The EMBO Journal vol.7 no.9 pp.2705-2713, 1988

Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme

N.Mantei, M.Villa, T.Enzler, H.Wacker, W.Boll, P.James[1], W.Hunziker[2] and G.Semenza

Laboratorium für Biochemie II and [1]III der Eidgenössischen Technischen Hochschule, ETH-Zentrum, CH-8092 Zürich, Switzerland
[2]Present address: Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

Communicated by G.Semenza

We report the primary structures of human and rabbit brush border membrane β-glycosidase complexes (pre-pro-lactase-phlorizin hydrolase, or pre-pro-LPH, EC 3.2.1.23-62), as deduced from cDNA sequences. The human and rabbit primary translation products contain 1927 and 1926 amino acids respectively. Based on these data, as well as on peptide sequences and further biochemical data, we conclude that the proteins comprise five domains: (i) a cleaved signal sequence of 19 amino acids; (ii) a large 'pro' portion of 847 amino acids (rabbit), none of which appears in mature, membrane-bound LPH; (iii) the mature LPH, which contains both the lactase and phlorizin hydrolase activities in a single polypeptide chain; (iv) a membrane-spanning hydrophobic segment near the carboxy terminus, which serves as membrane anchor; and (v) a short hydrophilic segment at the carboxy terminus, which must be cytosolic (i.e. the protein has an $N_{out}-C_{in}$ orientation). The genes have a 4-fold internal homology, suggesting that they evolved by two cycles of partial gene duplication. This repetition also implies that parts of the 'pro' portion are very similar to parts of mature LPH, and hence that the 'pro' portion may be a water-soluble β-glycosidase with another cellular location than LPH. Our results have implications for the decline of LPH after weaning and for human adult-type alactasia, and for the evolutionary history of LPH.

Key words: cDNA cloning/lactase-phlorizin hydrolase/membrane anchoring/partial gene duplication/proteolytic processing

Introduction

Small-intestinal lactase has the distinction of being the enzyme involved in the most frequent genetically based syndrome in man, i.e. adult type alactasia (lactose intolerance in the adult) (Auricchio *et al.*, 1963; Dahlqvist *et al.*, 1963; for reviews see Kretchmer, 1971; Simoons, 1973; Semenza and Auricchio, 1988). This condition, a decline in lactase in adulthood to levels 5-10% of those at birth, affects one-third to one-half of mankind. Investigating the regulation of synthesis and degradation of lactase requires the preparation and sequencing of the corresponding cDNA, which would in addition provide much-needed information on the biochemistry and cell biology of this enzyme (see below).

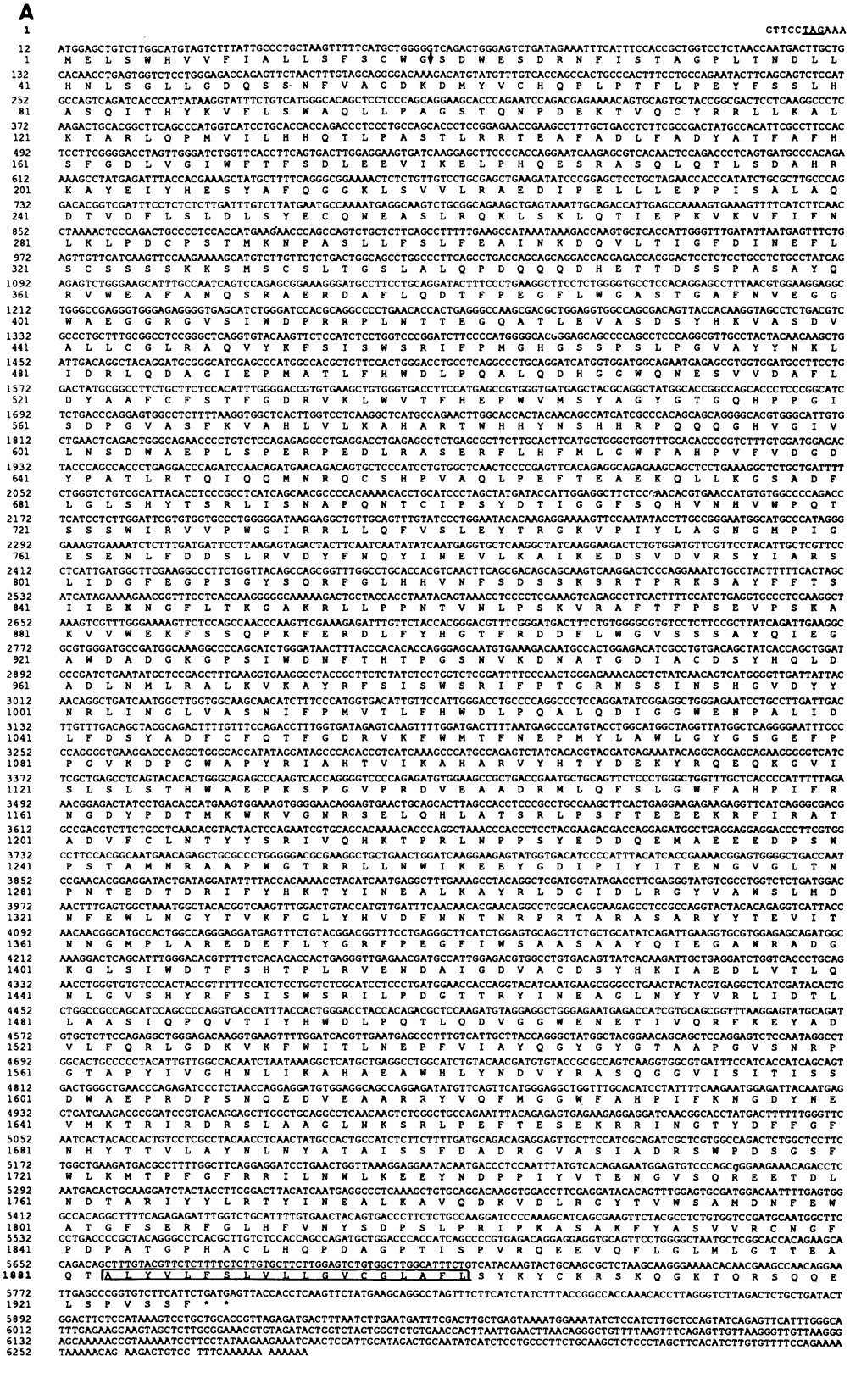
Lactase has been isolated from calf (Wallenfels and Fischer, 1960), rat (Schlegel-Haueter *et al.*, 1972), porcine (Skovbjerg *et al.*, 1982), simian (Ramaswamy and Radhakrishnan, 1975), human (Skovbjerg *et al.*, 1981; Potter *et al.*, 1985), and rabbit (H.Wacker *et al.*, in preparation) small intestine. The enzyme carries two activities; (i) lactase proper and (ii) an aryl- or alkyl-β-glucosidase ('phlorizin hydrolase') (Kraml *et al.*, 1972; Schlegel-Haueter *et al.*, 1972; Colombo *et al.*, 1973; Skovbjeg *et al.*, 1981), which is identical to glycosylceramidase (EC 3.2.1.45-46; Leese and Semenza, 1973). The two activities, which together are referred to as the β-glycosidase complex or lactase-phlorizin hydrolase (LPH; EC 3.2.1.23-62), can be distinguished from one another by their differential temperature sensitivity and by the very limited cross-inhibition by their respective substrates. These properties suggest the existence of two active sites, rather than a single one with broad substrate specificity. Evidence as to whether the two catalytic sites are located on a single polypeptide chain, or on two separate chains of identical size, has never been produced, although the former alternative is generally preferred. We show below that the LPH complex is, in fact, composed of a single type of polypeptide chain.

