

RTMP's copy

REPLACED SHEETS		SG
☒ Description	: paged 1 to paged 33	
☒ Claim(s)	: paged 34 to paged 37	
☒ Drawing(s)	: paged 4 to paged 4	
☒ Abstract	: paged 38 to paged	

Aktiebolaget Astra
Södertälje, Sweden

Inventors: Gunnar Bjursell, Lars Bläckberg, Peter Carlsson,
Sven Enerbäck, Olle Hernell, Jeanette Nilsson,
Thomas Olivecrona

HX 1067-1
1991-05-29
IH/IB

New Chemical Products

Field of the invention

The present invention relates to new DNA sequences, to new
5 proteins coded for by such DNA sequences, and to the use
as further described below of such proteins. The invention
also encompasses vectors, such as plasmid constructs,
comprising such DNA sequences, being capable of expressing
the desired enzyme. The invention also includes host
10 organisms transfected with such constructs e.g. bacteria
yeast, mammalian cells, and transgenic animals. The
invention also includes processes for the preparation of
the novel products of the invention. The new proteins of
the invention are related to an enzyme, known i.a. as
15 human bile salt-stimulated lipase.

20

25

30

35

This Sheet Has Been Replaced



Background to the invention

The human lactating mammary gland synthesizes and secretes
5 with the milk a bile salt-stimulated lipase (BSSL) [1]
that, after specific activation by primary bile salts [2,
57, 58], contributes to the breast-fed infant's endogenous
capacity of intestinal fat digestion [3-5]. This enzyme,
10 [6], is a non specific lipase; in vitro it hydrolyses not
only tri-, di- and monoacylglycerols, but also
cholesteryl-, and retinyl esters, and
lysophosphatidylglycerols [7-10]. Furthermore, its
activity is not restricted to emulsified substrates, but
15 micellar and soluble substrates are hydrolyzed at similar
rates [11].

BSSL is not degraded during passage with the milk through
the stomach, and in duodenal contents it is protected by
bile salts from inactivation by pancreatic proteases such
20 as trypsin and chymotrypsin [2,11]. It is, however,
inactivated when the milk is pasteurized, e.g. heated to
62.5 °C, 30 min [12]. Model experiments in vitro suggest
that the end products of triacylglycerol digestion are
different in the presence of BSSL [5,7]. Due to lower
25 intraluminal bile salt concentrations during the neonatal
period [13,14] this may be beneficial to product
absorption [5,15].

The carboxylic ester hydrolase (CEH) of human pancreatic
juice [16] seems functionally to be identical, or at least
30 very similar, to BSSL [8]. They also share common epitopes
[8,17], have identical N-terminal amino acid sequences
[17] and are inhibited by inhibitors of serine esterases,
e.g. eserine and diisopropylfluorophosphate [6,8,16]. It
has been hypothesized that the two enzymes are products of
35 the same gene [18,19]. The observed molecular size
difference [8,19] could be explained by different patterns
of glycosylation, as recently suggested [17].



Dietary lipids is an important source of energy. The energy-rich triacylglycerols constitute more than 95% of these lipids. Some of the lipids, e.g. certain fatty acids and the fat soluble vitamins, are essential dietary constituents. Before gastro-intestinal absorption the triacylglycerols as well as the minor components, i.e. esterified fat-soluble vitamins and cholesterol, and diacylphosphatidylglycerols, require hydrolysis of the ester bonds to give rise to less hydrophobic, absorbable products. These reactions are catalyzed by a specific group of enzymes called lipases.

In the human adult the essential lipases involved are considered to be gastric lipase, pancreatic colipase-dependent lipase (tri- and diacylglycerol hydrolysis), pancreatic phospholipase A2 (diacylphosphatidylglycerols) and carboxylic ester hydrolase (cholesteryl- and fat soluble vitamin esters). In the breast-fed newborn bile salt-stimulated lipase plays an essential part in the hydrolysis of several of the above mentioned lipids. Together with bile salts the products of lipid digestion form mixed micelles from which absorption occurs (3-5).

Common causes of lipid malabsorption, and hence malnutrition, are reduced intraluminal levels of pancreatic colipase-dependent lipase an/or bile salts. Typical examples of such lipase deficiency are patients suffering from cystic fibrosis, a common genetic disorder resulting in a life-long deficiency in som 80% of the patients, and chronic pancreatitis, often due to chronic alcoholism.

The pancreatic and liver functions are not fully developed at birth, most notably in infants born before term. Fat malabsorption, for physiological reasons, is a common finding and thought to result from low intraluminal pancreatic colipase-dependent lipase and bile salt concentrations (3,4,13-15). However, because of BSSL, such



malabsorption is much less frequent in breast-fed infants than in infants fed pasteurized human milk of infant formulas (3-5, 12, 59, 60, 61). This is one reason why it has been advocated that newborn infants, particularly
5 preterm infants, that cannot be fed their own mothers milk should be fed non-pasteurized milk from other mothers (12).

The present treatment of patients suffering from a
10 deficiency of pancreatic lipase is the oral administration of very large doses of a crude preparation of porcine pancreatic enzymes. Colipase-dependent pancreatic lipase is inactivated by low pH. Such conditions are prevalent in the stomach, with the result that orally administered
15 pancreatic lipase is virtually completely inactivated on the passage through the stomach to the gut. Therefore, this effect cannot be completely overcome by the use of large doses of enzyme. The large doses administered are inadequate for most patients, and the preparations are
20 impure and unpalatable. Certain tablets have been formulated which pass through the acid regions of the stomach and discharge the enzyme only in the relatively alkaline environment of the jejunum. However, many patients suffering from pancreatic disorders have an
25 abnormally acid jejunum and such tablets may fail to discharge the enzyme and may therefore be ineffective. Moreover, since the preparations presently on the market are of a non-human source there is a risk of immunoreactions that may cause harmful effects to the
30 patients or result in reduced therapy efficiency. A further drawback with the present preparations is that their content of other lipolytic activities than colipase-dependent lipase are not stated. In fact, most of them contain very low levels of CEH/BSSL-activity. This may be
35 one reason why many patients, suffering from cystic fibrosis in spite of supplementation therapy, suffer from deficiencies of fat soluble vitamins and essential fatty



acids.

Thus, there is a great need for products with properties and structure derived from human lipases and with a broad substrate specificity, which products may be orally
5 administered to patients suffering from deficiency of one or several of the pancreatic lipolytic enzymes. The products of the present invention fulfil this need by themselves or in combination with other lipases or in combination with preparations containing other lipases.
10 Furthermore, for some human infants there is an obvious need to improve fat utilization from conventional infant formulas, or pasteurized human milk from so-called milk banks.

BSSL has several unique properties that makes it ideally
15 suited for substitution and supplementation therapy: It has been designed by nature for oral administration. Thus, it resists passage through the stomach and is activated in contents of the small intestine.

Its specific activation mechanism should prevent
20 hazardous lipolysis of food or tissue lipids during storage and passage to its site of action.

Due to its broad substrate specificity it has the potential to, on its own, mediate complete digestion of most dietary lipids, including the fat soluble vitamin
25 esters.

BSSL may be superior to pancreatic colipase-dependent lipase to hydrolyze ester bonds containing long-chain polyunsaturated fatty acids.

In the presence of gastric lipase and in the absence of,
30 or at low levels of colipase-dependent lipase BSSL can ascertain a complete triacylglycerol digestion in vitro even if the bile salt levels are low such as in newborn infants. In the presence of BSSL the end products of triacylglycerol digestion become free fatty acids and free
35 glycerol rather than free fatty acids and monoacylglycerol generated by the other two lipases (5). This may favour product absorption particularly when the intraluminal bile



salt levels are low (3,15).

From a historical point of view infant formulas have been developed, and improved, from the concept that their composition should be as similar to that of human milk as possible. It is desirable to supplement such formulas.

The utilization for supplementation, substitution or therapy of bile salt-stimulated lipases (BSSL), or of proteins with the essential functions of BSSL, requires however access to quantities of the product on a large technical scale. It is not possible in factory scale to rely on natural sources such as milk as starting material. Besides the problem mentioned above with inactivation of BSSL during pasteurization, there is the additional risk of contamination of material from a natural source with infectious agents, e.g. vira such as HIV virus and CMV. There is, accordingly, a need for large scale access to products having BSSL properties. The present invention provides such products and methods for their preparation.

20

Prior art references are given later in this specification.

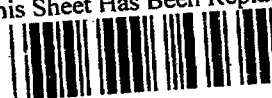
This Sheet Has Been Replaced



The invention

- The present invention is based on the cloning of cDNA coding for BSSL derived from human mammary gland. We have
- 5 also isolated, from human pancreas, a partial cDNA coding for CEH. Deduced amino acid sequences from the human cDNA's and comparison with CEH from other species, support the interpretation that BSSL and CEH are identical.
- 10 As will be further detailed below, it was surprisingly found that the structure of the protein as deduced from the cDNA sequence is quite different from the structure of other lipases. The structure proved unexpectedly to be more like the structure of typical esterases, such as
- 15 cholinesterase.

This Sheet Has Been Replaced



With reference to Figure II and Figure VII, products of the invention are:

- 5 a) a protein as defined by the amino acid sequence 1-722 in Fig. VII,
- b) a protein as defined by the amino acid sequence 1-535 in Fig. VII,
- 10 c) a protein as defined by the amino acid sequence 1-278 in Fig. VII,
- d) a protein as defined by the amino acid sequence 1-341 in Fig. VII,
- 15 e) a protein as defined by the amino acid sequence 1-409 in Fig. VII,
- f) a protein as defined by the amino acid sequence 1-474 in Fig. VII,
- 20 g) combinations of proteins defined under b) - f) e.g. as defined by the amino acid sequence in positions 1-278, 279-341, 279-409, 279-474, 342-409, 342-474 and 536-722.
- 25 h) combinations of proteins defined under b) - g) in combination with one or more of the repeats according to Figure V,
- 30 i) a protein as defined under a) - h) possessing an additional, N-terminal amino acid, namely methionine, and functionally equivalent variants and mutants of the proteins defined in a) - i) above;
- 35 It should be noted that the proteins under a, b, c, d, e, f, h and i above will not be identical in all respects to naturally occurring BSSL, but they will exhibit one or



more of the critical functions of naturally occurring BSSL. Critical functions are given below.

j) a DNA sequence coding for the proteins defined in a,
5 b, c, d, e, f, h and i above.

k) a DNA sequence according to Fig. II, defined by the following nucleotide numbers in Fig. II:

10

a) a DNA sequence 151-2316 according to Fig. II, coding for the protein defined by the amino acid sequence 1-722 in Fig. VII,

b) a DNA sequence 151-1755 according to Fig. II, coding
15 for the protein defined by the amino acid sequence 1-535 in Fig. VII,

c) a DNA sequence 151-985 according to Fig. II, coding for the protein defined by the amino acid sequence 1-278 in Fig. VII,

20 d) a DNA sequence 151-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 1-341 in Fig. VII,

e) a DNA sequence 151-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 1-409
25 in Fig. VII,

f) a DNA sequence 151-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 1-474 in Fig. VII,

g) a DNA sequence 986-1172 according to Fig. II, coding
30 for the protein defined by the amino acid sequence 279-341 in Fig. VII,

h) a DNA sequence 986-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 279-409 in Fig. VII,

35 i) a DNA sequence 986-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 279-474 in Fig. VII,

This Sheet Has Been Replaced



- j) a DNA sequence 1173-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 342-409 in Fig. VII,
- k) a DNA sequence 1173-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 342-474 in Fig. VII.

Significant functions of proteins of the invention are

- a) suitable for oral administration,
- b) being activated by specific bile salts,
- c) acting as a non-specific lipase in the contents of the small intestines, that is being able to hydrolyze lipids relatively independent of their chemical structure and physical state (emulsified, micellar, soluble).
- d) Ability to hydrolyze triacylglycerols with fatty acids of different chain-length and different degree of unsaturation.
- e) Ability to hydrolyze also diacylglycerol, monoacylglycerol, cholesteryl esters, lysophosphatidylacylglycerol, and retinyl and other fat soluble vitamin-esters.
- f) Ability to hydrolyze not only the sn-1(3) ester bonds in a triacylglycerol but also the sn-2 ester bond.
- g) Ability to interact with not only primary but also secondary bile salts.
- h) Dependency on bile salts for optimal activity.
- i) Stability so that gastric contents will not affect the catalytical efficiency to any substantial degree.



- j) Stability towards inactivation by pancreatic proteases, e.g. trypsin, provided bile salts are present.
- 5 k) Ability to bind to heparin and heparin derivatives, e.g. heparan sulphate.
- l) Ability to bind to lipid-water interphases.
- 10 m) Stability to permit lyophilization.
- n) Stability when mixed with food constituents such as in human milk, or milk formula.
- 15 The critical functions for supplementation, substitution, or therapy are these according to a), c), d), e), f), i), j) and l). For other purposes, not all critical functions may be necessary.
- 20 For expression of the proteins indicated above, the appropriate DNA sequence indicated above will be inserted into a suitable vector which then is introduced into a suitable host organism. The said vector will also have to comprise appropriate signal and other sequences enabling
- 25 the organism to express the desired protein.

Suitable expression organisms:

- With the recombinant DNA techniques it is possible to
- 30 clone and express a protein of interest in a variety of prokaryotic and eukaryotic host organisms. Possible expression organisms are bacteria, simple eukaryotes (yeast), animal cell cultures, insect cell cultures, plant cell cultures, plants and transgenic animals. Each
- 35 individual system has its own particular advantages and disadvantages. The simple conclusion is that every gene to be expressed is a unique problem and no standard solution



is available.

Commonly used bacterial systems are E. Coli, Bacillus subtilis, Streptomyces. Commonly used yeasts are
 5 saccharomyces, and Pichia Pastoris. Commonly used animal cells are CHO cells and COS cells. Commonly used insect cell cultures are Drosophila derived cells.

Commonly used plant is the tobacco plant. Possible
 10 transgenic animals are goat and cow.

Possible bacterial vectors are exemplified by pUC and protein A-vectors.

Possible yeast vector is exemplified by pMA 91.

15 Possible insect vectors are devired from Baculo-virus.

Possible animal cell vectors are derived from SV/40.

Possible plant vectors are derived from the Ti-plasmid.

In every system, both natural and synthetic promoters and terminators can be used.

20

It is understood that depending on the choice of expression system, the expressed protein may contain an additional N-terminal amino acid (methionine), contain a few extra amino acids, or be fused to a heterologous
 25 protein, (e.g. protein A), or differ from the naturally occurring protein with respect to glycosylation.

Furthermore, the vectors may also contain signal sequences in order to export the protein to the periplasm or to the culture medium.

30 Thus, further aspects of the invention are:

- a) a vector comprising a DNA sequence coding for a protein as specified above,
- b) a host organism comprising a DNA sequence as specified above,
- 35 c) a process for the production of a protein as specified above, by growing a host organism containing a vector as specified under a) above and isolating the protein.

This Sheet Has Been Replaced



Methods from purification are based on the expression system used (e.g. protein A/IgG) and/or on methods used for purification of the naturally occurring enzyme, as described in reference 6.

5

Additional aspects of the invention are:

- a pharmaceutical composition comprising a protein as specified above,

10

- the use of a protein as specified above for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency,

15

- the use of a protein as specified above for the manufacture of a medicament for the treatment of cystic fibrosis,

20 - the use of a protein as specified above as a supplement to an infant food formulation,

-the use of a protein as specified above for the manufacture of a medicament for the treatment of chronic
25 pancreatitis,

-the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption of any etiology,

30

-the use of a protein as specified above for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins,

35 -the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons, e.g. in new-



born infants.

The DNA sequence in Fig. II from position 151 up to and including position 2316 is the sequence coding for the entire protein. The sequence from position 2317 up to and including position 2415 is not translated to protein, but is included in exon d identified in Table 2 below.

In one embodiment of the invention, the protein as defined in paragraphs a) - i) above is provided in isolated form and/or in substantially pure form.

The DNA sequences as defined in paragraphs a) - k) above are in one embodiment of the invention provided in isolated form and/or in substantially pure form.

Experimental part

20 Abbreviations

aa, amino acid; bp, base pair; BSSL, bile salt-stimulated lipase; c-AMP, cyclic adenosine monophosphate; CEH, carboxylic ester hydrolase; Da, dalton; c ³GTP, 7-deaza-2-deoxyguanosine 5'triphosphate; EDTA, ethylene diamine tetraacetate; kb, kilobases; MOPS, 3-N-morpholino-propanesulfonic acid; nt, nucleotide; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; SSC, NaCl citrate, xGal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

30

Enzymes

Bile salt-stimulated lipase EC 3.1.1.3

Carboxylic ester hydrolase EC 3.1.1.1

35

This Sheet Has Been Replaced



Material and Methods

A. Enzyme and antibody preparation

- 5 BSSL was purified from human milk as previously described [6]. When used for antibody production the enzyme was further purified by SDS-PAGE. The protein band corresponding to the lipase was, after staining with Coomassie Brilliant blue, electroeluted from the gel.
- 10 Twentyfive μg of purified enzyme, together with an equal volume of Freund's complete adjuvant, was used for a first i.c. injection and the same amount of enzyme with incomplete adjuvant for the subsequent monthly booster injections. The rabbits were bled about two weeks after
- 15 each booster and sera prepared and stored at $-20\text{ }^{\circ}\text{C}$.

B. Preparation of tryptic fragments and amino acid sequence analysis

- 20 Three mg of purified BSSL was dissolved in 1 ml of 0.1M Tris-Cl buffer, pH 8.5, containing 6M guanidinium hydrochloride and 2 mM EDTA. Dithioerythritol was added to 5mM. After incubation at $37\text{ }^{\circ}\text{C}$ for 2h, 300 μl 50 mM iodoacetate was added. After 90 min incubation at $25\text{ }^{\circ}\text{C}$ in
- 25 darkness the reduced and carboxymethylated enzyme was desalted on a Sephadex G-25 column, equilibrated with 0.5M ammonium bicarbonate. Thirty μg of tosyl-L-phenylalanine chloromethane treated bovine trypsin (Worthington diagnostics system Inc., Freehold, NJ, USA)
- 30 was added before lyophilization. The lyophilized protein was dissolved in 4 ml 0.1M ammonium bicarbonate and an additional 90 μg of trypsin was added. After 5h incubation at $37\text{ }^{\circ}\text{C}$ the protein was again lyophilized. The tryptic digest was dissolved in 0.1% trifluoroacetic acid
- 35 (2mg/ml). Three hundred μg of trypsinated BSSL was chromatographed on HPLC using a C-18 reversed phase column and eluted with a gradient of 0-50% acetonitrile in 0.1%



trifluoroacetic acid. Peptide collection was monitored by continuous recording of the absorbance at 215 nm. Peptides to be sequenced were further purified by rechromatography using the same column with adjusted gradients.

- 5 Samples of peptide fragments to be sequenced were dried under nitrogen to remove acetonitrile and applied to the sequencer. For N-terminal sequence analysis, native BSSL was dissolved in 0.1% acetic acid. Sequence analyzes were performed on an Applied Biosystems Inc. 477A pulsed
10 liquid-phase sequencer and on-line PTH 120A analyzer with regular cycle programs and chemicals from the manufacturer. Initial and repetitive yields, calculated from a sequenced standard protein, β -lactoglobulin, were 47 and 97%, respectively.

15

C. Isolation of RNA

- Samples of human pancreatic adipose and lactating mammary gland tissues were obtained at surgery and immediately put
20 into guanidinium thiocyanate (1-5 g in 50 ml). Total RNA was extracted as described by Chirgwin [20]. Poly(A)-RNA was prepared by chromatography on oligo-deoxythymidilate-(oligo(DT))-cellulose column [21].

25 D. Construction and screening of cDNA libraries

- Approximately 15 μ g poly-adenylated RNA from human pancreas was denatured with methyl mercuric hydroxide [22] and primed with oligo (dT) ₁₂₋₁₈ primers (Pharmacia,
30 Uppsala, Sweden), and reversely transcribed using standard procedures [23]. Second-strand synthesis was carried out according to Gubler and Hoffman [24], except that DNA ligase and β -NAD were omitted, and the reaction temperature was set at 15 °C. Excess RNA was digested with
35 RNase A (50 μ g/ml), and the double-stranded cDNA was treated with EcoRI methylase [25]. Ends were blunted with Klenow enzyme. After ligation to EcoRI linkers and



cleavage with EcoRI the cDNA was fractionated on a Sepharose 4B-C1 column. The void volume fraction was precipitated with ethanol and the cDNA ligated into the EcoRI site of a phosphatase treated gtl1 vector [26]. In vitro packing yielded more than 7×10^5 recombinants.

A cDNA library from human mammary gland, derived from tissue obtained from a women at the eighth month of pregnancy, was purchased (Clontech Laboratories, Inc., Palo Alto, CA; USA).

Phages from the cDNA libraries were plated at 5×10^4 plaque forming units per 120-mm dish. The antiserum was diluted to a ratio of 1:3200 and screening was performed according to Young and Davis [27]. Alkaline-phosphatase-conjugated goat-anti-rabbit antibodies were used as second antibodies (Bio-Rad, Richmond, CA USA). To isolate clones corresponding to the 5'-end of the mRNA, nucleic acid hybridization was done under standard conditions [23] using a subcloned fragment from one of the immunopositive clones as a probe.

E. RNA analysis

Electrophoresis was carried out in a 1% agarose gel in 40mM MOPS buffer pH 7.0 after denaturation with glyoxal and dimethylsulfoxide [28]. Glyoxalated total RNA was then transferred to nitrocellulose filters [29]. The blots were probed with subclones of BSSL and CEH recombinants that were labelled by the oligo-labeling technique [30]. Prehybridization and hybridization were carried out with 50% formamid at 46 °C [23]. Posthybridization washes were performed at high stringency (0.1% SDS and 0.1xSSC at 60° C). (1xSSC, 0,15M NaCl, 0.0015M Na₃ citrate, pH 7.6).

35

F. Nucleotide sequence

This Sheet Has Been Replaced



cDNA inserts from BSSL and CEH recombinants were either directly cloned into M3mp18 and mp19 after sonication and size fractionation or some of them were further subcloned into pTZ19R after digestions with PstI, BstXI, NarI, SmaI and AhaII. The nucleotide sequence was determined by the dideoxy chain-termination method [31]. The GC-rich repeats (see below) were also sequenced with TaqI polymerase and dc ⁷GTP. Both strands were sequenced. Sequence information was retrieved from autoradiograms by use of the software MS-EdSeq as described by Sjöberg et al [55].

G. Amino acid sequence predictions and homologies.

To predict the corresponding amino acid sequence of the cDNA inserts, codon usage of different reading frames was compared according to Staden and gave one open reading frame [32]. Homologies were searched for with the programs of the UWCGG software package [33].

20 Results and Discussion

A. Sequences of tryptic fragments and the N-terminus of BSSL

25 Trypsin digestion of purified BSSL resulted in approximately 50 fragments as judged by the number of peaks obtained during the HPLC-chromatography (Fig. I). The peaks were collected and the indicated peaks which could be isolated in a highly purified state and in reasonable quantities, were sequenced. The resulting sequences are shown in Table I. In addition the 30 most N-terminal residues were sequenced (Fig. II), and they confirm the previously reported sequence of Abouakil et al. (30 residues) [17].

35

The 22 residues long N-terminal sequence reported from Wang & Johnson [34] is a glycine in our report and a



lysine in Wang & Johnson's.

B. Nucleotide sequence of BSSL

5 For construction of the λ gt11 cDNA library we used polyadenylated RNA from human pancreas. Initially four immunopositive clones were isolated, and then this pancreatic expression cDNA library was screened with antiserum against BSSL. Nucleotide sequence analysis of
10 the four clones showed that they are in perfect agreement and correspond to the 3'-end of the mRNA. They all begin with a poly A tail and differ only in length; the longest insert, designated ACEH, spans 996 bp.

A cDNA library from human mammary gland was screened with
15 antiserum, and the pancreas clone ACEH as probe. Positive clones were isolated from both screenings, which all originate from the 3'-end. The longest mammary gland clone, designated ABSSL, reaches 2100 bp upstream. It contains four of the sequenced tryptic fragments (Fig.
20 II), but do not include the N-terminal amino acid sequence. To extend the sequence beyond the translation start, the mammary gland cDNA library was rescreened with a 118 bases long probe derived from the most 5'proximal part of ABSSL. One clone was isolated that continued a
25 further 328 nucleotides upstream. It matched the N-terminal amino acid sequence, and contained the remaining tryptic fragment. As shown in Fig. II, the cDNA is 2428 nucleotides long and contains 81 bases upstream from the first ATG codon. The polyadenylation signal, AATAAA is
30 located 13 nucleotides upstream from the poly A tail and the termination codon TAG was found at nucleotide 2317 followed by a 3'-untranslated region of 112 bp.

A GC rich region consisting of 16 repeats of 33 bases was found in the 3'-end of the sequence between base 1756 and
35 2283. The nucleotide sequence of the repetition, shown in Fig. III, consists of six identical repetitions surrounded by ten repetitions with different number of



substitutions that have probably occurred after several duplications. The low number of substitutions suggests that these repetitions have appeared late during evolution.

5

C. Tissue distribution of expression

RNA from human lactating mammary gland, pancreas, adipose tissue and from a human hepatoma celline (HepG2) was
10 analyzed by Northern blotting. The size of the messenger was determined to be approximately 2.5 kb in both lactating mammary gland and pancreas. No signal could be detected in the lanes with RNA extracted from HepG2 or adipose tissue (Fig. IV).

15 Since the mRNA used for the mammary gland library was obtained from a female in her 8th month of pregnancy, it is evident that transcription and probably translation of the BSSL gene is turned on before partus, in agreement with previous findings on BSSL secretion before partus
20 [35]. See Figure IV.

D. Amino acid sequence of BSSL

Assessed by SDS-PAGE the molecular mass has been reported
25 to be 107-125 kDa [8,36] and by analytical ultracentrifugation to be 105 kDa [37]. The enzyme, as deduced from the cDNA, consists of 722 amino acid residues (Fig. II) which, giving a molecular mass of 76.271 Da, indicates that the enzyme contains at least 15-20%
30 carbohydrate. The leader sequence is 23 residues long. A tentative active site serine residue is localized to serine-217 (Fig. V). The sequence around this serine accord with the consensus active site sequence of serine-hydrolases [38]. It has recently been it was proposed that
35 basic residues found close to the active site serine may be involved in the cleavage of ester bonds in acylglycerols by lipases [39]. It is interesting to note that such



residues are not present in BSSL. The single tentative N-glycosylation site is localized only seven residues from the serine. The degree of glycosylation [6,16] suggests that the enzyme contains O-linked carbohydrate. There are
5 numerous sites where such glycosylation could have occurred. The amino acid composition based on purified enzyme has shown a high content of proline residues [6]. The amino acid sequence obtained from cDNA confirms this. Moreover, most of the proline residues are localized in
10 the 16 repeats of 11 residues each, constituting the main part of the C-terminal half of the enzyme.

E. Comparison of the enzymes in mammary gland (BSSL) and pancreas (CEH)

15 BSSL of human milk and human pancreas CEH have previously been shown to be similar, if not identical. The present data strongly suggests that the two enzymes are products of the same gene. The nucleotide sequence of the cDNA
20 clones shows that the pancreatic clone ACEH is identical with the mammary gland clone ABSSL from the poly A tail and 996 bases towards the 5'-end, including the sequence coding for the proline rich repeats. Northern blot gave a
25 single band of 2.5 kb in RNA from pancreas and lactating mammary gland (Fig. IV). Genomic Southern blots further support the idea that only one gene codes for BSSL and CEH. The difference in mobility on SDS-PAGE between BSSL and CEH can be explained as a consequence of different glycosylation or differential splicing.

30 The similarity of BSSL to the rat and bovine enzymes (see below) and to results from genomics blots support the possibility that differential splicing cannot account for the mobility difference. Since the C-terminal sequence has
35 not been confirmed on the protein level there is a less likely possibility that CEH may be processed by a proteolytic clearance in the C-terminal end.



So far as we know pancreatic enzymes that obviously correspond to CEH have often been named after species and the particular substrates used to determine their respective activities; lysophospholipase, cholesteryl
5 esterase, sterol ester hydrolase, non-specific lipase, carboxyl ester lipase and cholesteryl ester hydrolase. Available data are compatible with the view that all these activities described originates in one and the same functional entity [42,43]. This illustrates the broad
10 substrate specificity and the relevance of designating them as non-specific lipases. When the sequence of human BSSL/CEH is compared to the sequence of lysophospholipase from fat pancreas [40] and cholesterol esterase from bovine pancreas [41] extensive similarities are found that
15 extend about 530 residues from the N-terminal (Fig V); but they differ in the part of the molecule where the repeats occur. The rat enzyme has only four repeats and the bovine three. Hence the human enzyme is a considerable longer peptide.

20 Moreover, striking similarities were found between BSSL and a number of typical esterases, e.g. acetyl choline esterases from several species, including man and Drosophila, and carboxyl esterases (Fig. VI). These
25 similarities were restricted to the N-terminal 300 residues of BSSL which includes the tentative active site serine-residue. A similarity to acetyl choline esterase has been predicted from the fact that BSSL is inhibited by typical choline esterase inhibitors [6, 8, 16]. With the
30 possible exception of the rat liver carboxyl esterase [45], none of these similar enzymes has been shown to have the same bile-salt dependency as BSSL; this suggests that the structural basis for this property resides in the C-terminal part of the protein. Moreover, BSSL can
35 efficiently attack emulsified substrates which is not a known characteristic of the similar esterases. For this activity bile salt is a prerequisite.



The predicted sequence for human BSSL was compared with other well characterized mammalian lipases. Apart from the consensus sequence around the active site serine (G-X-S-X-G), no obvious similarities were found [44].

In addition to the similarities with other enzymes, there also significant similarities to one c-AMP dependent protein from Dictyostelium discoideum [46] as well as to thyroglobulin from several species (Fig VI) [47-49]. The similarities between BSSL and thyroglobulin, which comprise the active site region but not the active site itself, indicate that these highly conserved stretches of amino acids are of more generalized importance than merely supporting the enzymatic activity of esterases.

In conclusion, human milk BSSL consists of 722 amino acid residues. Available data strongly indicate that its peptide chain is identical to that of pancreatic CEH, and they are coded for by the same gene. The strongest evidence is that the nucleotide sequences of their 3 - ends and their N-terminal amino acid sequences are identical. The striking homologies found to rat pancreatic lysophospholipase and bovine pancreatic cholesterol esterase support the hypothesis that also these enzymes are functionally identical. However, as it has been suggested, the different molecular sizes found among species are not due to differences merely in glycosylation; instead they reflect a variable number of an eleven amino acid repeat. The similarity of the active site sequence between these esterases suggests that these proteins derive from a common ancestral gene.

With reference to Figures I-VII, the following legends are given.

This Sheet Has Been Replaced



Figure I: Separation of the tryptic digest of BSSL on HPLC

Purified BSSL was treated with trypsin and chromatographed on HPLC as described in Materials and Methods. The
 5 indicated peaks were collected and purified further by a rechromatograph and their amino acid sequence determined.

Figure II: The cDNA nucleotide sequence and the deduced amino acid sequence for human bile salt-stimulated lipase:

10

The cDNA is 2428 bases long. The N-terminal 23-codon sequence (nt82-150) starting with an ATG, is interpreted as a leader peptide since the N-terminal amino acid sequence of the mature protein starts at codon 24 (nt 151,
 15 Ala). The leader peptide is underlined. The sign * indicates the starting point of an exon. The sign # indicates the starting point of the repetition part.

Figure III: The nucleotide sequence of the C-terminal GC-rich repetitions in the bile salt-stimulated lipase:

20

Substitutions are indicated by a *.

Figure IV: Northern blot hybridization

25

Northern blot analysis of total RNA isolated from human lactating mammary gland, pancreas, adipose tissue and a human hepatoma cell line (HepG2). Total TNA (10µg) from lactating mammary gland (lane A), pancreas (lane B),
 30 adipose tissue (lane C) and HepG2 (lane D) were electrophoresed in a 1% agarose gel in 40mM MOPS buffer at Ph 7.0 after denaturation of RNA in 1M glyoxal, 50% dimethylsulfoxide and 40mM MOPS. The glyoxalated RNA was then transferred to nitrocellulose paper for hybridization
 35 with [³² P] labeled BSSL cDNA (ABSSL).



Figure V: Comparison of the deduced amino acid sequence from human milk BSSL, rat pancreatic lysophospholipase (Rat1pl) [40] and bovine pancreatic cholesterol esterase (Bovceh) [41]:

5

The serine residues involved in the active site are indicated by a *, and the # indicates the single possible N-glycosylation signal of the protein. The direct repeats of amino acid sequences are boxed. Matching sequences are denoted in capital letters, matching sequences between two
10 enzymes are denoted in small letters and mismatching with a dot.

Figure VI: Comparison of the primary structure of BSSL to
15 other esterases, thyroglobuline and to one c-AMP dependent enzyme from Dictyostelium discoideum:

BSSL: bile salt stimulated lipase from human, Cheshum: cholinesterase from fetal human tissue [50], Torpace:
20 acetylcholinesterase from Torpedo marmorata [51], Drosceh: carboxylic ester hydrolase from Drosophila melaogaster [52], Ratlivce: carboxyl esterase from rat liver [53], Drosace: acetylcholinesterase from Drosophila melaogaster [54], Thyrhum: thyroglobulin from human [49]
25 and Dict.Di: c-AMP dependent enzyme from Dictyostelium discoideum [46]. There are 7 different domains that show similarities between the enzymes. Boxes enclose residues which are identical and small letters in the consensus sequence indicate identical residues in all the enzymes
30 except for one. Dots indicate mismatches. The serine residue involved in the active site is indicated with *. The figure in the right hand corner shows how the domains are oriented.

35

This Sheet Has Been Replaced



Figure VII:

gives the amino acid sequence 1 - 722 for the entire protein (one letter code) and indicates exons a, b, c, 5 and d. The sign # indicates the starting point of the repetition part.

This Sheet Has Been Replaced



Table 1. Amino acid sequence of BSSL peptides.

Due to interfering peaks no positive identification of the residue in cycles 1 and 2 of the sequencing could be made in peptide number 26. The peptide numbers refer to the peaks in Fig I.

10

Tryptic fragments

15

TP16:LysValThrGluGluAspPheTyrLys

TP19:GlyIleProPheAlaAlaProThrLys

Tp20:LeuValSerGluPheThrIleThrLys

20

TP24:ThrTyrAlaTyrLeuPheSerHisProSerArg

TP26:PheAspValTyrThrGluSerTrpAlaGlnAsp

25

ProSerGlnGluAsnLys

This Sheet Has Been Replaced



**Table 2: Identification of the exons a, b, c and d
numbered as in Figures II and VII.**

5		<u>Location</u>	
10	exon	between nucleotide number	between aminoacid number
		-----	-----
	a	986-1172	279-341
	b	1173-1376	342-409
15	c	1377-1574	410-474
	d	1575-2415	475-722
	the entire protein	151-2316	1-722
20	the entire protein	151-1755	1-535
	excluding repetitions		
25	-----		-----

30

35

This Sheet Has Been Replaced



References

1. Bläckberg, L., Ångquist, K.A, & Hernell, O. (1987) FEBS Lett. 217, 37-41.
- 5 2. Hernell, O. (1975) Eur. J. Clin. Invest. 5, 267-272.
3. Hernell, O., Bläckberg, L. & Bernbäck, S. (1988) In Perinatal nutrition (Lindblad, B.S., ed.) Bristol-Myers Nutrition symposia vol. 6, pp. 259-272, Academic Press, New York.
- 10 4. Hernell, O, L. Bläckberg, B. Fredrikzon, and T. Olivecrona. 1981. Bile salt-stimulated lipase in human milk and lipid digestion in the neonatal period. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 465-471.
- 15 5. Bernbäck, S., Bläckberg, L. & Hernell, O. (1990) J. Clin. Invest. J. Clin. Invest. 221-226 (1990)
6. Bläckberg, L. & Hernell, O. (1981) Eur. J. Biochem 116, 221-225.
- 20 7. Hernell, O. & Bläckberg, L. (1982) Pediatr. Res. 16, 882-885.
8. Bläckberg, L. Lombardo, D., Hernell, O., Guy, O. & Olivecrona, T. (1981) FEBS Lett. 136, 284-288.
9. Fredrikzon, B., Hernell, O., Bläckberg, L. & Olivecrona, T. (1978) Pediatr. Res. 12, 1048-1052.
- 25 10. Wang, C.-S., Hartsuck, J.A. & Downs, D. (1988) Biochemistry 27, 4834-4840.
11. Bläckberg, L. & Hernell, O. (1983) FEBS Lett. 157, 337-341.
- 30 12. Björkstén, B., Burman, L.G., deChateau, P., Fredrikzon, B., Gotheffors, L. & Hernell, O. (1980) Br. Med. J. 201, 267-272.
13. Brueton, M.J., Berger, H.M., Brown, G.A., Ablitt, L., Iyangkaran, N. & Wharton B.A. (1978) Gut 19, 95-98.
- 35 14. Murphy, G.M. & Signer, E. (1975) Gut 15, 151-163.
15. Hernell, O., Staggers, J.E. & Carey, M.C. Biochemistry. (1990), 29:2041-2056.

This Sheet Has Been Replaced



16. Lombardo, D., Guy, O. & Figarella, C. (1978) *Biochim. Biophys. Acta* 527, 142-149.
17. Abouakil, N., Rogalska, E., Bonicel, J. & Lombardo, D. (1988) *Biochim. Biophys. Acta* 961, 299-308.
- 5 18. Hernell, O., Bläckberg, L. & Lindberg, T. (1988) in *Textbook of Gastroenterology and Nutrition in infancy* (Lebenthal, E., ed) pp. 209-217, Raven Press, New York.
- 10 19. Wang, C.-S. (1988) *Biochem. Biophys. Res. Com.* 155, 950-955.
20. Chirgwin, J.M. Przybyla, A.E., MacDonald R.J. & Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
21. Aviv, H. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 69, 5201-5205.
- 15 22. Bailey, J.M. & Davidson, N. (1976) *Anal. Biochem.* 70, 75-85.
23. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982): *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- 20 24. Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263-269.
25. Maniatis T., Haldison, R.C., Lacy, E., Lauer, J., O'Connell, C. & Qvon, D. (1978) *Cell* 15, 687-701.
26. Young, R.A. & Davis, R.W. (1983) *Science* 222, 778-782.
- 25 27. Young, R.A. & Davis, R.W. (1983) *Proc. Natl. Acad. Sci, USA* 80, 1194-1198.
28. McMaster, G.K. & Charmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 48 353-48 358.
29. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
- 30 30. Feinberg, A.P. & Vogelstein, B. (1983) *Analyt. Biochem.* 132, 6-13.
31. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74. 5463-5467.
- 35 32. Staden, R. (1984) *Nucl. Acids Res.* 12, 551-567.
33. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.

This Sheet Has Been Replaced



34. Wang, C.-S. & Johnson, K. (1983) *Anal. Biochem.* 133, 457-461.
35. Hamosh, M. (1986) in *Human Milk Infant Nutrition and Health* (Howell, R.R. Morriss, F.H. & Pickering, L.K. eds) pp. 66-97, Charles, C. Thomas, Springfield.
- 5 36. Wang, C.-S. (1980) *Anal. Biochem.* 105, 398-402.
37. Wang, C.-S & Lee, D.M. (1985) *J. Lipid. Res.* 26, 824-830.
38. Brenner, S. (1988) *Nature* 334, 528-530.
- 10 39. Yang, C.-Y., Gu, Z.-W., Yang, H.-H., Rohde, M.F., Gotto, Jr., A.M. & Pownall, H.J. (1989) *J. Biol. Chem.* 265, 16822-16827.
40. Han, J.H., Stratowa, C., & Rutter, W.J. (1987) *Biochemistry* 26, 1617-1625.
- 15 41. Kyger, E., Wiegand, R., & Lange, L. (1989) *Biochim. Biophys. Res. Com* 164, 1302-1309.
42. Bläckberg, L. (1981) *Fat digestion in newborn infant. Umeå University Medical Dissertations New Series No 71.*
- 20 43. Rudd, E.A. & Brockman, H.L. (1984) in *Lipases* (Borgström, B & Brockman, H.L. eds) pp. 184-204, Elsevier, Amsterdam.
44. Mickel, S., Weidenbach, F., Swarovsky, B., LaForge, S. & Scheele, G. (1989) *J. Biol. Chem.* 264, 12895-12901.
- 25 45. Cammulli, E.D., Linke, M.J., Brockman, H.L. & Hui, D.Y. (1989) *Biochim. Biophys. Acta* 1005, 177-182.
46. Mann, S.K.O. & Firtel, R.A. (1987) *Mol. Cell Biol.* 7, 458-469.
- 30 47. Mercken, L., Simons, M.J., Swillens, S., Masser, M. & Vassart, G. (1985) *Nature* 316, 647-651.
48. Lauro, R., Obici, S. Condliffe, D., Ursini, V.M., Musti, A., Moscatelli, C. & Avvedimento, V.E. (1985) *Eur. J. Biochem.* 148, 7-11.
- 35 49. Malthiery, Y. & Lissitzky, S. (1987) *Eur. J. Biochem.* 165, 491-498.
50. Prody, C., Zevin-Sonkin, D., Gnatt, A., Goldberg, O.,



- & Soreq, H. (1987) Proc. Natl. Acad. Sci. USA 84, 3555-3559.
51. Sikorav, J-L., Krejci, E. & Massoulie', J. (1987) EMBO J. 6, 1865-1873.
- 5 52. Oakeshott, J.G. Collet, C., Phillis, R.W., Nielsen, K.M., Russel, R.J., Cambers, G.K., Ross, V. & Richmond, R.C. (1987) Proc. Natl. Acad. Sci, USA 84, 3359-3363.
53. Long, R., Satho, H., Martin, B., Kimura, S., Gonzalez, F. & Pohl, L. (1988) Biochem. Biophys. Res. Comm. 156, 866-873.
- 10 54. Hall, L.M.C. & Spierer, P. (1986) EMBO J. 5, 2949-2954.
55. Sjöberg, S., Carlsson, P., Enerbäck, S., & Bjursell, G. (1989) CABIOS 5, 41-46.
- 15 56. EP-A-317355.
57. Hernell, O., and T. Olivecrona. 1974. Human milk lipases II. Bile salt-stimulated lipase. Biochim. Biophys. Acta 369:234-244.
- 20 58. Hernell, O., L. Bläckberg, and T. Olivecrona. 1981. Human milk lipases. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 347-354.
- 25 59. Atkinson, S.A., M.H. Bryan, and G.H. Andersson. 1981. Human milk feeding in premature infants: protein, fat and carbohydrate balances in the first two weeks of life. J.Pediatr. 99:617-624.
- 30 60. Chappell, J.E., M.T. Clandinin, C. Kearney-Volpe, B. Reichman and P.W. Swyer. 1986. Fatty acid balance studies in premature infants fed human milk or formula: effect of calcium supplementation. J. Pediatr. 108:438-447.
- 35



61. Williamson, S., E. Finucane, H. Ellis, and H.R. Gamsu. 1978. Effect of heat treatment of human milk on absorption of nitrogen, fat, sodium, calcium and phosphorus by preterm infants. Arch. Dis. Childhood 53:555-563.
- 5

10

15

This Sheet Has Been Replaced





What we claim is:

1. A protein as indicated in Figure VII from position 1
5 to position 722.
2. A protein as indicated in Figure VII from position 1
to position 535.
- 10 3. A protein as indicated in Figure VII from position 1
to position 278.
4. A protein as indicated in Figure VII from position 1
to position 341.
- 15 5. A protein as indicated in Figure VII from position 1
to position 409.
6. A protein as indicated in Figure VII from position 1
20 to position 474.
7. A protein as indicated in Figure VII from position 1
to position 278, position 279 to position 341,
position 279 to position 409, position 279 to
25 position 474, position 342 to position 409, position
342 to position 474, or position 536 to position 722.
8. A protein according to any of claims 1-7 in
combination with one or more of the repeats according
30 to Figure V.
9. A protein according to any of claims 1-8, possessing
methionine as additional N-terminal amino acid.
- 35 10. A protein according to any of claims 1-9, exhibiting
one or several of the critical functions of
naturally occurring bile salt-stimulated lipase.





11. A functionally equivalent variant or mutant of a protein according to any of claims 1-9.
- 5 12. A DNA sequence coding for a protein with the amino acid sequence according to claim 1.
13. A DNA sequence coding for a protein with the amino acid sequence according to claim 2.
- 10 14. A DNA sequence coding for a protein with the amino acid sequence according to any of claims 3, 4, 5, 6, 7, 8, 9, 10 and 11.
- 15 15. A DNA sequence defined by the following nucleotide numbers in Figure II:
- 151-2316
- 151-1755
- 151-985
- 20 151-1172
- 151-1376
- 151-1574
- 986-1172
- 986-1376
- 25 986-1574
- 1173-1376
- 1173-1574
- 30 16. A vector comprising a DNA sequence coding for a protein according to claims 1 - 11
17. A vector comprising a DNA sequence according to claims 12 - 15.
- 35 18. A host organism transformed with a vector according to claims 16 or 17.

This Sheet Has Been Replaced





19. A process for the production of a protein according to claims 1 - 11 by growing a host organism containing a vector according to claims 16 or 17, and isolating the protein.
20. A pharmaceutical composition comprising a protein according to any of claims 1 - 11.
21. A pharmaceutical composition comprising a protein according to any of claims 1-11 in combination with a lipase or in combination with preparations containing lipases.
22. The use of a protein according to any of claims 1 - 11, for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency.
23. The use of a protein according to any of claims 1 - 11 for the manufacture of a medicament for the treatment of cystic fibrosis.
24. The use of a protein according to any of claims 1 - 11 as a supplement to an infant food formulation.
25. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of chronic pancreatitis.
26. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption.
27. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins.





28. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons.
- 5.
29. The use, according to claims 22-28 of a protein according to any of claims 1-11 in combination with a lipase or lipases or in combination with preparations containing a lipase or lipases.
- 10
30. An infant food formulation, supplemented with a protein according to any of claims 1 - 11.
- 15
31. A protein according to claims 1 - 11 in substantially pure form.
32. A protein according to claims 1 - 11 in isolated form.
- 20
33. A DNA sequence according to claims 12 - 15 in substantially pure form.
34. A DNA sequence according to claims 12 - 15 in isolated form.
- 25
35. A process for the preparation of a pharmaceutical composition according to claims 20 or 21 by incorporating a protein according to claims 1 - 11, 31, or 32 in a pharmacologically acceptable carrier.
- 30
36. A process according to claim 35 for the preparation of a pharmaceutical composition for oral administration.
- 35
37. A process for the preparation of an infant food formulation according to claim 30, by supplementing an infant food formulation with a protein according to claims 1 -11, 31, or 32.

This Sheet Has Been Replaced



Abstract

The present invention relates to new DNA sequences, to new
5 proteins coded for by such DNA sequences, and to the use
as further described below of such proteins. The invention
also encompasses vectors, such as plasmid constructs,
comprising such DNA sequences, being capable of expressing
the desired enzyme. The invention also includes host
10 organisms transfected with such constructs e.g. bacteria
yeast, mammalian cells, and transgenic animals. The
invention also includes processes for the preparation of
the novel products of the invention. The new proteins of
the invention are related to an enzyme, known i.a. as
15 human bile salt-stimulated lipase.

