Classification of *Lactococcus lactis* cell envelope proteinase based on gene sequencing, peptides formed after hydrolysis of milk, and computer modeling

M. W. Børsting,*†¹ K. B. Qvist,† E. Brockmann,† J. Vindeløv,† T. L. Pedersen,† F. K. Vogensen,* and Y. Ardö*

*Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 28, 1958 Frederiksberg, Denmark

†Chr. Hansen A/S, 2970 Hørsholm, Denmark

ABSTRACT

Lactococcus lactis strains depend on a proteolytic system for growth in milk to release essential AA from casein. The cleavage specificities of the cell envelope proteinase (CEP) can vary between strains and environments and whether the enzyme is released or bound to the cell wall. Thirty-eight *Lc. lactis* strains were grouped according to their CEP AA sequences and according to identified peptides after hydrolysis of milk. Finally, AA positions in the substrate binding region were suggested by the use of a new CEP template based on *Streptococcus* C5a CEP. Aligning the CEP AA sequences of 38 strains of *Lc. lactis* showed that 21 strains, which were previously classified as group d, could be subdivided into 3 groups. Independently, similar subgroupings were found based on comparison of the *Lc. lactis* CEP AA sequences and based on normalized quantity of identified peptides released from α_{S1} -casein and β -casein. A model structure of *Lc. lactis* CEP based on the crystal structure of *Streptococcus* C5a CEP was used to investigate the AA positions in the substrate-binding region. New AA positions were suggested, which could be relevant for the cleavage specificity of CEP; however, these could only explain 2 out of 3 found subgroups. The third subgroup could be explained by 1 to 5 AA positions located opposite the substrate binding region.

Key words: cell envelope proteinase (CEP), *Lactococcus lactis* ssp. classification, computer modeling, substrate binding

INTRODUCTION

Milk contains a relatively small amount of free AA and small peptides; however, lactic acid bacteria (LAB) need 4 to 16 essential AA to grow (Chopin, 1993). This makes LAB dependent on a proteolytic system that can break down caseins in milk to AA and peptides. The

proteolytic system in *Lactococcus lactis* consists of a cell envelope proteinase (CEP), also called lactocepin or PrtP, which is located outside the cell and initiates the breakdown of casein into medium-sized peptides. These peptides are transported across the cell membrane by at least 3 different types of specialized transport systems. Inside the cells, intracellular peptidases hydrolyze medium-size peptides into smaller peptides and AA (Kunji et al., 1996; Savijoki et al., 2006).

One of the first classification systems divided the CEP into 3 groups based on their ability to degrade different parts of casein: CEP type PI preferentially hydrolyzes β -CN, whereas CEP type PIII also hydrolyzes α_{S1} -CN and κ -CN; an intermediate group PI/PIII was also defined (Visser et al., 1986). About 98% homology of CEP sequences was found between *Lc. lactis* ssp. *cremoris* SK11 (CEP type PIII) and *Lc. lactis* ssp. *cremoris* Wg2 (CEP type I; Vos et al., 1989), which made it possible, using recombinant DNA technology, to construct hybrid proteins by swapping regions of the 2 CEP with new proteolytic properties. The AA sequences and proteolytic properties were compared by using a computer model of SK11 CEP with a template based on the crystal structure of subtilisin (Vos et al., 1991). The homology between subtilisin and *Lc. lactis* CEP is seen especially in the catalytic domain around the residues of the active site aspartic acid (Asp30), histidine (His94), asparagine (Asn196), and serine (Ser433). Comparison between the cleavage sites and the CEP model made it possible to suggest 5 AA positions (131, 138, 142, 144, and 166) to be part of the substrate binding pocket and 2 other AA positions (747 and 748), which were thought to be important in electrostatic interactions with caseins (Vos et al., 1991).

Exterkate et al. (1993) provided the foundation of a new classification system of CEP. They compared the selected AA positions found by Vos et al. (1991) plus 2 extra residues (177 and 763) with the ability to hydrolyse the peptide α_{S1} -CN (f1–23) from 16 CEP purified from various *Lc. lactis* strains. This resulted in 7 classification groups (named a to g) that were later extended to 8 groups (a to h; Broadbent et al., 1998). This classification system is hereafter referred to as

Received June 24, 2014.

Accepted September 2, 2014.

¹Corresponding author: dkmew@chr-hansen.com

Exterkate's classification system. The specificity of *Lc. lactis* CEP cleavage of intact casein was reviewed in Kunji et al. (1996). Specificity and activity depend on whether purified or native-bound CEP is used (Exterkate and Alting, 1993; Flambard and Juillard, 2000). Environmental factors such as salt, pH, and water activity also influence the specificity (Reid and Coolbear, 1998, 1999). Finally, expression of CEP depends on the growth medium (Exterkate, 1979, 1985).

Cheddar cheese contains residual chymosin and plasmin activity that initiates the primary casein breakdown. The CEP from LAB hydrolyses the medium-sized peptides to smaller peptides, which are transported into the cell by the specialized transport systems. Intracellular peptidases continue the breakdown of peptides when *Lc. lactis* are grown in milk. Proteolysis in cheese is a major contributor to flavor formation and structure development of Cheddar cheese (Upadhyay et al., 2004).

More knowledge and new technology have been gained since the last update of the *Lc. lactis* CEP classification system. Faster and cheaper sequencing of LAB genomes indicates that the a to h classification system is not consistent and does not fully describe the variation of CEP.

The aim of this study was to consider a subdivision of Exterkate's classification system group d, strains of which are the most commonly used in Cheddar cheese production. This was based on comparison of the *Lc. lactis* CEP AA sequences and normalized quantity of identified peptides after hydrolysis in milk. Furthermore, new AA positions important in the substrate binding region were suggested based on a computer model of *Lc. lactis* ssp. *cremoris* SK11 CEP with *Streptococcus* C5a CEP as template.

