Identification of Food-grade Subtilisins as Gluten-degrading Enzymes to Treat Celiac Disease

Guoxian Wei1, #, Na Tian1, #, Roland Siezen2, Detlef Schuppan3,4, Eva J. Helmerhorst1,*

1 Department of Molecular and Cell biology, Henry M. Goldman School of Dental Medicine, Boston, MA 02118; 2Bacterial Genomics Group, Center for Molecular and Biomolecular Informatics, Radboud University Medical Centre, Nijmegen, the Netherlands, 3Division of Gastroenterology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215; 4 Institute of Translational Immunology and Research Center for Immunology (FZI), University Medical Center, Johannes-Gutenberg-University, 55131 Mainz, Germany

Short title: Subtilisins for Celiac Disease

Keywords: Celiac disease; gluten, subtilisin, Rothia, therapy, Bacillus

* Corresponding author
Eva J. Helmerhorst, M.S., Ph.D.
Dept. of Molecular and Cell Biology
Boston University
Henry M. Goldman School of Dental Medicine
700 Albany Street, CABR W202B
Boston, MA 02118
Phone: 617-414-1119
Fax: 617-638-4924
Email: helmer@bu.edu

# these authors contributed equally to this paper

Copyright © 2016 by the American Physiological Society.
Gluten are proline- and glutamine-rich proteins present in wheat, barley and rye, and contain the immunogenic sequences that drive celiac disease (CD). *Rothia mucilaginosa*, an oral microbial colonizer, can cleave these gluten epitopes. The aim was to isolate and identify the enzymes and evaluate their potential as novel enzyme therapeutics for CD. The membrane-associated *R. mucilaginosa* proteins were extracted and separated by DEAE chromatography. Enzyme activities were monitored with paranitroanilide derivatized- and FRET peptide substrates, and by gliadin zymography. Epitope elimination was determined in R5 and G12 ELISA assays. The gliadin-degrading *Rothia* enzymes were identified by LC-ESI-MS/MS as hypothetical proteins ROTMU0001_0241 (C6R5V9_9MICC), ROTMU0001_0243 (C6R5W1_9MICC) and ROTMU0001_240 (C6R5V8_9MICC). A blast search revealed that these are subtilisin-like serine proteases belonging to the peptidase S8 family. Alignment of the major *Rothia* subtilisins indicated that all contain the catalytic triad with Asp (D), His (H) and Ser (S) in the D-H-S order. They cleaved succinyl-Ala-Ala-Pro-Phe-paranitroanilide, a substrate for subtilisin with Pro in the P2 position, as in Tyr-Pro-Gln and Leu-Pro-Tyr in gluten, which are also cleaved. Consistently, FRET substrates of gliadin immunogenic epitopes comprising Xaa-Pro-Xaa motives were rapidly hydrolyzed. The *Rothia* subtilisins as well as two subtilisins from *Bacillus licheniformis*, subtilisin A and the food-grade Nattokinase efficiently degraded the immunogenic gliadin-derived 33-mer peptide as well as the immunodominant epitopes recognized by the R5 and G12 antibodies. This study identified *Rothia* and food-grade *Bacillus* subtilisins as promising new candidates for enzyme therapeutics in CD.
A novel class of gluten-degrading enzymes were isolated from *Rothia* bacteria, which are natural colonizers of the oral cavity. The enzymes were identified as subtilisins belonging to the S8 family of peptidases. Food-grade *Bacillus* species also produce such subtilisins, and these were also able to cleave and abolish gluten immunogenic epitopes. Subtilisins, cleaving after XPX↓ represent an as yet overlooked class of enzymes with great potential for enzyme therapeutic applications in celiac disease.
INTRODUCTION

Sensitivity to gluten-containing foods is widespread, and manifests predominantly in the form of celiac disease (CD). The increased awareness for CD in the medical community and the general public has led to improved diagnosis and earlier initiation of preventive strategies. The prevalence of CD ranges between ~1:100-1:200 in most populations (2). The development of CD is dependent on exogenous and host-associated factors, whereby ingested gluten is the direct trigger of the disease, and the presence of HLA-DQ2 or HLA-DQ8 alleles, and tissue transglutaminase (TG2) activity are the major contributing host-associated factors (21, 31). The disease is characterized by inflammation and flattening of the duodenal and jejunal villi, with a broad spectrum of symptoms, ranging from a clinically silent disease to severe malabsorption and a high risk for secondary autoimmune diseases (1).

To date, a strict gluten-free diet is the only treatment option for CD, which is difficult to maintain, and this poses a significant social and psychological burden to the patient. Traces of gluten are present in nearly all refined foods, and the felt quality of life of many patients equals that of hemodialysis patients (32). Therefore, novel therapies that would relieve patients from the need to adhere to the highly restrictive gluten-free diet are highly desired (27, 30).

