Leading article

Regional expression of intestinal genes for nutrient absorption

Many genes are specifically expressed in the intestine and produce the proteins that digest and absorb the diverse nutrient molecules in the diet. Despite apparent similarities of the mucosa in the duodenum, jejunum, and ileum, these regions differ in their abilities to absorb various nutrients. Clinical awareness of these differences is perhaps most obvious in vitamin B12 malabsorption following terminal ileal disease or resection, but there are many other examples including distal intestinal absorption of bile salts, proximal absorption of calcium and iron, and the various electrolyte transporters that produce net fluid secretion proximally but absorption distally.[1] These anatomical differences become important in the treatment of patients with the short bowel syndrome where certain, but not all, nutrient systems are capable of adaptation.

Recent work is giving a better understanding of the fundamental reasons for these regional variations in gene expression, which are believed to underpin differences in cell and tissue differentiation. This article will review patterns of differentiation in intestinal cells and the control of the expression of nutrient transport genes by specific transcription factors.

Intestinal cell differentiation

The basis for restriction of specialised transport systems to different parts of the gut may partly be found in studies of how cell fate is determined along the crypt-villus and longitudinal axes of the small intestine. Each crypt contains non-migratory, self renewing stem cells and zones of proliferation and differentiation as the cells migrate on to the villi. In adult mammals, the intestine is one of a select number of tissues that undergoes continuous regeneration and can be viewed as a microcosm of ongoing development, with a prodigiously high turnover of cells.

It is not surprising, therefore, that students of cell specification in the adult intestine have turned to embryology for some answers. The mucosa is derived from endoderm, in fore-gut, mid-gut, and hind-gut regions, and shares aspects of development with associated organs such as the pancreas and liver. During development, both prenatal and postnatal, some genes are expressed in most of these cells, others are restricted to certain regions. Several of these genes have been studied in detail and principles governing the transcriptional control of their expression by elements in the promoter and enhancer regions of the genes are now becoming apparent.

Fatty acid binding proteins (FABP)

Much has been learnt from these two small proteins that bind fatty acids in the cytosol of the enterocyte; the one known as liver FABP is expressed in both liver and intestine whereas intestinal FABP is found only in enterocytes. Using liver FABP as a marker in the developing rat embryo, Rubin[2] showed that this gene is silent until embryonic days 17–18 when it becomes activated and its mRNA is then readily detectable in proximal intestine. Activation proceeds distally in a wavelike fashion over one to two days providing evidence for a genetic developmental programme. What is it about this gene, or for that matter about any gene with a restricted intestinal distribution, that enables expression to be regulated in this way?

Detailed studies in transgenic mice of the intestinal FABP promoter region linked to a reporter gene have demonstrated specific regions that seem to be important in directing intestine specific expression.[3] For instance, the segment of nucleotides from −103 to +28 is able to direct appropriate transgene expression in late fetal life, with reporter expression restricted to cells of the enterocyte lineage, and with an appropriate duodenal to colonic gradient. The shape of this cephalo-caudal gradient was found to be influenced both positively and negatively by other sequences; elements positioned betwen nucleotides −1178 and −278 increased expression in ileum and colon, while deletion of those located between −277 and −185 resulted in precocious expression in crypts and the upper small intestine and also in the colon, both areas where FABP expression would not normally be expected. Thus, remarkably, nucleotide sequences of fewer than 100 base pairs include elements that can restrict reporter gene expression to proximal intestine or to villus cells. A conserved 14 bp element that binds the transcription factor HNF-4 had been identified in this promoter[4] and another 20 bp element with no known ligand appears to function as a suppressor of expression in proximal intestine.[5]

Lactase/phlorizin hydrolase

Two brush border membrane disaccharidases, lactase/phlorizin hydrolase (LPH) and sucrase-isomaltase, are found in mid-intestine. Lactase particularly declines in expression towards the ileum but another important aspect of lactase expression is the down regulation occurring after weaning. This temporal restriction in expression represents the usual mammalian pattern of regulation rather than a disease state in humans with adult lactose intolerance. Regions of the lactase promoter with intrinsic ability to down regulate gene expression have recently been identified. Although comparisons of the promoter in lactase restricted and non-restricted subjects have found no differences, a nuclear protein has been identified, designated NF-LPH1, which recognises and binds to a short nucleotide sequence 40 bp upstream of the transcriptional start site.[6] Furthermore, the post-weaning decline of lactase activity is paralleled by a reduction in NF-LPH1 in pigs,[7] though clearly the system will prove to be much more complex.