The complex is synthesized as a large single-chain precursor with apparent mol. wt 215 000-245 000 (Danielsen *et al.*, 1984; Skovbjerg *et al.*, 1984; Büller *et al.*, 1987; Naim *et al.*, 1987). It is not clear whether proteolytic processing occurs before (Naim *et al.*, 1987) or after (Danielsen *et al.*, 1984) complex glycosylation; this may be related to possible differences in the degree of compartmentalization of transglycosidases in the Golgi apparatus of enterocytes (Roth *et al.*, 1986; Taatjes *et al.*, 1988). The processing may be stepwise (215 000 to 160 000 to 130 000 in the rat; Büller *et al.*, 1987), or direct (215 000 to 160 000 in man; Naim *et al.*, 1987). When solubilized by detergents, 'final' LPH gives a single sharp band in SDS-PAGE under reducing conditions, with an apparent mol. wt of ~130 000-160 000. The native protein has an apparent mol. wt of ~320 000 by gel filtration (Skovbjerg *et al.*, 1982), suggesting that LPH dimerizes.

The catalytic mechanism of lactase has not yet been elucidated. However, the similarities with sucrase-isomaltase, e.g. identical pK_a'-values of the groups responsible for the pH optimum (Wallenfels and Fischer, 1960; Vasseur *et al.*, 1982 respectively), retention of configuration at C_1 of the product (see, for example, Semenza *et al.*, 1969), indicate a catalytic mechanism not unlike that of sucrase-isomaltase,

which involves an oxocarbonium ion temporarily stabilized by an Asp-COO^- (Cogoli and Semenza, 1975; Quaroni and Semenza, 1976; reviewed by Semenza, 1987).

We present here the cloning and complete sequences of cDNAs encoding both human and rabbit pre-pro-LPH. The deduced amino acid sequences suggest models (i) for the first steps in biosynthesis and membrane insertion; (ii) for the extensive proteolytic processing of the pre-pro-LPH into two or more polypeptides (of which only one, i.e. the 'final' LPH, is targeted to the brush border membrane); (iii) for the mode of anchoring of the complex to the brush border membrane; and (iv) for the location of both enzymatic activities in the single-chain LPH complex. They also allow inferences as to the evolutionary history of LPH.



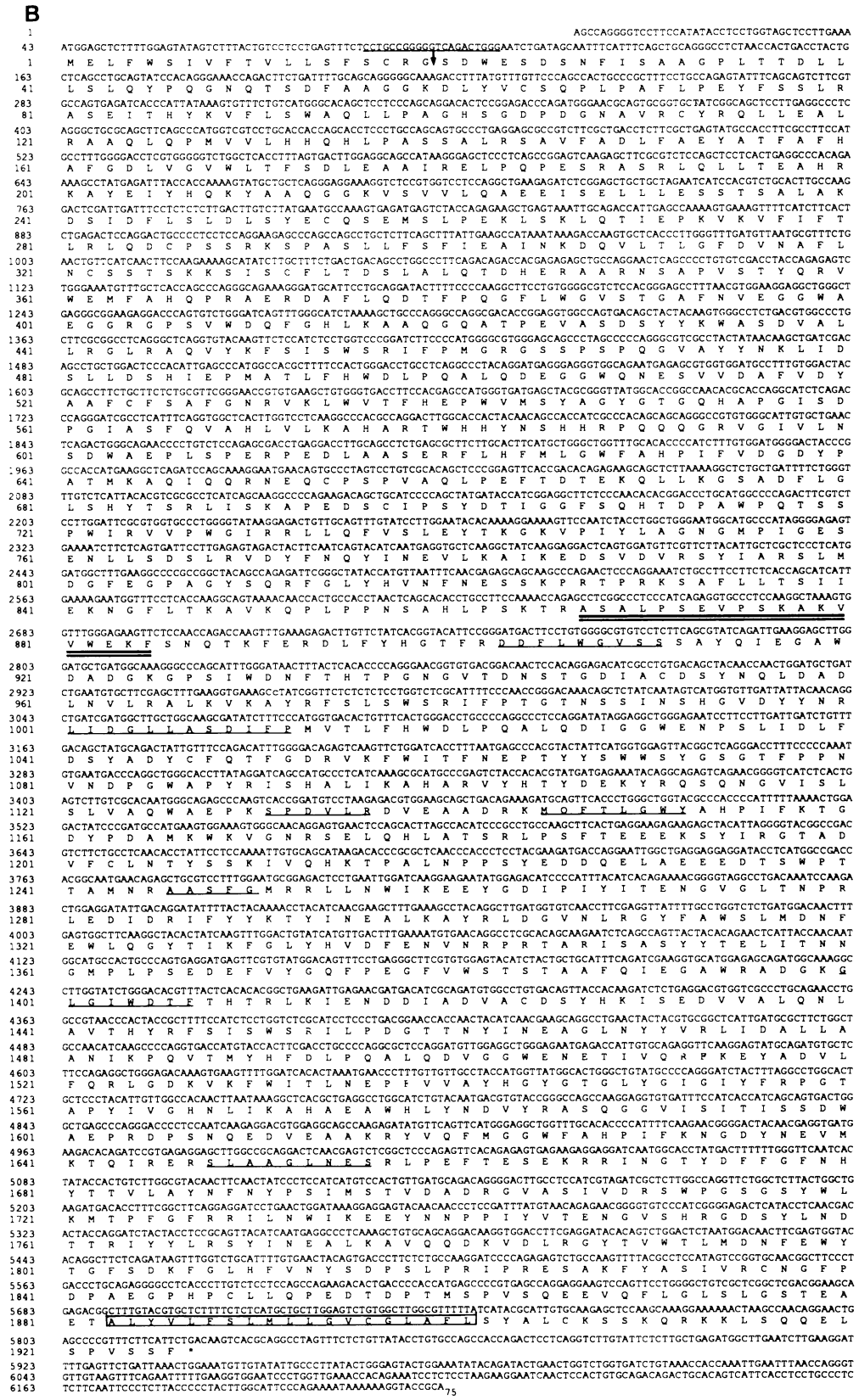


Fig. 1. Sequences of pre-pro-LPH cDNAs and deduced amino acid sequences. (**A**) Human sequence from plasmids pHLac-61 and pHLac-5. The stop codon preceding the initiaton codon is underlined. The vertical arrow indicates the putative signal peptidase cleavage site. The hydrophobic segment near the carboxy terminus is boxed. (**B**) Rabbit sequence from plasmids pRLac-8 and pRLac-122. The underlined DNA sequence, nucleotides 88−109, is the complement of the primer used for the experiment shown in Figure 3. The doubly underlined amino acid sequence (residues 867−885) is that determined by direct amino acid sequencing of untreated mature LPH. Singly underlined peptides are those found by sequencing tryptic peptides. Vertical arrow and box as in (**A**). Potential N-glycosylation sites (N-X-S/T): in human proLPH, 42, 368, 418, 512, 821, 934, 946, 989, 1174, 1340, 1508, 1656, 1672, 1761 and 1814; in rabbit proLPH, 49, 321, 510, 822, 887, 932, 944, 987, 1033 (X = P, weak site), 1172, 1506, 1654, 1670, 1759 and 1813.

