Research article

Proteolytic profiles of two isoforms of human AMBN expressed in *E. coli* by MMP-20 and KLK-4 proteases

Vetyskova Veronika [a], Hubalek Martin [a], Sulc Josef [a,b], Prochazka Jan [c],
Vondrasek Jiri [a,**], Kristyna Vydra Bousova [a,*]

[a] *Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo namesti 2, 16000, Prague, Czech Republic*
[b] *Department of Physical and Macromolecular Chemistry, Faculty of Natural Sciences, Charles University, Hlavova 8, 128 00 Prague 2, Czech Republic*
[c] *Institute of Molecular Genetics of the Czech Academy of Sciences, Videnska 5, 14000, Prague, Czech Republic*

A B S T R A C T

Ameloblastin is a protein in biomineralization of tooth enamel. However recent results indicate that this is probably not its only role in an organism. Enamel matrix formation represents a complex process enabled via specific crosslinking of two proteins – the most abundant amelogenin and the ameloblastin (AMBN). The human AMBN (hAMBN) gene possesses 13 protein coding exons with alternatively spliced transcripts and the longest isoform about 447 amino acid residues. It has been described that AMBN molecules *in vitro* assemble into oligomers via a sequence encoded by exon 5. Enamel is formed by the processing of enamel proteins by two specific proteases - enamelysin (MMP-20) and kallikrein 4 (KLK-4). The scaffold made of AMEL and non-amelogenin proteins is cleaved and removed from the developed tooth enamel. The hAMBN is expressed in two isoforms (ISO I and II), which could lead to their different utilization determined by distinct proteolytic profiles. In this study, we compared proteolytic profiles of both isoforms of hAMBN expressed in *E. coli* after proteolysis by MMP-20, KLK-4, and their 1:2 mixture. Proteolysis products were analysed and cleavage sites were identified by mass spectrometry. The proteolytic profiles of two AMBN isoforms showed different results, although we have to determine that the analysed AMBN was not post-translationally modified as expressed in prokaryotic cells. These results may lead to the suggestion of potentially divergent roles of AMBN isoforms cleavage products in various cell signalling pathways such as calcium buffering or signalling cascades.

1. Introduction

Enamel forms a specific chemical layer that creates a hard sheet on the dental surface. Enamel formation referred to as amelogenesis occurs inside the jaw before tooth hatches out [1–3]. Enamel is a mineralized tissue and its formation is co-processed by specific proteins and proteases [1,4–6]. The process specific proteins are synthesized by ameloblast cells into the enamel space where they nucleate and direct the growth of hydroxyapatite (HAP) crystals and their organization [2,4,5]. The enamel matrix is mainly

* Corresponding author.
** Corresponding author.
E-mail addresses: jiri.vondrasek@uochb.cas.cz (V. Jiri), kristyna.bousova@uochb.cas.cz (K. Vydra Bousova).

composed of two groups of proteins known as: amelogenins and non-amelogenins. The amelogenin (AMEL; Q99217) is the most abundant enamel matrix protein secreted by ameloblasts in 90% of the whole protein content [5,7]. Non-amelogenin proteins are represented by ameloblastin (Q9NP70), tuftelin (TUFT1; Q9NNX1), and enamelin (ENAM; Q9NRM1) and all of them fall into the category of intrinsically disordered proteins (IDPs), proteins lacking a cooperatively folded structure under the native conditions. Most of the proteins of the enamel matrix belong to secretory calcium-binding phosphoproteins [8–10]. Ameloblastin (AMBN) represents the most abundant enamel matrix protein from the non-amelogenin protein family [9,11,12]. AMBN also possesses additional functions such as the formation of long and cranial bones or its association with the progression of testicular germ cell tumour [13]. Defects in enamel formation lead to severe diseases, typically to the amelogenesis imperfecta, which is a group of genetically heterogenous hereditary disorders (characteristic for mutations in AMEL, AMBN, ENAM, MMP-20; O60882, KLK-4; Q9Y5K2, etc.) [14–16]. However, non-syndromic human amelogenesis imperfecta is caused only by AMBN mutations. Specifically by deletion of exon 6 of hAMBN [17].

AMBN is an IDP naturally assembling into size-heterogenous homo-oligomers (Fig. 1) [9,18,19]. The protein is expressed in two isoforms (ISO I: Q9NP70-1; and ISO II: Q9NP70-2) [9,19,20] that differ only by 15 amino acids (AAs) following the sequence encoded by exon 5 [9,21]. Human AMBN mRNA has been found to be expressed in various tissues, including those related to the nervous, reproductive, secretory, and internal systems [22]. AMBN was characterized and studied as a product of *E. coli* expression system [9, 18,23–26] therefore without native post-translational modifications such as phosphorylation or glycosylation. In AMBN several post-translational modifications have been identified (sulphated O-linked glycosylations of S112 and T387, hydroxylated prolines, and phosphorylations) [27]. The AMBN sequence also contains stretches of proline-rich motifs mostly present at the N-terminus, which could potentially lead to AMBN peptides signalling activity. Previous studies have shown that the oligomerization is guided by an evolutionary conserved motif encoded by exon 5 sequence: Y/F-x-x-Y/L/F-x-Y/F [8]. hAMBN consists of two parts based on physicochemical properties. The N-terminal part is enriched by proline-rich regions and has an approximately neutral charge. The total number of all positively charged residues in the N-terminal part is 12 and the total number of negatively charged residues is 15, which results in approximate neutral or more negative charge. The *C*-terminal part is massively negatively charged with 31 negatively charged residues [9,28], which has a relevancy to the calcium binding [9,10,29]. In addition, the calcium binding sites presented in AMBN along with heparin-binding domains and CD63 interaction domains can affect cell adhesion [13].