MATERIAL AND METHODS

Bacteria Preparation

Thirty-four different *Lc. lactis* ssp. *cremoris* and *Lc. lactis* ssp. *lactis* strains (Chr. Hansen A/S, Hørsholm, Denmark) were genome sequenced in this study. Four *Lc. lactis* ssp. *cremoris* strains (University of Copenhagen, Frederiksberg, Denmark) were not genome sequenced. However, published CEP sequences of strains (www.uniprot.com) with the same strain names (Wg2, HP, NCDO763, and SK11) were used in the analysis. Stocks were prepared by growing strains in reconstituted laboratory milk (9.5% total solids, autoclaved 15 min, 121°C; hereafter, **A-milk**) and incubated overnight at 30°C, diluted 1:10 with A-milk, and transferred to ampules, which were stored at -80°C until use.

Gene Sequences

Genome sequencing was performed at Geneservice (Cambridge, UK) or Beijing Genome Institute (BGI, Hong Kong) on purified DNA originating from single-strain samples using the next-generation sequencing (NGS) platform of Illumina Inc. (San Diego, CA). Sequencing strategies used were 36-bp single reads or 36-, 90-, or 100-bp paired-end reads with 500-bp spacer size. De novo assemblies of the genome sequences were performed using the algorithm in CLC Genomics Workbench (CLC Bio, Århus, Denmark), resulting in a variable number of contigs. The CEP sequences were identified by BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>; CLC Genomics Workbench) of the contigs resulting from de novo assembly against the nucleotide sequence of the CEP from *Lactococcus lactis* ssp. *cremoris* HP (GenBank accession no. AF247159). Only strains without sequencing ambiguities as revealed by the de novo conservation profile were included in the study. The gene sequences were aligned using the software Jalview (www.jalview.org).

Modeling of CEP

In the absence of an experimentally determined 3-dimensional structure for any *Lc. lactis* CEP, a homology model for *Lc. lactis* ssp. *cremoris* SK11 lactocepin (P15292) was constructed. First profile-profile alignment (HHPRED) at the Max Planck Institute for Developmental Biology website (<http://toolkit.tuebingen.mpg.de/hhpred>) was used to find the most suitable template available, the alignment of which was then passed to the MODELER (Sali and Blundell, 1993) server at the same website for determination of the predicted 3-dimensional model.

Identification of Milk-Derived Peptides

The A-milk was inoculated with 1% (vol/vol) of each strain from stock ampules for 20 h at 30°C, followed by reinoculation with 1% (vol/vol) in A-milk and incubation under the same conditions. The hydrolysis experiment was carried out in fresh skim milk, which was pasteurized (30 min at 64°C) before use. Then, 50 mL of skim milk was inoculated with 1% of each strain and incubated for 24 h at 30°C. The acidified milk was centrifuged (3,400 × *g*, 10 min, at 5°C), the supernatant was heated (10 min at 85°C) to stop potential enzyme reactions, and the samples were stored at -20°C until analysis. The experimental work was carried out in duplicate on 2 different days.

Fifteen microliters of each supernatants was analyzed by liquid chromatography-mass spectrometry (LC-MS)

using an Agilent 1290 Infinity LC system coupled to a 6540 Q-TOF MS system (Agilent Technologies A/S, Glostrup, Denmark). The equipment parameters were according to Jensen et al. (2009). The only modification was a shorter run time (46 min instead of 60 min) with a similar gradient within the first 46 min.

Raw agilent.d files were converted to mzML files using msconvert with 32-bit precision. Peptides were identified using MS-GF+ (v8597) according to Kim et al. (2010) with the following parameters: tolerance = 40 ppm, tda = 1, minimum peptide length = 3, maximum peptide length = 50, minimum charge = 1, maximum charge = 3, protease = unspecific, instrument = TOF. MS1 peaks were quantified using the software xcms according to Smith et al. (2006) with the following settings: centWave algorithm: (ppm = 15) peakwidth = c(20,50), snthresh = 5, prefilter = c(1,1000). In each sample, retention time was corrected against a sample containing a mixture of all samples using the obiwarp algorithm with a response of 10. Grouping of detected peaks across samples was done using the density-based approach (mzwid = 0.1, bw = 10, minfrac = 0), and nondetected peaks within groups were filled out by integrating the signal in the group region.

Differences between peptides released by the selected strains were identified by normalizing the area under the peak for each sample with regard to the most intense peak for each peptide. A value of 100 was allocated to the sample with the greatest intensity and the others were given as a percentage of this value. An average of the intensity between sample duplicates within each strain was used in the further data processing. Phosphorylated peptides were not identified.

Data Analysis

The software SIMCA V13.0 (Umetrics, Umeå, Sweden) was used for multivariate data analysis. Principal component analysis (PCA) was used to group CEP AA sequences and to group the identified peptides with subsequent hierarchical cluster analysis (HCA) of the PCA scores. The orthogonal projections to latent structures with more than one Y-variable (O2PLS) method was used to find the relationship between the CEP AA sequences and the identified peptides. An O2PLS allows partitioning of the systematic variability in X and Y into 3 parts: (1) the information that is simultaneously present in X and Y, which is only used in this study; (2) information that is only found in X; and (3) information that is only found in Y (Eriksson et al., 2013).

CEP AA Sequences Data Set for PCA. A data table was created so each row represented a strain (N observations) and each column represented an AA at a

given position (K variables). Cells were then populated with zeroes (0) if a given combination of strain and AA at a specific position was absent, and ones (1) if present.

Peptides Data Set for PCA. A data table was created so that each row represented a strain (N observations) and each column represented an identified peptide. Cells contained the relative intensity of that peptide across samples; that is 0 (not measured) to 100 (greatest intensity recorded). The intensity of the peptide is a measure of quantity of the peptide and can only be compared within the same peptide and not between different peptides.

