Characterization of the regulatory regions in the human desmoglein genes encoding the pemphigus foliaceus and pemphigus vulgaris antigens

Michael J. ADAMS^{*1}, Martin B. REICHEL[†], Ian A. KING^{*}, Mark D. MARSDEN^{*2}, Matthew D. GREENWOOD^{*}, Hilary THIRLWELL^{*2}, Joachim ARNEMANN[‡], Roger S. BUXTON^{*3} and Robin R. ALI[†]

^{*}Division of Membrane Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K., [†]Department of Molecular Genetics, Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL, U.K. and [‡]Institut für Humangenetik, Universitätsklinik, Theodor-Stern-Kai 7, Haus 9, D-60590 Frankfurt/M., Germany

The adhesive proteins in the desmosome type of cell junction consist of two members of the cadherin superfamily, the desmogleins and desmocollins. Both desmogleins and desmocollins occur as at least three different isoforms with various patterns of expression. The molecular mechanisms controlling the differential expression of the desmosomal cadherin isoforms are not yet known. We have begun an investigation of desmoglein gene expression by cloning and analysing the promoters of the human genes coding for the type 1 and type 3 desmogleins (DSG1 and DSG3). The type 1 isoform is restricted to the suprabasal layers of the epidermis and is the autoantigen in the autoimmune blistering skin disease pemphigus foliaceus. The type 3 desmoglein isoform is also expressed in the epidermis, but in lower layers than the type 1 isoform, and is the autoantigen in pemphigus vulgaris. Phage λ genomic clones were obtained containing 4.2 kb upstream of the translation start site of DSG1 and 517 bp upstream of the DSG3 start site. Sequencing of 660 bp upstream of DSG1 and 517 bp upstream of DSG3 revealed that there was no obvious TATA box, but a possible CAAT box was present at –238 in DSG1 and at –193 in DSG3 relative to the translation start site. Primer extension analysis and RNase protection experiments revealed four putative transcription initiation sites for DSG1 at positions –163, –151, –148 and –141, and seven closely linked sites for DSG3, the longest being at –140 relative to the translation start site. The

sequences at these possible sites at –166 to –159 in DSG1 (TTCAGTCC) and at –124 to –117 in DSG3 (CTTAGACT) have some similarity to the initiator sequence (CTCANTCT) described for a TATA-less promoter often from –3 to +5, and the true transcription initiator site might therefore be the A residue in these sequences. There were two regions of similarity between the DSG1 and DSG3 promoters just upstream of the transcription initiation sites, of 20 and 13 bp, separated by 41 bp in DSG1 and 36 bp in DSG3. The significance of these regions of similarity remains to be elucidated, but the results suggest that they represent a point at which these two desmoglein genes are co-ordinately regulated. Analysis of the upstream sequences revealed GC-rich regions and consensus binding sites for transcription factors including AP-1 and AP-2. Exon boundaries were conserved compared with the classical cadherin E-cadherin, but the equivalent of the second cadherin intron was lacking. A 4.2 kb region of the human DSG1 promoter sequence was linked to the *lacZ* gene reporter gene in such a way that there was only one translation start site, and this construct was used to generate transgenic mice. We present the first transgenic analysis of a promoter region taken from a desmosomal cadherin gene. Our results suggest that the 4.2 kb upstream region of DSG1 does not contain all the regulatory elements necessary for correct expression of this gene but might have elements that regulate activity during hair growth.

INTRODUCTION

Desmosomes are rivet-like points of intercellular contact that are a major type of junction especially in epithelial cells, but are also present in cardiac muscle, brain meninges and follicular dendritic cells [1]. Intermediate filaments, which provide a strengthening framework for the cytoplasm, are anchored at desmosomes and form a continuous network throughout the tissue via this structure. The adhesive proteins of the desmosome are the desmogleins and desmocollins, two different families of cadherin proteins [2,3]. Transfection experiments with non-adhesive fibroblasts have shown that the expression of a single type of desmosomal cadherin, for example with only a desmoglein, provides some cell–cell adhesion [4]. However, for strong intercellular

adhesion to occur, expression of both a desmoglein and a desmocollin together with the cytoplasmic protein plakoglobin is required (C. Marozzi, I. D. J. Burdett, R. S. Buxton and A. I. Magee, unpublished work). This latter protein is known to bind to desmogleins and desmocollins [5,6]. The desmogleins differ from other types of cadherin at the C-terminus, which contains an extra amino acid sequence; part of this is a repeated unique sequence that in the Dsg1 isoform occurs six times. This isoform is expressed in the epidermis and its clinical importance stems from its being the autoantigen in the autoimmune blistering skin disease pemphigus foliaceus. Another desmoglein isoform, Dsg3, is the autoantigen in the similar disease pemphigus vulgaris. In pemphigus foliaceus the blistering occurs higher up in the epidermis than in the more severe pemphigus vulgaris, reflecting

Abbreviations used: L-CAM, liver cell adhesion molecule; X-gal, 5-bromo-4-chloroindol-3-yl β -D-galactopyranoside.

