Pancreatology xxx (2017) 1–8

Contents lists available at ScienceDirect

Pancreatology

journal homepage: www.elsevier.com/locate/pan

The role of the carboxyl ester lipase (*CEL*) gene in pancreatic disease

Bente B. Johansson [a, b], Karianne Fjeld [a, c], Khadija El Jellas [a, d, e], Anny Gravdal [a, c, d], Monica Dalva [a, c, d], Erling Tjora [a, b], Helge Ræder [a, b], Rohit N. Kulkarni [f, g], Stefan Johansson [a, c], Pål R. Njølstad [a, b], Anders Molven [a, d, e, *]

[a] KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway
[b] Department of Pediatrics and Adolescent Medicine, Haukeland University Hospital, Bergen, Norway
[c] Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway
[d] Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Bergen, Norway
[e] Department of Pathology, Haukeland University Hospital, Bergen, Norway
[f] Islet Cell and Regenerative Biology, Joslin Diabetes Center, Harvard Medical School, Boston, MA, USA
[g] Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

Article history:
Received 17 September 2017
Received in revised form
28 November 2017
Accepted 1 December 2017
Available online xxx

Keywords:
Carboxyl ester lipase
MODY8
Chronic pancreatitis
Diabetes mellitus
Exocrine dysfunction

ABSTRACT

The enzyme carboxyl ester lipase (CEL), also known as bile salt-dependent or -stimulated lipase (BSDL, BSSL), hydrolyzes dietary fat, cholesteryl esters and fat-soluble vitamins in the duodenum. CEL is mainly expressed in pancreatic acinar cells and lactating mammary glands. The human *CEL* gene resides on chromosome 9q34.3 and contains a variable number of tandem repeats (VNTR) region that encodes a mucin-like protein tail. Although the number of normal repeats does not appear to significantly influence the risk for pancreatic disease, single-base pair deletions in the first VNTR repeat cause a syndrome of endocrine and exocrine dysfunction denoted MODY8. Hallmarks are low fecal elastase levels and pancreatic lipomatosis manifesting before the age of twenty, followed by development of diabetes and pancreatic cysts later in life. The mutant protein forms intracellular and extracellular aggregates, suggesting that MODY8 is a protein misfolding disease. Recently, a recombined allele between *CEL* and its pseudogene *CELP* was discovered. This allele (*CEL-HYB*) encodes a chimeric protein with impaired secretion increasing five-fold the risk for chronic pancreatitis. The *CEL* gene has proven to be exceptionally polymorphic due to copy number variants of the *CEL-CELP* locus and alterations involving the VNTR. Genome-wide association studies or deep sequencing cannot easily pick up this wealth of genetic variation. *CEL* is therefore an attractive candidate gene for further exploration of links to pancreatic disease.

Introduction

Carboxyl ester lipase (CEL, EC 3.1.1.13) is an enzyme present in all vertebrates examined to date [1–3]. The enzyme is one of four major lipases secreted by the pancreas to the duodenum, but only CEL has an absolute requirement for activation by bile salts [4–6]. The protein is therefore also referred to as bile salt-dependent lipase (BSDL) [7] or bile salt-stimulated lipase (BSSL) [8]. The official gene name assigned by the HUGO Gene Nomenclature Committee (www.genenames.org) is 'carboxyl ester lipase' with the synonym 'bile salt-stimulated lipase'. In the literature, also carboxyl ester hydrolase [1] as well as other names (e.g. cholesterol esterase, lysophospholipase) have been used. For simplicity, we will here refer to the protein as 'carboxyl ester lipase' with the unique gene symbol *CEL*.

The purpose of this paper is to review the role of carboxyl ester lipase in pancreatic disease with emphasis on genetic studies. For a broader introduction to the protein, the reader is referred to the two extensive reviews by Hui & Howles [5] and Lombardo [9].

