Characterisation of a human homologue of a yeast cell division cycle gene, MCM6, located adjacent to the 5′ end of the lactase gene on chromosome 2q21

Clare B. Harvey[a], Yangxi Wang[a], Dalila Darmoul[b], Alan Phillips[d], Ned Mantei[c], Dallas M. Swallow[a,*]

[a] *MRC Human Biochemical Genetics Unit, Wolfson House (UCL), 4 Stephenson Way, London NW1 2HE, UK*
[b] *INSERM U410, Faculté Xavier Bichat, 16 rue Henri Huchard, 75018 Paris, France*
[c] *ETH-Honggerberg, CH-8093 Zurich, Switzerland*
[d] *University Department of Pediatric Gastroenterology, Royal Free Hospital, Pond Street, London NW3 2QG, UK*

Received 12 September 1996

Abstract Four exons of a human homologue of a yeast cell division cycle gene (MCM6/mis5, which is thought to encode a DNA replication licensing factor) have been identified 3.3 kb upstream from the start of transcription of the intestinal lactase gene on human chromosome 2q21, initially by similarity to a rat 'intestinal crypt-cell replication factor'. RT-PCR analysis shows, that unlike lactase, MCM6 is not restricted in its tissue distribution and does not show person-to-person variation in the level of expression in adult intestine.

Key words: Lactase; MCM6; Human intestine; Gene expression; DNA replication

1. Introduction

In the course of our efforts to identify the *cis*-acting element which causes the lactase persistence polymorphism in adult humans [1], we have been sequencing upstream of the start of transcription of the lactase gene (LCT). This paper is concerned with the identification of a novel gene in close proximity to LCT, which has a high level of identity to a rat intestinal crypt-cell replication factor [2]. Both the human and rat genes appear to be mammalian homologues of a yeast cell division cycle gene identified as mis5 (in *Schizosaccharomyces pombe*), more recently named MCM6 [3,4].

MCM6 is one of a large group of genes which were first identified in yeast by the analysis of temperature-sensitive mutants, which are either blocked at some point in the cell cycle (cdc) or unable to maintain the presence of mini-chromosomes (MCM). The genes were separated into groups or families depending on where the arrest occurs. MCM6 falls into the group (MCM2-7 [3,4]) in which the arrest occurs before S-phase. The protein products appear to play a role in 'licensing' the replication of DNA [5,6] and are thought to bind to DNA after mitosis (G1) and to be removed from the DNA during replication (S phase). They demonstrate a 200 amino acid region of high similarity, which resembles DNA-dependent ATP-binding domains [7].

Identification of homologues of these genes in man is of great interest since they may play a role in cellular differentiation. Several of these genes have been identified in vertebrate DNA and homologues have so far been identified for MCM2 (BM28 [8]), MCM3 (P1 [9]), MCM4 (CDC54/cdc21 [10]), MCM5 (CDC46 [10]), and MCM7 (cdc47 [10]). A human MCM6 homologue has not so far been reported.

2. Materials and methods

2.1. Subcloning

The lambda clone hchrlac7 has been reported previously [11]. This clone was sequentially digested by *Eco*RI and *Hin*dIII using the conditions recommended by the manufacturer (BRL). The resulting fragments were subcloned into pGEM-3Z (Promega) which had been digested with both *Eco*RI and *Hin*dIII. Clones with the correct insert sizes were chosen for further analysis. Three subclones were obtained: plc7, insert size 1.2 kb; plc9, insert size 1.8 kb and plc1, insert size approx. 9 kb. Representative colonies of each clone were grown and DNA prepared using a Wizard kit (Promega).

2.2. PCR products for sequencing

PCR products over 1 kb in length were made to cover the gaps remaining from the subcloning approach. The primer sequences are available on request. The amplification was performed in 100 μl volume and using standard conditions [12]. Taq polymerase was obtained from Advanced Biotechnologies or Promega and amplification performed in a Hybaid Omnigene apparatus. Products were cleaned using a 4B Sepharose column (Pharmacia) and the resulting column fraction was used directly as sequencing template [13].

