2018

Yuxi Deng

Towards predicting enzymatic protein hydrolysis

Towards predicting enzymatic protein hydrolysis
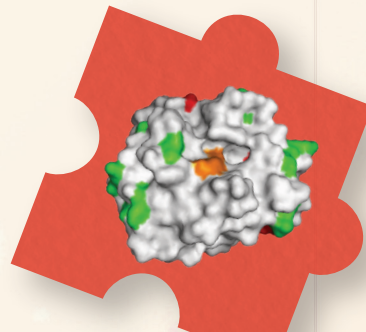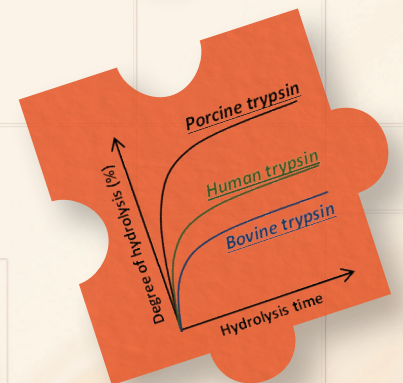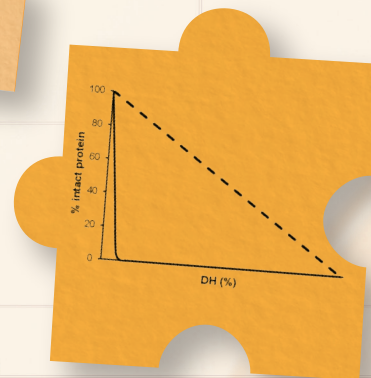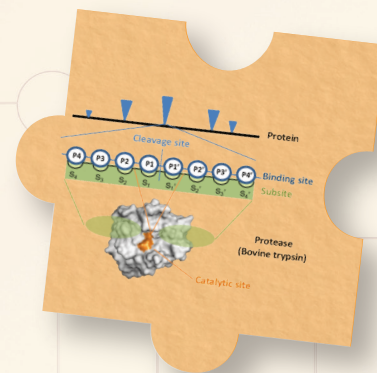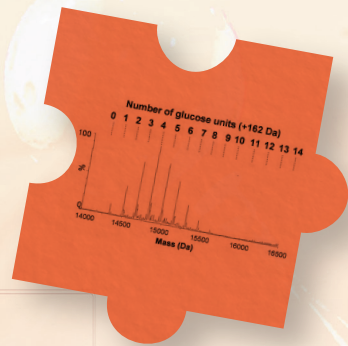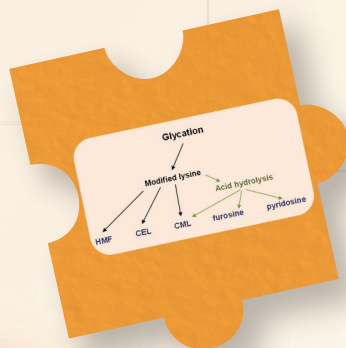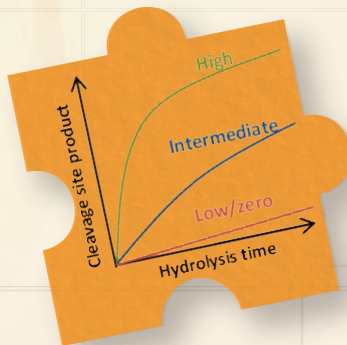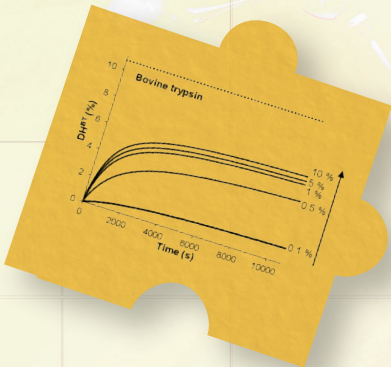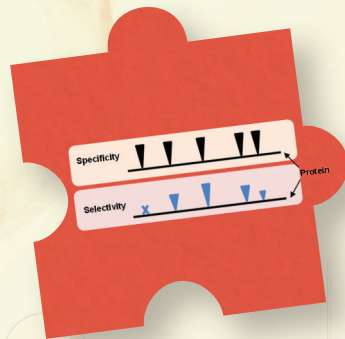
Yuxi Deng

Invitation

You are cordially invited to the public defence of my PhD thesis entitled:

Towards predicting enzymatic protein hydrolysis

on **Friday, 20th April 2018** at **13:30**, at **the Aula of Wageningen University & Research**

After the defence, there will be a buffet and party starting around **6 pm** at **Cafe Loburg, Molenstraat 6, 6701 DM, Wageningen**

Yuxi Deng

deng_yu_xi@hotmail.com

Paranymphs:
Yue Han
yue.han@wur.nl
Matthias Frommhagen
matthias.frommhagen@wur.nl

Propositions

1. The extent of enzymatic protein hydrolysis can be predicted for any given protein by any specific protease.
   (this thesis)

2. The Linderstrøm-Lang theory describes the condition dependent product inhibition.
   (this thesis)

3. Young obsessive-compulsive disorder patients should be checked for early onset Alzheimer's disease.

4. Co-extraction of lignin carbohydrate complexes is often misstated as a proof of covalent linkages.

5. People often wrongly ask for other people's experiences rather than advices to make decisions.

6. Memory is a cognitive skill that is often underestimated in scientific education.

<div align="center">

Propositions belonging to the thesis, entitled

"Towards predicting enzymatic protein hydrolysis"

Yuxi Deng

Wageningen, 20 April 2018

</div>

Towards predicting enzymatic protein hydrolysis

Yuxi Deng

Thesis committee

Promotor
Prof. Dr H. Gruppen
Professor of Food Chemistry
Wageningen University & Research

Co-promotors
Dr P.A. Wierenga
Assistant professor, Laboratory of Food Chemistry
Wageningen University & Research

Prof. Dr H.A. Schols
Personal chair at the Laboratory of Food Chemistry
Wageningen University & Research

Other members
Prof. Dr M.A.J.S. van Boekel, Wageningen University & Research
Prof. Dr R.J. FitzGerald, University of Limerick, Ireland
Prof. Dr W.J.H. van Berkel, Wageningen University & Research
Dr M.D. Vlasie, DSM, Delft, The Netherlands

Towards predicting enzymatic protein hydrolysis

Yuxi Deng

Thesis

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Friday 20 April 2018

at 1.30 p.m. in the Aula.

Abstract

The experimental maximum degree of hydrolysis ($DH_{max,exp}$) was typically lower than the theoretical maximum degree of hydrolysis ($DH_{max,theo}$), calculated using the number of cleavage sites (as based on protease specificity) on the protein. The aim of this thesis is, therefore, to develop a method to better define expectations of the $DH_{max,exp}$. The developed method is based on the molecular properties of the protein and the protease, and the effects of modifications on protein primary structure and changes in system conditions. Protein primary structure was modified by glycation, the initial stage of the Maillard reaction upon heating proteins in the presence of carbohydrates. The effect of glycation on $DH_{max,exp}$ for five different enzymes could be predicted based on the protease primary and secondary specificity, and the sensitivity of the enzyme to modifications of the binding site. Secondary specificity was determined by linking the molecular properties of the amino acids on the binding site positions of cleavage sites with the enzyme selectivity. The enzyme selectivity, i.e. the relative hydrolysis rate constants towards cleavage sites in a protein, of bovine, porcine and human trypsins was determined after full quantitative analysis of peptides formed during hydrolysis. The prediction of $DH_{max,exp}$ for bovine trypsin based on the secondary specificity was 5 times better than the prediction based on the primary specificity ($DH_{max,theo}$). The large differences found in $DH_{max,exp}$ by the three trypsins could be predicted based on enzyme secondary specificity. The substrate concentration, one of the system conditions varying in the *in vitro* digestion models, also showed large effect on $DH_{max,exp}$. This was found to be related to changes in the mechanism of hydrolysis, illustrated by the differences in the percentage of intact protein in the hydrolysates as a function of DH. Due to the changes in mechanism, different amounts of inhibitory peptides for the protease were proposed to be present in the hydrolysates, leading to various $DH_{max,exp}$. Based on these findings, a model to predict the $DH_{max,exp}$ of hydrolysis of any given protein by any specific protease was developed. The concept was shown to adequately explain the hydrolysis of proteins with various structural states and in hydrolysis by a sequential addition of enzyme.

Table of contents

CHAPTER 1

General introduction

The study of *in vivo* protein digestibility has received a lot of attention from food and feed producers. *In vitro* enzymatic protein hydrolysis is often used as a screening tool to investigate the protein digestibility since the *in vivo* protein digestion experiments are not always feasible options. Enzymatic protein hydrolysis is also used to produce protein hydrolysates, which are used as ingredients in food products. The outcome of an enzymatic protein hydrolysis is affected by many factors, including the structure of the substrate protein, the protease used and their sensitivities towards system conditions. For example, the modifications on the protein structure during industrial processing, e.g. through the Maillard reaction or heating, might influence the hydrolysis of the protein. Another challenge is that different parameters are used to describe the hydrolysis process, e.g. the degree of hydrolysis, the percentage of remaining intact protein, the molecular weight distribution of the hydrolysate, the amino acid (AA) sequence and the amount of peptides formed. The current state of knowledge on enzymatic protein hydrolysis crosses several fields of study. Some studies aim to reveal the effect of system conditions or protein sources on the extent of hydrolysis. To describe the hydrolysis curve, i.e. the degree of hydrolysis *versus* time, (phenomenological) kinetic models are used. Other studies focus on investigating the molecular interactions between proteins and proteases. In proteomics studies, the outcome of tryptic hydrolysis has received some attention because tryptic hydrolysis is applied to the samples before analysis. Although a large amount of data has been acquired, due to a lack of mechanistic understanding of the hydrolysis process so far, it is difficult to compare published results. Mechanistic understanding of the hydrolysis process can help to understand protein digestion, and -in the future- to control the hydrolysate composition during production. To explain results reported in literature and to define expectations for further experiments, this thesis aims to identify the molecular properties of the protein and the protease, as well as their interactions, which determine the extent of hydrolysis. The goal is to develop a method to define expectations of the extent of enzymatic protein hydrolysis, based on the intrinsic properties of the protein and the protease, and the effects of modifications on protein primary structure and changes in system conditions.