Results and discussion

Purification and partial sequence of rabbit LPH

LPH as purified from rabbit brush border membranes by antibody affinity chromatography showed a major band corresponding to mol. wt 135 000 in SDS gel electrophoresis, plus a band of lower mobility thought to represent pre-pro-LPH. Sequence analysis of this material (without further fragmentation) on a gas phase sequencer yielded a sequence of 19 amino acids (later identified with residues 867–885 in Figure 1B).

Cloning and characterization of LPH cDNAs

The partial amino acid sequence was used to design an oligonucleotide probe, which comprised a 64-fold degenerate pool of oligonucleotides, 17 residues in length and coding for amino acids 880–885 (except for the last nucleotide of codon 885) of the sequence in Figure 1B. This probe was used to screen a rabbit intestinal mucosa cDNA bank. Two positive clones were shown by DNA sequencing to code for the whole sequence of 19 amino acids. (Additional peptide sequences are discussed below.) To identify more certainly these clones as coding for LPH, a fragment derived from one of them was used as a probe in a Northern blot analysis of RNAs from various rabbit tissues. As shown in Figure 2A, the probe hybridizes to an RNA of ~6 kb, found in small intestine (lane 1), but not in liver, lung or kidney (lanes 2–4). This is the tissue distribution expected for LPH, and the relatively large size corresponds to that required to code for a protein of mol. wt 220 000. The controls in Figure 2B show hybridization of the same blot with cDNA coding for another small intestinal brush border protein, sucrase–isomaltase (Hunziker *et al.*, 1986) (giving rise to the band at 6 kb), and with tubulin cDNA, whose cognate RNA (1.8 kb) is expected to be found in all tissues.

Full-length clones of both rabbit and human LPH cDNA were isolated from libraries of >4 kb intestinal cDNA. Various cDNAs were sequenced partially or completely. For human, the sequence of 6274 nt (Figure 1A) is from clones pHLac-5, pHLac-61 and pHLac-1 (pHLac-5 was sequenced completely and contains all but 273 nucleotides at the 5′ end and 17 nt at the 3′ end). For rabbit, the sequence of 6221 nt (Figure 1B) is derived from clone pRLac-8 and pRLac-122. (pRLac-8 is complete except for a deletion of one nucleotide at position 3183; this nucleotide is present in pRLac-122 and is necessary to preserve the reading frame and the alignment with the human sequence.)

The human sequence includes 11 and 482 nt of 5′ and 3′ untranslated region respectively. Nucleotides 12–5792 show a single open reading frame coding for a peptide of 1927 amino acids. The methionine codon starting at nucleotide 12 must be the beginning of the coding region, because it is preceded two codons upstream by an in-frame termination codon (Figure 1A). The rabbit sequence comprises 42 nt of 5′ untranslated region, a coding region specifying a peptide of 1926 amino acids, and a 3′ untranslated region of 401 nt. To assess the completeness of the rabbit sequence at the 5′ end, the primer indicated in Figure 1B was hybridized to rabbit intestinal poly(A)$^+$ RNA and then extended with reverse transcriptase. As shown in Figure 3, essentially no products are found to extend further upstream than the end of clone RLac-8, save for a trace corresponding to a product extending about another 23 nt (which may be an artefact or may correspond to a 'minor upstream start' of transcription; for discussion of the latter see Frampton *et al.*, 1987). Further, 15 additional rabbit cDNA clones were isolated using a probe comprising the 138 nt at the 5′ end of the RLac-8 sequence. None of the cDNAs so obtained extended further in the 5′ direction than RLac-8. On the basis of these results and the strong homology to the human sequence, we conclude that the complete coding region is also present in the rabbit clone. Both sequences end in a poly(A) segment, preceded 20–25 nucleotides upstream by the 'AATAAA' sequence typically

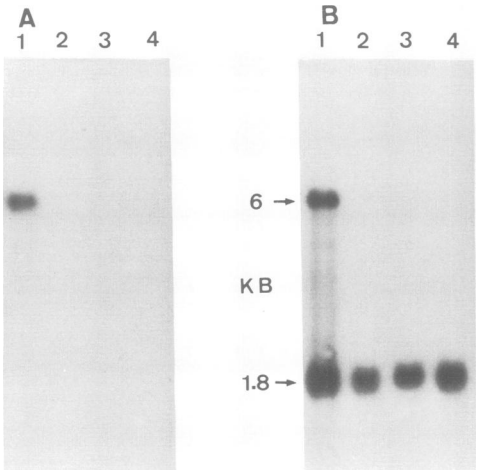


Fig. 2. Northern blot analysis of rabbit RNAs. (**A**) RNAs extracted from (**1**) small intestine, (**2**) liver, (**3**) lung and (**4**) kidney were hybridized with a DNA fragment comprising nucleotides 2603–2788 of pRLac-8. (**B**) The same blot, after removal of the LPH probe, was rehybridized with labeled sucrase–isomaltase (Hunziker *et al.*, 1986) and rabbit tubulin cDNAs. The sizes indicated are the known sizes of sucrase–isomaltase and tubulin mRNAs.

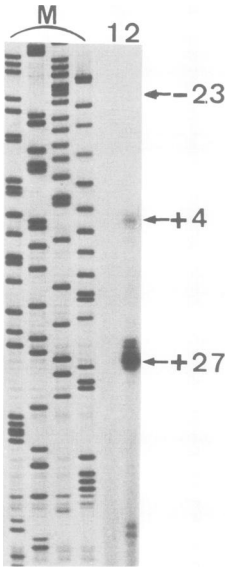


Fig. 3. Primer-extension analysis of rabbit LPH mRNA. An amount of size-selected (see Materials and methods) rabbit intestinal mucosa RNA corresponding to 2 μg total poly(A)$^+$ RNA was hybridized to 0.4 pmol of the primer shown in Figure 1B. The primer was elongated with reverse transcriptase at 50°C (**lane 2**) or incubated without enzyme (**lane 1**). **Lanes M**, M13mp19 sequencing ladder as marker: from left to right G, A, T, C. The numbers give the approximate endpoints of the products in the coordinates of the pRLac-8 sequence (Figure 1B).