The major therapies currently being pursued for CD target the immunogenic gluten peptides and the immune system, e.g. using a vaccine-based strategy, gluten-degrading enzymes, luminal gluten binders, and inhibitors of the body’s enzyme tissue transglutaminase which potentiates gluten antigenicity by deamidation (39). Enzyme therapies are aimed at cleaving gluten and abolishing immunogenic epitopes before they reach the lamina propria of the small intestine where T cell activation occurs (18, 19). This approach has distinct advantages. First, enzyme therapeutics are targeting the most upstream trigger, i.e. antigenic gluten peptides, rather than endogenous molecules or cells downstream in the cascade of intestinal inflammation; second, the enzymes are amenable to optimization of substrate specificity and pH activity (4).
Bacterial and barley-derived gluten-degrading enzymes have been isolated (34, 41) and are being explored for clinical application (19, 35, 42). Furthermore, gluten-degrading bacteria, mostly lactobacilli and bifidobacteria, have been found potentially useful for CD treatment (10, 20, 29), with a few reports indicating actual gluten digestion (6, 7). We have found that exceptionally high gluten-degrading enzyme activities are naturally associated with bacteria that colonize the oral cavity (15). Thus *Rothia* bacteria from human saliva can hydrolyze gluten domains that are highly immunogenic and resistant to mammalian digestive enzymes (44, 50). This discovery has identified these natural microbes as novel sources of gluten-degrading enzymes (9). Here we isolated the gluten-degrading enzymes from *R. mucilaginosa*, of which the complete genome sequence is available, and identified the proteinases as subtilisin family members. The discoveries highlight this group of enzymes, with cleavage specificities after Xaa-Pro-Xaa↓, and some with proven safety in the food industry, as hitherto unrecognized candidates for dietary enzyme therapeutics for CD.

**METHODS**

**Bacterial culturing**

*R. mucilaginosa* ATCC 25296 was routinely grown on a *Brucella*-agar plates (Hardy Diagnostics, Santa Maria, CA) at 37°C for 24 h under aerobic conditions. Individual bacterial colonies were transferred to 100 ml of Todd-Hewitt broth (Beckton Dickinson, Sparks, MD) supplemented with 0.5% Tween-80 (THT), and subcultured into 4 L of THT. All incubations were carried out in Erlenmeyer flasks, while shaking at 200 rpm for 48 h at 37°C.

**Preparation of a *R. mucilaginosa* cell extract**

Cells were harvested from the 4 L THT culture by centrifugation at 16,000 x g at 4°C for 30 min, washed two times with 20 mM TrisHCl buffer, pH 7.5, and then re-suspended in 150 ml of 20 mM TrisHCl buffer containing 596 kU/ml lysozyme (Sigma, St. Louis, MO, USA). After incubation at 37°C for 1 h, 25
mg/ml n-Octyl-β-D-glucopyranoside was added (Thermo Fisher, Waltham MA) as well as 0.3 mg/ml of L-cysteine. The cell suspension was frozen at -20°C, defrosted, and then sonicated on ice using a sonifier with a macro tip (Branson sonifier 450, VWR Scientific, Bridgeport, NJ). The chemical pretreatment of the cells combined with the sonication reduced the OD\textsubscript{620} of the suspension by 60%. The suspension was centrifuged for at 31,209 x g for 30 min at 4°C. The cloudy supernatant was harvested and then recentrifuged at 151,243 x g, for 1 h at 4°C. The resulting pellet, containing most of the activity, was resuspended in 4 ml of 20 mM TrisHCl, pH 7.5. The protein concentration was determined using the bicinchoninic acid assay (Pierce Biotechnology, Rockford, IL, USA). Samples were stored at -80°C.

DEAE anion exchange chromatography

Four ml of the resuspended pellet sample was thawed, and solubilized by the addition of 40 mg/ml n-Octyl-β-D-glucopyranoside. A 6 ml volume of 50 mM Tris-HCl containing 0.3 M NaCl, pH 7.0, (DEAE buffer A) was added. The 10 ml sample was then loaded onto an anion-exchange DEAE sepharose fast flow column with a column size of 2.6 cm diameter x 27 cm length with a column volume (CV) of 143 ml. The resin used was a cross-linked agarose with the diethylaminoethyl exchange group (-O-CH\textsubscript{2}CH\textsubscript{2}N+H(CH\textsubscript{2}CH\textsubscript{3})\textsubscript{2}; GE Healthcare, Bjorkgatan, Sweden). The column was coupled to a FPLC system (AKTApurifier 10, GE Healthcare, Bjorkgatan, Sweden) and the flow rate applied was 1 ml/min.

The buffers employed for protein elution were buffer A, containing 50 mM Tris-HCl, 0.3 M NaCl, pH 7.0, and buffer B, containing 50 mM Tris-HCl, 1 M NaCl, pH 7.0. Proteins were separated using a two-step gradient of 0% buffer B for 2.0 column volumes (286 ml) (isocratic conditions), followed by 100% elution buffer B for 1.5 column volume (215 ml). The absorbance was monitored at 219 nm and the eluate was collected in 10 ml fractions.
Analytical SDS-PAGE

The protein content of individual DEAE fractions (100 µl) and pooled DEAE fractions (400 µl) was determined using pre-casted discontinuous 4-12% SDS PAGE gels (NuPAGE, Thermofisher, Cambridge, MA) under reducing conditions. After electrophoresis gels were silver-stained as described (49).