The maximum degree of hydrolysis

The degree of hydrolysis (DH), defined as the percentage of hydrolysed peptide bonds of the total number of peptide bonds, is often used as an indication to express the extent of hydrolysis. Several methods are used to determine the DH, e.g. pH-stat, trinitrobenzenesulphonic acid (TNBS), o-phthaldialdehyde (OPA) and formol titration [2]. The pH-stat method is the most commonly used method to real-time monitor the DH as a function of time. The hydrolysis curve (DH *versus* time) is commonly used to study the hydrolysis kinetics. It must be noted that the traditionally used Michaelis-Menten model cannot be used to describe the kinetics of a protein hydrolysis, because it was developed to describe the conversion of a single substrate to a single product [3]. In many kinetic studies, a modified kinetic model based on the Michaelis-Menten model was developed to describe the experimental hydrolysis curves [4-7]. Although the

experimental hydrolysis curves were fitted nicely by those phenomenological models, they cannot identify the mechanisms that determine the final DH reached by a hydrolysis. Two characteristics of the hydrolysis are shown in the hydrolysis curve: the extent of hydrolysis (the plateau value) and the rate of hydrolysis. For hydrolysis of a given protein by a specific protease, the theoretical maximum degree of hydrolysis ($DH_{max,theo}$) can be estimated based on the AA sequence of the protein and the (primary) specificity of the protease (equation 1). Protease specificity is defined as the type(s) of AAs next to which the protease can hydrolyse a peptide bond, either at the carboxylic or amino side. The specificity of many proteases were summarised in the MEROPS database [1]. On a protein or a peptide, the AAs that corresponding to the enzyme specificity are referred to as the cleavage sites (CSs) of the protease.

$$(1) \quad DH_{max,\,theo}\left[\%\right] = \frac{\#\,CS}{\#\,peptide\,\,bonds} \times 100\,\%$$

#CS refers to the number of CSs of the protease on the substrate protein and the #peptide bonds refers to the number of peptide bonds of the substrate protein. In literature, in many cases, the reported experimental maximum degree of hydrolysis ($DH_{max,exp}$) did not reach the $DH_{max,theo}$, even when the hydrolysis was performed for an extensive amount of time (**Table 1.1**). For instance, after 24 h of hydrolysis of β-lactoglobulin by bovine trypsin, the $DH_{max,exp}$ only reached 8 %, which was ~71 % of the $DH_{max,theo}$ [8]. That the reached $DH_{max,exp}$ was lower than the $DH_{max,theo}$ means that not all CSs in the substrate protein were efficiently hydrolysed. This showed that the extent of hydrolysis cannot be solely estimated by the AA sequence of the protein and the specificity of the protease. One possible explanation for this could be that the protease did not cleave all CSs efficiently in the substrate protein. The difference in the enzyme preference towards CSs in a substrate protein could be due to the enzyme secondary specificity, defined as the preference of protease towards neighbouring AAs of the CSs [9]. Another possible explanation for this inefficient cleavage of CSs in a substrate protein could be that the protease lost its activity during hydrolysis. This could be due to the enzyme autolysis and/or the formation of inhibitory peptides for the protease during the hydrolysis.

Table 1.1: Overview of reported values of the experimental maximum degree of hydrolysis (DH$_{max,exp}$) under different hydrolysis conditions, compared to theoretical maximum degree of hydrolysis (DH$_{max,theo}$) calculated based on the protease specificity for the enzyme-substrate combinations used in this thesis.

Enzyme	Specificity#	Protein	#CS	DH$_{max,theo}$ [%]	DH$_{max,exp}$ [%]	DH$_{max,exp}$/DH$_{max,theo}$ [%]	Protein conc. [g/L]	pH	E:S	Temperature [°C]	Time [h]	Ref.
Bovine trypsin	KR	α-LA	13	10.7	9	84	/	7.8	1:25	37	5	10
		β-cas	15	7.2	5	69	/	7.6	1:100	30	/	11
		β-LG	18	11.2	8	71	45	8.0	1:650	37	24	8
		WPI	13 (α-LA) 18 (β-LG)	11.1	9.5	86	100	7.0	1:106	25	6	12
		BSA	82	14.1	7	50	25	/	1:10	37	0.4	13
Bovine α-chymotrypsin	FLY*	α-LA	21	17.2	10	58	/	7.8	1:25	37	5	10
		BSA	108	18.6	10	54	25	/	1:10	37	0.4	13
		WPI	21 (α-LA) 30 (β-LG)	17.9	7.6	42	50, 100	8.5	1:50	50	2	14
BLP	DE	WPI	20 (α-LA) 26 (β-LG)	16.0	8	50	10	8.0	1:100	40	3	15
					6.8	43	50	8.0	1:100	40	3	16

Not mentioned values were represented by "/";

The single letter codes for amino acids were used in this column.

* The specificity of α-chymotrypsin was taken from the three most frequently cleaved AAs in MEROPS database [1].

BLP: *Bacillus licheniformis* protease; α-LA: α-lactalbumin; β-cas: β-casein; β-LG: β-lactoglobulin; BSA: bovine serum albumin; WPI: whey protein isolate;

#CS: number of cleavage sites per protein; E:S: enzyme to substrate ratio

The interaction between the substrate protein and protease

Most studies of enzyme secondary specificity have built on the proposed theory in 1967 that the protease efficiency in hydrolysing a CS depends on the interactions between the catalytic environment (subsite) of the protease and the binding environment (binding site) of the CS [17] (**Figure 1.1**). In this theory, the P4-P4' positions refer to the AAs on the N- to C-terminal sides of the CS, respectively, and P1 is the AA for which the enzyme is specific [17]. The S4-S4' positions refer to the AAs near the catalytic site that bind to the P4-P4' positions, respectively [17]. Initially, the model was proposed for papain, but it has been taken as the guideline to study the secondary specificity for other proteases [18]. For example, in the MEROPS database, the enzyme preference towards AAs on the binding site positions was represented by the "specificity matrix", which contained 8 binding site positions (P4-P4') [1]. According to this theory, an unfavourable interaction between the binding site of a CS and the subsite of an enzyme would lead to an inefficiently hydrolysis of that CS.
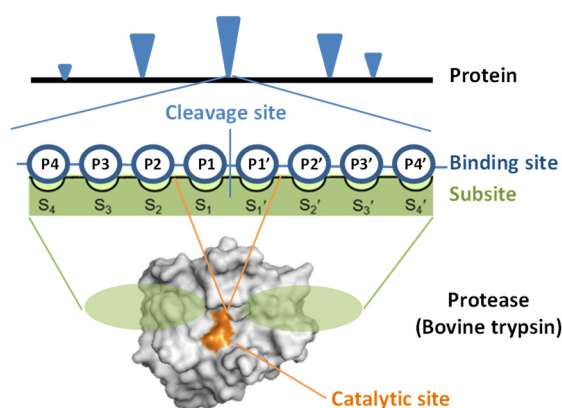
Figure 1.1: Schematic example of the interaction between the binding site (P4-P4') of a CS (the peptide bond between P1 and P1') and the subsite (S4-S4') of a protease (e.g. bovine trypsin, PDB 4I8L [19]), adapted from Schechter and Berger [17]. The black line represents the protein AA sequence and the orange area on the protease represents the catalytic site. The subsite refers to the AAs surrounding the catalytic site (green areas), of which the exact location is unknown so far. Differences in this interaction lead to different enzyme efficiencies in hydrolysing the CSs, which is represented by the blue triangles on the protein.

Enzyme secondary specificity

The secondary specificities of many proteases have been described using the model illustrated in **Figure 1.1**. For digestive enzymes, the secondary specificity of trypsin has been received a lot of attention, because of its extensive usage in proteomics [20] and in protein digestion studies [21]. For example, CSs with certain specific binding site sequences could not be hydrolysed at all by bovine trypsin [22]. This behaviour of bovine trypsin was referred to as the "*missed cleavage*" [22]. It has been proposed that the missed cleavage behaviour of bovine trypsin was due to the presence of charged AAs on the binding site sequences of the CSs [20, 22-27]. Another finding was

that bovine trypsin could not hydrolyse CSs with a proline at the P1' position of the binding site sequence [22, 28, 29]. However, these missed cleavage rules only indicated which CSs could not be hydrolysed by bovine trypsin at all, while perhaps there was a gradual scale of decreasing efficiency of the enzyme in hydrolysing CSs. These differences in enzyme efficiency of hydrolysing CSs have been observed in studies using synthetic peptides. For example, the hydrolysis rate constants of thrombin, a serine protease, of arginine (the CS) in 20 synthetic peptides with 10 different AAs on the P2' or P3' positions, were determined [30]. By altering the AA on the P2' position of the binding site sequence, the hydrolysis rate constants of the synthetic peptides ranged from $2.1 \cdot 10^6$ to $3.2 \cdot 10^3$ $M^{-1} \cdot s^{-1}$ [30]. By altering the AA on the P3' position, the hydrolysis rate constants ranged from $8.6 \cdot 10^6$ to $2.6 \cdot 10^5$ $M^{-1} \cdot s^{-1}$, meaning that thrombin was more sensitive to the AA on the P2' position than the P3' position [30]. Another study altered the P3, P2 and P1' positions with 16 different AAs of a synthetic peptide, and hydrolysed these peptides by a trypsin obtained from *Prerifplaneta americana* [31]. The hydrolysis rate constants of these peptides had a difference of 1-10 times [31]. Later on, this group applied the same strategy and expended their knowledge on enzyme secondary specificity of 4 other insect trypsins [32]. By altering the binding site sequences of the synthetic peptide, they proposed that these insect trypsins preferred CSs with hydrophobic AAs on the binding site positions [32].