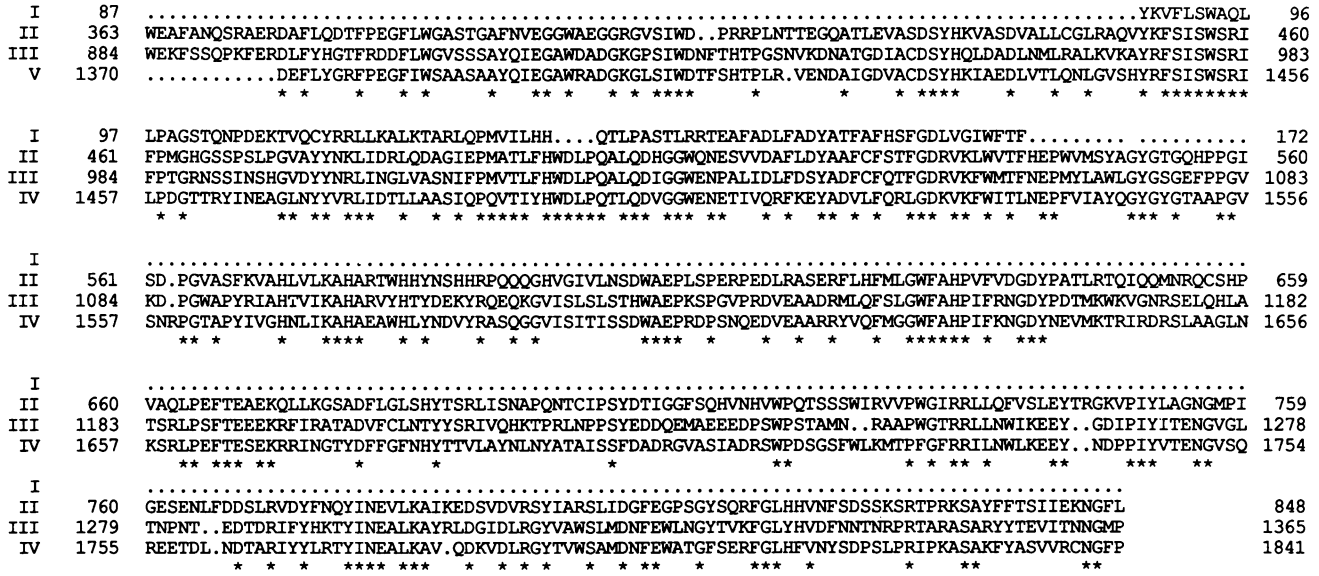


Fig. 4. Internal homologies within the human pro-LPH amino acid sequence. Regions I (residues 87–172), II (363–848), III (884–1365) and IV (1370–1841) of the human pre-pro-LPH sequence were aligned as described in Materials and methods. Residues which match in three or more sequences are indicated by asterisks.

found in this position (Proudfoot and Brownlee, 1976); the 3′ untranslated regions are therefore also complete.

The deduced primary structures of the human and rabbit pre-pro-LPH complexes are reported in Figure 1. As the two sequences share 83% identical amino acids (and 85% identical nucleotides), they will be discussed together.

Characteristics of the proteins

In the following, we will often make reference to the similarly long (1827 amino acids) sequence of another intestinal disaccharidase complex, that of pro-sucrase–isomaltase (Hunziker *et al.*, 1986). Pre-pro-LPH is processed proteolytically into fragments which separate from one another (see below). Pro-sucrase–isomaltase, apart from losing the initiator methionine, is split into two unequal subunits which remain associated (the 'sucrase–isomaltase complex', or SI).

The cleavable signal. Amino acids 1–19 (Figure 1) have in both human and rabbit pre-pro-LPH the characteristics of a cleaved signal peptide (von Heijne, 1985, 1986): a polar N-terminal region is followed by a short hydrophobic segment and then by a 'C' region, which is thought to determine the exact site of signal peptidase cleavage. The weight matrix method of von Heijne (1986) predicts that cleavage is very likely between amino acids 19 and 20 (scores of 9.8 and 8.2 for human and rabbit respectively, well in the range of scores calculated for known cleaved signal sequences; see Figure 2 in von Heijne, 1986). In contrast, the hydrophobic sequence of pro-SI is much longer (20 amino acids), and the maximal cleavage site score in this area (4.2 for cleavage between residues 26 and 27) is lower than the scores for almost all known signalase sites. The hydrophobic region of pro-SI is in fact part of an uncleaved signal and also serves to anchor SI in the membrane (Ghersa *et al.*, 1986; Hunziker *et al.*, 1986).

Mature LPH. As mentioned above, the amino-terminal sequence determined for mature rabbit LPH solubilized from brush border membranes corresponds to amino acids 867–885 in the deduced total amino acid sequence of rabbit pre-pro-LPH. We also determined the sequences of seven tryptic peptides derived from mature LPH. All seven sequences (underlined in Figure 1B) occur downstream from position 866. We conclude that of all the 1926 amino acids in pre-pro-LPH, only those from Ala-867 onwards occur in the brush border membrane as LPH. In other words, the final LPH is composed of a single polypeptide species. A notable feature in the sequence of LPH is its internal homology, shown (for the human sequence) as regions III and IV in Figure 4. Note that this repetition does not include either the signal or the hydrophobic segment in the C-terminal region (see below). In LPH the two active sites (see Introduction) must therefore reside on a single polypeptide chain. This situation differs from that found with sucrase–isomaltase, where only the pro-form contains both activities in one polypeptide chain.

Table I. Release of brush border hydrolase activities from intestinal brush border membrane vesicles upon incubation with phosphatidylinositol-specific phospholipase C

	Phospholipase C (U/ml)				
	0[a]	0	0.2	0.4	1.0
Sucrase	0.0	0.1	0.1	0.2	0.2
Lactase	0.0	5.9	6.8	7.4	7.4
Trehalase	0.3	0.6	27.1	32.4	39.1
Alkaline phosphatase	0.0	0.2	13.2	15.9	19.2

Rabbit brush border membrane vesicles (3 mg protein/ml) in 50 mM $NaPO_4$, pH 7.0, 1 mM EDTA, 0.02% NaN_3, were incubated for 1 h at 37°C with 0, 0.2, 0.4 and 1.0 U/ml phospholipase respectively. After centrifugation (200 000 *g*, 15 min) enzyme activities were determined in the supernatants, and are expressed in percentage of activities before centrifugation.
[a]Without incubation.

The membrane anchor of LPH and implications for membrane targeting. The segments comprising residues 1883–1901 of both human and rabbit LPH are highly

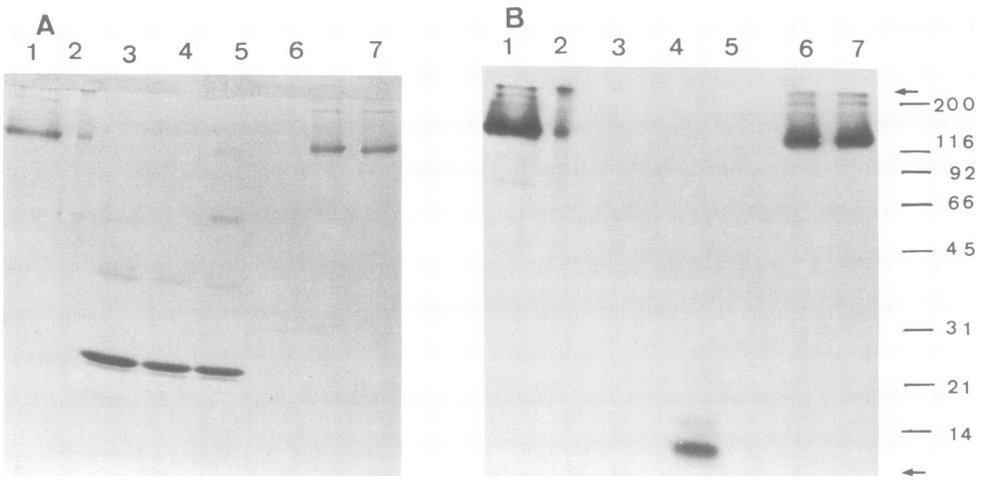


Fig. 5. SDS−PAGE of [^{125}I]TID-LPH. LPH was labeled with the hydrophobic photoaffinity reagent TID, subjected to treatment with nitrous acid, papain or phospholipase, and then analyzed by SDS−PAGE.
(**A**) Coomassie blue staining. (**B**) Autoradiography. **Lane 1**, LPH after nitrous acid treatment; **lane 2**, control without nitrous acid; **lanes 3** and **4**, papain, as used to digest LPH; **lane 5**, LPH after papain treatment; **lane 6**, LPH after treatment with phosphoinositol-specific phospholipase C; **lane 7**, control without phospholipase. The numbers at the side are mol. wts of marker proteins in thousands. Arrows indicate the start and front.

hydrophobic, are sufficiently long to cross a lipid bilayer in a helical conformation, and are the only hydrophobic stretches of such length occurring in the proteins. We thus identify these as the hydrophobic anchors of LPH.