Proteases belonging to the non-amelogenin group are represented by matrix metalloproteinase-20 (MMP-20), also called enamelysin, and kallikrein 4 (KLK-4). MMP-20 is secreted during amelogenesis with other enamel proteins [25,29–31]. MMP-20 as a tooth-specific protease is expressed in ameloblasts in their secretory and early maturation stages [29,32,33]. This processing provides space for enamel crystallization. Another role of MMP-20 is the activation of KLK-4 in the presence of zymogen [32,34]. KLK-4, a serine protease also known as a tooth enamel matrix serine protease 1 [29], is expressed at the transition between the secretory and maturation stage and continues to be expressed throughout the maturation stages of enamel development [29,35]. KLK-4 has a unique ability to interact with enamel proteins, thereby providing a broad variety of target sequence specificity [36,37]. The apparent role of MMP-20 and KLK-4 is to cleave the scaffold made of enamel proteins: AMBN, AMEL, and ENAM [29,38]. AMEL cleavage was first described by Yamakoshi et al. [36]. They used porcine AMEL for the study, (same as Nagano et al. [39]), and reported that MMP-20 catalysed the major cleavage of AMEL [29]. Therefore we may expect that the main role of KLK-4 during enamel development is to process peptides already cleaved by MMP-20 [40]. An interesting feature of KLK-4 represents its activity against all enamel proteins, thus KLK-4 seems to have a much wider range of sequence specificity [41]. AMBN-derived cleavage products in the presence of MMP-20 and KLK-4 were also studied using porcine AMBN by Chun et al. [29]. Their study demonstrated that porcine AMBN is cleaved at the *N*-terminal part during the secretory phase of enamel formation [29].

Here we report a proteolytic profile of a recombinant hAMBN expressed in *E. coli*. The proteolysis was carried out using tooth-specific proteases MMP-20 and KLK-4 by acting in their 1:2 mixture (MMP-20/KLK-4). Our study is focused on a comparison of two isoforms of AMBN, their proteolytic profiles and the identification of cleavage sites by MMP-20 and KLK-4. Differences in proteolytic profiles of both isoforms may eventually lead to different products indicating their different roles. AMBN ISO I is longer than ISO II by 15 AA residues following exon 5 to our knowledge there is no study providing any evidence of the different functions of the
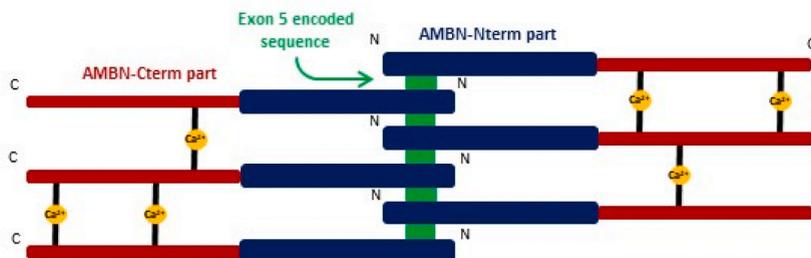

Fig. 1. Schematic view of a possible AMBN assembling into homo-oligomers. The blue rectangle illustrates AMBN N-terminal part. The green rectangle shows oligomerization induced by the sequence encoded by exon 5. The AMBN oligomerization is strictly driven by the sequence encoded by exon 5 without dependency on Ca^{2+} presence. The red lines show possible arrangements of the AMBN *C*-terminal part in the presence of Ca^{2+} (yellow circles). The scheme is modified from Vetyskova et al. [9]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

AMBN variants. The proteolytic profiles clearly indicated different cleavage products of AMBN isoforms and may therefore potentially lead to different functions.

2. Material and methods

2.1. Plasmid constructs of the AMBN ISO I and II

The expression vector encoding AMBN ISO I and ISO II (hAMBN: Q9NP70) was kindly donated by professor Michaela Rumlova from the University of Chemistry and Technology (Prague, Czechia). Both AMBN proteins were fused with three His-tags (6 x His) and a thioredoxin tag, which is a solubility enhancing tag and it prevents the accumulation of recombinant proteins in bacterial inclusion bodies. Both tags were designed to be cleavable by Tobacco Etch Virus (TEV) protease (TEVp: Q0GDU8). The cDNAs of the AMBN were subcloned to the pET28b vector (details [18]).

2.2. AMBN protein expression

AMBN ISO I and II were expressed as recombinant proteins in *Escherichia coli* BL21 (DE3) cells. The bacteria were grown at 37 °C in Luria-Bertani (LB) broth in a presence of kanamycin (30 μg/ml). For induction of expression, 0.5 mM isopropyl-β-ᴅ-thiogalactopyranoside was used and the cultivation continued for 20 h at 15 °C. Cells were harvested by centrifugation ($2500 \times g$, 20 min), resuspended in a phosphate buffer: 50 mM Na_2HPO_4 (pH = 8.0), 50 mM NaCl, 0.1% 2-mercaptoethanol, and stored at − 80 °C.

2.3. AMBN protein purification

Thawed cells were sonicated at 4 °C for 45 min (amplitude 30%, 5 s on, 10 s off) and centrifuged at $21,000 \times g$ for 1 h. The purified lysate with overexpressed recombinant AMBN was mixed with 8 M urea and separated by metal affinity chromatography with the usage of Chelating Sepharose Fast Flow resin (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The chelating slurry was charged with Ni^{2+} ions and equilibrated with (8 M urea, 50 mM Tris-HCl (pH = 7.4), 600 mM NaCl, and 20 mM imidazole) before the loading of proteins in the lysate. The His-tagged bounded proteins were eluted by buffer containing 8 M urea, 50 mM Tris-HCl (pH = 7.4), 600 mM NaCl and 600 mM imidazole. After the elution, AMBN was analysed by SDS-PAGE and refolded into a native state using dialysis in a buffer containing 50 mM Tris-HCl (pH = 7.4) and 500 mM NaCl. AMBN was mixed with in-house purified TEV protease [42] (2500 U per 10 mg of AMBN protein) kindly provided by Obsilova V. laboratory (Institute of Physiology of the Czech Academy of Sciences, Prague, Czechia) and the cleavage process was performed for 12 h at 4 °C and verified by the SDS-PAGE analysis. The AMBN mixture was purified by immobilized metal affinity (IMAC) chromatography on HisTrap HP column (Cytiva, Marlborough, USA) equilibrated by 10 mM Tris-HCl (pH = 7.4), 100 mM NaCl and the separation process was carried out on an ÄKTA pure purification system (Cytiva, Marlborough, USA). The proteins were eluted by the 10 mM Tris-HCl (pH = 7.4), 100 mM NaCl and analysed by SDS-PAGE and dialyzed against 10 mM Tris-HCl (pH = 7.4), 100 mM NaCl. The final AMBN protein was analysed by SDS-PAGE and mass spectrometry (MS). The mass of final recombinant AMBN proteins was confirmed by MS, for AMBN ISO I it was 46.71 kDa and for ISO II it was 45.00 kDa which corresponds to the calculated molecular weight.