2.3. Sequencing

Sequencing of clones was performed using Sequenase v2.0 (US Biochemicals) using double-stranded templates as described [14]. PCR products were initially boiled with primers and buffer for 5 min and then snap cooled on dry ice and ethanol. The primers used were either from vector sequence or designed to 'walk' out from new sequence.

2.4. Sequence analysis

The GENBANK and EMBL databases were searched using the BLAST program and additional comparisons were made using FASTA, Wordsearch, Bestfit, Gap and Pileup, all parts of the GCG package (University of Wisconsin) at the HGMP-RC computing centre. Both protein translations and nucleotide sequences were used in searches and comparisons.

2.5. RNA samples and RT-PCR analysis

The human tissues were as described previously [1,15–17]. Semi-quantitative RT-PCR was performed using the method described previously [15] except that the RT was performed at 37°C using 1 μg total RNA as starting material for 10 PCR reactions and, for MCM6 only, 30 rounds of amplification (which gave a linear dose response for this transcript). Primer sequences for dipeptidylpeptidase IV (DPP4) and lactase-phlorizin hydrolase (LCT) were as described previously. The primers for the MCM6 transcript were ZA1S and 4XA (see Table 1). Controls included the omission of reverse transcriptase, no template blanks and test amplifications of human genomic DNA.

*Corresponding author. Fax: (44) (171) 387 3496.
E-mail: dswallow@hgmp.mrc.ac.uk

PII S0014-5793(96)01189-1

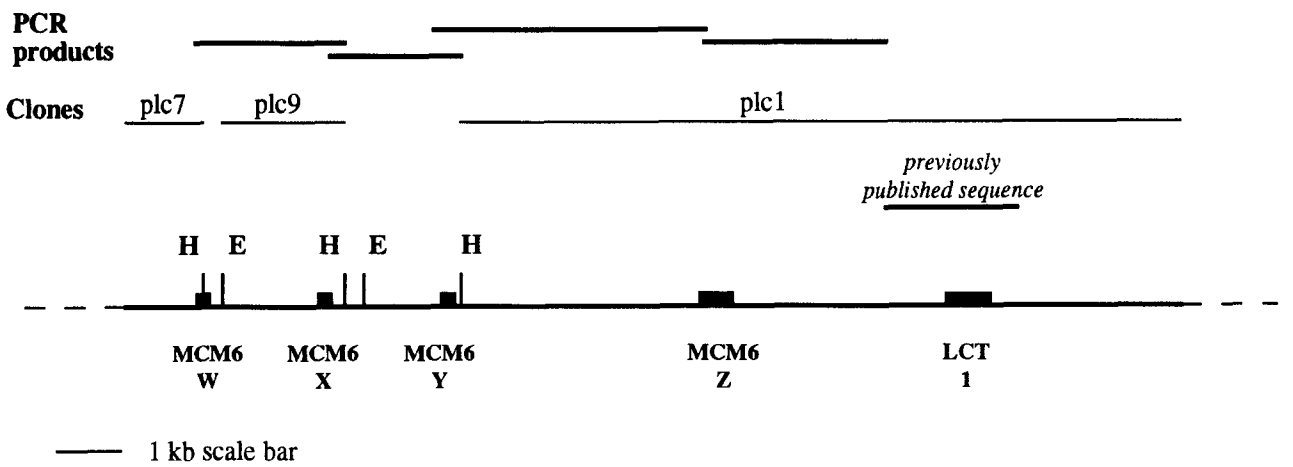

Fig. 1. Diagrammatic representation of the lambda hchrlac7 clone showing the positions of the MCM6 exons and LCT exon1 in relation to the *Eco*RI (E) and *Hin*dIII (H) sites, and to the plc1, 7 and 9 subclones and the PCR products covering the region. The previously published sequence [11,12,18] was obtained from more than five different individuals.