¹ This paper is dedicated to the memory of our friend and colleague Michael Adams, whose work it mainly describes, who tragically died from leukaemia during the course of his studentship.

² Present address: Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104-2092, U.S.A.

³ To whom correspondence should be addressed.

The human DSG1 nucleotide sequence reported here will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AJ001716 and partly overlaps the cDNA sequence under accession number X56654. The human DSG3 nucleotide sequence will appear under accession number 708432.

the site of desmoglein gene expression as shown by hybridization *in situ* on human foreskin sections [7]. The other isoform, Dsg2, is not restricted to the epidermis, being more widely distributed; for example it is the only desmoglein isoform found in the heart [8,9].

At present we do not know what components of these genes are involved in specifying differential expression of the various isoforms, or what parts are concerned with co-ordinate expression of the desmogleins and desmocollins. There might be a locus control region requirement, reminiscent of the β -globin genes [10]. In fact the order of the desmocollin and desmoglein genes (DSC and DSG) on chromosome 18 is correlated with the spatial order of gene expression observed in the developing mouse embryo [11], suggesting that gene order is important for gene expression. Here we describe genomic clones containing the immediate 5' upstream regions of the desmoglein type 1 and type 3 genes that contain the putative promoters. To determine whether the DSG1 sequence had promoter activity, a transgenic approach was adopted because cell lines do not always express all the isoforms (especially the more suprabasal Dsg1). We have cloned 4.2 kb of the 5' upstream region of the human DSG1 gene coding for the Dsg1 isoform, attached this to a *lacZ* reporter gene, and then generated transgenic mice.

MATERIALS AND METHODS

Genomic library screening

An amplified human male genomic DNA library (Stratagene, Cambridge, U.K.) was used to isolate clones containing the 5' region of the DSG1 and DSG3 genes. The library had been constructed from 9–22 kb fragments of a partial *Sau3A* digest of placental DNA inserted into the *XhoI* site of the λ FIXII cloning vector. Plaque lifts of 10^5 recombinants were hybridized with a 498 bp *HindIII*–*XbaI* 32 P-labelled fragment of the most 5' region of human DSG1 cDNA [12] (see Figure 1a) and with a 426 bp *PstI*–*HindIII* 5' DSG3 cDNA probe (see Figure 4a). Prehybridization was performed at 68 °C in a solution containing 1% (w/v) SDS, 4×SSC and Denhardt's solution. Before addition, the radiolabelled probe was boiled with salmon sperm DNA (200 μ g/ml) and hybridization proceeded overnight at 68 °C. The filters were then washed at 63 °C, first in 0.1% SDS/0.4×SSC, then in 0.4×SSC. Positive recombinant phage were purified by repetitive (three or four) rounds of hybridizations. Out of approximately 10^6 plaques screened, two positive DSG1 plaques were identified, D6 and D10, and four positive DSG3 plaques, P1, P5, P6 and P7.

Human genomic DNA isolation

This was performed with whole blood as described by Miller et al. [13], by using a saturated salt solution to precipitate the cellular proteins of nucleated blood cells.

DNA sequencing and analysis, and genomic blotting

These were performed as described previously [12].

DSG1 RNase protection

Total RNA was purified from primary human keratinocytes by precipitation with LiCl [14] or with Qiagen columns. RNase protection was performed on plasmid pMA93T3 (see below) with an anti-sense RNA probe spanning the region from the T7 promoter of pBluescript to the *BamHI* site (see Figure 1b). The plasmid was linearized with *BamHI*, and labelled anti-sense

transcripts were synthesized *in vitro* with T7 RNA polymerase (Boehringer Mannheim) and [α - 32 P]UTP (800 Ci/mmol; Amersham), and purified from a 6% (w/v) sequencing gel. The labelled transcript (10^5 c.p.m.) was hybridized with 20 μ g of RNA in 80% (v/v) formamide/40 mM Na/Pipes (pH 6.7)/0.4 M NaCl/1 mM EDTA at 49 °C for 15 h. The annealed RNA was then digested with 5 units of RNase T₁ and 5 μ g of RNase A (Boehringer) for 15 min at 37 °C. The resulting protected fragments were run on a 6% (w/v) denaturing polyacrylamide gel and detected by autoradiography.

Primer extensions

For DSG1, a 21-mer primer JA115, 5'-ACTCCAGTCCATCT-CTGCTGG-3', or a 25-mer MA22, 5'-CTGTTCTTTTCTTT-CTCCCACTCA-3', both complementary to the 5' upstream region of DSG1 (see Figure 5) were 5' end-labelled with [γ - 32 P]ATP (Amersham) and T4 polynucleotide kinase (Promega). The labelled probes were annealed to approx. 10 μ g of total RNA from cultured keratinocytes and hybridized overnight at temperatures varying from 37 to 45 °C. The extension reaction was performed with 8 units of AMV reverse transcriptase (Promega) for 60 min at 42 °C. The reaction products were then analysed on a 6% (w/v) denaturing polyacrylamide gel alongside a DNA sequencing ladder, and detected by autoradiography. DSG3 primer extension was performed in a similar way with a 25-mer primer MA23, 5'-CTGCTTCTATCCTGCCAGCCGT-CTG-3', and a 17-mer MA1, 5'-CTGTAGTTCTGGGGAAG-3', complementary to the 5' upstream region of DSG3.