2.4. Digestion of AMBN by MMP-20

Full-length AMBN (ISO I and ISO II) of 0.1 mg/ml was incubated with MMP-20 (MMP-20: O60882; Creative BioMart, Shirley, USA) according to the manufacturer's recommendations with a specific activity >50 U/μg. The MMP-20 manufacturer's concentration was 0.2 mg/ml and it was dissolved in 20 mM Tris (pH = 7.2), 10 mM $CaCl_2$, 0.1 mM $ZnCl_2$, 0.3 M NaCl, and 0.5 M Acetohydroxamic Acid. The incubation was performed for 4 h at 37 °C. Then the mixture was immediately measured by MS (viz. LC-MS/MS analysis). The cleavage of the AMBN was performed in 10 mM Tris-HCl (pH = 7.4), 100 mM NaCl.

2.5. Digestion of AMBN by KLK-4

The AMBN (ISO I and ISO II) about 0.5 mg/ml was incubated with KLK-4 (KLK4: Q9Y5K2) about 0.25 mg/ml at enzyme-to-substrate ratio 1:9 (Abcam, Cambridge, Great Britain) for 4 h at 37 °C. Then the mixture was immediately measured by MS (viz. LC-MS/MS analysis). The AMBN was digested in 10 mM Tris-HCl (pH = 7.4), 100 mM NaCl and the KLK-4 was in phosphate buffer saline.

2.6. Digestion of AMBN by a mixture of MMP-20 and KLK4

The AMBN (ISO I and ISO II) was incubated with a mixture of MMP-20 and KLK-4 for 4 h at 37 °C as well as in case of cleavage by a single protease (MMP-20 or KLK-4). The AMBN concentration was 0.5 mg/ml in presence of buffer containing 10 mM Tris-HCl (pH = 7.4), 100 mM NaCl. The digestion was performed by MMP-20 (Creative BioMart, Shirley, USA) and used according to the manufacturer's recommendations with a specific activity >50 U/μg. The KLK-4 was obtained from Abcam company (Cambridge, Great Britain) and used at enzyme-to-substrate ratio 1:9. The KLK-4 manufacture's concertation was 0.25 mg/ml and for MMP-20 it was 0.2 mg/ml.

2.7. LC-MS/MS analysis

The peptides after proteolytic digestion were separated on the UltiMate 3000 RSLCnano system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to the Mass Spectrometer Orbitrap Fusion Lumos (Thermo Fisher Scientific). The peptides were trapped and desalted with 2% acetonitrile in 0.1% formic acid at a flow rate of 30 µl/min on an Acclaim PepMap100 column (5 µm, 5 mm by 300-µm internal diameter (ID); Thermo Fisher Scientific). Eluted peptides were separated using an Acclaim PepMap100 analytical column (2 µm, 50-cm by 75-µm ID; ThermoFisher Scientific). The 30-min elution gradient at a constant flow rate of 300 nL/min was set to 5% phase B (0.1% formic acid in 99.9% acetonitrile) and 95% phase A (0.1% formic acid) for the first 1 min. Then, the content of acetonitrile was increased gradually. The orbitrap mass range was set from m/z 350 to 2000, in the MS mode, and the instrument acquired fragmentation spectra for ions of m/z 100 to 2000 in data dependent acquisition mode (DDA). The LC-MS/MS signal was processed with two procedures. One of the processing approaches involved the identification of peptides by standard proteomic identification of peptides involving the comparison of the fragmentation spectra of the peptide with the theoretical values of peptides sequences derived from AMBN (ISO I and ISO II) sequences. This step was performed in PEAKS Studio Xpro (Bioinformatics Solutions Inc., Ontario, Canada) [43]. The other processing step involved the analysis of only the MS signal combined over the time of the peptide elution (4–25 min) and automatic peptide signal deconvolution. This step generated a spectrum showing the peptide masses recorded in the sample intendent of retention time. This processing step was generated in FreeStyle 1.8 SP2 software (Thermo Fisher Scientific). The deconvoluted mass was assigned the peptide sequence identified in the first step of the processing. The data in PEAKS were also analysed in order to propose the probability of the MMP-20 or KLK-4 cleavage site.

The data presented in this study were measured by LC-MS/MS as is described above. All three experiments (digestion by MMP-20, KLK-4 and by their mixture) were repeated three times using UltrafleXtreme (Bruker Daltonics, Bremen, Germany), which is a MALDI-TOF Mass Spectrometer, to verify the reproducibility of the data.

3. Results

3.1. Preparation of AMBN cleavage products

Two AMBN isoforms (AMBN ISO I and ISO II) were expressed in *E. coli* as recombinant proteins with three individual His-tags located at the *N*- and *C*-termini of the protein. Purification of AMBN is quite complicated due to the formation of a heterogeneous mixture of homo-oligomers because of AMBN IDP character. This complication led to the construct containing a solubility-enhancing thioredoxin (Trx) tag at the *N*-terminus of both AMBN isoforms. Two Tobacco etch virus (TEV) protease cleavage sites were introduced in the design of recombinant AMBN isoforms to produce AMBN without tags after the purification procedure (in detail [9,18]). AMBN was purified by immobilized metal affinity chromatography (IMAC) in denaturing conditions (8 M urea) taking into account its oligomeric character. The protein was then re-natured and cleaved by TEV protease to obtain pure AMBN without the affinity and soluble tags. However, the efficiency of TEV cleavage was about 50% of cleaved AMBN reflecting most probably the oligomeric character of AMBN that prevents higher yields of pure protein.

Both recombinant AMBN isoforms were incubated with either MMP-20 or KLK-4 or with a mixture of both. Analysis of the products was performed by mass spectrometry (MS). The first experiment provided results of AMBN ISO I and ISO II cleaved with commercially prepared MMP-20 for 4 h and the mixture was promptly measured by LC-MS/MS. The second experiment provided results of AMBN isoforms cleaved for 4 h with commercially available KLK-4. The third experiment aimed to identify AMBN cleavage products with a mixture of MMP-20 and KLK-4 proteases for 4 h. The resulting peptides were separated on reverse phase and analysed by a mass spectrometer (LC-MS/MS). The aim was to identify the proteolytic products. The peptides of small or medium molecular weight (800–5000) are usually easy to fragment in collision cell (HCD) and assign the peptide sequence in the search algorithm. Peptides or

Fig. 2. Sequence of hAMBN ISO I (full-length sequence) and ISO II (the 15 AAs marked blue are missing in ISO II) with cleavage sites identified after treatment of MMP-20. Black arrows indicate unique cleavage sites in AMBN ISO I, blue ones in AMBN ISO II and red arrows mark cleavage sites the same for both AMBN isoforms cleaved with MMP-20. All of these cleavage sites were obtained using criteria of probability greater than 1.5%. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

proteins of higher M_W may not be possible to assign the sequence. This leads us to perform two stages of processing. First, we applied the PEAKS Studio Xpro algorithm [43] to identify the peptides. In addition, we processed the acquired data and deconvoluted the signal in order to get the molecular weight of the proteolytic products.