Besides studies using synthetic peptides, the differences in enzyme efficiency in hydrolysing CSs were also observed in hydrolysis of proteins. It was shown that each peptide formed during the bovine tryptic hydrolysis of α_{s2}-casein had an unique release kinetics, which suggested that the enzyme preference towards each CS was different [33]. In 2013, Fernández and Riera thoroughly followed the release kinetics of each peptide during bovine tryptic hydrolysis of β-lactoglobulin, and quantified the hydrolysis rate constant of the enzyme of each CS [8]. It was found that one CS (K69-70) was not hydrolysed at all, and for the hydrolysed CSs, the difference in hydrolysis rate constants between the fastest and the slowest hydrolysed CS was up to 300 times [8]. In 2014, a quantitative method was developed to compare the relative hydrolysis rate constants of an enzyme towards CSs in a protein, referred to as the "*enzyme selectivity*" [15] (**Figure 1.2**). Enzyme selectivity is a quantitative parameter that expresses the consequence of the enzyme secondary specificity on protein hydrolysis.
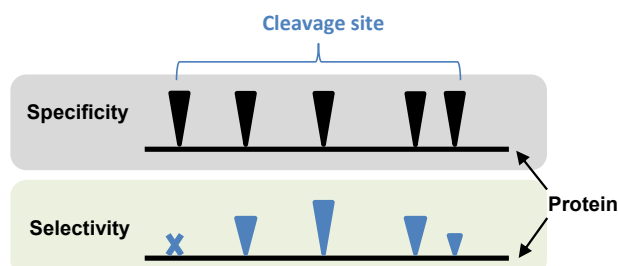

Figure 1.2: Schematic presentation of enzyme specificity and enzyme selectivity, based on the information from Butré, *et.al* [15]. The black lines represent the protein AA sequence. The size of the triangles illustrates the relative hydrolysis rate constants. The (×) refers to a missed cleavage site.

The hydrolysis rate constants of CSs in β-lactoglobulin by *Bacillus licheniformis* protease (BLP) were found to vary from 0 to $1.1 \cdot 10^{-6}$ $s^{-1} \cdot mg^{-1}_{enzyme}$, corresponding to selectivity values of 0 % to 17 % [15]. In principle, the difference in enzyme selectivity towards CSs in a protein should not directly lead to a lower extent of hydrolysis. However, because the hydrolysis rate constants of some CSs were 10 or even 100 times lower than the rate constants of the quickly hydrolysed CSs in the protein. In reality, these slowly hydrolysed CSs would not be hydrolysed completely during the time of experiments, which would lead to an apparent $DH_{max,exp}$ that is lower than $DH_{max,theo}$.

It should be noted that proteases with the same primary specificity might not have the same secondary specificity. In a bovine tryptic hydrolysate of human serum albumin, the number of peptides with CSs in the AA sequences was ~60 % higher than in a porcine tryptic hydrolysate [34]. This indicated that bovine and porcine trypsins might not have the same secondary specificity. The enzyme secondary specificity can also be modified by mutation of the enzymes. For example, the hydrolysis rate constants of CSs on β-casein by bovine trypsin and 5 trypsin mutants showed large differences [35]. In addition, chemical modifications on the enzyme might affect the enzyme secondary specificity. For instance, modifying lysine residues in bovine trypsin by succinylation led to a 8-10 times higher acceptance of positively charged AAs (KR) on the P1′ position of the binding site sequence, while the modification by guanylation showed no effects [36].

Formation of inhibitory peptides during hydrolysis

Within one enzyme-substrate combination, the extent of hydrolysis was still largely affected by system conditions (**Table 1.2**). For example, the $DH_{max,exp}$ of whey protein isolate (WPI) hydrolysis by BLP decreased from 13 % to 6 % with an increasing substrate concentration (0.1 %-10 %) [37]. The $DH_{max,exp}$ of 5 % WPI hydrolysis by BLP ranged from 4 % to 8 % with an increasing hydrolysis pH (7 to 9) [38].

Table 1.2: Overview of studies that investigated the effect of system conditions, i.e. temperature, pH, enzyme to substrate ratio (E:S) and protein concentration, on the extent of protein hydrolysis.

Protein	Enzyme	Temperature [°C]	pH	E:S (w/w)	Protein conc. [g/L]	$DH_{max,exp}$ [%]	Ref.
Chick pea protein	Alcalase	50	8	7→20 (mAU/g)	47	17→13	*4*
Bovine hemoglobin	Alcalase	35→55	7	20 (mAU/g)	2.5	13→23	*6*
		55	7	15→38 (mAU/g)	2.5	30→18	
Bovine serum albumin	Bovine trypsin	40	8	1:5→1:102	2.44	18→8	*7*
		40→55	8	1:102	0.244	18→21	
Whey protein isolate	Bovine trypsin	25	7→8.5	1:106	100	9.5→7	*12*
		25→50	8.5	1:106	100	7.1→3.9	
Whey protein isolate	Alcalase	40	8	1:100	10→300	23→13	*40*
	BLP	40	8	1:100	1→100	1→36	*37*
	BLP	40	7→9	1:100	50	4→9	*38*
Wheat gluten	Flavourzyme	50	5-8*	1:30	200→600	18→40	*39*

* The pH of the wheat gluten hydrolysis by Flavourzyme was not specified but between pH 5-8.

As discussed above, the enzyme secondary specificity is one mechanism that may explain the extent of protein hydrolysis. However, the enzyme secondary specificity should be independent of the system conditions. The results shown in **Table 1.2** indicated that the enzyme secondary specificity cannot be the only mechanism that determines the extent of protein hydrolysis. It is a strong indication that another mechanism was involved. The other possible explanation is that the protease lost its activity during the hydrolysis, which could be caused by the formation of inhibitory peptides of the protease during hydrolysis.

Actually, peptides formed during protein hydrolysis have been found to have inhibitory activities for many proteases (**Table 1.3**). It is well known that two tri-peptides, i.e. VPP and IPP, can inhibit the angiotensin-I converting enzyme (ACE) [41], a protease which is crucial in the regulation of blood pressure. Inhibitory peptides for ACE were identified in the hydrolysates of proteins from many sources, e.g. milk [42, 43], soy [44] and plant [45] proteins (**Table 1.3**). Besides ACE, inhibitory peptides for dipeptidyl peptidase IV (DPP-IV), a protease which plays a role on the regulation of blood glucose level, has also been found in many protein hydrolysates (**Table 1.3**). For example, the hydrolysates of milk protein isolate by Neutrase [46] or trypsin [47] were identified as the sources of inhibitory peptides for DPP-IV. The inhibitory activity of these hydrolysates was found to vary with the system conditions, which was most likely due to the different absolute concentrations of specific (inhibitory) peptides. Changing the pH, temperature and enzyme to substrate ratio (E:S) of the hydrolysis of whey protein concentrate was found to affect the inhibitory activity of the hydrolysates for ACE [48]. In addition, by altering the time, temperature and E:S of the hydrolysis, the IC_{50} (the concentration needed to achieve 50 % inhibition of the protease) of the milk protein isolate hydrolysates varied from 0.55 to 2.14 mg/mL for 3 nM DPP-IV [46].

Table 1.3: Overview of studies as an illustration of protein hydrolysates with inhibitory activity for proteases.

Inhibited protease	Protein source	Hydrolysis protease	DH [%]	IC_{50}*	Ref.
ACE	Buffalo milk proteins	Papain	23	30.5 µg/mL for 13.6 nM ACE	42
		Pepsin	33	38.9 µg/mL for 13.6 nM ACE	
		Trypsin	26	54.4 µg/mL for 13.6 nM ACE	
	Soy protein	Pepsin and pancreatin	40	0.28 mg/mL for 5.7 nM ACE	44
	Rice protein	Chymotrypsin & thermolysin	n.d.#	27 µg/mL for 68 nM ACE	45
	Wheat protein			40 µg/mL for 68 nM ACE	
	Pea protein			33 µg/mL for 68 nM ACE	
DPP-IV	Milk protein isolate	Neutrase	3.4-11.8	0.55-2.14 mg/mL for 3 nM DPP-IV	46
		Trypsin	7.0-13.5	0.62-1.59 mg/mL for 3 nM DPP-IV	47
	β-lactoglobulin	Porcine elastase	9, 13	1-1.5 mg/mL for 3 nM DPP-IV	49

#n.d.: not determined
*IC_{50} refers to the concentration of hydrolysates needed to achieve 50 % inhibition of the protease
ACE: angiotensin-I converting enzyme; DPP-IV: dipeptidyl peptidase IV

It is surprising that despite the widespread information on inhibitory activity of peptides, little attention is given to the possible consequence this has for the hydrolysis process. If peptides

formed by the enzyme during the hydrolysis inhibit that enzyme, the hydrolysis would stop. This would indicate that if the hydrolysate composition of a hydrolysis is changed, the amount of inhibitory peptides might be different, resulting in a different $DH_{max,exp}$ value. Differences in the hydrolysate composition can be obtained by modifying system conditions, e.g. the substrate concentration [40, 50], the enzyme to substrate ratio (E:S) [49, 51], the pH [12], etc. This could be the possible reason why changes in system condition result in different extents of hydrolysis.

In addition to the examples that specifically studied the inhibition of proteases by peptides, some studies focused on the kinetic modelling of protein hydrolysis. In those kinetic models of protein hydrolysis, the inactivation constant of enzyme and the substrate/product inhibition constant have been often used in kinetic equations [4-7]. For example, the experimental hydrolysis curve of bovine serum albumin by trypsin was described by an exponential equation with the enzyme inactivation constant of 9.358 min^{-1} and the substrate inhibition constant of 7.961 g/L [7]. This is another indication that the enzyme might be inhibited during the protein hydrolysis.

