. 1983. Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA*. 80:1194-1198.