3.2. Proteolytic profiles of AMBN ISO I and ISO II by MMP-20 and KLK-4

Peptide fragmentation spectra of cleavage of recombinant AMBN ISO I and ISO II by MMP-20 proteases were processed using PEAKS Studio Xpro providing peptide coverage of AMBN with all detected peptides. These peptides were used to thoroughly detect cleavage sites in AMBN ISO I and ISO II sequences. Each of the cleavage sites was evaluated reflecting the probability of occurrences which was calculated from the proportion of peptides cleaved in a given cleavage site to the number of all peptides covering the cleavage site. The cleavage sites were identified based on a threshold over 1.5% and displayed in the sequence of AMBN ISO I (Fig. 2). AMBN ISO I cleavage resulted in 18 and ISO II in 12 MMP-20 cleavage sites.

Peptide profiles after proteolytic cleavage with one protease (MMP-20 or KLK-4) are shown in Supplementary Figs. S1 and S2. The LC-MS/MS data of the LC elution profile of all peptides were combined into one spectrum and deconvoluted. Fig. S1 shows the processing of AMBN data cleaved by MMP-20. The spectra were combined from 5 to 14 min for AMBN ISO I and from 4 to 14 min for AMBN ISO II. Fig. S2 presents proteolytic spectra after cleavage with KLK-4.

To summarize, there are 11 identical MMP-20 cleavage sites identified in both AMBN isoforms, shown in Fig. 2.: Q102-S103; A195-R196; L197-I198; L216-Y217; M247-A248; Y249-G250; F276-T277; L310-E311; D358-P359; N395-S396; L397-Q398. However, noticeable differences between the cleavage of hAMBN ISO I and ISO II can be reported. Only one cleavage site L16-S17 has been identified in AMBN ISO II and not in AMBN ISO I. Seven unique cleavage sites were determined for AMBN ISO I only: F100-L101; T112-A113; L129-Q130; A229-R230; A314-F315; E410-M411; M412-H413.

The processing of both AMBN isoforms by KLK-4 protease is shown in Fig. 3 and it is apparent that almost no cleavage event occurs for either AMBN ISO I nor for AMBN ISO II proteins. The position of the only cleavage site is at the very beginning of exon 5 – the part of the sequence responsible for oligomerization. The position of the identified cleavage sites has probably no effect to destroy the oligomerization of the processed AMBN molecules, on the other hand, the very N-termini of both isoforms is probably important for the kinetics of the oligomerization, is cleaved out.

The SDS-PAGE method was used to analyse the cleavage of AMBN by KLK-4. As shown in Fig. 4, after 4 h of digestion, both AMBN ISO I protein and KLK-4 protease were present.

3.2.1. Proteolytic profiles of AMBN ISO I and ISO II by a mixture of MMP-20 and KLK-4

The original task of the study was to identify cleavage profiles of both isoforms of AMBN by MMP-20 and KLK-4 separately. Chun et al. performed a similar experiment with porcine AMBN and porcine proteases [29] which provided an orientation in proteolytic profiles based on high sequence similarity of both AMBN versions (human and porcine). In comparison with acquired spectra, Chun et al. identified more cleavage sites by KLK-4 giving rise to a question of used experimental setups and their similarity. To at least address the question of differences between both approaches we decided to use a complementary strategy. We exposed both AMBN isoforms to an equimolar mixture of MMP-20 and KLK-4 in a 1:2 ratio and the cleavage experiment was carried out for the same time as for the independent cleavage by one of the proteases for 4 h.

To verify that both proteases remain in solution after 4 h, the mixture of the proteases in the same ratio using the same experimental conditions was analysed (4 h, 37 °C) by SDS-PAGE. The MMP-20 is not visible due to its low concentration, however the KLK-4 is still visible after 4 h. Results are shown in Supplemental Fig. S4.

The products of proteolysis were analysed by LC-MS/MS and each chromatogram was combined into one spectrum and deconvoluted (Fig. S5). The deconvoluted spectra were focused on retention time from 5 to 22 min. The peptide fragmentation spectra of the processed AMBN ISO I and ISO II were processed in PEAKS Studio Xpro and peptide coverage of the AMBN by detected peptides was

[1]VPFFPQQSGTPGMASLSLETMRQLGSLQRLNTLSQYSRYGFGKSFNSLWMHGLLPPHSSL[60]

[61]PWMRPREHETQQYEYSLPVHPPPLPSQPSLKPQQPGLKPFLQSAAATTNQATALKEALQP[120]

[121]PIHLGHLPLQEGELPLVQQQVAPSDKPPKPELPGVDFADPQGPSLPGMDFPDPQGPSLPG[180]

[181]LDFADPQGSTIFQIARLISHGPMPQNKQSPLYPGMLYVPFGANQLNAPARLGIMSSEEVA[240]

[241]GGREDPMAYGAMFPGFGGMRPGFEGMPHNPAMGGDFTLEFDSPVAATKGPENEEGGAQGS[300]

[301]PMPEANPDNLENPAFLTELEPAPHAGLLALPKDDIPGLPRSPSGKMKGLPSVTPAAADPL[360]

[361]MTPELADVYRTYDADMTTSVDFQEEATMDTTMAPNSLQTSMPGNKAQEPEMMHDAWHFQEP[421]