3. Results and discussion

3.1. Identification of novel exons upstream from the lactase gene

The human genomic clone (lambda hchrlac7) contains 15 kb of insert sequence and includes the first exon and part of the first intron of the lactase gene [11]. 1 kb of sequence upstream of the start of transcription has been published [18]. To characterise the remainder of the clone, three *Eco*RI/*Hin*dIII subclones were partially sequenced, initially using vector primers and then further primers made to 'walk' along the sequence. Gaps not covered by the subclones were sequenced using a PCR-based strategy. Fig. 1 shows the positions of the subclones and the PCR products sequenced.

As they were obtained, sequences were used to search the databases. A novel exon (W) was identified through a short region of similarity to a rat intestinal crypt-cell replication factor (U17656) [2] and was shown to be 136 bp in length. Exons X (156 bp) and Y (140 bp) of the same transcript were identified by direct comparison of new sequence with the rat sequence.

The positions of each of these exons are shown in the map in Fig. 1.

3.2. Construction of a composite cDNA from EST sequences

A derived composite cDNA sequence was made by joining together the W, X and Y exon sequences and this was used to screen the databases, including the EST (expressed sequence tags) databases. The databases were also screened using the rat (U17565) and *S. pombe* (mis5, D31960) sequences (see below). The overlapping EST sequences identified formed a cDNA contig spanning 1265 bp and contained the three exons already identified in genomic DNA (Fig. 2).

The most 3′ exon (Z) of the transcript was localised in the genomic subclone plc1 by PCR using primers taken from the composite human cDNA sequence together with primers from sequence from the subclone itself. The exon was sequenced using these PCR products as template. This exon is 509 bp in length and includes the 3′UTR and polyadenylation signal. Its position is shown in Fig. 1. Exon W is the most 5′ exon present in the genomic clone.

The composite human cDNA sequence shows 99.5% identity (within the high-quality sequence ranges of the ESTs) to the genomic exons, right up to the start of the poly(A) tail. Very recently, two complete human cDNA sequences have appeared on the databases [19,20] and these both also show 99% identity to the sequences reported here.

3.3. Sequencing of the exon/intron boundaries

By comparison of the genomic and cDNA sequences it was possible to identify the intron/exon boundaries of all four exons. In each case, further sequencing primers were designed as necessary such that the complete exon and the exon/intron boundaries at either side were sequenced on both strands: these sequences will be available on the databases (U67281–4). Table 1 lists the intron primers used to amplify and/or sequence across the exons. Classical ag/gt splice sites were identified at the ends of each exon.

3.4. Sequence comparison with members of the MCM2-7 family

Since the rat intestinal crypt-cell replication factor identified by Sykes and Weiser [2] had been shown to have a high level of similarity to the cdc/MCM family of proteins, the composite human cDNA sequence and the derived protein sequence were used to investigate the relationship of this transcript to the MCM gene family. These analyses showed that both of these sequences are most similar to the mis5 gene of *S. pombe* [21] (Fig. 3), the conclusion also reached by Tsugura and colleagues who recently submitted their sequence to the databases [19]. The name MCM6 has been suggested for these sequences [3,4]. Thus, the human gene described in this paper is designated MCM6 (HUGO/GDB approved gene symbol).

3.5. Chromosomal location

The clone lambda hchrlac7 had previously been used for the chromosomal mapping of the lactase gene and gave only one signal in fluorescence in situ hybridisation experiments [22], thus mapping MCM6 to 2q21. A PCR approach was used to confirm the disposition of the MCM6 exons and their close proximity to LCT in human genomic DNA in vivo. PCR amplification of genomic DNA performed in overlapping segments across the whole region gave the expected size products (Fig. 1). In addition, a primer in the 3′ exon (within the 3′