20

25

30

35

This Sheet Has Been Replaced



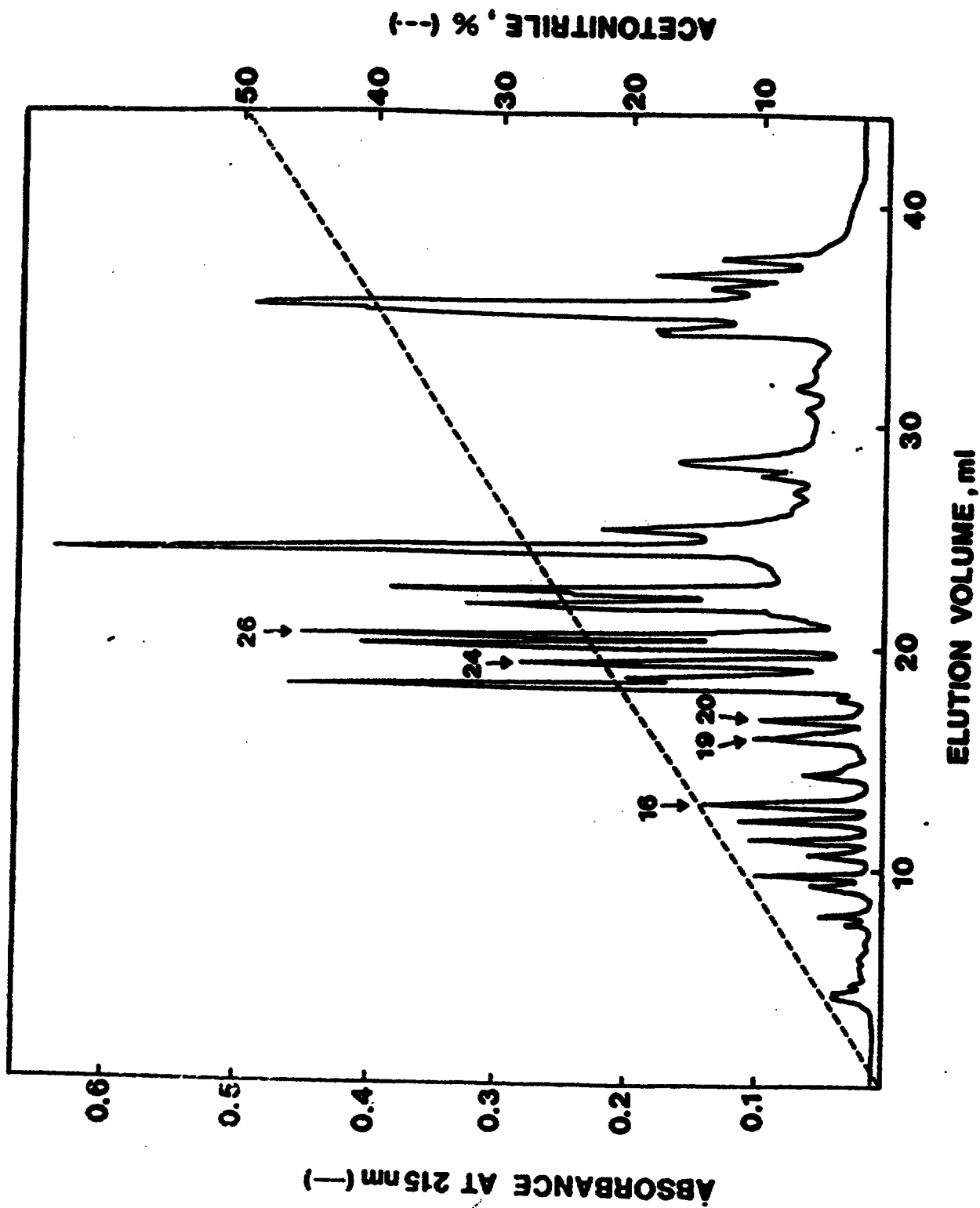


Figure I

This Sheet Has Been Replaced



10	30	50
ACCTTCTGTATCAGTTAAGTGTCAGATGGAAGGAACAGCAGTCTCAAGATAATGCAAAG		
70	90	110
AGTTTATTTCATCCAGAGGCTGATGCTCACCATGGGGCGCCTGCAACTGGTTGTGTTGGGC		
	<u>MetLeuThrMetGlyArgLeuGlnLeuValValLeuGly</u>	
130	150	170
CTCACCTGCTGCTGGGCAGTGGCGAGTGCCGCGAAGCTGGGGCGCCGTGTACACAGAAGGT		
<u>LeuThrCysCysTrpAlaValAlaSerAlaAlaLysLeuGlyAlaValTyrThrGluGly</u>		
190	210	230
GGGTTCGTGGAAGGCGTCAATAAGAAGCTCGGCCTCCTGGGTGACTCTGTGGACATCTTC		
GlyPheValGluGlyValAsnLysLysLeuGlyLeuLeuGlyAspSerValAspIlePhe		
250	270	290
AAGGGCATCCCCTTCGCAGCTCCACCAAGGCCCTGGAAAATCCTCAGCCACATCCTGGC		
LysGlyIleProPheAlaAlaProThrLysAlaLeuGluAsnProGlnProHisProGly		
310	330	350
TGGCAAGGGACCCTGAAGGCCAAGAACTTCAAGAAGAGATGCCTGCAGGCCACCATCACC		
TrpGlnGlyThrLeuLysAlaLysAsnPheLysLysArgCysLeuGlnAlaThrIleThr		
370	390	410
CAGGACAGCACCTACGGGGATGAAGACTGCCTGTACCTCAACATTTGGGTGCCCCAGGGC		
GlnAspSerThrTyrGlyAspGluAspCysLeuTyrLeuAsnIleTrpValProGlnGly		
430	450	470
AGGAAGCAAGTCTCCCGGGACCTGCCC GTTATGATCTGGATCTATGGAGGCGCCTTCCTC		
ArgLysGlnValSerArgAspLeuProValMetIleTrpIleTyrGlyGlyAlaPheLeu		
490	510	530
ATGGGGTCCGGCCATGGGGCCAACCTCCTCAACAACCTACCTGTATGACGGCGAGGAGATC		
MetGlySerGlyHisGlyAlaAsnPheLeuAsnAsnTyrLeuTyrAspGlyGluGluIle		

Figure II/1

This Sheet Has Been Replaced



550	570	590
GCCACACGCGGAAACGTCATCGTGGTCACCTTCAACTACCGTGTCGGCCCCCTTGGGTTC		
AlaThrArgGlyAsnValIleValValThrPheAsnTyrArgValGlyProLeuGlyPhe		
610	630	650
CTCAGCACTGGGGACGCCAATCTGCCAGGTAACCTATGGCCTTCGGGATCAGCACATGGCC		
LeuSerThrGlyAspAlaAsnLeuProGlyAsnTyrGlyLeuArgAspGlnHisMetAla		
670	690	710
ATTGCTTGGGTGAAGAGGAATATCGCGGCCTTCGGGGGGGACCCCAACAACATCACGCTC		
IleAlaTrpValLysArgAsnIleAlaAlaPheGlyGlyAspProAsnAsnIleThrLeu		
730	750	770
TTCGGGGAGTCTGCTGGAGGTGCCAGCGTCTCTCTGCAGACCCTCTCCCCCTACAACAAG		
PheGlyGluSerAlaGlyGlyAlaSerValSerLeuGlnThrLeuSerProTyrAsnLys		
790	810	830
GGCCTCATCCGGCGAGCCATCAGCCAGAGCGGCGTGGCCCTGAGTCCCTGGGTTCATCCAG		
GlyLeuIleArgArgAlaIleSerGlnSerGlyValAlaLeuSerProTrpValIleGln		
850	870	890
AAAAACCCACTCTTCTGGGCCAAAAGGTGGCTGAGAAGGTGGGTGGCCCTGTGGGTGAT		
LysAsnProLeuPheTrpAlaLysLysValAlaGluLysValGlyCysProValGlyAsp		
910	930	950
GCCGCCAGGATGGCCCAGTGTCTGAAGGTTACT ^{P24} GATCCCCGAGCCCTGACGCTGGCCTAT		
AlaAlaArgMetAlaGlnCysLeuLysValThrAsp ^{P24} ProArgAlaLeuThrLeuAlaTyr		
970	990	1010
AAGGTGCCGCTGGCAGGCCTGGAGTACCCCATGCTGCACTATGTGGGCTTCGTCCCTGTC		
LysValProLeuAlaGlyLeuGluTyrProMetLeuHisTyrValGlyPheValProVal		
1030	1050	1070
ATTGATGGAGACTTCATCCCCGCTGACCCGATCAACCTGTACGCCAACGCCGCCGACATC		
IleAspGlyAspPheIleProAlaAspProIleAsnLeuTyrAlaAsnAlaAlaAspIle		

Figure II/2

This Sheet Has Been Replaced



1090	1110	1130
GACTATATAGCAGGCACCAACAACATGGACGGCCACATCTTCGCCAGCATCGACATGCCT		
AspTyrIleAlaGlyThrAsnAsnMetAspGlyHisIlePheAlaSerIleAspMetPro		
1150	1170	1190
GCCATCAACAAGGGCAACAAGAAAGTCACGGAGGAGGACTTCTACAAGCTGGTCAGTGAG		
AlaIleAsnLysGlyAsnLysLysValThrGluGluAspPheTyrLysLeuValSerGlu		
1210	1230	1250
TTCACAATCACCAAGGGGCTCAGAGGCGCCAAGACGACCTTTGATGTCTACACCGAGTCC		
PheThrIleThrLysGlyLeuArgGlyAlaLysThrThrPheAspValTyrThrGluSer		
1270	1290	1310
TGGGCCCAGGACCCATCCCAGGAGAATAAGAAGAAGACTGTGGTGGACTTTGAGACCGAT		
TrpAlaGlnAspProSerGlnGluAsnLysLysLysThrValValAspPheGluThrAsp		
1330	1350	1370
GTCCTCTTCCTGGTGCCACCGAGATTGCCCTAGCCCAGCACAGAGCCAATGCCAAGAGT		
ValLeuPheLeuValProThrGluIleAlaLeuAlaGlnHisArgAlaAsnAlaLysSer		
1390	1410	1430
GCCAAGACCTACGCCTACCTGTTTTCCCATCCCTCTCGGATGCCCGTCTACCCCAAATGG		
AlaLysThrTyrAlaTyrLeuPheSerHisProSerArgMetProValTyrProLysTrp		
1450	1470	1490
GTGGGGGCGGACCATGCAGATGACATTCAGTACGTTTTTCGGGAAGCCCTTCGCCACCCCC		
ValGlyAlaAspHisAlaAspAspIleGlnTyrValPheGlyLysProPheAlaThrPro		
1510	1530	1550
ACGGGCTACCGGCCCCAAGACAGGACAGTCTCTAAGGCCATGATCGCCTACTGGACCAAC		
ThrGlyTyrArgProGlnAspArgThrValSerLysAlaMetIleAlaTyrTrpThrAsn		
1570	1590	1610
TTTGCCAAAACAGGGGACCCCAACATGGGCGACTCGGCTGTGCCACACACTGGGAACCC		
PheAlaLysThrGlyAspProAsnMetGlyAspSerAlaValProThrHisTrpGluPro		

Figure II/3

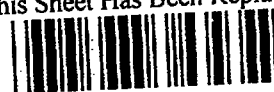
This Sheet Has Been Replaced



1630	1650	1670
TACACTACGGAAAACAGCGGCTACCTGGAGATCACCAAGAAGATGGGCAGCAGCTCCATG		
TyrThrThrGluAsnSerGlyTyrLeuGluIleThrLysLysMetGlySerSerSerMet		
1690	1710	1730
AAGCGGAGCCTGAGAACCAACTTCCTGCGCTACTGGACCCTCACCTATCTGGCGCTGCCC		
LysArgSerLeuArgThrAsnPheLeuArgTyrTrpThrLeuThrTyrLeuAlaLeuPro		
1750	1770	1790
ACAGTGACCGACCAGGAGGCCACCCCTGTGCCCCCACAGGGGACTCCGAGGCCACTCCC		
ThrValThrAspGlnGluAlaThrProValProProThrGlyAspSerGluAlaThrPro		
1810	1830	1850
GTGCCCCCACGGGTGACTCCGAGACCGCCCCCGTGCCGCCCACGGGTGACTCCGGGGCC		
ValProProThrGlyAspSerGluThrAlaProValProProThrGlyAspSerGlyAla		
1870	1890	1910
CCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCCACGGGTGACTCC		
ProProValProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSer		
1930	1950	1970
GGGGCCCCCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCCACGGGT		
GlyAlaProProValProProThrGlyAspSerGlyAlaProProValProProThrGly		
1990	2010	2030
GACTCCGGGGCCCCCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCC		
AspSerGlyAlaProProValProProThrGlyAspSerGlyAlaProProValProPro		
2050	2070	2090
ACGGGTGACTCCGGCGCCCCCCCCCGTGCCGCCCACGGGTGACGCCGGGCCCCCCCCCGTG		
ThrGlyAspSerGlyAlaProProValProProThrGlyAspAlaGlyProProProVal		
2110	2130	2150
CCGCCCACGGGTGACTCCGGCGCCCCCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCC		
ProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSerGlyAlaPro		

Figure II/4

This Sheet Has Been Replaced



2170	2190	2210
CCCGTGACCCCCACGGGTGACTCCGAGACCGCCCCCGTGCCGCCCACGGGTGACTCCGGG		
ProValThrProThrGlyAspSerGluThrAlaProValProProThrGlyAspSerGly		
2230	2250	2270
GCCCCCCTGTGCCCCCACGGGTGACTCTGAGGCTGCCCTGTGCCCCCACAGATGAC		
AlaProProValProProThrGlyAspSerGluAlaAlaProValProProThrAspAsp		
2290	2310	2330
TCCAAGGAAGCTCAGATGCCTGCAGTCATTAGGTTTTAGCGTCCCATGAGCCTTGGTATC		
SerLysGluAlaGlnMetProAlaValIleArgPhe		
2350	2370	2390
AAGAGGCCACAAGAGTGGGACCCCAGGGGCTCCCCTCCCATCTTGAGCTCTTCCTGAATA		
2410		
AAGCCTCATACCCCTAAAAAAAAAAAAA		

Figure II/5

This Sheet Has Been Replaced



		Number of substitutions	
GAGGCCACCCCTGTCCGCCACAGGGACTCC	1	0	
GAGGCCACTCCCGTCCGCCACCGGTGACTCC	2	4	
GAGCCGCCCCCGTCCCGCCACCGGTGACTCC	3	3	
GGGGCCCCCCCCGTCCCGCCACCGGTGACTCC	4	0	
GGGGCCCCCCCCGTCCCGCCACCGGTGACTCC	5	0	
GGGGCCCCCCCCGTCCCGCCACCGGTGACTCC	6	Identical to consensus	0
GGGGCCCCCCCCGTCCCGCCACCGGTGACTCC	7		0
GGGGCCCCCCCCGTCCCGCCACCGGTGACTCC	8		0
GGGGCCCCCCCCGTCCCGCCACCGGTGACTCC	9		0
GGGCCCCCCCCCGTCCCGCCACCGGTGACTCC	10	2	
GGGCCCCCCCCCGTCCCGCCACCGGTGACTCC	11	1	
GGGCCCCCCCCCGTCCCGCCACCGGTGACTCC	12	1	
GGGCCCCCCCCCGTCCCGCCACCGGTGACTCC	13	2	
GAGACCGCCCCCGTCCCGCCACCGGTGACTCC	14	3	
GGGGCCCCCGTGTCCCGCCACCGGTGACTCT	15	3	
GAGGCTGCCCCCTGTCCGCCACAGTACTCC	16	7	
GGGGCCCCCCCCGTCCCGCCACCGGTGACTCC	Consensus		

Figure III

This Sheet Has Been Replaced



A B C D Kb



3.0

2.0

1.5

Figure IV

This Sheet Has Been Replaced



1 30

Besl MLTHGRIQLVYGLTCCYAVASAAKLGAVTTGCGTVEGVNKKLGLG.DSVDIFKGIPTAAKALHPQPHSGHGGTLAKNKKERGLQATTITQDSTTG
 RatiplHGRIVYVIGLTC...AAACAATGALTTGCGTVEGVNKKLGLG.DSVDIFKGIPTAAKALHPQPHSGHGGTLAKNKKERGLQATTITQDSTTG
 BovcshHGCASRLGSPGCLAVASAAKLGAVTTGCGTVEGVNKKLGLG.DSVDIFKGIPTAAKALHPQPHSGHGGTLAKNKKERGLQATTITQDSTTG
 Consensusg:l....lgtcclaasaaAKLGavTTGCGTVEGVNKKLGLG.DSVDIFKGIPTAa..KaLePqrHPQWQGTLEak..FKKRCGLQATTITQDSTTG

101 150

Besl DEDCLYLNINVPQGRKqVShDLFVHLWIYGGAPLMGsggGANFL.NYLYDGKELATGnVIVVTFNTRVGPLGLSTGDnLPGHGLrDQHMAIANVKE
 Ratipl DEDCLYLNINVPQGRKqVShDLFVHLWIYGGAPLMGsggGANFL.NYLYDGKELATGnVIVVTFNTRVGPLGLSTGDnLPGHGLrDQHMAIANVKE
 Bovcsh DEDCLYLNINVPQGRKqVShDLFVHLWIYGGAPLMGsggGANFL.NYLYDGKELATGnVIVVTFNTRVGPLGLSTGDnLPGHGLrDQHMAIANVKE
 Consensus .EDCLYLNINVPQGRKqVShDLFVHLWIYGGAPLMGsggGANFL.NYLYDGKELATGnVIVVTFNTRVGPLGLSTGDnLPGHGLrDQHMAIANVKE

201 250

Besl NIAAPGCDPNITLPGESAGGASVSLQTLSPYHKGILIRRAISQSCVALSPHYIOKNEIHWAKYANKVGGCPVGDAAARMAOCLKYTDPRALTLATKVLPLAG
 Ratipl NIAAPGCDPNITLPGESAGGASVSLQTLSPYHKGILIRRAISQSCVALSPHYIOKNEIHWAKYANKVGGCPVGDAAARMAOCLKYTDPRALTLATKVLPLAG
 Bovcsh NIAAPGCDPNITLPGESAGGASVSLQTLSPYHKGILIRRAISQSCVALSPHYIOKNEIHWAKYANKVGGCPVGDAAARMAOCLKYTDPRALTLATKVLPLAG
 Consensus NIAAPGCDPNITLPGESAGGASVSLQTLSPYHKGILIRRAISQSCVALSPHYIOKNEIHWAKYANKVGGCPVGDAAARMAOCLKYTDPRALTLATKVLPLAG

301 350

Besl LKYPMLHYVGGTYPVIDGDFIPDPINLYANAADIDYIAGTNDGCHLFAASIDPAINEGHKKYTKEDFYKLVSEKTYITGCLGAKTITDVTYTESHAODPS
 Ratipl LKYPMLHYVGGTYPVIDGDFIPDPINLYANAADIDYIAGTNDGCHLFAASIDPAINEGHKKYTKEDFYKLVSEKTYITGCLGAKTITDVTYTESHAODPS
 Bovcsh LKYPMLHYVGGTYPVIDGDFIPDPINLYANAADIDYIAGTNDGCHLFAASIDPAINEGHKKYTKEDFYKLVSEKTYITGCLGAKTITDVTYTESHAODPS
 Consensus .KYP.LHYL.FvPVIDGDFIPDPINLYANAADIDYIAGTNDGCHLFAASIDPAINEGHKKYTKEDFYKLVSEKTYITGCLGAKTITDVTYTESHAODPS

401 450

Besl QENKKCTVVDFTDILFLVPTALAOHRANAKSAKTYATLPSHPSRMPIYFKHnCADHADDIQTVFGKFPATP1CYRAQDRTVSKAMIAIYWNFAKTGD
 Ratipl QENKKCTVVDFTDILFLVPTALAOHRANAKSAKTYATLPSHPSRMPIYFKHnCADHADDIQTVFGKFPATP1CYRAQDRTVSKAMIAIYWNFAKTGD
 Bovcsh QENKKCTVVDFTDILFLVPTALAOHRANAKSAKTYATLPSHPSRMPIYFKHnCADHADDIQTVFGKFPATP1CYRAQDRTVSKAMIAIYWNFAKTGD
 Consensus QEn.KKTvV4fETDILFLVPTALAOHRANAKSAKTY.ATLPSHPSRMPIYFKHnCADHADDIQTVFGKFPATP1CYRAQDRTVSKAMIAIYWNFAKTGD

501 550

Besl PNMGDSAVPTTHKPTITENSGLKIKKMGSSSMERSLRTNLYLHILALPTVTDQ
 Ratipl PNMGDSAVPTTHKPTITENSGLKIKKMGSSSMERSLRTNLYLHILALPTVTDQ
 Bovcsh PNMGDSAVPTTHKPTITENSGLKIKKMGSSSMERSLRTNLYLHILALPTVTDQ
 Consensus PNMG.S.VPthW.PYT.En.nYLeInKhm.S.SMK.hLRtnfL.fWt.Ty.aLPTVt....a...FPeddS.eA.PVFPtddS....pvPptgDS....p

600 645

Besl	VPPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS
Ratipl	VPPPTGDS	QA.....							
Bovcsh	VPPPTGDS	QA.....							

ConsensusS..A.....

691 736

Besl	GDS	GAPPVPPTGDS	ETAPVPPTGDS	GAPPVPPTGDS	EAAPVPPTGDS	KE.AQMPAVIRF
Ratipl	GDS					VE.AQMPGFIH
Bovcsh	GDS					EVAAQMPVIGI

Consensusds...e.AQMP.vIgF

This Sheet Has Been Replaced



Figure V

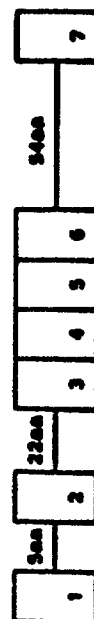
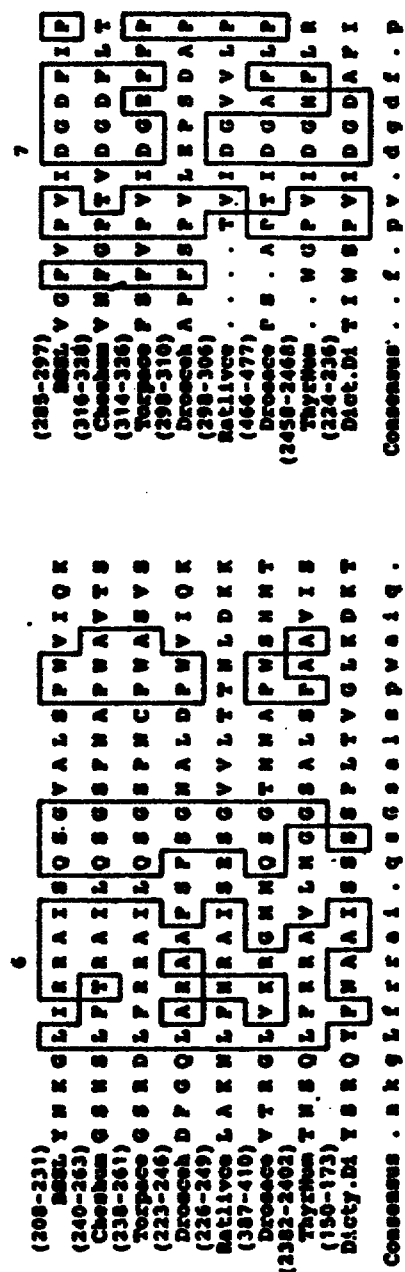
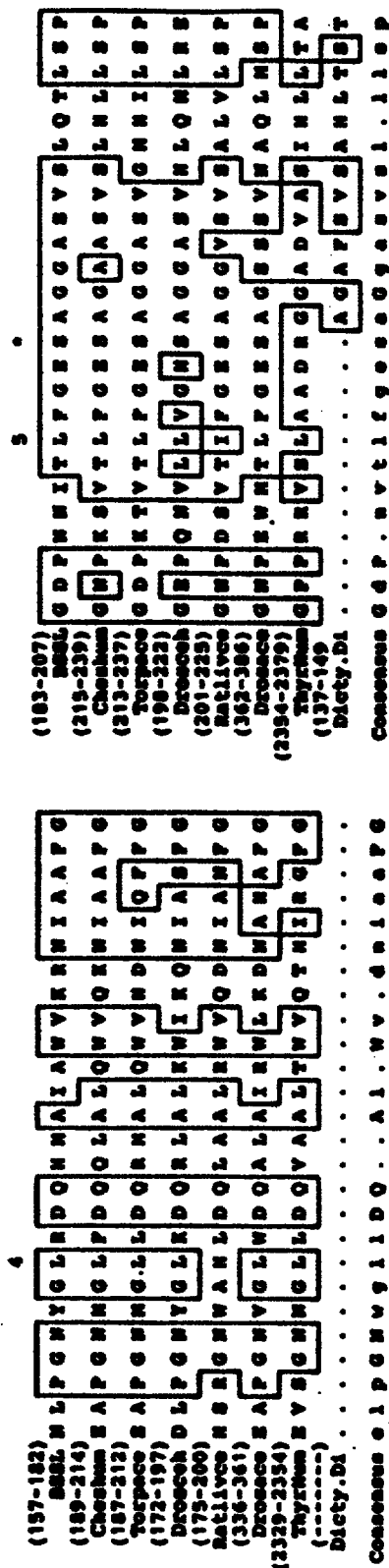
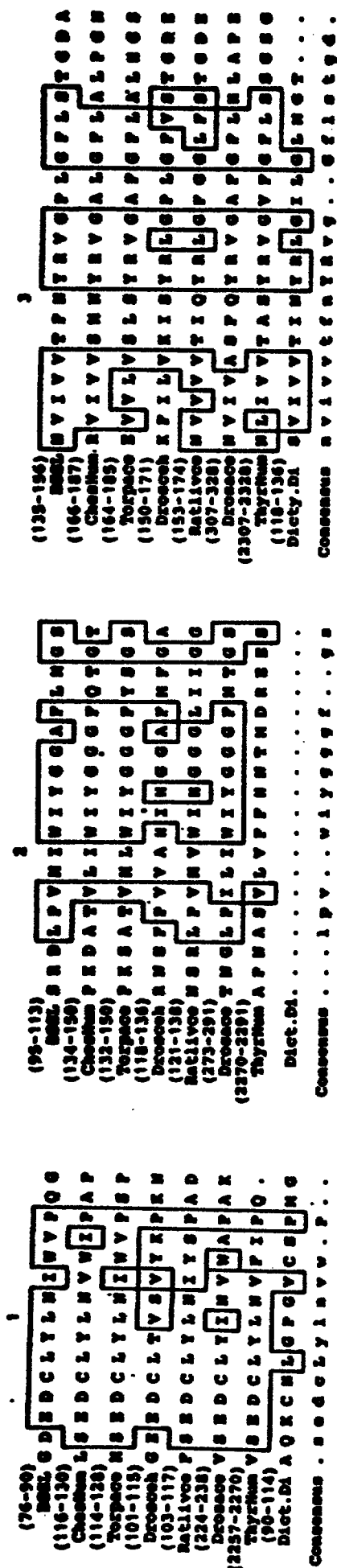


Fig. 1



10	30	50
AKLGAVYTEGGFVEGVNKKLGLLGDSVDIFKGIPFAAPTKALENPQPHPGWQGLKAKNF		
70	90	110
KKRCLQATITQDSTYGDEDCLYLNIWVPQGRKQVSRDLPVMIWIYGGAPLMGSGHGANFL		
130	150	170
NNYLYDGEEIATRGNVIVVTFNRYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKRNIAA		
190	210	230
FGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSPWVIQKNPLFWAKKV		
250	270	290
AEKVGCPVGDAAARMAQCLKVTDPRALTLAYKVPLAGLEYPMLHYVGFVPVIDGDFIPADP		
310	330	350
INLYANAADIDYIAGTNNMDGHIFASIDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGA		
370	390	410
KTTFDVYTESWAQDPSQENKKKTVDVDFETDVLFLVPTEIALAQHRANAKSAKTYAYLPSH		
-----exon b-----		
430	450	470
PSRMPVYPKWVGADHADDIQYVFGKPFATPTGYRPQDRTVSKAMIAWYTNFAKTGDPNMG		
-----exon c-----		
490	510	530
DSAG ⁴¹ PTHWEPTYTENSYLEITKKMGSSSMKRSRLRTNFLRYWTLTYLALPTVTDQBATPV		
-----exon d-----		
550	570	590
PPTGDSEATFPVPTGDSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD		
610	630	650
SGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDAGPPVPPTGDSGAPP		
670	690	710
VPPTGDSGAPPVTPTGDSETAPVPPTGDSGAPPVPPTGDSEAAPVPPTDDSKAQMPAVI		

RF

Figure VII

This Sheet Has Been Replaced:



REPLACED SHEETS		SG
☒ Description	: paged 1 to paged 33	
☒ Claim(s)	: page 34 to paged 37	
☒ Drawing(s)	: paged 1 to paged 15	
☒ Abstract	: paged 38 to paged	

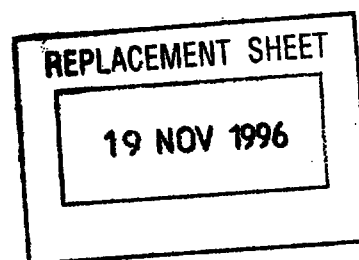
REPLACEMENT SHEETS		SG
☒ Description	: paged 1 to paged 33	
☒ Claim(s)	: paged 34 to paged 37	
☒ Drawing(s)	: paged 1 to paged 15	
☒ Abstract	: paged 38 to paged	
DATE OF RECEIPT		19 NOV 1996

1

**DERIVATIVES OF HUMAN BILE-SALT STIMULATED LIPASE,
AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM**

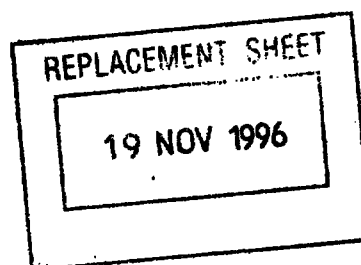
Field of the invention

The present invention relates to new DNA sequences, to new proteins coded for by such DNA sequences, and to the use as further described below of such proteins. The invention also encompasses vectors, such as plasmid constructs, comprising such DNA sequences, being capable of expressing the desired enzyme. The invention also includes host organisms transfected with such constructs e.g. bacteria yeast, mammalian cells, and transgenic animals. The invention also includes processes for the preparation of the novel products of the invention. The new proteins of the invention are related to an enzyme, known i.a. as human bile salt-stimulated lipase.



This Sheet Has Been Replaced





Background to the invention

The human lactating mammary gland synthesizes and secretes with the milk a bile salt-stimulated lipase (BSSL) [1] that, after specific activation by primary bile salts [2, 57, 58], contributes to the breast-fed infant's endogenous capacity of intestinal fat digestion [3-5]. This enzyme, which accounts for approximately 1% of total milk protein [6], is a non specific lipase; in vitro it hydrolyses not only tri-, di- and monoacylglycerols, but also cholesteryl-, and retinyl esters, and lysophosphatidylglycerols [7-10]. Furthermore, its activity is not restricted to emulsified substrates, but micellar and soluble substrates are hydrolyzed at similar rates [11].

BSSL is not degraded during passage with the milk through the stomach, and in duodenal contents it is protected by bile salts from inactivation by pancreatic proteases such as trypsin and chymotrypsin [2,11]. It is, however, inactivated when the milk is pasteurized, e.g. heated to 62.5 °C, 30 min [12]. Model experiments in vitro suggest that the end products of triacylglycerol digestion are different in the presence of BSSL [5,7]. Due to lower intraluminal bile salt concentrations during the neonatal period [13,14] this may be beneficial to product absorption [5,15].

The carboxylic ester hydrolase (CEH) of human pancreatic juice [16] seems functionally to be identical, or at least very similar, to BSSL [8]. They also share common epitopes [8,17], have identical N-terminal amino acid sequences [17] and are inhibited by inhibitors of serine esterases, e.g. eserine and diisopropylfluorophosphate [6,8,16]. It has been hypothesized that the two enzymes are products of the same gene [18,19]. The observed molecular size difference [8,19] could be explained by different patterns of glycosylation, as recently suggested [17].



19 NOV 1996

Dietary lipids is an important source of energy. The energy-rich triacylglycerols constitute more than 95% of these lipids. Some of the lipids, e.g. certain fatty acids and the fat soluble vitamins, are essential dietary constituents. Before gastro-intestinal absorption the triacylglycerols as well as the minor components, i.e. esterified fat-soluble vitamins and cholesterol, and diacylphosphatidylglycerols, require hydrolysis of the ester bonds to give rise to less hydrophobic, absorbable products. These reactions are catalyzed by a specific group of enzymes called lipases.

In the human adult the essential lipases involved are considered to be gastric lipase, pancreatic colipase-dependent lipase (tri- and diacylglycerol hydrolysis), pancreatic phospholipase A2 (diacylphosphatidylglycerols) and carboxylic ester hydrolase (cholesteryl- and fat soluble vitamin esters). In the breast-fed newborn bile salt-stimulated lipase plays an essential part in the hydrolysis of several of the above mentioned lipids. Together with bile salts the products of lipid digestion form mixed micelles from which absorption occurs (3-5).

Common causes of lipid malabsorption, and hence malnutrition, are reduced intraluminal levels of pancreatic colipase-dependent lipase and/or bile salts. Typical examples of such lipase deficiency are patients suffering from cystic fibrosis, a common genetic disorder resulting in a life-long deficiency in some 80% of the patients, and chronic pancreatitis, often due to chronic alcoholism.

The pancreatic and liver functions are not fully developed at birth, most notably in infants born before term. Fat malabsorption, for physiological reasons, is a common finding and thought to result from low intraluminal pancreatic colipase-dependent lipase and bile salt concentrations (3,4,13-15). However, because of BSSL, such



This Sheet Has Been Replaced



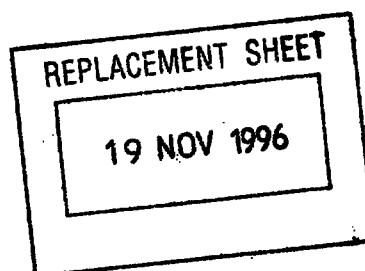
4

REPLACEMENT SHEET

19 NOV 1996

malabsorption is much less frequent in breast-fed infants than in infants fed pasteurized human milk of infant formulas (3-5, 12, 59, 60, 61). This is one reason why it has been advocated that newborn infants, particularly preterm infants, that cannot be fed their own mothers milk should be fed non-pasteurized milk from other mothers (12).

The present treatment of patients suffering from a deficiency of pancreatic lipase is the oral administration of very large doses of a crude preparation of porcine pancreatic enzymes. Colipase-dependent pancreatic lipase is inactivated by low pH. Such conditions are prevalent in the stomach, with the result that orally administered pancreatic lipase is virtually completely inactivated on the passage through the stomach to the gut. Therefore, this effect cannot be completely overcome by the use of large doses of enzyme. The large doses administered are inadequate for most patients, and the preparations are impure and unpalatable. Certain tablets have been formulated which pass through the acid regions of the stomach and discharge the enzyme only in the relatively alkaline environment of the jejunum. However, many patients suffering from pancreatic disorders have an abnormally acid jejunum and such tablets may fail to discharge the enzyme and may therefore be ineffective. Moreover, since the preparations presently on the market are of a non-human source there is a risk of immunoreactions that may cause harmful effects to the patients or result in reduced therapy efficiency. A further drawback with the present preparations is that their content of other lipolytic activities than colipase-dependent lipase are not stated. In fact, most of them contain very low levels of CEH/BSSL-activity. This may be one reason why many patients, suffering from cystic fibrosis in spite of supplementation therapy, suffer from deficiencies of fat soluble vitamins and essential fatty



acids.

Thus, there is a great need for products with properties and structure derived from human lipases and with a broad substrate specificity, which products may be orally administered to patients suffering from deficiency of one or several of the pancreatic lipolytic enzymes. The products of the present invention fulfil this need by themselves or in combination with other lipases or in combination with preparations containing other lipases. Furthermore, for some human infants there is an obvious need to improve fat utilization from conventional infant formulas, or pasteurized human milk from so-called milk banks.

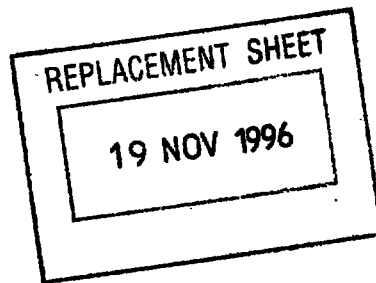
BSSL has several unique properties that makes it ideally suited for substitution and supplementation therapy: It has been designed by nature for oral administration. Thus, it resists passage through the stomach and is activated in contents of the small intestine.

Its specific activation mechanism should prevent hazardous lipolysis of food or tissue lipids during storage and passage to its site of action.

Due to its broad substrate specificity it has the potential to, on its own, mediate complete digestion of most dietary lipids, including the fat soluble vitamin esters.

BSSL may be superior to pancreatic colipase-dependent lipase to hydrolyze ester bonds containing long-chain polyunsaturated fatty acids.

In the presence of gastric lipase and in the absence of, or at low levels of colipase-dependent lipase BSSL can ascertain a complete triacylglycerol digestion in vitro even if the bile salt levels are low such as in newborn infants. In the presence of BSSL the end products of triacylglycerol digestion become free fatty acids and free glycerol rather than free fatty acids and monoacylglycerol generated by the other two lipases (5). This may favour product absorption particularly when the intraluminal bile



salt levels are low (3,15).

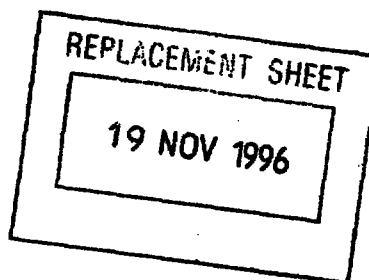
From a historical point of view infant formulas have been developed, and improved, from the concept that their composition should be as similar to that of human milk as possible. It is desirable to supplement such formulas.

The utilization for supplementation, substitution or therapy of bile salt-stimulated lipases (BSSL), or of proteins with the essential functions of BSSL, requires however access to quantities of the product on a large technical scale. It is not possible in factory scale to rely on natural sources such as milk as starting material. Besides the problem mentioned above with inactivation of BSSL during pasteurization, there is the additional risk of contamination of material from a natural source with infectious agents, e.g. vira such as HIV virus and CMV. There is, accordingly, a need for large scale access to products having BSSL properties. The present invention provides such products and methods for their preparation.

Prior art references are given later in this specification.

This Sheet Has Been Replaced





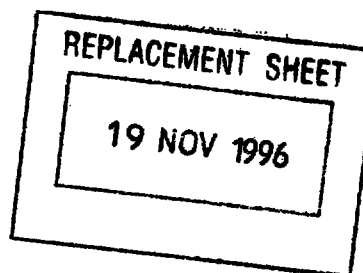
The invention

The present invention is based on the cloning of cDNA coding for BSSL derived from human mammary gland. We have also isolated, from human pancreas, a partial cDNA coding for CEH. Deduced amino acid sequences from the human cDNA's and comparison with CEH from other species, support the interpretation that BSSL and CEH are identical.

As will be further detailed below, it was surprisingly found that the structure of the protein as deduced from the cDNA sequence is quite different from the structure of other lipases. The structure proved unexpectedly to be more like the structure of typical esterases, such as cholinesterase.

This Sheet Has Been Replaced





With reference to Figure II and Figure VII, products of the invention are:

- a) a protein as defined by the amino acid sequence 1-722 in Fig. VII,
- b) a protein as defined by the amino acid sequence 1-535 in Fig. VII,
- c) a protein as defined by the amino acid sequence 1-278 in Fig. VII,
- d) a protein as defined by the amino acid sequence 1-341 in Fig. VII,
- e) a protein as defined by the amino acid sequence 1-409 in Fig. VII,
- f) a protein as defined by the amino acid sequence 1-474 in Fig. VII,
- g) combinations of proteins defined under b) - f) e.g. as defined by the amino acid sequence in positions 1-278, 279-341, 279-409, 279-474, 342-409, 342-474 and 536-722.
- h) combinations of proteins defined under b) - g) in combination with one or more of the repeats according to Figure V,
- i) a protein as defined under a) - h) possessing an additional, N-terminal amino acid, namely methionine, and functionally equivalent variants and mutants of the proteins defined in a) - i) above;

It should be noted that the proteins under a, b, c, d, e, f, h and i above will not be identical in all respects to naturally occurring BSSL, but they will exhibit one or



19 NOV 1996

more of the critical functions of naturally occurring BSSL. Critical functions are given below.

j) a DNA sequence coding for the proteins defined in a, b, c, d, e, f, h and i above.

k) a DNA sequence according to Fig. II, defined by the following nucleotide numbers in Fig. II:

a) a DNA sequence 151-2316 according to Fig. II, coding for the protein defined by the amino acid sequence 1-722 in Fig. VII,

b) a DNA sequence 151-1755 according to Fig. II, coding for the protein defined by the amino acid sequence 1-535 in Fig. VII,

c) a DNA sequence 151-985 according to Fig. II, coding for the protein defined by the amino acid sequence 1-278 in Fig. VII,

d) a DNA sequence 151-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 1-341 in Fig. VII,

e) a DNA sequence 151-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 1-409 in Fig. VII,

f) a DNA sequence 151-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 1-474 in Fig. VII,

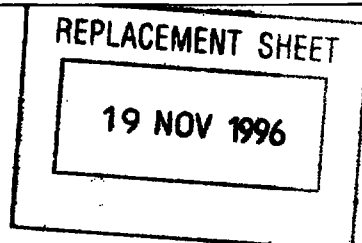
g) a DNA sequence 986-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 279-341 in Fig. VII,

h) a DNA sequence 986-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 279-409 in Fig. VII,

i) a DNA sequence 986-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 279-474 in Fig. VII,

This Sheet Has Been Replaced





This Sheet Has Been Replaced

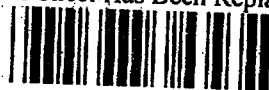
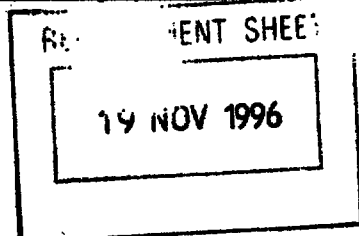


10

- j) a DNA sequence 1173-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 342-409 in Fig. VII,
- k) a DNA sequence 1173-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 342-474 in Fig. VII.

Significant functions of proteins of the invention are

- a) suitable for oral administration,
- b) being activated by specific bile salts,
- c) acting as a non-specific lipase in the contents of the small intestines, that is being able to hydrolyze lipids relatively independent of their chemical structure and physical state (emulsified, micellar, soluble).
- d) Ability to hydrolyze triacylglycerols with fatty acids of different chain-length and different degree of unsaturation.
- e) Ability to hydrolyze also diacylglycerol, monoacylglycerol, cholesteryl esters, lysophosphatidylacylglycerol, and retinyl and other fat soluble vitamin-esters.
- f) Ability to hydrolyze not only the sn-1(3) ester bonds in a triacylglycerol but also the sn-2 ester bond.
- g) Ability to interact with not only primary but also secondary bile salts.
- h) Dependency on bile salts for optimal activity.
- i) Stability so that gastric contents will not affect the catalytical efficiency to any substantial degree.



j) Stability towards inactivation by pancreatic proteases, e.g. trypsin, provided bile salts are present.

k) Ability to bind to heparin and heparin derivatives, e.g. heparan sulphate.

l) Ability to bind to lipid-water interphases.

m) Stability to permit lyophilization.

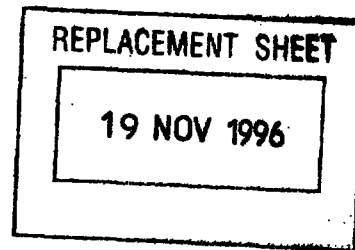
n) Stability when mixed with food constituents such as in human milk, or milk formula.

The critical functions for supplementation, substitution, or therapy are these according to a), c), d), e), f), i), j) and l). For other purposes, not all critical functions may be necessary.

For expression of the proteins indicated above, the appropriate DNA sequence indicated above will be inserted into a suitable vector which then is introduced into a suitable host organism. The said vector will also have to comprise appropriate signal and other sequences enabling the organism to express the desired protein.

Suitable expression organisms:

With the recombinant DNA techniques it is possible to clone and express a protein of interest in a variety of prokaryotic and eukaryotic host organisms. Possible expression organisms are bacteria, simple eukaryotes (yeast), animal cell cultures, insect cell cultures, plant cell cultures, plants and transgenic animals. Each individual system has its own particular advantages and disadvantages. The simple conclusion is that every gene to be expressed is a unique problem and no standard solution



is available.

Commonly used bacterial systems are *E. Coli*, *Bacillus subtilis*, *Streptomyces*. Commonly used yeasts are *saccharomyces*, and *Pichia Pastoris*. Commonly used animal cells are CHO cells and COS cells. Commonly used insect cell cultures are *Drosophila* derived cells.

Commonly used plant is the tobacco plant. Possible transgenic animals are goat and cow.

Possible bacterial vectors are exemplified by pUC and protein A-vectors.

Possible yeast vector is exemplified by pMA 91.

Possible insect vectors are devired from Baculo-virus.

Possible animal cell vectors are derived from SV/40.

Possible plant vectors are derived from the Ti-plasmid.

In every system, both natural and synthetic promoters and terminators can be used.

It is understood that depending on the choice of expression system, the expressed protein may contain an additional N-terminal amino acid (methionine), contain a few extra amino acids, or be fused to a heterologous protein, (e.g. protein A), or differ from the naturally occurring protein with respect to glycosylation. Furthermore, the vectors may also contain signal sequences in order to export the protein to the periplasm or to the culture medium.

Thus, further aspects of the invention are:

- a) a vector comprising a DNA sequence coding for a protein as specified above,
- b) a host organism comprising a DNA sequence as specified above,
- c) a process for the production of a protein as specified above, by growing a host organism containing a vector as specified under a) above and isolating the protein.

This Sheet Has Been Replaced



19 NOV 1996

13

Methods from purification are based on the expression system used (e.g. protein A/IgG) and/or on methods used for purification of the naturally occurring enzyme, as described in reference 6.

Additional aspects of the invention are:

- a pharmaceutical composition comprising a protein as specified above,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of cystic fibrosis,
- the use of a protein as specified above as a supplement to an infant food formulation,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of chronic pancreatitis,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption of any etiology,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons, e.g. in new-

This Sheet Has Been Replaced



19 NOV 1996

born infants.

The DNA sequence in Fig. II from position 151 up to and including position 2316 is the sequence coding for the entire protein. The sequence from position 2317 up to and including position 2415 is not translated to protein, but is included in exon d identified in Table 2 below.

In one embodiment of the invention, the protein as defined in paragraphs a) - i) above is provided in isolated form and/or in substantially pure form.

The DNA sequences as defined in paragraphs a) - k) above are in one embodiment of the invention provided in isolated form and/or in substantially pure form.

Experimental part

Abbreviations

aa, amino acid; bp, base pair; BSSL, bile salt-stimulated lipase; c-AMP, cyclic adenosine monophosphate; CEH, carboxylic ester hydrolase; Da, dalton; c ⁷GTP, 7-deaza-2-deoxyguanosine 5'triphosphate; EDTA, ethylene diamine tetraacetate; kb, kilobases; MOPS, 3-N-morpholino-propanesulfonic acid; nt, nucleotide; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; SSC, NaCl citrate, xGal, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside.

Enzymes

Bile salt-stimulated lipase EC 3.1.1.3

Carboxylic ester hydrolase EC 3.1.1.1

This Sheet Has Been Replaced



19 NOV 1996

Material and Methods

A. Enzyme and antibody preparation

BSSL was purified from human milk as previously described [6]. When used for antibody production the enzyme was further purified by SDS-PAGE. The protein band corresponding to the lipase was, after staining with Coomassie Brilliant blue, electroeluted from the gel. Twentyfive μg of purified enzyme, together with an equal volume of Freund's complete adjuvant, was used for a first i.c. injection and the same amount of enzyme with incomplete adjuvant for the subsequent monthly booster injections. The rabbits were bled about two weeks after each booster and sera prepared and stored at $-20\text{ }^{\circ}\text{C}$.

B. Preparation of tryptic fragments and amino acid sequence analysis

Three mg of purified BSSL was dissolved in 1 ml of 0.1M Tris-Cl buffer, pH 8.5, containing 6M guanidinium hydrochloride and 2 mM EDTA. Dithioerythritol was added to 5mM. After incubation at $37\text{ }^{\circ}\text{C}$ for 2h, 300 μl 50 mM iodoacetate was added. After 90 min incubation at $25\text{ }^{\circ}\text{C}$ in darkness the reduced and carboxymethylated enzyme was desalted on a Sephadex G-25 column, equilibrated with 0.5M ammonium bicarbonate. Thirty μg of tosyl-L-phenylalanine chloromethane treated bovine trypsin (Worthington diagnostics system Inc., Freehold, NJ, USA) was added before lyophilization. The lyophilized protein was dissolved in 4 ml 0.1M ammonium bicarbonate and an additional 90 μg of trypsin was added. After 5h incubation at $37\text{ }^{\circ}\text{C}$ the protein was again lyophilized. The tryptic digest was dissolved in 0.1% trifluoroacetic acid (2mg/ml). Three hundred μg of trypsinated BSSL was chromatographed on HPLC using a C-18 reversed phase column and eluted with a gradient of 0-50% acetonitrile in 0.1%



19 NOV 1996

16

trifluoroacetic acid. Peptide collection was monitored by continuous recording of the absorbance at 215 nm. Peptides to be sequenced were further purified by rechromatography using the same column with adjusted gradients.

Samples of peptide fragments to be sequenced were dried under nitrogen to remove acetonitrile and applied to the sequencer. For N-terminal sequence analysis, native BSSL was dissolved in 0.1% acetic acid. Sequence analyzes were performed on an Applied Biosystems Inc. 477A pulsed liquid-phase sequencer and on-line PTH 120A analyzer with regular cycle programs and chemicals from the manufacturer. Initial and repetitive yields, calculated from a sequenced standard protein, β -lactoglobulin, were 47 and 97%, respectively.

C. Isolation of RNA

Samples of human pancreatic adipose and lactating mammary gland tissues were obtained at surgery and immediately put into guanidinium thiocyanate (1-5 g in 50 ml). Total RNA was extracted as described by Chirgwin [20]. Poly(A)-RNA was prepared by chromatography on oligo-deoxythymidilate-(oligo(DT))-cellulose column [21].

D. Construction and screening of cDNA libraries

Approximately 15 μ g poly-adenylated RNA from human pancreas was denatured with methyl mercuric hydroxide [22] and primed with oligo (dT) ₁₂₋₁₈ primers (Pharmacia, Uppsala, Sweden), and reversely transcribed using standard procedures [23]. Second-strand synthesis was carried out according to Gubler and Hoffman [24], except that DNA ligase and β -NAD were omitted, and the reaction temperature was set at 15 °C. Excess RNA was digested with RNase A (50 μ g/ml), and the double-stranded cDNA was treated with EcoRI methylase [25]. Ends were blunted with Klenow enzyme. After ligation to EcoRI linkers and



19 NOV 1996

cleavage with EcoRI the cDNA was fractionated on a Sepharose 4B-C1 column. The void volume fraction was precipitated with ethanol and the cDNA ligated into the EcoRI site of a phosphatase treated gtl1 vector [26]. In vitro packing yielded more than 7×10^5 recombinants.

A cDNA library from human mammary gland, derived from tissue obtained from a women at the eighth month of pregnancy, was purchased (Clontech Laboratories, Inc., Palo Alto, CA; USA).

Phages from the cDNA libraries were plated at 5×10^4 plaque forming units per 120-mm dish. The antiserum was diluted to a ratio of 1:3200 and screening was performed according to Young and Davis [27]. Alkaline-phosphatase-conjugated goat-anti-rabbit antibodies were used as second antibodies (Bio-Rad, Richmond, CA USA). To isolate clones corresponding to the 5'-end of the mRNA, nucleic acid hybridization was done under standard conditions [23] using a subcloned fragment from one of the immunopositive clones as a probe.

E. RNA analysis

Electrophoresis was carried out in a 1% agarose gel in 40mM MOPS buffer pH 7.0 after denaturation with glyoxal and dimethylsulfoxide [28]. Glyoxalated total RNA was then transferred to nitrocellulose filters [29]. The blots were probed with subclones of BSSL and CEH recombinants that were labelled by the oligo-labeling technique [30]. Prehybridization and hybridization were carried out with 50% formamid at 46 °C [23]. Posthybridization washes were performed at high stringency (0.1% SDS and 0.1xSSC at 60° C). (1xSSC, 0,15M NaCl, 0.0015M Na₃ citrate, pH 7.6).

F. Nucleotide sequence

This Sheet Has Been Replaced



19 NOV 1996

cDNA inserts from BSSL and CEH recombinants were either directly cloned into M3mp18 and mp19 after sonication and size fractionation or some of them were further subcloned into pTZ19R after digestions with PstI, BstXI, NarI, SmaI and AhaII. The nucleotide sequence was determined by the dideoxy chain-termination method [31]. The GC-rich repeats (see below) were also sequenced with TaqI polymerase and dc ⁷GTP. Both strands were sequenced. Sequence information was retrieved from autoradiograms by use of the software MS-EdSeq as described by Sjöberg et al [55].

G. Amino acid sequence predictions and homologies.

To predict the corresponding amino acid sequence of the cDNA inserts, codon usage of different reading frames was compared according to Staden and gave one open reading frame [32]. Homologies were searched for with the programs of the UWGCG software package [33].

Results and Discussion

A. Sequences of tryptic fragments and the N-terminus of BSSL

Trypsin digestion of purified BSSL resulted in approximately 50 fragments as judged by the number of peaks obtained during the HPLC-chromatography (Fig. I). The peaks were collected and the indicated peaks which could be isolated in a highly purified state and in reasonable quantities, were sequenced. The resulting sequences are shown in Table I. In addition the 30 most N-terminal residues were sequenced (Fig. II), and they confirm the previously reported sequence of Abouakil et al. (30 residues) [17].

The 22 residues long N-terminal sequence reported from Wang & Johnson [34] is a glycine in our report and a



19 NOV 1996

lysine in Wang & Johnson's.

B. Nucleotide sequence of BSSL

For construction of the λ gt11 cDNA library we used polyadenylated RNA from human pancreas. Initially four immunopositive clones were isolated, and then this pancreatic expression cDNA library was screened with antiserum against BSSL. Nucleotide sequence analysis of the four clones showed that they are in perfect agreement and correspond to the 3'-end of the mRNA. They all begin with a poly A tail and differ only in length; the longest insert, designated ACEH, spans 996 bp.

A cDNA library from human mammary gland was screened with antiserum, and the pancreas clone ACEH as probe. Positive clones were isolated from both screenings, which all originate from the 3'-end. The longest mammary gland clone, designated ABSSL, reaches 2100 bp upstream. It contains four of the sequenced tryptic fragments (Fig. II), but do not include the N-terminal amino acid sequence. To extend the sequence beyond the translation start, the mammary gland cDNA library was rescreened with a 118 bases long probe derived from the most 5'proximal part of ABSSL. One clone was isolated that continued a further 328 nucleotides upstream. It matched the N-terminal amino acid sequence, and contained the remaining tryptic fragment. As shown in Fig. II, the cDNA is 2428 nucleotides long and contains 81 bases upstream from the first ATG codon. The polyadenylation signal, AATAAA is located 13 nucleotides upstream from the poly A tail and the termination codon TAG was found at nucleotide 2317 followed by a 3'-untranslated region of 112 bp.

A GC rich region consisting of 16 repeats of 33 bases was found in the 3'-end of the sequence between base 1756 and 2283. The nucleotide sequence of the repetition, shown in Fig. III, consists of six identical repetitions surrounded by ten repetitions with different number of



19 NOV 1996

substitutions that have probably occurred after several duplications. The low number of substitutions suggests that these repetitions have appeared late during evolution.

C. Tissue distribution of expression

RNA from human lactating mammary gland, pancreas, adipose tissue and from a human hepatoma cell line (HepG2) was analyzed by Northern blotting. The size of the messenger was determined to be approximately 2.5 kb in both lactating mammary gland and pancreas. No signal could be detected in the lanes with RNA extracted from HepG2 or adipose tissue (Fig. IV).

Since the mRNA used for the mammary gland library was obtained from a female in her 8th month of pregnancy, it is evident that transcription and probably translation of the BSSL gene is turned on before partus, in agreement with previous findings on BSSL secretion before partus [35]. See Figure IV.

D. Amino acid sequence of BSSL

Assessed by SDS-PAGE the molecular mass has been reported to be 107-125 kDa [8,36] and by analytical ultracentrifugation to be 105 kDa [37]. The enzyme, as deduced from the cDNA, consists of 722 amino acid residues (Fig. II) which, giving a molecular mass of 76.271 Da, indicates that the enzyme contains at least 15-20% carbohydrate. The leader sequence is 23 residues long. A tentative active site serine residue is localized to serine-217 (Fig. V). The sequence around this serine accord with the consensus active site sequence of serine-hydrolases [38]. It has recently been it was proposed that basic residues found close to the active site serine may be involved in the cleavage of ester bonds in acylglycerols by lipases [39]. It is interesting to note that such



19 NOV 1996

residues are not present in BSSL. The single tentative N-glycosylation site is localized only seven residues from the serine. The degree of glycosylation [6,16] suggests that the enzyme contains O-linked carbohydrate. There are numerous sites where such glycosylation could have occurred. The amino acid composition based on purified enzyme has shown a high content of proline residues [6]. The amino acid sequence obtained from cDNA confirms this. Moreover, most of the proline residues are localized in the 16 repeats of 11 residues each, constituting the main part of the C-terminal half of the enzyme.