O2PLS Data Set. Identified peptides from α_{S1} -CN and β -CN were used as X variables and the first 800 AA positions in the CEP as Y variables in one model, and suggested AA located in the substrate binding area of CEP as Y variables in another model.

One-way ANOVA using the software R (version 2.15.1, R Development Core Team, Vienna, Austria) was used to detect significant differences in peptide levels in the subgroups.

RESULTS AND DISCUSSION

Grouping of *Lactococcus* Based on CEP AA Sequence

Increased access, both faster and cheaper, to full genome sequencing, make it possible to compare CEP sequences in a larger number of strains. In this study, it was possible to align 1,608 AA residues, except for one strain where only 1,255 AA could be aligned. The cell wall spacer domain of *Lc. lactis* CEP, which has a length of 60 AA, is duplicated in some strains; for example, SK11 (Vos et al., 1989). Coverage data and conservation to CEP AA sequence of this region in the de novo assemblies strongly indicated that this was also the case for some strains in this study, especially for the sequence assembled from short single reads. The sequences of the duplicated region could not be resolved and the far C-terminal end was therefore excluded from the data analysis.

Comparing the CEP AA sequences of the *Lc. lactis* strains in this study with Exterkate's classification system showed that groups b and c were not represented among the strains in this study. Furthermore, 7 out of 38 strains showed AA variations outside Exterkate's classification, by combining features from several groups; for example, c-e, or h-e (Figure 1).

A PCA model of the first 800 AA positions of the N-terminal of *Lc. lactis* CEP, in which 70 positions varied, was carried out (model 1). The PCA model explained 68.6% of the AA variation with 4 principal