Modifications of the protein primary structure

During industrial processing of food products, modifications on the protein primary structure might occur. The modifications can be caused by chemical reactions, e.g. through the Maillard reaction [52], and by enzymatic reactions, e.g. using transglutaminase [53]. The Maillard reaction is sometimes applied on purpose to improve the (techno-)functional properties of proteins, e.g. emulsion stability [54] and foam stability [55]. In other cases, the reaction is a non-desired side effect of production and storage of food products [56]. In the early stage of the Maillard reaction, the reducing end of a carbohydrate reacts with a free amino group. This first step is referred to as the glycation. Because of the attachment of the carbohydrates, the protein primary structure is changed. If the glycation occurs in a protein/peptide, the amino group(s) can be the side chain of a lysine residue and at the N-terminal AA of the protein. The glycation of the side chain of arginine was also reported on the arginines R40, R124 and R148 in β-lactoglobulin [57]. In later stages of the Maillard reaction, secondary reactions start to take place, which lead to changes in the protein secondary [54, 58], tertiary [58, 59] and quaternary [60] structure. Protein cross-linking was reported to occur during this stage of the reaction, e.g. through lysine-arginine cross-links [61]. The heating process itself might also result in protein aggregation through lysinoalanyl cross-links [62] and/or non-covalent aggregation [63], depending on the method of heating.

The protein structural changes through the Maillard reaction can influence the susceptibility of the proteins to enzymatic hydrolysis. The overview of studies that investigated the influences of the Maillard reaction on protein hydrolysis were summarised (**Table 1.4**). Despite the amount of information that was obtained, the influences of the Maillard reaction induced glycation and aggregation on hydrolysis have not been clearly identified. This is because many studies did not have quantitative measurements of glycation (or aggregation) and/or hydrolysis (**Table 1.4**).

Table 1.4: Overview of studies that investigated the effects of the Maillard reaction on protein hydrolysis.

Sources of -NH$_2$ group	Saccharides	Glycation condition	Detection for glycation	Quantification of glycation	Detection for secondary reactions	Enzyme	Detection for digestibility	Quantification of hydrolysis	Ref.
β-Casein	Lactose	Wet heating, 120 °C	Fluorescamine method	0-40 % degree of lactosylation	SDS-PAGE Absorbance at 420 nm	Plasmin	HPLC-UV	Hydrolysis rate decreased from 100 % to 30 %	64
Soy protein isolate, casein ; milk enriched with whey	Corn syrup; sucrose; glucose; maltodextrin; lactose	Industrial processes (not shown)	OPA	33.2-81.1 % available lysine	n.d.	Trypsin; chymotrypsin ; intestinal peptidase	DH after 4 h hydrolysis	87.0-97.5 % digestibility	65
β-Lactoglobulin	Galacto-oligosaccharides	Dry heating, 40 ° C	Furosine; IEF-PAGE; MS	1.82-2.77 mg furosine per 100 mg of protein	SDS-PAGE	Pepsin; trypsin; chymotrypsin	SDS-PAGE; HPLC-UV	n.d.	66
Whey protein	Dextran of 10 KDa; dextran of 150 KDa	Wet heating, 60 ° C	n.d.	n.d.	SDS-PAGE	Pepsin; pancreatin	SDS-PAGE Fluorescence laser densitometry	Hydrolysis rate decreased from 0.02-0.01 min^{-1}	67
β-Lactoglobulin	Galactose; Tagatose; dextran of 10 KDa; dextran of 20 KDa	Dry heating, 40, 50 and 60 ° C	MS	14 galactose, 3 tagatose, 14 dextran (10 kDa), 15 dextran (20 kDa), per protein	SEC SDS-PAGE	Pepsin; trypsin; chymotrypsin	HPLC-UV	n.d.	60
β-Lactoglobulin; β-casein	Glucose	Wet heating, 90 ° C	n.d.	n.d.	Fluorescence; SDS-PAGE; SEC; TEM	Pepsin; trypsin; chymotrypsin	SDS-PAGE; TEM	n.d.	68
β-Lactoglobulin	Lactose	Wet heating, 60 ° C	MS	0.12-0.74 lactose per protein	n.d.	Trypsin	Peptide analysis	n.d.	69
Human serum albumin	Glucose	Wet heating, 37 ° C	Peptide analysis	n.d.	n.d.	Trypsin	Peptide analysis	n.d.	70
β-Lactoglobulin	Galacto-oligosaccharides	Dry heating, 40 ° C	MS	n.d.	n.d.	Pepsin; trypsin; chymotrypsin	Peptide analysis	n.d.	71

*SDS-PAGE: sodium dodecyl sulphate polyacrylamide gel electrophoresis; TEM: transmission electron microscopy; n.d. not determined.

One article reported that the proportion of total glycated amino groups in β-casein was up to ~40 %, while ~60 % of the proteins were covalently aggregated [64]. The initial rate of plasmin hydrolysis was observed to be linearly decreased by ~60 % with the maximal glycated sample [64]. However, a linear correlation could also be obtained between the amount of aggregation and the initial rate of hydrolysis. It is difficult to distinguish if the influence on hydrolysis is due to the changes in the enzyme secondary specificity or the accessibility of substrates.

Another challenge in this field is that various methods are used to quantify the extent of the Maillard reaction (**Table 1.4**). In literature, the extent of glycation is typically quantified using the following methods (**Figure 1.3**). The quantification of glycation can be directly performed by the *o*-phtaldialdehyde (OPA) method [72], which measures the amount of free amino groups, and mass spectrometry (MS) [73], which measures the addition of masses of carbohydrates on the masses of proteins. It can also be quantified by measuring the markers of the Maillard reaction after acid hydrolysis (6 M/8 M HCl at 110 °C) [74], e.g. furosine, of the samples. In this case, a conversion factor should be applied to convert the amount of markers formed to the amount of glycation present in the samples [75]. Using MS, a clear overview of the compounds that are attached to the protein molecules can be obtained. For example, the MS spectrum of the glycated α-lactalbumin with D-glucose which was heated at 60 °C with 65 % relative humidity for 8 h showed that there were 6-13 glucose attached [76]. However, when the proteins are denatured or aggregated, the MS does not give good signals. The advantage of the OPA method is that it can be used in these systems since the samples can be treated with sodium dodecyl sulphate and/or dithiothreitol prior to the analysis. However, since OPA method measures the amount of free amino groups in the system, if there is no reference sample, i.e. the same protein but does not have the glycation, to compare with, the amount of glycated lysine cannot be calculated. If the MS of the samples does not give good signals and there is no reference sample to compare with, acid hydrolysis is typically applied to the samples. The compounds formed after acid hydrolysis, e.g. furosine [77] and carboxymethyl-lysine [78], are taken as the markers for the Maillard reaction, which can be determined using the ultra-performance liquid chromatography coupled with MS. It should be noted that some articles also used the formation of a brown colour as the indication of the extent of glycation [79-81]. However, the brown colour formation is not an indication of glycation but an indication of the later stages of the Maillard reaction, e.g. the formation of Melanoidins.

Changes in the secondary and tertiary structure of proteins can be determined using circular dichroism (CD). The changes in protein quaternary structure can be analysed using size exclusion chromatography (SEC). The overview of the changes in protein structure due to the Maillard reaction or heating, and the commonly used detection techniques for these structural changes and the reaction markers, were summarised based on information provided in literature [61, 62, 75, 79, 82-86] (**Figure 1.3**). Because these methods actually determine different compounds formed during the Maillard reaction, it is difficult to combine existing data to make conclusions.
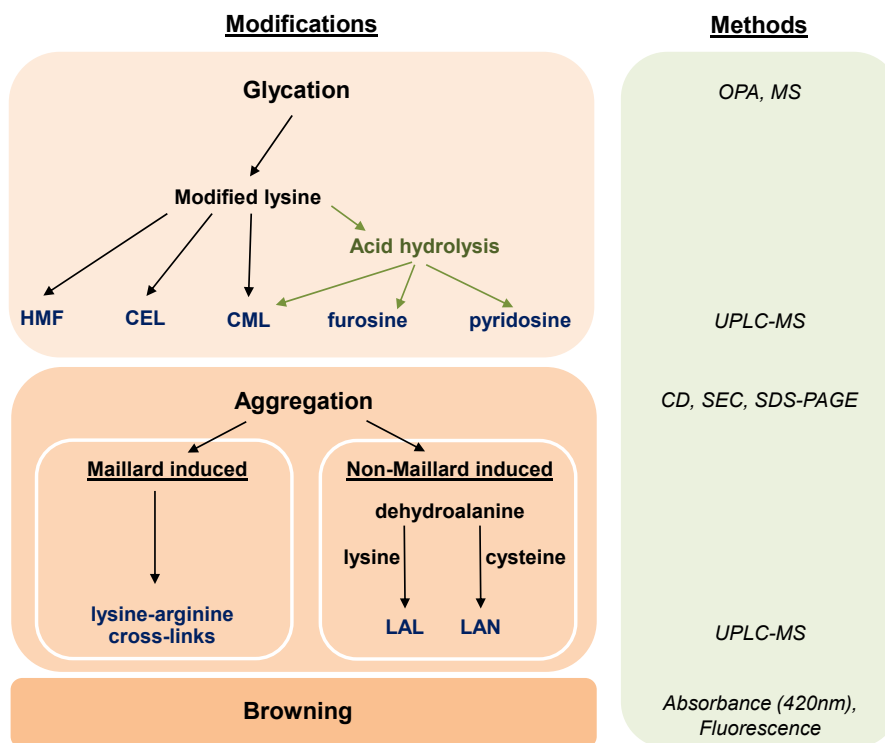
Modifications   **Methods**

Figure 1.3: Overview of the protein structural changes due to the Maillard reaction or heating, and the typically used techniques for the detection of these structural changes and the reaction markers, adapted from information provided in literature [61, 62, 75, 79, 82-86]. HMF: hydroxymethyl-furfural; CEL: carboxyethyl-lysine; CML: carboxymethyl-lysine; LAL: lysinoalanine; LAN: lanthionine. OPA: *o*-phtaldialdehyde method; MS: mass spectrometry; UPLC: ultra-high performance liquid chromatography; CD: circular dichroism; SEC: size exclusion chromatography; SDS-PAGE: sodium dodecyl sulphate polyacrylamide gel electrophoresis.