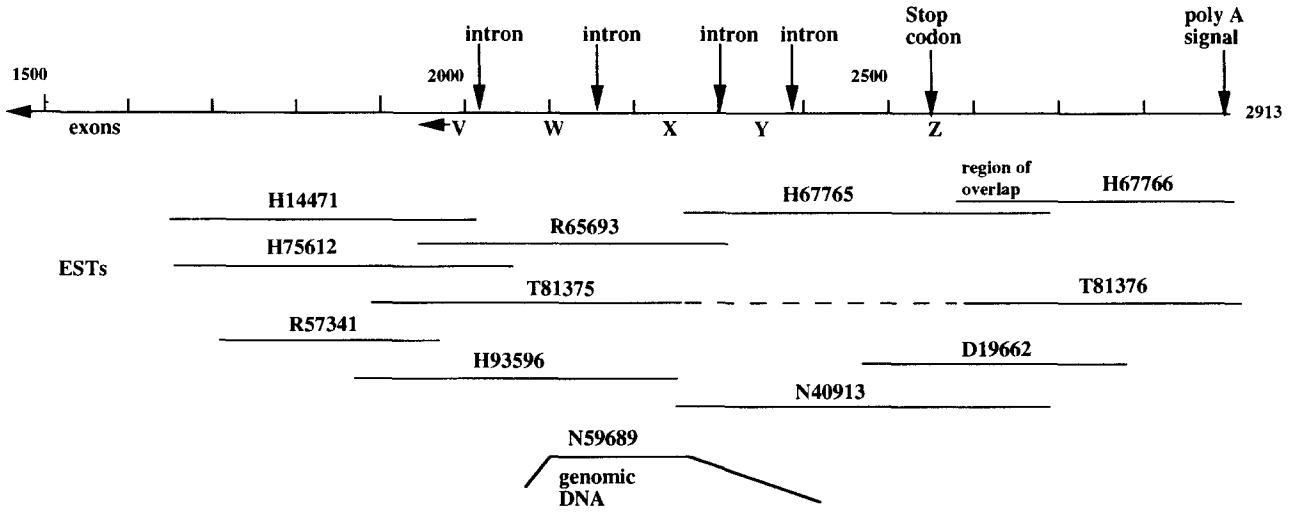

Fig. 2. Map of the contig made from EST sequences. The ESTs are identified by their Genbank/EMBL accession numbers. N.B. the sequences H67765/6 and T81375/6 are from opposite ends of the same clones. The nucleotide positions are numbered from one of the human cDNAs [20] which has recently appeared on the databases. The exon boundaries are indicated.

UTR) used together with a primer 200 bases upstream from the previously characterised lactase promoter sequence [12,18], produced the same 2.4 kb band both in the clone DNA and in 5 different genomic DNA samples (2 lactase persistent, 2 non-persistent and a known heterozygote).

Other homologues of the MCM2-7 genes have been mapped in humans, MCM7 (p85MCM, CDCL1) to 7q21 [23] and MCM3 has been mapped using the murine homologue as probe, to 6p21 [24]. Thus, these genes are dispersed in the human genome as they are in the genome of *S. cerevisiae* (SacchDB).

3.6. MCM6 expression in human tissues

We estimate that the start of transcription of LCT is 3.3 kb from the end of MCM6. It seems probable that there are elements involved in the developmental and tissue-specific expression of LCT upstream from the immediate promotor but whether or not they overlap with MCM6 remains to be determined. The fact that MCM6 was reported to be expressed in intestine in the rat while lactase itself is expressed specifically in the intestine was interesting. However, the clones containing the ESTs were derived from a variety of cDNA libraries from different tissues and included adult breast, female placenta, fetal heart, and mixed fetal liver and spleen, suggesting that this sequence is expressed in a wide variety of tissues other than intestine.