E. Comparison of the enzymes in mammary gland (BSSL) and pancreas (CEH)

BSSL of human milk and human pancreas CEH have previously been shown to be similar, if not identical. The present data strongly suggests that the two enzymes are products of the same gene. The nucleotide sequence of the cDNA clones shows that the pancreatic clone ACEH is identical with the mammary gland clone ABSSL from the poly A tail and 996 bases towards the 5'-end, including the sequence coding for the proline rich repeats. Northern blot gave a single band of 2.5 kb in RNA from pancreas and lactating mammary gland (Fig. IV). Genomic Southern blots further support the idea that only one gene codes for BSSL and CEH. The difference in mobility on SDS-PAGE between BSSL and CEH can be explained as a consequence of different glycosylation or differential splicing.

The similarity of BSSL to the rat and bovine enzymes (see below) and to results from genomics blots support the possibility that differential splicing cannot account for the mobility difference. Since the C-terminal sequence has not been confirmed on the protein level there is a less likely possibility that CEH may be processed by a proteolytic cleavage in the C-terminal end.

This Sheet Has Been Replaced



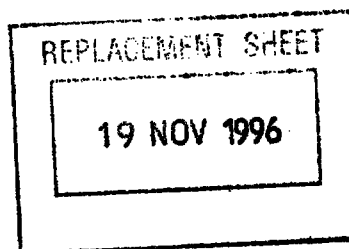
19 NOV 1996

So far as we know pancreatic enzymes that obviously correspond to CEH have often been named after species and the particular substrates used to determine their respective activities; lysophospholipase, cholesteryl esterase, sterol ester hydrolase, non-specific lipase, carboxyl ester lipase and cholesteryl ester hydrolase. Available data are compatible with the view that all these activities described originates in one and the same functional entity [42,43]. This illustrates the broad substrate specificity and the relevance of designating them as non-specific lipases. When the sequence of human BSSL/CEH is compared to the sequence of lysophospholipase from fat pancreas [40] and cholesterol esterase from bovine pancreas [41] extensive similarities are found that extend about 530 residues from the N-terminal (Fig V); but they differ in the part of the molecule where the repeats occur. The rat enzyme has only four repeats and the bovine three. Hence the human enzyme is a considerable longer peptide.

Moreover, striking similarities were found between BSSL and a number of typical esterases, e.g. acetyl choline esterases from several species, including man and Drosophila, and carboxyl esterases (Fig. VI). These similarities were restricted to the N-terminal 300 residues of BSSL which includes the tentative active site serine-residue. A similarity to acetyl choline esterase has been predicted from the fact that BSSL is inhibited by typical choline esterase inhibitors [6, 8, 16]. With the possible exception of the rat liver carboxyl esterase [45), none of these similar enzymes has been shown to have the same bile-salt dependency as BSSL; this suggests that the structural basis for this property resides in the C-terminal part of the protein. Moreover, BSSL can efficiently attack emulsified substrates which is not a known characteristic of the similar esterases. For this activity bile salt is a prerequisite.

This Sheet Has Been Replaced





The predicted sequence for human BSSL was compared with other well characterized mammalian lipases. Apart from the consensus sequence around the active site serine (G-X-S-X-G), no obvious similarities were found [44].

In addition to the similarities with other enzymes, there also significant similarities to one c-AMP dependent protein from Dictyostelium discoideum [46] as well as to thyroglobulin from several species (Fig VI) [47-49]. The similarities between BSSL and thyroglobulin, which comprise the active site region but not the active site itself, indicate that these highly conserved stretches of amino acids are of more generalized importance than merely supporting the enzymatic activity of esterases.

In conclusion, human milk BSSL consists of 722 amino acid residues. Available data strongly indicate that its peptide chain is identical to that of pancreatic CEH, and they are coded for by the same gene. The strongest evidence is that the nucleotide sequences of their 3' ends and their N-terminal amino acid sequences are identical. The striking homologies found to rat pancreatic lysophospholipase and bovine pancreatic cholesterol esterase support the hypothesis that also these enzymes are functionally identical. However, as it has been suggested, the different molecular sizes found among species are not due to differences merely in glycosylation; instead they reflect a variable number of an eleven amino acid repeat. The similarity of the active site sequence between these esterases suggests that these proteins derive from a common ancestral gene.

With reference to Figures I-VII, the following legends are given.

This Sheet Has Been Replaced



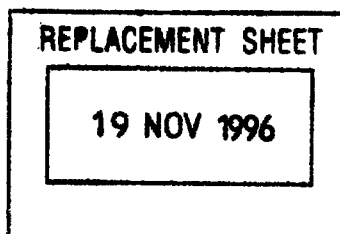


Figure I: Separation of the tryptic digest of BSSL on HPLC

Purified BSSL was treated with trypsin and chromatographed on HPLC as described in Materials and Methods. The indicated peaks were collected and purified further by a rechromatograph and their amino acid sequence determined.

Figure II: The cDNA nucleotide sequence and the deduced amino acid sequence for human bile salt-stimulated lipase:

The cDNA is 2428 bases long. The N-terminal 23-codon sequence (nt82-150) starting with an ATG, is interpreted as a leader peptide since the N-terminal amino acid sequence of the mature protein starts at codon 24 (nt 151, Ala). The leader peptide is underlined. The sign * indicates the starting point of an exon. The sign # indicates the starting point of the repetition part.

Figure III: The nucleotide sequence of the C-terminal GC-rich repetitions in the bile salt-stimulated lipase:

Substitutions are indicated by a *.

Figure IV: Northern blot hybridization

Northern blot analysis of total RNA isolated from human lactating mammary gland, pancreas, adipose tissue and a human hepatoma cell line (HepG2). Total TNA (10µg) from lactating mammary gland (lane A), pancreas (lane B), adipose tissue (lane C) and HepG2 (lane D) were electrophoresed in a 1% agarose gel in 40mM MOPS buffer at Ph 7.0 after denaturation of RNA in 1M glyoxal, 50% dimethylsulfoxide and 40mM MOPS. The glyoxalated RNA was then transferred to nitrocellulose paper for hybridization with [³² P] labeled BSSL cDNA (ABSSL).

This Sheet Has Been Replaced



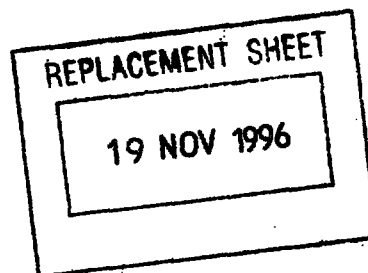


Figure V: Comparison of the deduced amino acid sequence from human milk BSSL, rat pancreatic lysophospholipase (Rat1pl) [40] and bovine pancreatic cholesterol esterase (Bovceh) [41]:

The serine residues involved in the active site are indicated by a *, and the # indicates the single possible N-glycosylation signal of the protein. The direct repeats of amino acid sequences are boxed. Matching sequences are denoted in capital letters, matching sequences between two enzymes are denoted in small letters and mismatching with a dot.

Figure VI: Comparison of the primary structure of BSSL to other esterases, thyroglobuline and to one c-AMP dependent enzyme from Dictyostelium discoideum:

BSSL: bile salt stimulated lipase from human, Cheshum: cholinesterase from fetal human tissue [50], Torpace: acetylcholinesterase from Torpedo marmorata [51], Drosceh: carboxylic ester hydrolase from Drosophila melaogaster [52], Ratlivce: carboxyl esterase from rat liver [53], Drosace: acetylcholinesterase from Drosophila melaogaster [54], Thyrhum: thyroglobulin from human [49] and Dict.Di: c-AMP dependent enzyme from Dictyostelium discoideum [46]. There are 7 different domains that show similarities between the enzymes. Boxes enclose residues which are identical and small letters in the consensus sequence indicate identical residues in all the enzymes except for one. Dots indicate mismatches. The serine residue involved in the active site is indicated with *. The figure in the right hand corner shows how the domains are oriented.

This Sheet Has Been Replaced



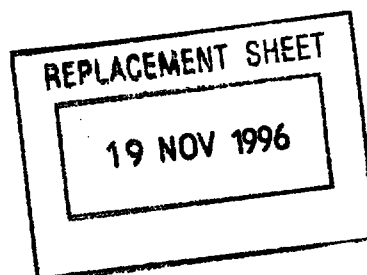


Figure VII:

gives the amino acid sequence 1 - 722 for the entire protein (one letter code) and indicates exons a, b, c, and d. The sign # indicates the starting point of the repetition part.

This Sheet Has Been Replaced



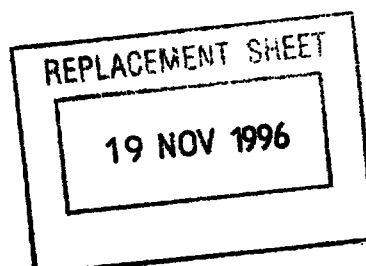


Table 1. Amino acid sequence of BSSL peptides.

Due to interfering peaks no positive identification of the residue in cycles 1 and 2 of the sequencing could be made in peptide number 26. The peptide numbers refer to the peaks in Fig I.

Tryptic fragments

TP16:LysValThrGluGluAspPheTyrLys

TP19:GlyIleProPheAlaAlaProThrLys

TP20:LeuValSerGluPheThrIleThrLys

TP24:ThrTyrAlaTyrLeuPheSerHisProSerArg

TP26:PheAspValTyrThrGluSerTrpAlaGlnAsp

ProSerGlnGluAsnLys

This Sheet Has Been Replaced



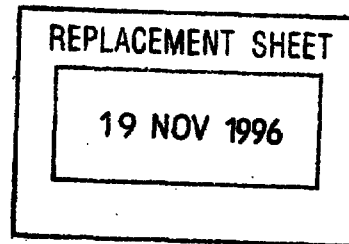


Table 2: Identification of the exons a, b, c and d numbered as in Figures II and VII.

exon	<u>Location</u>	
	between nucleotide number	between aminoacid number

a	986-1172	279-341
b	1173-1376	342-409
c	1377-1574	410-474
d	1575-2415	475-722
the entire protein	151-2316	1-722
the entire protein excluding repetitions	151-1755	1-535

This Sheet Has Been Replaced





References

1. Bläckberg, L., Ångquist, K.A., & Hernell, O. (1987) FEBS Lett. 217, 37-41.
2. Hernell, O. (1975) Eur. J. Clin. Invest. 5, 267-272.
3. Hernell, O., Bläckberg, L. & Bernbäck, S. (1988) In Perinatal nutrition (Lindblad, B.S., ed.) Bristol-Myers Nutrition symposia vol. 6, pp. 259-272, Academic Press, New York.
4. Hernell, O., L. Bläckberg, B. Fredrikzon, and T. Olivecrona. 1981. Bile salt-stimulated lipase in human milk and lipid digestion in the neonatal period. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 465-471.
5. Bernbäck, S., Bläckberg, L. & Hernell, O. (1990) J. Clin. Invest. J. Clin. Invest. 221-226 (1990)
6. Bläckberg, L. & Hernell, O. (1981) Eur. J. Biochem 116, 221-225.
7. Hernell, O. & Bläckberg, L. (1982) Pediatr. Res. 16, 882-885.
8. Bläckberg, L. Lombardo, D., Hernell, O., Guy, O. & Olivecrona, T. (1981) FEBS Lett. 136, 284-288.
9. Fredrikzon, B., Hernell, O., Bläckberg, L. & Olivecrona, T. (1978) Pediatr. Res. 12, 1048-1052.
10. Wang, C.-S., Hartsuck, J.A. & Downs, D. (1988) Biochemistry 27, 4834-4840.
11. Bläckberg, L. & Hernell, O. (1983) FEBS Lett. 157, 337-341.
12. Björkstén, B., Burman, L.G., deChateau, P., Fredrikzon, B., Gothefors, L. & Hernell, O. (1980) Br. Med. J. 201, 267-272.
13. Brueton, M.J., Berger, H.M., Brown, G.A., Ablitt, L., Iyangkaran, N. & Wharton B.A. (1978) Gut 19, 95-98.
14. Murphy, G.M. & Signer, E. (1975) Gut 15, 151-163.
15. Hernell, O., Staggars, J.E. & Carey, M.C. Biochemistry. (1990), 29:2041-2056.

This Sheet Has Been Replaced



19 NOV 1996

16. Lombardo, D., Guy, O. & Figarella, C. (1978) *Biochim. Biophys. Acta* 527, 142-149.
17. Abouakil, N., Rogalska, E., Bonicel, J. & Lombardo, D. (1988) *Biochim. Biophys. Acta* 961, 299-308.
18. Hernell, O., Bläckberg, L. & Lindberg, T. (1988) in *Textbook of Gastroenterology and Nutrition in infancy* (Lebenthal, E., ed) pp. 209-217, Raven Press, New York.
19. Wang, C.-S. (1988) *Biochem. Biophys. Res. Com.* 155, 950-955.
20. Chirgwin, J.M. Przybyla, A.E., MacDonald R.J. & Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
21. Aviv, H. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 69, 5201-5205.
22. Bailey, J.M. & Davidson, N. (1976) *Anal. Biochem.* 70, 75-85.
23. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982): *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
24. Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263-269.
25. Maniatis T., Haldison, R.C., Lacy, E., Lauer, J., O'Connell, C. & Qvon, D. (1978) *Cell* 15, 687-701.
26. Young, R.A. & Davis, R.W. (1983) *Science* 222, 778-782.
27. Young, R.A. & Davis, R.W. (1983) *Proc. Natl. Acad. Sci, USA* 80, 1194-1198.
28. McMaster, G.K. & Charnichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 48 353-48 358.
29. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
30. Feinberg, A.P. & Vogelstein, B. (1983) *Analyt. Biochem.* 132, 6-13.
31. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74. 5463-5467.
32. Staden, R. (1984) *Nucl. Acids Res.* 12, 551-567.
33. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.

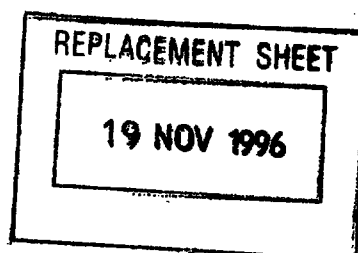
This Sheet Has Been Replaced



19 NOV 1996

34. Wang, C.-S. & Johnson, K. (1983) *Anal. Biochem.* 133, 457-461.
35. Hamosh, M. (1986) in *Human Milk Infant Nutrition and Health* (Howell, R.R. Morriss, F.H. & Pickering, L.K. eds) pp. 66-97, Charles, C. Thomas, Springfield.
36. Wang, C.-S. (1980) *Anal. Biochem.* 105, 398-402.
37. Wang, C.-S & Lee, D.M. (1985) *J. Lipid. Res.* 26, 824-830.
38. Brenner, S. (1988) *Nature* 334, 528-530.
39. Yang, C.-Y., Gu, Z.-W., Yang, H.-H., Rohde, M.F., Gotto, Jr., A.M. & Pownall, H.J. (1989) *J. Biol. Chem.* 265, 16822-16827.
40. Han, J.H., Stratowa, C., & Rutter, W.J. (1987) *Biochemistry* 26, 1617-1625.
41. Kyger, E., Wiegand, R., & Lange, L. (1989) *Biochim. Biophys. Res. Com* 164, 1302-1309.
42. Bläckberg, L. (1981) *Fat digestion in newborn infant. Umeå University Medical Dissertations New Series No 71.*
43. Rudd, E.A. & Brockman, H.L. (1984) in *Lipases* (Borgström, B & Brockman, H.L. eds) pp. 184-204, Elsevier, Amsterdam.
44. Mickel, S., Weidenbach, F., Swarovsky, B., LaForge, S. & Scheele, G. (1989) *J. Biol. Chem.* 264, 12895-12901.
45. Cammulli, E.D., Linke, M.J., Brockman, H.L. & Hui, D.Y. (1989) *Biochim. Biophys. Acta* 1005, 177-182.
46. Mann, S.K.O. & Firtel, R.A. (1987) *Mol. Cell Biol.* 7, 458-469.
47. Mercken, L., Simons, M.J., Swillens, S., Masser, M. & Vassart, G. (1985) *Nature* 316, 647-651.
48. Lauro, R., Obici, S. Condliffe, D., Ursini, V.M., Musti, A., Moscatelli, C. & Avvedimento, V.E. (1985) *Eur. J. Biochem.* 148, 7-11.
49. Malthiery, Y. & Lissitzky, S. (1987) *Eur. J. Biochem.* 165, 491-498.
50. Prody, C., Zevin-Sonkin, D., Gnatt, A., Goldberg, O.,





- & Soreq, H. (1987) Proc. Natl. Acad. Sci. USA 84, 3555-3559.
51. Sikorav, J-L., Krejci, E. & Massoulie', J. (1987) EMBO J. 6, 1865-1873.
 52. Oakeshott, J.G. Collet, C., Phillis, R.W., Nielsen, K.M., Russel, R.J., Cambers, G.K., Ross, V. & Richmond, R.C. (1987) Proc. Natl. Acad. Sci, USA 84, 3359-3363.
 53. Long, R., Satho, H., Martin, B., Kimura, S., Gonzalez, F. & Pohl, L. (1988) Biochem. Biophys. Res. Comm. 156, 866-873.
 54. Hall, L.M.C. & Spierer, P. (1986) EMBO J. 5, 2949-2954.
 55. Sjöberg, S., Carlsson, P., Enerbäck, S., & Bjursell, G. (1989) CABIOS 5, 41-46.
 56. EP-A-317355.
 57. Hernell, O., and T. Olivecrona. 1974. Human milk lipases II. Bile salt-stimulated lipase. Biochim. Biophys. Acta 369:234-244.
 58. Hernell, O., L. Bläckberg, and T. Olivecrona. 1981. Human milk lipases. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 347-354.
 59. Atkinson, S.A., M.H. Bryan, and G.H. Andersson. 1981. Human milk feeding in premature infants: protein, fat and carbohydrate balances in the first two weeks of life. J.Pediatr. 99:617-624.
 60. Chappell, J.E., M.T. Clandinin, C. Kearney-Volpe, B. Reichman and P.W. Swyer. 1986. Fatty acid balance studies in premature infants fed human milk or formula: effect of calcium supplementation. J. Pediatr. 108:438-447.



REPLACEMENT SHEET

19 NOV 1996

33

61. Williamson, S., E. Finucane, H. Ellis, and H.R. Gamsu. 1978. Effect of heat treatment of human milk on absorption of nitrogen, fat, sodium, calcium and phosphorus by preterm infants. Arch. Dis. Childhood 53:555-563.

This Sheet Has Been Replaced





34

REPLACEMENT SHEET

19 NOV 1996

What we claim is:

1. A protein as indicated in Figure VII from position 1 to position 722.
2. A protein as indicated in Figure VII from position 1 to position 535.
3. A protein as indicated in Figure VII from position 1 to position 278.
4. A protein as indicated in Figure VII from position 1 to position 341.
5. A protein as indicated in Figure VII from position 1 to position 409.
6. A protein as indicated in Figure VII from position 1 to position 474.
7. A protein as indicated in Figure VII from position 1 to position 278, position 279 to position 341, position 279 to position 409, position 279 to position 474, position 342 to position 409, position 342 to position 474, or position 536 to position 722.
8. A protein according to any of claims 1-7 in combination with one or more of the repeats according to Figure V.
9. A protein according to any of claims 1-8, possessing methionine as additional N-terminal amino acid.
10. A protein according to any of claims 1-9, exhibiting one or several of the critical functions of naturally occurring bile salt-stimulated lipase.

This Sheet Has Been Replaced



19 NOV 1996

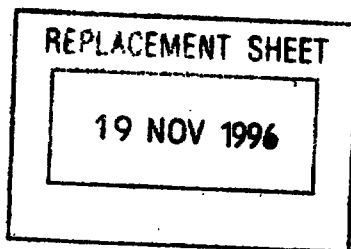


11. A functionally equivalent variant or mutant of a protein according to any of claims 1-9.
12. A DNA sequence coding for a protein with the amino acid sequence according to claim 1.
13. A DNA sequence coding for a protein with the amino acid sequence according to claim 2.
14. A DNA sequence coding for a protein with the amino acid sequence according to any of claims 3, 4, 5, 6, 7, 8, 9, 10 and 11.
15. A DNA sequence defined by the following nucleotide numbers in Figure II:

151-2316
151-1755
151-985
151-1172
151-1376
151-1574
986-1172
986-1376
986-1574
1173-1376
1173-1574
16. A vector comprising a DNA sequence coding for a protein according to claims 1 - 11
17. A vector comprising a DNA sequence according to claims 12 - 15.
18. A host organism transformed with a vector according to claims 16 or 17.

This Sheet Has Been Replaced





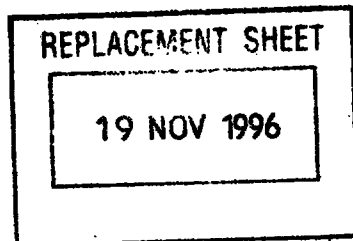
36



19. A process for the production of a protein according to claims 1 - 11 by growing a host organism containing a vector according to claims 16 or 17, and isolating the protein.
20. A pharmaceutical composition comprising a protein according to any of claims 1 - 11.
21. A pharmaceutical composition comprising a protein according to any of claims 1-11 in combination with a lipase or in combination with preparations containing lipases.
22. The use of a protein according to any of claims 1 - 11, for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency.
23. The use of a protein according to any of claims 1 - 11 for the manufacture of a medicament for the treatment of cystic fibrosis.
24. The use of a protein according to any of claims 1 - 11 as a supplement to an infant food formulation.
25. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of chronic pancreatitis.
26. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption.
27. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins.

This Sheet Has Been Replaced





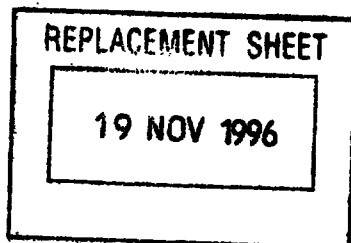
37



28. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons.
29. The use, according to claims 22-28 of a protein according to any of claims 1-11 in combination with a lipase or lipases or in combination with preparations containing a lipase or lipases.
30. An infant food formulation, supplemented with a protein according to any of claims 1 - 11.
31. A protein according to claims 1 - 11 in substantially pure form.
32. A protein according to claims 1 - 11 in isolated form.
33. A DNA sequence according to claims 12 - 15 in substantially pure form.
34. A DNA sequence according to claims 12 - 15 in isolated form.
35. A process for the preparation of a pharmaceutical composition according to claims 20 or 21 by incorporating a protein according to claims 1 - 11, 31, or 32 in a pharmacologically acceptable carrier.
36. A process according to claim 35 for the preparation of a pharmaceutical composition for oral administration.
37. A process for the preparation of an infant food formulation according to claim 30, by supplementing an infant food formulation with a protein according to claims 1 -11, 31, or 32.

This Sheet Has Been Replaced





DERIVATIVES OF HUMAN BILE-SALT STIMULATED LIPASE,
AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM

Abstract

The present invention relates to new DNA sequences, to new proteins coded for by such DNA sequences, and to the use as further described below of such proteins. The invention also encompasses vectors, such as plasmid constructs, comprising such DNA sequences, being capable of expressing the desired enzyme. The invention also includes host organisms transfected with such constructs e.g. bacteria yeast, mammalian cells, and transgenic animals. The invention also includes processes for the preparation of the novel products of the invention. The new proteins of the invention are related to an enzyme, known i.a. as human bile salt-stimulated lipase.

FIGURE 2.

This Sheet Has Been Replaced



REPLACEMENT SHEET

19 NOV 1996

- 1 -

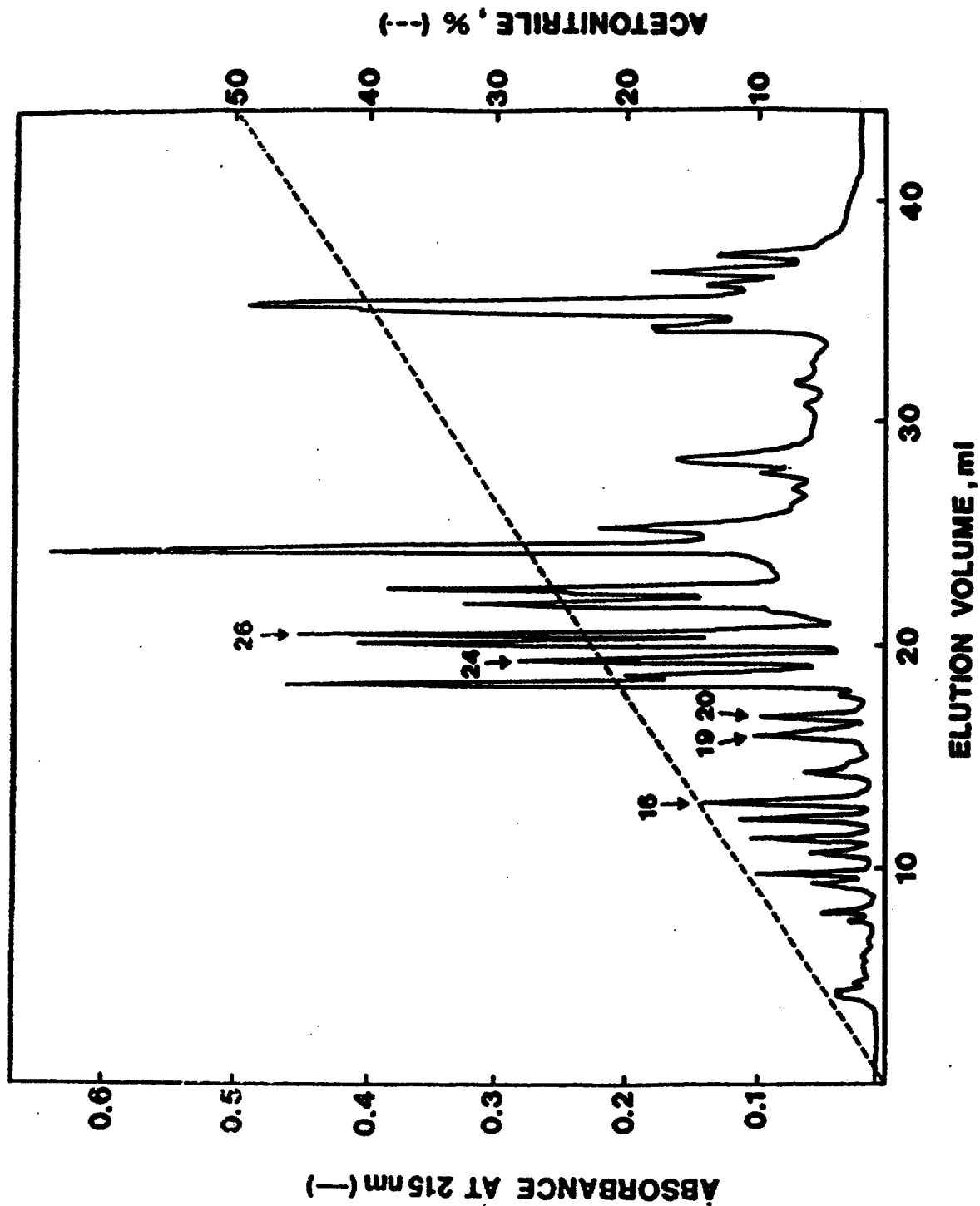
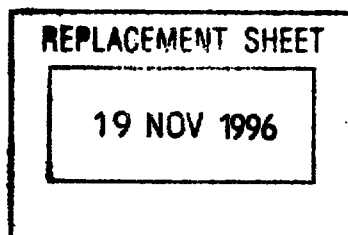


Figure I

This Sheet Has Been Replaced





- 2 -

10 30 50
ACCTTCTGTATCAGTTAAGTGTC AAGATGGAAGGAACAGCAGTCTCAAGATAATGCAAAG

70 90 110
AGTTTATTTCATCCAGAGGCTGATGCTCACCATGGGGCGCCTGCAACTGGTTGTGTTGGGC

MetLeuThrMetGlyArgLeuGlnLeuValValLeuGly

130 150 170
CTCACCTGCTGCTGGGCAGTGGCGAGTGCCGCGAAGCTGGGCGCCGTGTACACAGAAGGT

LeuThrCysCysTrpAlaValAlaSerAlaAlaLysLeuGlyAlaValTyrThrGluGly

190 210 230
GGGTTCGTGGAAGGCGTCAATAAGAAGCTCGGCCTCCTGGGTGACTCTGTGGACATCTTC

GlyPheValGluGlyValAsnLysLysLeuGlyLeuLeuGlyAspSerValAspIlePhe

250 270 290
AAGGGCATCCCCTTCGCAGCTCCCACCAAGGCCCTGGAAAATCCTCAGCCACATCCTGGC

LysGlyIleProPheAlaAlaProThrLysAlaLeuGluAsnProGlnProHisProGly

310 330 350
TGGCAAGGGACCCTGAAGGCCAAGAACTTCAAGAAGAGATGCCTGCAGGCCACCATCACC

TrpGlnGlyThrLeuLysAlaLysAsnPheLysLysArgCysLeuGlnAlaThrIleThr

370 390 410
CAGGACAGCACCTACGGGGATGAAGACTGCCTGTACCTCAACATTTGGGTGCCCCAGGGC

GlnAspSerThrTyrGlyAspGluAspCysLeuTyrLeuAsnIleTrpValProGlnGly

430 450 470
AGGAAGCAAGTCTCCCGGGACCTGCCCGTTATGATCTGGATCTATGGAGGCGCCTTCCTC

ArgLysGlnValSerArgAspLeuProValMetIleTrpIleTyrGlyGlyAlaPheLeu

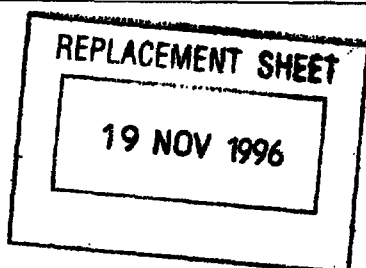
490 510 530
ATGGGGTCCGGCCATGGGGCCAACTTCCTCAACAACCTACCTGTATGACGGCGAGGAGATC

MetGlySerGlyHisGlyAlaAsnPheLeuAsnAsnTyrLeuTyrAspGlyGluGluIle

Figure II/1

This Sheet Has Been Replaced





- 3 -

550 570 590
GCCACACGCGGAAACGTCATCGTGGTCACCTTCAACTACCGTGTCGGCCCCCTTGGGTTC
AlaThrArgGlyAsnValIleValValThrPheAsnTyrArgValGlyProLeuGlyPhe
610 630 650
CTCAGCACTGGGGACGCCAATCTGCCAGGTA ACTATGGCCTTCGGGATCAGCACATGGCC
LeuSerThrGlyAspAlaAsnLeuProGlyAsnTyrGlyLeuArgAspGlnHisMetAla
670 690 710
ATTGCTTGGGTGAAGAGGAATATCGCGGCCTTCGGGGGGGACCCCAACAACATCACGCTC
IleAlaTrpValLysArgAsnIleAlaAlaPheGlyGlyAspProAsnAsnIleThrLeu
730 750 770
TTCGGGGAGTCTGCTGGAGGTGCCAGCGTCTCTCTGCAGACCCTCTCCCCCTACAACAAG
PheGlyGluSerAlaGlyGlyAlaSerValSerLeuGlnThrLeuSerProTyrAsnLys
790 810 830
GGCCTCATCCGGCGAGCCATCAGCCAGAGCGGCGTGGCCCTGAGTCCCTGGGTCATCCAG
GlyLeuIleArgArgAlaIleSerGlnSerGlyValAlaLeuSerProTrpValIleGln
850 870 890
AAAAACCCACTCTTCTGGGCCAAAAGGTGGCTGAGAAGGTGGGTTGCCCTGTGGGTGAT
LysAsnProLeuPheTrpAlaLysLysValAlaGluLysValGlyCysProValGlyAsp
910 930 950
GCCGCCAGGATGGCCCAGTGTCTGAAGGTTACTGATCCCCGAGCCCTGACGCTGGCCTAT
AlaAlaArgMetAlaGlnCysLeuLysValThrAspProArgAlaLeuThrLeuAlaTyr
970 990 1010
AAGGTGCCGCTGGCAGGCCTGGAGTACCCCATGCTGCACTATGTGGGCTTCGTCCCTGTC
LysValProLeuAlaGlyLeuGluTyrProMetLeuHisTyrValGlyPheValProVal
1030 1050 1070
ATTGATGGAGACTTCATCCCCGCTGACCCGATCAACCTGTACGCCAACGCCGCCGACATC
IleAspGlyAspPheIleProAlaAspProIleAsnLeuTyrAlaAsnAlaAlaAspIle

Figure II/2

This Sheet Has Been Replaced



REPLACEMENT SHEET

19 NOV 1996

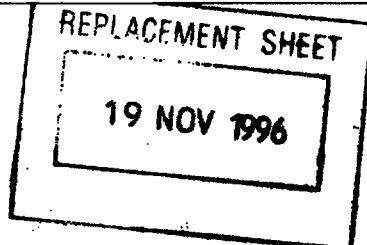
- 4 -

1090	1110	1130
GACTATATAGCAGGCACCAACAACATGGACGGCCACATCTTCGCCAGCATCGACATGCCT		
AspTyrIleAlaGlyThrAsnAsnMetAspGlyHisIlePheAlaSerIleAspMetPro		
1150	1170	1190
GCCATCAACAAGGGCAACAAGAAAGTCACGGAGGAGGACTTCTACAAGCTGGTCAGTGAG		
AlaIleAsnLysGlyAsnLysLysValThrGluGluAspPheTyrLysLeuValSerGlu		
1210	1230	1250
TTCACAATCACCAAGGGGCTCAGAGGCGCCAAGACGACCTTTGATGTCTACACCGAGTCC		
PheThrIleThrLysGlyLeuArgGlyAlaLysThrThrPheAspValTyrThrGluSer		
1270	1290	1310
TGGGCCCAGGACCCATCCCAGGAGAATAAGAAGAAGACTGTGGTGGACTTTGAGACCGAT		
TrpAlaGlnAspProSerGlnGluAsnLysLysLysThrValValAspPheGluThrAsp		
1330	1350	1370
GTCCTCTTCCTGGTGCCACCGAGATTGCCCTAGCCCAGCACAGAGCCAATGCCAAGAGT		
ValLeuPheLeuValProThrGluIleAlaLeuAlaGlnHisArgAlaAsnAlaLysSer		
1390	1410	1430
GCCAAGACCTACGCCTACCTGTTTTCCCATCCCTCTCGGATGCCCGTCTACCCCAAATGG		
AlaLysThrTyrAlaTyrLeuPheSerHisProSerArgMetProValTyrProLysTrp		
1450	1470	1490
GTGGGGGCGGACCATGCAGATGACATTCAGTACGTTTTCGGGAAGCCCTTCGCCACCCCC		
ValGlyAlaAspHisAlaAspAspIleGlnTyrValPheGlyLysProPheAlaThrPro		
1510	1530	1550
ACGGGCTACCGGCCCCAAGACAGGACAGTCTCTAAGGCCATGATCGCCTACTGGACCAAC		
ThrGlyTyrArgProGlnAspArgThrValSerLysAlaMetIleAlaTyrTrpThrAsn		
1570	1590	1610
TTTGCCAAAACAGGGGACCCCAACATGGGCGACTCGGCTGTGCCACACACTGGGAACCC		
PheAlaLysThrGlyAspProAsnMetGlyAspSerAlaValProThrHisTrpGluPro		

Figure II/3

This Sheet Has Been Replaced





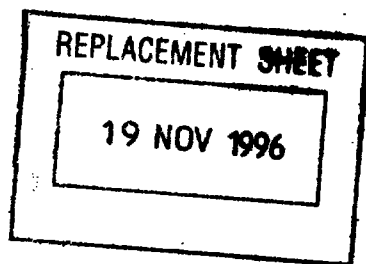
- 5 -

1630 1650 1670
TACACTACGGAAAACAGCGGCTACCTGGAGATCACCAAGAAGATGGGCAGCAGCTCCATG
TyrThrThrGluAsnSerGlyTyrLeuGluIleThrLysLysMetGlySerSerSerMet
1690 1710 1730
AAGCGGAGCCTGAGAACCAACTTCCTGCGCTACTGGACCCTCACCTATCTGGCGCTGCCC
LysArgSerLeuArgThrAsnPheLeuArgTyrTrpThrLeuThrTyrLeuAlaLeuPro
1750 1770 1790
ACAGTGACCGACCAGGAGGCCACCCCTGTGCCCCCACAGGGGACTCCGAGGCCACTCCC
ThrValThrAspGlnGluAlaThrProValProProThrGlyAspSerGluAlaThrPro
1810 1830 1850
GTGCCCCCACGGGTGACTCCGAGACCGCCCCCGTGCCGCCACGGGTGACTCCGGGGCC
ValProProThrGlyAspSerGluThrAlaProValProProThrGlyAspSerGlyAla
1870 1890 1910
CCCCCGTGCCGCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC
ProProValProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSer
1930 1950 1970
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCACGGGT
GlyAlaProProValProProThrGlyAspSerGlyAlaProProValProProThrGly
1990 2010 2030
GACTCCGGGGCCCCCCCCCGTGCCGCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCC
AspSerGlyAlaProProValProProThrGlyAspSerGlyAlaProProValProPro
2050 2070 2090
ACGGGTGACTCCGGCGCCCCCCCCCGTGCCGCCACGGGTGACGCCGGGCCCCCCCCCGTG
ThrGlyAspSerGlyAlaProProValProProThrGlyAspAlaGlyProProProVal
2110 2130 2150
CCGCCCACGGGTGACTCCGGCGCCCCCCCCCGTGCCGCCACGGGTGACTCCGGGGCCCC
ProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSerGlyAlaPro

Figure II/4

This Sheet Has Been Replaced





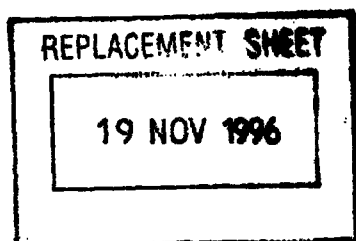
- 6 -

2170	2190	2210
CCCGTGACCCCCACGGGTGACTCCGAGACCGCCCCCGTGCCGCCCACGGGTGACTCCGGG		
ProValThrProThrGlyAspSerGluThrAlaProValProProThrGlyAspSerGly		
2230	2250	2270
GCCCCCCTGTGCCCCCACGGGTGACTCTGAGGCTGCCCTGTGCCCCCACAGATGAC		
AlaProProValProProThrGlyAspSerGluAlaAlaProValProProThrAspAsp		
2290	2310	2330
TCCAAGGAAGCTCAGATGCCTGCAGTCATTAGGTTTTAGCGTCCCATGAGCCTTGGTATC		
SerLysGluAlaGlnMetProAlaValIleArgPhe		
2350	2370	2390
AAGAGGCCACAAGAGTGGGACCCCAGGGGCTCCCTCCCATCTTGAGCTCTTCCTGAATA		
2410		
AAGCCTCATACCCCTAAAAAAAAAAAAA		

This Sheet Has Been Replaced



Figure II/5



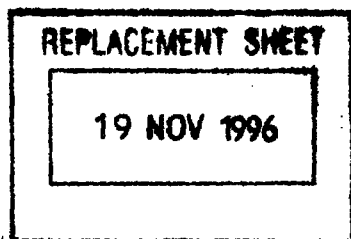
- 7 -

		Number of substitutions	
GAGGCCACCCCTGTGCCCCCACAAGGACTCC	1	0	
GAGGCCCTCCCGTCCCACCGGCTGACTCC	2	4	
GAGACCCCCCGTCCCCCACCCTGACTCC	3	3	
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	4	} Identical to consensus	0
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	5		0
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	6		0
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	7		0
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	8		0
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	9		0
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	10	2	
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	11	1	
GGGGCCCCCCCCGTCCCCCACCCTGACTCC	12	1	
GGGGCCCCCCCCGTGACCCCCACCGGCTGACTCC	13	2	
GAGACCCCCCGTCCCCCACCCTGACTCC	14	3	
GGGGCCCCCGTGTCCCCCACCCTGACTCT	15	3	
GAGGTCCTCCCTGTCCCACAGTACTCC	16	7	
GGGGCCCCCCCCGTCCCCCACCCTGACTCC Consensus			

Figure III

This Sheet Has Been Replaced





- 8 -

A B C D Kb

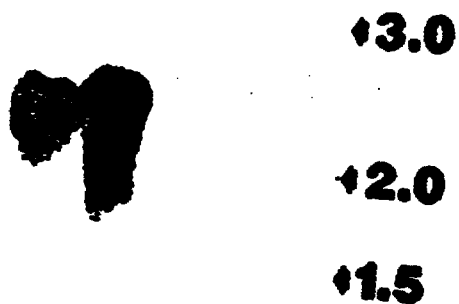


Figure IV

This Sheet Has Been Replaced



This Sheet Has Been Replaced

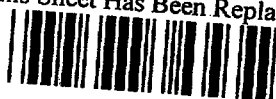


Fig. 5⁹(1/3)

REPLACEMENT SHEET

19 NOV 1996

50
Bssl MLTMGRLQLVVLGLTCCWAVASAAKLGAVYTEGGFVEGVNKKLGLLG.DS
Ratlpl ...MGRLEVLFGLTCCLAACAAGALYTEGGFVEGVNKKLSLGGDS
BovcehMLGASRLGPSPGCLAVASAAKLGSVYTEGGFVEGVNKKLSLFG.DS
Consensusgrl....lgltcclaaaaAAKLGavYTEGGFVEGVNKKLSLlg.DS

100
Bssl VDIFKGIPFAAPTKALENPQPHPGWQGTLLKAKNFKKRCLOATITQDSTYG
Ratlpl VDIFKGIPFA.TAKTLENPQRHDPGWQCTLKATQFKKRCLOATLTQDDTYG
Bovech VDIFKGIPFAAPKALEKPKRHPGWQGTLLKAKSFKKRCLOATLTQDSTYG
Consensus VDIFKGIPFAa..KaLEnPqrHPGWQGTLLKak.FKKRCLOATiTQDsTYG

150
Bssl DEDCLYLNIWVPQGRKQVSRDLPVMIWIYGGAFIMGSGBGANFLNNYLYD
Ratlpl QEDCLYLNIWVPQGRKQVSHDLPVMVWIYGGAFIMGSGQGANFLKNYLYD
Bovceh NEDCLYLNIWVPQGRKEVSHDLPVMIWIYGGAFIMGASQGANFLSNYLYD
Consensus .EDCLYLNIWVPQGRKqVShDLPVmiWIYGGAFIMGsgqGANFL.NYLYD

200
Bssl GKKIATRGNVIVVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKR
Ratlpl GKKIATRANVIVVTFNYRVGPLGFLSTGDANLPGNFGLRDQHMAIAWVKR
Bovech GKKIATRGNVIVVTFNYRVGPLGFLSTGDSNLPGNYGLWDQHMAIAWVKR
Consensus GKKIATRgNVIVVTFNYRVGPLGFLSTGDaNLPgNYGLrDQHMAIAWVKR

* 250
Bssl NIAAFGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSP
Ratlpl NIAAFGGDPNNITIFGESAGGAIVSLQTLSPYNKGLIRRAISQSGVALSP
Bovceh NIEAFGGDPDNITLFGESAGGASVSLQTLSPYNKGLIKRGISQSGVGLCP
Consensus NIaAFGCdPDNITLFGESAGGAsVSLQTLSPYNKGLIrRaISQSGVaLsP

300
Bssl WVIQKNPLFWAKKVAKKVGCVPVGDARMAQCLKVTDPRALTLAYKVPLAG
Ratlpl WAIQENPLFWAKTIAKKVGCPTEDTAKMAGCLKITDPRALTLAYRLPLKS
Bovech WAIQQDPLFWAKRIAKKVGCVPVDDTSKMAGCAKITDPRALTLAYKLPLGS
Consensus WaIQ.nPLFWAK.iAKKVGCpv.DtakMAGclKiTDPRALTLAYklPL.s

350
Bssl LEYPMLHYVGFVPVIDGDFIPADPINLYANAADIDYIAGTNNMDGHIFAS
Ratlpl QEYPIVEYLAFIPVVDGDFIPDDPINLYDNAADIDYLAGINDMDGHLFAT
Bovceh TEYPKLHYLSFVPVIDGDFIPDDPVNLYANAADVYIAGTNDMDGHLFVG
Consensus .EYP.lHYl.FvPViDGDFIPdDPiNLYaNAADiDYiAGtNdMDGHlPa.

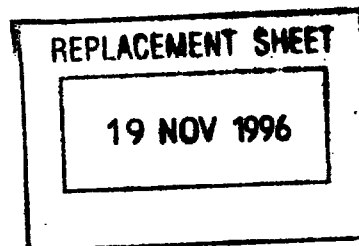


Fig. 5 (2/3)

400

Bssl	IDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGAKTTPDVYTESWAQDPS
Ratlpl	VDVPAIDKAKQDVTEEDFYRLVSGHTVAKGLKGTQATFDIYTESWAQDPS
Bovech	MDVPAINSNKQDVTEEDFYKLVSGLTVTKGLRGANATYEVYTEPWAQDSS

Consensus .DvPAInk.kqdVTEEDFYKLVSg.TvtKGLrGa.aTfdvYTESWAQDps

450

Bssl	QENKKKTVDVFETDVLFLVPTEIALAQHRANAKSAKTYAYLFSHPSRMPV
Ratlpl	QENMKKTVAFETDILFLIPTEMALAQHRAHAKSAKTYSYLFSHPSRMPi
Bovceh	QETRKKTMVDLETDILFLIPTKIAVAQHKSHAKSANTYTYLFSQPSRMPi

Consensus QEn.KKTvVdfETDiLFLiPTeialAQHrahAKSAkTY.YLFSHPSRMPi

500

Bssl	YPKWVGADHADDIQYVFGKPFATPTGYRPQDRTVSKAMIAWNTNFAKTGD
Ratlpl	YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAWNTNFAKSGD
Bovech	YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAWNTNFARTGD

Consensus YPKWmGADHADDlQYVFGKPFATPlGYRaQDRTVSKAMIAWNTNFAktGD

550

Bssl	PNMGDSAVPTHWEPTYTENS GYLEITKKMGSSSMKRS LR TNFLRYWTLTY
Ratlpl	PNMGNSPVPTHWPYPTMENGNYLDINKKITSTSMKEHLREKFLKFWAVTF
Bovceh	PNTGHSTVPANWDPYTLEDDNYLEINKQMDSNSMKLR TL TNYLQFWTQTY

Consensus PNnG.S.VPthW.PYT.Kn.nyleInKkm.S.SMK.hLRtnfL.fWt.Ty

596

Bssl	LALPTVTDQ	EATVPVPTGDS	EATVPVPTGDS	ETAPVPPTGDS	GAPP
Ratlpl	EMLPTV...	VGDHTPPKDDS	EAAPVPPTDDS	DGGVPVPTDDS	QTTP
Bovech	QALPTV TSA	GASLLPPEDNS	QASPVPPADNS	GAPTEPSAGDS	

Consensus .aLPTvt....a...PP.dds.eA.FVPpTdds....pvPptgDS....p

642

Bssl	VPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVP
Ratlpl	VPPTDNS	QA.....			
Bovceh					

Consensuss..a.....



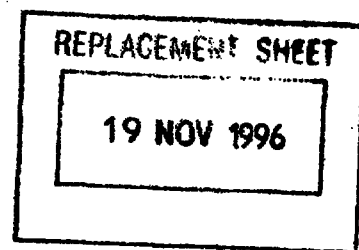


Fig.5 (3/3)

				689
Bssl	PTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GPPFVTPTGDS
Rar1pl				
Bovech				
Consensus				

				735
Bssl	DS	GAPPVTPTGDS	ETAPVPPTGDS	GAPPVPPTGD
Rat1pl	..			
Bovceh	..			
Consensus				ds

Bssl KE.AQMPAVIRF
 Rat1pl VE.AQMPGPIGF
 Bovech .EVAQMPVVIGF
 Consensus .e.AQMP.vIgF

This Sheet Has Been Replaced



19 NOV 1996

Fig.6(1/3)

This Sheet Has Been Replaced



1

- 12 -

(76-90)

BSSL

(116-130)

ChesHum

(114-128)

Torpace

(101-115)

Drosceh

(103-117)

Ratlivce

(224-238)

Drosace

(2257-2270)

ThryHum

(90-114)

Dict.Di

```

G D E D C L Y L N I W V P Q G
L S E D C L Y L N I W I P A P
M S E D C L Y L N I W V P S P
G E E D C L T V S V Y K P K N
F S E D C L Y I N V W A P A K
V S E D C L Y L N V W A P A K
V S E D C L Y L N V F I P Q .
A Q K C N L G P G V C S P M G

```

Consensus

. s e d c l y l n v w . P . .

2

(95-113)

BSSL

(134-150)

ChesHum

(132-150)

Torpace

(118-136)

Drosceh

(121-138)

Ratlivce

(273-291)

Drosace

(2270-2291)

ThryHum

```

S R D L P V M I W I Y G G A F L M G S
P K D A T V L I W I Y G G G F Q T G T
P K S A T V M L W I Y G G G F Y S G S
R N S F P V V A H I H G G A F M F G A
N S R L P V M V W I H G G G L I I G G
T N G L P I L I W I Y G G G F M T G S
A P N A S V L V F F H N T M D R E E S

```

Dict.Di

.

Consensus

. . . l p v . . w i y g g g f . . g s

3

(135-156)

BSSL

(166-187)

ChesHum

(164-185)

Torpace

(150-171)

Drosceh

(153-174)

Ratlivce

(307-328)

Drosace

(2307-2328)

ThryHum

(118-136)

Dict.Di

```

N V I V V T F N Y R V G P L G F L S T G D A
R V I V V S M N Y R V G A L G F L A L P G N
E V V L V S L S Y R V G A F G F L A L H G S
K F I L V K I S Y R L G P L G F V S T G D E
N V V V V T I Q Y R L G F G G L F S T G D E
N V I V A S F Q Y R V G A F G F L H L A P E
N L I V V T A S Y R V G V F G F L S S G S G
S V I V V T I N Y R L G I L G L M G T . . .

```

Consensus

n v i v v t f n Y R v g . . G f l s t a d .

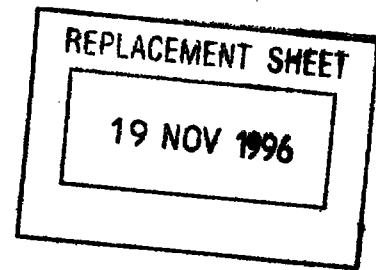


Fig.6(2/3)

4
(157-182)
BSSL
(189-214)
ChesHum
(187-212)
Torpace
(172-197)
Drosceh
(175-200)
Ratlivce
(336-361)
Drosace
(2329-2354)
ThryHum

N	L	P	G	N	Y	G	L	R	D	Q	H	M	A	I	A	W	V	K	R	N	I	A	A	F	G
E	A	P	G	N	M	G	L	F	D	Q	Q	L	A	L	Q	W	V	Q	K	N	I	A	A	F	G
E	A	P	G	N	M	G	L	L	D	Q	R	M	A	L	Q	W	V	H	D	N	I	Q	F	F	G
D	L	P	G	N	Y	G	L	K	D	Q	R	L	A	L	K	W	I	K	Q	N	I	A	S	F	G
H	S	R	G	N	W	A	H	L	D	Q	L	A	A	L	R	W	V	Q	D	N	I	A	N	F	G
E	A	P	G	N	V	G	L	W	D	Q	A	L	A	I	R	W	L	K	D	N	A	H	A	F	G
E	V	S	G	N	N	G	L	L	D	Q	V	A	A	L	T	W	V	O	T	H	I	R	G	F	G

Dict.Di

.....