					Varying amino acids in the <i>Lactococcus</i> CEP sequence within the first 800 amino acid positions																																		
Strain name	Exterkate/ Broadbents classification	Model 1 (800 AA)	Model 2 (peptides)	<i>L. lactis</i> subsp.	2	31	44	60	62	89	131 *	138 *	142 *	144 *	166 *	177 *	201	207	272	290	291	311	313	321	339	342	366	373	381	576	602	682	694	712	747 *	748 *	755	763 *	786
2871	d	D2	D2	<i>lactis</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	A	V	M	T	L	M	G	A	H	I	E	R	K	H	H	S
1915	d	D2	D2	<i>lactis</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	S	V	M	T	I	M	G	A	H	I	E	R	K	H	H	S
9867	d	D2	D2	<i>cremoris</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	S	V	M	T	I	M	G	A	H	I	E	R	K	H	H	S
11886	d	D2	D2	<i>lactis</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	A	V	M	T	L	M	G	A	H	I	E	R	K	H	H	S
11860	d	D2	D2	<i>lactis</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	A	V	M	T	L	M	G	A	H	I	E	R	K	H	H	S
270	d	D2	D2	<i>lactis</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	S	V	M	T	I	M	G	A	H	I	E	R	K	H	H	S
10651	d	D2	D2	<i>cremoris</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	S	V	M	T	I	M	G	A	H	I	E	R	K	H	H	S
15503	d	D2	X2	<i>cremoris</i>	A	S	D	D	V	T	T	T	A	L	D	L	G	V	D	K	V	T	A	S	V	M	T	I	M	G	A	H	I	E	R	K	H	H	S
2863	d	D2	D2	<i>cremoris</i>	A	S	D	D	V	K	T	T	A	L	D	L	G	V	D	K	V	T	A	S	V	M	T	I	M	G	A	H	I	E	R	K	H	H	S
3052	d	D1A	D1A	<i>lactis</i>	A	S	D	D	A	T	T	T	A	L	D	L	G	I	N	A	L	T	A	A	V	M	T	L	M	G	A	Y	T	D	R	K	H	H	S
4430	d	D1A	X2	<i>lactis</i>	A	S	D	D	A	T	T	T	A	L	D	L	G	I	N	A	L	T	A	A	V	M	T	L	M	G	A	Y	T	D	R	K	H	H	I
11290	d	D1A	D1A	<i>lactis</i>	A	S	D	N	A	T	T	T	A	L	D	L	G	V	N	A	L	T	A	A	V	M	T	L	M	G	A	Y	T	D	R	K	H	H	N
11277	d	D1A	D1A	<i>lactis</i>	A	S	D	N	A	T	T	T	A	L	D	L	G	V	N	A	L	T	A	A	V	M	T	L	M	G	A	Y	T	D	R	K	H	H	N
3142	d	D1A	D1A	<i>lactis</i>	A	S	D	N	A	T	T	T	A	L	D	L	G	V	N	A	L	T	A	A	V	M	T	L	M	G	A	Y	T	D	R	K	H	H	S
2281	d	D1A	D1A	<i>lactis</i>	A	S	D	N	A	T	T	T	A	L	D	L	G	V	N	A	L	T	A	A	V	M	T	L	M	G	A	Y	T	D	R	K	H	H	S
11625	d	D1A	D1A	<i>lactis</i>	A	S	D	N	A	T	T	T	A	L	D	L	G	V	N	A	L	T	A	A	V	M	T	L	M	G	A	Y	T	D	R	K	H	H	S
NCDO736	d	D1B	D1B	<i>cremoris</i>	A	S	D	D	A	T	T	T	A	L	D	L	G	V	D	K	V	T	A	A	V	M	A	L	L	G	A	Y	T	D	R	K	H	H	S
6289	d	D1B	D1B	<i>cremoris</i>	A	S	D	D	A	T	T	T	A	L	D	L	G	V	D	K	V	T	A	A	V	M	A	L	L	G	A	Y	T	D	R	K	H	H	S
278	d	D1B	D1B	<i>cremoris</i>	A	S	D	D	A	T	T	T	A	L	D	L	G	V	D	K	V	T	A	A	V	M	A	L	L	G	A	Y	T	D	R	K	H	H	S
15029	d	D1B	D1B	<i>cremoris</i>	A	S	D	D	A	T	T	T	A	L	D	L	G	V	D	K	V	T	A	A	V	M	A	L	L	G	A	Y	T	D	R	K	H	H	S
10812	d	D1B	X2	<i>lactis</i>	A	S	D	N	A	T	T	T	A	L	D	L	S	V	D	K	V	T	A	A	V	M	T	L	M	R	V	Y	T	D	R	K	H	H	S
11386	f	D1B	X2	<i>lactis</i>	A	S	D	D	V	K	T	T	A	L	D	L	S	V	D	K	V	T	A	A	V	M	T	L	M	R	V	H	T	E	L	T	N	H	S
11996	f	D1B	X2	<i>lactis</i>	A	S	N	D	A	T	T	T	A	L	D	L	G	V	N	K	V	T	A	A	V	M	T	L	M	R	V	Y	T	D	L	T	N	H	S
1182	f	D1B	X2	<i>lactis</i>	A	S	N	D	A	T	T	T	A	L	D	L	G	V	N	K	V	T	A	A	V	M	T	L	M	R	V	Y	T	D	L	T	N	H	S
6005	h	X1	X1	<i>lactis</i>	T	T	N	D	A	T	T	T	A	L	G	L	S	V	D	K	V	T	A	A	V	M	T	L	M	R	V	H	T	E	L	T	N	H	S
3912	h	X1	X1	<i>lactis</i>	T	T	N	D	A	T	T	T	A	L	G	L	S	V	D	K	V	T	A	A	V	I	T	L	M	R	V	H	T	E	L	T	N	H	S
11154	e	X1	X1	<i>cremoris</i>	S	T	N	D	A	T	T	T	A	L	G	L	S	V	D	A	V	T	A	A	V	M	T	I	M	G	A	H	I	E	L	T	H	N	S
10778	h-c-e	X1	X1	<i>lactis</i>	A	T	N	D	A	A	T	T	A	L	G	I	S	V	D	T	V	T	A	A	V	I	T	L	L	G	A	H	T	E	L	T	N	N	S
10675	h-e	X2	X2	<i>lactis</i>	A	S	N	D	A	T	T	T	A	L	G	L	S	V	D	A	L	S	D	A	L	I	T	L	Q	G	A	Y	T	D	L	T	N	N	S
10776	h-e	X2	X2	<i>lactis</i>	A	S	N	D	A	T	T	T	A	L	G	L	S	V	D	A	L	S	D	A	L	I	T	L	Q	G	A	Y	T	D	L	T	N	N	S
SK11	a	X2	X2	<i>cremoris</i>	A	S	D	D	V	K	S	K	S	V	D	K	A	V	N	L	G	V	S	A	E	H	D	F	I	A	L	F	S	D	K	A	T	M	S
HP-1040	g	X2	X2	<i>cremoris</i>	A	S	D	D	V	K	T	T	D	L	D	I	G	V	D	K	V	T	D	A	V	M	T	L	L	G	A	Y	T	D	L	T	H	H	S
207	g	X2	X2	<i>cremoris</i>	A	S	D	D	V	K	T	T	D	L	D	I	G	V	D	K	V	T	D	A	V	M	T	L	L	G	A	Y	T	D	L	T	H	H	S
3366	c-e	X2	X2	<i>cremoris</i>	A	S	D	D	V	K	T	T	A	L	D	I	G	V	D	A	L	T	D	A	V	I	T	L	L	G	A	Y	T	D	L	T	N	N	S
3053	e	X2	X2	<i>cremoris</i>	A	S	D	D	A	T	T	T	S	L	D	L	G	V	D	T	V	S	D	A	V	M	T	L	L	G	A	Y	T	D	L	T	H	N	S
Wg2	e	X2	X2	<i>cremoris</i>	A	S	D	D	A	T	T	T	S	L	D	L	G	V	D	A	L	S	D	A	V	M	T	L	L	G	A	Y	T	D	L	T	H	N	S
1844	e	X2	X2	<i>lactis</i>	A	S	D	D	A	T	T	T	S	L	D	L	G	V	D	A	L	S	D	A	V	M	T	L	L	G	A	Y	T	D	L	T	H	N	S
10155	c-e	X2	X2	<i>lactis</i>	A	S	D	D	V	K	T	T	A	L	D	I	G	I	T	A	L	T	A	A	V	M	T	I	M	R	A	H	I	E	L	T	H	N	S

Figure 1. Amino acid positions suggested as being important in a principal components analysis model based on the variation in the first 800 AA positions of the cell envelope proteinase (CEP) sequence. Asterisks (*) indicate AA positions from Exterkate's classification group a-h (Exterkate et al., 1993; Broadbent et al., 1998); Gray shading indicates AA not found in the multivariate model because of low variation. Groups D and X represent the new classification in the current study defined in the text.

Figure 2. Dendrogram of a principal components analysis model based on the first 800 AA sequence of the N-terminal of the cell envelope proteinase (CEP; model 1). Four principal components were used to explain 68.6% of the AA variation and the model robustness was 46.9%. (s) refers to the groupings based on the CEP AA sequences. Groups D and X are new classification groups described in the text. First letter in the sample name refers to Exterkate's classification (Exterkate et al., 1993; Broadbent et al., 1998).

components (**PC**) and a model robustness of 46.9%. Furthermore, the model suggested that 31 AA positions were important to group the *Lc. lactis* strains (Figure 1). Including more AA in the model, from position 800 to 1,608, resulted in a decrease of the explained variation and model robustness (data not shown). This is in accordance with Siezen (1999), who found that the C-terminal region from AA positions 800 to 1,608 is not involved in substrate binding.