DSG1 reporter gene constructs

Genomic DNA from the clone pMA40T3 was excised from the Bluescript vector with *Sall* and *SmaI* (vector site) and cloned into the reporter gene vector pPD46.21 containing a multiple cloning site upstream of the *lacZ* gene encoding β -galactosidase [15] to make the construct pMA44T8. However, this construct contains two ATG translational start sites that could cause premature initiation of translation before the translational start site of the *lacZ* reporter gene. This problem was overcome by designing a primer upstream of the first ATG codon that contained an *EcoRI* site in a 5' extension (MA4, 5'-GACGTG-GAATTCTGCTGGATTCTCTTAC-3'). This primer was used for PCR with a primer 20 bp upstream of the *BamHI* site in the promoter region (MA5, 5'-GATCTTGTCAGCATTTC-3') to amplify the intervening region from pMA40T3. The resulting product was digested with *BamHI* and *EcoRI* to give a 191 bp fragment; this was cloned into pBluescript to give pMA93T3, cut out with *BamHI* and *EcoRV* from the multiple cloning site and ligated back into pMA44T8 cut with *BamHI* and *SmaI* to give the construct pMA108T5. This construct was checked by sequencing.

Transgenic mice

A 7.3 kb *Sall*–*NotI* fragment from construct pMA108T5 containing the regulatory region of the human DSG1 gene followed by the *lacZ* reporter gene was injected into pronuclei of fertilized mouse eggs to generate transgenic mice [16]. CBA $\times$ B.10 F₁ were used for egg collection, stud males and pseudopregnant females. Mature CBA $\times$ B.10 were used for breeding. High-molecular-mass DNA from tail biopsies was prepared with the QIAamp Tissue kit (Qiagen, Dorking, Surrey, U.K.) and used for Southern blot analysis with a *lacZ* probe or for PCR analysis with a pair of *lacZ* primers. Adult animals were killed by cervical dislocation; embryos and neonatal mice were killed by decapitation. The

Figure 1 Restriction maps of human DSG1

(a) The cDNA probes used for Southern blots and to screen for genomic clones (the numbers refer to bp in the sequence); (b) genomic DNA and cloned fragments (the exons are numbered).

Figure 2 DNA sequence of the 5' upstream region of the human DSG1 gene

The sites of possible initiation of transcription are shown by arrows. The predicted amino acid sequence in the first exon is also shown, and the DNA sequence of part of the first intron (underlined). Also shown are the positions of the primers used for primer extension experiments (JA115 and MA22), possible transcription initiation sites as determined by primer extension, possible CAAT and transcription factor-binding sites, a sequence with similarity to the 5' upstream region of the human DSC2 gene and the position of the *Bam*HI recognition site.

organs were dissected, flash-frozen in liquid nitrogen and embedded in 2-chlorotoluene (OCT) and then 8 μ m cryosections were cut and air-dried. Slides were incubated overnight at room temperature in a humidified chamber with X-gal (5-bromo-4-chloroindol-3-yl β -D-galactopyranoside; Calbiochem) in a solution containing 10 mM $K_3Fe(CN)_6$, 10 mM $K_4Fe(CN)_6$ and 2 mM $MgCl_2$ in PBS. After incubation with X-gal, the sections were counterstained with nuclear Fast Red, and the slides were mounted and examined by light microscopy. To examine reporter gene expression in the whole skin, mice were killed and the fur was removed from the body by plucking. The hairless pelts were

then washed for 10 min in 1% (w/v) SDS, pre-fixed for 45 min in 4% (w/v) neutral buffered paraformaldehyde and then stained overnight with X-gal before being fixed with 4% (w/v) neutral buffered paraformaldehyde.

RESULTS

Cloning and characterization of human genomic DNA containing the DSG1 5' upstream region

Southern blots of human genomic DNA revealed that a 498 bp *Hind*III-*Xba*I fragment (clone pMA3T2) of the most 5' region of

Figure 3 Determination of transcription initiation site of human DSG1

(a) In RNase protection assays total human keratinocyte RNA (lane +RNA), no RNA (lane B) or yeast RNA (lane C) was hybridized to the anti-sense probe generated from the T7 promoter shown in the diagram. After treatment with RNase the protected products were run on a denaturing gel along with size markers. Lane P, probe only. (b) Primer extension was performed with anti-sense oligonucleotides JA115 and MA22 shown in the diagram, end-labelled, and annealed to total human keratinocyte RNA or to yeast RNA. The complementary DNA strand was extended with reverse transcriptase and the length of the extended products was determined on a denaturing gel alongside a sequencing ladder. Lane B, no RNA; lane C, yeast tRNA; lane +RNA, keratinocyte RNA. The hybridizations were performed at 42 °C.