Gliadin zymography

Pooled DEAE fractions F1-F7 were desalted using centrifugal tubes with 30 kDa molecular weight cut-off membranes (Amicon Ultra-15, EMD Millipore, Bellerica, MA) and aliquots of 200 µl were analyzed for gluten-degrading enzyme activities on a 6% gliadin zymogram gel. The zymogram gel composition and renaturing and developing conditions were reported previously (14, 50).

PAGE and Casein zymography

The F2 fraction, containing the highest enzyme activity (hereafter called *R. mucilaginosa* enzyme preparation or Rmep) was applied in amounts ranging from 2-32 µg on a 6% PAGE gel under non-reducing conditions. The composition of the gel was the same as the 6% gliadin zymogram gel, but without the incorporated gliadin. After electrophoresis the gel was divided in half. One half of the gel was silver-stained, the other half was developed as a zymogram gel using externally added casein as the enzyme substrate, as described (49).

LC-ESI-MS/MS

Proteins of interest were excised from the silver-stained gel half, and in-gel digested with sequencing-grade trypsin (Promega, Madison, WI), as described (49). The peptides were eluted from the gel, separated by in line C18 chromatography and sequenced using an LTQ Orbitrap mass spectrometer (ThermoFinnigan, San Jose, CA). The obtained b- and y-ion spectra were searched against a database of *R. mucilaginosa* ATCC 25296, containing 1737 Rothia protein entries, and 132 non-Rothia decoy
proteins. Applied filtering criteria were X-corr values > 1.5, 2.2 and 3.5 for Z = 1, 2, and 3, respectively. The deltaCn and peptide probability values selected were >0.1 and <0.01, resp.

Hydrolysis of paranitroanilide-derivatized substrates

Paranitroanilide (pNA)-derivatized tripeptide substrates were chemically synthesized at >90% purity (21st Century Biochemicals, Marlborough, MA). The substrates obtained were Z-YPQ-pNA, Z-QQP-pNA, Z-LPY-pNA, Z-PFP-pNA, Z-PPF-pNA, where Z=benzyloxycarbonyl, Y=tyrosine, P=proline, Q=glutamine, L=leucine, and F=phenylalanine. Suc-AAPF-pNA was obtained from Sigma, where Suc=N-succinyl and A=alanine. The peptides were dissolved in 75-100% DMSO at 10 mM, and were used at a final concentration of 200 µM in 50 mM TrisHCl pH 8.0. Purified *Rothia* enzyme or subtilisin A from *Bacillus licheniformis* (Sigma) or Nattokinase (extracted from a dietary food supplement NSK-SD, Pure Encapsulations, Sudbury MA), were tested at final concentrations of 1 µg/ml. Some experiments were conducted in the presence of inhibitors 4-(2-Aminoethyl) benzenesulfonyl fluoride (AEBSF), Aprotinin, E-64, EDTA, phenylmethanesulfonyl fluoride (PMSF) and Eglin C, which were tested at final concentrations of 10 mM, 0.08 mM, 0.1 mM, 1.5 mM, 1 mM, and 0.06-1.2 µM, respectively. Enzymes were preincubated with the inhibitors for 20 min before adding Suc-AAPF-pNA. Substrate hydrolysis was monitored for 10 h at 405 nm using a Genios microtiter plate reader (Tecan, Männedorf, Switzerland) in the kinetic mode at 37°C.

FRET substrate hydrolysis

Three fluorescence resonance energy transfer (FRET) substrates comprising the hexapeptides QPQLPY, PQPQPQ and QGSFQP were synthesized with at the N-terminus the HiLyte Fluor™ 488 label, and at the C-terminus with K(QXL520) (Anaspec, Fremont, CA). *Rmep* (1 µg/ml) and subtilisin A (0.5 µg/ml) were incubated at 37°C with the substrates (100 µM) in 50 mM TrisHCl pH 8.0. Fluorescence increase, indicating substrate hydrolysis, was measured every 5 min for 30 min, and every 10 min for the next 30 min, at λ_{ex} 485 nm and λ_{em} 520 nm, using a Genios microtiter plate reader in the kinetic mode.
Degradation of mixed gliadins and the 33-mer peptide

Mixed gliadins were obtained from Sigma. A synthetic highly immunogenic α-gliadin derived 33-mer peptide (33) was synthesized at a purity of >90% (21st Century Biochemicals). Mixed gliadins or the 33-mer peptide (both at final concentration 250 µg/ml) were incubated with Rmep or subtilisin A (each at 57 µg/ml) in 50 mM TrisHCl, pH 8.0. After t=0, 15 min, 30 min, and 2 h, 100 µl sample aliquots were removed and boiled. Gliadin degradation was assessed by 4-12% SDS-PAGE followed by Coomassie staining. Degradation of the 33-mer was determined by RP-HPLC.