Consensus

e l p G N w q l l D Q . . A l . W v . d n i a a F G

5
(183-207)
BSSL
(215-239)
ChesHum
(213-237)
Torpace
(198-222)
Drosceh
(201-225)
Ratlivce
(362-386)
Drosace
(2354-2379)
ThryHum
(137-149)
Dict.Di

G	D	P	N	N	I	T	L	F	G	E	S	A	G	G	A	S	V	S	L	Q	T	L	S	P
G	N	P	K	S	V	T	L	F	G	E	S	A	G	A	A	S	V	S	L	H	L	L	S	P
G	D	P	K	T	V	T	L	F	G	E	S	A	G	G	A	S	V	G	M	H	I	L	S	P
G	E	P	O	N	V	L	L	V	G	H	S	A	G	G	A	S	V	H	L	Q	M	L	R	E
G	N	P	D	S	V	T	I	F	G	E	S	A	G	G	V	S	V	S	A	L	V	L	S	P
G	N	P	E	W	M	T	L	F	G	E	S	A	G	S	S	S	V	N	A	Q	L	M	S	P
G	P	P	R	R	V	S	L	A	A	D	R	G	G	A	D	V	A	S	I	H	L	L	T	A
										 A G A F S V S A H L T S T														

Consensus

G d P . n v t l f g e s a G g a s v s l . l l s p

This Sheet Has Been Replaced



REPLACEMENT SHEET

19 NOV 1996

Fig. 6 (3/3)

6

(208-231)

BSSL

(240-263)

ChesHum

(238-261)

Torpac

(223-246)

Drosceh

(226-249)

Ratlivce

(387-410)

Drosace

(2382-2402)

ThryHum

(150-173)

Dict.Di

Y	N	K	G	L	I	R	R	A	I	S	Q	S	G	V	A	L	S	P	W	V	I	Q	K
G	S	H	S	L	F	T	R	A	I	L	Q	S	G	S	F	N	A	P	W	A	V	T	S
G	S	R	D	L	F	R	R	A	I	L	Q	S	G	S	P	N	C	P	W	A	S	V	S
D	F	G	Q	L	A	R	A	A	F	S	F	S	G	N	A	L	D	P	W	V	I	Q	K
L	A	K	N	L	F	H	R	A	I	S	E	S	G	V	V	L	T	T	N	L	D	K	K
V	T	R	G	L	V	K	R	G	M	M	Q	S	G	T	M	N	A	P	W	S	H	M	T
T	N	S	Q	L	F	R	R	A	V	L	M	G	G	S	A	L	S	P	A	A	V	I	S
Y	S	R	Q	Y	F	N	A	A	I	S	S	S	S	P	L	T	V	G	L	K	D	K	T

Consensus

n k g L f r r a i . q s G s a l s p w a i q .

7

(285-297)

BSSL

(316-328)

ChesHum

(314-326)

Torpac

(298-310)

Drosceh

(298-306)

Ratlivce

(466-477)

Drosace

(2458-2468)

ThryHum

(224-236)

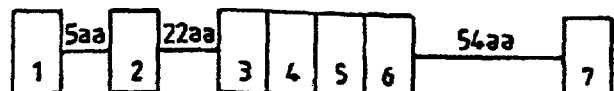
Dict.Di

V	G	F	V	P	V	I	D	G	D	F	I	P
V	N	F	G	P	T	V	D	G	D	F	L	T
F	S	F	V	P	V	I	D	G	E	F	F	P
A	P	F	S	P	V	L	E	P	S	D	A	P
.	.	.	.	T	V	I	D	G	V	V	L	P
P	S	.	A	P	T	I	D	G	A	F	L	P
.	.	W	G	P	V	I	D	G	H	F	L	R
T	I	W	S	P	V	I	D	G	D	A	F	I

Consensus

. . f . p v . d g d f . p

This Sheet Has Been Replaced



19 NOV 1996

- 15 -

10	30	50
AKLGAVYTEGGFVEGVNKKLGLLGDSVDIFKGIPFAAPTKALENPQPHPGWQGTLLKAKNF		
70	90	110
KKRCLQATITQDSTYGDEDCLYLNIWVPQGRKQVSRDLPVMIWIYGGAPLMGSGHGANFL		
130	150	170
NNYLYDGEEIATRGNVIVVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKRNIAA		
190	210	230
FGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSPWVIQKNPLFWAKKV		
250	270	290
AEKVGCPVGDAAARMAQCLKVTDPRALTLAYKVPLAGLEYPMLHYVGFVPVIDGDFIPADP		
310	330	350
INLYANAADIDYIAGTNNMDGHIFASIDMPAINKGKVKVTEEDFYKLVSEFTITKGLRGA		
370	390	410
KTTFDVYTESWAQDPSQENKKKTVDVDFETDVLFLVPTEIALAQHRANAKSAKTYAYLPSH		
430	450	470
PSRMPVYPKWVGADHADDIQYVFGKPFATPTGYRPQDRTVSKAMIAWNTNFAKTGDPNMG		
490	510	530
DSAG ^{for} PTHWEPYTTENSGYLEITKKMGSSSMKRSRLRTNFLRYWTLTYLALPTVTDQEATPV		
550	570	590
PPTGDSEATPVPPTGDSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD		
610	630	650
SGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAPP		
670	690	710
VPPTGDSGAPPVTPTGDSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPP		

RF

Figure VII

This Sheet Has Been Replaced



REPLACED SHEETS		SG
☒ Description	: paged 1 to paged 33	
☒ Claim(s)	: paged 34 to paged 37	
☒ Drawing(s)	: paged 1 to paged 15	
☒ Abstract	: paged 38 to paged	

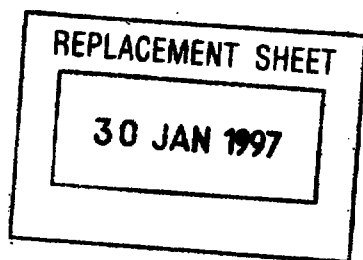
1

REPLACE ^{ment} SHEETS		SG
☒ Description	: paged 1 to paged 33	
☒ Claim(s)	: paged 34 to paged 37	
☒ Drawing(s)	: paged 1 to paged 15	
☒ Abstract	: paged 38 to paged	

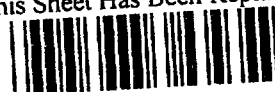
DERIVATIVES OF HUMAN BILE-SALT STIMULATED LIPASE,
AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM 30 JAN 1997

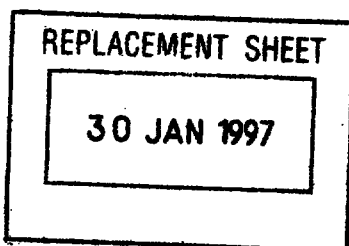
Field of the invention

The present invention relates to new DNA sequences, to new proteins coded for by such DNA sequences, and to the use as further described below of such proteins. The invention also encompasses vectors, such as plasmid constructs, comprising such DNA sequences, being capable of expressing the desired enzyme. The invention also includes host organisms transfected with such constructs e.g. bacteria yeast, mammalian cells, and transgenic animals. The invention also includes processes for the preparation of the novel products of the invention. The new proteins of the invention are related to an enzyme, known i.a. as human bile salt-stimulated lipase.



This Sheet Has Been Replaced





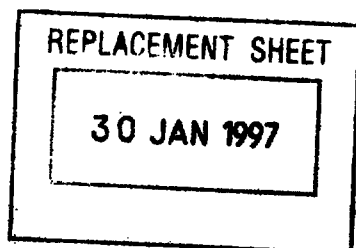
Background to the invention

The human lactating mammary gland synthesizes and secretes with the milk a bile salt-stimulated lipase (BSSL) [1] that, after specific activation by primary bile salts [2, 57, 58], contributes to the breast-fed infant's endogenous capacity of intestinal fat digestion [3-5]. This enzyme, which accounts for approximately 1% of total milk protein [6], is a non specific lipase; in vitro it hydrolyses not only tri-, di- and monoacylglycerols, but also cholesteryl-, and retinyl esters, and lysophosphatidylglycerols [7-10]. Furthermore, its activity is not restricted to emulsified substrates, but micellar and soluble substrates are hydrolyzed at similar rates [11].

BSSL is not degraded during passage with the milk through the stomach, and in duodenal contents it is protected by bile salts from inactivation by pancreatic proteases such as trypsin and chymotrypsin [2,11]. It is, however, inactivated when the milk is pasteurized, e.g. heated to 62.5 °C, 30 min [12]. Model experiments in vitro suggest that the end products of triacylglycerol digestion are different in the presence of BSSL [5,7]. Due to lower intraluminal bile salt concentrations during the neonatal period [13,14] this may be beneficial to product absorption [5,15].

The carboxylic ester hydrolase (CEH) of human pancreatic juice [16] seems functionally to be identical, or at least very similar, to BSSL [8]. They also share common epitopes [8,17], have identical N-terminal amino acid sequences [17] and are inhibited by inhibitors of serine esterases, e.g. eserine and diisopropylfluorophosphate [6,8,16]. It has been hypothesized that the two enzymes are products of the same gene [18,19]. The observed molecular size difference [8,19] could be explained by different patterns of glycosylation, as recently suggested [17].





This Sheet Has Been Replaced



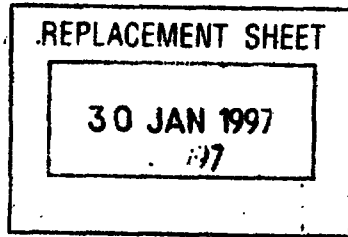
3

Dietary lipids is an important source of energy. The energy-rich triacylglycerols constitute more than 95% of these lipids. Some of the lipids, e.g. certain fatty acids and the fat soluble vitamins, are essential dietary constituents. Before gastro-intestinal absorption the triacylglycerols as well as the minor components, i.e. esterified fat-soluble vitamins and cholesterol, and diacylphosphatidylglycerols, require hydrolysis of the ester bonds to give rise to less hydrophobic, absorbable products. These reactions are catalyzed by a specific group of enzymes called lipases.

In the human adult the essential lipases involved are considered to be gastric lipase, pancreatic colipase-dependent lipase (tri- and diacylglycerol hydrolysis), pancreatic phospholipase A2 (diacylphosphatidylglycerols) and carboxylic ester hydrolase (cholesteryl- and fat soluble vitamin esters). In the breast-fed newborn bile salt-stimulated lipase plays an essential part in the hydrolysis of several of the above mentioned lipids. Together with bile salts the products of lipid digestion form mixed micelles from which absorption occurs (3-5).

Common causes of lipid malabsorption, and hence malnutrition, are reduced intraluminal levels of pancreatic colipase-dependent lipase and/or bile salts. Typical examples of such lipase deficiency are patients suffering from cystic fibrosis, a common genetic disorder resulting in a life-long deficiency in some 80% of the patients, and chronic pancreatitis, often due to chronic alcoholism.

The pancreatic and liver functions are not fully developed at birth, most notably in infants born before term. Fat malabsorption, for physiological reasons, is a common finding and thought to result from low intraluminal pancreatic colipase-dependent lipase and bile salt concentrations (3,4,13-15). However, because of BSSL, such



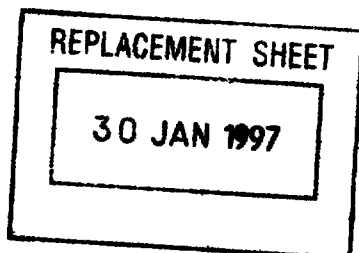
This Sheet Has Been Replaced



4

malabsorption is much less frequent in breast-fed infants than in infants fed pasteurized human milk of infant formulas (3-5, 12, 59, 60, 61). This is one reason why it has been advocated that newborn infants, particularly preterm infants, that cannot be fed their own mothers milk should be fed non-pasteurized milk from other mothers (12).

The present treatment of patients suffering from a deficiency of pancreatic lipase is the oral administration of very large doses of a crude preparation of porcine pancreatic enzymes. Colipase-dependent pancreatic lipase is inactivated by low pH. Such conditions are prevalent in the stomach, with the result that orally administered pancreatic lipase is virtually completely inactivated on the passage through the stomach to the gut. Therefore, this effect cannot be completely overcome by the use of large doses of enzyme. The large doses administered are inadequate for most patients, and the preparations are impure and unpalatable. Certain tablets have been formulated which pass through the acid regions of the stomach and discharge the enzyme only in the relatively alkaline environment of the jejunum. However, many patients suffering from pancreatic disorders have an abnormally acid jejunum and such tablets may fail to discharge the enzyme and may therefore be ineffective. Moreover, since the preparations presently on the market are of a non-human source there is a risk of immunoreactions that may cause harmful effects to the patients or result in reduced therapy efficiency. A further drawback with the present preparations is that their content of other lipolytic activities than colipase-dependent lipase are not stated. In fact, most of them contain very low levels of CEH/BSSL-activity. This may be one reason why many patients, suffering from cystic fibrosis in spite of supplementation therapy, suffer from deficiencies of fat soluble vitamins and essential fatty



acids.

Thus, there is a great need for products with properties and structure derived from human lipases and with a broad substrate specificity, which products may be orally administered to patients suffering from deficiency of one or several of the pancreatic lipolytic enzymes. The products of the present invention fulfil this need by themselves or in combination with other lipases or in combination with preparations containing other lipases. Furthermore, for some human infants there is an obvious need to improve fat utilization from conventional infant formulas, or pasteurized human milk from so-called milk banks.

BSSL has several unique properties that makes it ideally suited for substitution and supplementation therapy: It has been designed by nature for oral administration. Thus, it resists passage through the stomach and is activated in contents of the small intestine.

Its specific activation mechanism should prevent hazardous lipolysis of food or tissue lipids during storage and passage to its site of action.

Due to its broad substrate specificity it has the potential to, on its own, mediate complete digestion of most dietary lipids, including the fat soluble vitamin esters.

BSSL may be superior to pancreatic colipase-dependent lipase to hydrolyze ester bonds containing long-chain polyunsaturated fatty acids.

In the presence of gastric lipase and in the absence of, or at low levels of colipase-dependent lipase BSSL can ascertain a complete triacylglycerol digestion in vitro even if the bile salt levels are low such as in newborn infants. In the presence of BSSL the end products of triacylglycerol digestion become free fatty acids and free glycerol rather than free fatty acids and monoacylglycerol generated by the other two lipases (5). This may favour product absorption particularly when the intraluminal bile

REPLACEMENT SHEET

30 JAN 1997

6

salt levels are low (3,15).

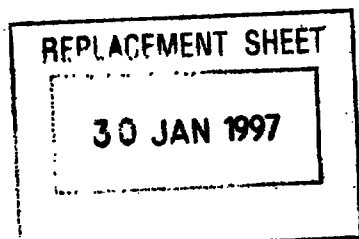
From a historical point of view infant formulas have been developed, and improved, from the concept that their composition should be as similar to that of human milk as possible. It is desirable to supplement such formulas.

The utilization for supplementation, substitution or therapy of bile salt-stimulated lipases (BSSL), or of proteins with the essential functions of BSSL, requires however access to quantities of the product on a large technical scale. It is not possible in factory scale to rely on natural sources such as milk as starting material. Besides the problem mentioned above with inactivation of BSSL during pasteurization, there is the additional risk of contamination of material from a natural source with infectious agents, e.g. vira such as HIV virus and CMV. There is, accordingly, a need for large scale access to products having BSSL properties. The present invention provides such products and methods for their preparation.

Prior art references are given later in this specification.

This Sheet Has Been Replaced





The invention

The present invention is based on the cloning of cDNA coding for BSSL derived from human mammary gland. We have also isolated, from human pancreas, a partial cDNA coding for CEH. Deduced amino acid sequences from the human cDNA's and comparison with CEH from other species, support the interpretation that BSSL and CEH are identical.

As will be further detailed below, it was surprisingly found that the structure of the protein as deduced from the cDNA sequence is quite different from the structure of other lipases. The structure proved unexpectedly to be more like the structure of typical esterases, such as cholinesterase.

This Sheet Has Been Replaced



30 JAN 1997



With reference to Figure II and Figure VII, products of the invention are:

- a) a protein as defined by the amino acid sequence 1-722 in Fig. VII,
- b) a protein as defined by the amino acid sequence 1-535 in Fig. VII,
- c) a protein as defined by the amino acid sequence 1-278 in Fig. VII,
- d) a protein as defined by the amino acid sequence 1-341 in Fig. VII,
- e) a protein as defined by the amino acid sequence 1-409 in Fig. VII,
- f) a protein as defined by the amino acid sequence 1-474 in Fig. VII,
- g) combinations of proteins defined under b) - f) e.g. as defined by the amino acid sequence in positions 1-278, 279-341, 279-409, 279-474, 342-409, 342-474 and 536-722.
- h) combinations of proteins defined under b) - g) in combination with one or more of the repeats according to Figure V,
- i) a protein as defined under a) - h) possessing an additional, N-terminal amino acid, namely methionine, and functionally equivalent variants and mutants of the proteins defined in a) - i) above;

It should be noted that the proteins under a, b, c, d, e, f, h and i above will not be identical in all respects to naturally occurring BSSL, but they will exhibit one or

30 JAN 1997

more of the critical functions of naturally occurring BSSL. Critical functions are given below.

j) a DNA sequence coding for the proteins defined in a, b, c, d, e, f, h and i above.

k) a DNA sequence according to Fig. II, defined by the following nucleotide numbers in Fig. II:

a) a DNA sequence 151-2316 according to Fig. II, coding for the protein defined by the amino acid sequence 1-722 in Fig. VII,

b) a DNA sequence 151-1755 according to Fig. II, coding for the protein defined by the amino acid sequence 1-535 in Fig. VII,

c) a DNA sequence 151-985 according to Fig. II, coding for the protein defined by the amino acid sequence 1-278 in Fig. VII,

d) a DNA sequence 151-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 1-341 in Fig. VII,

e) a DNA sequence 151-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 1-409 in Fig. VII,

f) a DNA sequence 151-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 1-474 in Fig. VII,

g) a DNA sequence 986-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 279-341 in Fig. VII,

h) a DNA sequence 986-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 279-409 in Fig. VII,

i) a DNA sequence 986-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 279-474 in Fig. VII,



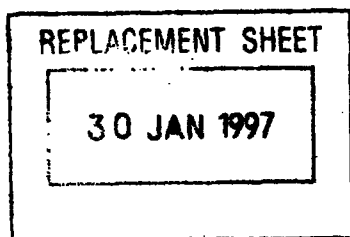
30 JAN 1997

- j) a DNA sequence 1173-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 342-409 in Fig. VII,
- k) a DNA sequence 1173-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 342-474 in Fig. VII.

Significant functions of proteins of the invention are

- a) suitable for oral administration,
- b) being activated by specific bile salts,
- c) acting as a non-specific lipase in the contents of the small intestines, that is being able to hydrolyze lipids relatively independent of their chemical structure and physical state (emulsified, micellar, soluble).
- d) Ability to hydrolyze triacylglycerols with fatty acids of different chain-length and different degree of unsaturation.
- e) Ability to hydrolyze also diacylglycerol, monoacylglycerol, cholesteryl esters, lysophosphatidylacylglycerol, and retinyl and other fat soluble vitamin-esters.
- f) Ability to hydrolyze not only the sn-1(3) ester bonds in a triacylglycerol but also the sn-2 ester bond.
- g) Ability to interact with not only primary but also secondary bile salts.
- h) Dependency on bile salts for optimal activity.
- i) Stability so that gastric contents will not affect the catalytical efficiency to any substantial degree.





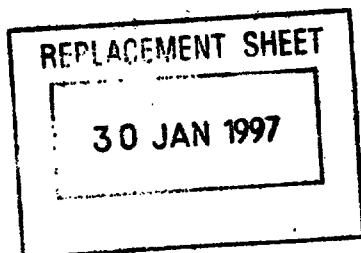
- j) Stability towards inactivation by pancreatic proteases, e.g. trypsin, provided bile salts are present.
- k) Ability to bind to heparin and heparin derivatives, e.g. heparan sulphate.
- l) Ability to bind to lipid-water interphases.
- m) Stability to permit lyophilization.
- n) Stability when mixed with food constituents such as in human milk, or milk formula.

The critical functions for supplementation, substitution, or therapy are these according to a), c), d), e), f), i), j) and l). For other purposes, not all critical functions may be necessary.

For expression of the proteins indicated above, the appropriate DNA sequence indicated above will be inserted into a suitable vector which then is introduced into a suitable host organism. The said vector will also have to comprise appropriate signal and other sequences enabling the organism to express the desired protein.

Suitable expression organisms:

With the recombinant DNA techniques it is possible to clone and express a protein of interest in a variety of prokaryotic and eukaryotic host organisms. Possible expression organisms are bacteria, simple eukaryots (yeast), animal cell cultures, insect cell cultures, plant cell cultures, plants and transgenic animals. Each individual system has its own particular advantages and disadvantages. The simple conclusion is that every gene to be expressed is a unique problem and no standard solution



12

This Sheet Has Been Replaced



is available.

Commonly used bacterial systems are *E. Coli*, *Bacillus subtilis*, *Streptomyces*. Commonly used yeasts are *saccharomyces*, and *Pichia Pastoris*. Commonly used animal cells are CHO cells and COS cells. Commonly used insect cell cultures are *Drosophila* derived cells.

Commonly used plant is the tobacco plant. Possible transgenic animals are goat and cow.

Possible bacterial vectors are exemplified by pUC and protein A-vectors.

Possible yeast vector is exemplified by pMA 91.

Possible insect vectors are devired from Baculo-virus.

Possible animal cell vectors are derived from SV/40.

Possible plant vectors are derived from the Ti-plasmid.

In every system, both natural and synthetic promoters and terminators can be used.

It is understood that depending on the choice of expression system, the expressed protein may contain an additional N-terminal amino acid (methionine), contain a few extra amino acids, or be fused to a heterologous protein, (e.g. protein A), or differ from the naturally occurring protein with respect to glycosylation.

Furthermore, the vectors may also contain signal sequences in order to export the protein to the periplasm or to the culture medium.

Thus, further aspects of the invention are:

- a) a vector comprising a DNA sequence coding for a protein as specified above,
- b) a host organism comprising a DNA sequence as specified above,
- c) a process for the production of a protein as specified above, by growing a host organism containing a vector as specified under a) above and isolating the protein.

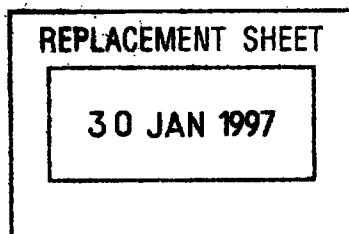


30 JAN 1997

Methods from purification are based on the expression system used (e.g. protein A/IgG) and/or on methods used for purification of the naturally occurring enzyme, as described in reference 6.

Additional aspects of the invention are:

- a pharmaceutical composition comprising a protein as specified above,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of cystic fibrosis,
- the use of a protein as specified above as a supplement to an infant food formulation,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of chronic pancreatitis,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption of any etiology,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons, e.g. in new-



14

born infants.

The DNA sequence in Fig. II from position 151 up to and including position 2316 is the sequence coding for the entire protein. The sequence from position 2317 up to and including position 2415 is not translated to protein, but is included in exon d identified in Table 2 below.

In one embodiment of the invention, the protein as defined in paragraphs a) - i) above is provided in isolated form and/or in substantially pure form.

The DNA sequences as defined in paragraphs a) - k) above are in one embodiment of the invention provided in isolated form and/or in substantially pure form.

Experimental part

Abbreviations

aa, amino acid; bp, base pair; BSSL, bile salt-stimulated lipase; c-AMP, cyclic adenosine monophosphate; CEH, carboxylic ester hydrolase; Da, dalton; c⁷GTP, 7-deaza-2-deoxyguanosine 5'triphosphate; EDTA, ethylene diamine tetraacetate; kb, kilobases; MOPS, 3-N-morpholino-propanesulfonic acid; nt, nucleotide; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; SSC, NaCl citrate, xGal, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside.

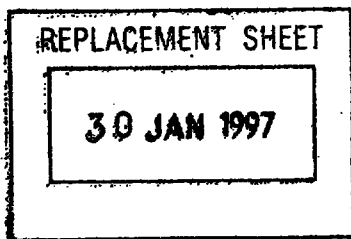
Enzymes

Bile salt-stimulated lipase EC 3.1.1.3

Carboxylic ester hydrolase EC 3.1.1.1

This Sheet Has Been Replaced





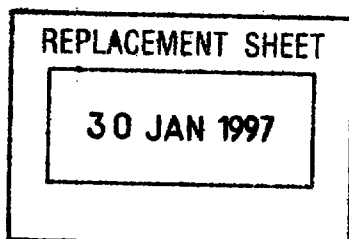
Material and Methods

A. Enzyme and antibody preparation

BSSL was purified from human milk as previously described [6]. When used for antibody production the enzyme was further purified by SDS-PAGE. The protein band corresponding to the lipase was, after staining with Coomassie Brilliant blue, electroeluted from the gel. Twentyfive μg of purified enzyme, together with an equal volume of Freund's complete adjuvant, was used for a first i.c. injection and the same amount of enzyme with incomplete adjuvant for the subsequent monthly booster injections. The rabbits were bled about two weeks after each booster and sera prepared and stored at $-20\text{ }^{\circ}\text{C}$.

B. Preparation of tryptic fragments and amino acid sequence analysis

Three mg of purified BSSL was dissolved in 1 ml of 0.1M Tris-Cl buffer, pH 8.5, containing 6M guanidinium hydrochloride and 2 mM EDTA. Dithioerythritol was added to 5mM. After incubation at $37\text{ }^{\circ}\text{C}$ for 2h, 300 μl 50 mM iodoacetate was added. After 90 min incubation at $25\text{ }^{\circ}\text{C}$ in darkness the reduced and carboxymethylated enzyme was desalted on a Sephadex G-25 column, equilibrated with 0.5M ammonium bicarbonate. Thirty μg of tosyl-L-phenylalanine chloromethane treated bovine trypsin (Worthington diagnostics system Inc., Freehold, NJ, USA) was added before lyophilization. The lyophilized protein was dissolved in 4 ml 0.1M ammonium bicarbonate and an additional 90 μg of trypsin was added. After 5h incubation at $37\text{ }^{\circ}\text{C}$ the protein was again lyophilized. The tryptic digest was dissolved in 0.1% trifluoroacetic acid (2mg/ml). Three hundred μg of trypsinated BSSL was chromatographed on HPLC using a C-18 reversed phase column and eluted with a gradient of 0-50% acetonitrile in 0.1%



trifluoroacetic acid. Peptide collection was monitored by continuous recording of the absorbance at 215 nm. Peptides to be sequenced were further purified by rechromatography using the same column with adjusted gradients.

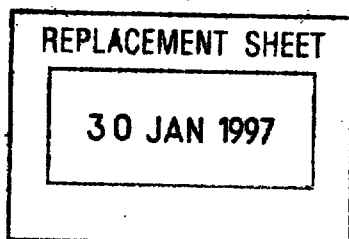
Samples of peptide fragments to be sequenced were dried under nitrogen to remove acetonitrile and applied to the sequencer. For N-terminal sequence analysis, native BSSL was dissolved in 0.1% acetic acid. Sequence analyzes were performed on an Applied Biosystems Inc. 477A pulsed liquid-phase sequencer and on-line PTH 120A analyzer with regular cycle programs and chemicals from the manufacturer. Initial and repetitive yields, calculated from a sequenced standard protein, β -lactoglobulin, were 47 and 97%, respectively.

C. Isolation of RNA

Samples of human pancreatic adipose and lactating mammary gland tissues were obtained at surgery and immediately put into guanidinium thiocyanate (1-5 g in 50 ml). Total RNA was extracted as described by Chirgwin [20]. Poly(A)-RNA was prepared by chromatography on oligo-deoxythymidilate-(oligo(DT))-cellulose column [21].

D. Construction and screening of cDNA libraries

Approximately 15 μ g poly-adenylated RNA from human pancreas was denatured with methyl mercuric hydroxide [22] and primed with oligo (dT) ₁₂₋₁₈ primers (Pharmacia, Uppsala, Sweden), and reversely transcribed using standard procedures [23]. Second-strand synthesis was carried out according to Gubler and Hoffman [24], except that DNA ligase and β -NAD were omitted, and the reaction temperature was set at 15 °C. Excess RNA was digested with RNase A (50 μ g/ml), and the double-stranded cDNA was treated with EcoRI methylase [25]. Ends were blunted with Klenow enzyme. After ligation to EcoRI linkers and



cleavage with EcoRI the cDNA was fractionated on a Sepharose 4B-C1 column. The void volume fraction was precipitated with ethanol and the cDNA ligated into the EcoRI site of a phosphatase treated gtl1 vector [26]. In vitro packing yielded more than 7×10^5 recombinants.

A cDNA library from human mammary gland, derived from tissue obtained from a woman at the eighth month of pregnancy, was purchased (Clontech Laboratories, Inc., Palo Alto, CA; USA).

Phages from the cDNA libraries were plated at 5×10^4 plaque forming units per 120-mm dish. The antiserum was diluted to a ratio of 1:3200 and screening was performed according to Young and Davis [27]. Alkaline-phosphatase-conjugated goat-anti-rabbit antibodies were used as second antibodies (Bio-Rad, Richmond, CA USA). To isolate clones corresponding to the 5'-end of the mRNA, nucleic acid hybridization was done under standard conditions [23] using a subcloned fragment from one of the immunopositive clones as a probe.

E. RNA analysis

Electrophoresis was carried out in a 1% agarose gel in 40mM MOPS buffer pH 7.0 after denaturation with glyoxal and dimethylsulfoxide [28]. Glyoxalated total RNA was then transferred to nitrocellulose filters [29]. The blots were probed with subclones of BSSL and CEH recombinants that were labelled by the oligo-labeling technique [30]. Prehybridization and hybridization were carried out with 50% formamid at 46 °C [23]. Posthybridization washes were performed at high stringency (0.1% SDS and 0.1xSSC at 60° C). (1xSSC, 0,15M NaCl, 0.0015M Na₃ citrate, pH 7.6).

F. Nucleotide sequence

This Sheet Has Been Replaced



30 JAN 1997

cDNA inserts from BSSL and CEH recombinants were either directly cloned into M3mp18 and mp19 after sonication and size fractionation or some of them were further subcloned into pTZ19R after digestions with PstI, BstXI, NarI, SmaI and AhaII. The nucleotide sequence was determined by the dideoxy chain-termination method [31]. The GC-rich repeats (see below) were also sequenced with TaqI polymerase and dc ⁷GTP. Both strands were sequenced. Sequence information was retrieved from autoradiograms by use of the software MS-EdSeq as described by Sjöberg et al [55].

G. Amino acid sequence predictions and homologies.

To predict the corresponding amino acid sequence of the cDNA inserts, codon usage of different reading frames was compared according to Staden and gave one open reading frame [32]. Homologies were searched for with the programs of the UWCGG software package [33].

Results and Discussion

A. Sequences of tryptic fragments and the N-terminus of BSSL

Trypsin digestion of purified BSSL resulted in approximately 50 fragments as judged by the number of peaks obtained during the HPLC-chromatography (Fig. I). The peaks were collected and the indicated peaks which could be isolated in a highly purified state and in reasonable quantities, were sequenced. The resulting sequences are shown in Table I. In addition the 30 most N-terminal residues were sequenced (Fig. II), and they confirm the previously reported sequence of Abouakil et al. (30 residues) [17].

The 22 residues long N-terminal sequence reported from Wang & Johnson [34] is a glycine in our report and a

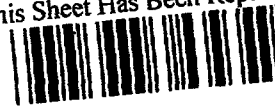


REPLACEMENT SHEET

30 JAN 1997

19

This Sheet Has Been Replaced



lysine in Wang & Johnson's.

B. Nucleotide sequence of BSSL

For construction of the λ gt11 cDNA library we used polyadenylated RNA from human pancreas. Initially four immunopositive clones were isolated, and then this pancreatic expression cDNA library was screened with antiserum against BSSL. Nucleotide sequence analysis of the four clones showed that they are in perfect agreement and correspond to the 3'-end of the mRNA. They all begin with a poly A tail and differ only in length; the longest insert, designated ACEH, spans 996 bp.

A cDNA library from human mammary gland was screened with antiserum, and the pancreas clone ACEH as probe. Positive clones were isolated from both screenings, which all originate from the 3'-end. The longest mammary gland clone, designated ABSSL, reaches 2100 bp upstream. It contains four of the sequenced tryptic fragments (Fig. II), but do not include the N-terminal amino acid sequence. To extend the sequence beyond the translation start, the mammary gland cDNA library was rescreened with a 118 bases long probe derived from the most 5'proximal part of ABSSL. One clone was isolated that continued a further 328 nucleotides upstream. It matched the N-terminal amino acid sequence, and contained the remaining tryptic fragment. As shown in Fig. II, the cDNA is 2428 nucleotides long and contains 81 bases upstream from the first ATG codon. The polyadenylation signal, AATAAA is located 13 nucleotides upstream from the poly A tail and the termination codon TAG was found at nucleotide 2317 followed by a 3'-untranslated region of 112 bp.

A GC rich region consisting of 16 repeats of 33 bases was found in the 3'-end of the sequence between base 1756 and 2283. The nucleotide sequence of the repetition, shown in Fig. III, consists of six identical repetitions surrounded by ten repetitions with different number of

30 JAN 1997



substitutions that have probably occurred after several duplications. The low number of substitutions suggests that these repetitions have appeared late during evolution.

C. Tissue distribution of expression

RNA from human lactating mammary gland, pancreas, adipose tissue and from a human hepatoma cell line (HepG2) was analyzed by Northern blotting. The size of the messenger was determined to be approximately 2.5 kb in both lactating mammary gland and pancreas. No signal could be detected in the lanes with RNA extracted from HepG2 or adipose tissue (Fig. IV).

Since the mRNA used for the mammary gland library was obtained from a female in her 8th month of pregnancy, it is evident that transcription and probably translation of the BSSL gene is turned on before partus, in agreement with previous findings on BSSL secretion before partus [35]. See Figure IV.

D. Amino acid sequence of BSSL

Assessed by SDS-PAGE the molecular mass has been reported to be 107-125 kDa [8,36] and by analytical ultracentrifugation to be 105 kDa [37]. The enzyme, as deduced from the cDNA, consists of 722 amino acid residues (Fig. II) which, giving a molecular mass of 76.271 Da, indicates that the enzyme contains at least 15-20% carbohydrate. The leader sequence is 23 residues long. A tentative active site serine residue is localized to serine-217 (Fig. V). The sequence around this serine accord with the consensus active site sequence of serine-hydrolases [38]. It has recently been it was proposed that basic residues found close to the active site serine may be involved in the cleavage of ester bonds in acylglycerols by lipases [39]. It is interesting to note that such

30 JAN 1997

This Sheet Has Been Replaced



residues are not present in BSSL. The single tentative N-glycosylation site is localized only seven residues from the serine. The degree of glycosylation [6,16] suggests that the enzyme contains O-linked carbohydrate. There are numerous sites where such glycosylation could have occurred. The amino acid composition based on purified enzyme has shown a high content of proline residues [6]. The amino acid sequence obtained from cDNA confirms this. Moreover, most of the proline residues are localized in the 16 repeats of 11 residues each, constituting the main part of the C-terminal half of the enzyme.

E. Comparison of the enzymes in mammary gland (BSSL) and pancreas (CEH)

BSSL of human milk and human pancreas CEH have previously been shown to be similar, if not identical. The present data strongly suggests that the two enzymes are products of the same gene. The nucleotide sequence of the cDNA clones shows that the pancreatic clone ACEH is identical with the mammary gland clone ABSSL from the poly A tail and 996 bases towards the 5'-end, including the sequence coding for the proline rich repeats. Northern blot gave a single band of 2.5 kb in RNA from pancreas and lactating mammary gland (Fig. IV). Genomic Southern blots further support the idea that only one gene codes for BSSL and CEH. The difference in mobility on SDS-PAGE between BSSL and CEH can be explained as a consequence of different glycosylation or differential splicing.

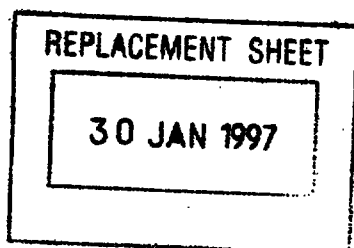
The similarity of BSSL to the rat and bovine enzymes (see below) and to results from genomics blots support the possibility that differential splicing cannot account for the mobility difference. Since the C-terminal sequence has not been confirmed on the protein level there is a less likely possibility that CEH may be processed by a proteolytic cleavage in the C-terminal end.

30 JAN 1997



So far as we know pancreatic enzymes that obviously correspond to CEH have often been named after species and the particular substrates used to determine their respective activities; lysophospholipase, cholesteryl esterase, sterol ester hydrolase, non-specific lipase, carboxyl ester lipase and cholesteryl ester hydrolase. Available data are compatible with the view that all these activities described originates in one and the same functional entity [42,43]. This illustrates the broad substrate specificity and the relevance of designating them as non-specific lipases. When the sequence of human BSSL/CEH is compared to the sequence of lysophospholipase from fat pancreas [40] and cholesterol esterase from bovine pancreas [41] extensive similarities are found that extend about 530 residues from the N-terminal (Fig V); but they differ in the part of the molecule where the repeats occur. The rat enzyme has only four repeats and the bovine three. Hence the human enzyme is a considerable longer peptide.

Moreover, striking similarities were found between BSSL and a number of typical esterases, e.g. acetyl choline esterases from several species, including man and Drosophila, and carboxyl esterases (Fig. VI). These similarities were restricted to the N-terminal 300 residues of BSSL which includes the tentative active site serine-residue. A similarity to acetyl choline esterase has been predicted from the fact that BSSL is inhibited by typical choline esterase inhibitors [6, 8, 16]. With the possible exception of the rat liver carboxyl esterase [45], none of these similar enzymes has been shown to have the same bile-salt dependency as BSSL; this suggests that the structural basis for this property resides in the C-terminal part of the protein. Moreover, BSSL can efficiently attack emulsified substrates which is not a known characteristic of the similar esterases. For this activity bile salt is a prerequisite.



The predicted sequence for human BSSL was compared with other well characterized mammalian lipases. Apart from the consensus sequence around the active site serine (G-X-S-X-G), no obvious similarities were found [44].

In addition to the similarities with other enzymes, there also significant similarities to one c-AMP dependent protein from Dictyostelium discoideum [46] as well as to thyroglobulin from several species (Fig VI) [47-49]. The similarities between BSSL and thyroglobulin, which comprise the active site region but not the active site itself, indicate that these highly conserved stretches of amino acids are of more generalized importance than merely supporting the enzymatic activity of esterases.

In conclusion, human milk BSSL consists of 722 amino acid residues. Available data strongly indicate that its peptide chain is identical to that of pancreatic CRH, and they are coded for by the same gene. The strongest evidence is that the nucleotide sequences of their 3' - ends and their N-terminal amino acid sequences are identical. The striking homologies found to rat pancreatic lysophospholipase and bovine pancreatic cholesterol esterase support the hypothesis that also these enzymes are functionally identical. However, as it has been suggested, the different molecular sizes found among species are not due to differences merely in glycosylation; instead they reflect a variable number of an eleven amino acid repeat. The similarity of the active site sequence between these esterases suggests that these proteins derive from a common ancestral gene.

With reference to Figures I-VII, the following legends are given.

This Sheet Has Been Replaced



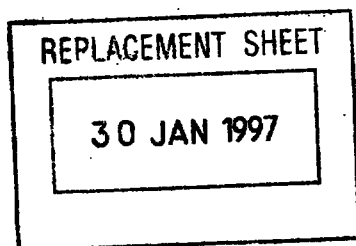


Figure I: Separation of the tryptic digest of BSSL on HPLC

Purified BSSL was treated with trypsin and chromatographed on HPLC as described in Materials and Methods. The indicated peaks were collected and purified further by a rechromatograph and their amino acid sequence determined.

Figure II: The cDNA nucleotide sequence and the deduced amino acid sequence for human bile salt-stimulated lipase:

The cDNA is 2428 bases long. The N-terminal 23-codon sequence (nt82-150) starting with an ATG, is interpreted as a leader peptide since the N-terminal amino acid sequence of the mature protein starts at codon 24 (nt 151, Ala). The leader peptide is underlined. The sign * indicates the starting point of an exon. The sign # indicates the starting point of the repetition part.

Figure III: The nucleotide sequence of the C-terminal GC-rich repetitions in the bile salt-stimulated lipase:

Substitutions are indicated by a *.

Figure IV: Northern blot hybridization

Northern blot analysis of total RNA isolated from human lactating mammary gland, pancreas, adipose tissue and a human hepatoma cell line (HepG2). Total TNA (10µg) from lactating mammary gland (lane A), pancreas (lane B), adipose tissue (lane C) and HepG2 (lane D) were electrophoresed in a 1% agarose gel in 40mM MOPS buffer at Ph 7.0 after denaturation of RNA in 1M glyoxal, 50% dimethylsulfoxide and 40mM MOPS. The glyoxalated RNA was then transferred to nitrocellulose paper for hybridization with [³² P] labeled BSSL cDNA (ABSSL).



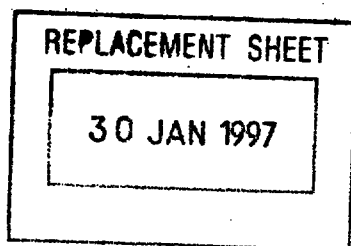


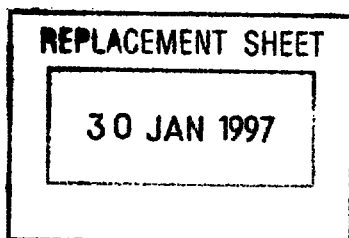
Figure V: Comparison of the deduced amino acid sequence from human milk BSSL, rat pancreatic lysophospholipase (Ratlpl) [40] and bovine pancreatic cholesterol esterase (Bovceh) [41]:

The serine residues involved in the active site are indicated by a *, and the # indicates the single possible N-glycosylation signal of the protein. The direct repeats of amino acid sequences are boxed. Matching sequences are denoted in capital letters, matching sequences between two enzymes are denoted in small letters and mismatching with a dot.

Figure VI: Comparison of the primary structure of BSSL to other esterases, thyroglobuline and to one c-AMP dependent enzyme from Dictyostelium discoideum:

BSSL: bile salt stimulated lipase from human, Cheshum: cholinesterase from fetal human tissue [50], Torpace: acetylcholinesterase from *Torpedo marmorata* [51], Drosceh: carboxylic ester hydrolase from *Drosophila melaogaster* [52], Ratlivce: carboxyl esterase from rat liver [53], Drosace: acetylcholinesterase from *Drosophila melaogaster* [54], Thyrrhum: thyroglobulin from human [49] and Dict.Di: c-AMP dependent enzyme from *Dictyostelium discoideum* [46]. There are 7 different domains that show similarities between the enzymes. Boxes enclose residues which are identical and small letters in the consensus sequence indicate identical residues in all the enzymes except for one. Dots indicate mismatches. The serine residue involved in the active site is indicated with *. The figure in the right hand corner shows how the domains are oriented.



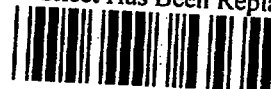


26

Figure VII:

gives the amino acid sequence 1 - 722 for the entire protein (one letter code) and indicates exons a, b, c, and d. The sign # indicates the starting point of the repetition part.

This Sheet Has Been Replaced



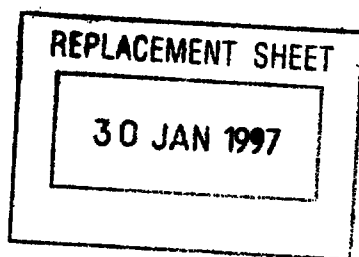


Table 1. Amino acid sequence of BSSL peptides.

Due to interfering peaks no positive identification of the residue in cycles 1 and 2 of the sequencing could be made in peptide number 26. The peptide numbers refer to the peaks in Fig I.

Tryptic fragments

TP16:LysValThrGluGluAspPheTyrLys

TP19:GlyIleProPheAlaAlaProThrLys

TP20:LeuValSerGluPheThrIleThrLys

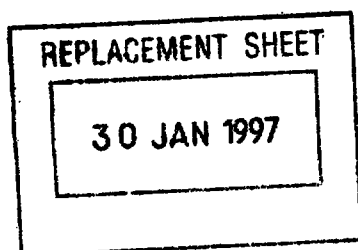
TP24:ThrTyrAlaTyrLeuPheSerHisProSerArg

TP26:PheAspValTyrThrGluSerTrpAlaGlnAsp

ProSerGlnGluAsnLys

This Sheet Has Been Replaced





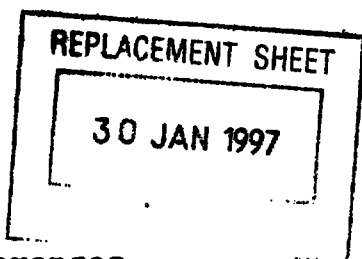
28

Table 2: Identification of the exons a, b, c and d
numbered as in Figures II and VII.

exon	<u>Location</u>	
	between nucleotide number	between aminoacid number
a	986-1172	279-341
b	1173-1376	342-409
c	1377-1574	410-474
d	1575-2415	475-722
the entire protein	151-2316	1-722
the entire protein excluding repetitions	151-1755	1-535

This Sheet Has Been Replaced

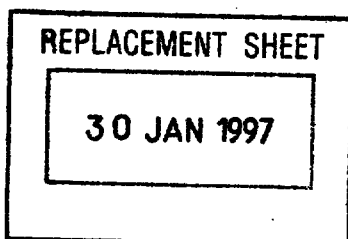




References

1. Bläckberg, L., Ångquist, K.A., & Hernell, O. (1987) FEBS Lett. 217, 37-41.
2. Hernell, O. (1975) Eur. J. Clin. Invest. 5, 267-272.
3. Hernell, O., Bläckberg, L. & Bernbäck, S. (1988) In Perinatal nutrition (Lindblad, B.S., ed.) Bristol-Myers Nutrition symposia vol. 6, pp. 259-272, Academic Press, New York.
4. Hernell, O., L. Bläckberg, B. Fredrikzon, and T. Olivecrona. 1981. Bile salt-stimulated lipase in human milk and lipid digestion in the neonatal period. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 465-471.
5. Bernbäck, S., Bläckberg, L. & Hernell, O. (1990) J. Clin. Invest. 221-226 (1990)
6. Bläckberg, L. & Hernell, O. (1981) Eur. J. Biochem 116, 221-225.
7. Hernell, O. & Bläckberg, L. (1982) Pediatr. Res. 16, 882-885.
8. Bläckberg, L. Lombardo, D., Hernell, O., Guy, O. & Olivecrona, T. (1981) FEBS Lett. 136, 284-288.
9. Fredrikzon, B., Hernell, O., Bläckberg, L. & Olivecrona, T. (1978) Pediatr. Res. 12, 1048-1052.
10. Wang, C.-S., Hartsuck, J.A. & Downs, D. (1988) Biochemistry 27, 4834-4840.
11. Bläckberg, L. & Hernell, O. (1983) FEBS Lett. 157, 337-341.
12. Björkstén, B., Burman, L.G., deChateau, P., Fredrikzon, B., Gotheffors, L. & Hernell, O. (1980) Br. Med. J. 201, 267-272.
13. Brueton, M.J., Berger, H.M., Brown, G.A., Ablitt, L., Iyangkaran, N. & Wharton B.A. (1978) Gut 19, 95-98.
14. Murphy, G.M. & Signer, E. (1975) Gut 15, 151-163.
15. Hernell, O., Staggars, J.E. & Carey, M.C. Biochemistry. (1990), 29:2041-2056.



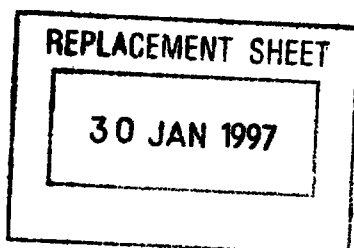


This Sheet Has Been Replaced

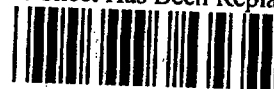


30

16. Lombardo, D., Guy, O. & Figarella, C. (1978) *Biochim. Biophys. Acta* 527, 142-149.
17. Abouakil, N., Rogalska, E., Bonicel, J. & Lombardo, D. (1988) *Biochim. Biophys. Acta* 961, 299-308.
18. Hernell, O., Bläckberg, L. & Lindberg, T. (1988) in *Textbook of Gastroenterology and Nutrition in infancy* (Lebenthal, E., ed) pp. 209-217, Raven Press, New York.
19. Wang, C.-S. (1988) *Biochem. Biophys. Res. Com.* 155, 950-955.
20. Chirgwin, J.M. Przybyla, A.E., MacDonald R.J. & Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
21. Aviv, H. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 69, 5201-5205.
22. Bailey, J.M. & Davidson, N. (1976) *Anal. Biochem.* 70, 75-85.
23. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982): *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
24. Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263-269.
25. Maniatis T., Haldison, R.C., Lacy, E., Lauer, J., O'Connell, C. & Qvon, D. (1978) *Cell* 15, 687-701.
26. Young, R.A. & Davis, R.W. (1983) *Science* 222, 778-782.
27. Young, R.A. & Davis, R.W. (1983) *Proc. Natl. Acad. Sci, USA* 80, 1194-1198.
28. McMaster, G.K. & Charnichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 48 353-48 358.
29. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
30. Feinberg, A.P. & Vogelstein, B. (1983) *Analyt. Biochem.* 132, 6-13.
31. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74. 5463-5467.
32. Staden, R. (1984) *Nucl, Acids Res.* 12, 551-567.
33. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.



This Sheet Has Been Replaced



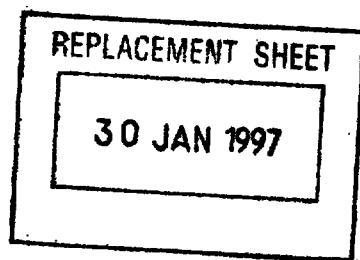
34. Wang, C.-S. & Johnson, K. (1983) *Anal. Biochem.* 133, 457-461.
35. Hamosh, M. (1986) in *Human Milk Infant Nutrition and Health* (Howell, R.R. Morriss, F.H. & Pickering, L.K. eds) pp. 66-97, Charles, C. Thomas, Springfield.
36. Wang, C.-S. (1980) *Anal. Biochem.* 105, 398-402.
37. Wang, C.-S & Lee, D.M. (1985) *J. Lipid. Res.* 26, 824-830.
38. Brenner, S. (1988) *Nature* 334, 528-530.
39. Yang, C.-Y., Gu, Z.-W., Yang, H.-H., Rohde, M.F., Gotto, Jr., A.M. & Pownall, H.J. (1989) *J. Biol. Chem.* 265, 16822-16827.
40. Han, J.H., Stratowa, C., & Rutter, W.J. (1987) *Biochemistry* 26, 1617-1625.
41. Kyger, E., Wiegand, R., & Lange, L. (1989) *Biochim. Biophys. Res. Com* 164, 1302-1309.
42. Bläckberg, L. (1981) *Fat digestion in newborn infant. Umeå University Medical Dissertations New Series No 71.*
43. Rudd, E.A. & Brockman, H.L. (1984) in *Lipases* (Borgström, B & Brockman, H.L. eds) pp. 184-204, Elsevier, Amsterdam.
44. Mickel, S., Weidenbach, F., Swarovsky, B., LaForge, S. & Scheele, G. (1989) *J. Biol. Chem.* 264, 12895-12901.
45. Cammulli, E.D., Linke, M.J., Brockman, H.L. & Hui, D.Y. (1989) *Biochim. Biophys. Acta* 1005, 177-182.
46. Mann, S.K.O. & Firtel, R.A. (1987) *Mol. Cell Biol.* 7, 458-469.
47. Mercken, L., Simons, M.J., Swillens, S., Masser, M. & Vassart, G. (1985) *Nature* 316, 647-651.
48. Lauro, R., Obici, S. Condliffe, D., Ursini, V.M., Musti, A., Moscatelli, C. & Avvedimento, V.E. (1985) *Eur. J. Biochem.* 148, 7-11.
49. Malthiery, Y. & Lissitzky, S. (1987) *Eur. J. Biochem.* 165, 491-498.
50. Prody, C., Zevin-Sonkin, D., Gnatt, A., Goldberg, O.,

30 JAN 1997

- & Soreq, H. (1987) Proc. Natl. Acad. Sci. USA 84, 3555-3559.
51. Sikorav, J-L., Krejci, E. & Massoulie', J. (1987) EMBO J. 6, 1865-1873.
52. Oakeshott, J.G. Collet, C., Phillis, R.W., Nielsen, K.M., Russel, R.J., Cambers, G.K., Ross, V. & Richmond, R.C. (1987) Proc. Natl. Acad. Sci, USA 84, 3359-3363.
53. Long, R., Satho, H., Martin, B., Kimura, S., Gonzalez, F. & Pohl, L. (1988) Biochem. Biophys. Res. Comm. 156, 866-873.
54. Hall, L.M.C. & Spierer, P. (1986) EMBO J. 5, 2949-2954.
55. Sjöberg, S., Carlsson, P., Enerbäck, S., & Bjursell, G. (1989) CABIOS 5, 41-46.
56. EP-A-317355.
57. Hernell, O., and T. Olivecrona. 1974. Human milk lipases II. Bile salt-stimulated lipase. Biochim. Biophys. Acta 369:234-244.
58. Hernell, O., L. Bläckberg, and T. Olivecrona. 1981. Human milk lipases. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 347-354.
59. Atkinson, S.A., M.H. Bryan, and G.H. Andersson. 1981. Human milk feeding in premature infants: protein, fat and carbohydrate balances in the first two weeks of life. J.Pediatr. 99:617-624.
60. Chappell, J.E., M.T. Clandinin, C. Kearney-Volpe, B. Reichman and P.W. Swyer. 1986. Fatty acid balance studies in premature infants fed human milk or formula: effect of calcium supplementation. J. Pediatr. 108:438-447.