The carboxyl ester lipase protein

The main expression site of CEL is the acinar cells of the pancreas (Fig. 1A and B), and the enzyme has been estimated to represent around 4% of total proteins detected in pancreatic juice [1]. CEL has

* Corresponding author. Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Bergen, Norway.
E-mail address: anders.molven@uib.no (A. Molven).

https://doi.org/10.1016/j.pan.2017.12.001
1424-3903/

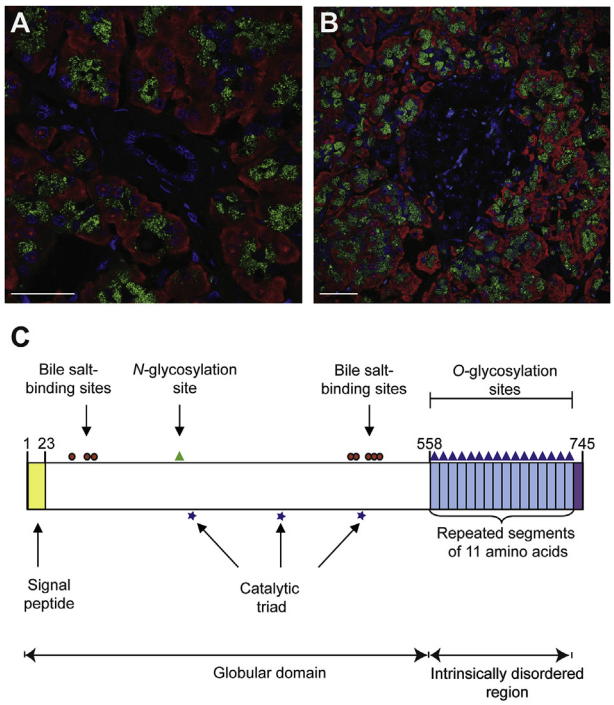

Fig. 1. Expression of human CEL in the acinar cells of normal pancreatic tissue surrounding a duct **(A)** and an islet of Langerhans **(B)**. Ducts and islets are negative. The images were produced by combined fluorescent *in situ* hybridization (red) and immunostaining (green), visualizing CEL mRNA and protein, respectively. For *in situ* hybridization, a probe covering exons 2–7 of the *CEL* gene was used (RNAscope 2.0 assay with alkaline phosphatase; Advanced Cell Diagnostics) in conjugation with Vector Red fluorescent substrate. For immunostaining, an anti-CEL antibody (Sigma HPA052701) was used, followed by treatment with FITC-conjugated, donkey anti-rabbit secondary antibody. Detailed protocol is available on request. Cell nuclei were stained blue with DAPI. Scale bars = 100 μm. **C)** Schematic structure of the CEL protein with functional domains and sites indicated. The drawing shows the most common variant with 16 tandem repeats in the C-terminal VNTR region. Numbers refer to amino acid position.

wide substrate specificity [10,11]. Stimulated by bile salts in the duodenum, it can hydrolyze dietary fat, cholesteryl esters and fat-soluble vitamins [1,5,11,12]. Of these activities, it is only the hydrolysis of cholesteryl esters that seems to be an activity that can be uniquely assigned to this enzyme in the digestive tract [5]. Moreover, it was recently suggested that CEL degrades branched fatty acid esters of hydroxyl fatty acids (FAHFAs), a newly discovered class of metabolites with anti-diabetic and anti-inflammatory properties [13].

CEL is also abundantly expressed in lactating mammary glands, being secreted with the mother's milk and estimated to comprise 1–2% of total milk protein [4,14]. In concert with pancreatic lipase-related protein 2, CEL may aid breastfed infants in the digestion of milk triglycerides and cholesteryl/retinyl esters [15,16]. In addition to the high levels in the pancreas and mammary glands, CEL expression have been documented in fetal liver [17], pituitary gland [18], macrophages [19], eosinophils [20] and endothelial cells [21].

The CEL protein (Fig. 1C) has two major structural domains: a globular N-terminal catalytic domain of 535 amino acid residues (excluding the signal peptide), followed by a C-terminal, intrinsically disordered and extended arm of repeated 11-amino acid segments [22,23]. The amino acid sequence of the N-terminal region including the catalytic triad Ser194-His435-Asp320 is well conserved in all vertebrate species examined to date [3]. This region also bears similarities to a number of other lipases and esterases [24,25].