Reversed-phase high performance liquid chromatography (RP-HPLC)

Separation of the 33-mer peptide and its fragments was achieved by RP-HPLC, using buffer A (0.1% trifluoroacetic acid) and buffer B (0.1% trifluoroacetic acid in 80% acetonitrile) at a gradient of 0-55% buffer B over 75 min time interval. The equipment used was an HPLC Model 715 (Gilson, Middleton, WI) and a C-18 column (TSK-GEL 5 mm, ODS-120T, TOSOHaas, Montgomeryville, PA) (14, 50).

R5 and G12 ELISA assays

Gluten epitope elimination in the gliadin-Rmep and gliadin-subtilisin A digests was assessed using two sandwich ELISA assays employing the R5 monoclonal antibody (RIDASCREEN Gliadin, R-Biopharm, Darmstadt, Germany) or the G12 monoclonal antibody (AgraQuant® ELISA Gluten G12, Romer labs, Union, MO). To prevent high background values, the R5 plate wells were blocked with 1% skim milk in PBS prior to incubation with the samples. The assays were performed according to the manufacturers’ instructions and as described (44, 49).

RESULTS

Purification of the enzyme from Rothia mucilaginosa
We observed that the gluten-degrading enzyme activity of *R. mucilaginosa* was primarily cell associated. Therefore, to isolate the enzyme(s), the cells were treated with lysozyme, an N-acetylmuramidase to break down the peptidoglycan layer, and then sonicated and centrifuged. The supernatant, which was turbid and contained microscopic vesicles, was ultracentrifuged. Enzyme activities, monitored with the substrate Z-YPQ-pNA, were primarily localized in the (vesicular) pellet. The pellet was dissolved to clarity with the mild detergent n-Octyl-β-D-glucopyranoside and then subjected to DEAE chromatography applying an isocratic gradient. Figure 1A shows the DEAE chromatogram, Figure 1B the protein content in fraction aliquots analyzed by SDS-PAGE, and Figure 1C the enzyme activity, determined with the substrate Z-YPQ-pNA. Most activity was contained in fractions 7 and 8 with some activity trailing in fractions 9-14. Fractions 7 and 8 contained proteins migrating between 75 and 150 kDa. Based on protein patterns and enzyme activities the fractions were pooled into seven fractions as follows: F1: fractions 1-5, F2: fractions 6-8, F3, fractions 9-12, F4, fractions 13-20, F5, fractions 21-52, F6, fractions 53-56 and F7: fractions 57-65.

The protein composition in F1-F7 is shown in Figure 2A. F2, containing the enzyme activity and hereafter called *R. mucilaginosa* enzyme preparation (*Rm*ep), displayed a major band between 100-150 kDa, and a double band between 75 and 100 kDa. Gliadin zymography of F1-F7 showed evidence for a double enzyme band in F2 of ~75-80 kDa (Fig. 2B). As expected, the highest specific activity was associated with F2 (not shown). *Rm*ep was subsequently analyzed at 4 different concentrations on a 6% SDS gel under non-reducing conditions. After electrophoresis half of the gel was silver stained (Fig. 2C) and the other half was developed as a zymogram with externally added casein as the substrate (Fig. 2D). This permitted a true comparison in electrophoretic mobilities of proteins in both gel halves. Enzyme activities were again associated with the ~75-80 kDa double band, and not with the also prominent ~125 kDa band. Several bands were excised from the SDS- and the zymogram gel for protein identification by LC-ESI-MS/MS. The inactive 125 kDa band was labeled as band a, the active ~80 kDa band as b/d and the active ~75 kDa band as c/e. The major proteins identified in bands a-e were ROTMU0001_0241 (C6R5V9_9MICC), ROTMU0001_0243 (C6R5W1_9MICC) and ROTMU0001_0240.
(C6R5V8_9MICC) (Fig. 2E). The relative abundances of the proteins in the respective bands, determined at the MS1 level, are indicated. It shows that the ~125 kDa band contains a mixture of primarily C6R5V9_9MICC and C6R5W1_9MICC, in a 32%/67% ratio. The ~80 kDa protein is a fragment of C6R5V9_9MICC, and the ~75 kDa band is derived from C6R5W1_9MICC. Identifications were made based on 49 and 62 matching tryptic peptides, resp., and >50% sequence coverage. As a control, three other bands were excised from the gel shown in Figure 1B (fraction 13, lowest band, and fraction 14 two middle bands), yielding identifications other than subtilisin enzymes (data not shown).

The amino acid sequences of the identified subtilisins ROTMU0001_0241, ROTMU0001_0243 and ROTMU0001_0240 are shown in Supplemental Figure 1. They are listed in UniProt as uncharacterized proteins. A blast search revealed that all are members of the S8 peptidase family, which share a catalytic triad comprising Asp (D), His (H) and Ser (S). The signal peptide cleavage sites as well as the catalytic triads are indicated. The protein sequences of all subtilisin genes in the three known members of the Rothia genus, Rothia aeria, Rothia mucilaginosa and Rothia dentocariosa were aligned using Clustal Omega (Supplemental Fig. 2). The three variants identified by LC-ESI-MS/MS are highlighted. A phylogenetic tree was drawn using Clustal Omega, showing evidence for 8-9 different Rothia subtilisin sequence types based on sequence conservation around the catalytic residues D, H and S (data not shown).