Some brush border hydrolases—alkaline phosphatase, trehalase, dipeptidase and aminopeptidase P—are known to be anchored via phosphatidylinositol (PI) (Ikezawa *et al.*, 1976; Takesue *et al.*, 1986; Hooper *et al.*, 1987a; Hooper and Turner, 1988a). Prior to formation of the PI anchor, proteins of this kind are apparently associated with the endoplasmic reticulum membrane through a hydrophobic amino acid sequence, which is split off from the rest during the transfer of most of the polypeptide to the PI anchor (reviewed by Low, 1987; Ferguson and Williams, 1988). One may wonder, therefore, whether the hydrophobic sequence in pro-LPH may not be such a temporary anchor. This is not so, however, as LPH proved not be anchored through PI: the enzyme is not solubilized by PI-specific phospholipase (Table I); it is readily solublized by the detergent Triton X-100 (which is indicative of it not being anchored by PI; see Hooper and Turner, 1988b). Further, when Triton-solubilized LPH is labeled with the hydrophobic photoreagent [^{125}I]TID (Brunner and Semenza, 1981), the label is retained even after the action of PI-specific phospholipase or nitrous acid (Figure 5), either of which would be expected to remove a PI anchor. All the evidence, therefore, trends to rule out PI as a possible membrane anchor, and points to the hydrophobic sequence of residues 1883−1901 as serving this function. As this sequence is too short to span the membrane twice, we have to place the C terminus on the cytosolic side of the brush border membrane.

Apart from the suggestion put forward for the angiotensin-converting enzyme (Hooper *et al.*, 1987b), LPH is the first brush border hydrolase found to be anchored via a hydrophobic amino acid stretch in the $C_{in}-N_{out}$ orientation. Sucrase−isomaltase (Semenza, 1979; Hunziker *et al.*, 1986), maltase−glucoamylase (Norén *et al.*, 1986) and a number of peptidases, including endopeptidase 24.11 (Devault *et al.*, 1987; Malfroy *et al.*, 1987), dipeptidyl-peptidase IV (MacNair and Kenny, 1979; Booth and Kenny, 1980), aminopeptidase N (Booth and Kenny, 1980; Norén and Sjöström, 1980; Ferracci *et al.*, 1982) and γ-glutamyl-transpeptidase (Frielle and Curthoys, 1983; Matsuda *et al.*, 1983; Laperche *et al.*, 1986; Olsen *et al.*, submitted), are known to be, or are likely to be, anchored near the N terminus. As mentioned above, a few more microvillar enzymes are anchored at the C terminus, but via phosphatidylinositol rather than hydrophobic peptides. (The transfer to the PI anchor most probably already takes place in the endoplasmic reticulum, see Low, 1987). Thus, the nature of the hydrophobic anchors, or the size, the polarity (and even the very existence) of the cytosolic domains cannot play an essential role in membrane targeting, i.e. in deciding which proteins are sent to the brush border membrane and which to the basolateral membrane in the polarized enterocytes.

The catalytic sites of LPH. H.Wacker *et al.* (in preparation) have recently labeled the active sites in mature LPH using the affinity label conduritol-B-epoxide. Two mols of label were bound per mol of LPH, each site being partially protected from inactivation by the respective substrate. This provides further evidence that mature LPH contains both lactase and phlorizin hydrolase active sites in a single polypeptide chain.

The 'pro' portion of LPH. Residues 20−866 (which we will designate 'XβGly'), comprising >40% of pro-LPH, must be split from mature LPH during intracellular processing. [This fits well with results obtained by pulse labeling in organ cultures of pig intestine (Danielsen *et al.*, 1984). However, at that time this interpretation was considered unlikely, as no fragment corresponding to XβGly was detected.] In contrast, pro-SI and mature SI have the same mass (see, for example, Hauri *et al.*, 1980; Sjöström *et al.*, 1980; Hu *et al.*, 1987). The separation of XβGly from mature LPH may occur through splitting of one or a few peptide bonds immediately upstream of position 867, as suggested for human LPH by Naim *et al.* (1987), or by cleavage both here and around residue 500, as suggested by the results of Büller

et al. (1987) with the rat enzyme.

Within XβGly, the segment designated 'II' in Figure 4 shows especially strong similarity to segments III and IV, which comprise mature LPH. We surmise, therefore, that XβGly may be a glycosidase, probably of the β-type.

The lack of hydrophobic sequence of any length in XβGly indicates, but does not prove, that it is not targeted to a membrane. A number of possibilities can be envisaged. (i) It could be secreted into the intestinal lumen—indeed, a 'soluble' lactase occurs in the succus entericus (see, for example, Aramayo *et al.*, 1983). At present, however, it cannot be excluded that this soluble lactase originates from the 'usual' brush border LPH via solubilization by the proteases of the intestinal lumen. (ii) XβGly could be secreted into the blood. (iii) It could be targeted to some intracellular compartment, e.g. to the lysosomes, which are known to contain α- and β-glycosidases. No similarity was found with them, nor with any other sequence in the NBRF database, nor with glucocerebrosidase (Sorge *et al.*, 1985). (iv) The enterocytes are known to be endowed with a cytosolic (?) β-galactosidase (Asp and Dahlqvist, 1968), which has, however, not yet been well characterized. (v) Finally, XβGly might be rapidly degraded. Whatever the destiny of XβGly, it is clear that the same translational unit gives rise to at least two polypeptides, XβGly and LPH, which are targeted to different cellular locations.

Possible physiological role(s) of the proteolytic processing of pro-LPH to LPH

This processing may serve one or more physiological functions. For example, it may be necessary for LPH (and perhaps XβGly) to fold in an enzymatically active conformation; to date, however, we have no information on whether pro-LPH is enzymatically active. It is also possible that this proteolytic processing may be play some role in the sorting and eventual delivery of LPH to the brush border membrane (e.g. by allowing the LPH to oligomerize while still in the ER membranes). However, leupeptin, which inhibits most if not all of the proteolytic processing, does not prevent a large mol. wt precursor from reaching the brush border membrane (Danielsen *et al.*, 1984).

The homing and perhaps the proteolytic processing and/or glycosylation associated with it may well be involved in the decline of lactase activity which takes place after weaning. Nsi Emvo *et al.* (1987), have reported the presence in the enterocytes of adult rats of a 300-kd, high-mannose, high-fucose polypeptide, which is devoid of lactase activity but immunologically cross-reacts with LPH. This 300-kd polypeptide is present only in traces in the enterocytes of suckling rats, which of course have high lactase activity. Administration of thyroxine to baby rats leads to a sharp decline in lactase, with simultaneous increase of the enzymatically inactive 300-kd polypeptide. These observations are compatible with the idea that the post-weaning decline in intestinal LPH, which occurs in the vast majority of mammals, is related to the processing—and homing—of this 300-kd polypeptide, which in all likelihood can be equated with an enzymatically inactive pro-LPH.

This regulatory scheme is also attractive in light of recent observations (G.Sebastio *et al.*, 1988; personal communication) that the decline of lactase in adult rabbits is not accompanied by a parallel decline of lactase mRNA, thus placing the key regulatory event at the translational or post-translational level.