In order to explore the pattern of expression further and to compare with the pattern of lactase expression we performed RT-PCR experiments. Primers designed to amplify the MCM6 transcript were taken from separate exons such that a different sized product would be amplified from genomic DNA from that obtained from cDNA. Two different pairs of primers (5XS located in the next exon upstream, exon V, and taken from the rat sequence, together with 3UXA in exon Z; and ZA1S in exon X together with 4XA in exon Z, Table 1) were shown to amplify cDNA from fetal intestine to produce products of predicted size, confirming that these sequences are part of an expressed gene. One pair (ZA1S and 4XA) was selected to examine the tissue distribution. Lactase (LCT) and dipeptidyl peptidase IV primers were used on the same cDNAs. Examples of the results are shown in Fig. 4. The MCM6 transcript was expressed in all the tissues examined in both fetuses and adults. From adults these were: liver, small intestine (n=4), colon (n=4), skeletal muscle, heart muscle and lung. From fetuses these were: liver, small intestine (n=7), colon (n=2), skeletal and limb muscle, stomach, pancreas, kidney, spleen, and brain. We also tested term placenta. In general, a lower level of transcript was observed in adult

Table 1
Exon primers used for RT-PCR and intron primers used for sequencing and/or PCR of the exons

Primer name	Location	Comments	Sequence
5XS	exon V sense (rat)	RT-PCR	5′ gca ctg ctg tga tga gg 3′
3UXA	exon Z antisense	RT-PCR	5′ cac agt tcc tca gct ctg g 3′
ZA1S	exon X sense	RT-PCR	5′ gag tac tgc cga atc tct acc 3′
4XA	exon Z antisense	RT-PCR	5′ cac tgg agc ctt tca gtc c 3′
7AS	intron V sense	forward through exon W	5′ cta ctg tga ttt cat tgg gag 3′
9Z2S	intron W antisense	back through exon W	5′ caa cag cag gtg tac ttc tac 3′
ZA3S	intron W sense	forward through exon X	5′ gta aat gct tga caa ggg 3′
ZA1A	intron X antisense	back through exon X	5′ cag gct ggc agg tat gct g 3′
RQ2	intron X sense	forward through exon Y	5′ ctg aca gta cga gtt gct ag 3′
1AS	intron Y antisense	back through exon Y	5′ ggt gac agc agt acg acc aag 3′
5ZXS	intron Y sense	forward through exon Z	5′ ggt cag cat cct act agg 3′
3ZXA	intron Z antisense	back through exon Z	5′ cca cag ttc tac ctc ttc tc 3′