Fig. 3. Sequence of hAMBN ISO I (full-length sequence) and ISO II (the 15 AAs marked blue are missing in ISO II) with identified cleavage sites of AMBN ISO I and ISO II by KLK-4. Cleavage sites in AMBN ISO I are marked by a black arrow, blue arrow indicates cleavage site in AMBN ISO. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
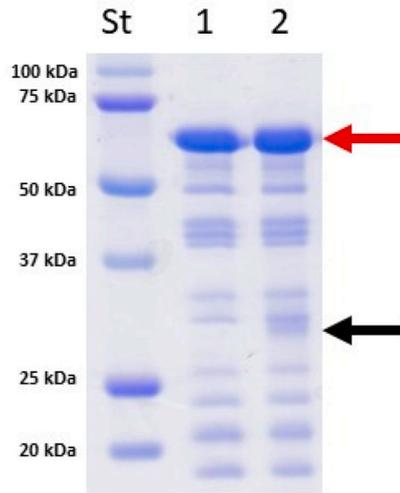
Fig. 4. Analysis of AMBN cleavage with KLK-4 for 4 h by SDS-PAGE using 12% polyacrylamide gel stained with Coomassie Blue (magnified, see original gel in Supplementary material as Fig. S3). Lane St represents molecular mass standards; lane 1: AMBN ISO I without KLK-4; lane 2, AMBN ISO I with KLK4 after 4 h of digestion. Red arrow indicates the AMBN protein. Black arrow indicates the KLK-4 protease. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

obtained. Identified peptides were used to determine the positions of cleavage sites based on criteria of probability threshold over 1.5% (Fig. 5). For AMBN ISO I, 14 cleavage sites were identified, and for AMBN ISO II there were 17 cleavage sites identified. Consequently, we compared the results of proteolysis for porcine and hAMBN by MMP-20 and KLK-4 (Figs. S6 and S7). Prevailing numbers of identical cleavage sites were identified; however, a distinct cleavage pattern was observed at the *C*-terminus of the hAMBN.

4. Discussion

The processing of hAMBN by the proteases MMP-20 and KLK-4 during amelogenesis represents a crucial step for the formation of

Fig. 5. Sequence of hAMBN ISO I (full-length sequence) and ISO II (the 15 AAs marked blue are missing in ISO II) with identified cleavage sites generated by cleavage of a mixture of MMP-20 and KLK-4. Black arrows indicate cleavage sites generated from AMBN ISO I, blue AMBN ISO II, and red both AMBN isoforms cleaved by a mixture of MMP-20 and KLK-4. Arrows with stars indicate cleavage sites that were created by cleavage of AMBN isoforms with a mixture of MMP-20 and KLK-4, but these sites were not detected after cleavage of AMBN ISO I and ISO II by MMP-20 or by KLK-4 separately. Sequence of porcine AMBN is marked by grey. Pink arrows correspond to the cleavage of porcine AMBN by porcine MMP-20. Yellow arrows correspond to the cleavage of porcine AMBN by porcine KLK-4. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

fully developed enamel and the withdrawal of the organic part from the fully organized mineral layer [37,44]. However, the role of the products of the proteolysis remains questionable. Using bioinformatics methods, many sites in the AMBN sequence are predicted to play a role in cellular or extracellular processes (Fig. 6). Based on our previous study [45], we can localize short linear motifs in hAMBN sequence regarding their further role based on such predictions. The motifs are mostly involved in protein–protein interactions. The proteins contain multiple SH2 or SH3 domain interaction sites (114–121, 131–137, and 131–138 in AMBN). In addition to protein interaction surfaces, the conserved sequence also possesses multiple sites for phosphorylation from signalling regulatory kinase. This kinase is phosphoinositide 3-kinase-related kinase (PIKK) [46] and is predicted to phosphorylate motifs on positions 62–68, 98–104, and 380–386 in AMBN. A calcineurin binding site motif located at 84–87 and 216–219 of AMBN may be involved in dephosphorylation.

The recombinant AMBN was already studied as a product of *E. coli* expression as in this project [9,18,23–26], therefore the interpretation of the data has to consider potential differences of AMBN proteolytic profile when protein purified from eukaryotic expression system. It has previously been observed that AMBN isolated from porcine enamel harbors several post-translational modifications, including hydroxylation of P11 and P324, phosphorylation of S17, and O-linked glycosylation of S86 and T361, respectively [47].

Processing of AMBN isoforms by MMP-20 alone confirmed our expectation regarding different proteolytic products. Both isoforms of hAMBN, ISO I, and ISO II, share 11 identical cleavage sites (Q102-S103; A195-R196; L197-I198; L216-Y217; M247-A248; Y249-G250; F276-T277; L310-E311; D358-P359; N395-S396; L397-Q398) that are distributed along the entire sequences of AMBN isoforms (Fig. 2). Interestingly, the processing of ISO I of the human AMBN provided an extra seven cleavage sites, four of them located at the *N*-terminus of the AMBN and three of them at the *C*-terminus. These positions are unique for AMBN ISO I and there is no identity with AMBN ISO II cleavage sites. On the contrary, the only extra cleavage site produced by proteolysis of AMBN ISO II by MMP-20 takes place at the very end of the *N*-terminus of the AMBN sequence and is not cleaved by MMP-20 in AMBN ISO I.

The overall outcome of proteolysis of recombinant hAMBN ISO I and ISO II by human KLK-4 was unexpectedly low. Only one cleavage site was identified in both hAMBN isoforms and these were not identical for both isoforms (Fig. 3) Comparison of cleavage sites in both isoforms of recombinant hAMBN (Figs. 5 and S7) by human KLK-4 with those published by Chun et al. [29] with porcine AMBN and porcine KLK-4 provides an interesting result. There is only one site in each of the hAMBN isoforms and this position between K43-S44 in hAMBN ISO I

correspond to the same site in porcine AMBN. The AMBN ISO II cleavage site is positioned closer to the *N*-terminus and it is located between R38-Y39.