human DSG1 cDNA [12] hybridized to a 6 kb *EcoRI* fragment, which was absent when hybridized with a more 3' 351 bp *EcoRI*–*XbaI* fragment (results not shown) (see Figure 1a). To isolate DNA containing 5' upstream elements of the DSG1 gene, clone pMA3T2 was used to probe a human genomic library (see the Materials and methods section) yielding two positive plaques, D6 and D10, which hybridized strongly to it and showed identical restriction patterns (results not shown). A 5.5 kb *EcoRI*–*SalI* fragment of clone D6 showed strong hybridization and was likely to be within the 6 kb *EcoRI* fragment identified from the genomic blot. This 5.5 kb *EcoRI*–*SalI* fragment was subcloned from the phage into Bluescript to give clone pMA40T3. In addition two *XbaI* fragments of 6 and 1.8 kb were subcloned from phage clone D10 to give clones pMA41T17 and pMA65T6 respectively. Southern analysis resulted in the map shown in Figure 1(b), revealing that pMA40T3 contains approx. 4.2 kb of sequence upstream of the translation start and also the first exon and 1.2 kb into the first intron; clone pMA41T17 contains only

2.4 kb of upstream sequence, whereas pMA65T6 contains the second and third exons. Southern blots of cloned DNA and genomic DNA resulted in identical sizes, suggesting that the cloned DNA was not a pseudogene (M. A. Adams, unpublished work).

Structural features of the DSG1 5'-flanking region

Sequencing with primers derived from the cDNA sequence revealed the sites of the first and second introns and showed that the first intron is very large, being approx. 8 kb, whereas the second is smaller (less than 1 kb). The sites of the first and second introns were conserved compared with bovine DSG1 [17]. The first intron is located within the putative signal peptide as it is in bovine DSG1, human DSG3, murine E-cadherin, human and murine N-cadherin, murine P-cadherin and chicken liver cell adhesion molecule (L-CAM) [17–24]. Because the prosegment is much shorter in the desmogleins than in the classical cadherins,

A computer-assisted transcription factor database search with the SIGNAL SCAN program [29] of the genomic sequence for putative nuclear protein-binding sites upstream of the transcription start sites revealed a number of consensus motifs for DNA-binding proteins in the 5'-flanking region of the gene. Potential binding sites were present for AP-1 and AP-2 (Figure 2). A functional AP-1 element has been found in the mouse gene coding for the loricrin protein expressed in terminally differentiating keratinocytes [30], and AP-1 seems to be an important regulator for a number of keratin genes [31]. AP-2-binding sites are found in the promoters of a large number of epidermally expressed genes, including keratins K1, K5, K6, K14, K16 and K17 [32,33] and BPAG1 coding for the 230 kDa bullous pemphigoid antigen [34]. The significance of these consensus transcription factor-binding sites remains to be tested. The sequence GCCTGCATGT was similar to the Ker1-binding motif GCCTGCAGGC described by Leask [35], which binds the AP-2 transcription factor, shown to be important in the regulation of the K14 gene [32].

Finally a region of 13 nucleotides, marked on Figure 5, has 92% identity with a sequence in the 5'-upstream region of the human DSC2 gene [27].

Determination of the transcription initiation site of DSG1

The transcription initiation site of the human DSG1 gene was mapped by RNase protection and primer extension experiments. In the RNase protection experiments, a 243 nt RNA probe upstream of intron 1 was annealed to total RNA from human cultured keratinocytes, digested with RNase and analysed by electrophoresis. Six bands between 210 and 150 nt were obtained when the experiment was performed with RNA from differentiated keratinocytes (Figure 3a), corresponding to transcription initiation sites between -152 and -132. No product was obtained when the experiment was performed without RNA or with yeast tRNA. To confirm that there were no introns upstream of intron 1, primer extension experiments were performed with two primers complementary to the 5' end of the mRNA, JA115, at the site of translation initiation, and MA22, approx. 60 nt upstream of this site. This yielded four extension products that started at -141, -148, -151 and -163 nt from the translation initiation site, in reasonably good agreement with the results of the RNase protection experiments (Figure 3b).