This Sheet Has Been Replaced



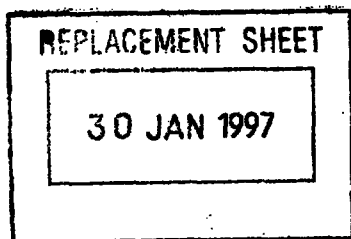


33

61. Williamson, S., E. Finucane, H. Ellis, and H.R. Gamsu. 1978. Effect of heat treatment of human milk on absorption of nitrogen, fat, sodium, calcium and phosphorus by preterm infants. Arch. Dis. Childhood 53:555-563.

This Sheet Has Been Replaced





This Sheet Has Been Replaced

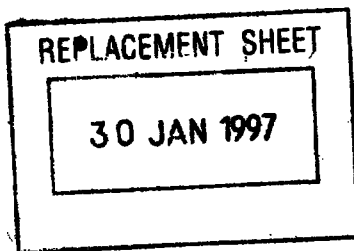


34



What we claim is:

1. A protein as indicated in Figure VII from position 1 to position 722.
2. A protein as indicated in Figure VII from position 1 to position 535.
3. A protein as indicated in Figure VII from position 1 to position 278.
4. A protein as indicated in Figure VII from position 1 to position 341.
5. A protein as indicated in Figure VII from position 1 to position 409.
6. A protein as indicated in Figure VII from position 1 to position 474.
7. A protein as indicated in Figure VII from position 1 to position 278, position 279 to position 341, position 279 to position 409, position 279 to position 474, position 342 to position 409, position 342 to position 474, or position 536 to position 722.
8. A protein according to any of claims 1-7 in combination with one or more of the repeats according to Figure V.
9. A protein according to any of claims 1-8, possessing methionine as additional N-terminal amino acid.
10. A protein according to any of claims 1-9, exhibiting one or several of the critical functions of naturally occurring bile salt-stimulated lipase.



35

11. A functionally equivalent variant or mutant of a protein according to any of claims 1-9.
12. A DNA sequence coding for a protein with the amino acid sequence according to claim 1.
13. A DNA sequence coding for a protein with the amino acid sequence according to claim 2.
14. A DNA sequence coding for a protein with the amino acid sequence according to any of claims 3, 4, 5, 6, 7, 8, 9, 10 and 11.
15. A DNA sequence defined by the following nucleotide numbers in Figure II:

151-2316
151-1755
151-985
151-1172
151-1376
151-1574
986-1172
986-1376
986-1574
1173-1376
1173-1574
16. A vector comprising a DNA sequence coding for a protein according to claims 1 - 11
17. A vector comprising a DNA sequence according to claims 12 - 15.
18. A host organism transformed with a vector according to claims 16 or 17.

This Sheet Has Been Replaced



REPLACEMENT SHEET

30 JAN 1997

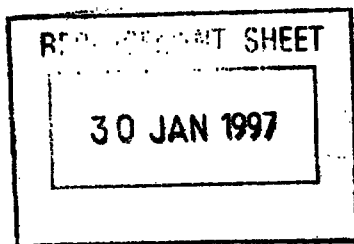
36



19. A process for the production of a protein according to claims 1 - 11 by growing a host organism containing a vector according to claims 16 or 17, and isolating the protein.
20. A pharmaceutical composition comprising a protein according to any of claims 1 - 11.
21. A pharmaceutical composition comprising a protein according to any of claims 1-11 in combination with a lipase or in combination with preparations containing lipases.
22. The use of a protein according to any of claims 1 - 11, for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency.
23. The use of a protein according to any of claims 1 - 11 for the manufacture of a medicament for the treatment of cystic fibrosis.
24. The use of a protein according to any of claims 1 - 11 as a supplement to an infant food formulation.
25. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of chronic pancreatitis.
26. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption.
27. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins.

This Sheet Has Been Replaced





37



28. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons.
29. The use, according to claims 22-28 of a protein according to any of claims 1-11 in combination with a lipase or lipases or in combination with preparations containing a lipase or lipases.
30. An infant food formulation, supplemented with a protein according to any of claims 1 - 11.
31. A protein according to claims 1 - 11 in substantially pure form.
32. A protein according to claims 1 - 11 in isolated form.
33. A DNA sequence according to claims 12 - 15 in substantially pure form.
34. A DNA sequence according to claims 12 - 15 in isolated form.
35. A process for the preparation of a pharmaceutical composition according to claims 20 or 21 by incorporating a protein according to claims 1 - 11, 31, or 32 in a pharmacologically acceptable carrier.
36. A process according to claim 35 for the preparation of a pharmaceutical composition for oral administration.
37. A process for the preparation of an infant food formulation according to claim 30, by supplementing an infant food formulation with a protein according to claims 1 -11, 31, or 32.

This Sheet Has Been Replaced



30 JAN 1997

DERIVATIVES OF HUMAN BILE-SALT STIMULATED LIPASE,
AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEMAbstract

The present invention relates to new DNA sequences, to new proteins coded for by such DNA sequences, and to the use as further described below of such proteins. The invention also encompasses vectors, such as plasmid constructs, comprising such DNA sequences, being capable of expressing the desired enzyme. The invention also includes host organisms transfected with such constructs e.g. bacteria yeast, mammalian cells, and transgenic animals. The invention also includes processes for the preparation of the novel products of the invention. The new proteins of the invention are related to an enzyme, known i.a. as human bile salt-stimulated lipase.

FIGURE 2.

This Sheet Has Been Replaced



REPLACEMENT SHEET

30 JAN 1997

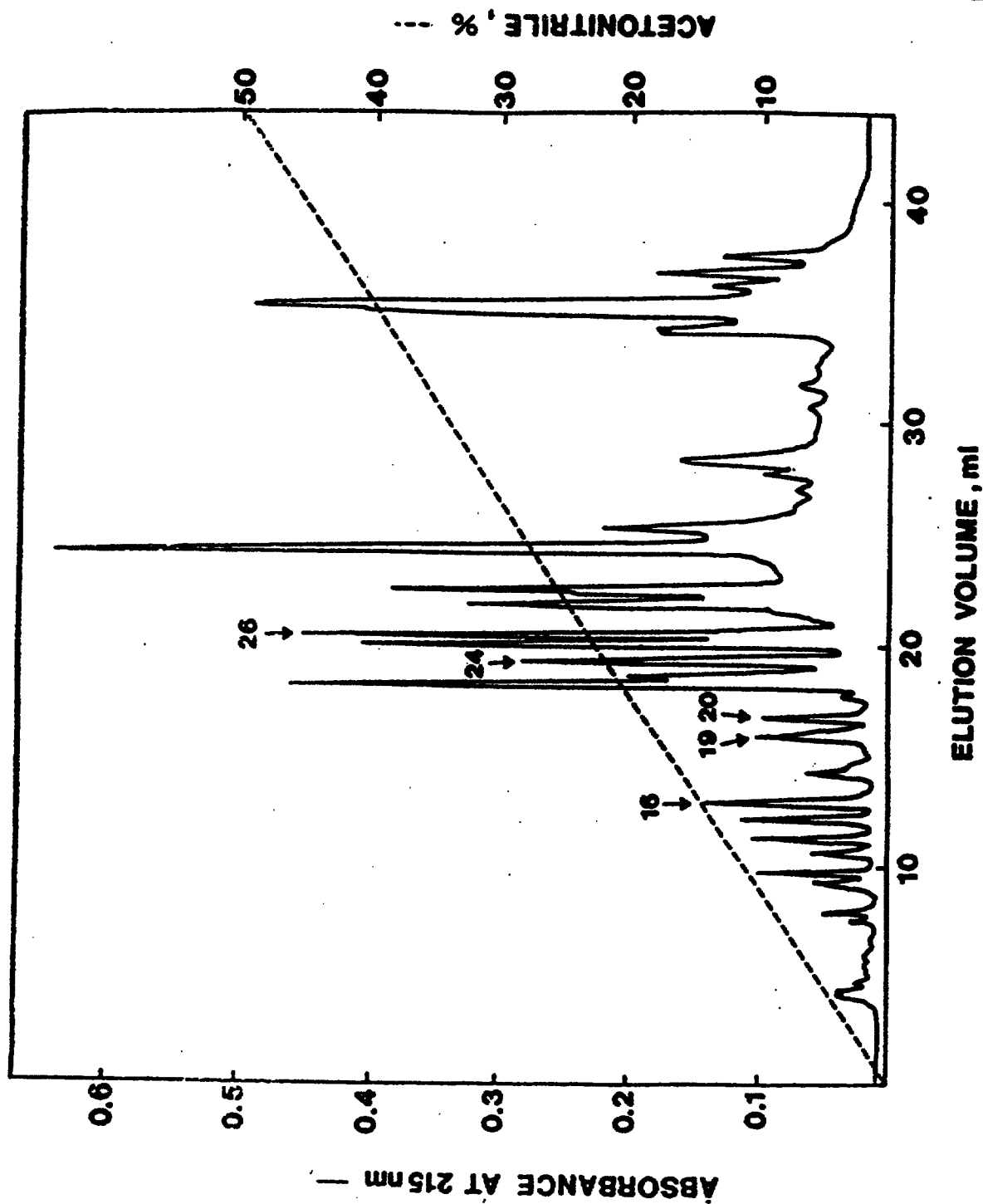
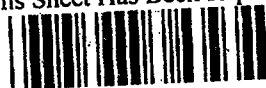


FIGURE 1

This Sheet Has Been Replaced



30 JAN 1997

- 2 -

10	30	50
ACCTTCTGTATCAGTTAAGTGTC AAGATGGAAGGAACAGCAGTCTCAAGATAATGCAAAG		
70	90	110
AGTTTATTCATCCAGAGGCTGATGCTCACCATGGGGCGCCTGCAACTGGTTGTGTTGGGC		
	<u>MetLeuThrMetGlyArgLeuGlnLeuValValLeuGly</u>	
130	150	170
CTCACCTGCTGCTGGGCAGTGGCGAGTGCCGCGAAGCTGGGCGCCGTGTACACAGAAGGT		
<u>LeuThrCysCysTrpAlaValAlaSerAlaAlaLysLeuGlyAlaValTyrThrGluGly</u>		
190	210	230
GGGTTCGTGGAAGGCGTCAATAAGAAGCTCGGCCTCCTGGGTGACTCTGTGGACATCTTC		
GlyPheValGluGlyValAsnLysLysLeuGlyLeuLeuGlyAspSerValAspIlePhe		
250	270	290
AAGGGCATCCCCTTCGCAGCTCCCACCAAGGCCCTGGAAAATCCTCAGCCACATCCTGGC		
LysGlyIleProPheAlaAlaProThrLysAlaLeuGluAsnProGlnProHisProGly		
310	330	350
TGGCAAGGGACCCTGAAGGCCAAGAACTTCAAGAAGAGATGCCTGCAGGCCACCATCACC		
TrpGlnGlyThrLeuLysAlaLysAsnPheLysLysArgCysLeuGlnAlaThrIleThr		
370	390	410
CAGGACAGCACCTACGGGGATGAAGACTGCCTGTACCTCAACATTGGGTGCCCCAGGGC		
GlnAspSerThrTyrGlyAspGluAspCysLeuTyrLeuAsnIleTrpValProGlnGly		
430	450	470
AGGAAGCAAGTCTCCCGGGACCTGCCC GTTATGATCTGGATCTATGGAGGCGCCTTCCTC		
ArgLysGlnValSerArgAspLeuProValMetIleTrpIleTyrGlyGlyAlaPheLeu		
490	510	530
ATGGGGTCCGGCCATGGGGCCAACCTCCTCAACAACCTACCTGTATGACGGCGAGGAGATC		
MetGlySerGlyHisGlyAlaAsnPheLeuAsnAsnTyrLeuTyrAspGlyGluGluIle		

FIGURE 2 - 1/5

This Sheet Has Been Replaced



30 JAN 1997

- 3 -

550 570 590
 GCCACACGCGGAAACGTCATCGTGGTCACCTTCAACTACCGTGTCGGCCCCCTTGGGTTC
 AlaThrArgGlyAsnValIleValValThrPheAsnTyrArgValGlyProLeuGlyPhe
 610 630 650
 CTCAGCACTGGGGACGCCAATCTGCCAGGTAACCTATGGCCTTCGGGATCAGCACATGGCC
 LeuSerThrGlyAspAlaAsnLeuProGlyAsnTyrGlyLeuArgAspGlnHisMetAla
 670 690 710
 ATTGCTTGGGTGAAGAGGAATATCGCGGCCTTCGGGGGGGACCCCAACAACATCACGCTC
 IleAlaTrpValLysArgAsnIleAlaAlaPheGlyGlyAspProAsnAsnIleThrLeu
 730 750 770
 TTCGGGGAGTCTGCTGGAGGTGCCAGCGTCTCTCTGCAGACCCTCTCCCCCTACAACAAG
 PheGlyGluSerAlaGlyGlyAlaSerValSerLeuGlnThrLeuSerProTyrAsnLys
 790 810 830
 GGCCTCATCCGGCGAGCCATCAGCCAGAGCGGCGTGGCCCTGAGTCCCTGGGTTCATCCAG
 GlyLeuIleArgArgAlaIleSerGlnSerGlyValAlaLeuSerProTrpValIleGln
 850 870 890
 AAAAACCCTCTTCTGGGCCAAAAGGTGGCTGAGAAGGTGGGTGGCCCTGTGGGTGAT
 LysAsnProLeuPheTrpAlaLysLysValAlaGluLysValGlyCysProValGlyAsp
 910 930 950
 GCCGCCAGGATGGCCCAGTGTCTGAAGGTTACTGATCCCCGAGCCCTGACGCTGGCCTAT
 AlaAlaArgMetAlaGlnCysLeuLysValThrAspProArgAlaLeuThrLeuAlaTyr
 970 990 1010
 AAGGTGCCGCTGGCAGGCCTGGAGTACCCCATGCTGCACTATGTGGGCTTCGTCCCTGTC
 LysValProLeuAlaGlyLeuGluTyrProMetLeuHisTyrValGlyPheValProVal
 1030 1050 1070
 ATTGATGGAGACTTCATCCCCGCTGACCCGATCAACCTGTACGCCAACGCCGCCGACATC
 IleAspGlyAspPheIleProAlaAspProIleAsnLeuTyrAlaAsnAlaAlaAspIle

FIGURE 2 - 2/5

This Sheet Has Been Replaced



30 JAN 1997

- 4 -

1090	1110	1130
GACTATATAGCAGGCACCAACAACATGGACGGCCACATCTTCGCCAGCATCGACATGCCT		
AspTyrIleAlaGlyThrAsnAsnMetAspGlyHisIlePheAlaSerIleAspMetPro		
1150	1170	1190
GCCATCAACAAGGGCAACAAGAAAGTCACGGAGGAGGACTTCTACAAGCTGGTCAGTGAG		
AlaIleAsnLysGlyAsnLysLysValThrGluGluAspPheTyrLysLeuValSerGlu		
1210	1230	1250
TTCACAATCACCAAGGGGCTCAGAGGCGCCAAGACGACCTTTGATGTCTACACCGAGTCC		
PheThrIleThrLysGlyLeuArgGlyAlaLysThrThrPheAspValTyrThrGluSer		
1270	1290	1310
TGGGCCCAGGACCCATCCCAGGAGAATAAGAAGAAGACTGTGGTGGACTTTGAGACCGAT		
TrpAlaGlnAspProSerGlnGluAsnLysLysLysThrValValAspPheGluThrAsp		
1330	1350	1370
GTCCTCTTCTGGTGCCACCGAGATTGCCCTAGCCCAGCACAGAGCCAATGCCAAGAGT		
ValLeuPheLeuValProThrGluIleAlaLeuAlaGlnHisArgAlaAsnAlaLysSer		
1390	1410	1430
GCCAAGACCTACGCCTACCTGTTTTCCCATCCCTCTCGGATGCCCCGTCTACCCCAAATGG		
AlaLysThrTyrAlaTyrLeuPheSerHisProSerArgMetProValTyrProLysTrp		
1450	1470	1490
GTGGGGGCGGACCATGCAGATGACATTACGTATTCGGGAAGCCCTTCGCCACCCCC		
ValGlyAlaAspHisAlaAspAspIleGlnTyrValPheGlyLysProPheAlaThrPro		
1510	1530	1550
ACGGGCTACCGGCCCCAAGACAGGACAGTCTCTAAGGCCATGATCGCCTACTGGACCAAC		
ThrGlyTyrArgProGlnAspArgThrValSerLysAlaMetIleAlaTyrTrpThrAsn		
1570	1590	1610
TTTGCCAAAACAGGGGACCCCAACATGGGCGACTCGGCTGTGCCACACACTGGGAACCC		
PheAlaLysThrGlyAspProAsnMetGlyAspSerAlaValProThrHisTrpGluPro		

FIGURE 2 - 3/5

This Sheet Has Been Replaced



30 JAN 1997

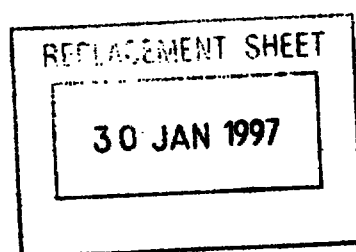
- 5 -

1630 1650 1670
TACACTACGGAAAACAGCGGCTACCTGGAGATCACCAAGAAGATGGGCAGCAGCTCCATG
TyrThrThrGluAsnSerGlyTyrLeuGluIleThrLysLysMetGlySerSerSerMet
1690 1710 1730
AAGCGGAGCCTGAGAACCAACTTCCTGCGCTACTGGACCCTCACCTATCTGGCGCTGCCC
LysArgSerLeuArgThrAsnPheLeuArgTyrTrpThrLeuThrTyrLeuAlaLeuPro
1750 1770 1790
ACAGTGACCGACCAGGAGGCCACCCCTGTGCCCCCACAGGGGACTCCGAGGCCACTCCC
ThrValThrAspGlnGluAlaThrProValProProThrGlyAspSerGluAlaThrPro
1810 1830 1850
GTGCCCCCACGGGTGACTCCGAGACCGCCCCCGTGCCGCCACGGGTGACTCCGGGGCC
ValProProThrGlyAspSerGluThrAlaProValProProThrGlyAspSerGlyAla
1870 1890 1910
CCCCCGTGCCGCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC
ProProValProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSer
1930 1950 1970
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCACGGGT
GlyAlaProProValProProThrGlyAspSerGlyAlaProProValProProThrGly
1990 2010 2030
GACTCCGGGGCCCCCCCCCGTGCCGCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCC
AspSerGlyAlaProProValProProThrGlyAspSerGlyAlaProProValProPro
2050 2070 2090
ACGGGTGACTCCGGCGCCCCCCCCCGTGCCGCCACGGGTGACGCCGGGCCCCCCCCCGTG
ThrGlyAspSerGlyAlaProProValProProThrGlyAspAlaGlyProProProVal
2110 2130 2150
CCGCCCACGGGTGACTCCGGCGCCCCCCCCCGTGCCGCCACGGGTGACTCCGGGGCCCCC
ProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSerGlyAlaPro

FIGURE 2 - 4/5

This Sheet Has Been Replaced





- 6 -

2170	2190	2210
CCCGTGACCCCCACGGGTGACTCCGAGACCGCCCCCGTGCCGCCCACGGGTGACTCCGGG		
ProValThrProThrGlyAspSerGluThrAlaProValProProThrGlyAspSerGly		
2230	2250	2270
GCCCCCCTGTGCCCCCAGGGGTGACTCTGAGGCTGCCCTGTGCCCCCAGATGAC		
AlaProProValProProThrGlyAspSerGluAlaAlaProValProProThrAspAsp		
2290	2310	2330
TCCAAGGAAGCTCAGATGCCTGCAGTCATTAGGTTTTAGCGTCCCATGAGCCTTGGTATC		
SerLysGluAlaGlnMetProAlaValIleArgPhe		
2350	2370	2390
AAGAGGCCACAAGAGTGGGACCCCAGGGGCTCCCCTCCCATCTTGAGCTCTTCCTGAATA		
2410		
AAGCCTCATACCCCTAAAAAAAAAAAAA		

FIGURE 2 - 5/5

This Sheet Has Been Replaced



30 JAN 1997

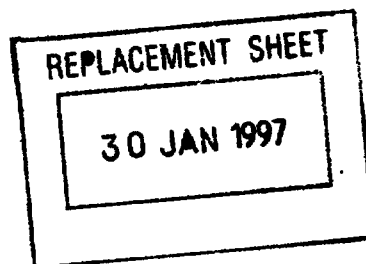
- 7 -

		Number of substitutions
GAGGCCACCCCTGTGCCCCCAGCGGACTCC	1	0
GAGGCCACTCCCGTCCGCCCCACGGGTACTCC	2	4
GAGACCGCCCCCGTCCCCCCACCGGTACTCC	3	3
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	4	0
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	5	0
GGGGCCCCCCCCGTCCGCCCCACGGGTACTCC	6	Identical to consensus 0
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	7	0
GGGGCCCCCCCCGTCCGCCCCACGGGTACTCC	8	0
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	9	0
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	10	2
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	11	1
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	12	1
GGGGCCCCCCCCGTACCCCCACGGGTACTCC	13	2
GAGACCGCCCCCGTCCCCCCACGGGTACTCC	14	3
GGGGCCCCCGTGTCCCCCCACGGGTACTCC	15	3
GAGGCTGCCCCGTGTCCCCCCACATGACTCC	16	7
GGGGCCCCCCCCGTCCCCCCACGGGTACTCC	Consensus	

FIGURE 3

This Sheet Has Been Replaced





A B C D Kb



FIGURE 4

This Sheet Has Been Replaced



This Sheet Has Been Replaced



Fig. 5-1/3

REPLACEMENT SHEET

30 JAN 1997

50
Bssl MLTMGRLQLVVLGLTCCWAVASAAKLGAVYTEGGFVEGVNKKLGLIG.DS
Ratlpl ...MGRLEVLFLGLTCCLAAACAAGALYTEGGFVEGVNKKLSLIGGDS
BovcehMLGASRLGSPSPGCLAVASAAKLGSVYTEGGFVEGVNKKLSLFG.DS
Consensusgrl....lgltcclaaaaAAKLGavYTEGGFVEGVNKKLSLlG.DS

100
Bssl VDIFKGIPFAAPT KALENPQHPGWQGT LKAKNFKKRCLQATITQDSTYG
Ratlpl VDIFKGIPFA.TAKTLENPQRHPGWQGT LKATQFKKRCLQATLTQDDTYG
Bovech VDIFKGIPFAAAPKALEKPKRHPGWQGT LKAKSFKKRCLQATLTQDSTYG
Consensus VDIFKGIPFAa..KaLEnPqrHPGWQGT LKak.FKKRCLQATiTQDsTYG

150
Bssl DEDCLYLNIWVPQGRKQVSRDLPVMIWIYGGAFILMGSGBGANFLNNYLYD
Ratlpl QEDCLYLNIWVPQGRKQVSHDLPVMVWIYGGAFILMGSGQGANFLKNYLYD
Bovceh NEDCLYLNIWVPQGRKEVSHDLPVMIWIYGGAFILMGASQGANFLSNYLYD
Consensus .EDCLYLNIWVPQGRKqVShDLPVMiWIYGGAFILMGsgqGANFL.NYLYD

200
Bssl GKKIATRGNVIVVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKR
Ratlpl GKKIATRANVIVVTFNYRVGPLGFLSTGDANLPGNFGLRDQHMAIAWVKR
Bovech GKKIATRGNVIVVTFNYRVGPLGFLSTGDSNLPGNYGLWDQHMAIAWVKR
Consensus GKKIATRgNVIVVTFNYRVGPLGFLSTGDaNLPgNyGLrDQHMAIAWVKR

* 250
Bssl NIAAFGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSP
Ratlpl NIAAFGGDPNNITIFGESAGGAIVSLQTLSPYNKGLIRRAISQSGVALSP
Bovceh NIEAFGGDPDNITLFGESAGGASVSLQTLSPYNKGLIKRGISQSGVGLCP
Consensus NIaAFGCDPDNITLFGESAGGAsVSLQTLSPYNKGLIrRaISQSGVaLsP

300
Bssl WVIQKNPLFWAKKVAKKVGCPVGDAARMAQCLKVTDPRALTLAYKVPLAG
Ratlpl WAIQENPLFWAKTIAKKVGCPTEDTAKMAGCLKITDPRALTLAYRLPLKS
Bovech WAIQQDPLFWAKRIAKKVGCVPDDTSKMAGCAKITDPRALTLAYKLPLGS
Consensus WaIQ.nPLFWAK.iAKKVGCPv.DtakMagClKiTDPRALTLAYk1PL.s

350
Bssl LEYFMLHYVGFVPVIDGDFIPADPINLYANAADIDYIAGTNNMDGHIFAS
Ratlpl QEYPIVEYLA FIPVVDGDFIPDDPINLYDNAADIDYLAGINDMDGHLFAT
Bovceh TEYPKLHYLSFVPVIDGDFIPDDPVNLYANAADVYIAGTNDMDGHLFVG
Consensus .EYP.lHYl.FvPViDGDfIPdDPiNLYaNAADiDYiAGtNdMDGHlPa.

30 JAN 1997

Fig. 5-2/3

400

Bssl	IDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGAKTTFDVYTESWAQDPS
Ratlpl	VDVPAIDKAKQDVTEEDFYRLVSGHTVAKGLKGTQATFDIYTESWAQDPS
Bovech	MDVPAINSNKQDVTEEDFYKLVSGLTVTKGLRGANATYEVYTEPWAQDSS

Consensus .DvPAInk.kqdVTEEDFYkLVsG.TvtKGLrGa.atfdvYTESWAQDps

450

Bssl	QENKKKTVDVFETDVLFLVPTETALAQHRANAKSAKTYAYLFSHPSRMPV
Ratlpl	QENMKKTVAFETDILFLIPTEMALAQHRAHAKSAKTSYLFShPSRMPi
Bovceh	QETRKKTMVDLETDILFLIPTKIAVAQHKSHAKSANTYTYLFsQPSRMPi

Consensus QEn.KKtVdFETDiLFLiPTeAlAQHrahAKSAkTY.YLFSHPSRMPi

500

Bssl	YPKWVGADHADDIQQYVFGKPFATPTGYRPQDRTVSKAMIAWTFNAKTGD
Ratlpl	YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAWTFNAKSGD
Bovech	YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAWTFNFARTGD

Consensus YPKWmGADHADDlQYVFGKPFATPlGYRaQDRTVSKAMIAWTFNAktGD

550

Bssl	PNMGDSAVPTHWEPTYTENSGLYLEITKKMGSSSMKRSRLTNFLRYWTLTY
Ratlpl	PNMGNSPVPTHWYPYTMENGNYLDINKKITSTSMKEHLREKFLKFWAVTF
Bovceh	PNTGHSTVPANWDPYTLEDDNYLEINKQMDNSNMKLRLTLNLYLQFWTQTY

Consensus PNnG.S.VPthW.PYT.Kn.nYleInKkm.S.SMK.hLRtnfL.fwt.Ty

596

Bssl	LALPTVTDQ	EATVPVPTGDS	EATVPVPTGDS	ETAPVPPTGDS	GAPP
Ratlpl	EMLPTV...	VGDHTPPKDDs	EAAPVPPTDDs	DGGVPVPTDDs	QTPP
Bovech	QALPTVtSA	GASLLPPEDNS	QASVPVPPADNS	GAPTEPSAGDS	

Consensus .aLPTvt.....a....PP.ddS.eA.PVPptddS....pvPptgDS....p

642

Bssl	VPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVP
Ratlpl	VPPTDNS	QA.....			
Bovceh					

ConsensusS..a.....



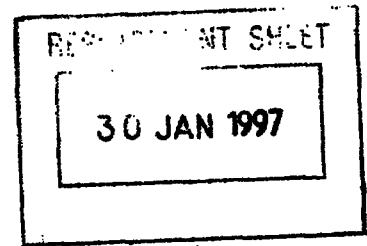


Fig. 5-3/3

					689
Bssl	PTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GPPVPTPTGDS	GAPPVPPTG
Rarlpl					
Bovech					
Consensus					

					735
Bssl	DS	GAPPVTPTGDS	ETAPVPPTGDS	GAPPVPPTGD	SEAAPVPPTDDS
Ratlpl	..				GDS
Bovceh	..				
Consensus					ds

Bssl KE.AQMPAVIRF
 Ratlpl VE.AQMPGPIGF
 Bovech .EVAQMPVVIGF
 Consensus .e.AQMP.vIgF

This Sheet Has Been Replaced



30 JAN 1997

1

(76-90)

BSSL

(116-130)

ChesHum

(114-128)

Torpace

(101-115)

Drosceh

(103-117)

Ratlivce

(224-238)

Drosace

(2257-2270)

ThryHum

(90-114)

Dict.Di

Consensus

- 12 -

G	D	E	D	C	L	Y	L	N	I	W	V	P	Q	G
L	S	E	D	C	L	Y	L	N	I	W	I	P	A	P
M	S	E	D	C	L	Y	L	N	I	W	V	P	S	P
G	E	E	D	C	L	T	V	S	V	Y	K	P	K	N
F	S	E	D	C	L	Y	I	N	V	W	A	P	A	K
V	S	E	D	C	L	Y	L	N	V	W	A	P	A	K
V	S	E	D	C	L	Y	L	N	V	F	I	P	Q	.
A	Q	K	C	N	L	G	P	G	V	C	S	P	M	G

Fig.6-1/3

This Sheet Has Been Replaced



2

(95-113)

BSSL

(134-150)

ChesHum

(132-150)

Torpace

(118-136)

Drosceh

(121-138)

Ratlivce

(273-291)

Drosace

(2270-2291)

ThryHum

Dict.Di

Consensus

S	R	D	L	P	V	M	I	W	I	Y	G	G	A	F	L	M	G	S
P	K	D	A	T	V	L	I	W	I	Y	G	G	G	F	Q	T	G	T
P	K	S	A	T	V	M	L	W	I	Y	G	G	G	F	Y	S	G	S
R	N	S	F	P	V	V	A	H	I	H	G	G	A	F	M	F	G	A
N	S	R	L	P	V	M	V	W	I	H	G	G	G	L	I	I	G	G
T	N	G	L	P	I	L	I	W	I	Y	G	G	G	F	M	T	G	S
A	P	N	A	S	V	L	V	F	F	H	N	T	M	D	R	E	E	S

3

(135-156)

BSSL

(166-187)

ChesHum

(164-185)

Torpace

(150-171)

Drosceh

(153-174)

Ratlivce

(307-328)

Drosace

(2307-2328)

ThryHum

(118-136)

Dict.Di

Consensus

N	V	I	V	V	T	F	N	Y	R	V	G	P	L	G	F	L	S	T	G	D	A
R	V	I	V	V	S	M	N	Y	R	V	G	A	L	G	F	L	A	L	P	G	N
E	V	V	L	V	S	L	S	Y	R	V	G	A	F	G	F	L	A	L	H	G	S
K	F	I	L	V	K	I	S	Y	R	L	G	P	L	G	F	V	S	T	G	D	E
N	V	V	V	V	T	I	Q	Y	R	L	G	F	G	G	L	F	S	T	G	D	E
N	V	I	V	A	S	F	Q	Y	R	V	G	A	F	G	F	L	H	L	A	P	E
N	L	I	V	V	T	A	S	Y	R	V	G	V	F	G	F	L	S	S	G	S	G
S	V	I	V	V	T	I	N	Y	R	L	G	I	L	G	L	M	G	T	.	.	.

REPLACEMENT SHEET

30 JAN 1997

Fig.6-2/3

4

(157-182)
BSSL
(189-214)
ChesHum
(187-212)
Torpac
(172-197)
Drosceh
(175-200)
Ratlivce
(336-361)
Drosace
(2329-2354)
ThryHum

N	L	P	G	N	Y	G	L	R	D	Q	H	M	A	I	A	W	V	K	R	N	I	A	A	F	G
E	A	P	G	N	M	G	L	F	D	Q	Q	L	A	L	Q	W	V	Q	K	N	I	A	A	F	G
E	A	P	G	N	M	G	L	L	D	Q	R	M	A	L	Q	W	V	H	D	N	I	Q	P	F	G
D	L	P	G	N	Y	G	L	K	D	Q	R	L	A	L	K	W	I	K	Q	N	I	A	S	F	G
H	S	R	G	N	W	A	H	L	D	Q	L	A	A	L	R	W	V	Q	D	N	I	A	N	F	G
E	A	P	G	N	V	G	L	W	D	Q	A	L	A	I	R	W	L	K	D	N	A	H	A	F	G
E	V	S	G	N	N	G	L	L	D	Q	V	A	A	L	T	W	V	O	T	H	I	R	G	F	G

Dict.Di

.....

Consensus

e l p G N w q l l D Q . . A l . W v . d n i a a F G

5

(183-207)
BSSL
(215-239)
ChesHum
(213-237)
Torpac
(198-222)
Drosceh
(201-225)
Ratlivce
(362-386)
Drosace
(2354-2379)
ThryHum
(137-149)
Dict.Di

G	D	P	N	N	I	T	L	F	G	E	S	A	G	G	A	S	V	S	L	Q	T	L	S	P
G	N	P	K	S	V	T	L	F	G	E	S	A	G	A	A	S	V	S	L	H	L	L	S	P
G	D	P	K	T	V	T	L	F	G	E	S	A	G	G	A	S	V	G	M	H	I	L	S	P
G	E	P	O	N	V	L	L	V	G	H	S	A	G	G	A	S	V	H	L	Q	M	L	R	E
G	N	P	D	S	V	T	I	F	G	E	S	A	G	G	V	S	V	S	A	L	V	L	S	P
G	N	P	E	W	M	T	L	F	G	E	S	A	G	S	S	S	V	N	A	Q	L	M	S	P
G	P	P	R	R	V	S	L	A	A	D	R	G	G	A	D	V	A	S	I	H	L	L	T	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	A	F	S	V	S	A	H	L	T

Consensus

G d P . n v t l f g e s a G g a s v s l . l l s p

This Sheet Has Been Replaced



30 JAN 1997

Fig. 6-3/3

6

(208-231)
BSSL
(240-263)
ChesHum
(238-261)
Torpace
(223-246)
Drosceh
(226-249)
Ratlivce
(387-410)
Drosace
(2382-2402)
ThryHum
(150-173)
Dict.Di

Y	N	K	G	L	I	R	R	A	I	S	Q	S	G	V	A	L	S	P	W	V	I	Q	K
G	S	H	S	L	F	T	R	A	I	L	Q	S	G	S	F	N	A	P	W	A	V	T	S
G	S	R	D	L	F	R	R	A	I	L	Q	S	G	S	P	N	C	P	W	A	S	V	S
D	F	G	Q	L	A	R	A	A	F	S	F	S	G	N	A	L	D	P	W	V	I	Q	K
L	A	K	N	L	F	H	R	A	I	S	E	S	G	V	V	L	T	T	N	L	D	K	K
V	T	R	G	L	V	K	R	G	M	M	Q	S	G	T	M	N	A	P	W	S	H	M	T
T	N	S	Q	L	F	R	R	A	V	L	M	G	G	S	A	L	S	P	A	A	V	I	S
Y	S	R	Q	Y	F	N	A	A	I	S	S	S	S	P	L	T	V	G	L	K	D	K	T

Consensus n k g L f r r a i . q s G s a l s p w a i q .

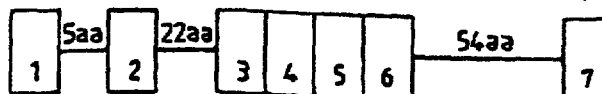
7

(285-297)
BSSL
(316-328)
ChesHum
(314-326)
Torpace
(298-310)
Drosceh
(298-306)
Ratlivce
(466-477)
Drosace
(2458-2468)
ThryHum
(224-236)
Dict.Di

V	G	F	V	P	V	I	D	G	D	F	I	P
V	N	F	G	P	T	V	D	G	D	F	L	T
F	S	F	V	P	V	I	D	G	E	F	F	P
A	P	F	S	P	V	L	E	P	S	D	A	P
.	.	.	.	T	V	I	D	G	V	V	L	P
P	S	.	A	P	T	I	D	G	A	F	L	P
.	.	W	G	P	V	I	D	G	H	F	L	R
T	I	W	S	P	V	I	D	G	D	A	F	I

Consensus . . f . p v . d g d f . p

This Sheet Has Been Replaced



REPLACEMENT SHEET

30 JAN 1997

- 15 -

10	30	50
AKLGAVYTEGGFVEGVNKKLGLLGDSVDIFKGIPFAAPTKALENPQPHPGWQGTAKNF		
70	90	110
KKRCLQATITQDSTYGDEDCLYLNIVWPQGRKQVSRDLPVMIWIYGGAFLMGSGHGANFL		
130	150	170
NNYLYDGEEIATRGNVIVVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKRNIAA		
190	210	230
FGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSPWVIQKNPLFWAKKV		
250	270	290
AEKVGCPVGDAAARMAQCLKVTDPRALTAYKVPLAGLEY PMLHYGVFPVIDGDFIPADP		
310	330	350
exon a		
INLYANAADIDYIAGTNNMDGHIFASIDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGA		
370	390	410
KTTFDVYTESWAQDPSQENKKKTVDVFETDVLFLVPTEIALAQHRANAKSAKTYAYLPSH		
430	450	470
exon b		
PSRMPVYPKWVGADHADDIQYVFGKPFATPTGYRPQDRTVSKAMIAYWTFNFAKTGDPNMG		
490	510	530
exon c		
DSAG ^{for} PTHWEPYTTENSGYLEITKKMGSSSMKRSRLRTNFLRYWTLTYLALPTVTDQEATPV		
550	570	590
exon d		
PPTGDSEATVPPTGDSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD		
610	630	650
SGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDAGPPVPPTGDSGAPP		
670	690	710
VPPTGDSGAPPVTPPTGDSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD		

RF

FIGURE 7

This Sheet Has Been Replaced



REPLACEMENT SHEET

03 FEB 1998

REPLACEMENT SHEETS SG

☒ Description : paged 1 to paged 33
☒ Claim(s) : paged 34 to paged 37
☒ Drawing(s) : paged 1 to paged 14
☒ Abstract : paged to paged

DATE OF RECEIPT

03 FEB 1998

DERIVATIVES OF HUMAN BILE-SALT STIMULATED LIPASE AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM

Field of the invention

The present invention relates to new DNA sequences, to new
 5 proteins coded for by such DNA sequences, and to the use
 as further described below of such proteins. The invention
 also encompasses vectors, such as plasmid constructs,
 comprising such DNA sequences, being capable of expressing
 the desired enzyme. The invention also includes host
 10 organisms transfected with such constructs e.g. bacteria
 yeast, mammalian cells, and transgenic animals. The
 invention also includes processes for the preparation of
 the novel products of the invention. The new proteins of
 the invention are related to an enzyme, known i.a. as
 15 human bile salt-stimulated lipase.

20

25

30

35

This Sheet Has Been Replaced



03 FEB 1998

Background to the invention

The human lactating mammary gland synthesizes and secretes
5 with the milk a bile salt-stimulated lipase (BSSL) [1]
that, after specific activation by primary bile salts [2,
57, 58], contributes to the breast-fed infant's endogenous
capacity of intestinal fat digestion [3-5]. This enzyme,
10 [6], is a non specific lipase; in vitro it hydrolyses not
only tri-, di- and monoacylglycerols, but also
cholesteryl-, and retinyl esters, and
lysophosphatidylglycerols [7-10]. Furthermore, its
activity is not restricted to emulsified substrates, but
15 micellar and soluble substrates are hydrolyzed at similar
rates [11].

BSSL is not degraded during passage with the milk through
the stomach, and in duodenal contents it is protected by
bile salts from inactivation by pancreatic proteases such
20 as trypsin and chymotrypsin [2,11]. It is, however,
inactivated when the milk is pasteurized, e.g. heated to
62.5 °C, 30 min [12]. Model experiments in vitro suggest
that the end products of triacylglycerol digestion are
different in the presence of BSSL [5,7]. Due to lower
25 intraluminal bile salt concentrations during the neonatal
period [13,14] this may be beneficial to product
absorption [5,15].

The carboxylic ester hydrolase (CEH) of human pancreatic
juice [16] seems functionally to be identical, or at least
30 very similar, to BSSL [5]. They also share common epitopes
[8,17], have identical N-terminal amino acid sequences
[17] and are inhibited by inhibitors of serine esterases,
e.g. eserine and diisopropylfluorophosphate [6,8,16]. It
has been hypothesized that the two enzymes are products of
35 the same gene [18,19]. The observed molecular size
difference [8,19] could be explained by different patterns
of glycosylation, as recently suggested [17].



03 FEB 1998

Dietary lipids is an important source of energy. The energy-rich triacylglycerols constitute more than 95% of these lipids. Some of the lipids, e.g. certain fatty acids and the fat soluble vitamins, are essential dietary constituents. Before gastro-intestinal absorption the triacylglycerols as well as the minor components, i.e. esterified fat-soluble vitamins and cholesterol, and diacylphosphatidylglycerols, require hydrolysis of the ester bonds to give rise to less hydrophobic, absorbable products. These reactions are catalyzed by a specific group of enzymes called lipases.

In the human adult the essential lipases involved are considered to be gastric lipase, pancreatic colipase-dependent lipase (tri- and diacylglycerol hydrolysis), pancreatic phospholipase A2 (diacylphosphatidylglycerols) and carboxylic ester hydrolase (cholesteryl- and fat soluble vitamin esters). In the breast-fed newborn bile salt-stimulated lipase plays an essential part in the hydrolysis of several of the above mentioned lipids. Together with bile salts the products of lipid digestion form mixed micelles from which absorption occurs (3-5).

Common causes of lipid malabsorption, and hence malnutrition, are reduced intraluminal levels of pancreatic colipase-dependent lipase an/or bile salts. Typical examples of such lipase deficiency are patients suffering from cystic fibrosis, a common genetic disorder resulting in a life-long deficiency in som 80% of the patients, and chronic pancreatitis, often due to chronic alcoholism.

The pancreatic and liver functions are not fully developed at birth, most notably in infants born before term. Fat malabsorption, for physiological reasons, is a common finding and thought to result from low intraluminal pancreatic colipase-dependent lipase and bile salt concentrations (3,4,13-15). However, because of BSSL, such



U3 FEB 1998

4

malabsorption is much less frequent in breast-fed infants than in infants fed pasteurized human milk of infant formulas (3-5, 12, 59, 60, 61). This is one reason why it has been advocated that newborn infants, particularly
5 preterm infants, that cannot be fed their own mothers milk should be fed non-pasteurized milk from other mothers (12).

The present treatment of patients suffering from a
10 deficiency of pancreatic lipase is the oral administration of very large doses of a crude preparation of porcine pancreatic enzymes. Colipase-dependent pancreatic lipase is inactivated by low pH. Such conditions are prevalent in the stomach, with the result that orally administered
15 pancreatic lipase is virtually completely inactivated on the passage through the stomach to the gut. Therefore, this effect cannot be completely overcome by the use of large doses of enzyme. The large doses administered are inadequate for most patients, and the preparations are
20 impure and unpalatable. Certain tablets have been formulated which pass through the acid regions of the stomach and discharge the enzyme only in the relatively alkaline environment of the jejunum. However, many patients suffering from pancreatic disorders have an
25 abnormally acid jejunum and such tablets may fail to discharge the enzyme and may therefore be ineffective. Moreover, since the preparations presently on the market are of a non-human source there is a risk of immunoreactions that may cause harmful effects to the
30 patients or result in reduced therapy efficiency. A further drawback with the present preparations is that their content of other lipolytic activities than colipase-dependent lipase are not stated. In fact, most of them contain very low levels of CEH/BSSL-activity. This may be
35 one reason why many patients, suffering from cystic fibrosis in spite of supplementation therapy, suffer from deficiencies of fat soluble vitamins and essential fatty

This Sheet Has Been Replaced



03 FEB 1998

5

This Sheet Has Been Replaced



acids.

Thus, there is a great need for products with properties and structure derived from human lipases and with a broad substrate specificity, which products may be orally administered to patients suffering from deficiency of one or several of the pancreatic lipolytic enzymes. The products of the present invention fulfil this need by themselves or in combination with other lipases or in combination with preparations containing other lipases. Furthermore, for some human infants there is an obvious need to improve fat utilization from conventional infant formulas, or pasteurized human milk from so-called milk banks.

BSSL has several unique properties that makes it ideally suited for substitution and supplementation therapy: It has been designed by nature for oral administration. Thus, it resists passage through the stomach and is activated in contents of the small intestine. Its specific activation mechanism should prevent hazardous lipolysis of food or tissue lipids during storage and passage to its site of action. Due to its broad substrate specificity it has the potential to, on its own, mediate complete digestion of most dietary lipids, including the fat soluble vitamin esters.

BSSL may be superior to pancreatic colipase-dependent lipase to hydrolyze ester bonds containing long-chain polyunsaturated fatty acids.

In the presence of gastric lipase and in the absence of, or at low levels of colipase-dependent lipase BSSL can ascertain a complete triacylglycerol digestion in vitro even if the bile salt levels are low such as in newborn infants. In the presence of BSSL the end products of triacylglycerol digestion become free fatty acids and free glycerol rather than free fatty acids and monoacylglycerol generated by the other two lipases (5). This may favour product absorption particularly when the intraluminal bile

REPLACEMENT SHEET

03 FEB 1998

6

salt levels are low (3,15).

From a historical point of view infant formulas have been developed, and improved, from the concept that their composition should be as similar to that of human milk as possible. It is desirable to supplement such formulas.

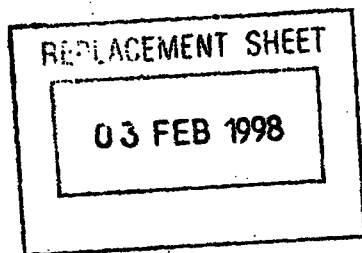
The utilization for supplementation, substitution or therapy of bile salt-stimulated lipases (BSSL), or of proteins with the essential functions of BSSL, requires however access to quantities of the product on a large technical scale. It is not possible in factory scale to rely on natural sources such as milk as starting material. Besides the problem mentioned above with inactivation of BSSL during pasteurization, there is the additional risk of contamination of material from a natural source with infectious agents, e.g. vira such as HIV virus and CMV. There is, accordingly, a need for large scale access to products having BSSL properties. The present invention provides such products and methods for their preparation.

20

Prior art references are given later in this specification.

This Sheet Has Been Replaced





The invention

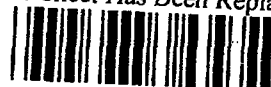
- The present invention is based on the cloning of cDNA coding for BSSL derived from human mammary gland. We have
- 5 also isolated, from human pancreas, a partial cDNA coding for CEH. Deduced amino acid sequences from the human cDNA's and comparison with CEH from other species, support the interpretation that BSSL and CEH are identical.
- 10 As will be further detailed below, it was surprisingly found that the structure of the protein as deduced from the cDNA sequence is quite different from the structure of other lipases. The structure proved unexpectedly to be more like the structure of typical esterases, such as
- 15 cholinesterase.

This Sheet Has Been Replaced



03 FEB 1998

This Sheet Has Been Replaced



8

With reference to Figure II and Figure VII, products of the invention are:

- 5 a) a protein as defined by the amino acid sequence 1-722 in Fig. VII,
- b) a protein as defined by the amino acid sequence 1-535 in Fig. VII,
- 10 c) a protein as defined by the amino acid sequence 1-278 in Fig. VII,
- d) a protein as defined by the amino acid sequence 1-341 in Fig. VII,
- 15 e) a protein as defined by the amino acid sequence 1-409 in Fig. VII,
- f) a protein as defined by the amino acid sequence 1-474 in Fig. VII,
- 20 g) combinations of proteins defined under b) - f) e.g. as defined by the amino acid sequence in positions 1-278, 279-341, 279-409, 279-474, 342-409, 342-474 and 536-722.
- 25 h) combinations of proteins defined under b) - g) in combination with one or more of the repeats according to Figure V,
- 30 i) a protein as defined under a) - h) possessing an additional, N-terminal amino acid, namely methionine, and functionally equivalent variants and mutants of the proteins defined in a) - i) above;
- 35 It should be noted that the proteins under a, b, c, d, e, f, h and i above will not be identical in all respects to naturally occurring BSSL, but they will exhibit one or

03 FEB 1998

more of the critical functions of naturally occurring BSSL. Critical functions are given below.

j) a DNA sequence coding for the proteins defined in a,
5 b, c, d, e, f, h and i above.

k) a DNA sequence according to Fig. II, defined by the following nucleotide numbers in Fig. II:

10

a) a DNA sequence 151-2316 according to Fig. II, coding for the protein defined by the amino acid sequence 1-722 in Fig. VII,

b) a DNA sequence 151-1755 according to Fig. II, coding
15 for the protein defined by the amino acid sequence 1-535 in Fig. VII,

c) a DNA sequence 151-985 according to Fig. II, coding for the protein defined by the amino acid sequence 1-278 in Fig. VII,

20 d) a DNA sequence 151-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 1-341 in Fig. VII,

e) a DNA sequence 151-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 1-409
25 in Fig. VII,

f) a DNA sequence 151-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 1-474 in Fig. VII,

g) a DNA sequence 986-1172 according to Fig. II, coding
30 for the protein defined by the amino acid sequence 279-341 in Fig. VII,

h) a DNA sequence 986-1375 according to Fig. II, coding for the protein defined by the amino acid sequence 279-409 in Fig. VII,

35 i) a DNA sequence 986-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 279-474 in Fig. VII,

This Sheet Has Been Replaced



03 FEB 1998

This Sheet Has Been Replaced

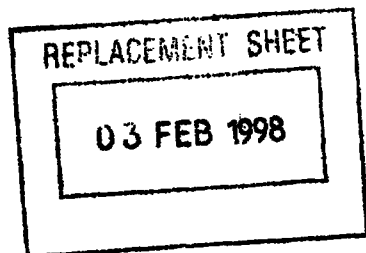


10

- j) a DNA sequence 1173-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 342-409 in Fig. VII,
- k) a DNA sequence 1173-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 342-474 in Fig. VII.

Significant functions of proteins of the invention are

- 10 a) suitable for oral administration,
- b) being activated by specific bile salts,
- c) acting as a non-specific lipase in the contents of the small intestines, that is being able to hydrolyze lipids relatively independent of their chemical structure and physical state (emulsified, micellar, soluble).
- 15 d) Ability to hydrolyze triacylglycerols with fatty acids of different chain-length and different degree of unsaturation.
- e) Ability to hydrolyze also diacylglycerol, monoacylglycerol, cholesteryl esters,
- 25 lysophosphatidylacylglycerol, and retinyl and other fat soluble vitamin-esters.
- f) Ability to hydrolyze not only the sn-1(3) ester bonds in a triacylglycerol but also the sn-2 ester bond.
- 30 g) Ability to interact with not only primary but also secondary bile salts.
- h) Dependency on bile salts for optimal activity.
- 35 i) Stability so that gastric contents will not affect the catalytical efficiency to any substantial degree.



- j) Stability towards inactivation by pancreatic proteases, e.g. trypsin, provided bile salts are present.
- 5 k) Ability to bind to heparin and heparin derivatives, e.g. heparan sulphate.
- l) Ability to bind to lipid-water interphases.
- 10 m) Stability to permit lyophilization.
- n) Stability when mixed with food constituents such as in human milk, or milk formula.
- 15 The critical functions for supplementation, substitution, or therapy are these according to a), c), d), e), f), i), j) and l). For other purposes, not all critical functions may be necessary.
- 20 For expression of the proteins indicated above, the appropriate DNA sequence indicated above will be inserted into a suitable vector which then is introduced into a suitable host organism. The said vector will also have to comprise appropriate signal and other sequences enabling
- 25 the organism to express the desired protein.