Many studies [29,34,37] describe that MMP-20 is first produced during a tooth enamel formation and followed by the action of KLK-4 protease. MMP-20 is expressed at the secretory and early maturation stages of ameloblasts while KLK-4 is mostly expressed during the maturation stage [37]. Nevertheless, both proteases MMP-20 and KLK-4 are expressed while the tooth enamel matures and they have to act simultaneously, so we decided to mimic the effect of both proteases on AMBN by their mixture solution. Furthermore, our study revealed that the cooperative action of both proteases on hAMBN resulted in additional cleavage sites not observed in experiments with separated proteases processing either ISO I or ISO II isoform. Identical positions obtained by cleavage of either ISO I or ISO II or by a mixture of proteases were low. No cleavage sites identified by KLK-4 are recognized by a mixture of both proteases, three of the identified cleavage sites by MMP-20/KLK-4 were the same as for the processing of AMBN ISO I by MMP-20 alone. One additional MMP-20/KLK-4 cleavage site was identified for ISO II only corresponding with previously identified cleavage site by MMP-20 found for both isoforms. All of the other identified cleavage sites by MMP-20/KLK-4 were unique and never observed for products obtained by the action of particular protease – 10 for ISO I and 12 for ISO II – 7 of them shared in both isoforms. To summarize – ISO I is cleaved at 14 positions, ISO II at 9 positions suggesting different proteolytic profiles. It is apparent that the level of synergy of
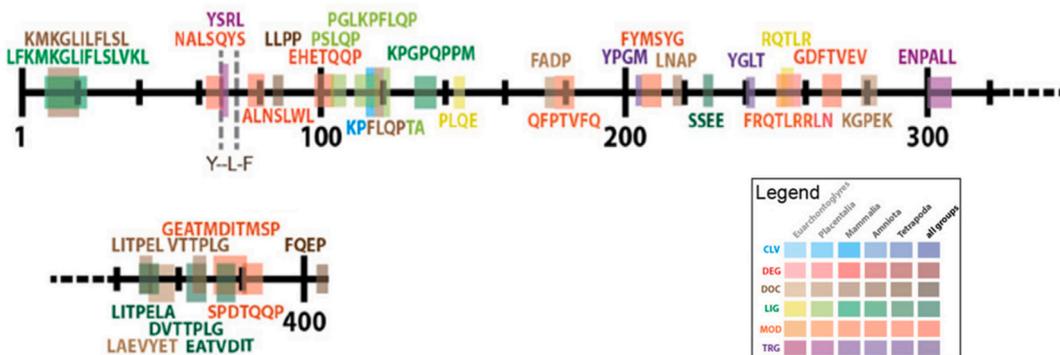

Fig. 6. Identification of functional motifs of extracellular matrix proteins. Detected functional motifs of AMBN (*N*-terminus to left, *C*-terminus to right. The position and motif structure correspond to *Mus musculus* sequences for protein AAB93765.1. (with permission of Spoutil et al. [45]). Colour correspond to the motif type (CLV = cleavage, DEG = degradation, DOC = docking, LIG = ligand, MOD = modifier, TRG = transport). Colour saturation corresponds to the minimal age of the motif. Which is for how large taxonomical unit is it common (Euarchontoglyres < Placentalia < Mammalia < Amniota < Tetrapoda < Sarcopterygii/Euteleostomi). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

both proteases in AMBN processing is an important factor in AMBN proteolysis. The most interesting question is what is the timing of proteolytic action of both proteases and the evolution of the process regarding the increasing concentration of KLK-4 [48].

Previous work on porcine AMBN cleavage has been published by Chun et al. [29]. It is necessary to say that the published results were not generated by cleavage with a mixture of MMP-20 and KLK-4. Instead, both proteases separately processed the full sequence of the porcine AMBN. The sequence of porcine AMBN is shorter in comparison with hAMBN (two substantial deletions, percentage of identity is 64%), so the identification of cleaved products and their identity in both AMBNs was an interesting question. To simplify the comparison of porcine and human cleavage profiles the alignment of both sequences using NLM BLAST [49] was performed. As follows from the alignment visible in Fig. 5, there are two deletions in porcine AMBN localized almost symmetrically at both *C*- and *N*- termini parts of the porcine AMBN, still preserving different characters of *N*- and -*C* domains from physical-chemical properties point of view. An interesting observation in cleavage profiles is the processing of *N*-terminal parts of both AMBN - human and porcine. Three of five cleavage sites identified in porcine AMBN sequence either by MMP-20 or KLK-4 are located at the *N*-terminus between position 25–50, one of these sites is identical with hAMBN cleavage site (M21-R22) obtained by processing of mixture of MMP-20/KLK-4, the other two are at some extent corresponding to cleavage site in hAMBN at *N*-terminal domain on both isoforms by human KLK-4 only. There are only two identified cleavage sites in both AMBN isoforms localized in similar sequence environments at the *C*-terminal domain (M247-A248; L310-E311; A314-F315). The obtained results give rise a question on both proteases specificity and the role of the processed peptides in consequent biological processes.

Fig. 6 shows cleavage sites that are generated by controlled proteolysis of porcine AMBN with either MMP-20 or KLK-4 as well as proteolysis of hAMBN by a mixture of human MMP-20 and human KLK-4. The cleavage site which is generated by KLK-4 processing hAMBN ISO II is located between Q28-R29 exactly matches the cleavage site created by porcine KLK-4 on porcine AMBN (Q28-G29) and makes this site really noteworthy because it is the only site being cleaved in both porcine and hAMBN. Other cleavage sites are distributed approximately to the porcine AMBN corresponding regions but no identity was detected. The first corresponding region can be found in both hAMBN isoforms starting with positions G42-K43 by a mixture of MMP-20/KLK-4 and a similar pattern is provided by porcine KLK-4 on porcine AMBN. The second region of similar processing corresponds to the part on hAMBN starting with position M247-A248. Finally, the third one starting with T371-Y372 corresponds to porcine KLK-4 and was detected only in the case of hAMBN ISO I.

From all the results, we can conclude that KLK-4 alone can digest at the *N*-terminal part of AMBN. The MMP-20 itself is probably not able to process the *N*-terminal part of AMBN which could eventually lead to the dissociation of AMBN oligomers and exposure of protein parts to be accessible for further processing. Processing of AMBN *N*-terminus particularly stabilized by exon 5 (36–72) during oligomerization could be probably a reason for distinct proteolytic profiles of hAMBN caused by the action of both proteases simultaneously.

Accession codes

AMBN: Q9NP70
AMEL: Q99217
ENAM: Q9NRM1
TEVp: Q0GDU8
TUFT1: Q9NNX1
MMP-20: O60882
KLK4: Q9Y5K2

Data availability statement

Raw data are available in the hyperlink https://zenodo.org/badge/DOI/10.5281/zenodo.7981657.svg; https://doi.org/10.5281/zenodo.7981657.

CRediT authorship contribution statement

Vetyskova Veronika: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Hubalek Martin:** Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis, Data curation. **Sulc Josef:** Validation, Formal analysis, Data curation. **Prochazka Jan:** Visualization, Validation, Investigation, Data curation. **Vondrasek Jiri:** Writing – original draft, Validation, Supervision, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Kristyna Vydra Bousova:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Abbreviations

AA amino acid
AMBN Ameloblastin
AMBN ISO I Ameloblastin isoform I
AMBN ISO II Ameloblastin isoform II
ENAM Enamelin
hAMBN human Ameloblastin
HAP hydroxyapatite
MMP-20 Enamelysin
MS mass spectrometry
KLK-4 Kallikrein-4
LC liquid chromatography
IMAC immobilized metal affinity chromatography
SDS-PAGE sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TEV Tobacco etch virus

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24564.