Cloning and characterization of human genomic DNA containing the DSG3 5' upstream region and comparison with DSG1

Screening of the genomic library with a 5' *Pst*I-*Hind*III cDNA probe (Figure 4a) resulted in the isolation of four genomic clones. Two *Eco*RI-*Sal*I restriction fragments, 2.0 and 4.2 kb, failed to hybridize to a more 3' cDNA probe, suggesting that they were at the 5' end of the gene. The 4.2 kb fragment contained 0.5 kb upstream of the start site of translation with 3.5 kb into the first intron, whereas the 2.0 kb clone contained all of the second exon and part of the third exon (Figure 4b). The sequence of the putative DSG3 promoter was obtained by using primers derived from the cDNA sequence. This sequence was identical with that obtained by Silos et al. [18] with the exception of bp 70, which was a T residue in our clone but a C in that of Silos et al., which presumably represents a polymorphism (Figure 5). The major transcription initiation site of the human DSG3 gene was mapped by using primer extension experiments with primers MA1 and MA23 (Figure 5). This yielded a number of extension products, the longest being 100 nt with MA23, and 167 nt with MA1. These results (Figure 6) demonstrated that transcription within the cloned fragment occurred at a number of sites, the furthest being 140 bp upstream of the translation initiation site, with major sites at -132, -121, -110 and -84 bp. RNase protection experiments confirmed these results (M. A. Adams, unpublished work). These results are in good agreement with those of Silos et al. [18], who found a major site at -109 and minor sites at -140 and -88. In our experiments the major site seemed to be at -84.

Comparison of the DSG3 sequence with the DSG1 sequence obtained above revealed that there were two regions of similarity to the DSG1 promoter, of 20 and 13 bp, separated by 41 bp in DSG1 and 36 bp in DSG3 (Figure 7). This region also contains the possible CAAT box. The region between these sequences in DSG3, which still has some partial similarity to DSG1, has been found to confer epidermis-specific expression on a reporter gene [18]. As noted above, the transcription initiation sites at -163 in DSG1 and -121 in DSG3 start at an A residue in a region with some similarity between the two sequences, and which has some similarity to the initiator sequence (CTCANTCT) described for

Figure 6 Determination of transcription initiation site of human DSG3

Primer extension was performed with anti-sense oligonucleotides MA23 and MA1 shown in the diagram, end-labelled, and annealed to total human keratinocyte RNA or to yeast RNA. The complementary DNA strand was extended with reverse transcriptase and the length of the extended products was determined on a denaturing gel alongside a sequencing ladder. Lane B, no RNA; lane C, yeast tRNA; lane +RNA, keratinocyte RNA. The hybridizations were performed at 45 °C for MA23 and 37 °C for MA1.

a TATA-less promoter often from -3 to +5 [28]. It seems probable that this is the major transcription start site, at least in DSG1. This site is 25 bp downstream of the 13 bp region of similarity. The significance of these regions of similarity remains to be elucidated but it might represent a point at which these two desmoglein genes are co-ordinately regulated.

Analysis of DSG1 4.2 kb promoter activity in transgenic mice

The heterologous construct pMA108T5, containing the 4.2 kb DSG1 promoter upstream of the *lacZ* reporter gene, was microinjected into fertilized mouse eggs, and transgenic lines were generated. Twelve transgenic founders were identified by Southern blot analysis and the number of copies of the integrated transgene was estimated. Ten lines were established (one founder died, another did not transmit).

Adult transgenic mice from all of the established lines were killed; skin from the flank was then removed and cryosectioned. Slides were incubated overnight in X-gal. Only one line (no. 42)

Figure 7 Comparison of the sequences of human DSG1 and human DSG3 with the GAP program

The most similar regions are boxed. The possible sites of transcription initiation as determined by the primer extension experiments are shown. The translation start codons are underlined; the numbering is from the A in this sequence.

with the highest number of integrated transgenes (30) showed positive staining. A variety of organs (tongue, epidermis of paw and nose, bladder, oesophagus, parotid gland, pancreas, lung, kidney, heart, bowel, liver, eye, stomach) were taken from each of these mice and examined for *lacZ* expression; they were all negative. *lacZ* expression in epidermal appendages of transgenic line 42 was then analysed in more detail and the time course of expression was investigated. We examined skin from the flanks of embryos and neonatal, juvenile and adult mice. We observed a few positive epidermal cells in 16.5-day embryos, the earliest time at which we were able to do so. The proportion of positive cells in the epidermis increased over the next few days; by 2 days after birth, most of the epidermis in the flanks was positive. At this stage there were also a few positive cells in the hair follicle inner root sheath (Figure 8a). Over the following 3 weeks the epidermal staining was still prominent but was accompanied by increasingly intensive and extensive staining within the hair follicle (Figures 8b to 8d). By 3 weeks the epidermis, follicular canal, inner and outer root sheath, and bulb were all intensely stained. Thereafter the level of X-gal staining decreased (results not shown). An indication of the expression pattern across the whole animal was obtained by staining adult pelts (Figure 9). After incubating whole pieces of skin we discovered that penetration of X-gal through the surface was very inefficient. Removal of hair by plucking and gentle washing of skin with detergent increased X-gal staining of the whole organ by improving the penetration of reagents. We observed an expression pattern in adult mice that was patchy but consistent in general distribution and intensity among animals of the same age. We observed that with age the level of *lacZ* expression decreased but was always extremely sparse on the ventral and dorsal midlines (Figure 9). The low level of X-gal staining on the ventral and dorsal midlines and in the skin of older animals was not due to poor permeability to reagents: cryosections taken from older

animals and from the midline regions also showed sparse X-gal staining compared with cryosections taken from the flanks of mice that were 1 month old (results not shown).