Suitable expression organisms:

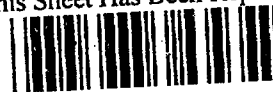
- With the recombinant DNA techniques it is possible to
- 30 clone and express a protein of interest in a variety of prokaryotic and eukaryotic host organisms. Possible expression organisms are bacteria, simple eukaryots (yeast), animal cell cultures, insect cell cultures, plant cell cultures, plants and transgenic animals. Each
- 35 individual system has its own particular advantages and disadvantages. The simple conclusion is that every gene to be expressed is a unique problem and no standard solution



03 FEB 1998

12

This Sheet Has Been Replaced



is available.

Commonly used bacterial systems are *E. Coli*, *Bacillus subtilis*, *Streptomyces*. Commonly used yeasts are
5 *saccharomyces*, and *Pichia Pastoris*. Commonly used animal cells are CHO cells and COS cells. Commonly used insect cell cultures are *Drosophila* derived cells.

Commonly used plant is the tobacco plant. Possible
10 transgenic animals are goat and cow.

Possible bacterial vectors are exemplified by pUC and protein A-vectors.

Possible yeast vector is exemplified by pMA 91.

15 Possible insect vectors are derived from Baculo-virus.
Possible animal cell vectors are derived from SV/40.
Possible plant vectors are derived from the Ti-plasmid.
In every system, both natural and synthetic promoters and terminators can be used.

20

It is understood that depending on the choice of expression system, the expressed protein may contain an additional N-terminal amino acid (methionine), contain a few extra amino acids, or be fused to a heterologous
25 protein, (e.g. protein A), or differ from the naturally occurring protein with respect to glycosylation.
Furthermore, the vectors may also contain signal sequences in order to export the protein to the periplasm or to the culture medium.

30 Thus, further aspects of the invention are:

- a) a vector comprising a DNA sequence coding for a protein as specified above,
- b) a host organism comprising a DNA sequence as specified above,
- 35 c) a process for the production of a protein as specified above, by growing a host organism containing a vector as specified under a) above and

03 FEB 1998



Methods from purification are based on the expression system used (e.g. protein A/IgG) and/or on methods used for purification of the naturally occurring enzyme, as described in reference 6.

5

Additional aspects of the invention are:

- a pharmaceutical composition comprising a protein as specified above,

10

- the use of a protein as specified above for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency,

15

- the use of a protein as specified above for the manufacture of a medicament for the treatment of cystic fibrosis,

20 - the use of a protein as specified above as a supplement to an infant food formulation,

-the use of a protein as specified above for the manufacture of a medicament for the treatment of chronic
25 pancreatitis,

-the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption of any etiology,

30

-the use of a protein as specified above for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins,

35 -the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons, e.g. in new-

03 FEB 1998

born infants.

5 The DNA sequence in Fig. II from position 151 up to and including position 2316 is the sequence coding for the entire protein. The sequence from position 2317 up to and including position 2415 is not translated to protein, but is included in exon d identified in Table 2 below.

10 In one embodiment of the invention, the protein as defined in paragraphs a) - i) above is provided in isolated form and/or in substantially pure form.

15 The DNA sequences as defined in paragraphs a) - k) above are in one embodiment of the invention provided in isolated form and/or in substantially pure form.

Experimental part

20 Abbreviations

aa, amino acid; bp, base pair; BSSL, bile salt-stimulated lipase; c-AMP, cyclic adenosine monophosphate; CEH, carboxylic ester hydrolase; Da, dalton; c ⁷GTP, 7-deaza-2-deoxyguanosine 5'triphosphate; EDTA, ethylene diamine tetraacetate; kb, kilobases; MOPS, 3-N-morpholino-
25 propanesulfonic acid; nt, nucleotide; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; SSC, NaCl citrate, xGal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

30

Enzymes

Bile salt-stimulated lipase EC 3.1.1.3

Carboxylic ester hydrolase EC 3.1.1.1

35

This Sheet Has Been Replaced



03 FEB 1998

Material and Methods**A. Enzyme and antibody preparation**

- 5 BSSL was purified from human milk as previously described [6]. When used for antibody production the enzyme was further purified by SDS-PAGE. The protein band corresponding to the lipase was, after staining with Coomassie Brilliant blue, electroeluted from the gel.
- 10 Twentyfive μ g of purified enzyme, together with an equal volume of Freund's complete adjuvant, was used for a first i.c. injection and the same amount of enzyme with incomplete adjuvant for the subsequent monthly booster injections. The rabbits were bled about two weeks after
- 15 each booster and sera prepared and stored at -20°C .

B. Preparation of tryptic fragments and amino acid sequence analysis

- 20 Three mg of purified BSSL was dissolved in 1 ml of 0.1M Tris-Cl buffer, pH 8.5, containing 6M guanidinium hydrochloride and 2 mM EDTA. Dithioerythritol was added to 5mM. After incubation at 37°C for 2h, 300 μ l 50 mM iodoacetate was added. After 90 min incubation at 25°C in
- 25 darkness the reduced and carboxymethylated enzyme was desalted on a Sephadex G-25 column, equilibrated with 0.5M ammonium bicarbonate. Thirty μ g of tosyl-L-phenylalanine chloromethane treated bovine trypsin (Worthington diagnostics system Inc., Freehold, NJ, USA)
- 30 was added before lyophilization. The lyophilized protein was dissolved in 4 ml 0.1M ammonium bicarbonate and an additional 90 μ g of trypsin was added. After 5h incubation at 37°C the protein was again lyophilized. The tryptic digest was dissolved in 0.1% trifluoroacetic acid
- 35 (2mg/ml). Three hundred μ g of trypsinated BSSL was chromatographed on HPLC using a C-18 reversed phase column and eluted with a gradient of 0-50% acetonitrile in 0.1%



03 FEB 1998

16

trifluoroacetic acid. Peptide collection was monitored by continuous recording of the absorbance at 215 nm. Peptides to be sequenced were further purified by rechromatography using the same column with adjusted gradients.

- 5 Samples of peptide fragments to be sequenced were dried under nitrogen to remove acetonitrile and applied to the sequencer. For N-terminal sequence analysis, native BSSL was dissolved in 0.1% acetic acid. Sequence analyzes were performed on an Applied Biosystems Inc. 477A pulsed
10 liquid-phase sequencer and on-line PTH 120A analyzer with regular cycle programs and chemicals from the manufacturer. Initial and repetitive yields, calculated from a sequenced standard protein, β -lactoglobulin, were 47 and 97%, respectively.

15

C. Isolation of RNA

- Samples of human pancreatic adipose and lactating mammary gland tissues were obtained at surgery and immediately put
20 into guanidinium thiocyanate (1-5 g in 50 ml). Total RNA was extracted as described by Chirgwin [20]. Poly(A)-RNA was prepared by chromatography on oligo-deoxythymidilate-(oligo(DT))-cellulose column [21].

25 D. Construction and screening of cDNA libraries

- Approximately 15 μ g poly-adenylated RNA from human pancreas was denatured with methyl mercuric hydroxide [22] and primed with oligo (dT) ₁₂₋₁₈ primers (Pharmacia,
30 Uppsala, Sweden), and reversely transcribed using standard procedures [23]. Second-strand synthesis was carried out according to Gubler and Hoffman [24], except that DNA ligase and β -NAD were omitted, and the reaction temperature was set at 15 °C. Excess RNA was digested with
35 RNase A (50 μ g/ml), and the double-stranded cDNA was treated with EcoRI methylase [25]. Ends were blunted with Klenow enzyme. After ligation to EcoRI linkers and

This Sheet Has Been Replaced



03 FEB 1998

17

cleavage with EcoRI the cDNA was fractionated on a Sepharose 4B-C1 column. The void volume fraction was precipitated with ethanol and the cDNA ligated into the EcoRI site of a phosphatase treated gtl1 vector [26]. In
5 vitro packing yielded more than 7×10^3 recombinants.

A cDNA library from human mammary gland, derived from tissue obtained from a women at the eighth month of pregnancy, was purchased (Clontech Laboratories, Inc.,
10 Palo Alto, CA; USA).

Phages from the cDNA libraries were plated at 5×10^4 plaque forming units per 120-mm dish. The antiserum was diluted to a ratio of 1:3200 and screening was performed according
15 to Young and Davis [27]. Alkaline-phosphatase-conjugated goat-anti-rabbit antibodies were used as second antibodies (Bio-Rad, Richmond, CA USA). To isolate clones corresponding to the 5'-end of the mRNA, nucleic acid hybridization was done under standard conditions [23]
20 using a subcloned fragment from one of the immunopositive clones as a probe.

E. RNA analysis

25 Electrophoresis was carried out in a 1% agarose gel in 40mM MOPS buffer pH 7.0 after denaturation with glyoxal and dimethylsulfoxide [28]. Glyoxalated total RNA was then transferred to nitrocellulose filters [29]. The blots were probed with subclones of BSSL and CEH recombinants that
30 were labelled by the oligo-labeling technique [30]. Prehybridization and hybridization were carried out with 50% formamid at 46 °C [23]. Posthybridization washes were performed at high stringency (0.1% SDS and 0.1xSSC at 60° C). (1xSSC, 0.15M NaCl, 0.0015M Na, citrate, pH 7.6).
35

F. Nucleotide sequence

This Sheet Has Been Replaced



U3 FEB 1998

cdNA inserts from BSSL and CEH recombinants were either directly cloned into M3mp18 and mp19 after sonication and size fractionation or some of them were further subcloned into pTZ19R after digestions with PstI, BstXI, NarI, SmaI and AhaII. The nucleotide sequence was determined by the dideoxy chain-termination method [31]. The GC-rich repeat (see below) were also sequenced with TaqI polymerase and dc ³GTP. Both strands were sequenced. Sequence information was retrieved from autoradiograms by use of the software MS-EdSeq as described by Sjöberg et al [55].

G. Amino acid sequence predictions and homologies.

To predict the corresponding amino acid sequence of the cdNA inserts, codon usage of different reading frames was compared according to Staden and gave one open reading frame [32]. Homologies were searched for with the program of the UWCGG software package [33].

20 Results and Discussion

A. Sequences of tryptic fragments and the N-terminus of BSSL

Trypsin digestion of purified BSSL resulted in approximately 50 fragments as judged by the number of peaks obtained during the HPLC-chromatography (Fig. I). The peaks were collected and the indicated peaks which could be isolated in a highly purified state and in reasonable quantities, were sequenced. The resulting sequences are shown in Table I. In addition the 30 most N-terminal residues were sequenced (Fig. II), and they confirm the previously reported sequence of Abouakil et al. (30 residues) [17].

35

The 27 residues long N-terminal sequence reported from Wang & Johnson [34] is a glycine in our report and a



03 FEB 1998

lysine in Wang & Johnson's.

B. Nucleotide sequence of BSSL

5 For construction of the λ gt11 cDNA library we used polyadenylated RNA from human pancreas. Initially four immunopositive clones were isolated, and then this pancreatic expression cDNA library was screened with antiserum against BSSL. Nucleotide sequence analysis of
10 the four clones showed that they are in perfect agreement and correspond to the 3'-end of the mRNA. They all begin with a poly A tail and differ only in length; the longest insert, designated ACEH, spans 996 bp.

A cDNA library from human mammary gland was screened with
15 antiserum, and the pancreas clone ACEH as probe. Positive clones were isolated from both screenings, which all originate from the 3'-end. The longest mammary gland clone, designated ABSSL, reaches 2100 bp upstream. It contains four of the sequenced tryptic fragments (Fig.
20 II), but do not include the N-terminal amino acid sequence. To extend the sequence beyond the translation start, the mammary gland cDNA library was rescreened with a 118 bases long probe derived from the most 5'proximal part of ABSSL. One clone was isolated that continued a
25 further 328 nucleotides upstream. It matched the N-terminal amino acid sequence, and contained the remaining tryptic fragment. As shown in Fig. II, the cDNA is 2428 nucleotides long and contains 81 bases upstream from the first ATG codon. The polyadenylation signal, AATAAA is
30 located 13 nucleotides upstream from the poly A tail and the termination codon TAG was found at nucleotide 2317 followed by a 3'-untranslated region of 112 bp.

A GC rich region consisting of 16 repeats of 33 bases was found in the 3'-end of the sequence between base 1756 and
35 2283. The nucleotide sequence of the repetition, shown in Fig. III, consists of six identical repetitions surrounded by ten repetitions with different number of





03 FEB 1998

substitutions that have probably occurred after several duplications. The low number of substitutions suggests that these repetitions have appeared late during evolution.

5

C. Tissue distribution of expression

RNA from human lactating mammary gland, pancreas, adipose tissue and from a human hepatoma cellline (HepG2) was
10 analyzed by Northern blotting. The size of the messenger was determined to be approximately 2.5 kb in both lactating mammary gland and pancreas. No signal could be detected in the lanes with RNA extracted from HepG2 or adipose tissue (Fig. IV).

15 Since the mRNA used for the mammary gland library was obtained from a female in her 8th month of pregnancy, it is evident that transcription and probably translation of the BSSL gene is turned on before partus, in agreement with previous findings on BSSL secretion before partus
20 [35]. See Figure IV.

D. Amino acid sequence of BSSL

Assessed by SDS-PAGE the molecular mass has been reported
25 to be 107-125 kDa [8,36] and by analytical ultracentrifugation to be 105 kDa [37]. The enzyme, as deduced from the cDNA, consists of 722 amino acid residues (Fig. II) which, giving a molecular mass of 76.271 Da, indicates that the enzyme contains at least 15-20%
30 carbohydrate. The leader sequence is 23 residues long. A tentative active site serine residue is localized to serine-217 (Fig. V). The sequence around this serine accord with the consensus active site sequence of serine-hydrolases [38]. It has recently been it was proposed that
35 basic residues found close to the active site serine may be involved in the cleavage of ester bonds in acylglycerols by lipases [39]. It is interesting to note that such

03 FEB 1998

21

residues are not present in BSSL. The single tentative N-glycosylation site is localized only seven residues from the serine. The degree of glycosylation [6,16] suggests that the enzyme contains O-linked carbohydrate. There are
5 numerous sites where such glycosylation could have occurred. The amino acid composition based on purified enzyme has shown a high content of proline residues [6]. The amino acid sequence obtained from cDNA confirms this. Moreover, most of the proline residues are localized in
10 the 16 repeats of 11 residues each, constituting the main part of the C-terminal half of the enzyme.

2. Comparison of the enzymes in mammary gland (BSSL) and pancreas (CEH)

15

BSSL of human milk and human pancreas CEH have previously been shown to be similar, if not identical. The present data strongly suggests that the two enzymes are products of the same gene. The nucleotide sequence of the cDNA
20 clones shows that the pancreatic clone ACEH is identical with the mammary gland clone ABSSL from the poly A tail and 996 bases towards the 5'-end, including the sequence coding for the proline rich repeats. Northern blot gave a single band of 2.5 kb in RNA from pancreas and lactating
25 mammary gland (Fig. IV). Genomic Southern blots further support the idea that only one gene codes for BSSL and CEH. The difference in mobility on SDS-PAGE between BSSL and CEH can be explained as a consequence of different glycosylation or differential splicing.

30

The similarity of BSSL to the rat and bovine enzymes (see below) and to results from genomics blots support the possibility that differential splicing cannot account for the mobility difference. Since the C-terminal sequence has
35 not been confirmed on the protein level there is a less likely possibility that CEH may be processed by a proteolytic cleavage in the C-terminal end.

This Sheet Has Been Replaced



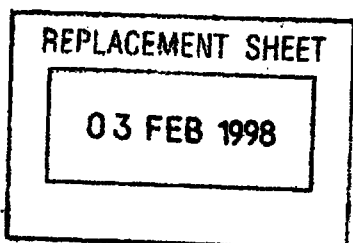
03 FEB 1998

So far as we know pancreatic enzymes that obviously correspond to CEH have often been named after species and the particular substrates used to determine their respective activities; lysophospholipase, cholesteryl
5 esterase, sterol ester hydrolase, non-specific lipase, carboxyl ester lipase and cholesteryl ester hydrolase. Available data are compatible with the view that all these activities described originates in one and the same functional entity [42,43]. This illustrates the broad
10 substrate specificity and the relevance of designating them as non-specific lipases. When the sequence of human BSSL/CEH is compared to the sequence of lysophospholipase from fat pancreas [40] and cholesterol esterase from bovine pancreas [41] extensive similarities are found that
15 extend about 530 residues from the N-terminal (Fig V); but they differ in the part of the molecule where the repeats occur. The rat enzyme has only four repeats and the bovine three. Hence the human enzyme is a considerable longer peptide.

20

Moreover, striking similarities were found between BSSL and a number of typical esterases, e.g. acetyl choline esterases from several species, including man and Drosophila, and carboxyl esterases (Fig. VI). These
25 similarities were restricted to the N-terminal 300 residues of BSSL which includes the tentative active site serine-residue. A similarity to acetyl choline esterase has been predicted from the fact that BSSL is inhibited by typical choline esterase inhibitors [6, 8, 16]. With the
30 possible exception of the rat liver carboxyl esterase [45], none of these similar enzymes has been shown to have the same bile-salt dependency as BSSL; this suggests that the structural basis for this property resides in the C-terminal part of the protein. Moreover, BSSL can
35 efficiently attack emulsified substrates which is not a known characteristic of the similar esterases. For this activity bile salt is a prerequisite.





The predicted sequence for human BSSL was compared with other well characterized mammalian lipases. Apart from the consensus sequence around the active site serine (G-X-S-X-G), no obvious similarities were found [44].

In addition to the similarities with other enzymes, there also significant similarities to one c-AMP dependent protein from Dictyostelium discoideum [46] as well as to thyroglobulin from several species (Fig VI) [47-49]. The similarities between BSSL and thyroglobulin, which comprise the active site region but not the active site itself, indicate that these highly conserved stretches of amino acids are of more generalized importance than merely supporting the enzymatic activity of esterases.

In conclusion, human milk BSSL consists of 722 amino acid residues. Available data strongly indicate that its peptide chain is identical to that of pancreatic CEH, and they are coded for by the same gene. The strongest evidence is that the nucleotide sequences of their 3 - ends and their N-terminal amino acid sequences are identical. The striking homologies found to rat pancreatic lysophospholipase and bovine pancreatic cholesterol esterase support the hypothesis that also these enzymes are functionally identical. However, as it has been suggested, the different molecular sizes found among species are not due to differences merely in glycosylation; instead they reflect a variable number of an eleven amino acid repeat. The similarity of the active site sequence between these esterases suggests that these proteins derive from a common ancestral gene.

With reference to Figures I-VII, the following legends are given.



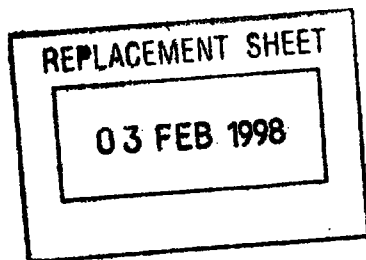


Figure I: Separation of the tryptic digest of BSSL on HPLC

Purified BSSL was treated with trypsin and chromatographed on HPLC as described in Materials and Methods. The indicated peaks were collected and purified further by a rechromatograph and their amino acid sequence determined.

Figure II: The cDNA nucleotide sequence and the deduced amino acid sequence for human bile salt-stimulated lipase:

10

The cDNA is 2428 bases long. The N-terminal 23-codon sequence (nt82-150) starting with an ATG, is interpreted as a leader peptide since the N-terminal amino acid sequence of the mature protein starts at codon 24 (nt 151, Ala). The leader peptide is underlined. The sign * indicates the starting point of an exon. The sign ‡ indicates the starting point of the repetition part.

15

Figure III: The nucleotide sequence of the C-terminal GC-rich repetitions in the bile salt-stimulated lipase:

20

Substitutions are indicated by a *.

Figure IV: Northern blot hybridization

25

Northern blot analysis of total RNA isolated from human lactating mammary gland, pancreas, adipose tissue and a human hepatoma cell line (HepG2). Total RNA (10µg) from lactating mammary gland (lane A), pancreas (lane B), adipose tissue (lane C) and HepG2 (lane D) were electrophoresed in a 1% agarose gel in 40mM MOPS buffer at Ph 7.0 after denaturation of RNA in 1M glyoxal, 50% dimethylsulfoxide and 40mM MOPS. The glyoxalated RNA was then transferred to nitrocellulose paper for hybridization with [³² P] labeled BSSL cDNA (ABSSL).

30

35



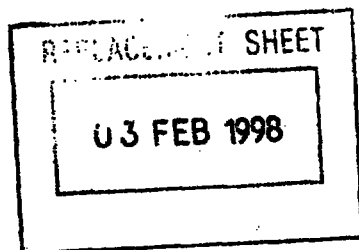


Figure V: Comparison of the deduced amino acid sequence from human milk BSSL, rat pancreatic lysophospholipase (Rat1pl) [40] and bovine pancreatic cholesterol esterase (Bovceh) [41]:

5

The serine residues involved in the active site are indicated by a *, and the # indicates the single possible N-glycosylation signal of the protein. The direct repeats of amino acid sequences are boxed. Matching sequences are denoted in capital letters, matching sequences between two enzymes are denoted in small letters and mismatching with a dot.

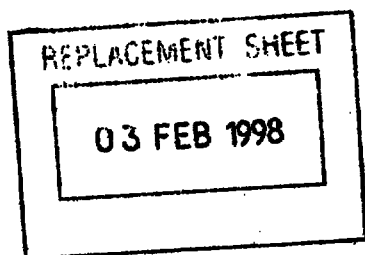
Figure VI: Comparison of the primary structure of BSSL to other esterases, thyroglobuline and to one c-AMP dependent enzyme from Dictyostelium discoideum:

BSSL: bile salt stimulated lipase from human, Cheshum: cholinesterase from fetal human tissue [50], Torpace: acetylcholinesterase from Torpedo marmorata [51], Drosceh: carboxylic ester hydrolase from Drosophila melaogaster [52], Ratlivce: carboxyl esterase from rat liver [53], Drosace: acetylcholinesterase from Drosophila melaogaster [54], Thyrrhum: thyroglobulin from human [49] and Dict.Di: c-AMP dependent enzyme from Dictyostelium discoideum [46]. There are 7 different domains that show similarities between the enzymes. Boxes enclose residues which are identical and small letters in the consensus sequence indicate identical residues in all the enzymes except for one. Dots indicate mismatches. The serine residue involved in the active site is indicated with *. The figure in the right hand corner shows how the domains are oriented.

35

This Sheet Has Been Replaced





26

Figure VII:

gives the amino acid sequence 1 - 722 for the entire protein (one letter code) and indicates exons a, b, c, 5 and d. The sign # indicates the starting point of the repetition part.

This Sheet Has Been Replaced



03 FEB 1998

Table 1. Amino acid sequence of BSSL peptides.

Due to interfering peaks no positive identification of the residue in cycles 1 and 2 of the sequencing could be made
5 in peptide number 26. The peptide numbers refer to the peaks in Fig I.

10

Tryptic fragments

15

TP16:LysValThrGluGluAspPheTyrLys

TP19:GlyIleProPheAlaAlaProThrLys

Tp20:LeuValSerGluPheThrIleThrLys

20

TP24:ThrTyrAlaTyrLeuPheSerHisProSerArg

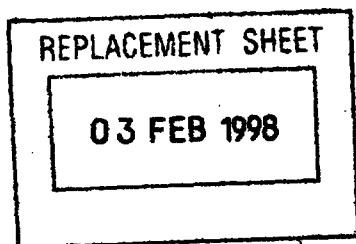
TP26:PheAspValTyrThrGluSerTrpAlaGlnAsp

25

ProSerGlnGluAsnLys

This Sheet Has Been Replaced





28

Table 2: Identification of the exons a, b, c and d
numbered as in Figures II and VII.

5

exon	<u>Location</u>	
10	between nucleotide number	between aminoacid number

a	986-1172	279-341
b	1173-1376	342-409
15 c	1377-1574	410-474
d	1575-2415	475-722
the entire protein	151-2316	1-722
20 the entire protein excluding repetitions	151-1755	1-535
25	-----	

30

35

This Sheet Has Been Replaced



03 FEB 1998

References

1. Bläckberg, L., Ångquist, K.A. & Hernell, O. (1987) FEBS Lett. 217, 37-41.
- 5 2. Hernell, O. (1975) Eur. J. Clin. Invest. 5, 267-272.
3. Hernell, O., Bläckberg, L. & Bernbäck, S. (1988) In Perinatal nutrition (Lindblad, B.S., ed.) Bristol-Myers Nutrition symposia vol. 6, pp. 259-272, Academic Press, New York.
- 10 4. Hernell, O., L. Bläckberg, B. Fredrikzon, and T. Olivecrona. 1981. Bile salt-stimulated lipase in human milk and lipid digestion in the neonatal period. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 465-471.
- 15 5. Bernbäck, S., Bläckberg, L. & Hernell, O. (1990) J. Clin. Invest. J. Clin. Invest. 221-226 (1990)
6. Bläckberg, L. & Hernell, O. (1981) Eur. J. Biochem 116, 221-225.
- 20 7. Hernell, O. & Bläckberg, L. (1982) Pediatr. Res. 16, 882-885.
8. Bläckberg, L. Lombardo, D., Hernell, O., Guy, O. & Olivecrona, T. (1981) FEBS Lett. 136, 284-288.
9. Fredrikzon, B., Hernell, O., Bläckberg, L. & Olivecrona, T. (1978) Pediatr. Res. 12, 1048-1052.
- 25 10. Wang, C.-S., Hartsuck, J.A. & Downs, D. (1988) Biochemistry 27, 4834-4840.
11. Bläckberg, L. & Hernell, O. (1983) FEBS Lett. 157, 337-341.
- 30 12. Björkstén, B., Burman, L.G., deChateau, P., Fredrikzon, B., Gotheffors, L. & Hernell, O. (1980) Br. Med. J. 201, 267-272.
13. Brueton, M.J., Berger, H.M., Brown, G.A., Ablitt, L., Iyangkaran, N. & Wharton B.A. (1978) Gut 19, 95-98.
- 35 14. Murphy, G.M. & Signer, E. (1975) Gut 15, 151-163.
15. Hernell, O., Staggars, J.E. & Carey, M.C. Biochemistry. (1990), 29:2041-2056.

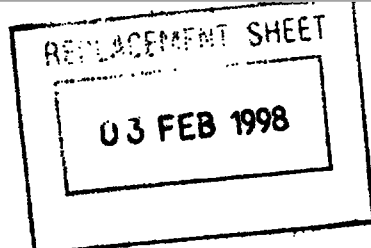
This Sheet Has Been Replaced



03 FEB 1998

16. Lombardo, D., Guy, O. & Figarella, C. (1978) *Biochim. Biophys. Acta* 527, 142-149.
17. Abouakil, N., Rogalska, E., Bonicel, J. & Lombardo, D. (1988) *Biochim. Biophys. Acta* 961, 299-308.
- 5 18. Hernell, O., Bläckberg, L. & Lindberg, T. (1988) in *Textbook of Gastroenterology and Nutrition in infancy* (Lebenthal, E., ed) pp. 209-217, Raven Press, New York.
19. Wang, C.-S. (1988) *Biochem. Biophys. Res. Com.* 155, 950-955.
- 10 20. Chirgwin, J.M. Przybyla, A.E., MacDonald R.J. & Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
21. Aviv, H. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 69, 5201-5205.
- 15 22. Bailey, J.M. & Davidson, N. (1976) *Anal. Biochem.* 70, 75-85.
23. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982): *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- 20 24. Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263-269.
25. Maniatis T., Haldison, R.C., Lacy, E., Lauer, J., O'Connell, C. & Qvon, D. (1978) *Cell* 15, 687-701.
26. Young, R.A. & Davis, R.W. (1983) *Science* 222, 778-782.
- 25 27. Young, R.A. & Davis, R.W. (1983) *Proc. Natl. Acad. Sci, USA* 80, 1194-1198.
28. McMaster, G.K. & Charmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 48 353-48 358.
29. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
- 30 30. Feinberg, A.P. & Vogelstein, B. (1983) *Analyt. Biochem.* 132, 6-13.
31. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74. 5463-5467.
- 35 32. Staden, R. (1984) *Nucl, Acids Res.* 12, 551-567.
33. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.





34. Wang, C.-S. & Johnson, K. (1983) *Anal. Biochem.* 133, 457-461.
35. Hamosh, M. (1986) in *Human Milk Infant Nutrition and Health* (Howell, R.R. Morriss, F.H. & Pickering, L.K. eds) pp. 66-97, Charles, C. Thomas, Springfield.
- 5 36. Wang, C.-S. (1980) *Anal. Biochem.* 105, 398-402.
37. Wang, C.-S & Lee, D.M. (1985) *J. Lipid. Res.* 26, 824-830.
38. Brenner, S. (1988) *Nature* 334, 528-530.
- 10 39. Yang, C.-Y., Gu, Z.-W., Yang, H.-H., Rohde, M.F., Gotto, Jr., A.M. & Pownall, H.J. (1989) *J. Biol. Chem.* 265, 16822-16827.
40. Han, J.H., Stratowa, C., & Rutter, W.J. (1987) *Biochemistry* 26, 1617-1625.
- 15 41. Kyger, E., Wiegand, K., & Lange, L. (1989) *Biochim. Biophys. Res. Com* 164, 1302-1309.
42. Bläckberg, L. (1981) *Fat digestion in newborn infant. Umeå University Medical Dissertations New Series No 71.*
- 20 43. Rudd, E.A. & Brockman, H.L. (1984) in *Lipases* (Borgström, B & Brockman, H.L. eds) pp. 184-204, Elsevier, Amsterdam.
44. Mickel, S., Weidenbach, F., Swarovsky, B., LaForge, S. & Scheele, G. (1989) *J. Biol. Chem.* 264, 12895-12901.
- 25 45. Cammulli, E.D., Linke, M.J., Brockman, H.L. & Hui, D.Y. (1989) *Biochim. Biophys. Acta* 1005, 177-182.
46. Mann, S.K.O. & Firtel, R.A. (1987) *Mol. Cell Biol.* 7, 458-469.
- 30 47. Mercken, L., Simons, M.J., Swillens, S., Masser, M. & Vassart, G. (1985) *Nature* 316, 647-651.
48. Lauro, R., Obici, S. Condliffe, D., Ursini, V.M., Musti, A., Moscatelli, C. & Avvedimento, V.E. (1985) *Eur. J. Biochem.* 148, 7-11.
- 35 49. Malthiery, Y. & Lissitzky, S. (1987) *Eur. J. Biochem.* 165, 491-498.
50. Prody, C., Zevin-Sonkin, D., Gnatt, A., Goldberg, O.,



03 FEB 1998

- & Soreq, H. (1987) Proc. Natl. Acad. Sci. USA 84, 3555-3559.
51. Sikorav, J-L., Krejci, E. & Massoulie', J. (1987) EMBO J. 6, 1865-1873.
- 5 52. Oakeshott, J.G. Collet, C., Phillis, R.W., Nielsen, K.M., Russel, R.J., Cambers, G.K., Ross, V. & Richmond, R.C. (1987) Proc. Natl. Acad. Sci, USA 84, 3359-3363.
53. Long, R., Satho, H., Martin, B., Kimura, S., Gonzalez, F. & Pohl, L. (1988) Biochem. Biophys. Res. Comm. 156, 866-873.
- 10 54. Hall, L.M.C. & Spierer, P. (1986) EMBO J. 5, 2949-2954.
55. Sjöberg, S., Carlsson, P., Enerbäck, S., & Bjursell, G. (1989) CABIOS 5, 41-46.
- 15 56. EP-A-317355.
57. Hernell, O., and T. Olivecrona. 1974. Human milk lipases II. Bile salt-stimulated lipase. Biochim. Biophys. Acta 369:234-244.
- 20 58. Hernell, O., L. Bläckberg, and T. Olivecrona. 1981. Human milk lipases. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 347-354.
- 25 59. Atkinson, S.A., M.H. Bryan, and G.H. Andersson. 1981. Human milk feeding in premature infants: protein, fat and carbohydrate balances in the first two weeks of life. J.Pediatr. 99:617-624.
- 30 60. Chappell, J.E., M.T. Clandinin, C. Kearney-Volpe, B. Reichman and P.W. Swyer. 1986. Fatty acid balance studies in premature infants fed human milk or formula: effect of calcium supplementation. J. Pediatr. 108:438-447.
- 35

This Sheet Has Been Replaced



REPLACEMENT SHEET

03 FEB 1998

33

61. Williamson, S., E. Finucane, H. Ellis, and H.R. Gamsu. 1978. Effect of heat treatment of human milk on absorption of nitrogen, fat, sodium, calcium and phosphorus by preterm infants. Arch. Dis. Childhood 53:555-563.
- 5

10

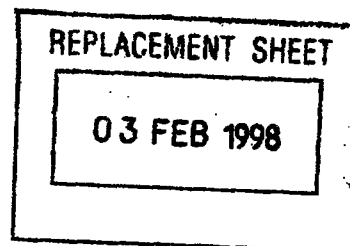
15

This Sheet Has Been Replaced





34

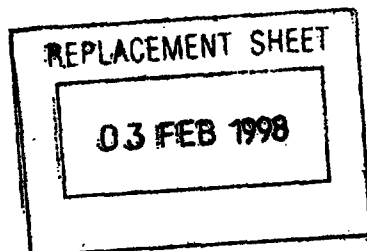


What we claim is:

- 5 1. A protein as indicated in Figure VII from position 1 to position 722.
2. A protein as indicated in Figure VII from position 1 to position 535.
- 10 3. A protein as indicated in Figure VII from position 1 to position 278.
4. A protein as indicated in Figure VII from position 1 to position 341.
- 15 5. A protein as indicated in Figure VII from position 1 to position 409.
6. A protein as indicated in Figure VII from position 1 to position 474.
- 20 7. A protein as indicated in Figure VII from position 1 to position 278, position 279 to position 341, position 279 to position 409, position 279 to position 474, position 342 to position 409, position 342 to position 474, or position 536 to position 722.
- 25 8. A protein according to any of claims 1-7 in combination with one or more of the repeats according to Figure V.
- 30 9. A protein according to any of claims 1-8, possessing methionine as additional N-terminal amino acid.
- 35 10. A protein according to any of claims 1-9, exhibiting one or several of the critical functions of naturally occurring bile salt-stimulated lipase.

This Sheet Has Been Replaced



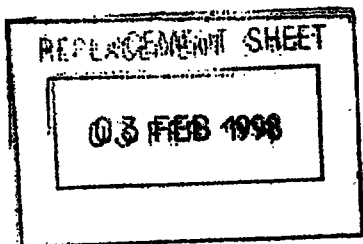


35

11. A functionally equivalent variant or mutant of a protein according to any of claims 1-9.
12. A DNA sequence coding for a protein with the amino acid sequence according to claim 1.
13. A DNA sequence coding for a protein with the amino acid sequence according to claim 2.
14. A DNA sequence coding for a protein with the amino acid sequence according to any of claims 3, 4, 5, 6, 7, 8, 9, 10 and 11.
15. A DNA sequence defined by the following nucleotide numbers in Figure II:
- 151-2316
- 151-1755
- 151-985
- 151-1172
- 151-1376
- 151-1574
- 986-1172
- 986-1376
- 986-1574
- 1173-1376
- 1173-1574
16. A vector comprising a DNA sequence coding for a protein according to claims 1 - 11
17. A vector comprising a DNA sequence according to claims 12 - 15.
18. A host organism transformed with a vector according to claims 16 or 17.

This Sheet Has Been Replaced





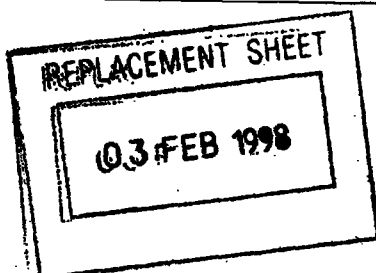
36



19. A process for the production of a protein according to claims 1 - 11 by growing a host organism containing a vector according to claims 16 or 17, and isolating the protein.
20. A pharmaceutical composition comprising a protein according to any of claims 1 - 11.
21. A pharmaceutical composition comprising a protein according to any of claims 1-11 in combination with a lipase or in combination with preparations containing lipases.
22. The use of a protein according to any of claims 1 - 11, for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency.
23. The use of a protein according to any of claims 1 - 11 for the manufacture of a medicament for the treatment of cystic fibrosis.
24. The use of a protein according to any of claims 1 - 11 as a supplement to an infant food formulation.
25. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of chronic pancreatitis.
26. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption.
27. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins.

This Sheet Has Been Replaced





37



28. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons.
- 5
29. The use, according to claims 22-28 of a protein according to any of claims 1-11 in combination with a lipase or lipases or in combination with preparations containing a lipase or lipases.
- 10
30. An infant food formulation, supplemented with a protein according to any of claims 1 - 11.
- 15
31. A protein according to claims 1 - 11 in substantially pure form.
- 20
32. A protein according to claims 1 - 11 in isolated form.
33. A DNA sequence according to claims 12 - 15 in substantially pure form.
- 25
34. A DNA sequence according to claims 12 - 15 in isolated form.
- 35
35. A process for the preparation of a pharmaceutical composition according to claims 20 or 21 by incorporating a protein according to claims 1 - 11, 31, or 32 in a pharmacologically acceptable carrier.
- 30
36. A process according to claim 35 for the preparation of a pharmaceutical composition for oral administration.
- 35
37. A process for the preparation of an infant food formulation according to claim 30, by supplementing an infant food formulation with a protein according to claims 1 -11, 31, or 32.

This Sheet Has Been Replaced



REPLACEMENT SHEET
03 FEB 1998

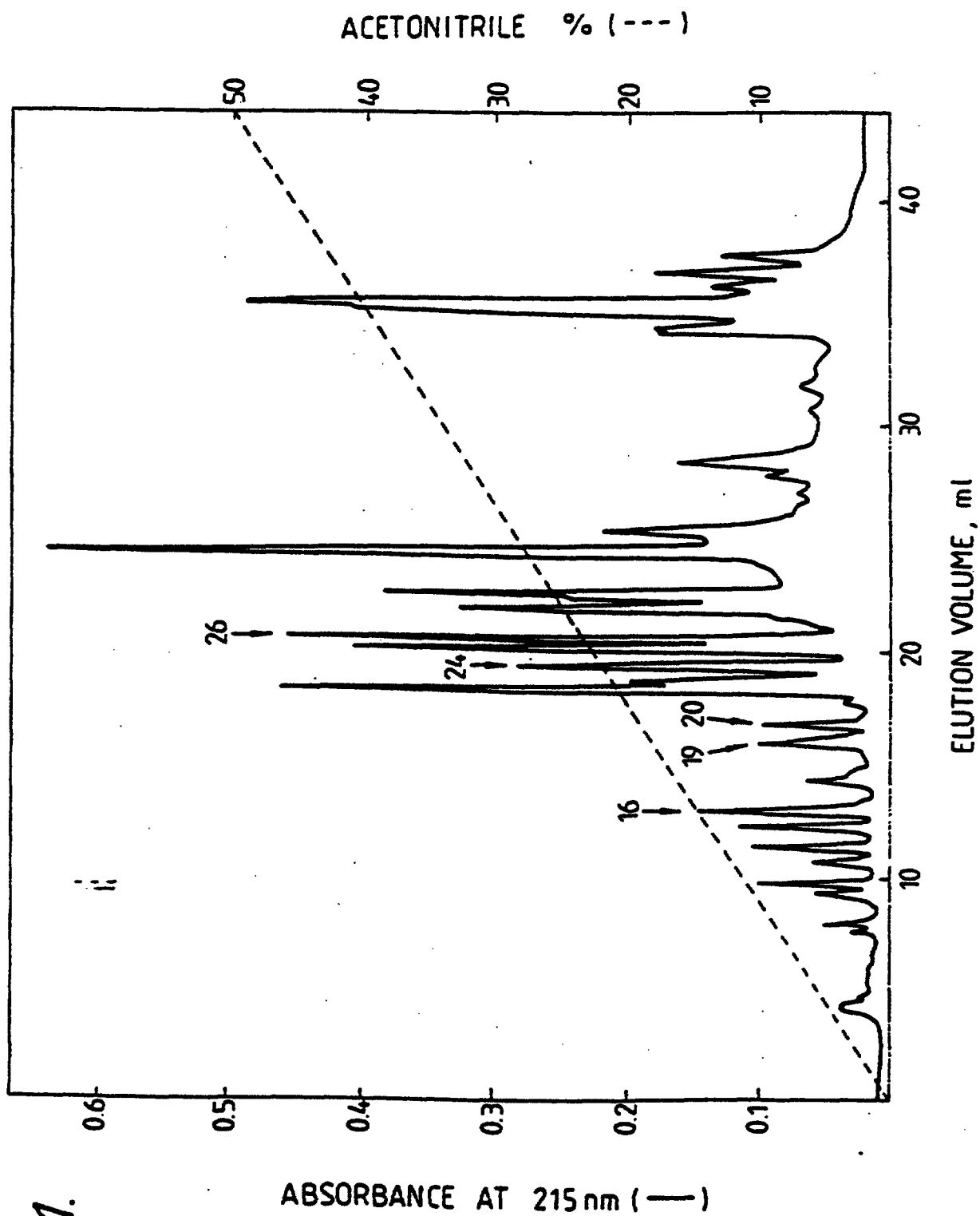
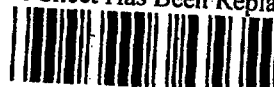
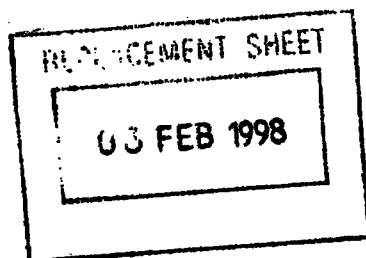


Fig. 1.

This Sheet Has Been Replaced





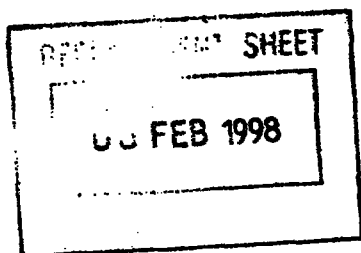
2,

This Sheet Has Been Replaced



Fig.2 (1/4)

ACCTTCIGTA TCAGTTAAGT GTCAAGATGG AAGGAACAGC AGTCTCAAGA TAATGCAAAG	60
AGTTTATTCA TCCAGAGGCT G ATG CTC ACC ATG GGG CGC CTG CAA CTG GTT	111
Met Leu Thr Met Gly Arg Leu Gln Leu Val	
1 5 10	
GTG TTG GGC CTC ACC TGC TGC TGG GCA GTG GCG AGT GGC GCG AAG CTG	159
Val Leu Gly Leu Thr Cys Cys Trp Ala Val Ala Ser Ala Ala Lys Leu	
15 20 25	
GGC GGC GTG TAC ACA GAA GGT GGG TTC GTG GAA GGC GTC AAT AAG AAG	207
Gly Ala Val Tyr Thr Glu Gly Gly Phe Val Glu Gly Val Asn Lys Lys	
30 35 40	
CTC GGC CTC CTG GGT GAC TCT GTG GAC ATC TTC AAG GGC ATC CCC TTC	255
Leu Gly Leu Leu Gly Asp Ser Val Asp Ile Phe Lys Gly Ile Pro Phe	
45 50 55	
GCA GCT CCC ACC AAG GGC CTG GAA AAT OCT CAG CCA CAT OCT GGC TGG	303
Ala Ala Pro Thr Lys Ala Leu Glu Asn Pro Gln Pro His Pro Gly Trp	
60 65 70	
CAA GGG ACC CTG AAG GCC AAG AAC TTC AAG AAG AGA TGC CTG CAG GCC	351
Gln Gly Thr Leu Lys Ala Lys Asn Phe Lys Lys Arg Cys Leu Gln Ala	
75 80 85 90	
ACC ATC ACC CAG GAC AGC ACC TAC GGG GAT GAA GAC TGC CTG TAC CTC	399
Thr Ile Thr Gln Asp Ser Thr Tyr Gly Asp Glu Asp Cys Leu Tyr Leu	
95 100 105	
AAC ATT TGG GTG CCC CAG GGC AGG AAG CAA GTC TCC CGG GAC CTG CCC	447
Asn Ile Trp Val Pro Gln Gly Arg Lys Gln Val Ser Arg Asp Leu Pro	
110 115 120	
GTT ATG ATC TGG ATC TAT GGA GGC GCC TTC CTC ATG GGG TCC GGC CAT	495
Val Met Ile Trp Ile Tyr Gly Gly Ala Phe Leu Met Gly Ser Gly His	
125 130 135	
GGG GCC AAC TTC CTC AAC AAC TAC CTG TAT GAC GGC GAG GAG ATC GCC	543
Gly Ala Asn Phe Leu Asn Asn Tyr Leu Tyr Asp Gly Glu Glu Ile Ala	
140 145 150	
ACA CGC GGA AAC GTC ATC GTG GTC ACC TTC AAC TAC CGT GTC GGC CCC	591
Thr Arg Gly Asn Val Ile Val Val Thr Phe Asn Tyr Arg Val Gly Pro	
155 160 165 170	
CIT GGG TTC CTC AGC ACT GGG GAC GCC AAT CTG CCA GGT AAC TAT GGC	639
Leu Gly Phe Leu Ser Thr Gly Asp Ala Asn Leu Pro Gly Asn Tyr Gly	
175 180 185	



This Sheet Has Been Replaced

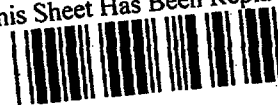
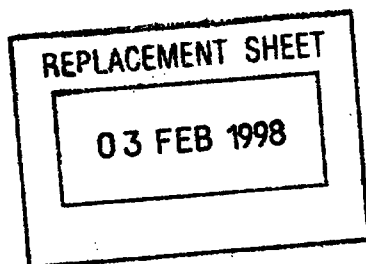


Fig. 2(2/4)

CTT CCG GAT CAG CAC ATG GGC ATT GCT TGG GTG AAG AGG AAT ATC GCG Leu Arg Asp Gln His Met Ala Ile Ala Trp Val Lys Arg Asn Ile Ala 190 195 200	627
GCC TTC GGG GGG GAC CCC AAC AAC ATC ACG CTC TTC GGG GAG TCT GCT Ala Phe Gly Gly Asp Pro Asn Asn Ile Thr Leu Phe Gly Glu Ser Ala 205 210 215	735
GGA GGT GCC AGC GTC TCT CTG CAG ACC CTC TCC CCC TAC AAC AAG GGC Gly Gly Ala Ser Val Ser Leu Gln Thr Leu Ser Pro Tyr Asn Lys Gly 220 225 230	783
CTC ATC CCG CGA GGC ATC AGC CAG AGC GGC GTG GCC CTG AGT CCC TGG Leu Ile Arg Arg Ala Ile Ser Gln Ser Gly Val Ala Leu Ser Pro Trp 235 240 245 250	831
GTC ATC CAG AAA AAC CCA CTC TTC TGG GCC AAA AAG GTG GCT GAG AAG Val Ile Gln Lys Asn Pro Leu Phe Trp Ala Lys Lys Val Ala Glu Lys 255 260 265	879
GTG GGT TGC CCT GTG GGT GAT GGC GCC AGG ATG GCC CAG TGT CTG AAG Val Gly Cys Pro Val Gly Asp Ala Ala Arg Met Ala Gln Cys Leu Lys 270 275 280	927
GTT ACT GAT CCC CGA GGC CTG ACG CTG GCC TAT AAG GTG CCG CTG GCA Val Thr Asp Pro Arg Ala Leu Thr Leu Ala Tyr Lys Val Pro Leu Ala 285 290 295	975
*	
GGC CTG GAG TAC CCC ATG CTG CAC TAT GTG GGC TTC GTC CCT GTC ATT Gly Leu Glu Tyr Pro Met Leu His Tyr Val Gly Phe Val Pro Val Ile 300 305 310	1033
GAT GGA GAC TTC ATC CCC GCT GAC CCG ATC AAC CTG TAC GCC AAC GCC Asp Gly Asp Phe Ile Pro Ala Asp Pro Ile Asn Leu Tyr Ala Asn Ala 315 320 325 330	1071
GCC GAC ATC GAC TAT ATA GCA GGC ACC AAC AAC ATG GAC GGC CAC ATC Ala Asp Ile Asp Tyr Ile Ala Gly Thr Asn Asn Met Asp Gly His Ile 335 340 345	1119
TTC GCC AGC ATC GAC ATG CCT GGC ATC AAC AAG GGC AAC AAG AAA GTC Phe Ala Ser Ile Asp Met Pro Ala Ile Asn Lys Gly Asn Lys Lys Val 350 355 360	1167
*	
ACG GAG GAG GAC TTC TAC AAG CTG GTC AGT GAG TTC ACA ATC ACC AAG Thr Glu Glu Asp Phe Tyr Lys Leu Val Ser Glu Phe Thr Ile Thr Lys 365 370 375	1215
GGG CTC AGA GGC GCC AAG ACG ACC TTT GAT GTC TAC ACC GAG TCC TGG Gly Leu Arg Gly Ala Lys Thr Thr Phe Asp Val Tyr Thr Glu Ser Trp 380 385 390	1263



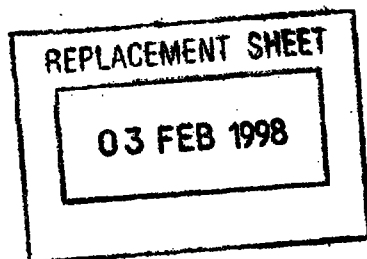
4

This Sheet Has Been Replaced



Fig.2(3/4)

GCC CAG GAC OCA TCC CAG GAG AAT AAG AAG AAG ACT GTG GTG GAC TTT Ala Gln Asp Pro Ser Gln Glu Asn Lys Lys Lys Thr Val Val Asp Phe 395 400 405 410	1311
GAG ACC GAT GTC CTC TTC CTG GTG CCC ACC GAG ATT GCC CTA GCC CAG Glu Thr Asp Val Leu Phe Leu Val Pro Thr Glu Ile Ala Leu Ala Gln 415 420 425	1359
*	
CAC AGA GCC AAT GCC AAG AGT GCC AAG ACC TAC GCC TAC CTG TTT TCC His Arg Ala Asn Ala Lys Ser Ala Lys Thr Tyr Ala Tyr Leu Phe Ser 430 435 440	1407
CAT CCC TCT CGG ATG CCC GTC TAC CCC AAA TGG GTG GGG GCC GAC CAT His Pro Ser Arg Met Pro Val Tyr Pro Lys Trp Val Gly Ala Asp His 445 450 455	1455
GCA GAT GAC ATT CAG TAC GTT TTC GGG AAG CCC TTC GCC ACC CCC ACG Ala Asp Asp Ile Gln Tyr Val Phe Gly Lys Pro Phe Ala Thr Pro Thr 460 465 470	1503
GGC TAC CGG CCC CAA GAC AGG ACA GTC TCT AAG GCC ATG ATC GCC TAC Gly Tyr Arg Pro Gln Asp Arg Thr Val Ser Lys Ala Met Ile Ala Tyr 475 480 485 490	1551
*	
TGG ACC AAC TTT GCC AAA ACA GGG GAC CCC AAC ATG GCC GAC TCG GCT Trp Thr Asn Phe Ala Lys Thr Gly Asp Pro Asn Met Gly Asp Ser Ala 495 500 505	1599
GTG CCC ACA CAC TGG GAA CCC TAC ACT ACG GAA AAC AGC GCC TAC CTG Val Pro Thr His Trp Glu Pro Tyr Thr Thr Glu Asn Ser Gly Tyr Leu 510 515 520	1647
GAG ATC ACC AAG AAG ATG GGC AGC AGC TCC ATG AAG CGG AGC CTG AGA Glu Ile Thr Lys Lys Met Gly Ser Ser Ser Met Lys Arg Ser Leu Arg 525 530 535	1695
ACC AAC TTC CTG CGC TAC TGG ACC CTC ACC TAT CTG GCG CTG CCC ACA Thr Asn Phe Leu Arg Tyr Trp Thr Leu Thr Tyr Leu Ala Leu Pro Thr 540 545 550	1743
*	
GTG ACC GAC CAG GAG GCC ACC CCT GTG CCC CCC ACA GGG GAC TCC GAG Val Thr Asp Gln Glu Ala Thr Pro Val Pro Pro Thr Gly Asp Ser Glu 555 560 565 570	1791
GCC ACT CCC GTG CCC CCC ACG GGT GAC TCC GAG ACC GCC CCC GTG CCG Ala Thr Pro Val Pro Pro Thr Gly Asp Ser Glu Thr Ala Pro Val Pro 575 580 585	1839
CCC ACG GGT GAC TCC GGG GGC CCC CCC GTG CCG CCC ACG GGT GAC TCC Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser 590 595 600	1887



5.