The dendrogram (Figure 2) depicts the *Lc. lactis* strains as 3 major groups. All 21 strains that were defined as CEP group d according to Exterkate's classification system were arranged in 2 of the major groups, called D1(s) and D2(s), where (s) refers to the groupings based on the CEP AA sequences. Group D1(s) was explained in the first PC and D2(s) was explained in second PC. Group D1 was further subgrouped into D1A and D1B, which were explained in the third PC. Together with group D1B(s) were 3 strains, which Exterkate et al. (1993) defined as group f. The third major group, called X(s), contained the rest of the investigated *Lc. lactis* strains. This group could be further subdivided into 2 groups, X1(s) and X2(s). Group X1(s) contained 2 strains, which Broadbent et al. (1998), who extended Exterkate's classification system, defined as group h, and 2 strains that were defined as group h-c-e and h-e in this study (Figure 1). Group X2 contained a mix of 6 strains, which Exterkate defined as group a, e, g together with 3 strains, which were defined as h-e and c-e in this study.

The major group of bacteria, 21 strains, that were classified as group d according to Exterkate's classification system can displace the groupings by eliminating important information. An example of this was seen in the grouping of *Lc. lactis* ssp. *cremoris* SK11, which was the only strain in this study defined as group a by Exterkate's classification system; SK11 grouped together with strains that were defined as groups e and g according to Exterkate's classification system. Furthermore, the AA positions 131, 138, and 144, which Exterkate et al. (1993) used to define group a, were eliminated in this study because of low variation.

Modeling of CEP

Using subtilisin, only a relatively small part of *Lc. lactis* CEP can be modeled, as the latter has a mature length of 1,715 to 1,775 residues (Kunji et al., 1996), whereas mature subtilisin only has 275 AA. In contrast, using the more recently released C5a peptidase structure as a template, it was possible here to model the first 922 AA positions of mature *Lc. lactis* CEP, thus covering the protease domain as well as the insert domain and much of the A-domain (Siezen, 1999; Figure 3). Further, when doing a BlastP search of SK11 CEP against C5a peptidase and subtilisin, we obtained much better results (based on scores and E-value) for the former; that is, scores of 646 versus 195, and E-values of 4×10^{-68} versus 5×10^{-15} , respectively, where E-value is the expectation value and a lower E-value

is better. A limitation when using the structure of the C5a peptidase as template is that positions 185 to 205, including the catalytic His-193, seem to be somewhat displaced (Brown et al., 2005), which translates to a similar uncertainty about the exact location of residues 82 to 106 in the derived CEP computer model. Consequently, caution is needed when referring to this part of the model.

We then combined information from the sequence analysis of our CEP with the 3-dimensional model by marking all positions showing AA variation on the 3-dimensional model. Several of the variable positions were located near the active site or in positions that have previously been suggested as being involved in specificity of the proteases (Vos et al., 1991; Exterkate et al., 1993). Of these, based on location, we considered 5 positions (31, 201, 222, 712, and 755) as the most likely potential additional contributors to specificity (Figure 2). A further 5 positions (134, 207, 427, 666, and 745) were also selected as potential contributors to specificity. Because position 89 was variable and its

exact location uncertain (see above), this position was also selected. Figure 1 does not show the AA positions 134, 222, 427, 666, and 745 because of their very low variation within the different *Lc. lactis* strains.

This study identified the structure (filename: 1xf1.pdb) of the cell wall peptidase C5a from *Streptococcus agalactiae* (Brown et al., 2005) as a more suitable template for *Lc. lactis* CEP modeling than the previously used structure of subtilisin (Vos et al., 1991); however, the model still contains limitations. The major difference between *Lc. lactis* CEP and *Streptococcus* CEP is the lack of the B-domain (Savijoki et al., 2006); although this domain does not seem to be close to the substrate binding region, we cannot exclude the possibility that the B-domain is involved in other ways.

Grouping of *Lc. lactis* Based on Identified Peptides After Hydrolysis in Milk

The peptide profile analysis identified 460 different peptides divided into 108 peptides released from α_{S1} -

Figure 3. Three-dimensional model of cell envelope proteinase (CEP) structure of subtilisin (1LW6) on the left, and model of *Lactococcus lactis* ssp. *cremoris* SK11 CEP, based on C5a protease (1xf1) on the right. Blue = active site residues, green = 9 AA positions in Exterkate's classification (Exterkate et al., 1993; Broadbent et al., 1998), red = 5 new suggested AA positions in red, black = 5 additional suggested AA positions (further information in the text).

Figure 4. Dendrogram of a principal components analysis model based identified peptides released from α_{S1} - and β -casein (model 2). Five principal components were used to explain 67.2% of the variation within the identified peptides released from α_{S1} -CN and β -CN and model robustness was 42.5%. (s) refers to the groupings based on the CEP AA sequences. Groups D and X are new classification groups described in the text. First letter in the sample name refers to Exterkate's classification (Exterkate et al., 1993; Broadbent et al., 1998).

CN, 89 peptides released from α_{S2} -CN, 187 peptides released from β -CN, and 76 peptides released from κ -CN. The PCA models of α_{S1} -CN, α_{S2} -CN, β -CN, and κ -CN were compared separately and together.