LPH was used as a marker in intestinal transplant experiments, which demonstrated elegantly that temporal and spatial developmental patterns are programmed by the original identity of the transplanted segment and, in particular, by the endoderm.[8]

Sucrase-isomaltase

The understanding of the factors regulating expression of sucrase-isomaltase (SI) is also proving to be exciting. In transient transfection experiments (which provide a more immediate analysis than transgenesis for promoter mapping) proximal and distal regulatory elements were identified. Constructs containing bases from −303 to +60 produced sevenfold to 11-fold greater activity compared with the promoterless construct, and addition of bases −3580 to +60 gave an increase of over 30-fold.[9]

Three nuclear protein binding sites within 183 bp of the transcription start site were identified in studies using DNase I footprint analysis.[10] Deletion of each footprint in transfection experiments led to significant reductions in reporter gene expression and mutations within the sequence known as SIF1 (sucrase-isomaltase footprint 1) produced similar results. Comparison of human and mouse sequences of each putative regulatory sequence showed close homology, with 100% identity in the case of SIF1. Using the SIF1 sequence as a probe to screen a mouse jejunal cDNA expression library led to the identification of a gene product that had the same pattern of binding to SIF1 as the crude nuclear extract.[11] Sequencing this gene showed it to be identical to mouse Cdx-2, a homeobox protein (see later). It is probable that NF-LPH1 found in the lactase promoter is identical or closely related.[12]

Bile acid transport genes

The molecular mechanisms underlying bile acid absorption, which is restricted to the distal small intestine, have been a recent focus of attention. Both membrane bound and cytosolic transport proteins have been identified, and the cytoplasmic ileal bile acid binding protein (IBABP) is of particular interest because of significant sequence homology with the fatty acid binding proteins. Sequences for IBABP are now known in several species and abundant, but restricted, IBABP mRNA expression in distal small intestine has been shown.[13 14] This restricted expression makes studies of the regulatory region of its gene particularly illuminating.[15]

Calcium transport genes

Two genes involved in calcium absorption, the cytoplasmic vitamin D dependent calcium binding protein calbindin-D9k and the plasma membrane calcium ATPase isoform 1 (PMCA1), have coordinated patterns of expression in proximal jejunum and duodenum, with suppression more distally in the jejunum and ileum.[16 17] This mimics the well established pattern of calcium absorption by whole tissue. Studies are investigating the transcriptional regulation of calbindin-D9k in rat and human.[18 19] Vitamin D responsive elements have been shown empirically in both species; DNase I footprint studies have identified several other regions of potential transcription factor binding. Although some of these regions are homologous to footprints in the SI genes, the basis of the quite different regional expression of these genes is not yet known. Homology also exists with the distal suppressor sequence in the intestinal FABP gene.[5] Studies of the PMCA1 promoter are also in progress, which should identify how this widely expressed, TATA-less gene is regulated in the intestinal regions.

Transcription factors in the intestine (Table)

An alternative way of approaching the problem of regional gene expression is to investigate which transcription factors are preferentially expressed in the intestine to discover if any show regional variation coinciding with that of the digestive and transport proteins. Many ubiquitous factors are also found in the intestine; these include steroid hormone receptors such as the thyroid hormone or vitamin D receptors but this discussion will focus on those which appear to be specific for intestine and other tissues derived from endoderm.