**Enzymatic characteristics of Rmep**

*Rmep* was further tested for enzyme activities and specificities relevant for CD. The enzyme was active over a pH range of 6.0-10.0, with low or negligible activities at pH values lower than 5.0 or higher than 11.0 (Fig. 3A). Activity was strongly inhibited by PMSF and AEBSF (serine protease inhibitors), with partial inhibition by aprotinin, but not by E64 (cysteine protease inhibitor) or EDTA (metalloprotease inhibitor) (Fig. 3B). These findings are consistent with subtilisins being serine proteases. *Rmep* was only partly inhibited by high concentrations of Eglin C, which is an effective inhibitor of *B. licheniformis* subtilisin A (Fig. 3C, D). This difference in sensitivity is likely due to the significant overall structural
differences between Rmep and subtilisin A, including in domains flanking their active sites which participate in binding of the inhibitor (16).

In order to investigate enzyme specificities relevant for gliadin degradation the substrates Z-YPQ-pNA, Z-QQP-pNA, Z-LPY-pNA, Z-PFP-pNA, Z-PPF-pNA were employed in which the tripeptides are all contained with high frequency in the immunogenic gliadin domains. Suc-AAPF-pNA was used as it is a preferred substrate for subtilisins. Results showed that Z-QQP-pNA and Z-PFP-pNA were not hydrolyzed by either subtilisin A or Rmep (data not shown). In contrast, substrates with a P residue in the P2 position were cleaved, albeit at quite different rates; Suc-AAPF-pNA was hydrolyzed most rapidly (Fig. 4A, B), followed by Z-YPQ-pNA. The rapid hydrolysis of Suc-AAPF-pNA by Rmep further supports the identification of subtilisin-like enzymes. LPY hydrolysis by Rmep could be demonstrated at five-fold increased enzyme concentrations and longer incubation times, and PPF was hydrolyzed by subtilisin A at fifty-fold increased enzyme concentrations, but not by Rmep (data not shown).

Activities were also determined towards three gliadin-derived FRET substrates (Fig. 4C, D). The substrates were QPQLPY, contained in three immunogenic gliadin epitopes; PQQPQ, a theoretical substrate for Rothia enzymes, and QGSFQP, contained in two gliadin epitopes. The first two peptides, containing XPX motifs, were rapidly hydrolyzed by Rmep (Fig. 4C). In contrast, QGSFQP, not containing XPX, was not cleaved. Subtilisin A showed the highest activity towards the PQQPQ substrate (Fig. 4D).

**Elimination of immunogenic epitopes by Rmep**

Hydrolysis of the FRET substrates would suggest CD-relevant epitope degradation. To further investigate this, the time courses of degradation of mixed gliadins and of the highly immunogenic 33-mer α-gliadin peptide were determined. Epitope elimination was monitored by two ELISA assays employing the R5 antibody recognizing QQPFP and related pentapeptides (28, 46) and the G12 antibody recognizing QPQLPY contained in the 33-mer peptide (24, 25). Data obtained with Rmep are shown in Figure 5, and with subtilisin A in Figure 6.
Rmep rapidly degraded mixed gliadins within 15 min incubation and completed 33-mer peptide degradation within 30 min of incubation (Fig. 5A and 5B, resp.). R5 epitope abolishment in the Rmep digest of gliadins—could already be observed in the t=0 sample (46% reduction in antibody binding compared to the control, corrected for enzyme only values). After 15 min virtually all R5 epitopes were eliminated (Fig. 5C). After 30 min the G12 antibody binding had decreased by 65% and showed a time-dependent decline to 85% reduction after 2 h of incubation (Fig. 5D). Subtilisin A degraded gliadins even faster, with near complete degradation in the t=0 sample (Fig. 6A). However, the 33-mer peptide was not as efficiently cleaved, and after 2 h of incubation residual 33-mer still remained (74% degradation; Fig. 6B). The R5 epitopes in the gliadin-subtilisin A digest were rapidly eliminated, by 90% at t=0, and by 100% within 15 min of incubation (Fig. 6C). In contrast, the G12 epitopes were quite resistant to subtilisin A, and were degraded by 22% at t=0 and 45% at 2 h (Fig. 6D). This agrees with the modest activity observed towards the 33-mer.

Lastly, we tested B. subtilis subtilisin NAT, or Nattokinase, which is available commercially as a dietary supplement for human consumption. Like Rmep and subtilisin A, Nattokinase degraded gliadins effectively, down to ng/ml concentrations (Fig. 7A, B). Nattokinase hydrolyzed Z-AAPF-pNA (Fig. 7C), and degraded the 33-mer peptide (Fig. 7D), although a large fragment eluting just prior to the 33-mer in the RP-HPLC chromatogram remained. It abolished the R5 epitopes (Fig. 7E), but was less effective in eliminating the G12 epitopes, consistent with observations made with subtilisin A (Fig. 7F).