We decided to continue with a model based on only α_{S1} -CN and β -CN (model 2) because of the low contributions from α_{S2} -CN and κ -CN. Model 2 explained 67.2% with 5 PC and had a robustness of 42.5% (model 2). A good model was also found when peptides from β -CN were modeled alone, where 6% more variations was explained but with the use of 1 additional PC and the same robustness. The grouping of the peptides released from α_{S1} -CN and β -CN showed 2 major groups, called D(p) and X(p), where p refers to the groupings based on the identified peptides. Group D(p) was subgrouped into D1(p) and D2(p) and further subgrouping of D1(p) into D1A(p) and D1B(p) was similar to that for D1(s), D1A(s), and D1B(s) in model 1 (Figure 4). The only exceptions in model 2 compared with model 1 were that 1 strain from each of the 3 D-subgroup arranged into group X2(p). The 3 subgroupings D1A, D1B, and D2 created on the basis of the identified peptides (model 2) was also created based on subgrouping of the 800 AA positions in the CEP (model 1), independently of each other. Three strains defined as group f by Exterkate's classification system that were arranged in group D1B(s) in model 1 were arranged in group X2(p) in model 2. Group X1 was identical in the 2 models, and subgrouping of X2 was possible in both models.

A low correlation was found when *Lc. lactis* strains were grouped based on peptides released from α_{S1} -CN (f1–23), which were the foundation peptides in Exterkate's classification system. This could be due to the changes of CEP activity due to different media and changes in the specificity between purified and native bound CEP (Exterkate and Alting, 1993; Flambard and Juillard, 2000). The current study only identified the peptides after 24 h of incubation, which included cell uptake of peptide fragments and further breakdown of peptides.

Correlation Between Key AA Positions in the CEP and Identified Peptides

A comparison was made between the 2 O2PLS models that had the identified peptides from α_{S1} -CN and β -CN as X variables, model 3 with the first 800 AA positions as Y variable, and model 4 with the 9 suggested AA positions used in Exterkate's classification system together with the 11 new (suggested) AA positions from the 3-D modeling in this study as Y variables (data not shown). Model 3 indicated that the same 31 AA positions as in model 1 were important for the specificity, whereas model 4 indicated that 12 out of the 20 AA positions were important. The 12 AA positions (31, 89, 142, 166, 177, 201, 207, 712, 747, 748, 755, 763) suggested in model 4 were part of model 3. Models 3 and 4 explained 84.4% and 60.7%, respectively, of the variations of peptides from α_{S1} -CN and β -CN and

Table 1. Overview of peptides released from α_{S1} -CN and β -casein that were significant ($P < 0.05$) between the groups

Peptides	Subgroup ¹			Peptides	Subgroup ¹		
	D1A	D1B	D2		D1A	D1B	D2
Peptide in group D1A				Peptide in group D2			
α_{S1} (1–6) ²	13	1	7	α_{S1} (126–135) ²	12	13	8
α_{S1} (1–8)	8	3	2	α_{S1} (179–192)	20	61	33
α_{S1} (1–18) ^{3,4}	59	38	38	α_{S1} (194–199)	9	13	10
α_{S1} (1–21)	13	10	5	β (1–14) ⁴	11	16	10
α_{S1} (14–25)	21	8	2	β (25–34)	10	17	10
α_{S1} (17–23) ^{3,4}	55	43	38	β (44–57)	15	34	24
α_{S1} (24–32)	69	37	44	β (69–82) ^{2,4}	29	37	19
α_{S1} (24–34) ³	2	1	0	β (81–86) ^{2,3,4}	2	5	1
α_{S1} (25–36)	1	0	0	β (81–87) ^{3,4,5}	0	1	0
α_{S1} (33–43) ³	80	31	31	β (92–98) ³	40	46	28
α_{S1} (51–55) ⁴	26	14	23	β (101–107)	28	35	21
α_{S1} (85–89) ³	55	30	16				
α_{S1} (96–101) ³	26	15	11	α_{S1} (1–13) ^{2,4}	22	20	32
α_{S1} (97–104) ^{3,4,5}	74	47	31	α_{S1} (1–14) ^{3,4}	31	36	57
α_{S1} (100–105)	37	37	24	α_{S1} (14–19) ^{3,4}	7	10	44
α_{S1} (124–128)	30	15	25	α_{S1} (24–30) ²	50	48	61
α_{S1} (131–139) ³	77	40	38	α_{S1} (31–38) ⁴	6	5	10
β (21–28) ⁴	17	12	7	α_{S1} (31–40) ³	37	32	45
β (22–33) ^{3,5}	30	19	11	α_{S1} (55–60) ^{3,4}	25	21	46
β (29–39)	60	29	36	α_{S1} (55–61)	37	18	47
β (38–45)	42	25	26	α_{S1} (56–63) ⁴	21	22	37
β (40–48) ³	17	6	13	α_{S1} (85–93) ⁴	39	22	47
β (42–46) ⁵	72	36	44	α_{S1} (93–98) ³	45	50	64
β (43–47) ⁴	34	22	26	α_{S1} (96–102) ³	13	9	16
β (44–52) ^{2,3,5}	45	30	23	α_{S1} (96–103)	25	22	35
β (44–53)	20	18	12	α_{S1} (99–108) ²	2	5	8
β (69–77) ^{2,3}	56	35	40	α_{S1} (99–109) ^{3,4}	32	25	59
β (78–93) ⁴	75	40	43	α_{S1} (109–114) ^{2,3,4}	4	4	11
β (92–100)	9	9	2	α_{S1} (118–124)	11	11	15
β (92–101) ^{3,4}	7	3	2	α_{S1} (124–130) ³	58	40	62
β (94–102)	22	8	2	α_{S1} (136–142)	5	5	12
β (94–106) ³	68	28	35	α_{S1} (143–148) ²	48	34	57
β (94–121)	35	9	22	α_{S1} (176–189) ³	50	26	51
β (94–122)	63	19	41	β (1–6) ^{3,4}	46	29	53
β (94–123) ⁴	48	10	35	β (22–27)	20	20	59
β (97–105) ²	46	32	21	β (25–31)	9	9	14
β (109–119) ^{2,3,5}	6	2	1	β (25–32)	38	27	48
β (120–131) ⁴	12	6	3	β (28–36)	46	42	65
β (126–132)	35	24	22	β (46–56) ⁴	24	27	33
β (132–139) ²	6	2	4	β (56–68) ^{2,4}	48	37	52
β (132–142) ³	44	18	18	β (58–72) ^{2,4}	29	24	32
β (134–141)	66	50	40	β (83–91) ^{2,3}	4	5	8
β (134–143) ^{3,5}	50	42	25	β (92–97)	35	48	66
β (144–154) ^{2,3}	29	17	16	β (168–176) ^{3,4}	5	5	9
β (144–160) ^{2,4}	31	16	13				
β (145–156)	37	20	17				
β (155–163) ⁴	31	19	18				
β (161–168) ^{3,4}	89	65	58				
β (167–175) ²	87	71	67				
β (176–182) ²	73	60	57				
β (176–188) ³	6	2	1				
β (183–190) ^{3,4,5}	62	31	38				
β (183–192)	41	23	25				
β (183–193) ⁵	74	64	51				
β (194–209)	76	57	51				
β (199–208)	11	5	10				
β (201–206)	66	0	18				
Peptide in group D1B							
α_{S1} (1–9) ^{3,4}	23	32	15				
α_{S1} (10–16)	19	33	17				
α_{S1} (18–24) ^{3,4,5}	28	30	20				
α_{S1} (22–35) ⁴	11	21	2				
α_{S1} (80–90) ⁴	19	24	14				