The C-terminal region of CEL is comprised of 11-amino acid repeats, which are tailed by the unique sequence KEAQMPAVIRF. The region is enriched in the amino acids proline, glutamate, serine and threonine (so-called PEST sequences [26]) and bears similarities to mucinous proteins [27]. Due to a varying number of C-terminal repeats in the general population and differences in posttranslational modification, the observed molecular weight of the human CEL protein fluctuates significantly. The most common version of human CEL has 16 repeated segments, corresponding to a mature protein of 722 amino acids with a theoretical molecular mass of 79 kDa in the unmodified state. The number of 11-amino acid segments also varies considerably between species, spanning from no repeats in cod, frog, lizard and chicken, 3–4 repeats in rodents, 13 in dog, to as many as 39 repeats in gorilla (see comparison in Ref. [3]). CEL's C-terminus has been postulated to be important for secretion and activity [28]. However, other reports indicate that deletion of the repeat region has no effect on the catalytic properties of CEL [29,30].

After cleavage of the N-terminal hydrophobic signal peptide of 23 amino acid residues, the polypeptide chain is released into the lumen of ER where it forms a folding complex with several chaperones [9]. CEL is *N*-glycosylated at the conserved residue Asn187, a modification important for correct folding and secretion [31]. Subsequently, CEL is transported to the Golgi where the amino acid repeats of the C-terminus are heavily *O*-glycosylated [32]. PEST sequences, such as those found in the C-terminal region of CEL, can be a signal for rapid protein degradation [26]. *O*-glycosylation of the CEL repeats could possibly mask these PEST sequences, thereby serving to increase the stability of the enzyme by preventing proteolytic damage [33].

Once fully glycosylated in the *trans*-Golgi network, CEL is phosphorylated at Thr340, which allows final translocation through the secretory pathway [34], followed by co-storage with other digestive enzymes in zymogen granules. Finally, when secreted into the duodenal lumen, a substantial part of the CEL-molecules remain in complex with the molecular chaperone GRP94 [35]. It has been reported that a fraction of the CEL/GRP94 complex can be endocytosed from the intestines and transported to the blood compartment where it partly associates with apolipoprotein B-containing lipoproteins [36,37]. The enzyme is suggested to be cleared from the circulation by renal filtration and can be detected in the urine of healthy individuals [38].

CEL – the carboxyl ester lipase gene

The human *CEL* gene is located on chromosome band 9q34.13 [39], covers around 10 kilobase pairs (kb) of genomic sequence and consists of 11 exons (Fig. 2A) [40]. The last exon contains a very GC-rich variable number of tandem repeats (VNTR) region, made up of nearly identical 33-base pair (bp) segments that encode the 11-amino acid repeats of the protein tail. The VNTR makes the gene highly polymorphic, and repeat lengths between 3 and 23 have been observed [41]. The most frequent repeat numbers are those between 13 and 17, with 16 being the predominating length in all materials studied (Table 1) [41–44].

A *CEL*-like pseudogene (*CELP*) is located in tandem with the human *CEL* gene, about 11 kb downstream of the latter (Fig. 2A) [40,45]. *CELP* is reportedly transcribed in many tissues of the body, but it lacks exons 2–7, contains several base pairs different from *CEL* in the remaining exons, and is not expected to be translated into protein [46]. Apparently, the gene duplication event giving rise to *CEL* and *CELP* occurred relatively late in mammalian evolution, followed by a pseudogene-creating deletion in the lineage leading to great apes and man [47]. It has been suggested that *CELP* is, in fact, the original gene because the promoter region of the mouse *Cel* gene is most similar to that of the human *CELP* gene [45,48].

Please cite this article in press as: Johansson BB, et al., The role of the carboxyl ester lipase (*CEL*) gene in pancreatic disease, Pancreatology (2017), https://doi.org/10.1016/j.pan.2017.12.001