Hepatocyte nuclear factors

The beginning of the study of transcription factors in gastroenterology was in the mid-1980s, when work on DNA binding proteins in rat hepatocyte nuclear extracts began. In 1987, a nuclear protein, designated hepatocyte nuclear factor 1 (HNF-1), was identified[20] and shown to bind to a sequence required for hepatocyte specific transcription of the β-chain of fibrinogen. In transfection experiments with 5' deletion mutations of this promoter and in DNase I footprint studies, the key sequence was identified between bases −102 to −75 as ATTAAC; this was present also in the α fibrinogen and α_1 antitrypsin promoters. Purified HNF-1 protein[21] was found to interact with essential promoter regions of several other hepatic genes including albumin, α fetoprotein, and transthyretin. Cloning of HNF-1 identified it as a homeobox protein[22] (see below) placing it in a family of already well characterised transcription factors. HNF-1 is not, despite its name, liver specific. It is present in similar amounts in liver and kidney, and at substantially lower levels, in intestine, spleen, and thymus. Significantly, all these tissues may express hepatic markers. For example the kidney synthesises α and β fibrinogen, and α fetoprotein is expressed throughout the gastrointestinal tract, though restricted to enteroendocrine cells. There are two genes for HNF, HNF-1α and HNF-1β, and their expression was studied in the gut to find out if they colocalised with α fetoprotein[23]; these results showed that HNF-1α had substantially higher expression in ileum and jejunum, while HNF-1β was greatest in the colon. A crypt-villus gradient also was established for both these factors, with highest levels of expression in the intestinal crypts.

One of the CCAAT/enhancer binding proteins (C/EBP) was initially identified independently as HNF-2. C/EBPα, but no other member of this widely expressed family, is found in murine duodenal and jejunal villus enterocytes but is not present in crypts or in the ileum or colon.[24] Elements that bind C/EBPα have been identified in several intestinal genes.[3 18 24] Other hepatocyte nuclear factors, HNF-3 and HNF-4, are also found in intestine as well as

Summary of some transcription factors described in intestinal tissues

Transcription factors	*Abbreviation*	*Features*
Hepatocyte nuclear factor 1 α and β	HNF-1α, HNF-1β	Homeobox proteins Liver, kidney, intestine, etc
Hepatocyte nuclear factor 4	HNF-4	Zinc-finger protein Intestine, liver, and kidney
CCAAT-enhancer binding protein α	C/EBPα	Intestine, liver, etc (binds to the nucleotide sequence CCAAT)
Members of the homeobox gene clusters A, B, C, D	For example, Hox-A6, Hox-B3, Hox-C8	Widespread including intestine Important in gradients of developmental differentation
Caudal family proteins	Cdx-1	Homeobox protein Large and small intestine
	Cdx-2/Cdx-3	Identical homeobox proteins Large and small intestine
	PDX-1 (Idx-1/Ipf-1/Stf-1)	Identical homeobox proteins Duodenum and pancreas
Steroid hormone receptors (glucocorticoid, thyroid, etc)		Widespread members of the zinc-finger protein family
Other ubiquitous factors		Including those involved in responses to growth factors

in liver.[25 26] HNF-4 is a member of the steroid hormone receptor superfamily and is present in kidney and intestine as well as liver but is absent from other tissues. We have shown it to be distributed throughout the intestine with somewhat greater expression distally.[27] The gene for HNF-4 has been 'knocked out' in mice with profound effects on embryonic development, including that of endodermal tissues.[28]

Homeobox genes

Perhaps the best known group of DNA binding proteins are encoded by the homeobox genes. These were first identified in *Drosophila* where copious evidence shows that they control segmentation and body part development. The members of this large family share a 60 amino acid motif, termed the homeodomain or homeobox. This sequence has been shown to bind to DNA and is strikingly conserved even in evolutionarily remote species. In mammals, four clusters of Hox genes (A, B, C and D) are found on four different chromosomes. Other homeobox genes map outside the clusters of Hox genes. One of these in *Drosophila* is the gene known as caudal. Expression of caudal concentrates posteriorly and later is persistently expressed in the posterior midgut cells and Malphigian tubules in *Drosophila*. Well conserved caudal-like genes are present in other species and are increasingly being shown to be important in gastrointestinal differentiation.