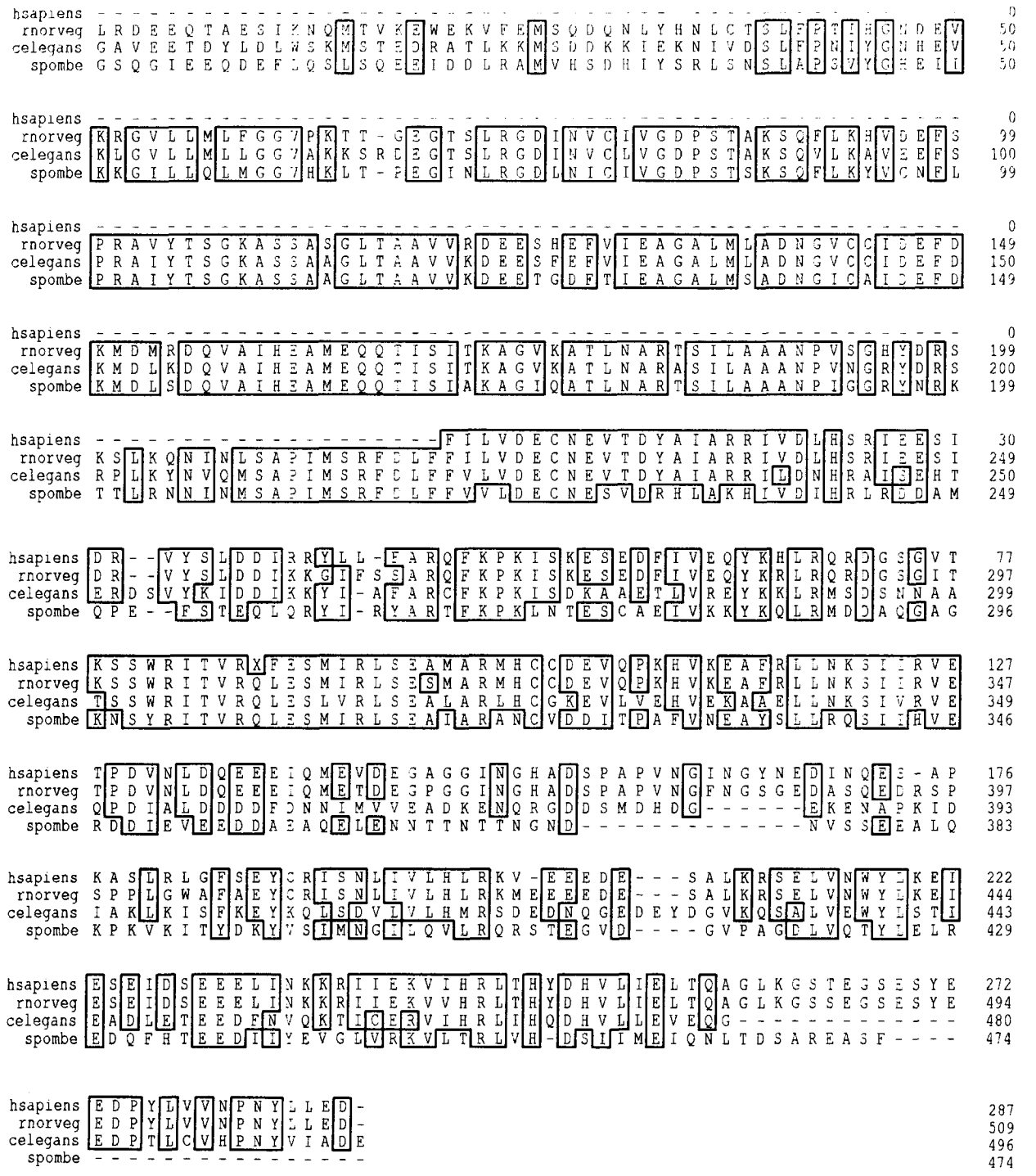

Fig. 3. Comparison of the deduced amino-acid sequence of the MCM6 contig (hsapiens) and other MCM transcripts (from the rat (rnorveg): accession no. U17565; from *C. elegans*: accession no. P34647; from *S. pombe*: accession no. D31960).

tissues than in fetal tissues. Small intestine samples from 7 children (age range 0–8 years) were also tested. The level of MCM6 transcript in the small intestine of children was comparable to that of the adults, both children and adults showing a lower level of expression than the fetuses. This pattern of expression contrasts with that of lactase which shows a very restricted distribution, occurring at detectable levels only in the intestine, and at high levels only in the small intestine of children and lactase persistent adults [15,17] (Fig. 4).

We also explored the possibility that the level of expression of MCM6 in adult intestine might show person-to-person variation as observed for the lactase polymorphism. We thus examined lactase and MCM6 mRNA levels in adult small intestine from 10 different individuals of known lactase persistence status (5 lactase persistent and 5 non-persistent).