**DISCUSSION**

The gluten-degrading enzymes of *Rothia* were identified as members of the subtilisin protease family, and the gluten-degrading activities of this class of enzymes were shown to extend beyond the *Rothia* genus. The cleavage specificity of subtilisins is XPX↓ with X in the P1 position preferably being a hydrophobic amino acid. In accordance, substrates with Q in the P1 position were found to be highly susceptible to cleavage by Rmep and subtilisin A. XPQ is present in the majority of the antigenic gluten
epitopes relevant in CD (40). A subset of such Q residues in gliadins are deamidated by the enzyme tissue transglutaminase 2 (TG2) in the lamina propria, a key step in the pathogenesis of CD (47). This deamidation increases the peptides’ affinity for HLA-DQ2 and HLA-DQ8 expressed on antigen-presenting cells and triggers the destructive mucosal T cell response. Glutamine residues in the XPQXP context are particularly prone to such deamidation (38, 45). Selective cleavage of this sequence will prevent Q deamidation and thus both Rothia and Bacillus subtilisins can be anticipated to be efficient in preventing T cell activation in the lamina propria. This is supported by our finding of the elimination of the major gliadin epitopes in two independent ELISA assays, even without pre-digesting gliadins with the mammalian digestive enzymes pepsin and trypsin.

The three major Rothia species are R. aeria, R. mucilaginosa and R. dentocariosa. Analysis of the gluten degrading Rothia enzymes showed that they belong to the D-H-S class of subtilisins (36, 37) and not to the family of kexins (an S8 protease sub-family). All contain a signal peptide for secretion, which is presumed to be cleaved off as most contain the consensus sequence AxA/A which is recognized by a signal peptidase. Using NCBI BlastP (or InterPro) analysis it was found that none of the sequences contain a C-terminal LPxTG-type peptidoglycan anchor. However, most Rothia subtilisins did contain two to three C-terminal SLH (surface layer homology) domains, called Pfam00395 (22). Such bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving cell wall polysaccharide pyruvylation. The theoretical cell-envelope association agrees with our finding that Rothia subtilisins could be isolated from cell-derived vesicles that were harvested by ultracentrifugation.

The mass spectrometric identification of subtilisins prompted us to explore subtilisins from other microbes, specifically Bacillus species, for activity, and they were identified also to degrade gluten. Like Rmep, subtilisins A and NAT (Nattokinase) indeed degraded gluten efficiently. Like Rothia, the Bacillus genomes encode for several subtilisins. In general, bacteria that express multiple extracellular proteases reside in a protein-rich environment, and use proteases to degrade proteins into peptides that can subsequently be utilized for growth. This is the case for the natural habitat of Rothia species, the oral cavity, where XPQ-containing substrates are prevalent. Interestingly, in the oral cavity, the XPQ-rich
proteins are not only represented by ingested gluten proteins, but also by salivary proline-rich proteins that are produced constitutively by the salivary glands (3, 12, 17, 43). In this context it is of interest to note that, like *Rothia*, *B. subtilis* also colonizes the oral cavity (8). The salivary proline-rich proteins undergo extensive proteolytic fragmentation, with primary cleavage after XPQ↓ (13, 23, 48). Based on the data presented here, these cleavages are carried out most likely by oral bacterial subtilisins.

Some interesting observations were made when the substrate specificities of subtilisin A and *Rm*ep were compared. First, both enzymes cleaved AAPF rapidly, but not PPF. This suggests that a P in the P3 position interferes with substrate recognition by both enzymes. In most gluten immunogenic domains containing the XPF sequence, X is represented by a Q, and not a P. The FRET results showed that QPF is efficiently cleaved. Also abundant in gliadin immunogenic domains is the sequence LPY. *R. mucilaginosa* cleaves the 33-mer at LPY↓, QPQ↓ and YPQ↓ (44), and it is very likely that all cleavages are carried out by the same subtilisin enzyme.

An important observation made was that the molecular weight of the active subtilisins, migrating in the ~75-80 kDa gel area was substantially shorter than the parent gene product of ~140 kDa. It is well known that subtilisins from *Bacillus* species undergo autocatalytic activation to produce a shorter mature enzyme (11). In the Rothia subtilisins the catalytic triad is located at the N-terminal portion of the protein. We postulate that the parent inactive precursor proteins encoded by C6R5V9_9MICC and C6R5W1_9MICC are processed at the N-termini to remove the inhibitory propeptide, and at the C-terminus between the Ig-like segment and SLH domains, generating the ~75 and ~80 kDa mature enzymes, as outlined in Figure 8. Further structural analysis should reveal the exact processing sites and the secondary and tertiary structures of active Rothia subtilisins.