The post-weaning decline of LPH occurring in most mammals is very similar to adult-type alactasia, i.e. to the genetic condition leading to lactose intolerance in man. It is thus possible that adult-type alactasia may be related to an alteration of the processing of pro-LPH to LPH, which would lead to decreased appearance of LPH in the brush border membrane.,

Internal homology

The internal homology noted above for LPH is also found in XβGly and, indeed, in pro-LPH as a whole (Figure 4). For human pro-LPH four homologous regions are apparent, comprising approximately residues 87–172, 363–848, 883–1365 and 1370–1841. In pairwise comparisons they contain 38–55% identical residues (with 1–5 small gaps). Essentially the same regions can be found in rabbit pro-LPH, with 35–50% identical residues (not shown).

Evolution of pro-LPH

Save for their large size and similar catalytic mechanism, pro-SI and pre-pro-LPH are quite different proteins. They differ in having a cleavable (pre-pro-LPH) versus an uncleavable (pro-SI) signal, in the direction their polypepitde chains are inserted into the membrane, in the location of their hydrophobic anchors and in their different proteolytic processing. Their regulation (developmental, hormonal, and dietary controls, reviewed, for example, by Koldovský, 1981; Henning, 1985; Semenza and Auricchio, 1988) are also different. No sequence similarities are evident at either the DNA or amino acid sequence level. The similarities in the catalytic mechanisms and in the membrane targeting are thus indicative of a convergent rather than a divergent evolution.

Pro-SI shows a double inner homology (Hunziker *et al.*, 1986) as predicted by the 'two sites, one chain precursor mechanism' (discussed in its ontogenetic and evolutionary aspects elsewhere by Hunziker *et al.*, 1986; Semenza, 1986). Pre-pro-LPH has a 4-fold inner homology (Figure 4), with regions I and II assigned to XβGly and III and IV to LPH. These repeats, which include neither the cleavable signal nor the hydrophobic anchor of LPH, indicate that a partial gene duplication occurred twice.

Whereas lactase activity is confined to mammals, aryl-β-glucosidase ('phlorizin hydrolase') activity has been found in the intestines of all vertebrates where it has been sought (reviewed by Semenza, 1981). The 'natural' substrates of this enzyme, the glycosyl ceramides (Leese and Semenza, 1973), in fact occur in the diet of most vertebrates. It is reasonable to suppose, therefore, that phlorizin hydrolase is the phylogenetic progenitor of both catalytic sites in the LPH complex (regions III and IV in the 4-fold homology of pre-pro-LPH).

The homologies between XβGly and LPH indicate an additional interesting possibility. It is conceivable that XβGly (probably a water-soluble protein of still unknown final location) may have arisen through gene duplication from LPH, a brush border enzyme. The converse is also possible, i.e. that LPH may have arisen from XβGly with the further addition of the hydrophobic membrane anchor, and also, perhaps, of the 'address' targeting it to the brush border. This interesting relationship between a brush border and a lysosomal enzyme (if XβGly is lysosomal) is similar to, but not identical with, that suggested by the work of Hoefsloot *et al.* (1988): lysosomal α-glucosidase is homologous to the

brush border sucrase–isomaltase complex. These enzymes do not, however, belong to the same translational unit. Clearly, more information on the cellular location of XβGly and its biology is badly needed.

Materials and methods

Enzyme purification and sequencing

Brush border membranes from 9-day-old rabbits were prepared using precipitation with Mg^{2+}/EGTA (Hauser *et al.*, 1980). LPH was solubilized with Triton X-100 and purified over columns of Sephacryl S-400, DEAE-cellulose, Sephadex G-200 and anti-aminopeptidase-M-Sepharose 4B respectively, as described elsewhere (H.Wacker *et al.*, in preparation). The purified LPH showed one band in SDS–PAGE and was used for raising antibodies in guinea-pigs. The IgG fraction was purified from the serum by chromatography with Affi Gel Blue CM (Bio-Rad) followed by precipitaton with ammonium sulfate (Gee *et al.*, 1979), and was coupled to BrCN-activated Sepharose 4B (Cuatrecasas *et al.*, 1969). This affinity matrix was used for the purification of LPH from the brush border membrane of adult rabbits after solubilization with Triton X-100. After application of the Triton extract the column was washed extensively with high salt buffer (500 mM NaCl) containing 1% Triton, and the LPH was eluted with 1 mM EDTA, 0.1% Triton, yielding 25 nmol protein. About 0.7 nmol was used directly for sequencing. A further 4 nmol were reduced and alkylated by treatment with 10 mM DTT for 2 h at 37°C followed by 15 mM iodoacetic acid for 4 h at 4°C. After precipitation with 15% trichloroacetic acid, the pellet was washed twice with 50 mM ammonium bicarbonate, resuspended by sonication and adjusted to pH 8.0. Protein was digested by trypsin at a ratio of 50:1 w/w for 24 h at 37°C. Peptides were purified by reverse-phase HPLC on a Brownlee RP300 C-8 column eluted with a gradient of 0–50% iso-propanol in 0.1% heptafluorobutyric acid. Amino acid sequences were determined using an Applied Biosystems 470A gas phase sequencer with a Model 120 on-line PTH amino acid detection system. LPH was also chemically deglycosylated (Edge *et al.*, 1981) and sequenced directly and after tryptic digestion.

cDNA synthesis and cloning

RNA was purified from adult rabbit and human small intestinal mucosa using a urea–lithium chloride method (Auffray and Rougeon, 1980). cDNA was prepared from poly(A)$^+$ RNA according to Gubler and Hoffman (1983) and cloned via *Eco*RI linkers. For the first library, rabbit cDNA was cloned into plasmid p91023B (Wong *et al.*, 1985), and DNA prepared from 225 pools of 100–200 clones. These pools were screened (Wood *et al.*, 1985) using a dot-blot procedure with the oligonucleotide mixture described in the text as probe. Bacteria from positive pools were subcloned and single positive plasmids isolated. Two positive clones (pRLac-16 and pRLac-122) were identified by DNA sequencing. To prepare a second library, poly(A)$^+$ RNA from human and rabbit mucosa was fractionated by sucrose gradient centrifugation, the peak fractions located by dot blot hybridization with a labeled fragment of pRLac-16 DNA, and cDNA prepared as before. A fraction ~4–8 kb in size was isolated by electrophoresis in low gelling temperature agarose and cloned in λZAP bacteriophage (Strategene). About 2% of the rabbit clones and 0.02% of the human clones were positive when screened by hybridization with labeled pRLac-122 insert. Plasmid rescue was carried out by co-transfection with λZAP and helper phage as described in the supplier's protocol. We noted a tendency for deleted plasmids to arise upon preparing larger amounts of plasmid; we recommend retransfecting plasmid DNA from a small-scale preparation to eliminate possible contaminating helper phage RF DNA.

Sequencing and analysis

Fragments produced by cleavage of cDNAs with *Alu*I, *Pal*I, *Rsa*I and *Sau*3a (separately) were cloned (Struhl, 1985) into M13mp18 or M13mp19 (Norrander *et al.*, 1983) and sequenced using the dideoxy method (Sanger *et al.*, 1977) with the Sequenase (US Biochemicals) reagents and protocol. Various other fragments, suggested by the preliminary data, were sequenced to resolve doubtful regions and to complete all the sequences on both strands. 'Compressions' could be alleviated by running sequencing gels in a 65°C oven. We sequenced a total of ~60 kb.