The first mammalian caudal-related gene[29] was cloned in the mouse in 1988 and termed Cdx-1. Transcripts were localised to the epithelium of the intestine at 14 days gestation; the expression of Cdx-1 continued into adulthood. It could not be detected in any other tissues and is most abundant in colon. Studies of the transcriptional regulation of Cdx-1 are underway following cloning of its gene.[30] In a study of homeobox gene expression in the mouse using a PCR based strategy, a second member of the Cdx family, Cdx-2, was found and several known Hox genes with greater expression in colon compared with small intestine were identified.[31] One, now known as Hox-C8, was abundant only in ileum. Later, Cdx-3 was described in a hamster insulinoma cell line and demonstrated in intestine[32]; it is the homologue of Cdx-2 in that species. Cdx-4 was described in mouse embryonic endoderm but not adult intestine.[33]

Both Cdx-1 and Cdx-2 are present in proximal and distal areas of small intestine and at a somewhat higher level in colon.[31] The expression of Cdx-2 has been shown to be in all epithelial cells in the colon and ileum[34] and has been shown to interact with the sucrase-isomaltase SIF1 footprint[11] and with the rat calbindin-D9k promoter.[18] For these genes, at least, Cdx-2 is unlikely to account for their regional expression in small intestine.

The best candidate as the homeobox gene that promotes proximal small intestinal gene expression is the one now known as PDX-1 (formerly IDX-1 in rat,[35] and as IPF-1 or STF-1 in mouse.[36 37]) This is present in pancreas and duodenum, but not in jejunum, ileum or colon.[35] Human PDX-1 sequences are also abundant in duodenal but not ileal RNA.[38 39] When the gene was targeted in 'knock-out' mouse models, the pancreas and proximal duodenum failed to develop, and as the animals died shortly after birth, gene expression in adults could not be studied.[40 41] In situ hybridisation studies have demonstrated transcripts of PDX-1 in duodenal crypt epithelium, though further analysis of transcript levels along the intestine are needed to expand on the degree and extent of these proximal to distal differences.

The study of transcriptional regulation of gene expression in mammalian intestine is still very much in its infancy but current findings are consistent with the hypothesis that certain homeobox genes play a critical part in establishing regional identity in the gut. It seems likely that soon we will know the reasons why duodenum, jejunum, and ileum are what they are.

CHARLES J SHAW-SMITH
JULIAN R F WALTERS
Gastroenterology Unit, Royal Postgraduate Medical School, Hammersmith Hospital, Ducane Road, London W12 0NN