DISCUSSION

The type 1 desmoglein isoform, Dsg1, the autoantigen in the blistering skin disease pemphigus foliaceus, exhibits a stratification-dependent mode of expression in human epidermis [7]. We have cloned the 5' upstream region of the human DSG1 gene and determined the major sites of transcriptional initiation. In this paper we have presented the first transgenic analysis of a promoter region taken from a desmosomal gene. Our results indicate that the 4.2 kb DSG1 promoter is not sufficient to direct copy number-dependent, integration site-independent gene expression. Tissue-specific reporter gene expression was found in only one out of ten transgenic mouse lines, which was also the line with the highest number of integrated transgenes. Although it is possible that, in this line, the transgenes had integrated close to an epidermally expressed murine gene, the chance of this happening must be extremely low. It therefore seems that some, but not all, of the elements responsible for tissue-specific DSG1 gene expression are contained in the 4.2 kb 5' upstream region.

Cell type-specific expression elements have been described for a number of keratin genes that, because they are part of the desmosome-cytokeratin complex, are pertinent comparisons. For example, the promoter region of the human keratin 14 gene faithfully mimics the expression of the endogenous murine keratin 14 gene in the basal layers of stratified epithelia [36]. Similar tissue-specific and cell type-specific expression of *lacZ* linked to the promoter of keratin 5 has been reported [37]. For other keratin genes, for example keratin 1, control elements that regulate epidermal-specific expression have been located in the 3'-flanking region of the gene [38,39]. All of the known DSG and

Figure 8 Histological examination of *lacZ* expression in DSG1 transgenic mice

Cryosections ($8\ \mu\text{m}$) through the dermis and hair follicles originating from the flank of DSG1-*lacZ* transgenic mice of different ages from line 42. Sections were incubated in X-gal overnight and counterstained with nuclear Fast Red. **(a)** At 2 days X-gal staining is restricted mostly to the epidermis, with only a few positive cells in the hair follicle inner root sheath ($\times 2680$). **(b)** By 6 days the epidermal staining is still prominent but accompanied by more intense staining within the upper half of the inner root sheath. Blue cells are apparent in the follicular canal ($\times 1070$). **(c)** At day 10 there is also staining of the outer and inner root sheath further down towards the follicular bulb ($\times 1070$). **(d)** At 3 weeks the epidermis, follicular canal, inner and outer root sheath and bulb are all intensely stained ($\times 1070$).

Figure 9 Expression of *lacZ* in pelts of DSG1 transgenic mice

Pelts taken from DSG1-*lacZ* transgenic mice from line 42 after careful shaving and 24 h of X-gal incubation. Left panel: at 3 weeks, blue staining of the skin is predominantly on the flanks of the animal but is sparse on the ventral and dorsal midlines. Right panel: at 6 months the staining is much weaker and is more patchy, but the dorsal and ventral midlines are completely spared.

DSC genes are very closely linked on human chromosome 18q12.1 [40–46], occupying a region of not more than 700 kb and being transcribed outwards from the region between the desmocollin and desmoglein clusters. The expression of the type 1 desmoglein isoform differs quite significantly from that of the type 1 desmocollin during embryonic development, with DSG1 transcripts found before DSC1 in keratinizing tissue [47]. Thus it seems quite likely that there are additional control elements, perhaps in the form of locus control regions [10], that are not in the present transgenic construct and might be points of common control between the DSG and DSC genes.

Although elements of the regulatory region of DSG1 are certainly missing from the flanking 4.2 kb 5' region, the transgenic construct, at least in high copy number and perhaps depending on chromosomal location, contains regulatory elements that are able to produce an expression pattern that partly reflects that of DSG1. The reporter gene was expressed specifically in the epidermis at around the correct time (1 day later than DSG1 mRNA is detected in the mouse by hybridization *in situ* [47]) and in the correct cell types. However, reporter gene expression was not detected in other tissues in which we would have expected to see DSG1. For example, mouse DSG1 has been detected by hybridization *in situ*, in the nose, tongue and paws. In fact the pattern of X-gal staining that we observed coincided with activity of the hair follicle, which undergoes regular cycles of growth (anagen), atrophy (catagen) and inactivity (telogen). In neonatal mice the cycles follow each other rapidly so that telogen lasts for only a few days [48–50]. However, telogen becomes progressively extended in subsequent cycles. Thus, for example, for hair follicles on the back of the mouse the first telogen lasts for a few

days, the second for 50 days and subsequent telogens for several months. New cycles of growth are very rare in old animals. During anagen the dermal papilla at the base of the fully differentiated follicle is almost enclosed by rapidly dividing matrix cells. Daughter cells move up and differentiate to form the concentric layers of the hair itself and of the inner root sheath, and the tip of the developing hair becomes keratinized. It seems that expression of the *lacZ* transgene coincides with this keratinization during hair growth. Thus β -galactosidase is readily detected in young animals with active hair follicles but is present in fewer cells in old animals, which have decreased hair renewal. Our results suggest that the 5' promoter region of DSG1 might contain an element that up-regulates DSG1 expression during the hair growth cycle. This might reflect the additional stress placed on follicles during this period.