*B. subtilis* is food safe and has been consumed for decades, e.g. in a product called natto, a Japanese fermented soy bean dish. In natto the active enzyme is Nattokinase, a 27.7 kDa subtilisin enzyme (26) that was used in our studies. Despite a long history of consumption of *B. subtilis* and its products, there are very few reports of adverse events. The food-grade status of *B. subtilis*, and the already widely consumed natto products, open new avenues for potential therapeutic applications of the subtilisin enzymes. The *R.*
mucilaginosa subtilisins have superior epitope-detoxifying capacities compared to Bacillus subtilisins, and they are predicted to be more effective in neutralizing gluten epitopes, as based on our in vitro assessments. The usefulness and safety of subtilisins in vivo should be further explored. These include assessment of enzyme activity in buffers and systems that mimic gastro-duodenal conditions, and that take into account the variable pH conditions in the stomach and duodenum, the presence of endogenous digestive enzymes, and competing food proteins. In addition, functional assays should reveal if the biochemical observations of epitope abolishment in vitro can be validated in cell-based assays, e.g. in CD biopsy-derived T cell lines, and in in vivo mouse models for CD. Furthermore, it should be established if subtilisins doses required for in vivo gluten digestion are safe for long term consumption. Going forward, since gluten-degrading enzymes are the preferred therapy of choice for CD (5), and given the exceptional activity of the subtilisins and their association with natural human microbial colonizers, they are worthy of further exploration for clinical applications in CD and potentially other gluten-intolerance disorders.

Authorship

Guarantor of the article: Eva J. Helmerhorst

Author contributions: GW: Experimental design, acquisition of data; analysis and interpretation of data; NT: Experimental design, acquisition of data; analysis and interpretation of data; RS: Critical revision of the manuscript for important intellectual content; DS: Study concept, critical revision of the manuscript for important intellectual content; EJH: study concept and design, acquisition of data, manuscript writing, obtained funding.

Acknowledgements

The authors thank Dr. Maram Zamakhchari for her contributions to the project, and Ross Tomaino for conducting the mass spectrometric analyses.

Declaration of personal interests: we declare no conflict of interest for this manuscript.
Declaration of funding interests: These studies were supported by NIH/NIAID grants AI087803 and AI101067 to EJH.
Figure legends

Fig. 1. Isolation of *R. mucilaginosa* gluten-degrading enzymes by DEAE chromatography. *R. mucilaginosa* cells were lysed, sonicated, the supernatant ultracentrifuged, and the pellet dissolved in n-Octyl-β-D-glucopyranoside. A, Separation of proteins by DEAE chromatography applying an isocratic gradient containing 0.3M NaCl and 50 mM TrisHCl pH 7.0; B, Protein content in 100 µl fraction aliquots investigated by 4-12% SDS PAGE; C, Enzyme activity in 50 µl fraction aliquots investigated with Z-YPQ-pNA as the substrate. The data shown are representative of two independent experiments.

Fig. 2. Identification of the gluten-degrading enzymes of *R. mucilaginosa*. DEAE fractions 1-65 were pooled into 7 fractions designated F1-F7 (see text for details). A, protein content in 400 µl aliquots of F1-F7 analyzed on a 4-12% SDS gel; B, Gliadin-degrading enzyme activity in 200 µl desalted aliquots of F1-F7 analyzed on a 6% gliadin zymogram gel. C, PAGE under non-reducing conditions of *Rm*ep (F2), loaded amounts 4, 8, 16, 32 µg protein in lanes 1-4, resp. D, casein zymogram of F2, loaded amounts 2, 4, 8, and 16 µg in lanes 1-4, resp. E, proteins identified in excised gel bands labeled a-e. The data shown are representative of at least four independent experiments.

Fig. 3. Effect of pH and inhibitors on *Rm*ep activity. A, activity of *Rm*ep in 0.1M Citric acid/0.2M phosphate buffer mixtures, pH 2-12, measured with Suc-AAPF-pNA (200 µM) as the substrate. B, Inhibitor profile of *Rm*ep. Inhibitors AEBSF, Aprotinin, E-64, EDTA, PMSF and Eglin C were tested at final concentrations of 10 mM, 0.08 mM, 0.1 mM, 1.5 mM, 1 mM, and 0.06 µM, respectively. *Rm*ep was added at a final concentration of 5.7 µg/ml. The percent inhibition was determined from the ratio of the initial velocities of hydrolysis of Suc-AAPF-pNA in the absence and presence of inhibitor. C and D, Inhibition of subtilisin A and *Rm*ep, respectively, at higher Eglin C concentrations (1.2 µM).
Fig. 4. Cleavage specificities of Rmep and B. subtilis subtilisin A. A and B, hydrolysis of Suc-AAPF-pNA, Z-YPQ-pNA, and Z-LYP-pNA, each at 200 µM, by Rmep and subtilisin A, respectively, each at 1 µg/ml. C, D, Hydrolysis of gliadin-derived FRET substrates. FRET substrates containing QPQLPY, PQPQPQ and QGSFQP, each at 100 µM, were incubated with Rmep (C) and subtilisin A (D), at 1 µg/ml and 0.5 µg/ml, respectively. Controls were boiled Rmep and subtilisin A incubated with PQPQLPY (+), PQPQPQ (◊) and QGSFQP (*) (no activity, baseline).