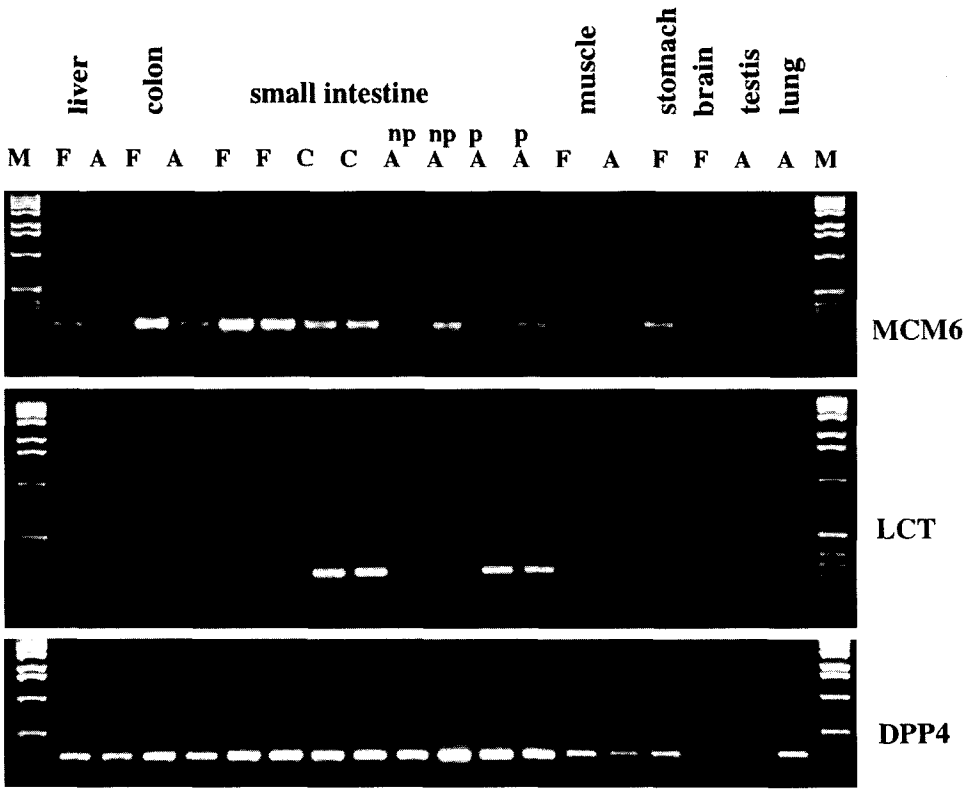

Fig. 4. Sample results of RT-PCR analysis showing the widespread expression of MCM6 in comparison with the restricted distribution of lactase (LCT). Dipeptidyl peptidase 4 (DPP4) which is more widely expressed was also tested on the same cDNA as 'positive' control. The faint additional bands detected with the MCM6 primers were shown to due to the presence of trace amounts of genomic DNA in some of the total RNA preparations. The tissues were derived from fetuses (F), children (C) or adults (A). The small-intestinal samples were taken from lactase persistent (P) or non-persistent (NP) individuals. M=kb ladder (BRL).

There was no apparent correlation in the levels of the MCM6 and lactase transcripts.

These results thus suggest that there are no functionally significant tissue specificity or developmental *cis*-acting elements shared by the two genes. There can, however, be overlap of the regulatory regions of one sequence with the end of another transcript over substantial distances. For example, a critical *cis*-acting tissue specific element for the α-globin cluster is located within intron 5 of the so-called '−14' gene transcript at 40 kb from the start of transcription of the ζ gene [25]. Interestingly, the −14 gene, like MCM6, shows ubiquitous expression.

The role of MCM6 in vertebrate cells is not known. It is tempting to speculate that in higher organisms it could play a role in terminal differentiation as well as licensing DNA replication. It is noteworthy that the transcript is expressed at relatively high levels both in tissues which contain many proliferating cells, such as testis and intestine, and in tissues where there is less proliferation such as muscle. Studies are in progress to determine whether, as reported in the rat, the transcript is more abundant in the dividing intestinal crypt cells than in the differentiated cells of the villi.

Acknowledgements: We are grateful to the MRC Tissue Bank and clinical collaborators from various hospitals, especially the Departments of Gastroenterology and Histopathology, UCLH, and the Department of Paediatric Gastroenterology, Queen Elizabeth Hackney for tissues, to Amy Brar-Rai for technical assistance, Adrian Jenkins for the sequencing protocol, and to the MRC HGMP Resource Centre for computing facilities. C.B.H. was funded by a British Digestive Foundation Training Fellowship.