References

[1] A. Gil-Bona, F.B. Bidlack, Tooth enamel and its dynamic protein matrix, Int. J. Mol. Sci. 21 (12) (2020) 4458.
[2] R.S. Lacruz, S. Habelitz, J.T. Wright, M.L. Paine, Dental enamel formation and implications for oral health and disease, Physiol. Rev. 97 (3) (2017) 939–993.
[3] D. Heller, E.J. Helmerhorst, F.G. Oppenheim, Saliva and serum protein exchange at the tooth enamel surface, J. Dent. Res. 96 (4) (2017) 437–443.
[4] H. Margolis, E. Beniash, C. Fowler, Role of macromolecular assembly of enamel matrix proteins in enamel formation, J. Dent. Res. 85 (9) (2006) 775–793.
[5] M. Zeichner-David, Is there more to enamel matrix proteins than biomineralization? Matrix Biol. 20 (5–6) (2001) 307–316.
[6] V.P. Thompson, N.R.F.A. Silva, in: P. Vallittu (Ed.), 1 - Structure and properties of enamel and dentin, Non-Metallic Biomaterials for Tooth Repair and Replacement, Woodhead Publishing, 2013, pp. 3–19.
[7] J.-Y. Sire, S. Delgado, D. Fromentin, M. Girondot, Amelogenin: lessons from evolution, Arch. Oral Biol. 50 (2) (2005) 205–212.
[8] T. Wald, F. Spoutil, A. Osickova, M. Prochazkova, O. Benada, P. Kasparek, L. Bumba, O.D. Klein, R. Sedlacek, P. Sebo, J. Prochazka, R. Osicka, Intrinsically disordered proteins drive enamel formation via an evolutionarily conserved self-assembly motif, Proc. Natl. Acad. Sci. USA 114 (9) (2017) E1641–E1650.
[9] V. Vetyskova, M. Zouharova, L. Bednarova, O. Vaněk, P. Sázelová, V. Kašička, J. Vymetal, J. Srp, M. Rumlová, T. Charnavets, K. Postulkova, J.E. Reseland, K. Bousova, J. Vondrasek, Characterization of AMBN I and II isoforms and study of their Ca2+-binding properties, Int. J. Mol. Sci. 21 (23) (2020) 9293.
[10] O. Stakkestad, S.P. Lyngstadaas, B. Thiede, J. Vondrasek, B.S. Skalhegg, J.E. Reseland, Phosphorylation modulates ameloblastin self-assembly and Ca2+ binding, Front. Physiol. 8 (2017) 10.
[11] T. Lu, M. Li, X. Xu, J. Xiong, C. Huang, X. Zhang, A. Hu, L. Peng, D. Cai, L. Zhang, Whole exome sequencing identifies an AMBN missense mutation causing severe autosomal-dominant amelogenesis imperfecta and dentin disorders, Int. J. Oral Sci. 10 (3) (2018) 1–9.
[12] R.M. Wazen, P. Moffatt, S.F. Zalzal, Y. Yamada, A. Nanci, A mouse model expressing a truncated form of ameloblastin exhibits dental and junctional epithelium defects, Matrix Biol. 28 (5) (2009) 292–303.
[13] T. Geng, C.A. Heyward, X. Chen, M. Zheng, Y. Yang, J.E. Reseland, Comprehensive analysis identifies ameloblastin-related competitive endogenous RNA as a prognostic biomarker for testicular germ cell tumour, Cancers 14 (8) (2022) 1870.
[14] D. Ozdemir, P. Hart, O. Ryu, S. Choi, M. Ozdemir-Karatas, E. Firatli, N. Piesco, T. Hart, MMP20 active-site mutation in hypomaturation amelogenesis imperfecta, J. Dent. Res. 84 (11) (2005) 1031–1035.
[15] M.M.I. Sabandal, E. Schäfer, Amelogenesis imperfecta: review of diagnostic findings and treatment concepts, Odontology 104 (3) (2016) 245–256.
[16] V. Simancas-Escorcia, A. Natera, M.G. Acosta-de-Camargo, Genes involved in amelogenesis imperfecta. Part I, Rev. Fac. Odontol. Univ. Antioquia 30 (1) (2018) 105–120.
[17] J.A. Poulter, G. Murillo, S.J. Brookes, C.E.L. Smith, D.A. Parry, S. Silva, J. Kirkham, C.F. Inglehearn, A.J. Mighell, Deletion of ameloblastin exon 6 is associated with amelogenesis imperfecta, Hum. Mol. Genet. 23 (20) (2014) 5317–5324.
[18] T. Wald, A. Osickova, M. Sulc, O. Benada, A. Semeradtova, L. Rezabkova, V. Veverka, L. Bednarova, J. Maly, P. Macek, Intrinsically disordered enamel matrix protein ameloblastin forms ribbon-like supramolecular structures via an N-terminal segment encoded by exon 5, J. Biol. Chem. 288 (31) (2013) 22333–22345.
[19] T. Wald, L. Bednarova, R. Osicka, P. Pachl, M. Sulc, S.P. Lyngstadaas, I. Slaby, J. Vondrasek, Biophysical characterization of recombinant human ameloblastin, Eur. J. Oral Sci. 119 (2011) 261–269.
[20] M. MacDougall, D. Simmons, T.T. Gu, K. Forsman-Semb, C. Kärrman Mårdh, M. Mesbah, N. Forest, P.H. Krebsbach, Y. Yamada, A. Berdal, Cloning, characterization and immunolocalization of human ameloblastin, Eur. J. Oral Sci. 108 (4) (2000) 303–310.
[21] T. Wald, L. Bednarova, R. Osicka, P. Pachl, M. Sulc, S.P. Lyngstadaas, I. Slaby, J. Vondrasek, Biophysical characterization of recombinant human ameloblastin, Eur. J. Oral Sci. 119 (Suppl 1) (2011) 261–269.
[22] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T.I. Stein, R. Nudel, I. Lieder, Y. Mazor, The GeneCards suite: from gene data mining to disease genome sequence analyses, Current protocols in bioinformatics 54 (1) (2016) 1, 30. 1-1.30. 33.