1 Walters JRF. Mechanisms of intestinal absorption. In: Weatherall D, Ledingham JCG, Warrell DA, eds. *Oxford Textbook of medicine.* 3rd ed. Oxford: Oxford University Press, 1995: 1899–904.
2 Rubin DC. Spatial analysis of transcriptional activation in fetal rat jejunal and ileal gut epithelium. *Am J Physiol* 1992; **263:** G853–63.
3 Cohn SM, Simon TC, Roth KA, Birkenmeier EH, Gordon JI. Use of transgenic mice to map cis-acting elements in the intestinal fatty acid binding protein gene (Fabpi) that control its cell lineage-specific and regional patterns of expression along the duodenal-colonic and crypt-villus axes of the gut epithelium. *J Cell Biol* 1992; **119:** 27–44.
4 Rottman JN, Gordon JI. Comparison of the patterns of expression of rat intestinal fatty acid binding protein/human growth hormone fusion genes in cultured intestinal epithelial cell lines and in the gut epithelium of transgenic mice. *J Biol Chem* 1993; **268:** 11994–2002.
5 Simon TC, Roberts LJJ, Gordon JI. A 20-nucleotide element in the intestinal fatty acid binding protein gene modulates its cell lineage-specific, differentiation-dependent, and cephalocaudal patterns of expression in transgenic mice. *Proc Natl Acad Sci USA* 1995; **92:** 8685–9.
6 Troelsen JT, Olsen J, Noren O, Sjostrom H. A novel intestinal trans-factor (NF-LPH1) interacts with the lactase-phlorizin hydrolase promoter and co-varies with the enzyme activity. *J Biol Chem* 1992; **267:** 20407–11.
7 Troelsen JT, Mehlum A, Olsen J, Spodsberg N, Hansen GH, Prydz H, *et al.* 1kb of the lactase-phlorizin hydrolase promoter directs post-weaning decline and small intestinal-specific expression in trangenic mice. *FEBS Lett* 1994; **342:** 291–6.
8 Duluc I, Freund JN, Leberquier C, Kedinger M. Fetal endoderm primarily holds the temporal and postional information required for mammalian intestinal development. *J Cell Biol* 1994; **126:** 211–21.
9 Wu GD, Wang W, Traber PG. Isolation and characterization of the human sucrase-isomaltase gene and demonstration of intestine-specific transcriptional elements. *J Biol Chem* 1992; **267:** 7863–70.
10 Traber PG, Wu GD, Wang W. Novel DNA-binding proteins regulate intestine-specific transcription of the sucrase-isomaltase gene. *Mol Cell Biol* 1992; **12:** 3614–27.
11 Suh E, Chen L, Taylor J, Traber PG. A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol Cell Biol* 1994; **14:** 7340–51.
12 Troelsen JT, Olsen J, Mitchelmore C, Hansen GH, Sjöström H, Norén O. Two intestinal specific unclear factors binding to the lactase-phlorizin hydrolase and sucrase-isomaltase promoters are functionally related oligomeric molecules. *FEBS Lett* 1994; **342:** 297–301.
13 Gong YZ, Everett ET, Schwartz DA, Norris JS, Wilson FA. Molecular cloning, tissue distribution, and expression of a 14-kDa bile acid-binding protein from rat ileal cytosol. *Proc Natl Acad Sci USA* 1994; **91:** 4741–5.
14 Shaw-Smith CJ, Howard A, Legon S, Walters JRF. Human ileal bile acid binding protein; partial sequence and tissue expression. *Gut* 1995; **37** (suppl 2): A17.
15 Crossman MW, Hauft SM, Gordon JI. The mouse ileal lipid-binding protein gene: a model for studying axial patterning during gut morphogenesis. *J Cell Biol* 1994; **126:** 1547–64.
16 Howard A, Legon S, Walters JRF. Human and rat intestinal plasma membrane calcium pump isoforms. *Am J Physiol* 1993; **265:** G917–25.
17 Freeman TC, Howard A, Bentsen BS, Legon S, Walters JRF. Cellular and regional expression of transcripts of the plasma membrane calcium pump PMCA1 in rabbit intestine. *Am J Physiol* 1995; **269:** G126–31.
18 Lambert M, Colnot S, Suh ER, L'Horset F, Blin C, Calliot ME, *et al.* Cis-acting elements and transcription factors involved in the intestinal specific expression of the rat calbindin-D9k gene. *Eur J Biochem* 1996; **236:** 778–88.
19 Howard A, Legon S, Moghrabi N, Walters JRF. Studies of the human calbindin-D9k gene. In: Norman AW, Bouillon R, Thomasset M, eds. *Vitamin D: gene regulation, structure-function analysis and clinical application.* Berlin: Walter de Gruyter, 1994: 396–7.
20 Courtois G, Morgan JG, Campbell LA, Fourel G, Crabtree GR. Interaction of a liver-specific nuclear factor with the fibrinogen and alpha 1-antitrypsin promoters. *Science* 1987; **238:** 688–92.
21 Courtois G, Baumheuter S, Crabtree G. Purified hepatocyte nuclear factor 1 interacts with a family of hepatocyte-specific factors. *Proc Natl Acad Sci USA* 1988; **85:** 7937–41.
22 Baumhueter S, Mendel DB, Conley PB, Kuo CJ, Turk C, Graves MK, *et al.* HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. *Genes Dev* 1990; **4:** 372–9.
23 Serfas MS, Tyner AL. HNF-1 alpha and HNF-1 beta expression in mouse intestinal crypts. *Am J Physiol* 1993; **265:** G506–13.
24 Chandrasekaran C, Gordon JI. Cell lineage-specific and differentiation-dependent patterns of CCAAT/enhancer binding protein alpha expression in the gut epithelium of normal and transgenic mice. *Proc Natl Acad Sci USA* 1993; **90:** 8871–5.
25 Lai E, Prezioso VR, Smith E, Livin O, Costa RH, Darnell JE. HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev* 1990; **4:** 1427–36.
26 Sladek FM, Zhong W, Lai E, Darnell JE. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev* 1990; **4:** 2353–65.