References

[1] Wang, Y., Harvey, C.B., Pratt, W.S., Sams, V.R., Sarner, M., Rossi, M., Auricchio, S. and Swallow, D.M. (1995) Hum. Mol. Genet. 4, 657–662.
[2] Sykes, D.E. and Weiser, M.M. (1995) Gene 163, 243–247.
[3] Chong, J.P.J., Thommes, P. and Blow, J.J. (1996) Trends Biochem. Sci. 21, 102–106.
[4] Kearsey, S.E., Maiorano, D., Holmes, E.C. and Todorov, I.T. (1996) Bioessays 18, 183–190.
[5] Blow, J.J. and Laskey, R.A. (1988) Nature 332, 546–548.
[6] Tye, B.-K. (1994) Trends Cell Biol. 4, 160–166.
[7] Koonin, E.V. (1993) Nucleic Acids Res. 21, 2541–2547.
[8] Todorov, I.T., Pepperkok, R., Philipova, R.N., Kearsey, S.E., Ansorge, W. and Werner, D. (1994) J. Cell Sci. 107, 253–265.
[9] Thommes, P., Fett, R., Schray, B., Burkhart, R., Barnes, M., Kennedy, C., Brown, N.C. and Knippers, R. (1992) Nucleic Acids Res. 20, 1069–1074.
[10] Hu, B., Burkhart, R., Schulte, D., Musahl, C. and Knippers, R. (1993) Nucleic Acids Res. 21, 5289–5293.
[11] Boll, W., Wagner, P. and Mantei, N. (1991) Am. J. Hum. Genet. 48, 889–902.
[12] Harvey, C.B., Pratt, W.S., Islam, I., Whitehouse, D.B. and Swallow, D.M. (1995) Eur. J. Hum. Genet. 3, 27–41.
[13] Almond, N., Jenkins, A., Slade, A., Heath, A., Cranage, M. and Kitchin, P. (1992) AIDS Res. Hum. Retroviruses 8, 77–87.
[14] Pratt, W.S., Islam, I. and Swallow, D.M. (1996) Ann. Hum. Genet. 60, 21–28.
[15] Wang, Y., Harvey, C., Rousset, M. and Swallow, D.M. (1994) Pediatr. Res. 36, 514–521.
[16] Harvey, C.B., Wang, Y., Hughes, L.A., Swallow, D.M., Thurrell,

W.P., Sams, V.R., Barton, R., Lanzon-Miller, S. and Sarner, M. (1995) Gut 36, 28–33.
[17] Harvey, C.B., Wang, Y., Phillips, A.D., Walker-Smith, J.A. and Swallow, D.M. (1996) Gut 38 (Suppl.), A21.
[18] Lloyd, M., Mevissen, G., Fischer, M., Olsen, W., Goodspeed, D., Genini, M., Boll, W., Semenza, G. and Mantei, N. (1992) J. Clin. Invest. 89, 524–529.
[19] Tsuruga, H., Yabuta, N., Hosoya, S. and Nojima, H. (1996) Accession no. D84557.
[20] Holthoff, H.P. and Knippers, R. (1996) Accession no. U46838.
[21] Takahashi, K., Yamada, H. and Yanagida, M. (1994) Mol. Biol. Cell 5, 1145–1158.
[22] Harvey, C.B., Fox, M.F., Jeggo, P.A., Mantei, N., Povey, S. and Swallow, D.M. (1993) Ann. Hum. Genet. 57, 179–185.
[23] Nakatsuru, S., Sudo, K. and Nakamura, Y. (1995) Cytogenet. Cell Genet. 68, 226–230.
[24] Starborg, M., Brundell, E., Gell, K., Larsson, C., White, I., Daneholt, B. and Hoog, C. (1995) J. Cell Sci. 108, 927–934.
[25] Vyas, P., Vickers, M.A., Picketts, D.J. and Higgs, D.R. (1995) Genomics 29, 679–689.