[23] J. Su, R.A. Bapat, J. Moradian-Oldak, The expression and purification of recombinant mouse ameloblastin in E. coli, Methods Mol. Biol. (Clifton, NJ) 1922 (2019) 229.
[24] R.A. Bapat, J. Su, J. Moradian-Oldak, Co-immunoprecipitation reveals interactions between amelogenin and ameloblastin via their self-assembly domains, Front. Physiol. (2020) 1721.
[25] C. Shao, R.A. Bapat, J. Su, J. Moradian-Oldak, Regulation of Hydroxyapatite Nucleation in Vitro through Ameloblastin–Amelogenin Interactions, ACS Biomaterials Science & Engineering, 2022.
[26] V. Vetyskova, M. Zouharova, K. Bousova, Production of recombinant human ameloblastin by a fully native purification pathway, Protein Expr. Purif. 198 (2022) 106133.
[27] N. Kobayashi, R. Arai, Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks, Curr. Opin. Biotechnol. 46 (2017) 57–65.
[28] M. Zouharova, J. Vymetal, L. Bednarova, O. Vanek, P. Herman, V. Vetyskova, K. Postulkova, P.S. Lingstaadas, J. Vondrasek, K. Bousova, Intrinsically disordered protein domain of human ameloblastin in synthetic fusion with calmodulin increases calmodulin stability and modulates its function, Int. J. Biol. Macromol. 168 (2021) 1–12.
[29] Y. Chun, Y. Yamakoshi, F. Yamakoshi, M. Fukae, J. Hu, J. Bartlett, J. Simmer, Cleavage site specificity of MMP-20 for secretory-stage ameloblastin 89 (8) (2010) 785–790.
[30] Y. Yamakoshi, J.C.C. Hu, M. Fukae, F. Yamakoshi, J.P. Simmer, How do enamelysin and kallikrein 4 process the 32-kDa enamelin? Eur. J. Oral Sci. 114 (2006) 45–51.
[31] S. Shahid, J.D. Bartlett, in: M. Goldberg, P. Den Besten (Eds.), Proteinases in enamel development, Extracellular Matrix Biomineralization of Dental Tissue Structures, Springer International Publishing, Cham, 2021, pp. 261–270.
[32] Y. Yamakoshi, J.P. Simmer, J.D. Bartlett, T. Karakida, S. Oida, MMP20 and KLK4 activation and inactivation interactions in vitro, Arch. Oral Biol. 58 (11) (2013) 1569–1577.
[33] K. Koli, G. Saxena, K.U. Ogbureke, Expression of matrix metalloproteinase (MMP)-20 and potential interaction with dentin sialophosphoprotein (DSPP) in human major salivary glands, J. Histochem. Cytochem. 63 (7) (2015) 524–533.
[34] O. Ryu, J.C.C. Hu, Y. Yamakoshi, J.L. Villemain, X. Cao, C. Zhang, J.D. Bartlett, J.P. Simmer, Porcine kallikrein-4 activation, glycosylation, activity, and expression in prokaryotic and eukaryotic hosts, Eur. J. Oral Sci. 110 (5) (2002) 358–365.
[35] J.P. Simmer, J.C.C. Hu, Expression, structure, and function of enamel proteinases, Connect. Tissue Res. 43 (2–3) (2002) 441–449.
[36] Y. Yamakoshi, T. Tanabe, M. Fukae, M. Shimizu, Porcine amelogenins, Calcif. Tissue Int. 54 (1) (1994) 69–75.
[37] Y. Lu, P. Papagerakis, Y. Yamakoshi, J. Hu, J. Bartlett, J. Simmer, Functions of KLK4 and MMP-20 in dental enamel formation, Biol. Chem. 389 (2008) 695–700.
[38] J.P. Simmer, A.S. Richardson, C.E. Smith, Y. Hu, J.C.C. Hu, Expression of kallikrein-related peptidase 4 in dental and non-dental tissues, Eur. J. Oral Sci. 119 (2011) 226–233.
[39] T. Nagano, A. Kakegawa, Y. Yamakoshi, S. Tsuchiya, J.C.-C. Hu, K. Gomi, T. Arai, J.D. Bartlett, J.P. Simmer, Mmp-20 and Klk4 cleavage site preferences for amelogenin sequences, J. Dent. Res. 88 (9) (2009) 823–828.
[40] Y. Yamakoshi, A.S. Richardson, S.M. Nunez, F. Yamakoshi, R.N. Milkovich, J.C.C. Hu, J.D. Bartlett, J.P. Simmer, Enamel proteins and proteases in Mmp20 and Klk4 null and double-null mice, Eur. J. Oral Sci. 119 (2011) 206–216.
[41] E. Del Nery, J.R. Chagas, M.A. Juliano, L. Juliano, E.S. Prado, Comparison of human and porcine tissue kallikrein substrate specificities, Immunopharmacology 45 (1) (1999) 151–157.
[42] P.G. Blommel, B.G. Fox, A combined approach to improving large-scale production of tobacco etch virus protease, Protein Expr. Purif. 55 (1) (2007) 53–68.
[43] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry, Rapid Commun. Mass Spectrom. 17 (20) (2003) 2337–2342.
[44] R.S. Lacruz, Enamel: molecular identity of its transepithelial ion transport system, Cell Calcium 65 (2017) 1–7.
[45] F. Spoutil, G. Aranaz-Novaliches, M. Prochazkova, T. Wald, V. Novosadova, P. Kasparek, R. Osicka, J.E. Reseland, S.P. Lyngstadaas, H. Tiainen, K. Bousova, J. Vondrasek, R. Sedlacek, J. Prochazka, Early evolution of enamel matrix proteins is reflected by pleiotropy of physiological functions, Sci. Rep.-Uk 13 (1) (2023) 1471.
[46] R.T. Abraham, PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways, DNA Repair 3 (8) (2004) 883–887.
[47] T. Wald, A. Osickova, M. Sulc, O. Benada, A. Semeradtova, L. Rezabkova, V. Veverka, L. Bednarova, J. Maly, P. Macek, P. Sebo, I. Slaby, J. Vondrasek, R. Osicka, Intrinsically disordered enamel matrix protein ameloblastin forms ribbon-like supramolecular structures via an N-terminal segment encoded by exon 5, J. Biol. Chem. 288 (31) (2013) 22333–22345.
[48] Y. Lu, P. Papagerakis, Y. Yamakoshi, J.C.-C. Hu, J.D. Bartlett, J.P. Simmer, Functions of KLK4 and MMP-20 in Dental Enamel Formation, 2008.
[49] E.W. Sayers, E.E. Bolton, J.R. Brister, K. Canese, J. Chan, D.C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B.W. Trawick, K.D. Pruitt, S.T. Sherry, Database resources of the national center for biotechnology information, Nucleic Acids Res. 50 (D1) (2022) D20–d26.