27 Shaw-Smith CJ, Poulsom R, Hanby AM, Howard A, Walters JRF. Distribution of transcription factor HNF-4 in human and rat tissue. *Clin Sci* 1996; **90:** 34P.
28 Chen WS, Manova K, Weinstein DC, Duncan SA, Plump AS, Prezioso VR, *et al.* Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Genes Dev* 1994; **8:** 2466–77.
29 Duprey P, Chowdhury K, Dressler GR, Balling R, Simon D, Guenet JL, *et al.* A mouse gene homologous to the Drosophila gene caudal is expressed in epithelial cells from the embryonic intestine. *Genes Dev* 1988; **2:** 1647–54.
30 Hu Y, James R, Kazenwadel J. Isolation and characterization of the murine homeobox gene Cdx-1. Regulation of expression in intestinal epithelial cells. *J Biol Chem* 1993; **268:** 27214–25.
31 James R, Kazenwadel J. Homeobox gene expression in the intestinal epithelium of adult mice. *J Biol Chem* 1991; **266:** 3246–51.
32 German MS, Wang J, Chadwick RB, Rutter WJ. Synergistic activation of the insulin gene by a LIM-homeodomain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. *Genes Dev* 1992; **6:** 2165–76.
33 Gamer LW, Wright CVE. Murine Cdx-4 bears striking similarities to the Drosophila caudal gene in its homeodomain sequence and early expression pattern. *Mech Dev* 1993; **43:** 71–81.
34 James E, Erler T, Kazenwadel J. Structure of the murine homeobox gene cdx-2. Expression in embryonic and adult intestinal epithelium. *J Biol Chem* 1994; **269:** 15229–37.
35 Miller CP, McGehee RE, Habener JF. IDX-1: a new homeodomain transcription factor expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene. *EMBO J* 1994; **13:** 1145–56.
36 Ohlsson H, Karlsson K, Edlund T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J* 1993; **12:** 4251–9.
37 Leonard J, Peers B, Johnson T, Ferreri K, Lee S, Montminy MR. Characterization of somatostatin transactivating factor-1, a novel homeobox factor that stimulates somatostatin expression in pancreatic islet cells. *Mol Endocrinol* 1993; **7:** 1275–83.
38 Walters JRF, Rumble HEE, Shaw-Smith CJ, Howard A, Legon S. Identification of human small intestinal homeodomain transcription factors. *Clin Sci* 1996; **90:** 19P.
39 Howard A, Rumble H, Prathalingram SR, Shaw-Smith CJ, Legon S, Walters JRF. Expression of homeobox transcription factors differs in human duodenum and ileum. *Gut* 1996; **38** (suppl 1): A33.
40 Jonsson J, Carlsson L, Edlund T, Edlund H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 1994; **371:** 606–9.
41 Offield MF, Jetton JL, Labosky PA, Ray M, Stein RW, Magnuson MA, *et al.* PDX-1 is required for pancreatic outgrowth and differentiation of the rostal duodenum. *Development* 1996; **122:** 983–95.