This Sheet Has Been Replaced



Fig.2 (4/4)

GGG GGC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC GTG Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val 605 610 615	1925
CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC GTG CCG CCC ACG GGT GAC Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp 620 625 630	1983
TCC GGG GGC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro 635 640 645 650	2031
GTG CCG CCC ACG GGT GAC TCC GGC GGC CCC CCC GTG CCG CCC ACG GGT Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly 655 660 665	2079
GAC GGC GGG CCC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGC GGC CCC Asp Ala Gly Pro Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro 670 675 680	2127
CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC GTG ACC CCC ACG Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Thr Pro Thr 685 690 695	2175
GGT GAC TCC GAG ACC GGC CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC Gly Asp Ser Glu Thr Ala Pro Val Pro Pro Thr Gly Asp Ser Gly Ala 700- 705 710	2223
CCC CCT GTG CCC CCC ACG GGT GAC TCT GAG GCT GGC CCT GTG CCC CCC Pro Pro Val Pro Pro Thr Gly Asp Ser Glu Ala Ala Pro Val Pro Pro 715 720 725 730	2271
ACA GAT GAC TCC AAG GAA GCT CAG ATG CCT GCA GTC ATT AGG TTT Thr Asp Asp Ser Lys Glu Ala Gln Met Pro Ala Val Ile Arg Phe 735 740 745	2316
TAGGTCCCA TGAGCCTTGG TATCAAGAGG CCACAAGAGT GGGACCCAG GGGCTCCCT	2376
CCCATCTTGA GCTCTTCTG AATAAAGCCT CATACCCCTA AAAAAAAAAA AA	2428

REPLACEMENT SHEET

03 FEB 1998

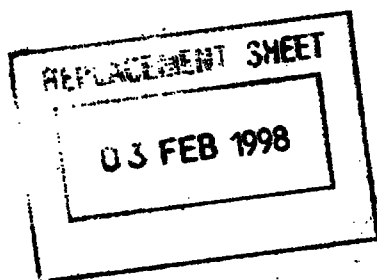
6.

Fig. 3.

		Number of substitutions
GAGGCCACCCCTGTGCCCCCAGGGGACTCC	1	6
GAGGCCACTCCCGTGCCCCCAGGGTGACTCC	2	4
GAGACCGCCCCCGTGCCGCCACGGGTGACTCC	3	3
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	4	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	5	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	6	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	7	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	8	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	9	0
GGCGCCCCCCCCCGTGCCGCCACGGGTGACGCC	10	2
GGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	11	1
GGCGCCCCCCCCCGTGCCGCCACGGGTGACTCC	12	1
GGGGCCCCCCCCCGTGACCCCCACGGGTGACTCC	13	2
GAGACCGCCCCCGTGCCGCCACGGGTGACTCC	14	3
GGGGCCCCCCTGTGCCCCCAGGGTGACTCT	15	3
GAGGCTGCCCCCTGTGCCCCCAGATGACTCC	16	7
GGGGCCCCCCCC GTGCCGCCACGGGTGACTCC	Consensus	

This Sheet Has Been Replaced





7.

Fig. 4.

A B C D Kb



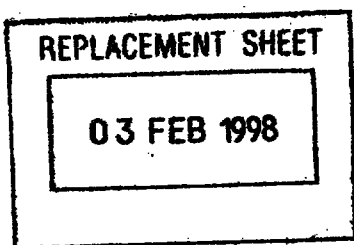
03.0

02.0

01.5

This Sheet Has Been Replaced





8



Fig.5 (1/3)

50
Bssl MLTMGRLQLVVLGLTCCWAVASAAKLGAVYTEGGFVEGVNKKLGLLG.DS
Rat1pl ...MGRLEVLFLGLTCCLAACAALGALYTEGGFVEGVNKKLSLLGGDS
BovcehMLGASRLGPSPGCLAVASAAKLGSVYTEGGFVEGVNKKLSLFG.DS
Consensusgrl....lgltcclaaaaAAKLGavYTEGGFVEGVNKKLSLlG.DS

100
Bssl VDIFKGIPFAAPTCALENPQHPGWQGTLLKAKNFKKRCLQATITQDSTYG
Rat1pl VDIFKGIPFA.TAKTLENPQRHPGWQGTLLKATQFKKRCLQATITQDDTYG
Bovech VDIFKGIPFAAAPKALEKPKRHPGWQGTLLKAKSFKKRCLQATLTQDSTYG
Consensus VDIFKGIPFAa..KaLEnPqrHPGWQGTLLKak.FKKRCLQATiTQDsTYG

150
Bssl DEDCLYLNIWVPQGRKQVSRDLPVMIWIYGGAFLMGSGBGANFLNNYLYD
Rat1pl QEDCLYLNIWVPQGRKQVSHDLPVMVWIYGGAFLMGSGQGANFLKNYLYD
Bovceh NEDCLYLNIWVPQGRKEVSHDLPVMIWIYGGAFLMGASQGANFLSNYLYD
Consensus .EDCLYLNIWVPQGRKqVShDLPVMIWIYGGAFLMGsqqGANFL.NYLYD

200
Bssl GKKIATRGNVIVVTFNRYRGPLGFLSTGDANLPGNYGLRDQHMAIAWVKR
Rat1pl GKKIATRANVIVVTFNRYRGPLGFLSTGDANLPGNFGLRDQHMAIAWVKR
Bovech GKKIATRGNVIVVTFNRYRGPLGFLSTGDSNLPNGYGLWDQHMAIAWVKR
Consensus GKKIATRgNVIVVTFNRYRGPLGFLSTGDaNLPNGyGLrDQHMAIAWVKR

250
Bssl NIAAFGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSP
Rat1pl NIAAFGGDPNNITIFGESAGGAIVSLQTLSPYNKGLIRRAISQSGVALSP
Bovceh NIEAFGGDPDNITLFGESAGGASVSLQTLSPYNKGLIKRGISQSGVGLCP
Consensus NIaAFGCDPDNITLFGESAGGASVSLQTLSPYNKGLIrRaISQSGVaLsP

300
Bssl WVIQKNPLFWAKKVAKKVGCVPVGDAAARMAQCLKVTDPRALTLAYKVPLAG
Rat1pl WAIQENPLFWAKTIKKVGCPTEDTAKMAGCLKITDPRALTLAYRLPLKS
Bovech WAIQQDPLFWAKRIKKVGCVPDDTSKMAGCAKITDPRALTLAYKLPLGS
Consensus WaIQ.nPLFWAK.iAKKVGCpV.DtakMAgClKiTDPRALTLAYklPL.s

350
Bssl LEYPM LHYVGFVPVIDGDFIPADPINLYANAADIDYIAGTNNMDGHIFAS
Rat1pl QEYPIVEYLA FIPVVDGDFIPDDPINLYDNAADIDYLAGINDMDGHLFAT
Bovceh TEYPKLHYLSFVPVIDGDFIPDDPVNLYANAADV DYIAGTNDMDGHLFVG
Consensus .EYP.lHYl.FvPVIdGDFIPdDPiNLYaNAADiDYIAGtNdMDGHlFa.

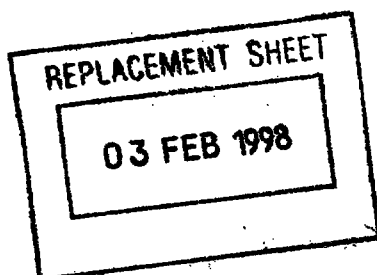


Fig. 5 (2/3)

400
Bssl IDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGAKTTFDVYTESWAQDPS
Ratlpl VDVPAIDKAKQDVTEEDFYRLVSGHTVAKGLKGTQATFDIYTESWAQDPS
Bovech MDVPAINSNKQDVTEEDFYKLVSGTLVTKGLRGANATYEVYTEPWAQDSS

Consensus .DvPAInk.kqdVTEEDFYkLVsg.TvtKGLrGa.aTfdvYTESWAQDps

450
Bssl QENKKKTVVDFETDVLFLVPTEIALAQHRANAKSAKTYAYLFSHPSRMPV
Ratlpl QENMKKTVVAFETDILFLIPTEMALAQHRAHAKSAKTYSYLFShPSRMPi
Bovceh QETRKKTMVDLETDILFLIPTKIAVAQHAKSHAKSANTYTYLFSQPSRMPi

Consensus QEn.KKtVvdfETDiLFLiPteIAlAQHrahAKSAkTY.YLFSHPSRMPi

500
Bssl YPKWVGADHADDIQQYVFGKPFATPTGYRPQDRTVSKAMIAWNTNFAKTGD
Ratlpl YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAWNTNFAKSGD
Bovech YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAWNTNFARTGD

Consensus YPKWmGADHADDlQYVFGKPFATPlGYRaQDRTVSKAMIAWNTNFAktGD

550
Bssl PNMGDSAVPTHWEPTYTENS GYLEITKKMGSSSMKRS LR TNFLRYWTLTY
Ratlpl PNMGNSPVPTHWYPYTMENGN YLDINKKITSTSMKEHLREKFLKFWAVTF
Bovceh PNTGHSTVPANWDPYTLEDDNYLEINKQMDSNSMKLRTL TNYLQFWTQTY

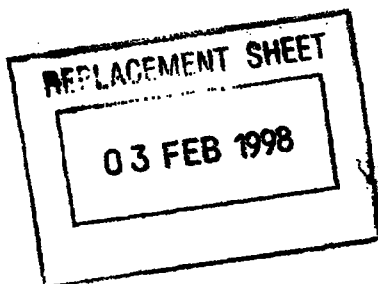
Consensus PnNg.S.VPthW.PYT.Kn.nyleInKkm.S.SMK.hLRtnfL.fWt.Ty

596
Bssl LALPTVTDQ EATPVPTGDS EATPVPTGDS ETAPVPTGDS GAPP
Ratlpl EMLPTV... VGDHTPPKDDS EAAPVPTDDS DGGPVPTDDS QTTP
Bovech QALPTV TSA GASLLPPEDNS QASPVPPADNS GAPTEPSAGDS

Consensus .aLPTVt....a...PP.dds.eA.PVPptddS....pvPptgDS....p

642
Bssl VPPTGDS GAPPVPPTGDS GAPPVPPTGDS GAPPVPPTGDS GAPPVP
Ratlpl VPPTDNS QA.....
Bovceh

Consensuss..a.....



10,

Fig.5 (3/3)

	ER9				
Bssl	PTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GPPPVTPTGDS	GAPPVPPTG
Rar1pl					
Bovech					
Consensus					

	735				
Bssl	DS	GAPPVTPTGDS	ETAPVPPTGDS	GAPPVPPTGDS	EAAPVPPTDDS
Rat1pl	..				GDS
Bovceh	..				
Consensus	ds				

Bssl	KE.AQMPAVIRF
Rat1pl	VE.AQMPGPIGF
Bovech	.EVAQMPVVIGF
Consensus	.e.AQMP.vIgF

This Sheet Has Been Replaced





Fig. 6(1/3)

1

(76-90)
BSSL
(116-130)
ChesHum
(114-128)
Torpace
(101-115)
Drosceh
(103-117)
Ratlivce
(224-238)
Drosace
(2257-2270)
ThryHum
(90-114)
Dict.Di

```

G D E D C L Y L N I W V P Q G
L S E D C L Y L N V W I P A P
M S E D C L Y L N I W V P S P
G E E D C L T V S V Y K P K N
F S E D C L Y L N I Y S P A D
V S E D C L Y I N V W A P A K
V S E D C L Y L N V F I P Q .
A Q K C N L G P G V C S P M G
  
```

Consensus . s e d c l y l n v w . P . .

2

(95-113)
BSSL
(134-150)
ChesHum
(132-150)
Torpace
(118-136)
Drosceh
(121-138)
Ratlivce
(273-291)
Drosace
(2270-2291)
ThryHum

```

S R D L P V M I W I Y G G A F L M G S
P K D A T V L I W I Y G G G F Q T G T
P K S A T V M L W I Y G G G F Y S G S
R N S F P V V A H I H G G A F M F G A
N S R L P V M V W I H G G G L I I G G
T N G L P I L I W I Y G G G F M T G S
A P N A S V L V F F H N T M D R E E S
  
```

Dict.Di

Consensus . . . l p v . . w i y g g g f . . g s

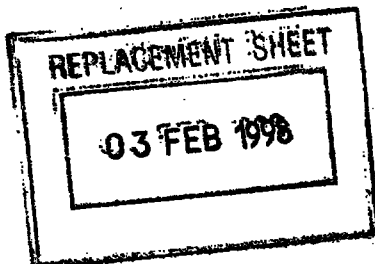
3

(135-156)
BSSL
(166-187)
ChesHum
(164-185)
Torpace
(150-171)
Drosceh
(153-174)
Ratlivce
(307-328)
Drosace
(2307-2328)
ThryHum
(118-136)
Dict.Di

```

N V I V V T F N Y R V G P L G F L S T G D A
R V I V V S M N Y R V G A L G F L A L P G N
E V V L V S L S Y R V G A F G F L A L H G S
K F I L V K I S Y R L G P L G F V S T G D E
N V V V V T I Q Y R L G F G G L F S T G D E
N V I V A S F Q Y R V G A F G F L H L A P E
N L I V V T A S Y R V G V F G F L S S G S G
S V I V V T I N Y R L G I L G L M G T . . .
  
```

Consensus n v i v v t f n Y R v g . . G f l s t g d .



12.

Fig.6 (2/3)

4

(157-182)
BSSL
(189-214)
ChesHum
(187-212)
Torpace
(172-197)
Drosceh
(175-200)
Ratlivce
(336-361)
Drosace
(2329-2354)
ThryHum

N	L	P	G	N	Y	G	L	R	D	Q	H	M	A	I	A	W	V	K	R	N	I	A	A	F	G
E	A	P	G	N	M	G	L	F	D	Q	Q	L	A	L	Q	W	V	Q	K	N	I	A	A	F	G
E	A	P	G	N	M	G	L	L	D	Q	R	M	A	L	Q	W	V	H	D	N	I	Q	F	F	G
D	L	P	G	N	Y	G	L	K	D	Q	R	L	A	L	K	W	I	K	Q	N	I	A	S	F	G
H	S	R	G	N	W	A	H	L	D	Q	L	A	A	L	R	W	V	Q	D	N	I	A	N	F	G
E	A	P	G	N	V	G	L	W	D	Q	A	L	A	I	R	W	L	K	D	N	A	H	A	F	G
E	V	S	G	N	N	G	L	L	D	Q	V	A	A	L	T	W	V	O	T	H	I	R	G	F	G

Dict.Di

.....

Consensus

e l p G N w q l l D Q . . A l . W v . d n i a a F G

5

(183-207)
BSSL
(215-239)
ChesHum
(213-237)
Torpace--
(198-222)
Drosceh
(201-225)
Ratlivce
(362-386)
Drosace
(2354-2379)
ThryHum
(137-149)
Dict.Di

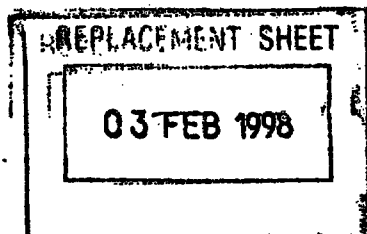
G	D	P	N	N	I	T	L	F	G	E	S	A	G	G	A	S	V	S	L	Q	T	L	S	P
G	N	P	K	S	V	T	L	F	G	E	S	A	G	A	A	S	V	S	L	H	L	L	S	P
G	D	P	K	T	V	T	L	F	G	E	S	A	G	G	A	S	V	G	M	H	I	L	S	P
G	E	P	O	N	V	L	L	V	G	H	S	A	G	G	A	S	V	H	L	Q	M	L	R	E
G	N	P	D	S	V	T	I	F	G	E	S	A	G	G	V	S	V	S	A	L	V	L	S	P
G	N	P	E	W	M	T	L	F	G	E	S	A	G	S	S	S	V	N	A	Q	L	M	S	P
G	P	P	R	R	V	S	L	A	A	D	R	G	G	A	D	V	A	S	I	H	L	L	T	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	A	F	S	V	S	A	H	L	T
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Consensus

G d P . n v t l f g e s a G g a s v s l . l l s p

This Sheet Has Been Replaced





13.

Fig. 6 (3/3)

6
(208-231)
BSSL
(240-263)
ChesHum
(238-261)
Torpace
(223-246)
Drosceh
(226-249)
Ratlivce
(387-410)
Drosace
(2382-2402)
ThryHum
(150-173)
Dict.Di
Consensus

Y	N	K	G	L	I	R	R	A	I	S	Q	S	G	V	A	L	S	P	W	V	I	Q	K
G	S	H	S	L	F	T	R	A	I	L	Q	S	G	S	F	N	A	P	W	A	V	T	S
G	S	R	D	L	F	R	R	A	I	L	Q	S	G	S	P	N	C	P	W	A	S	V	S
D	F	G	Q	L	A	R	A	A	F	S	F	S	G	N	A	L	D	P	W	V	I	Q	K
L	A	K	N	L	F	H	R	A	I	S	E	S	G	V	V	L	T	T	N	L	D	K	K
V	T	R	G	L	V	K	R	G	M	M	Q	S	G	T	M	N	A	P	W	S	H	M	T
T	N	S	Q	L	F	R	R	A	V	L	M	G	G	S	A	L	S	P	A	A	V	I	S
Y	S	R	Q	Y	F	N	A	A	I	S	S	S	S	P	L	T	V	G	L	K	D	K	T
n k g L f r r a i . q s G s a l s p w a i q .																							

7
(285-297)
BSSL
(316-328)
ChesHum
(314-326)
Torpace
(298-310)
Drosceh
(298-306)
Ratlivce
(466-477)
Drosace
(2458-2468)
ThryHum
(224-236)
Dict.Di
Consensus

V	G	F	V	P	V	I	D	G	D	F	I	P
V	N	F	G	P	T	V	D	G	D	F	L	T
F	S	F	V	P	V	I	D	G	E	F	F	P
A	P	F	S	P	V	L	E	P	S	D	A	P
.	.	.	.	T	V	I	D	G	V	V	L	P
P	S	.	A	P	T	I	D	G	A	F	L	P
.	.	W	G	P	V	I	D	G	H	F	L	R
T	I	W	S	P	V	I	D	G	D	A	F	I
. . f . p v . d g d f . p												

This Sheet Has Been Replaced



1 5aa 2 22aa 3 4 5 6 54aa 7

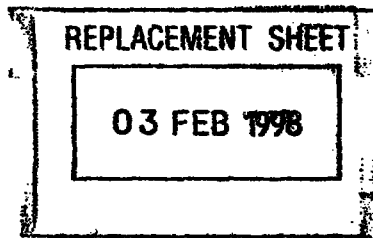
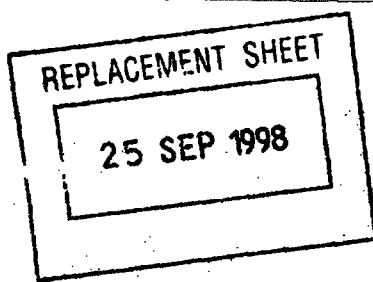


Fig. 7.

10	30	50
AKLGAVYTEGGFVEGVNKKLGLLGDSVDIFKGIPFAAPTKALENPQHPGWQGTILKAKNF		
70	90	110
KKRCLQATITQDSTYGDEDCLYLNWVPQGRKQVSRDLPVMIWIYGGAFIMGSGHGANFL		
130	150	170
NNYLYDGEEIATRGNVIVVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKRNIAA		
190	210	230
FGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSPWVIQKNPLFWAKKV		
250	270	290
AEKVGCPVGDAARMAQCLKVTDPRALTAYKVPLAGLEYPMLHYVGFPVIDGDFIPADP		
310	330	350
INLYANAADIDYIAGTNNMDGHIFASIDMPAINKGKVKVTEEDFYKLVSEFTITKGLRGA		
370	390	410
KTIQDVYTESWAQDPSQENKKKTVVDFEIDVLFVPTETIALAQHRANAKSAKTYAYLFSH		
430	450	470
PSRMPVYPKWVGADHADDIQYVFGKPFATPTGYRQDRTVSKAMIAWTFNAKTGDPNMG		
490	510	530
DSAVPTEWEPYTTENSGYLEITKMGSSSMKRSRLTNFLRYWTLTYLALPTVIDQEATPV		
550	570	590
PPTGDSEATPVPTGDSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD		
610	630	650
SGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAPP		
670	690	710
VPPTGDSGAPPVPTGSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPP		





*Latest amended
specification.*

1

DERIVATIVES OF HUMAN BILE-SALT STIMULATED LIPASE, AND PHARMACEUTI-
CAL COMPOSITIONS CONTAINING THEM

Field of the invention

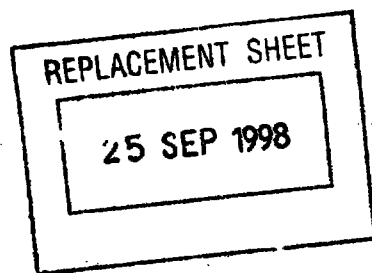
The present invention relates to new DNA sequences, to new
5 proteins coded for by such DNA sequences, and to the use
as further described below of such proteins. The invention
also encompasses vectors, such as plasmid constructs,
comprising such DNA sequences, being capable of expressing
the desired enzyme. The invention also includes host
10 organisms transfected with such constructs e.g. bacteria
yeast, mammalian cells, and transgenic animals. The
invention also includes processes for the preparation of
the novel products of the invention. The new proteins of
the invention are related to an enzyme, known i.a. as
15 human bile salt-stimulated lipase.

20

25

30

35



Background to the invention

The human lactating mammary gland synthesizes and secretes
5 with the milk a bile salt-stimulated lipase (BSSL) [1]
that, after specific activation by primary bile salts [2,
57, 58], contributes to the breast-fed infant's endogenous
capacity of intestinal fat digestion [3-5]. This enzyme,
which accounts for approximately 1% of total milk protein
10 [6], is a non specific lipase; in vitro it hydrolyses not
only tri-, di- and monoacylglycerols, but also
cholesteryl-, and retinyl esters, and
lysophosphatidylglycerols [7-10]. Furthermore, its
activity is not restricted to emulsified substrates, but
15 micellar and soluble substrates are hydrolyzed at similar
rates [11].

BSSL is not degraded during passage with the milk through
the stomach, and in duodenal contents it is protected by
bile salts from inactivation by pancreatic proteases such
20 as trypsin and chymotrypsin [2,11]. It is, however,
inactivated when the milk is pasteurized, e.g. heated to
62.5 °C, 30 min [12]. Model experiments in vitro suggest
that the end products of triacylglycerol digestion are
different in the presence of BSSL [5,7]. Due to lower
25 intraluminal bile salt concentrations during the neonatal
period [13,14] this may be beneficial to product
absorption [5,15].

The carboxylic ester hydrolase (CEH) of human pancreatic
juice [16] seems functionally to be identical, or at least
30 very similar, to BSSL [5]. They also share common epitopes
[8,17], have identical N-terminal amino acid sequences
[17] and are inhibited by inhibitors of serine esterases,
e.g. eserine and diisopropylfluorophosphate [6,8,16]. It
has been hypothesized that the two enzymes are products of
35 the same gene [18,19]. The observed molecular size
difference [8,19] could be explained by different patterns
of glycosylation, as recently suggested [17].

25 SEP 1998

Dietary lipids is an important source of energy. The energy-rich triacylglycerols constitute more than 95% of these lipids. Some of the lipids, e.g. certain fatty acids and the fat soluble vitamins, are essential dietary constituents. Before gastro-intestinal absorption the triacylglycerols as well as the minor components, i.e. esterified fat-soluble vitamins and cholesterol, and diacylphosphatidylglycerols, require hydrolysis of the ester bonds to give rise to less hydrophobic, absorbable products. These reactions are catalyzed by a specific group of enzymes called lipases.

In the human adult the essential lipases involved are considered to be gastric lipase, pancreatic colipase-dependent lipase (tri- and diacylglycerol hydrolysis), pancreatic phospholipase A2 (diacylphosphatidylglycerols) and carboxylic ester hydrolase (cholesteryl- and fat soluble vitamin esters). In the breast-fed newborn bile salt-stimulated lipase plays an essential part in the hydrolysis of several of the above mentioned lipids. Together with bile salts the products of lipid digestion form mixed micelles from which absorption occurs (3-5).

Common causes of lipid malabsorption, and hence malnutrition, are reduced intraluminal levels of pancreatic colipase-dependent lipase an/or bile salts. Typical examples of such lipase deficiency are patients suffering from cystic fibrosis, a common genetic disorder resulting in a life-long deficiency in som 80% of the patients, and chronic pancreatitis, often due to chronic alcoholism.

The pancreatic and liver functions are not fully developed at birth, most notably in infants born before term. Fat malabsorption, for physiological reasons, is a common finding and thought to result from low intraluminal pancreatic colipase-dependent lipase and bile salt concentrations (3,4,13-15). However, because of BSSL, such

25 SEP 1998

malabsorption is much less frequent in breast-fed infants than in infants fed pasteurized human milk of infant formulas (3-5, 12, 59, 60, 61). This is one reason why it has been advocated that newborn infants, particularly
5 preterm infants, that cannot be fed their own mothers milk should be fed non-pasteurized milk from other mothers (12).

The present treatment of patients suffering from a
10 deficiency of pancreatic lipase is the oral administration of very large doses of a crude preparation of porcine pancreatic enzymes. Colipase-dependent pancreatic lipase is inactivated by low pH. Such conditions are prevalent in the stomach, with the result that orally administered
15 pancreatic lipase is virtually completely inactivated on the passage through the stomach to the gut. Therefore, this effect cannot be completely overcome by the use of large doses of enzyme. The large doses administered are inadequate for most patients, and the preparations are
20 impure and unpalatable. Certain tablets have been formulated which pass through the acid regions of the stomach and discharge the enzyme only in the relatively alkaline environment of the jejunum. However, many patients suffering from pancreatic disorders have an
25 abnormally acid jejunum and such tablets may fail to discharge the enzyme and may therefore be ineffective. Moreover, since the preparations presently on the market are of a non-human source there is a risk of immunoreactions that may cause harmful effects to the
30 patients or result in reduced therapy efficiency. A further drawback with the present preparations is that their content of other lipolytic activities than colipase-dependent lipase are not stated. In fact, most of them contain very low levels of CEH/BSSL-activity. This may be
35 one reason why many patients, suffering from cystic fibrosis in spite of supplementation therapy, suffer from deficiencies of fat soluble vitamins and essential fatty

25 SEP 1998

acids.

Thus, there is a great need for products with properties and structure derived from human lipases and with a broad substrate specificity, which products may be orally administered to patients suffering from deficiency of one or several of the pancreatic lipolytic enzymes. The products of the present invention fulfil this need by themselves or in combination with other lipases or in combination with preparations containing other lipases. Furthermore, for some human infants there is an obvious need to improve fat utilization from conventional infant formulas, or pasteurized human milk from so-called milk banks.

BSSL has several unique properties that makes it ideally suited for substitution and supplementation therapy: It has been designed by nature for oral administration. Thus, it resists passage through the stomach and is activated in contents of the small intestine.

Its specific activation mechanism should prevent hazardous lipolysis of food or tissue lipids during storage and passage to its site of action.

Due to its broad substrate specificity it has the potential to, on its own, mediate complete digestion of most dietary lipids, including the fat soluble vitamin esters.

BSSL may be superior to pancreatic colipase-dependent lipase to hydrolyze ester bonds containing long-chain polyunsaturated fatty acids.

In the presence of gastric lipase and in the absence of, or at low levels of colipase-dependent lipase BSSL can ascertain a complete triacylglycerol digestion in vitro even if the bile salt levels are low such as in newborn infants. In the presence of BSSL the end products of triacylglycerol digestion become free fatty acids and free glycerol rather than free fatty acids and monoacylglycerol generated by the other two lipases (5). This may favour product absorption particularly when the intraluminal bile

25 SEP 1998

salt levels are low (3,15).

From a historical point of view infant formulas have been developed, and improved, from the concept that their composition should be as similar to that of human milk as possible. It is desirable to supplement such formulas.

The utilization for supplementation, substitution or therapy of bile salt-stimulated lipases (BSSL), or of proteins with the essential functions of BSSL, requires however access to quantities of the product on a large technical scale. It is not possible in factory scale to rely on natural sources such as milk as starting material. Besides the problem mentioned above with inactivation of BSSL during pasteurization, there is the additional risk of contamination of material from a natural source with infectious agents, e.g. vira such as HIV virus and CMV. There is, accordingly, a need for large scale access to products having BSSL properties. The present invention provides such products and methods for their preparation.

20

Prior art references are given later in this specification.

25 SEP 1998

7

The invention

- The present invention is based on the cloning of cDNA coding for BSSL derived from human mammary gland. We have
- 5 also isolated, from human pancreas, a partial cDNA coding for CEH. Deduced amino acid sequences from the human cDNA's and comparison with CEH from other species, support the interpretation that BSSL and CEH are identical.
- 10 As will be further detailed below, it was surprisingly found that the structure of the protein as deduced from the cDNA sequence is quite different from the structure of other lipases. The structure proved unexpectedly to be more like the structure of typical esterases, such as
- 15 cholinesterase.

25 SEP 1998

8

With reference to Figure II and Figure VII, products of the invention are:

- 5 a) a protein as defined by the amino acid sequence 1-722 in Fig. VII,
- b) a protein as defined by the amino acid sequence 1-535 in Fig. VII,
- 10 c) a protein as defined by the amino acid sequence 1-278 in Fig. VII,
- d) a protein as defined by the amino acid sequence 1-341 in Fig. VII,
- 15 e) a protein as defined by the amino acid sequence 1-409 in Fig. VII,
- f) a protein as defined by the amino acid sequence 1-474 in Fig. VII,
- 20 g) combinations of proteins defined under b) - f) e.g. as defined by the amino acid sequence in positions 1-278, 279-341, 279-409, 279-474, 342-409, 342-474 and 536-722.
- 25 h) combinations of proteins defined under b) - g) in combination with one or more of the repeats according to Figure V,
- 30 i) a protein as defined under a) - h) possessing an additional, N-terminal amino acid, namely methionine, and functionally equivalent variants and mutants of the proteins defined in a) - i) above;
- 35 It should be noted that the proteins under a, b, c, d, e, f, h and i above will not be identical in all respects to naturally occurring BSSL, but they will exhibit one or

25 SEP 1998

more of the critical functions of naturally occurring BSSL. Critical functions are given below.

5 j) a DNA sequence coding for the proteins defined in a, b, c, d, e, f, h and i above.

k) a DNA sequence according to Fig. II, defined by the following nucleotide numbers in Fig. II:

10

a) a DNA sequence 151-2316 according to Fig. II, coding for the protein defined by the amino acid sequence 1-722 in Fig. VII,

15 b) a DNA sequence 151-1755 according to Fig. II, coding for the protein defined by the amino acid sequence 1-535 in Fig. VII,

c) a DNA sequence 151-985 according to Fig. II, coding for the protein defined by the amino acid sequence 1-278 in Fig. VII,

20 d) a DNA sequence 151-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 1-341 in Fig. VII,

e) a DNA sequence 151-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 1-409 in Fig. VII,

25 f) a DNA sequence 151-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 1-474 in Fig. VII,

g) a DNA sequence 986-1172 according to Fig. II, coding for the protein defined by the amino acid sequence 279-341 in Fig. VII,

30 h) a DNA sequence 986-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 279-409 in Fig. VII,

35 i) a DNA sequence 986-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 279-474 in Fig. VII,

25 SEP 1998

- j) a DNA sequence 1173-1376 according to Fig. II, coding for the protein defined by the amino acid sequence 342-409 in Fig. VII,
- k) a DNA sequence 1173-1574 according to Fig. II, coding for the protein defined by the amino acid sequence 342-474 in Fig. VII.

Significant functions of proteins of the invention are

- a) suitable for oral administration,
- b) being activated by specific bile salts,
- c) acting as a non-specific lipase in the contents of the small intestines, that is being able to hydrolyze lipids relatively independent of their chemical structure and physical state (emulsified, micellar, soluble).
- d) Ability to hydrolyze triacylglycerols with fatty acids of different chain-length and different degree of unsaturation.
- e) Ability to hydrolyze also diacylglycerol, monoacylglycerol, cholesteryl esters, lysophosphatidylacylglycerol, and retinyl and other fat soluble vitamin-esters.
- f) Ability to hydrolyze not only the sn-1(3) ester bonds in a triacylglycerol but also the sn-2 ester bond.
- g) Ability to interact with not only primary but also secondary bile salts.
- h) Dependency on bile salts for optimal activity.
- i) Stability so that gastric contents will not affect the catalytical efficiency to any substantial degree.

25 SEP 1998

- j) Stability towards inactivation by pancreatic proteases, e.g. trypsin, provided bile salts are present.
- 5 k) Ability to bind to heparin and heparin derivatives, e.g. heparan sulphate.
- l) Ability to bind to lipid-water interphases.
- 10 m) Stability to permit lyophilization.
- n) Stability when mixed with food constituents such as in human milk, or milk formula.
- 15 The critical functions for supplementation, substitution, or therapy are these according to a), c), d), e), f), i), j) and l). For other purposes, not all critical functions may be necessary.
- 20 For expression of the proteins indicated above, the appropriate DNA sequence indicated above will be inserted into a suitable vector which then is introduced into a suitable host organism. The said vector will also have to comprise appropriate signal and other sequences enabling
- 25 the organism to express the desired protein.

Suitable expression organisms:

- With the recombinant DNA techniques it is possible to
- 30 clone and express a protein of interest in a variety of prokaryotic and eukaryotic host organisms. Possible expression organisms are bacteria, simple eukaryots (yeast), animal cell cultures, insect cell cultures, plant cell cultures, plants and transgenic animals. Each
- 35 individual system has its own particular advantages and disadvantages. The simple conclusion is that every gene to be expressed is a unique problem and no standard solution

25 SEP 1998

is available.

Commonly used bacterial systems are E. Coli, Bacillus subtilis, Streptomyces. Commonly used yeasts are
5 saccharomyces, and Pichia Pastoris. Commonly used animal cells are CHO cells and COS cells. Commonly used insect cell cultures are Drosophila derived cells.

Commonly used plant is the tobacco plant. Possible
10 transgenic animals are goat and cow.

Possible bacterial vectors are exemplified by pUC and protein A-vectors.

Possible yeast vector is exemplified by pMA 91.

15 Possible insect vectors are devired from Baculo-virus. Possible animal cell vectors are derived from SV/40. Possible plant vectors are derived from the Ti-plasmid. In every system, both natural and synthetic promoters and terminators can be used.

20

It is understood that depending on the choice of expression system, the expressed protein may contain an additional N-terminal amino acid (methionine), contain a few extra amino acids, or be fused to a heterologous
25 protein, (e.g. protein A), or differ from the naturally occurring protein with respect to glycosylation. Furthermore, the vectors may also contain signal sequences in order to export the protein to the periplasm or to the culture medium.

30 Thus, further aspects of the invention are:

- a) a vector comprising a DNA sequence coding for a protein as specified above,
- b) a host organism comprising a DNA sequence as specified above,
- 35 c) a process for the production of a protein as specified above, by growing a host organism containing a vector as specified under a) above and

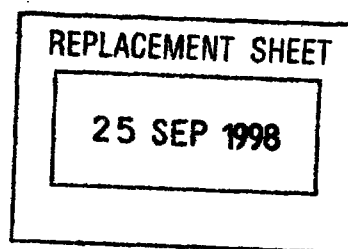
25 SEP 1998

Methods from purification are based on the expression system used (e.g. protein A/IgG) and/or on methods used for purification of the naturally occurring enzyme, as described in reference 6.

5

Additional aspects of the invention are:

- a pharmaceutical composition comprising a protein as specified above,
- 10
- the use of a protein as specified above for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency,
- 15
- the use of a protein as specified above for the manufacture of a medicament for the treatment of cystic fibrosis,
- 20
- the use of a protein as specified above as a supplement to an infant food formulation,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of chronic
- 25 pancreatitis,
- the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption of any etiology,
- 30
- the use of a protein as specified above for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins,
- 35
- the use of a protein as specified above for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons, e.g. in new-



born infants.

5 The DNA sequence in Fig. II from position 151 up to and including position 2316 is the sequence coding for the entire protein. The sequence from position 2317 up to and including position 2415 is not translated to protein, but is included in exon d identified in Table 2 below.

10 In one embodiment of the invention, the protein as defined in paragraphs a) - i) above is provided in isolated form and/or in substantially pure form.

15 The DNA sequences as defined in paragraphs a) - k) above are in one embodiment of the invention provided in isolated form and/or in substantially pure form.

Experimental part

20 Abbreviations

aa, amino acid; bp, base pair; BSSL, bile salt-stimulated lipase; c-AMP, cyclic adenosine monophosphate; CEH, carboxylic ester hydrolase; Da, dalton; c ⁷GTP, 7-deaza-2-deoxyguanosine 5'triphosphate; EDTA, ethylene diamine tetraacetate; kb, kilobases; MOPS, 3-N-morpholino-
25 propanesulfonic acid; nt, nucleotide; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; SSC, NaCl citrate, xGal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

30

Enzymes

Bile salt-stimulated lipase EC 3.1.1.3

Carboxylic ester hydrolase EC 3.1.1.1

35

25 SEP 1998

Material and Methods

A. Enzyme and antibody preparation

- 5 BSSL was purified from human milk as previously described [6]. When used for antibody production the enzyme was further purified by SDS-PAGE. The protein band corresponding to the lipase was, after staining with Coomassie Brilliant blue, electroeluted from the gel.
- 10 Twentyfive μg of purified enzyme, together with an equal volume of Freund's complete adjuvant, was used for a first i.c. injection and the same amount of enzyme with incomplete adjuvant for the subsequent monthly booster injections. The rabbits were bled about two weeks after
- 15 each booster and sera prepared and stored at $-20\text{ }^{\circ}\text{C}$.

B. Preparation of tryptic fragments and amino acid sequence analysis

- 20 Three mg of purified BSSL was dissolved in 1 ml of 0.1M Tris-Cl buffer, pH 8.5, containing 6M guanidinium hydrochloride and 2 mM EDTA. Dithioerythritol was added to 5mM. After incubation at $37\text{ }^{\circ}\text{C}$ for 2h, 300 μl 50 mM iodoacetate was added. After 90 min incubation at $25\text{ }^{\circ}\text{C}$ in
- 25 darkness the reduced and carboxymethylated enzyme was desalted on a Sephadex G-25 column, equilibrated with 0.5M ammonium bicarbonate. Thirty μg of tosyl-L-phenylalanine chloromethane treated bovine trypsin (Worthington diagnostics system Inc., Freehold, NJ, USA)
- 30 was added before lyophilization. The lyophilized protein was dissolved in 4 ml 0.1M ammonium bicarbonate and an additional 90 μg of trypsin was added. After 5h incubation at $37\text{ }^{\circ}\text{C}$ the protein was again lyophilized. The tryptic digest was dissolved in 0.1% trifluoroacetic acid
- 35 (2mg/ml). Three hundred μg of trypsinated BSSL was chromatographed on HPLC using a C-18 reversed phase column and eluted with a gradient of 0-50% acetonitrile in 0.1%

25 SEP 1998

trifluoroacetic acid. Peptide collection was monitored by continuous recording of the absorbance at 215 nm. Peptides to be sequenced were further purified by rechromatography using the same column with adjusted gradients.

- 5 Samples of peptide fragments to be sequenced were dried under nitrogen to remove acetonitrile and applied to the sequencer. For N-terminal sequence analysis, native BSSL was dissolved in 0.1% acetic acid. Sequence analyses were performed on an Applied Biosystems Inc. 477A pulsed
- 10 liquid-phase sequencer and on-line PTH 120A analyzer with regular cycle programs and chemicals from the manufacturer. Initial and repetitive yields, calculated from a sequenced standard protein, β -lactoglobulin, were 47 and 97%, respectively.

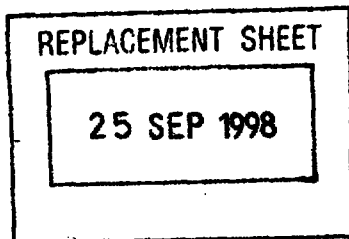
15

C. Isolation of RNA

- Samples of human pancreatic adipose and lactating mammary gland tissues were obtained at surgery and immediately put
- 20 into guanidinium thiocyanate (1-5 g in 50 ml). Total RNA was extracted as described by Chirgwin [20]. Poly(A)-RNA was prepared by chromatography on oligo-deoxythymidilate-(oligo(DT))-cellulose column [21].

25 D. Construction and screening of cDNA libraries

- Approximately 15 μ g poly-adenylated RNA from human pancreas was denatured with methyl mercuric hydroxide [22] and primed with oligo (dT) ₁₂₋₁₈ primers (Pharmacia,
- 30 Uppsala, Sweden), and reversely transcribed using standard procedures [23]. Second-strand synthesis was carried out according to Gubler and Hoffman [24], except that DNA ligase and β -NAD were omitted, and the reaction temperature was set at 15 °C. Excess RNA was digested with
- 35 RNase A (50 μ g/ml), and the double-stranded cDNA was treated with EcoRI methylase [25]. Ends were blunted with Klenow enzyme. After ligation to EcoRI linkers and



cleavage with EcoRI the cDNA was fractionated on a Sepharose 4B-C1 column. The void volume fraction was precipitated with ethanol and the cDNA ligated into the EcoRI site of a phosphatase treated gtl1 vector [26]. In
5 vitro packing yielded more than 7×10^5 recombinants.

A cDNA library from human mammary gland, derived from tissue obtained from a women at the eighth month of pregnancy, was purchased (Clontech Laboratories, Inc.,
10 Palo Alto, CA; USA).

Phages from the cDNA libraries were plated at 5×10^4 plaque forming units per 120-mm dish. The antiserum was diluted to a ratio of 1:3200 and screening was performed according
15 to Young and Davis [27]. Alkaline-phosphatase-conjugated goat-anti-rabbit antibodies were used as second antibodies (Bio-Rad, Richmond, CA USA). To isolate clones corresponding to the 5'-end of the mRNA, nucleic acid hybridization was done under standard conditions [23]
20 using a subcloned fragment from one of the immunopositive clones as a probe.

E. RNA analysis

25 Electrophoresis was carried out in a 1% agarose gel in 40mM MOPS buffer pH 7.0 after denaturation with glyoxal and dimethylsulfoxide [28]. Glyoxalated total RNA was then transferred to nitrocellulose filters [29]. The blots were probed with subclones of BSSL and CEH recombinants that
30 were labelled by the oligo-labeling technique [30]. Prehybridization and hybridization were carried out with 50% formamid at 46 °C [23]. Posthybridization washes were performed at high stringency (0.1% SDS and 0.1xSSC at 60° C). (1xSSC, 0,15M NaCl, 0.0015M Na, citrate, pH 7.6).

35

F. Nucleotide sequence

25 SEP 1998

18

cdNA inserts from BSSL and CEH recombinants were either directly cloned into M3mp18 and mp19 after sonication and size fractionation or some of them were further subcloned into pTZ19R after digestions with PstI, BstXI, NarI, SmaI and AhaII. The nucleotide sequence was determined by the dideoxy chain-termination method [31]. The GC-rich repeats (see below) were also sequenced with TaqI polymerase and dGTP. Both strands were sequenced. Sequence information was retrieved from autoradiograms by use of the software MS-EdSeq as described by Sjöberg et al [55].

G. Amino acid sequence predictions and homologies.

To predict the corresponding amino acid sequence of the cdNA inserts, codon usage of different reading frames was compared according to Staden and gave one open reading frame [32]. Homologies were searched for with the programs of the UWGCG software package [33].

20 Results and Discussion

A. Sequences of tryptic fragments and the N-terminus of BSSL

25 Trypsin digestion of purified BSSL resulted in approximately 50 fragments as judged by the number of peaks obtained during the HPLC-chromatography (Fig. I). The peaks were collected and the indicated peaks which could be isolated in a highly purified state and in reasonable quantities, were sequenced. The resulting sequences are shown in Table I. In addition the 30 most N-terminal residues were sequenced (Fig. II), and they confirm the previously reported sequence of Abouakil et al. (30 residues) [17].

35

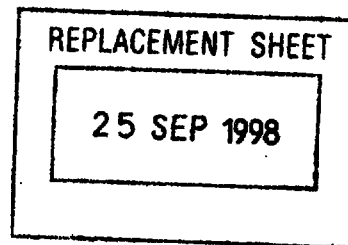
The 22 residues long N-terminal sequence reported from Wang & Johnson [34] is a glycine in our report and a

25 SEP 1998

lysine in Wang & Johnson's.

B. Nucleotide sequence of BSSL

- 5 For construction of the λ gt11 cDNA library we used polyadenylated RNA from human pancreas. Initially four immunopositive clones were isolated, and then this pancreatic expression cDNA library was screened with antiserum against BSSL. Nucleotide sequence analysis of
- 10 the four clones showed that they are in perfect agreement and correspond to the 3'-end of the mRNA. They all begin with a poly A tail and differ only in length; the longest insert, designated ACEH, spans 996 bp.
- A cDNA library from human mammary gland was screened with
- 15 antiserum, and the pancreas clone ACEH as probe. Positive clones were isolated from both screenings, which all originate from the 3'-end. The longest mammary gland clone, designated ABSSL, reaches 2100 bp upstream. It contains four of the sequenced tryptic fragments (Fig.
- 20 II), but do not include the N-terminal amino acid sequence. To extend the sequence beyond the translation start, the mammary gland cDNA library was rescreened with a 118 bases long probe derived from the most 5'proximal part of ABSSL. One clone was isolated that continued a
- 25 further 328 nucleotides upstream. It matched the N-terminal amino acid sequence, and contained the remaining tryptic fragment. As shown in Fig. II, the cDNA is 2428 nucleotides long and contains 81 bases upstream from the first ATG codon. The polyadenylation signal, AATAAA is
- 30 located 13 nucleotides upstream from the poly A tail and the termination codon TAG was found at nucleotide 2317 followed by a 3'-untranslated region of 112 bp.
- A GC rich region consisting of 16 repeats of 33 bases was found in the 3'-end of the sequence between base 1756 and
- 35 2283. The nucleotide sequence of the repetition, shown in Fig. III, consists of six identical repetitions surrounded by ten repetitions with different number of



substitutions that have probably occurred after several duplications. The low number of substitutions suggests that these repetitions have appeared late during evolution.

5

C. Tissue distribution of expression

RNA from human lactating mammary gland, pancreas, adipose tissue and from a human hepatoma cellline (HepG2) was
10 analyzed by Northern blotting. The size of the messenger was determined to be approximately 2.5 kb in both lactating mammary gland and pancreas. No signal could be detected in the lanes with RNA extracted from HepG2 or adipose tissue (Fig. IV).

15 Since the mRNA used for the mammary gland library was obtained from a female in her 8th month of pregnancy, it is evident that transcription and probably translation of the BSSL gene is turned on before partus, in agreement with previous findings on BSSL secretion before partus
20 [35]. See Figure IV.

D. Amino acid sequence of BSSL

Assessed by SDS-PAGE the molecular mass has been reported
25 to be 107-125 kDa [8,36] and by analytical ultracentrifugation to be 105 kDa [37]. The enzyme, as deduced from the cDNA, consists of 722 amino acid residues (Fig. II) which, giving a molecular mass of 76.271 Da, indicates that the enzyme contains at least 15-20%
30 carbohydrate. The leader sequence is 23 residues long. A tentative active site serine residue is localized to serine-217 (Fig. V). The sequence around this serine accord with the consensus active site sequence of serine-hydrolases [38]. It has recently been it was proposed that
35 basic residues found close to the active site serine may be involved in the cleavage of ester bonds in acylglycerols by lipases [39]. It is interesting to note that such

25 SEP 1998

21

residues are not present in BSSL. The single tentative N-glycosylation site is localized only seven residues from the serine. The degree of glycosylation [6,16] suggests that the enzyme contains O-linked carbohydrate. There are numerous sites where such glycosylation could have occurred. The amino acid composition based on purified enzyme has shown a high content of proline residues [6]. The amino acid sequence obtained from cDNA confirms this. Moreover, most of the proline residues are localized in the 16 repeats of 11 residues each, constituting the main part of the C-terminal half of the enzyme.

3. Comparison of the enzymes in mammary gland (BSSL) and pancreas (CEH)

BSSL of human milk and human pancreas CEH have previously been shown to be similar, if not identical. The present data strongly suggests that the two enzymes are products of the same gene. The nucleotide sequence of the cDNA clones shows that the pancreatic clone ACEH is identical with the mammary gland clone ABSSL from the poly A tail and 996 bases towards the 5'-end, including the sequence coding for the proline rich repeats. Northern blot gave a single band of 2.5 kb in RNA from pancreas and lactating mammary gland (Fig. IV). Genomic Southern blots further support the idea that only one gene codes for BSSL and CEH. The difference in mobility on SDS-PAGE between BSSL and CEH can be explained as a consequence of different glycosylation or differential splicing.

The similarity of BSSL to the rat and bovine enzymes (see below) and to results from genomics blots support the possibility that differential splicing cannot account for the mobility difference. Since the C-terminal sequence has not been confirmed on the protein level there is a less likely possibility that CEH may be processed by a proteolytic clearance in the C-terminal end.

25 SEP 1998

22

So far as we know pancreatic enzymes that obviously correspond to CEH have often been named after species and the particular substrates used to determine their respective activities; lysophospholipase, cholesteryl
5 esterase, sterol ester hydrolase, non-specific lipase, carboxyl ester lipase and cholesteryl ester hydrolase. Available data are compatible with the view that all these activities described originates in one and the same functional entity [42,43]. This illustrates the broad
10 substrate specificity and the relevance of designating them as non-specific lipases. When the sequence of human BSSL/CEH is compared to the sequence of lysophospholipase from fat pancreas [40] and cholesterol esterase from bovine pancreas [41] extensive similarities are found that
15 extend about 530 residues from the N-terminal (Fig V); but they differ in the part of the molecule where the repeats occur. The rat enzyme has only four repeats and the bovine three. Hence the human enzyme is a considerable longer peptide.

20

Moreover, striking similarities were found between BSSL and a number of typical esterases, e.g. acetyl choline esterases from several species, including man and Drosophila, and carboxyl esterases (Fig. VI). These
25 similarities were restricted to the N-terminal 300 residues of BSSL which includes the tentative active site serine-residue. A similarity to acetyl choline esterase has been predicted from the fact that BSSL is inhibited by typical choline esterase inhibitors [6, 8, 16]. With the
30 possible exception of the rat liver carboxyl esterase [45], none of these similar enzymes has been shown to have the same bile-salt dependency as BSSL; this suggests that the structural basis for this property resides in the C-terminal part of the protein. Moreover, BSSL can
35 efficiently attack emulsified substrates which is not a known characteristic of the similar esterases. For this activity bile salt is a prerequisite.

25 SEP 1998

The predicted sequence for human BSSL was compared with other well characterized mammalian lipases. Apart from the consensus sequence around the active site serine (G-X-S-X-G), no obvious similarities were found [44].

In addition to the similarities with other enzymes, there also significant similarities to one c-AMP dependent protein from Dictyostelium discoideum [46] as well as to thyroglobulin from several species (Fig VI) [47-49]. The similarities between BSSL and thyroglobulin, which comprise the active site region but not the active site itself, indicate that these highly conserved stretches of amino acids are of more generalized importance than merely supporting the enzymatic activity of esterases.

In conclusion, human milk BSSL consists of 722 amino acid residues. Available data strongly indicate that its peptide chain is identical to that of pancreatic CEH, and they are coded for by the same gene. The strongest evidence is that the nucleotide sequences of their 3 - ends and their N-terminal amino acid sequences are identical. The striking homologies found to rat pancreatic lysophospholipase and bovine pancreatic cholesterol esterase support the hypothesis that also these enzymes are functionally identical. However, as it has been suggested, the different molecular sizes found among species are not due to differences merely in glycosylation; instead they reflect a variable number of an eleven amino acid repeat. The similarity of the active site sequence between these esterases suggests that these proteins derive from a common ancestral gene.