Besides different methods to quantify glycation, the proteases used in literature to perform the hydrolysis were also different (**Table 1.4**), which might lead to a different outcome of the effect of glycation on protein hydrolysis obtained. For proteases that are specific for lysine residues (the glycation site), e.g. trypsin and plasmin, the modifications through glycation correspond to the modifications of the CSs (P1 position) in a protein. It was shown that replacing the charged side chain of lysine by the neutral side chain of heptyline (similar structure as lysine but without the ε-amino group) resulted in a decrease of hydrolysis rate constants by a factor of 10^2-10^6 [87]. The chemical reactions that alter both the charge and the size of the AA side chain on the P1 position, e.g. glycation with maltodextrins, might have an even larger influence. For enzymes that are not specificity for lysine residues, the modification through glycation may be on another part of the binding site. In these situations, it might depend on the sensitivity of the enzyme towards that binding site position.

Aim and outline of the thesis

The aim of this thesis is to develop a model to predict the extent of protein hydrolysis, based on the intrinsic properties of the protein and protease (primary and secondary specificity), and the effects of modifications on protein primary structure and changes in system conditions. To achieve this, it is necessary to identify the mechanisms that determine the $DH_{max,exp}$. In **Chapter 2**, the effects of modifications on protein primary structure (through glycation) on the $DH_{max,exp}$ of hydrolysis by proteases with different primary and secondary specificity are described. In **Chapter 3**, the bovine trypsin selectivity, determined by full quantitative analysis of peptides formed during hydrolysis, is linked with the binding site sequence of the substrate. The link is used to predict the $DH_{max,exp}$ based on enzyme selectivity. In **Chapter 4**, the enzyme selectivity of bovine, porcine and human trypsins are compared, and are used to explain their differences in the $DH_{max,exp}$. In **Chapter 5**, the effects of substrate concentration on the $DH_{max,exp}$ of hydrolysis by these digestive proteases are studied, and the hypothesis about the formation of inhibitory peptides is tested. In **Chapter 6**, the quantitative peptide analysis is also used to follow the glycation kinetics of individual peptides in protein hydrolysates. In **Chapter 7**, the industrial and societal relevance of this thesis, i.e. the risk analysis and technology assessment, is discussed. In the general discussion (**Chapter 8**), a prediction model that defines the expectation of the $DH_{max,exp}$ of hydrolysis of a given protein by a specific protease is derived, based on the outcome of the previous chapters. The effects of modifications on protein primary structure and changes in substrate concentration are also included in the equation. Using this model, the effects of glycation and substrate concentration on the $DH_{max,exp}$ are combined. The joint effect on hydrolysis by several proteases is shown in a three-dimensional graph (**Figure 1.4**). Furthermore, the boundary of this prediction model is tested, and the obtained knowledge in this thesis was extrapolated to systems that are more complex.
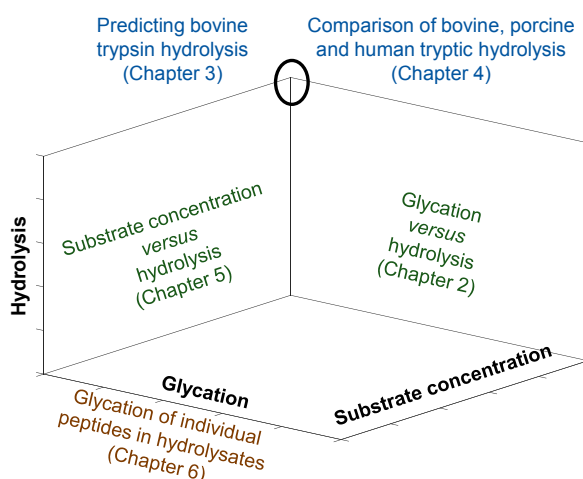

Figure 1.4: Outline of the thesis.

References

[1] N.D. Rawlings, M. Waller, A.J. Barrett, A. Bateman, MEROPS: The database of proteolytic enzymes, their substrates and inhibitors, Nucleic Acids Research, 42 (2014) D503-D509.

[2] S.M. Rutherfurd, Methodology for determining degree of hydrolysis of proteins in hydrolysates: A review, Journal of AOAC International, 93 (2010) 1515-1522.

[3] K.A. Johnson, R.S. Goody, The original Michaelis constant: Translation of the 1913 Michaelis–Menten paper, Biochemistry, 50 (2011) 8264-8269.

[4] M.C. Marquez Moreno, V. Fernandez Cuadrado, Enzymic hydrolysis of vegetable proteins: Mechanism and kinetics, Process Biochemistry, 28 (1993) 481-490.

[5] P. Gonzàlez-Tello, F. Camacho, E. Jurado, M.P. Páez, E.M. Guadix, Enzymatic hydrolysis of whey proteins: I. Kinetic models, Biotechnology and Bioengineering, 44 (1994) 523-528.

[6] M.C. Márquez, M.A. Vázquez, Modeling of enzymatic protein hydrolysis, Process Biochemistry, 35 (1999) 111-117.

[7] W. Qi, Z. He, Enzymatic hydrolysis of protein: Mechanism and kinetic model, Frontiers of Chemistry in China, 1 (2006) 308-314.

[8] A. Fernández, F. Riera, β-Lactoglobulin tryptic digestion: A model approach for peptide release, Biochemical Engineering Journal, 70 (2013) 88-96.

[9] C. Boudier, M.L. Jung, N. Stambolieva, J.G. Bieth, Importance of secondary enzyme-substrate interactions in human cathepsin G and chymotrypsin II catalysis, Archives of Biochemistry and Biophysics, 210 (1981) 790-793.

[10] M. Salami, R. Yousefi, M.R. Ehsani, S.H. Razavi, J.-M. Chobert, T. Haertlé, A.A. Saboury, M.S. Atri, A. Niasari-Naslaji, F. Ahmad, A.A. Moosavi-Movahedi, Enzymatic digestion and antioxidant activity of the native and molten globule states of camel α-lactalbumin: Possible significance for use in infant formula, International Dairy Journal, 19 (2009) 518-523.

[11] D. Panyam, A. Kilara, Emulsifying peptides from the tryptic hydrolysis of casein, Journal of Food Science, 69 (2004) FCT154-FCT163.

[12] S.C. Cheison, E. Leeb, J. Toro-Sierra, U. Kulozik, Influence of hydrolysis temperature and pH on the selective hydrolysis of whey proteins by trypsin and potential recovery of native alpha-lactalbumin, International Dairy Journal, 21 (2011) 166-171.

[13] P. Maresca, G. Ferrari, Modelling of the kinetics of bovine serum albumin enzymatic hydrolysis assisted by high hydrostatic pressure, Food and Bioproducts Processing, 105 (2017) 1-11.

[14] K. Lisak, J. Toro-Sierra, U. Kulozik, R. Božanić, S.C. Cheison, Chymotrypsin selectively digests β-lactoglobulin in whey protein isolate away from enzyme optimal conditions: Potential for native α-lactalbumin purification, Journal of Dairy Research, 80 (2013) 14-20.

[15] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Introducing enzyme selectivity: A quantitative parameter to describe enzymatic protein hydrolysis, Analytical and Bioanalytical Chemistry, 406 (2014) 5827-5841.

[16] N. Creusot, H. Gruppen, Hydrolysis of whey protein isolate with *Bacillus licheniformis* protease: Fractionation and identification of aggregating peptides, Journal of Agricultural and Food Chemistry, 55 (2007) 9241-9250.

[17] I. Schechter, A. Berger, On the size of the active site in proteases. I. Papain, Biochemical and Biophysical Research Communications, 27 (1967) 157-162.

[18] N.D. Rawlings, Protease families, evolution and mechanism of action, Proteases: Structure and function, Springer Vienna, Vienna, 2013, pp. 1-36.

[19] D. Liebschner, M. Dauter, A. Brzuszkiewicz, Z. Dauter, On the reproducibility of protein crystal structures: Five atomic resolution structures of trypsin, Acta Crystallographica Section D: Biological Crystallography, 69 (2013) 1447-1462.

[20] E. Vandermarliere, M. Mueller, L. Martens, Getting intimate with trypsin, the leading protease in proteomics, Mass Spectrometry Reviews, 32 (2013) 453-465.

[21] S.J. Hur, B.O. Lim, E.A. Decker, D.J. McClements, *In vitro* human digestion models for food applications, Food Chemistry, 125 (2011) 1-12.

[22] B. Keil, Specificity of proteolysis, Springer-Verlag Berlin Heidelberg, New York, USA, 1992, pp. 66-69.

[23] C.Y. Yen, S. Russell, A.M. Mendoza, K. Meyer-Arendt, S. Sun, K.J. Cios, N.G. Ahn, K.A. Resing, Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: Protease cleavage and SCX elution rules from data mining of MS/MS spectra, Analytical Chemistry, 78 (2006) 1071-1084.

[24] J.A. Siepen, E.-J. Keevil, D. Knight, S.J. Hubbard, Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics, Journal of Proteome Research, 6 (2007) 399-408.

[25] P.D. Gershon, Cleaved and missed sites for trypsin, Lys-C, and Lys-N can be predicted with high confidence on the basis of sequence context, Journal of Proteome Research, 13 (2014) 702-709.