DNA and protein sequences were assembled and analyzed with the Genetics Computer Group program package (Devereux *et al.*, 1984) on a VAX 8700. The regions aligned as shown in Figure 4 were first identified with a 'dot plot' comparing the sequence to itself. The alignment was carried out using the program 'Gap' interactively (Needleman and Wunsch, 1970). The appropriateness of the gaps was also determined visually; as they are few in number, the alignment shown is very likely optimal. Signal peptidase cleavage sites were predicted with the weight matrix method of von Heijne (1986) with a program written for the Macintosh computer.

RNA analysis

For blot analysis, ~6 μg each poly(A)$^+$ RNA was run on a 1.2% agarose gel with 1 M formaldehyde and transferred to Zeta-Probe (Bio-Rad) membrane (Thomas, 1983). The membrane was irradiated with UV (Khandjian, 1983) before hybridization with labeled DNA (Feinberg and Vogelstein, 1983). The filter was washed in 0.3 × SSC, 0.1% SDS at 65°C. Primer extension analysis was performed by annealing an end-labeled primer with size-fractionated RNA (see above), and elongating with AMV reverse transcriptase (Pharmacia) at 50°C (Geliebter *et al.*, 1986).

Labeling with [^{125}I]TID

LPH, 0.6 nmol at 1 mg/ml in 0.1% Triton X-100, was equilibrated with [^{125}I]TID and illuminated under nitrogen for 1 min with a 350-W medium-pressure mercury lamp (Spiess *et al.*, 1982). Labeled LPH was separated from labeled Triton and photolysis products by gel filtration on a column (2.0 × 40 cm) of Sepharose 4B equilibrated and eluted with buffer containing 2% sodium cholate. LPH was separated from the bulk of the radioactivity and concentrated by filtration (Amicon, PM-10). Sodium cholate was removed by dialysis against buffer containing 0.1% Triton X-100.

Degradation of TID-labeled LPH

Incubation with papain. TID-labeled LPH (0.1 nmol) was incubated under nitrogen for 30 min at 37°C with 39 μg of preactivated (0.05 μmol cysteine) papain. The reaction was terminated by the addition of 0.1 μmol of iodoacetamide followed by boiling in SDS sample buffer.

Incubation with phosphoinositol-specific phospholipase C. TID-labeled LPH (0.1 nmol) was incubated (30 min, 37°C) with 250 mU of *B.thuringiensis* lipase (Hooper *et al.*, 1987). The reaction was stopped by boiling in SDS sample buffer.

Degradation with nitrous acid (Ferguson et al., 1985). TID-labeled LPH (0.1 nmol, dialyzed against sodium acetate, pH 3.5), was incubated with 160 mM $NaNO_3$ in 25 mM acetate, pH 3.5, for 4.5 h at 23°C; the reaction mixture was then neutralized with NaOH and boiled in SDS sample buffer.

SDS–PAGE. Electrophoresis was performed using a discontinuous sulfate–borate system (Neville, 1971) as described in Wacker *et al.* (1981).

Acknowledgements

We thank Ms B.O'Neill for technical assistance, Professor F.Lagiardèr (Universität Zürich) and an anonymous organ donor for a sample of human small intestine, Ms A.von Kieckebush-Gück for rabbit kidney, lung and liver RNA samples, Dr M.Low for phosphatidylinositol-specific phospholipase C, and Dr S.Moroney and Mr A.Müller for oligonucleotides. We thank O.Norén and H.Sjöström for important contributions at the beginning of this work. Supported by the Swiss National Science Foundation, Berne.

References

Aramayo,L.A., De Silva,D.G.H., Hughes,C.A., Brown,G.A. and McNeish,A.S. (1983) *Arch. Dis. Childhood,* **58**, 686–691.
Asp,N.G. and Dahlqvist,A. (1968) *Biochem. J.*, **106**, 841–845.
Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.
Auricchio,S., Rubino,A., Landolt,M., Semenza,G. and Prader,A. (1963) *Lancet,* **ii**, 324–326.
Berger,J., Garattini,E., Hua,J.-C. and Udenfriend,S. (1987) *Proc. Natl. Acad. Sci. USA,* **84**, 695–698.
Booth,A.G. and Kenny,A.J. (1980) *Biochem. J.*, **187**, 31–44.
Brunner,J. and Semenza,G. (1981) *Biochemistry,* **20**, 7174–7182.
Büller,H.A., Montgomery,R.K., Sasak,W.V. and Grand,R.J. (1987) *J. Biol. Chem.*, **262**, 17206–17211.
Cogoli,A. and Semenza,G. (1975) *J. Biol. Chem.*, **250**, 7802–7809.
Colombo,V., Lorenz-Meyer,H. and Semenza,G. (1973) *Biochim. Biophys. Acta,* **327**, 412–424.
Cuatrecasas,P., Wilchek,M. and Anfinsen,C.B. (1969) *Proc. Natl. Acad. Sci. USA,* **61**, 636–643.
Dahlqvist,A., Hammond,J.B., Crane,R.K., Dunphy,J.V. and Littman,A. (1963) *Gastroenterology,* **45**, 488–491.
Danielsen,E.M., Skovbjerg,H., Norén,O. and Sjöström,H. (1984) *Biochem. Biophys. Res. Commun.*, **122**, 82–90.
Devault,A., Lazure,C., Nault,C., Moual,H.L., Seidah,N.G., Chrétien,M.,