The similarity between the DSG1 and DSG3 promoter sequences starts approx. 25–35 bp upstream of the major transcriptional start sites and could be where there is a common control of the transcription of these two desmoglein genes or, perhaps more probably because their expression patterns are different, similar but not identical transcription factors might bind in the two genes. In either case, it seems likely that this sequence could be a point at which transcription initiation is modulated.

A mouse mutation, *bal* (balding), has been mapped on murine chromosome 18 [51], closely linked to the position of the DSG and DSC genes [52,53]. This is a recessive mutation; affected mice are smaller than their normal littermates at birth and lose much of their hair before weaning. Although the hair regrows later, bare patches continue to appear on the head and body throughout adult life. The restricted distribution of DSG1

expression around the hair follicles lead us to suppose that *bal* could be a mutation of one of the DSG or DSC genes [52]; recent work has shown that this is indeed the case because *bal* mice have a chain-terminating mutation in the DSG3 gene [54], leading to decreased expression of DSG3, particularly in the matrix cells of active hair follicle bulbs (K. H. Sullivan, I. A. King, A. I. Magee, I. D. J. Burdett and R. S. Buxton, unpublished work).

We thank Carol Irving for her help in the preparation of this manuscript; Tony Magee for useful comments during this work; and Andrew Fire for the gift of *lacZ* reporter gene vector obtained via Siu-Pok Yee, whom, together with Peter Rigby, we also thank for help with these vectors. M.J.A. was supported by an MRC research studentship. M.D.G. and M.D.M. were supported by collaborative MRC research studentships jointly funded by Unilever Research plc, whose support we gratefully acknowledge. This work was funded by the Medical Research Council.