[26] Y. Pan, K. Cheng, J. Mao, F. Liu, J. Liu, M. Ye, H. Zou, Quantitative proteomics reveals the kinetics of trypsin-catalyzed protein digestion, Analytical and Bioanalytical Chemistry, 406 (2014) 6247-6256.

[27] C. Lawless, S.J. Hubbard, Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics, OMICS A Journal of Integrative Biology, 16 (2012) 449-456.

[28] J.V. Olsen, S.E. Ong, M. Mann, Trypsin cleaves exclusively C-terminal to arginine and lysine residues, Molecular and Cellular Proteomics, 3 (2004) 608-614.

[29] J. Rodriguez, N. Gupta, R.D. Smith, P.A. Pevzner, Does trypsin cut before proline?, Journal of Proteome Research, 7 (2008) 300-305.

[30] B.F. Le Bonniec, T. Myles, T. Johnson, C.G. Knight, C. Tapparelli, S.R. Stone, Characterization of the P2′ and P3′ specificities of thrombin using fluorescence-quenched substrates and mapping of the subsites by mutagenesis, Biochemistry, 35 (1996) 7114-7122.

[31] S.R. Marana, A.R. Lopes, L. Juliano, M.A. Juliano, C. Ferreira, W.R. Terra, Subsites of trypsin active site favor catalysis or substrate binding, Biochemical and Biophysical Research Communications, 290 (2002) 494-497.

[32] A.R. Lopes, M.A. Juliano, S.R. Marana, L. Juliano, W.R. Terra, Substrate specificity of insect trypsins and the role of their subsites in catalysis, Insect Biochemistry and Molecular Biology, 36 (2006) 130-140.

[33] J. Tauzin, L. Miclo, S. Roth, D. Mollé, J.L. Gaillard, Tryptic hydrolysis of bovine α_{s2}-casein: Identification and release kinetics of peptides, International Dairy Journal, 13 (2003) 15-27.

[34] S.J. Walmsley, P.A. Rudnick, Y. Liang, Q. Dong, S.E. Stein, A.I. Nesvizhskii, Comprehensive analysis of protein digestion using six trypsins reveals the origin of trypsin as a significant source of variability in proteomics, Journal of Proteome Research, 12 (2013) 5666-5680.

[35] M.M. Vorob'ev, M. Dalgalarrondo, J.M. Chobert, T. Haertlé, Kinetics of β-casein hydrolysis by wild-type and engineered trypsin, Biopolymers, 54 (2000) 355-364.

[36] C. Elsner, S. Grahn, S. Bauer, D. Ullmann, T. Kurth, H.-D. Jakubke, Effects of chemical modification of lysine residues in trypsin, Journal of Molecular Catalysis B: Enzymatic, 8 (2000) 193-200.

[37] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Determination of the influence of substrate concentration on enzyme selectivity using whey protein isolate and *Bacillus licheniformis* protease, Journal of Agricultural and Food Chemistry, 62 (2014) 10230-10239.

[38] C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Determination of the influence of the pH of hydrolysis on enzyme selectivity of *Bacillus licheniformis* protease towards whey protein isolate, International Dairy Journal, 44 (2015) 44-53.

[39] C.I. Butré, P.A. Wierenga, H. Gruppen, Effects of ionic strength on the enzymatic hydrolysis of diluted and concentrated whey protein isolate, Journal of Agricultural and Food Chemistry, 60 (2012) 5644-5651.

[40] N.A. Hardt, A.E.M. Janssen, R.M. Boom, A.J. van der Goot, Factors impeding enzymatic wheat gluten hydrolysis at high solid concentrations, Biotechnology and Bioengineering, 111 (2014) 1304-1312.

[41] Y. Nakamura, N. Yamamoto, K. Sakai, A. Okubo, S. Yamazaki, T. Takano, Purification and characterization of angiotensin I-converting enzyme inhibitors from sour milk, Journal of Dairy Science, 78 (1995) 777-783.

[42] M. Abdel-Hamid, J. Otte, C. De Gobba, A. Osman, E. Hamad, Angiotensin I-converting enzyme inhibitory activity and antioxidant capacity of bioactive peptides derived from enzymatic hydrolysis of buffalo milk proteins, International Dairy Journal, 66 (2017) 91-98.

[43] R.J. FitzGerald, H. Meisel, Milk protein-derived peptide inhibitors of angiotensin-I converting enzyme, British Journal of Nutrition, 84 (2000) 33-37.

[44] W.M.Y. Lo, E.C.Y. Li-Chan, Angiotensin I converting enzyme inhibitory peptides from *in vitro* pepsin-pancreatin digestion of soy protein, Journal of Agricultural and Food Chemistry, 53 (2005) 3369-3376.

[45] S. Rudolph, D. Lunow, S. Kaiser, T. Henle, Identification and quantification of ACE-inhibiting peptides in enzymatic hydrolysates of plant proteins, Food Chemistry, 224 (2017) 19-25.

[46] A.B. Nongonierma, M. Lalmahomed, S. Paolella, R.J. FitzGerald, Milk protein isolate (MPI) as a source of dipeptidyl peptidase IV (DPP-IV) inhibitory peptides, Food Chemistry, 231 (2017) 202-211.

[47] A.B. Nongonierma, C. Mazzocchi, S. Paolella, R.J. FitzGerald, Release of dipeptidyl peptidase IV (DPP-IV) inhibitory peptides from milk protein isolate (MPI) during enzymatic hydrolysis, Food Research International, 94 (2017) 79-89.

[48] Y. Guo, D. Pan, M. Tanokura, Optimisation of hydrolysis conditions for the production of the angiotensin-I converting enzyme (ACE) inhibitory peptides from whey protein using response surface methodology, Food Chemistry, 114 (2009) 328-333.

[49] S. Le Maux, A.B. Nongonierma, R.J. FitzGerald, Peptide composition and dipeptidyl peptidase IV inhibitory properties of β-lactoglobulin hydrolysates having similar extents of hydrolysis while generated using different enzyme-to-substrate ratios, Food Research International, 99 (2017) 84-90.

[50] N.A. Hardt, A.J. van der Goot, R.M. Boom, Influence of high solid concentrations on enzymatic wheat gluten hydrolysis and resulting functional properties, Journal of Cereal Science, 57 (2013) 531-536.

[51] P.G. Tello, F. Camacho, E. Jurado, M.P. Páez, E.M. Guadix, Enzymatic hydrolysis of whey proteins. II. Molecular-weight range, Biotechnology and Bioengineering, 44 (1994) 529-532.

[52] J.A. O'Mahony, K.P. Drapala, E.M. Mulcahy, D.M. Mulvihill, Controlled glycation of milk proteins and peptides: Functional properties, International Dairy Journal, 67 (2017) 16-34.

[53] M. Motoki, K. Seguro, Transglutaminase and its use for food processing, Trends in Food Science & Technology, 9 (1998) 204-210.

[54] M. Darewicz, J. Dziuba, The effect of glycosylation on emulsifying and structural properties of bovine β-casein, Food/Nahrung, 45 (2001) 15-20.

[55] P.A. Wierenga, L. Van Norél, E.S. Basheva, Reconsidering the importance of interfacial properties in foam stability, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 344 (2009) 72-78.

[56] C.G.A. Davies, F.M. Netto, N. Glassenap, C.M. Gallaher, T.P. Labuza, D.D. Gallaher, Indication of the Maillard reaction during storage of protein isolates, Journal of Agricultural and Food Chemistry, 46 (1998) 2485-2489.

[57] F. Fenaille, F. Morgan, V. Parisod, J.-C. Tabet, P.A. Guy, Solid-state glycation of β-lactoglobulin by lactose and galactose: Localization of the modified amino acids using mass spectrometric techniques, Journal of Mass Spectrometry, 39 (2004) 16-28.

[58] K. Broersen, A.G.J. Voragen, R.J. Hamer, H.H.J. De Jongh, Glycoforms of β-lactoglobulin with improved thermostability and preserved structural packing, Biotechnology and Bioengineering, 86 (2004) 78-87.

[59] F. Chevalier, J.M. Chobert, M. Dalgalarrondo, Y. Choiset, T. Haertlé, Maillard glycation of β-lactoglobulin induces conformation changes, Nahrung, 46 (2002) 58-63.

[60] M. Corzo-Martínez, A.C. Soria, J. Belloque, M. Villamiel, F.J. Moreno, Effect of glycation on the gastrointestinal digestibility and immunoreactivity of bovine β-lactoglobulin, International Dairy Journal, 20 (2010) 742-752.

[61] K.M. Biemel, O. Reihl, J. Conrad, M.O. Lederer, Formation pathways for lysine-arginine cross-links derived from hexoses and pentoses by Maillard processes: Unraveling the structure of a pentosidine precursor, Journal of Biological Chemistry, 276 (2001) 23405-23412.

[62] R. Fritsch, H. Hoffmann, H. Klostermeyer, Formation of lysinoalanine during heat treatment of milk, Zeitschrift für Lebensmittel-Untersuchung und Forschung, 176 (1983) 341-345.

[63] I.B. O'Loughlin, B.A. Murray, P.M. Kelly, R.J. Fitzgerald, A. Brodkorb, Enzymatic hydrolysis of heat-induced aggregates of whey protein isolate, Journal of Agricultural and Food Chemistry, 60 (2012) 4895-4904.

[64] H. Bhatt, A. Cucheval, C. Coker, H. Patel, A. Carr, R. Bennett, Effect of lactosylation on plasmin-induced hydrolysis of β-casein, International Dairy Journal, 38 (2014) 213-218.

[65] A.S. Pereyra Gonzáles, G.B. Naranjo, L.S. Malec, M.S. Vigo, Available lysine, protein digestibility and lactulose in commercial infant formulas, International Dairy Journal, 13 (2003) 95-99.

[66] M.L. Sanz, M. Corzo-Martínez, R.A. Rastall, A. Olano, F.J. Moreno, Characterization and *in vitro* digestibility of bovine β-lactoglobulin glycated with galactooligosaccharides, Journal of Agricultural and Food Chemistry, 55 (2007) 7916-7925.