Fig. 5. Degradation of mixed gliadins by Rmep and abolishment of immunogenic epitopes. A, Gliadins (250 µg/ml) in 50 mM Tris/HCl pH 8.0 were incubated with Rmep at 57 µg/ml. After 0, 15 min, 30 min and 2 h incubation 100 µl aliquots were removed, boiled and analyzed by SDS-PAGE and stained with Coomassie Brilliant Blue. The bold arrow (left) points to the position of the major band in the gliadin preparation; the thin arrow (right) to the 140 kDa band in the Rmep preparation, and the dotted arrows (right) to the gliadin degradation fragments. B, RP-HPLC of degradation of the immunogenic 33-mer peptide from α-gliadin. Arrow (top) points to the intact 33-mer; C, D assessment of the survival of epitopes in the gliadin-Rmep degradation mixture employing the R5 ELISA (C) or G12 ELISA (D) assays.

Fig. 6. Degradation of mixed gliadins by B. subtilis subtilisin A and abolishment of immunogenic epitopes. A, Gliadins (250 µg/ml) in 50 mM Tris/HCl pH 8.0 were incubated with subtilisin A at 57 µg/ml. Experiments were conducted as described in the legend of Figure 5 with subtilisin A instead of Rmep.

Fig. 7. Gliadin degradation and epitope abolishment by nattokinase (NattoK) from B. subtilis. A, mixed gliadins (G, 250 µg/ml) were incubated with NattoK (Pure encapsulations, Sudbury, MA; 57 µg/ml) for 0, 15, 30 and 120 min. Left four lanes: controls without enzyme or gliadins, respectively, each at t=0 and 120 min. Arrow points to the major components in the gliadin mixture represented by α and γ-gliadins.
containing most of the immunogenic epitopes. B, dilution series of NattoK (3.5-0.06 µg/ml) incubated for 30 min with mixed gliadins. Right lanes, gliadin (G) and NattoK (NK) control. C, dilution series of NattoK incubated with Suc-AAPF-pNA. Hydrolysis was measured at 405 nm. D, RP-HPLC analysis of the gliadin-derived 33-mer (250 µg/ml) incubated for 0, 15, 30 and 120 min with NattoK 57 µg/ml. E and F, Epitope abolishment in mixed gliadins (250 µg/ml) incubated for 0, 15, 30 and 120 min with NattoK (57 µg/ml) assessed with the R5 ELISA (E) and G12 ELISA (F) assays.

**Fig. 8.** Domain composition of C6R5V9_9MICC. The protein of 1328 amino acids contains a peptidase S8 propeptide domain (also called proteinase inhibitor I9), a peptidase S8/S53 family domain, an immunoglobulin-like fold domain, and 3 surface layer homology domains. Figure based on InterPro analysis.
REFERENCES


LC-ESI-MS/MS of proteins in bands a-e

<table>
<thead>
<tr>
<th>Protein</th>
<th>MW gene product</th>
<th>Band a</th>
<th>Band b/d</th>
<th>Band c/e</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6R5V9_9MICC</td>
<td>141.8</td>
<td>32%</td>
<td>90%</td>
<td>10%</td>
</tr>
<tr>
<td>C6R5W1_9MICC</td>
<td>141.8</td>
<td>67%</td>
<td>10%</td>
<td>90%</td>
</tr>
<tr>
<td>C6R5V8_9MICC</td>
<td>142.4</td>
<td>1%</td>
<td>~0%</td>
<td>~0%</td>
</tr>
</tbody>
</table>

1 Full sequences are shown in Supplemental Figure 1. The 125 kDa band is the original gene product minus the signal peptide (38-50 amino acids) and possibly the predicted propeptide region (approximately 120 amino acids), thus migrating around 125 kDa.
Figure 5

Table:

<table>
<thead>
<tr>
<th>Incubation Time</th>
<th>Gliadin</th>
<th>Rmep</th>
<th>Gliadin + Rmep</th>
</tr>
</thead>
<tbody>
<tr>
<td>0h</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0.25h</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0.5h</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2h</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Graphs:

A) Gel electrophoresis

B) 33-mer

C) R5

D) G12
Glia
Subt A
Glia + Subt A

Table:

<table>
<thead>
<tr>
<th></th>
<th>0h</th>
<th>2h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gliadin</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Subt A</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Graphs:

A: Gel electrophoresis showing bands for Glia, Subt A, and Glia + Subt A across time points.

B: Graph depicting absorbance (450nm) over time with 33-mer indicated.

C: Graph showing absorbance (450nm) for R5 over time.

D: Graph showing absorbance (450nm) for G12 over time.
Figure 7
Figure 8

- **Signal peptide cleavage**
- **Propeptide cleavage**
- **Peptidase S8/S53 domain**
- **Ig-like domain**
- **SLH domains**
- **Autoprocessing?**

Peptidase S8
Propeptide/inhibitor