REFERENCES

- Collins, J. E. and Garrod, D. R. (eds.) (1994) Molecular Biology Intelligence Unit: Molecular Biology of Desmosomes and Hemidesmosomes, R. G. Landes, Austin, TX.
- Magee, A. I. and Buxton, R. S. (1991) *Curr. Opin. Cell Biol.* **3**, 854–861.
- Buxton, R. S. and Magee, A. I. (1992) *Semin. Cell Biol.* **3**, 157–167.
- Amagai, M., Kärpät, S., Klaus-Kovtun, V., Udey, M. C. and Stanley, J. R. (1994) *J. Invest. Dermatol.* **102**, 402–408.
- Troyanovsky, S. M., Troyanovsky, R. B., Eshkind, L. G., Krutovskikh, V. A., Leube, R. E. and Franke, W. W. (1994) *J. Cell Biol.* **127**, 151–160.
- Troyanovsky, S. M., Troyanovsky, R. B., Eshkind, L. G., Leube, R. E. and Franke, W. W. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10790–10794.
- Arnemmann, J., Sullivan, K. H., Magee, A. I., King, I. A. and Buxton, R. S. (1993) *J. Cell Sci.* **104**, 741–750.
- Schäfer, S., Koch, P. J. and Franke, W. W. (1994) *Exp. Cell Res.* **211**, 391–399.
- Angst, B. D., Buxton, R. S. and Magee, A. I. (1995) *Ann. N. Y. Acad. Sci.* **752**, 101–104.
- Grosveld, F., Blom van Assendelft, G., Greaves, D. R. and Kollias, G. (1987) *Cell* **51**, 975–985.
- King, I. A., Angst, B. D., Hunt, D. M., Krüger, M., Arnemann, J. and Buxton, R. S. (1997) *Differentiation* **62**, 83–96.
- Wheeler, G. N., Parker, A. E., Thomas, C. L., Ataliotis, P., Poynter, D., Arnemann, J., Rutman, A. J., Pidsley, S. C., Watt, F. M., Rees, D. A. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4796–4800.
- Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988) *Nucleic Acids Res.* **16**, 1215.
- Auffray, C. and Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314.
- Fire, A., Harrison, S. W. and Dixon, D. (1990) *Gene* **93**, 189–198.
- Palmiter, R. D. and Brinster, R. L. (1986) *Annu. Rev. Genet.* **20**, 465–499.
- Puttagunta, S., Mathur, M. and Cowin, P. (1994) *J. Biol. Chem.* **269**, 1949–1955.
- Silos, S. A., Tamai, K., Li, K. H., Kivirikko, S., Kouba, D., Christiano, A. M. and Uitto, J. (1996) *J. Biol. Chem.* **271**, 17504–17511.
- Ringwald, M., Baribault, H., Schmidt, C. and Kemler, R. (1991) *Nucleic Acids Res.* **19**, 6533–6539.
- Miyatani, S., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. and Takeichi, M. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8443–8447.
- Sorkin, B. C., Hemperly, J. J., Edelman, G. M. and Cunningham, B. A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7617–7621.
- Hatta, M., Miyatani, S., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. and Takeichi, M. (1991) *Nucleic Acids Res.* **19**, 4437–4441.
- Faraldo, M. L. M. and Cano, A. (1993) *J. Mol. Biol.* **231**, 935–941.
- Wallis, J., Fox, M. F. and Walsh, F. S. (1994) *Genomics* **22**, 172–179.
- Greenwood, M. D., Marsden, M. D., Cowley, C. M. E., Sahota, V. K. and Buxton, R. S. (1997) *Genomics* **44**, 330–335.
- Behrens, J., Löwrick, O., Klein-Hitpass, L. and Birchmeier, W. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11495–11499.
- Marsden, M. D., Collins, J. E., Greenwood, M. D., Adams, M. J., Fleming, T., Magee, A. I. and Buxton, R. S. (1997) *Gene* **186**, 237–247.
- Smale, S. T., Schmidt, M. C., Berk, A. J. and Baltimore, D. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4509–4513.
- Prestridge, D. S. (1991) *CABIOS* **7**, 203–206.
- DiSepio, D., Jones, A., Longley, M. A., Bundman, D., Rothnagel, J. A. and Roop, D. R. (1995) *J. Biol. Chem.* **270**, 10792–10799.
- Eckert, R. L. and Welter, J. F. (1996) *Mol. Biol. Rep.* **23**, 59–70.
- Leask, A., Byrne, C. and Fuchs, E. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7948–7952.
- Snape, A. M., Winning, R. S. and Sargent, T. D. (1991) *Development* **113**, 283–293.
- Tamai, K., Li, K. and Uitto, J. (1994) *J. Biol. Chem.* **269**, 493–502.
- Leask, A., Rosenberg, M., Vassar, R. and Fuchs, E. (1990) *Genes Dev.* **4**, 1985–1998.
- Vassar, R., Rosenberg, M., Ross, S., Tyner, A. and Fuchs, E. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1563–1567.
- Ramirez, A., Bravo, A., Jorcano, J. L. and Vidal, M. (1994) *Differentiation* **58**, 53–64.
- Huff, C. A., Yuspa, S. H. and Rosenthal, D. (1993) *J. Biol. Chem.* **268**, 377–384.
- Rothnagel, J. A., Greenhalgh, D. A., Gagne, T. A., Longley, M. A. and Roop, D. R. (1993) *J. Invest. Dermatol.* **101**, 506–513.
- Arnemmann, J., Spurr, N. K., Wheeler, G. N., Parker, A. E. and Buxton, R. S. (1991) *Genomics* **10**, 640–645.
- Arnemmann, J., Spurr, N. K., Magee, A. I. and Buxton, R. S. (1992) *Genomics* **13**, 484–486.
- Arnemmann, J., Spurr, N. K. and Buxton, R. S. (1992) *Hum. Genet.* **89**, 347–350.
- King, I. A., Arnemann, J., Spurr, N. K. and Buxton, R. S. (1993) *Genomics* **18**, 185–194.
- King, I. A., Sullivan, K. H., Bennett, R. and Buxton, R. S. (1995) *J. Invest. Dermatol.* **105**, 314–321.
- Simrak, D., Cowley, C. M. E., Buxton, R. S. and Arnemann, J. (1995) *Genomics* **25**, 591–594.
- Cowley, C. M. E., Simrak, D., Marsden, M. D., King, I. A., Arnemann, J. and Buxton, R. S. (1997) *Genomics* **42**, 208–217.
- King, I. A., O'Brien, T. J. and Buxton, R. S. (1996) *J. Invest. Dermatol.* **107**, 531–538.
- Dry, F. W. (1926) *J. Genet.* **16**, 287–340.
- Borum, K. (1954) *Acta Pathol. Microbiol. Scand.* **34**, 521–541.
- Johnson, E. (1965) in *Comparative Physiology and Pathology of the Skin* (Rook, A. J. and Walton, G. S., eds.), pp. 137–143, Blackwell Scientific, Oxford.
- Davisson, M. T., Cook, S. A., Johnson, K. R. and Eicher, E. M. (1994) *J. Hered.* **85**, 134–136.
- Buxton, R. S., Wheeler, G. N., Pidsley, S. C., Marsden, M. D., Adams, M. J., Jenkins, N. A., Gilbert, D. J. and Copeland, N. G. (1994) *Genomics* **21**, 510–516.
- Ishikawa, H., Silos, S. A., Tamai, K., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. and Uitto, J. (1994) *Mamm. Genome* **5**, 803–804.
- Koch, P. J., Mahoney, M. G., Ishikawa, H., Pulkkinen, L., Uitto, J., Shultz, L., Murphy, G. F., Whitaker-Menezes, D. and Stanley, J. R. (1997) *J. Cell Biol.* **137**, 1091–1102.