[67] F.H. Böttger, M.R. Etzel, J.A. Lucey, *In vitro* infant digestion of whey protein–dextran glycates, Food Digestion, 4 (2013) 76-84.

[68] M.S. Pinto, J. Léonil, G. Henry, C. Cauty, A.F. Carvalho, S. Bouhallab, Heating and glycation of β-lactoglobulin and β-casein: Aggregation and *in vitro* digestion, Food Research International, 55 (2014) 70-76.

[69] F. Morgan, J. Léonil, D. Mollé, S. Bouhallab, Nonenzymatic lactosylation of bovine β-lactoglobulin under mild heat treatment leads to structural heterogeneity of the glycoforms, Biochemical and Biophysical Research Communications, 236 (1997) 413-417.

[70] A. Lapolla, D. Fedele, R. Reitano, N.C. Aricò, R. Seraglia, P. Traldi, E. Marotta, R. Tonani, Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides, Journal of the American Society for Mass Spectrometry, 15 (2004) 496-509.

[71] F.J. Moreno, J.E. Quintanilla-López, R. Lebrón-Aguilar, A. Olano, M.L. Sanz, Mass spectrometric characterization of glycated β-lactoglobulin peptides derived from galacto-oligosaccharides surviving the *in vitro* gastrointestinal digestion, Journal of the American Society for Mass Spectrometry, 19 (2008) 927-937.

[72] M.S. Vigo, L.S. Malec, R.G. Gomez, R.A. Llosa, Spectrophotometric assay using *o*-phthaldialdehyde for determination of reactive lysine in dairy products, Food Chemistry, 44 (1992) 363-365.

[73] C.M. Oliver, Insight into the glycation of milk proteins: An ESI- and MALDI-MS perspective (review), Critical Reviews in Food Science and Nutrition, 51 (2011) 410-431.

[74] A.D. Troise, A. Fiore, M. Wiltafsky, V. Fogliano, Quantification of Nε-(2-Furoylmethyl)-l-lysine (furosine), Nε-(Carboxymethyl)-l-lysine (CML), Nε-(Carboxyethyl)-l-lysine (CEL) and total lysine through stable isotope dilution assay and tandem mass spectrometry, Food Chemistry, 188 (2015) 357-364.

17

[75] R. Krause, K. Knoll, T. Henle, Studies on the formation of furosine and pyridosine during acid hydrolysis of different Amadori products of lysine, European Food Research and Technology, 216 (2003) 277-283.

[76] R. ter Haar, H.A. Schols, H. Gruppen, Effect of saccharide structure and size on the degree of substitution and product dispersity of α-lactalbumin glycated via the Maillard reaction, Journal of Agricultural and Food Chemistry, 59 (2011) 9378-9385.

[77] H.F. Erbersdobler, V. Somoza, Forty years of furosine - Forty years of using Maillard reaction products as indicators of the nutritional quality of foods, Molecular Nutrition and Food Research, 51 (2007) 423-430.

[78] C. Delgado-Andrade, Carboxymethyl-lysine: Thirty years of investigation in the field of AGE formation, Food and Function, 7 (2016) 46-57.

[79] M. Friedman, Food browning and its prevention: An overview, Journal of Agricultural and Food Chemistry, 44 (1996) 631-653.

[80] E.-J. Kwak, S.-I. Lim, The effect of sugar, amino acid, metal ion, and NaCl on model Maillard reaction under pH control, Amino Acids, 27 (2004) 85-90.

[81] G. Münch, D. Schicktanz, A. Behme, M. Gerlach, P. Riederer, D. Palm, R. Schinzel, Amino acid specificity of glycation and protein-AGE crosslinking reactivities determined with a dipeptide SPOT library, Nature Biotechnology, 17 (1999) 1006-1010.

[82] J.E. Hodge, Dehydrated foods, chemistry of browning reactions in model systems, Journal of Agricultural and Food Chemistry, 1 (1953) 928-943.

[83] M.O. Lederer, H.P. Bühler, Cross-linking of proteins by Maillard processes—characterization and detection of a lysine-arginine cross-link derived from D-glucose, Bioorganic & Medicinal Chemistry, 7 (1999) 1081-1088.

[84] M.W. Poulsen, R.V. Hedegaard, J.M. Andersen, B. de Courten, S. Bügel, J. Nielsen, L.H. Skibsted, L.O. Dragsted, Advanced glycation endproducts in food and their effects on health, Food and Chemical Toxicology, 60 (2013) 10-37.

[85] M. Hellwig, T. Henle, Baking, ageing, diabetes: A short history of the maillard reaction, Angewandte Chemie - International Edition, 53 (2014) 10316-10329.

[86] S.I.F.S. Martins, W.M.F. Jongen, M.A.J.S. van Boekel, A review of Maillard reaction in food and implications to kinetic modelling, Trends in Food Science & Technology, 11 (2000) 364-373.

[87] B.M. Sanborn, G.E. Hein, The interaction of trypsin with neutral substrates and modifiers, Biochemistry, 7 (1968) 3616-3624.

CHAPTER 2

Effect of Maillard induced glycation on protein hydrolysis by lysine/arginine and non-lysine/arginine specific proteases

Enzymatic protein hydrolysis is sensitive to modifications of protein structure, e.g. Maillard reaction. In early stages of the reaction, glycation takes place, modifying the protein primary structure. In later stages protein aggregation occurs. The specific effect of glycation on protein hydrolysis was studied using α-lactalbumin glycated with D-glucose at 50 °C (0-10 h). This resulted in proteins with different average degrees of glycation (DG_{av}=0-63 %) without changes in secondary, tertiary and quaternary structure. These glycated proteins were hydrolysed by lysine/arginine specific proteases (bovine and porcine trypsins) or by non-lysine/arginine specific proteases (*Bacillus licheniformis* protease (BLP), α-chymotrypsin and subtilisin A). For bovine and porcine trypsins, the maximal degree of hydrolysis decreased linearly by 65 % from untreated to maximal glycated protein (DG_{av}=63 %). This means trypsin cannot hydrolyse glycated cleavage sites. BLP and subtilisin A hydrolyses are independent of glycation, while α-chymotrypsin cannot hydrolyse cleavage sites with glycated binding sites. This means for non-lysine/arginine specific proteases, the effect of glycation depends on the enzyme sensitivity towards modifications on binding sites. Since not all cleavage sites are used efficiently by the enzymes, the extent of the effects depends on the enzyme selectivity towards cleavage sites (for trypsin) or cleavage sites near glycation sites (for α-chymotrypsin). Combining the results of all proteases, an equation was derived describing the effect of modification of protein primary structure on the extent of hydrolysis based on the enzyme specificity, selectivity and binding site sensitivity.

Introduction

Enzymatic protein hydrolysis is sensitive to modifications of the protein structure, which can be induced by the industrial processes, e.g. Maillard reaction. In food production, the Maillard reaction is sometimes applied to improve the techno-functional properties of proteins, e.g. emulsion stability [1] and foam stability [2]. In other cases, the reaction is a non-desired side effect of production and storage of food products. The Maillard reaction results in glycation of lysine residues, but the proteins can also undergo unfolding or aggregation. These factors may all affect protein hydrolysis during digestion, as well as during industrial production of protein hydrolysates. The effects may further depend on the enzymes used for the protein hydrolysis. This study focusses on a quantitative understanding of the effects of glycation on enzymatic protein hydrolysis.

The early stage of the Maillard reaction changes the protein primary structure. In the first step of the Maillard reaction, also called glycation, the reducing end of a carbohydrate reacts with a free amino group, which results in Amadori compounds. This amino group can be the side chain of a lysine residue, an arginine residue or at the N-terminal amino acid in a protein. For example, the glycation of α-lactalbumin with glucose at 60 °C with 65 % relative humidity for 8 h resulted in a mixture of variants with 6-13 glucoses on the α-lactalbumin molecule [3]. In perspective of protein hydrolysis, the glycation can change the protein primary structure. In later stages of the reaction, secondary reactions take place, which lead to protein aggregation [4], e.g. through lysine-arginine cross-links [5]. In addition, the heating itself may result in aggregation through lysinoalanyl cross-links [6] and/or non-covalent aggregation. These reactions can influence the secondary, tertiary and quaternary structure of the protein. Each of these structural changes can influence the susceptibility of the protein to enzymatic hydrolysis. The occurrences of the secondary reactions also confuse published results of the influence of glycation on protein hydrolysis. For example, one study showed that the pancreatic hydrolysis rate of whey protein was decreased by glycation with dextrans [7]. Also, the hydrolysis of two infant formulas made of soy protein isolates by a mixture of trypsin, chymotrypsin and intestinal peptidase showed that the sample with the highest degree of glycation resulted in the lowest degree of hydrolysis [8]. In contrast to these findings, other studies showed that glycation has been shown to increase the extent of protein hydrolysis. In the same study, when the hydrolysis of two infant formulas made of bovine milk powder were compared, the sample with the highest glycation had the highest degree of hydrolysis [8]. Another indication of increased hydrolysis with increased glycation was obtained from the comparison of the hydrolysis of ultra-high-temperature (UHT) milk compared to pasteurized milk by pepsin and pancreatin. The amount of intact proteins remaining after hydrolysis was lower in the UHT milk than in the pasteurized milk, despite higher levels of Maillard reaction indicators (carboxymethyllysine and lactulosylysine) in the UHT milk [9]. The inconsistent reports on the effects of Maillard reaction on subsequent enzymatic hydrolysis may be in part due to the fact that not only glycation, but also aggregation may have occurred in the

samples. It is not clear whether the dominant factor of the influence was glycation or aggregation because aggregation was also reported to affect hydrolysis [10]. One study reported that the proportion of total glycated amino groups in β-casein samples was up to ~40 %, while ~60 % of the proteins were covalently aggregated [11]. It was observed that the initial rate of plasmin hydrolysis linearly decreased by 60 % with the maximal glycated sample, which also had the highest amounts of aggregates [11]. Another reason for the variety in reported effects of glycation on hydrolysis could be that different enzymes were used.