With reference to Figures I-VII, the following legends are given.

Figure I: Separation of the tryptic digest of BSSL on HPLC

Purified BSSL was treated with trypsin and chromatographed on HPLC as described in Materials and Methods. The indicated peaks were collected and purified further by a rechromatograph and their amino acid sequence determined.

Figure II: The cDNA nucleotide sequence and the deduced amino acid sequence for human bile salt-stimulated lipase:

10

The cDNA is 2428 bases long. The N-terminal 23-codon sequence (nt82-150) starting with an ATG, is interpreted as a leader peptide since the N-terminal amino acid sequence of the mature protein starts at codon 24 (nt 151, Ala). The leader peptide is underlined. The sign * indicates the starting point of an exon. The sign ‡ indicates the starting point of the repetition part.

Figure III: The nucleotide sequence of the C-terminal GC-rich repetitions in the bile salt-stimulated lipase:

20

Substitutions are indicated by a *.

Figure IV: Northern blot hybridization

25

Northern blot analysis of total RNA isolated from human lactating mammary gland, pancreas, adipose tissue and a human hepatoma cell line (HepG2). Total TNA (10µg) from lactating mammary gland (lane A), pancreas (lane B), adipose tissue (lane C) and HepG2 (lane D) were electrophoresed in a 1% agarose gel in 40mM MOPS buffer at Ph 7.0 after denaturation of RNA in 1M glyoxal, 50% dimethylsulfoxide and 40mM MOPS. The glyoxalated RNA was then transferred to nitrocellulose paper for hybridization with [³² P] labeled BSSL cDNA (ABSSL).

35

25 SEP 1998

25

Figure V: Comparison of the deduced amino acid sequence from human milk BSSL, rat pancreatic lysophospholipase (Rat1pl) [40] and bovine pancreatic cholesterol esterase (Bovceh) [41]:

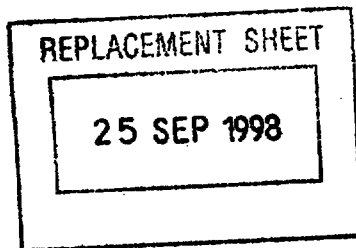
5

The serine residues involved in the active site are indicated by a *, and the # indicates the single possible N-glycosylation signal of the protein. The direct repeats of amino acid sequences are boxed. Matching sequences are denoted in capital letters, matching sequences between two enzymes are denoted in small letters and mismatching with a dot.

Figure VI: Comparison of the primary structure of BSSL to other esterases, thyroglobuline and to one c-AMP dependent enzyme from Dictyostelium discoideum:

BSSL: bile salt stimulated lipase from human, Cheshum: cholinesterase from fetal human tissue [50], Torpace: acetylcholinesterase from Torpedo marmorata [51], Drosceh: carboxylic ester hydrolase from Drosophila melaogaster [52], Ratlivce: carboxyl esterase from rat liver [53], Drosace: acetylcholinesterase from Drosophila melaogaster [54], Thyrrhum: thyroglobulin from human [49] and Dict.Di: c-AMP dependent enzyme from Dictyostelium discoideum [46]. There are 7 different domains that show similarities between the enzymes. Boxes enclose residues which are identical and small letters in the consensus sequence indicate identical residues in all the enzymes except for one. Dots indicate mismatches. The serine residue involved in the active site is indicated with *. The figure in the right hand corner shows how the domains are oriented.

35



26

Figure VII:

gives the amino acid sequence 1 - 722 for the entire protein (one letter code) and indicates exons a, b, c,
5 and d. The sign # indicates the starting point of the repetition part.

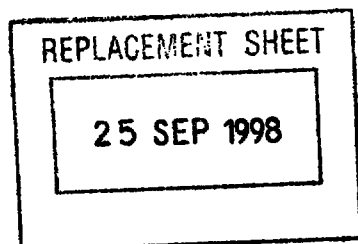


Table 1. Amino acid sequence of BSSL peptides.

Due to interfering peaks no positive identification of the residue in cycles 1 and 2 of the sequencing could be made
5 in peptide number 26. The peptide numbers refer to the peaks in Fig I.

10

Tryptic fragments

15

TP16:LysValThrGluGluAspPheTyrLys

TP19:GlyIleProPheAlaAlaProThrLys

Tp20:LeuValSerGluPheThrIleThrLys

20

TP24:ThrTyrAlaTyrLeuPheSerHisProSerArg

TP26:PheAspValTyrThrGluSerTrpAlaGlnAsp

25

ProSerGlnGluAsnLys

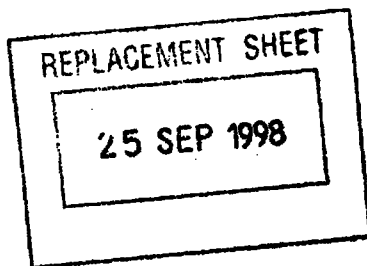


Table 2: Identification of the exons a, b, c and d numbered as in Figures II and VII.

5

exon	<u>Location</u>		
10	between nucleotide number	between aminoacid number	

	a	986-1172	279-341
	b	1173-1376	342-409
15	c	1377-1574	410-474
	d	1575-2415	475-722
	the entire protein	151-2316	1-722
20	the entire protein excluding repetitions	151-1755	1-535
25	-----		

30

35

25 SEP 1998

References

1. Bläckberg, L., Ångquist, K.A., & Hernell, O. (1987) FEBS Lett. 217, 37-41.
- 5 2. Hernell, O. (1975) Eur. J. Clin. Invest. 5, 267-272.
3. Hernell, O., Bläckberg, L. & Bernbäck, S. (1988) In Perinatal nutrition (Lindblad, B.S., ed.) Bristol-Myers Nutrition symposia vol. 6, pp. 259-272, Academic Press, New York.
- 10 4. Hernell, O., L. Bläckberg, B. Fredrikzon, and T. Olivecrona. 1981. Bile salt-stimulated lipase in human milk and lipid digestion in the neonatal period. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 465-471.
- 15 5. Bernbäck, S., Bläckberg, L. & Hernell, O. (1990) J. Clin. Invest. J. Clin. Invest. 221-226 (1990)
6. Bläckberg, L. & Hernell, O. (1981) Eur. J. Biochem 116, 221-225.
- 20 7. Hernell, O. & Bläckberg, L. (1982) Pediatr. Res. 16, 882-885.
8. Bläckberg, L. Lombardo, D., Hernell, O., Guy, O. & Olivecrona, T. (1981) FEBS Lett. 136, 284-288.
9. Fredrikzon, B., Hernell, O., Bläckberg, L. & Olivecrona, T. (1978) Pediatr. Res. 12, 1048-1052.
- 25 10. Wang, C.-S., Hartsuck, J.A. & Downs, D. (1988) Biochemistry 27, 4834-4840.
11. Bläckberg, L. & Hernell, O. (1983) FEBS Lett. 157, 337-341.
- 30 12. Björkstén, B., Burman, L.G., deChateau, P., Fredrikzon, B., Gothevors, L. & Hernell, O. (1980) Br. Med. J. 201, 267-272.
13. Brueton, M.J., Berger, H.M., Brown, G.A., Ablitt, L., Iyangkaran, N. & Wharton B.A. (1978) Gut 19, 95-98.
- 35 14. Murphy, G.M. & Signer, E. (1975) Gut 15, 151-163.
15. Hernell, O., Staggars, J.E. & Carey, M.C. Biochemistry. (1990), 29:2041-2056.

25 SEP 1998

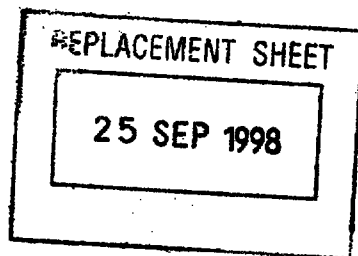
16. Lombardo, D., Guy, O. & Figarella, C. (1978) *Biochim. Biophys. Acta* 527, 142-149.
17. Abouakil, N., Rogalska, E., Bonicel, J. & Lombardo, D. (1988) *Biochim. Biophys. Acta* 961, 299-308.
- 5 18. Hernell, O., Bläckberg, L. & Lindberg, T. (1988) in *Textbook of Gastroenterology and Nutrition in infancy* (Lebenthal, E., ed) pp. 209-217, Raven Press, New York.
19. Wang, C.-S. (1988) *Biochem. Biophys. Res. Com.* 155, 950-955.
- 10 20. Chirgwin, J.M. Przybyla, A.E., MacDonald R.J. & Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
21. Aviv, H. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 69, 5201-5205.
- 15 22. Bailey, J.M. & Davidson, N. (1976) *Anal. Biochem.* 70, 75-85.
23. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982): *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- 20 24. Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263-269.
25. Maniatis T., Haldison, R.C., Lacy, E., Lauer, J., O'Connell, C. & Qvon, D. (1978) *Cell* 15, 687-701.
26. Young, R.A. & Davis, R.W. (1983) *Science* 222, 778-782.
- 25 27. Young, R.A. & Davis, R.W. (1983) *Proc. Natl. Acad. Sci, USA* 80, 1194-1198.
28. McMaster, G.K. & Charmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 48 353-48 358.
29. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
- 30 30. Feinberg, A.P. & Vogelstein, B. (1983) *Analyt. Biochem.* 132, 6-13.
31. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74. 5463-5467.
- 35 32. Staden, R. (1984) *Nucl. Acids Res.* 12, 551-567.
33. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.

25 SEP 1998

34. Wang, C.-S. & Johnson, K. (1983) *Anal. Biochem.* 133, 457-461.
35. Hamosh, M. (1986) in *Human Milk Infant Nutrition and Health* (Howell, R.R. Morriss, F.H. & Pickering, L.K. eds) pp. 66-97, Charles, C. Thomas, Springfield.
- 5 36. Wang, C.-S. (1980) *Anal. Biochem.* 105, 398-402.
37. Wang, C.-S & Lee, D.M. (1985) *J. Lipid. Res.* 26, 824-830.
38. Brenner, S. (1988) *Nature* 334, 528-530.
- 10 39. Yang, C.-Y., Gu, Z.-W., Yang, H.-H., Rohde, M.F., Gotto, Jr., A.M. & Pownall, H.J. (1989) *J. Biol. Chem.* 265, 16822-16827.
40. Han, J.H., Stratowa, C., & Rutter, W.J. (1987) *Biochemistry* 26, 1617-1625.
- 15 41. Kyger, E., Wiegand, K., & Lange, L. (1989) *Biochim. Biophys. Res. Com* 164, 1302-1309.
42. Bläckberg, L. (1981) *Fat digestion in newborn infant. Umeå University Medical Dissertations New Series No 71.*
- 20 43. Rudd, E.A. & Brockman, H.L. (1984) in *Lipases* (Borgström, B & Brockman, H.L. eds) pp. 184-204, Elsevier, Amsterdam.
44. Mickel, S., Weidenbach, F., Swarovsky, B., LaForge, S. & Scheele, G. (1989) *J. Biol. Chem.* 264, 12895-12901.
- 25 45. Cammulli, E.D., Linke, M.J., Brockman, H.L. & Hui, D.Y. (1989) *Biochim. Biophys. Acta* 1005, 177-182.
46. Mann, S.K.O. & Firtel, R.A. (1987) *Mol. Cell Biol.* 7, 458-469.
- 30 47. Mercken, L., Simons, M.J., Swillens, S., Masser, M. & Vassart, G. (1985) *Nature* 316, 647-651.
48. Lauro, R., Obici, S. Condliffe, D., Ursini, V.M., Musti, A., Moscatelli, C. & Avvedimento, V.E. (1985) *Eur. J. Biochem.* 148, 7-11.
- 35 49. Malthiery, Y. & Lissitzky, S. (1987) *Eur. J. Biochem.* 165, 491-498.
50. Prody, C., Zevin-Sonkin, D., Gnatt, A., Goldberg, O.,

25 SEP 1998

- & Soreq, H. (1987) Proc. Natl. Acad. Sci. USA 84, 3555-3559.
51. Sikorav, J-L., Krejci, E. & Massoulie', J. (1987) EMBO J. 6, 1865-1873.
- 5 52. Oakeshott, J.G. Collet, C., Phillis, R.W., Nielsen, K.M., Russel, R.J., Cambers, G.K., Ross, V. & Richmond, R.C. (1987) Proc. Natl. Acad. Sci, USA 84, 3359-3363.
53. Long, R., Satho, H., Martin, B., Kimura, S., Gonzalez, F. & Pohl, L. (1988) Biochem. Biophys. Res. Comm. 156, 866-873.
- 10 54. Hall, L.M.C. & Spierer, P. (1986) EMBO J. 5, 2949-2954.
55. Sjöberg, S., Carlsson, P., Enerbäck, S., & Bjursell, G. (1989) CABIOS 5, 41-46.
- 15 56. EP-A-317355.
57. Hernell, O., and T. Olivecrona. 1974. Human milk lipases II. Bile salt-stimulated lipase. Biochim. Biophys. Acta 369:234-244.
- 20 58. Hernell, O., L. Bläckberg, and T. Olivecrona. 1981. Human milk lipases. In Textbook of gastroenterology and nutrition in infancy. E. Lebenthal, editor. Raven Press, New York. 347-354.
- 25 59. Atkinson, S.A., M.H. Bryan, and G.H. Andersson. 1981. Human milk feeding in premature infants: protein, fat and carbohydrate balances in the first two weeks of life. J.Pediatr. 99:617-624.
- 30 60. Chappell, J.E., M.T. Clandinin, C. Kearney-Volpe, B. Peichman and P.W. Swyer. 1986. Fatty acid balance studies in premature infants fed human milk or formula: effect of calcium supplementation. J. Pediatr. 108:438-447.
- 35



33

61. Williamson, S., E. Finucane, H. Ellis, and H.R. Gamsu. 1978. Effect of heat treatment of human milk on absorption of nitrogen, fat, sodium, calcium and phosphorus by preterm infants. Arch. Dis. Childhood 53:555-563.

5

10

15

25 SEP 1998



What we claim is:

1. A protein as indicated in Figure VII from position 1
5 to position 722.
2. A protein as indicated in Figure VII from position 1
to position 535.
- 10 3. A protein as indicated in Figure VII from position 1
to position 278.
4. A protein as indicated in Figure VII from position 1
to position 341.
- 15 5. A protein as indicated in Figure VII from position 1
to position 409.
6. A protein as indicated in Figure VII from position 1
20 to position 474.
7. A protein as indicated in Figure VII from position 1
to position 278, position 279 to position 341,
position 279 to position 409, position 279 to
25 position 474, position 342 to position 409, position
342 to position 474, or position 536 to position 722.
8. A protein according to any of claims 1-7 in
combination with one or more of the repeats according
30 to Figure V.
9. A protein according to any of claims 1-8, possessing
methionine as additional N-terminal amino acid.
- 35 10. A protein according to any of claims 1-9, exhibiting
one or several of the critical functions of
naturally occurring bile salt-stimulated lipase.

25 SEP 1998

35

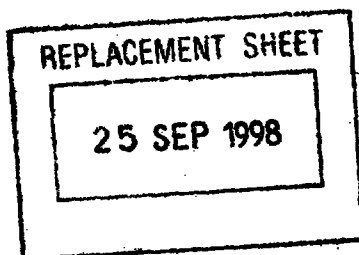


11. A functionally equivalent variant or mutant of a protein according to any of claims 1-9.
12. A DNA sequence coding for a protein with the amino acid sequence according to claim 1.
13. A DNA sequence coding for a protein with the amino acid sequence according to claim 2.
14. A DNA sequence coding for a protein with the amino acid sequence according to any of claims 3, 4, 5, 6, 7, 8, 9, 10 and 11.
15. A DNA sequence defined by the following nucleotide numbers in Figure II:
- 151-2316
151-1755
151-985
151-1172
151-1376
151-1574
986-1172
986-1376
986-1574
1173-1376
1173-1574
16. A vector comprising a DNA sequence coding for a protein according to claims 1 - 11
17. A vector comprising a DNA sequence according to claims 12 - 15.
18. A host organism transformed with a vector according to claims 16 or 17.

25 SEP 1998

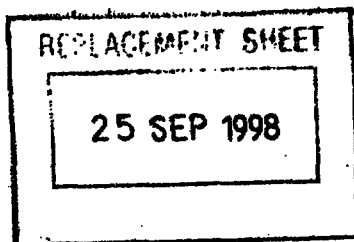


19. A process for the production of a protein according to claims 1 - 11 by growing a host organism containing a vector according to claims 16 or 17, and isolating the protein.
20. A pharmaceutical composition comprising a protein according to any of claims 1 - 11.
21. A pharmaceutical composition comprising a protein according to any of claims 1-11 in combination with a lipase or in combination with preparations containing lipases.
22. The use of a protein according to any of claims 1 - 11, for the manufacture of a medicament for the treatment of a pathological condition related to exocrine pancreatic insufficiency.
23. The use of a protein according to any of claims 1 - 11 for the manufacture of a medicament for the treatment of cystic fibrosis.
24. The use of a protein according to any of claims 1 - 11 as a supplement to an infant food formulation.
25. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of chronic pancreatitis.
26. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption.
27. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of malabsorption of fat soluble vitamins.



37

28. The use of a protein according to any of claims 1-11 for the manufacture of a medicament for the treatment of fat malabsorption due to physiological reasons.
- 5
29. The use, according to claims 22-28 of a protein according to any of claims 1-11 in combination with a lipase or lipases or in combination with preparations containing a lipase or lipases.
- 10
30. An infant food formulation, supplemented with a protein according to any of claims 1 - 11.
- 15
31. A protein according to claims 1 - 11 in substantially pure form.
32. A protein according to claims 1 - 11 in isolated form.
- 20
33. A DNA sequence according to claims 12 - 15 in substantially pure form.
34. A DNA sequence according to claims 12 - 15 in isolated form.
- 25
35. A-process for the preparation of a pharmaceutical composition according to claims 20 or 21 by incorporating a protein according to claims 1 - 11, 31, or 32 in a pharmacologically acceptable carrier.
- 30
36. A process according to claim 35 for the preparation of a pharmaceutical composition for oral administration.
- 35
37. A process for the preparation of an infant food formulation according to claim 30, by supplementing an infant food formulation with a protein according to claims 1 -11, 31, or 32.



DERIVATIVES OF HUMAN BILE-SALT STIMULATED LIPASE,
AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM

Abstract

The present invention relates to new DNA sequences, to new proteins coded for by such DNA sequences, and to the use as further described below of such proteins. The invention also encompasses vectors, such as plasmid constructs, comprising such DNA sequences, being capable of expressing the desired enzyme. The invention also includes host organisms transfected with such constructs e.g. bacteria yeast, mammalian cells, and transgenic animals. The invention also includes processes for the preparation of the novel products of the invention. The new proteins of the invention are related to an enzyme, known i.a. as human bile salt-stimulated lipase.

FIGURE 2.

REPLACEMENT SHEET
SEP 1998

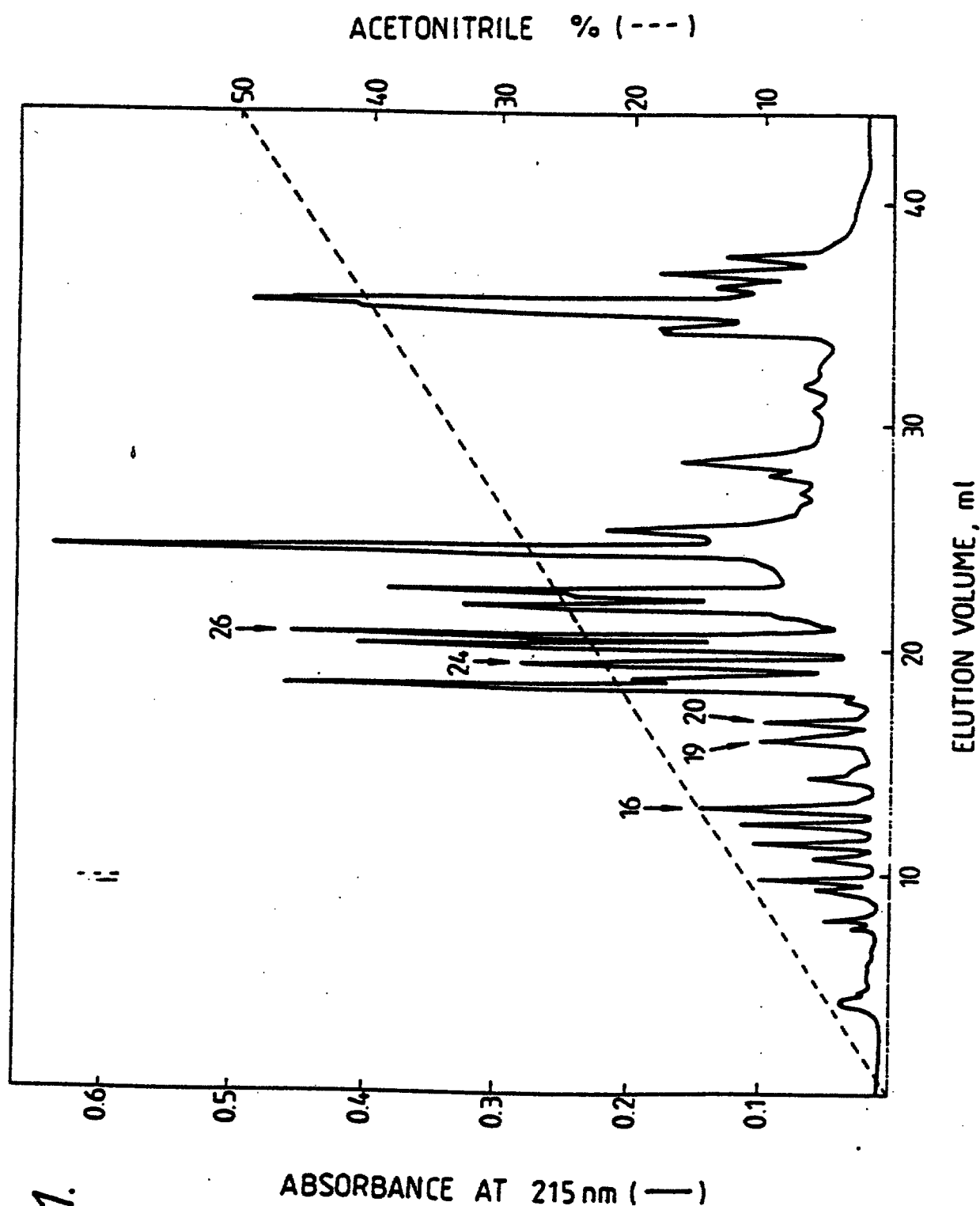
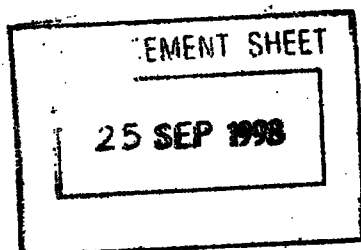


Fig. 1.



2,

Fig.2 (1/4)

ACCTTCTGTGTA TCAGTTAAGT GTCAAGATGG AAGGAACAGC AGTCTCAAGA TAATGCAAAG	60
AGTTTATTCA TCCAGAGGCT G ATG CTC ACC ATG GGG CGC CTG CAA CTG GTT	111
Met Leu Thr Met Gly Arg Leu Gln Leu Val	
1 5 10	
GTG TTG GGC CTC ACC TGC TGC TGG GCA GTG GCG AGT GGC GCG AAG CTG	159
Val Leu Gly Leu Thr Cys Cys Trp Ala Val Ala Ser Ala Ala Lys Leu	
15 20 25	
GGC GCC GTG TAC ACA GAA GGT GGG TTC GTG GAA GGC GTC AAT AAG AAG	207
Gly Ala Val Tyr Thr Glu Gly Gly Phe Val Glu Gly Val Asn Lys Lys	
30 35 40	
CTC GGC CTC CTG GGT GAC TCT GTG GAC ATC TTC AAG GGC ATC CCC TTC	255
Leu Gly Leu Leu Gly Asp Ser Val Asp Ile Phe Lys Gly Ile Pro Phe	
45 50 55	
GCA GCT CCC ACC AAG GCC CTG GAA AAT CCT CAG CCA CAT CCT GGC TGG	303
Ala Ala Pro Thr Lys Ala Leu Glu Asn Pro Gln Pro His Pro Gly Trp	
60 65 70	
CAA GGG ACC CTG AAG GCC AAG AAC TTC AAG AAG AGA TGC CTG CAG GCC	351
Gln Gly Thr Leu Lys Ala Lys Asn Phe Lys Lys Arg Cys Leu Gln Ala	
75 80 85 90	
ACC ATC ACC CAG GAC AGC ACC TAC GGG GAT GAA GAC TGC CTG TAC CTC	399
Thr Ile Thr Gln Asp Ser Thr Tyr Gly Asp Glu Asp Cys Leu Tyr Leu	
95 100 105	
AAC ATT TGG GTG CCC CAG GGC AGG AAG CAA GTC TCC CGG GAC CTG CCC	447
Asn Ile Trp Val Pro Gln Gly Arg Lys Gln Val Ser Arg Asp Leu Pro	
110 115 120	
GTT ATG ATC TGG ATC TAT GGA GGC GCC TTC CTC ATG GGG TCC GGC CAT	495
Val Met Ile Trp Ile Tyr Gly Gly Ala Phe Leu Met Gly Ser Gly His	
125 130 135	
GGG GCC AAC TTC CTC AAC AAC TAC CTG TAT GAC GGC GAG GAG ATC GCC	543
Gly Ala Asn Phe Leu Asn Asn Tyr Leu Tyr Asp Gly Glu Glu Ile Ala	
140 145 150	
ACA GGC GGA AAC GTC ATC GTG GTC ACC TTC AAC TAC CGT GTC GGC CCC	591
Thr Arg Gly Asn Val Ile Val Val Thr Phe Asn Tyr Arg Val Gly Pro	
155 160 165 170	
CIT GGG TTC CTC AGC ACT GGG GAC GCC AAT CTG CCA GGT AAC TAT GGC	639
Leu Gly Phe Leu Ser Thr Gly Asp Ala Asn Leu Pro Gly Asn Tyr Gly	
175 180 185	

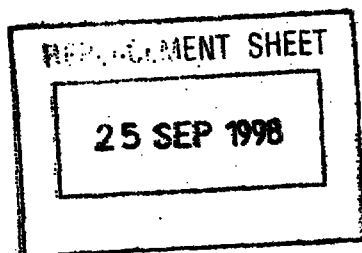


Fig. 2(2/4)

CTT CCG GAT CAG CAC ATG GCC ATT GCT TGG GTG AAG AGG AAT ATC GCG Leu Arg Asp Gln His Met Ala Ile Ala Trp Val Lys Arg Asn Ile Ala 190 195 200	627
GCC TTC GGG GGG GAC CCC AAC AAC ATC ACG CTC TTC GGG GAG TCT GCT Ala Phe Gly Gly Asp Pro Asn Asn Ile Thr Leu Phe Gly Glu Ser Ala 205 210 215	735
GGA GGT GCC AGC GTC TCT CTG CAG ACC CTC TCC CCC TAC AAC AAG GGC Gly Gly Ala Ser Val Ser Leu Gln Thr Leu Ser Pro Tyr Asn Lys Gly 220 225 230	783
CTC ATC CCG CGA GCC ATC AGC CAG AGC GGC GTG GCC CTG AGT CCC TGG Leu Ile Arg Arg Ala Ile Ser Gln Ser Gly Val Ala Leu Ser Pro Trp 235 240 245 250	831
GTC ATC CAG AAA AAC CCA CTC TTC TGG GCC AAA AAG GTG GCT GAG AAG Val Ile Gln Lys Asn Pro Leu Phe Trp Ala Lys Lys Val Ala Glu Lys 255 260 265	879
GTG GGT TGC OCT GTG GGT GAT GCC GCC AGG ATG GCC CAG TGT CTG AAG Val Gly Cys Pro Val Gly Asp Ala Ala Arg Met Ala Gln Cys Leu Lys 270 275 280	927
GTT ACT GAT CCC CGA GCC CTG ACG CTG GCC TAT AAG GTG CCG CTG GCA Val Thr Asp Pro Arg Ala Leu Thr Leu Ala Tyr Lys Val Pro Leu Ala 285 290 295	975
*	
GGC CTG GAG TAC CCC ATG CTG CAC TAT GTG GGC TTC GTC OCT GTC ATT Gly Leu Glu Tyr Pro Met Leu His Tyr Val Gly Phe Val Pro Val Ile 300 305 310	1023
GAT GGA GAC TTC ATC CCC GCT GAC CCG ATC AAC CTG TAC GCC AAC GCC Asp Gly Asp Phe Ile Pro Ala Asp Pro Ile Asn Leu Tyr Ala Asn Ala 315 320 325 330	1071
GCC GAC ATC GAC TAT ATA GCA GGC ACC AAC AAC ATG GAC GGC CAC ATC Ala Asp Ile Asp Tyr Ile Ala Gly Thr Asn Asn Met Asp Gly His Ile 335 340 345	1119
TTC GCC AGC ATC GAC ATG OCT GCC ATC AAC AAG GGC AAC AAG AAA GTC Phe Ala Ser Ile Asp Met Pro Ala Ile Asn Lys Gly Asn Lys Lys Val 350 355 360	1167
*	
ACG GAG GAG GAC TTC TAC AAG CTG GTC AGT GAG TTC ACA ATC ACC AAG Thr Glu Glu Asp Phe Tyr Lys Leu Val Ser Glu Phe Thr Ile Thr Lys 365 370 375	1215
GGG CTC AGA GGC GCC AAG ACG ACC TTT GAT GTC TAC ACC GAG TCC TGG Gly Leu Arg Gly Ala Lys Thr Thr Phe Asp Val Tyr Thr Glu Ser Trp 380 385 390	1263

25 SEP 1998

Fig.2(3/4)

GCC CAG GAC CCA TCC CAG GAG AAT AAG AAG AAG ACT GTG GTG GAC TTT	1311
Ala Gln Asp Pro Ser Gln Glu Asn Lys Lys Lys Thr Val Val Asp Phe	
395 400 405 410	
GAG ACC GAT GTC CTC TTC CTG GTG CCC ACC GAG ATT GCC CTA GCC CAG	1359
Glu Thr Asp Val Leu Phe Leu Val Pro Thr Glu Ile Ala Leu Ala Gln	
415 420 425	
*	
CAC AGA GCC AAT GCC AAG AGT GCC AAG ACC TAC GCC TAC CTG TTT TCC	1407
His Arg Ala Asn Ala Lys Ser Ala Lys Thr Tyr Ala Tyr Leu Phe Ser	
430 435 440	
CAT CCC TCT CGG ATG CCC GTC TAC CCC AAA TGG GTG GGG GCC GAC CAT	1455
His Pro Ser Arg Met Pro Val Tyr Pro Lys Trp Val Gly Ala Asp His	
445 450 455	
GCA GAT GAC ATT CAG TAC GTT TTC GGG AAG CCC TTC GCC ACC CCC ACG	1503
Ala Asp Asp Ile Gln Tyr Val Phe Gly Lys Pro Phe Ala Thr Pro Thr	
460 465 470	
GGC TAC CGG CCC CAA GAC AGG ACA GTC TCT AAG GCC ATG ATC GCC TAC	1551
Gly Tyr Arg Pro Gln Asp Arg Thr Val Ser Lys Ala Met Ile Ala Tyr	
475 480 485 490	
*	
TGG ACC AAC TTT GCC AAA ACA GGG GAC CCC AAC ATG GCC GAC TCG GCT	1599
Trp Thr Asn Phe Ala Lys Thr Gly Asp Pro Asn Met Gly Asp Ser Ala	
495 500 505	
GTG CCC ACA CAC TGG GAA CCC TAC ACT ACG GAA AAC AGC GCC TAC CTG	1647
Val Pro Thr His Trp Glu Pro Tyr Thr Thr Glu Asn Ser Gly Tyr Leu	
510 515 520	
GAG ATC ACC AAG AAG ATG GGC AGC AGC TCC ATG AAG CGG AGC CTG AGA	1695
Glu Ile Thr Lys Lys Met Gly Ser Ser Ser Met Lys Arg Ser Leu Arg	
525 530 535	
ACC AAC TTC CTG CGC TAC TGG ACC CTC ACC TAT CTG GCG CTG CCC ACA	1743
Thr Asn Phe Leu Arg Tyr Trp Thr Leu Thr Tyr Leu Ala Leu Pro Thr	
540 545 550	
*	
GTG ACC GAC CAG GAG GCC AOC OCT GTG CCC CCC ACA GGG GAC TCC GAG	1791
Val Thr Asp Gln Glu Ala Thr Pro Val Pro Pro Thr Gly Asp Ser Glu	
555 560 565 570	
GCC ACT CCC GTG CCC CCC ACG GGT GAC TCC GAG ACC GCC CCC GTG CCG	1839
Ala Thr Pro Val Pro Pro Thr Gly Asp Ser Glu Thr Ala Pro Val Pro	
575 580 585	
CCC ACG GGT GAC TCC GGG GCC CCC CCC GTG CCG CCC ACG GGT GAC TCC	1887
Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser	
590 595 600	

REPLACEMENT SHEET

25 SEP 1998

Fig.2 (4/4)

GGG GGC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC GTG Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val 605 610 615	1935
CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC GTG CCG CCC ACG GGT GAC Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp 620 625 630	1983
TCC GGG GGC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro 635 640 645 650	2031
GTG CCG CCC ACG GGT GAC TCC GGC GGC CCC CCC GTG CCG CCC ACG GGT Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly 655 660 665	2079
GAC GGC GGG CCC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGC GGC CCC Asp Ala Gly Pro Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro 670 675 680	2127
CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC CCC CCC GTG ACC CCC ACG Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Thr Pro Thr 685 690 695	2175
GGT GAC TCC GAG ACC GGC CCC GTG CCG CCC ACG GGT GAC TCC GGG GGC Gly Asp Ser Glu Thr Ala Pro Val Pro Pro Thr Gly Asp Ser Gly Ala 700 705 710	2223
CCC CCT GTG CCC CCC ACG GGT GAC TCT GAG GCT GCC CCT GTG CCC CCC Pro Pro Val Pro Pro Thr Gly Asp Ser Glu Ala Ala Pro Val Pro Pro 715 720 725 730	2271
ACA GAT GAC TCC AAG GAA GCT CAG ATG CCT GCA GTC ATT AGG TTT Thr Asp Asp Ser Lys Glu Ala Gln Met Pro Ala Val Ile Arg Phe 735 740 745	2316
TAGGTCCCA TGAGCCTTGG TATCAAGAGG CCACAAGAGT GGGACCCCAG GGGCTCCCC	2376
CCCATCTTGA GCTCTTCTG AATAAAGCCT CATACCCCTA AAAAAAAAAA AA	2428

25 SEP 1998

Fig. 3.

		Number of substitutions
GAGGCCACCCCTGTGCCCCCACAGGGGACTCC	1	6
* * *		
GAGGCCACTCCCGTGCCCCCACGGGTGACTCC	2	4
* * *		
GAGACCGCCCCCGTGCCGCCACGGGTGACTCC	3	3
* * *		
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	4	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	5	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	6	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	7	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	8	0
GGGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	9	0
GGCGCCCCCCCCCGTGCCGCCACGGGTGACGCC	10	2
* *		
GGGCCCCCCCCCGTGCCGCCACGGGTGACTCC	11	1
*		
GGCGCCCCCCCCCGTGCCGCCACGGGTGACTCC	12	1
*		
GGGGCCCCCCCCCGTGACCCCCACGGGTGACTCC	13	2
* *		
GAGACCGCCCCCGTGCCGCCACGGGTGACTCC	14	3
* * *		
GGGGCCCCCCCCCTGTGCCCCCACGGGTGACTCT	15	3
* *		
GAGGCTGCCCCCTGTGCCCCCACAGATGACTCC	16	7
* ** * * **		
GGGGCCCCCCCC GTGCCGCCACGGGTGACTCC	Consensus	

Identical
to
consensus

REPLACEMENT SHEET

25 SEP 1998

7.

Fig. 4.

A B C D Kb

Ø3.0

Ø2.0

Ø1.5



25 SEP 1998

8

Fig. 5 (1/3)

Bssl MLTMGRLQLVVLGLTCCWAVASAAKLGAVYTEGGFVEGVNKKLGLLG.DS 50
 Ratlpl ...MGRLEVLFGLTCCLAACAAGALYTEGGFVEGVNKKLSLLGGDS
 BovcehMLGASRLGPSGCLAVASAAKLGSVYTEGGFVEGVNKKLSLFG.DS
 Consensusgrl....lgltcclaaaaAAKLGavYTEGGFVEGVNKKLSLLG.DS

Bssl VDIFKGIPFAAPTCALENPQPHPGWQGTLLKAKNFKKRCLOATITQDSTYG 100
 Ratlpl VDIFKGIPFA.TAKTLENPQRHPGWQGTLLKATQFKKRCLOATITQDOTYG
 Bovech VDIFKGIPFAAAPKALEKPKRHPGWQGTLLKAKSFKKRCLOATITQDSTYG
 Consensus VDIFKGIPFAa..KaLEnPqrHPGWQGTLLKak.FKKRCLOATITQDSTYG

Bssl DEDCLYLNIWVPQGRKQVSRDLPVMIWIYGGAFLMGSGBGANFLNNYLYD 150
 Ratlpl QEDCLYLNIWVPQGRKQVSHDLPVMVWIYGGAFLMGSGQGANFLKNYLYD
 Bovceh NEDCLYLNIWVPQGRKEVSHDLPVMIWIYGGAFLMGASQGANFLSNYLYD
 Consensus .EDCLYLNIWVPQGRKqVshDLPVmiWiYGGAFLMGsgqGANFL.NYLYD

Bssl GKKIATRGNVIVVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKR 200
 Ratlpl GKKIATRANVIVVTFNYRVGPLGFLSTGDANLPGNFGLRDQHMAIAWVKR
 Bovech GKKIATRGNVIVVTFNYRVGPLGFLSTGDSNLPNGYGLWDQHMAIAWVKR
 Consensus GKKIATRgNVIVVTFNYRVGPLGFLSTGDaNLPNGyGLrDQHMAIAWVKR

Bssl NIAAFGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSP 250
 Ratlpl NIAAFGGDPNNITIFGESAGGAIVSLQTLSPYNKGLIRRAISQSGVALSP
 Bovceh NIEAFGGDPDNITLFGESAGGASVSLQTLSPYNKGLIKRGISQSGVGLCP
 Consensus NIaAFGCDPDNITLFGESAGGAsVSLQTLSPYNKGLIrRaISQSGVaLsP

Bssl WVIQKNPLFWAKKVAKKVGCPCVDAARMAQCLKVTDPRALTLAYKVPLAG 300
 Ratlpl WAIQENPLFWAKTIKKVGCPTEDTAKMAGCLKITDPRALTLAYRLPLKS
 Bovech WAIQQDPLFWAKRIKKVGCPCVDDTSKMAGCAKITDPRALTLAYKLPLGS
 Consensus WaIQ.nPLFWAK.iAKKVGCpV.DtakMAgClKiTDPRALTLAYkIPL.s

Bssl LEYPMHYVGFVPVIDGDFIPADPINLYANAADIDYIAGTNNMDGHIFAS 350
 Ratlpl QEYPIVEYLAFIGVVDGDFIPDDPINLYDNAADIDYLAGINDMDGHLFAT
 Bovceh TEYPKLHYLSFVPVIDGDFIPDDPNLYANAADVYIAGTNDMDGHLFVG
 Consensus .EYP.lHYl.FvPViDGDFIPdDPiNLYaNAADiDYiAGtNdMDGHlFa.

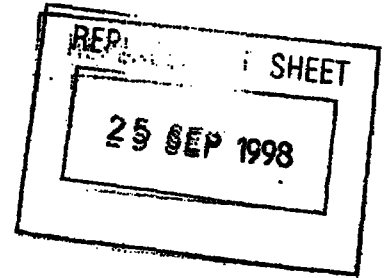


Fig. 5 (2/3)

400

Bssl IDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGAKTTFDVYTESWAQDPS
 Ratlpl VDVPAIDKAKQDVTEEDFYRLVSGHTVAKGLKGTQATFDIYTESWAQDPS
 Bovech MDVPAINSNKQDVTEEDFYKLVSGLTVTKGLRGANATYEVYTEPWAQDSS

Consensus .DvPAInk.kqdvTEEDFYkLVsg.TvtKGLrGa.atfdvYTESWAQDps

450

Bssl QENKKKTVVDFETDVLFLVPTEIALAQHRANAKSAKTYAYLFSHPSRMPV
 Ratlpl QENMKKTVVAFETDILFLIPTEMALAQHRAHAKSAKTYSYLFSHPSRMPi
 Bovceh QETRKKTMVDLETDILFLIPTKIAVAQHAKSHAKSANTYTYLFSQPSRMPi

Consensus QEn.KKtVvdfETDiLFLiPteIAAQHrahAKSAkTY.YLFSHPSRMPi

500

Bssl YPKWVGADHADDIQYVFGKPFATPTGYRPQDRTVSKAMIAYWTNFAKTGD
 Ratlpl YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAYWTNFAKSGD
 Bovech YPKWMGADHADDLQYVFGKPFATPLGYRAQDRTVSKAMIAYWTNFARTGD

Consensus YPKWmGADHADDlQYVFGKPFATPlGYRaQDRTVSKAMIAYWTNFAktGD

550

Bssl PNMGDSAVPTHWEPTYTTENSGYLEITKKMGSSSMKRSRLRTNFLRYWTLTY
 Ratlpl PNMGNSPVPTHWYPYTMENGNLYLDINKKITSTSMKEHLREKFLKFWAVTF
 Bovceh PNTGHSTVPANWDPYTLLEDDNYLEINKQMDSNSMKLRTLNTNYLQFWTQTY

Consensus PnNg.S.VPthW.PYT.Kn.nyleInKkm.S.SMK.hLRtnfL.fwt.Ty

596

Bssl	LALPTVTDQ	EATPVPPTGDS	EATPVPPTGDS	ETAPVPPTGDS	GAPP
Ratlpl	EMLPTV...	VGDHTPPKDDS	EAAPVPPTDDS	DGGVPVPTDDS	QTPP
Bovech	QALPTV TSA	GASLLPPEDNS	QASPVPPADNS	GAPTEPSAGDS	

Consensus .aLPTVt....a...PP.ddS.eA.PVPptddS....pvPptgDS....p

642

Bssl	VPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GAPPVP
Ratlpl	VPPTDNS	QA.....			
Bovceh					

Consensuss..a.....

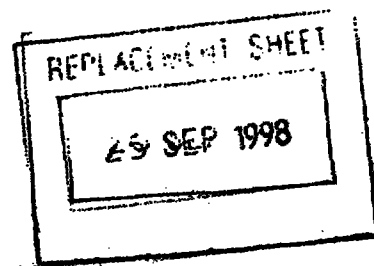


Fig.5 (3/3)

					689
Bssl	PTGDS	GAPPVPPTGDS	GAPPVPPTGDS	GPPPVPTPTGDS	GAPPVPPTG
Rarlpl					
Bovech					
Consensus					

					735
Bssl	DS	GAPPVTPTGDS	ETAPVPPTGDS	GAPPVPPTGDS	EAAPVPPTDDS
Ratlpl	..				GDS
Bovceh	..				
Consensus	ds				

Bssl	KE.AQMPAVIRF
Ratlpl	VE.AQMPGPIGF
Bovech	.EVAQMPVVIGF
Consensus	.e.AQMP.vIgF

25 SEP 1998

11.

1

(76-90)

BSSL

(116-130)

ChesHum

(114-128)

Torpace

(101-115)

Drosceh

(103-117)

Ratlivce

(224-238)

Drosace

(2257-2270)

ThryHum

(90-114)

Dict.Di

Consensus

```

G D E D C L Y L N I W V P Q G
L S E D C L Y L N V W I P A P
M S E D C L Y L N I W V P S P
G E E D C L T V S V Y K P K N
F S E D C L Y L N I Y S P A D
V S E D C L Y I N V W A P A K
V S E D C L Y L N V F I P Q .
A Q K C N L G P G V C S P M G
. s e d c l y l n v w . P . .

```

Fig.6(1/3)

2

(95-113)

BSSL

(134-150)

ChesHum

(132-150)

Torpace

(118-136)

Drosceh

(121-138)

Ratlivce

(273-291)

Drosace

(2270-2291)

ThryHum

Dict.Di

Consensus

```

S R D L P V M I W I Y G G A F L M G S
P K D A T V L I W I Y G G G F Q T G T
P K S A T V M L W I Y G G G F Y S G S
R N S F P V V A H I H G G A F M F G A
N S R L P V M V W I H G G G L I I G G
T N G L P I L I W I Y G G G F M T G S
A P N A S V L V F F H N T M D R E E S
. . . . .
. . . l p v . . w i y g g g f . . g s

```

3

(135-156)

BSSL

(166-187)

ChesHum

(164-185)

Torpace

(150-171)

Drosceh

(153-174)

Ratlivce

(307-328)

Drosace

(2307-2328)

ThryHum

(118-136)

Dict.Di

Consensus

```

N V I V V T F N Y R V G P L G F L S T G D A
R V I V V S M N Y R V G A L G F L A L P G N
E V V L V S L S Y R V G A F G F L A L H G S
K F I L V K I S Y R L G P L G F V S T G D E
N V V V V T I Q Y R L G F G G L F S T G D E
N V I V A S F Q Y R V G A F G F L H L A P E
N L I V V T A S Y R V G V F G F L S S G S G
S V I V V T I N Y R L G I L G L M G T . . .
n v i v v t f n Y R v g . . G f l s t g d .

```

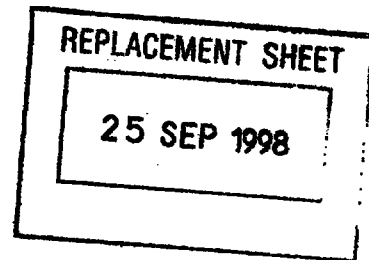


Fig. 6 (2/3)

4

(157-182)
 BSSL
 (189-214)
 ChesHum
 (187-212)
 Torpace
 (172-197)
 Drosceh
 (175-200)
 Ratlivce
 (336-361)
 Drosace
 (2329-2354)
 ThryHum

N	L	P	G	N	Y	G	L	R	D	Q	H	M	A	I	A	W	V	K	R	N	I	A	A	F	G
E	A	P	G	N	M	G	L	F	D	Q	Q	L	A	L	Q	W	V	Q	K	N	I	A	A	F	G
E	A	P	G	N	M	G	L	L	D	Q	R	M	A	L	Q	W	V	H	D	N	I	Q	F	F	G
D	L	P	G	N	Y	G	L	K	D	Q	R	L	A	L	K	W	I	K	Q	N	I	A	S	F	G
H	S	R	G	N	W	A	H	L	D	Q	L	A	A	L	R	W	V	Q	D	N	I	A	N	F	G
E	A	P	G	N	V	G	L	W	D	Q	A	L	A	I	R	W	L	K	D	N	A	H	A	F	G
E	V	S	G	N	N	G	L	L	D	Q	V	A	A	L	T	W	V	O	T	H	I	R	G	F	G

Dict.Di

.....

Consensus

e l p G N w q l l D Q . . A l . W v . d n i a a F G

5

(183-207)
 BSSL
 (215-239)
 ChesHum
 (213-237)
 Torpace--
 (198-222)
 Drosceh
 (201-225)
 Ratlivce
 (362-386)
 Drosace
 (2354-2379)
 ThryHum
 (137-149)
 Dict.Di

G	D	P	N	N	I	T	L	F	G	E	S	A	G	G	A	S	V	S	L	Q	T	L	S	P
G	N	P	K	S	V	T	L	F	G	E	S	A	G	A	A	S	V	S	L	H	L	L	S	P
G	D	P	K	T	V	T	L	F	G	E	S	A	G	G	A	S	V	G	M	H	I	L	S	P
G	E	P	O	N	V	L	L	V	G	H	S	A	G	G	A	S	V	H	L	Q	M	L	R	E
G	N	P	D	S	V	T	I	F	G	E	S	A	G	G	V	S	V	S	A	L	V	L	S	P
G	N	P	E	W	M	T	L	F	G	E	S	A	G	S	S	S	V	N	A	Q	L	M	S	P
G	P	P	R	R	V	S	L	A	A	D	R	G	G	A	D	V	A	S	I	H	L	L	T	A
										 A G A F S V S A H L T S T														

Consensus

G d P . n v t l f g e s a G g a s v s l . l l s p

25 SEP 1998

Fig.6 (3/3)

6
 (208-231)
 BSSL
 (240-263)
 ChesHum
 (238-261)
 Torpace
 (223-246)
 Drosceh
 (226-249)
 Ratlivce
 (387-410)
 Drosace
 (2382-2402)
 ThryHum
 (150-173)
 Dict.Di

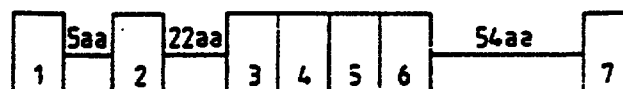
Y	N	K	G	L	I	R	R	A	I	S	Q	S	G	V	A	L	S	P	W	V	I	Q	K
G	S	H	S	L	F	T	R	A	I	L	Q	S	G	S	F	N	A	P	W	A	V	T	S
G	S	R	D	L	F	R	R	A	I	L	Q	S	G	S	P	N	C	P	W	A	S	V	S
D	F	G	Q	L	A	R	A	A	F	S	F	S	G	N	A	L	D	P	W	V	I	Q	K
L	A	K	N	L	F	H	R	A	I	S	E	S	G	V	V	L	T	T	N	L	D	K	K
V	T	R	G	L	V	K	R	G	M	M	Q	S	G	T	M	N	A	P	W	S	H	M	T
T	N	S	Q	L	F	R	R	A	V	L	M	G	G	S	A	L	S	P	A	A	V	I	S
Y	S	R	Q	Y	F	N	A	A	I	S	S	S	S	P	L	T	V	G	L	K	D	K	T

Consensus n k g L f r r a i . q s G s a l s p w a i q .

7
 (285-297)
 BSSL
 (316-328)
 ChesHum
 (314-326)
 Torpace
 (298-310)
 Drosceh
 (298-306)
 Ratlivce
 (466-477)
 Drosace
 (2458-2468)
 ThryHum
 (224-236)
 Dict.Di

V	G	F	V	P	V	I	D	G	D	F	I	P
V	N	F	G	P	T	V	D	G	D	F	L	T
F	S	F	V	P	V	I	D	G	E	F	F	P
A	P	F	S	P	V	L	E	P	S	D	A	P
.	.	.	.	T	V	I	D	G	V	V	L	P
P	S	.	A	P	T	I	D	G	A	F	L	P
.	.	W	G	P	V	I	D	G	H	F	L	R
T	I	W	S	P	V	I	D	G	D	A	F	I

Consensus . . f . p v . d g d f . p



10 30 50
AKLGAVYTEGGFVEGVNKKLGLLGDSVDIFKGIPFAAPTKALENPQPHPGWQGTILKAKNF
70 90 110
KKRCLQATITQDSTYGDEDCLYLNIVWPQGRKQVSRDLFVMIWIYGGAFLMGSGHGANFL
130 150 170
NNLYLDGEEIATRGNVIVVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKRNIAA
190 210 230
FGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVALSPWVIQKNPLFWAKKV
250 270 290
AEKVGCPVGDAARMAQCLKVIDPRALTILAYKVPLAGLEYPMLEHYGVFPVIDGDFIPADP
310 330 350
INLYANAADIDYIAGTNNMDGHIFASIDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGA
370 390 410
KTTFDVYTESWAQDPSQENKKKTIVVDFETDVLFLVPTEIALAQHRANAKSAKTYAYLFSH
430 450 470
PSRMPVYPKWVGADHADDIQYVFGKPFATPTGYRPQDRIVSKAMIAWYNFAKTGDPNMG
490 510 530
DSAVPTHWEPTYTENSYLEITKKMGSSSMKRSRLTNFLRYWTLTYLALPTVIDQEATPV
550 570 590
PPTGDS EATPV PPTGDS EATPV PPTGDS GAPPV PPTGDS GAPPV PPTGDS GAPPV PPTGDS
610 630 650
SGAPPV PPTGDS GAPPV PPTGDS GAPPV PPTGDS GAPPV PPTGDS GAPPV PPTGDS GAPPV
670 690 710
VPPTGDS GAPPV PPTGDS EATPV PPTGDS GAPPV PPTGDS EATPV PPTGDS EATPV PPTGDS