Continued

¹The 3 subgroups represent *Lactococcus lactis* strains in the groups in both model 1 and model 2: D2 (n = 8), D1A (n = 3), D1B (n = 6). The peptides are represented in the subgroup by their peak intensity. The peak intensity is an average of 2 replications and normalized from 0 to 100 between the samples.

²Proline in P2 (nomenclature according to Schechter and Berger, 1967).

³Hydrophobic nonaromatic AA in P4 (Ala, Ile, Leu, or Val; nomenclature according to Schechter and Berger, 1967).

⁴Hydrophobic nonaromatic AA in P2' (nomenclature according to Schechter and Berger, 1967).

⁵Hydrophobic aromatic side chain in P1 (Phe, Trp, or Tyr; nomenclature according to Schechter and Berger, 1967).

81.7% and 61.6% of the AA variations with 5 and 3 PC, respectively. The robustness of the 2 models was 54.9% and 45.7%, respectively. The major difference between the 2 models was that model 3 was able to arrange group D into 3 subgroups, D1A, D1B, and D2, as described in models 1 and 2. Model 4 arranged the group D into 2 subgroups, D1 and D2.

Five AA positions (272, 290, 291, 366, 381) were able to distinguish group D1A and D1B; however, none of the AA positions were conserved in the major group X (Figure 1). The 3-dimensional model of CEP showed that these 5 positions were opposite the substrate binding region (Figure 3); however, they could be important in substrate specificity. This correlated with findings of Bruinenberg et al. (1994), who found that the removal of the external loop insertion (residues 238–388) substantially altered the caseinolytic specificity of the enzyme but did not inhibit folding and activity of the *Lc. lactis* ssp. *cremoris* SK11. Together, these findings indicate that the specificity of the CEP is highly complex.

Distinction Between Subgroups of Group D

It was not possible to separate the 3 D-subgroups with unique peptides. All groups contained overall the same peptides but with varying intensities. Table 1 shows the average intensity of each significant peptide within the 3 groups. Each subgroup is represented by the peptides that had the highest intensity of the respective peptide. Group D1A contained the greatest number of peptides with highest intensity, especially within peptides released from β -CN. Group D1B contained the smallest number of peptides. However, no exact pattern between the peptides was seen, which could be due to peptidase activity during hydrolysis.

A tendency of general specificities of cleavage with the nomenclature of Schechter and Berger (1967) could be seen in this study (Table 1). In this current study, only the CEP cleavage specificity of the peptide C-terminal was investigated. Aminopeptidase activities from whole cells were present in the milk and responsible for the cleavage of the peptides N-terminal and thereby excluded. In general, higher cleavage was observed when proline was in position P2, which agrees with previous studies (Exterkate, 1995) but this could also be due to the large number of proline residues in β -CN (Fox and McSweeney, 1998). No cleavage of a peptide was found if proline was placed in P1 or P1'. In general, higher cleavage specificity was observed when an AA with a hydrophobic nonaromatic side chain (e.g., alanine, isoleucine, leucine, or valine) was placed in P4, in agreement with Exterkate (1995), or in position P2'. Group DA1 had higher cleavage specificity if an AA

with a hydrophobic aromatic side chain (e.g., phenylalanine, tryptophan, or tyrosine) was placed in position P1 compared with group D2.

It was not possible to explain variation of the identified peptides within the subgroups of D; however, similar subgroups were seen in both the derived peptides from α_{S1} -CN and β -CN and CEP AA sequences, independent of each other. The grouping of *Lc. lactis* could be changed or further clarified if more strains from groups other than d were included and if phosphorylated peptides were included in the peptide identification. This study indicates that the proteolytic specificity of *Lc. lactis* is reflected in the CEP AA sequence. In the future, as genome sequencing is more frequently implemented, a new screening method of *Lc. lactis* proteolytic activity could be conducted by investigating the CEP AA sequence.

ACKNOWLEDGMENTS

The authors thank the Danish Agency for Science, Technology and Innovation (Copenhagen) for partly financing the PhD studies of Mette Winther Børsting.