Kahn,P., Powell,J., Mallet,J., Beaumont,A., Roques,B.P., Crine,P. and Boileau,G. (1987) *EMBO J.*, **6**, 1317–1322.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Edge,A.S.B., Faltynek,C.R., Hof,L., Reichert,L.E. and Weber,P. (1981) *Anal. Biochem.*, **118**, 131–137.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
Ferguson,M.A.J., Low,M.G. and Cross,G.A.M. (1985) *J. Biol. Chem.*, **260**, 14547–14555.
Ferguson,M.A.J. and Williams,J. (1988) *Annu. Rev. Biochem.*, in press.
Ferracci,J., Maroux,S., Bonical,J. and Desnuelle,P. (1982) *Biochim. Biophys. Acta*, **684**, 133–136.
Frielle,T. and Curthoys,N.P. (1983) *Biochemistry*, **22**, 5709–5714.
Frampton,J., Conkie,D., Chambers,I., McBain,W., Dexter,M. and Harrison,P. (1987) *Nucleic Acids Res.*, **15**, 3671–3688.
Gee,A.P., Borsos,T. and Boyle,M.D.P. (1979) *J. Immunol. Methods*, **30**, 119–126.
Geliebter,J., Zeff,R.A., Melvold,R.W. and Nathenson,S.G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3371–3375.
Ghersa,P., Huber,P., Semenza,G. and Wacker,H. (1986) *J. Biol. Chem.*, **261**, 7969–7974.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Hauri,H.-P., Quaroni,A. and Isselbacher,K.J. (1980) *Proc. Natl Acad. Sci. USA*, **77**, 6629–6633.
Hauser,H., Howell,K., Dawson,R.M.C. and Bower,D.E. (1980) *Biochim. Biophys. Acta*, **602**, 567–577.
Henning,S.J. (1985) *Annu. Rev. Physiol.*, **47**, 231–245.
Hoefsloot,L.H., Hoogeven-Westerveld,M., Kroos,M.A., van Beeumen,J., Reuser,A.J.J. and Oostra,B.A. (1988) *EMBO J.*, **7**, 1697–1704.
Hooper,N.M., Low,M.G. and Turner,A.J. (1987a) *Biochem. J.*, **244**, 465–469.
Hooper,N.M., Keen,J., Pappin,D.J.C. and Turner,A.J. (1987b) *Biochem. J.*, **247**, 85–93.
Hooper,N.M. and Turner,A.J. (1988a) *FEBS Lett.*, **229**, 340–344.
Hooper,N.M. and Turner,A.J. (1988b) *Biochem. J.*, **250**, 865–869.
Hu,Ch.-B., Spiess,M. and Semenza,G. (1987) *Biochim. Biophys. Acta*, **896**, 275–286.
Hunziker,W., Spiess,M., Semenza,G. and Lodish,H.F. (1986) *Cell*, **46**, 227–234.
Ikezawa,H., Yamanegi,M., Taguchi,R., Miyashita,T. and Ohyabu,T. (1976) *Biochim. Biophys. Acta*, **450**, 154–164.
Khandjian,E.W. (1986) *Mol. Biol. Rep.*, **11**, 107–115.
Koldovský, O. (1981) In Randle,J.P., Steiner,D.F. and Whelan,W.J. (eds), *Carbohydrate Metabolism and its Disorders*. Academic Press, London, Vol. 3, pp. 418–522.
Kraml,J., Kolínská,J., Ellederova,H. and Hiršová,D. (1972) *Biochim. Biophys. Acta*, **258**, 520–530.
Kretchmer,N. (1971) *Gastroenterology*, **61**, 805–813.
Laperche,Y., Bulle,F., Aissani,T., Chobert,M.-N., Aggerbeck,M., Hanoune,J. and Guellaën,G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 837–941.
Leese,H.J. and Semenza,G. (1973) *J. Biol. Chem.*, **248**, 8170–8173.
Low,M.G. (1987) *Biochem. J.*, **244**, 1–13.
MacNair,R.D. and Kenny,A.J. (1979) *Biochem. J.*, **179**, 379–395.
Malfroy,B., Schofield,P.R., Kuang,W.-J., Seeburg,P.H., Mason,A.J. and Henzel,W.J. (1987) *Biochem. Biophys. Res. Commun.*, **144**, 59–66.
Matsuda,Y., Tsuji,A. and Katunuma,N. (1983) *J. Biochem.*, **93**, 1427–1433.
Naim,H.Y., Sterchi,E.E. and Lentze,M.G. (1987) *Biochem. J.*, **241**, 427–434.
Needleman,S.B. and Wunsch,C.D. (1970) *J. Mol. Biol.*, **48**, 443–453.
Neville,D.M.,Jr (1971) *J. Biol. Chem.*, **246**, 6328–6334.
Norén,O. and Sjöström,H. (1980) *Eur. J. Biochem.*, **104**, 25–31.
Norén,O., Sjöström,H., Cowell,G.M., Tranum-Jensen,J., Hansen,O.C. and Welinder,K.G. (1986) *J. Biol. Chem.*, **261**, 12306–12309.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.
Nsi-Emvo,E., Launay,J.-F. and Raul,F. (1987) *Cell. Mol. Biol.*, **33**, 335–344.
Potter,J., Ho,M.-W., Bolton,H., Furth,A.J., Swallow,D.M. and Griffiths,B. (1985) *Biochem. Genet.*, **23**, 423–439.
Proudfoot,N. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Quaroni,A. and Semenza,G. (1976) *J. Biol. Chem.*, **251**, 3250–3253.
Ramaswamy,S. and Radhakrishnan,A.N. (1975) *Biochim. Biophys. Acta*, **403**, 446–455.
Roth,J., Taatjes,D.J., Weinstein,J., Paulson,J.C., Greenwell,P. and Watkins,W.M. (1986) *J. Biol. Chem.*, **261**, 14307–14312.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5468.
Schlegel-Haueter,S., Hore,P., Kerry,K.R. and Semenza,G. (1972) *Biochim. Biophys. Acta*, **258**, 506–519.
Semenza,G. (1979) In Elliott,K. and Whelan,J. (eds), *Development of Mammalian Absorptive Processes*, Ciba Foundation Symp., Excerpta Medica, Amsterdam, Vol. 70, pp. 133–144.
Semenza,G. (1981) In Randle,J.P., Steiner,D.F. and Whelan,W. (eds), *Carbohydrate Metabolism and its Disorders*. Academic Press, London, pp. 425–479.
Semenza,G. (1986) *Annu. Rev. Cell Biol.*, **2**, 255–313.
Semenza,G. (1987) In Kenny,A.J. and Turner,A.J. (eds), *Mammalian Ectoenzymes*. Elsevier, Amsterdam. pp. 265–287.
Semenza,G. and Auricchio,S. (1988) In Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle,D. (eds), *The Metabolic Basis of Inherited Disease*. McGraw-Hill, New York. 6th edn, in press.
Semenza,G., Curtius,H.C., Raunhardt,O., Hore,P. and Müller,M. (1969) *Carbohydr. Res.*, **10**, 417–428.
Simoons,F.J. (1973) *Am. J. Dig. Dis.*, **18**, 595–611.
Skovbjerg,H., Danielsen,E.M., Norén,O. and Sjöström,H. (1984) *Biochim. Biophys. Acta*, **798**, 247–251.
Skovbjerg,H., Sjöström,H. and Norén,O. (1981) *Eur. J. Biochem.*, **114**, 653–661.
Skovbjerg,H., Norén,O., Sjöström,H., Danielsen,E.M. and Enevoldsen,B.S. (1982) *Biochem. Biophys. Acta*, **707**, 89–97.
Sorge,J., West,C., Westwood,B. and Beutler,E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7289–7293.
Spiess,M., Brunner,J. and Semenza,G. (1982) *J. Biol. Chem.*, **257**, 2370–2377.
Struhl,K. (1985) *BioTechniques*, **3**, 452–453.
Taatjes,D.J., Roth,J., Weinstein,J. and Paulson,J.C. (1988) *J. Biol. Chem.*, **263**, 6302–6309.
Takesue,Y., Yokota,K., Nishi,Y., Taguchi,R. and Ikezawa,H. (1986) *FEBS Lett.*, **201**, 5–8.
Thomas,P. (1983) *Methods Enzymol.*, **100**, 255–266.
Vasseur,M., Tellier,K.C. and Alvarado,F. (1982) *Arch. Biochem. Biophys.*, **218**, 263–274.
von Heijne,G. (1985) *J. Mol. Biol.*, **184**, 99–105.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.
Wacker,H., Jaussi,R., Sonderegger,P., Dokow,M., Ghersa,P., Hauri,H.P., Christen,P. and Semenza,G. (1981) *FEBS Lett.*, **136**, 329–332.
Wallenfels,K. and Fischer,J. (1960) *Hoppe Seyler's Z. Physiol. Chem.*, **321**, 223–245.
Wong,G.G., Witek,J.S., Temple,P.A., Wilkens,K.M., Leary,A.C., Luxenberg,D.P., Jones,S.S., Brown,E.L., Kay,R.M., Orr,E.C., Shoemaker,C., Golde,D.W., Kaufman,R.J., Hewick,R.M., Wang,E.A. and Clark,S.C. (1985) *Science*, **228**, 810–815.
Wood,W.I., Gitschier,J., Lasky,L.A. and Lawn,R.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1585–1588.

Received on May 30, 1988

Note added in proof

These sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession numbers X07994 (HUMLPH) and X07995 (RABLPH).