REFERENCES

- Broadbent, J. R., M. Strickland, B. C. Weimer, M. E. Johnson, and J. L. Steele. 1998. Peptide accumulation and bitterness in Cheddar cheese made using single-strain *Lactococcus lactis* starters with distinct proteinase specificities. *J. Dairy Sci.* 81:327–337.
- Brown, C., Z. Gu, Y. Matsuka, S. Purushothaman, L. Winter, P. Cleary, S. Olmsted, D. Ohlendorf, and C. Earhart. 2005. Structure of the streptococcal cell wall C5a peptidase. *Proc. Natl. Acad. Sci. USA* 102:18391–18396.
- Bruinenberg, P. G., P. Doesburg, A. C. Alting, F. A. Exterkate, W. M. de Vos, and R. J. Siezen. 1994. Evidence for a large dispensable segment in the subtilisin-like catalytic domain of the *Lactococcus lactis* cell-envelope proteinase. *Protein Eng.* 7:991–996.
- Chopin, A. 1993. Organization and regulation of genes for amino acid biosynthesis in lactic acid bacteria. *FEMS Microbiol. Rev.* 12:21–37.
- Eriksson, L., T. Byrne, E. Johansson, J. Trygg, and C. Vikström. 2013. Multi- and Megavariate Data Analysis. 3rd ed. Umetrics Academy, Malmö, Sweden.
- Exterkate, F. A. 1979. Accumulation of proteinase in the cell wall of *Streptococcus cremoris* strain AM1 and regulation of its production. *Arch. Microbiol.* 120:247–254.
- Exterkate, F. A. 1985. A dual-directed control of cell wall proteinase production in *Streptococcus cremoris* AM1: A possible mechanism of regulation during growth in milk. *J. Dairy Sci.* 68:562–571.
- Exterkate, F. A. 1995. The lactococcal cell envelope proteinases: Differences, calcium-binding effects and role in cheese ripening. *Int. Dairy J.* 5:995–1018.
- Exterkate, F. A., and A. C. Alting. 1993. The conversion of the α_{S1} -casein-(1–23)-fragment by the free and bound forms of the cell envelope proteinase of *Lactococcus lactis* ssp. *cremoris* under conditions prevailing in cheese. *Syst. Appl. Microbiol.* 16:1–8.
- Exterkate, F. A., A. C. Alting, and P. G. Bruinenberg. 1993. Diversity of cell envelope proteinase specificity among strains of *Lactococcus lactis* and its relationship to charge characteristics of the substrate-binding region. *Appl. Environ. Microbiol.* 59:3640–3647.
- Flambard, B., and V. Juillard. 2000. The autoproteolysis of *Lactococcus lactis* lactocepin III affects its specificity towards beta-casein. *Appl. Environ. Microbiol.* 66:5134–5140.

- Fox, P. F., and P. L. H. McSweeney. 1998. Milk proteins. Pages 146–238 in *Dairy Chemistry and Biochemistry*. P. F. Fox and P. L. H. McSweeney, ed. Kluwer Academic/Plenum Publishers, New York, NY.
- Jensen, M. P., F. K. Vogensen, and Y. Ardö. 2009. Variation in caseinolytic properties of six cheese related *Lactobacillus helveticus* strains. *Int. Dairy J.* 19:661–668.
- Kim, S., N. Mischerikow, N. Bandeira, J. D. Navarro, L. Wich, S. Mohammed, A. J. R. Heck, and P. A. Pevzner. 2010. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: Applications to database search. *Mol. Cell. Proteomics* 9:2840–2852.
- Kunji, E. R. S., I. Mierau, A. Hagting, B. Poolman, and W. N. Konings. 1996. The proteolytic systems of lactic acid bacteria. *Antonie van Leeuwenhoek* 70:187–221.
- Reid, J. R., and T. Coolbear. 1998. Altered specificity of lactococcal proteinase PI (lactocepin I) in humectant systems reflecting the water activity and salt content of cheddar cheese. *Appl. Environ. Microbiol.* 64:588–593.
- Reid, J. R., and T. Coolbear. 1999. Specificity of *Lactococcus lactis* ssp. *cremoris* SK11 proteinase, lactocepin III, in low-water-activity, high-salt-concentration humectant systems and its stability compared with that of lactocepin I. *Appl. Environ. Microbiol.* 65:2947–2953.
- Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.
- Savijoki, K., H. Ingmer, and P. Varmanen. 2006. Proteolytic systems of lactic acid bacteria. *Appl. Microbiol. Biotechnol.* 71:394–406.
- Schechter, I., and A. Berger. 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27:157–162.
- Siezen, R. J. 1999. Multi-domain, cell-envelope proteinases of lactic acid bacteria. *Antonie van Leeuwenhoek* 76:139–155.
- Smith, C., E. Want, G. O'Maille, R. Abagyan, and G. Siuzdak. 2006. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78:779–787.
- Upadhyay, V. K., P. L. H. McSweeney, A. A. A. Magboul, and P. F. Fox. 2004. Proteolysis in cheese during ripening. Page 391–433 in *Cheese: Chemistry, physics and microbiology*. Vol. 1. General Aspects. 3rd ed. P. F. Fox, P. L. H. McSweeney, T. M. Cogan, T. P. Guinee, ed. Elsevier Academic Press, London, UK.
- Visser, S., F. A. Exterkate, C. J. Slangen, and G. J. C. M. de Veer. 1986. Comparative study of action of cell wall proteinases from various strains of *Streptococcus cremoris* on bovine α_{S1} -, β -, and κ -casein. *Appl. Environ. Microbiol.* 52:1162–1166.
- Vos, P., I. J. Boerrigter, G. Buist, A. J. Haandrikmann, M. Nijhuis, M. B. Dereuver, R. J. Siezen, G. Venema, W. M. de Vos, and J. Kok. 1991. Engineering of the *Lactococcus lactis* serine proteinase by construction of hybrid enzymes. *Protein Eng.* 4:479–484.
- Vos, P., G. Simons, R. J. Siezen, and W. M. Devos. 1989. Primary structure and organization of the gene for a procaryotic, cell envelope-located serine proteinase. *J. Biol. Chem.* 264:13579–13585.