For enzymes that are specific for lysine and arginine residues, e.g. trypsin and plasmin, it is commonly assumed that the hydrolysis is negatively influenced by glycation because the glycated cleavage site cannot be utilized by trypsin. As a consequence, the extent of trypsin hydrolysis is expected to decrease by glycation. Most studies that annotated peptides after tryptic hydrolysis of glycated proteins did not find peptides that were cleaved after glycated lysine/arginine residues [12, 13]. Only two previous studies reported cleavage peptide bonds after glycated lysines, based on annotation of the formed peptides. They reported that 1 out 19 [14] and 3 out of 16 [15] of the peptides formed during the hydrolysis were cleaved after glycated lysines, respectively. A study using model peptides showed that the replacement of the charged side chain of lysine by the neutral side chain of heptyline resulted in a decrease of the hydrolysis rate constants by a factor of 10^2-10^6 [16]. The attachment of a carbohydrate on the lysine side chain by glycation might have an even larger influence because not only the charge, but also the size of the side chain of lysine is modified.

For enzymes that are not specific for lysine and arginine residues, information on the influence of glycated residues proximate to the cleavage sites on protein hydrolysis has not been reported. However, it was mentioned that the binding of the protease towards the cleavage site was dependent on the interaction between the catalytic environment of the enzyme and the binding environment of the substrate [17]. According to this information, the P4 to P4' positions (binding site) are the amino acids that come before and after the cleavage site (P1 position) on the primary sequence [17]. Hence, if the glycation sites are on the binding site positions, it is possible that the glycation influences protein hydrolysis catalysed by non-lysine/arginine specific enzymes as well.

In addition to the above, it should be noted that during hydrolysis of native proteins, not all cleavage sites are hydrolysed at an equal rate. This means that when hydrolysing a protein by an enzyme, the final DH might not reach the theoretical maximum. It has been shown for instance that bovine trypsin has different hydrolysis rates towards the various lysine and arginine residues within the native β-lactoglobulin and β-casein [18, 19]. Recently, the relative hydrolysis rate constants of the various cleavage sites in a protein were determined quantitatively and defined as selectivity [20]. It was shown that 6 out of 26 of the cleavage sites in β-lactoglobulin were not cleaved by *Bacillus licheniformis* protease (BLP) and 9 cleavage sites were responsible for ~95 %

of the total hydrolysis rate constant [20]. This means that the glycation of some sites might have a larger effect on the total hydrolysis than the glycation of other sites.

In the present study, the effects of glycation on enzymatic hydrolysis were studied using α-lactalbumin and D-glucose. The protein hydrolysis was performed using lysine/arginine specific enzymes (i.e. bovine and porcine trypsins) and non-lysine/arginine specific enzymes (i.e. BLP, bovine α-chymotrypsin and subtilisin A). The aim is to determine whether altering protein primary structure by glycation influences the protein hydrolysis catalysed by enzymes with various specificities and the quantitative relationship between the degree of glycation and the degree of hydrolysis.

Materials and methods

Materials

α-Lactalbumin (α-LA) was obtained from Davisco Foods International Inc. (Le Sueur, MN, USA). The monoisotopic mass of the protein is 14,178 Da, as found in ExPASy PeptideMass (P00711) (http://web.expasy.org/peptide_mass). Mass spectrometry (MS) analysis showed that there were no naturally glycosylated variants of the protein present in the sample. Of the α-LA powder, the total protein content is ~92.8 % [w/w], as determined using Dumas method (N×6.25, based on the amino acid composition of the α-LA as described in Uniprot accession code P00711, www.uniprot.org). Of the total protein content, ~90 % is α-LA, based on the proportion of α-LA UV_{214} peak area of the total UV_{214} area using reversed phase ultra-high performance liquid chromatography (RP-UHPLC) [21]. The other proteins present were β-lactoglobulin and bovine serum albumin according to MS results (data not shown). The α-LA was free of lactose and glucose, as determined by high performance anion exchange chromatography (HPAEC, data not shown). This was also confirmed by the fact that no glycation by glucose or lactose were found after heating the protein for 10 h at 50 °C with 65 % relative humidity (determined using MS, data not shown). Based on the results from circular dichroism [22], ~72 % of the α-LA was in the apo form (data not shown).

Bovine trypsin (EC 3.4.21.4, Sigma-Aldrich, St. Louis, MO, USA) was treated with tosyl-phenylalanyl-chloromethyl ketone (TPCK) to inactivate any chymotrypsin activity present (chymotrypsin activity is ≤0.1 N-benzoyl-$_L$-tyrosine-ethyl-ester (BTEE) units/mg protein). Porcine trypsin (EC 3.4.21.4, Sigma-Aldrich) was also treated with TPCK and the chymotrypsin activity is ≤1 BTEE units/mg protein. Both bovine and porcine trypsin powders are essentially salt-free freeze-dried powder. According to the manufacturer, the activities of bovine trypsin and porcine trypsin were ≥10,000 and 13,000-20,000 benzoyl-L-arginine-ethyl-ester (BAEE) units/mg protein, respectively. The protein content of the bovine trypsin powder is ~80 % as based on Dumas results (N×5.97, based on the amino acid composition of the protein as described in Uniprot accession code P00760), from which 100 % of the protein is bovine trypsin (based on UV_{280} peak

area from RP-UHPLC). Based on the UV_{214} peak area, 93.4 % is the intact enzyme and the remaining 6.6 % were identified using MS as autolysis products of bovine trypsin (data not shown). The protein content of porcine trypsin powder is ~86 % as based on Dumas results (N×5.84, based on the amino acid composition of the protein as described in Uniprot accession code P00761), from which 100 % of the protein is trypsin (based on UV_{280} peak area from RP-UHPLC). Based on the UV_{214} peak area, 92.6 % is the intact enzyme and the remaining 7.4 % were identified as autolysis products (data not shown). *Bacillus licheniformis* protease (BLP) (batch NS-37005) was obtained from Novozymes (Bagsvaerd, Denmark). The BLP powder was partly insoluble. Therefore, the suspension was centrifuged (10 min, 4,000×*g*, 25 °C) and the supernatant was subsequently dialyzed (cut-off 12-14 kDa) against 150 mM sodium chloride solution and subsequently demineralized water. The freeze dried BLP powder contained 60 % [w/w] protein (N×5.93, based on the amino acid composition of the protein as described in Uniprot accession code P80057) and an activity of 3.9 AU mg^{-1} min^{-1} determined by the azocasein assay [23]. The purity of BLP was determined to be 100 % from the UV_{280} peak area coupled with RP-UHPLC and 92 % from the UV_{214} peak area. In the UV_{214} chromatogram, two peaks were found, of which the main peak (78 % of total peak area) was identified using MS to be BLP (23.6 kDa) and the minor peak (14 % of total peak area) was the pro-peptide (6.9 kDa) [24]. Bovine α-chymotrypsin (EC 3.4.21.1, Sigma-Aldrich) was treated with N-p-tosyl-L-phenylalanine chloromethyl ketone (TPCK) to inactive any trypsin activity present. The α-chymotrypsin is essentially a salt-free freeze-dried powder, which had ≥40 BTEE units/mg protein. The protein content of α-chymotrypsin is ~86 % based on Dumas results (N×5.99, based on the amino acid composition of the protein as described in Uniprot accession code P00766), from which 100 % of the protein is α-chymotrypsin (based on UV_{280} and UV_{214} peak area from RP-UHPLC). Subtilisin A (EC 3.4.21.62, Sigma-Aldrich) is essentially a salt-free freeze-dried powder with a protein content of ~86 % as based on Dumas results (N×5.97, based on the amino acid composition of the protein as described in Uniprot accession code P00780). Since subtilisin A did not bind to the RP column, sodium dodecyl sulfate polyacrylamide gel electrophoresis was used to confirm that impurities in the subtilisin A were neglectable (data not shown). The cleavage sites of these enzymes and the glycation sites in α-lactalbumin were indicated in **Figure 2.1**. All other chemicals were of analytical grade and purchased from Sigma or Merck.

Figure 2.1: Amino acid sequence of α-lactalbumin. (▽) indicates BLP cleavage sites (DE) without glycation sites on the binding site (P4 to P4' positions); (▼) indicates BLP cleavage sites with glycation sites on the binding site. (◇) indicates α-chymotrypsin cleavage sites (FLY) without glycation sites on the binding site. (◆) indicates α-chymotrypsin cleavage sites with glycation sites on the binding site. Amino acids in bold are glycation sites as well as trypsin cleavage sites.

Methods

Glycation of α-lactalbumin (α-LA)

Glycation of α-LA was performed essentially as described previously [3]. α-LA (10 mg powder/mL) and D-glucose (3.5 mg/mL) were each dissolved separately in 10 mM sodium phosphate buffer pH 8.0. These two solutions were mixed to reach a molar ratio of total free amino groups (12 lysines, 1 arginine and the N-terminal): sugar reducing ends of 1:2 and subsequently freeze dried (in 30 mL batches). The freeze dried powder was incubated in a humidity control chamber HCP108 (Memmert, Schwabach, Germany) at 50 °C with 65 % relative humidity for 0, 1, 2, 3, 4, 6, 8 and 10 h and labelled as AG_0-AG_{10}. Samples were immediately cooled down and stored at -20 °C before dialysis. To remove the buffer salts and free glucose, the samples were dialyzed (cut-off 12-14 kDa) against demineralized water for ≥72 h at 4 °C, freeze dried again and then stored at -20 °C. The concentration of remaining glucose in the sample was ≤0.001 g glucose/g protein based on HPAEC analysis (data now shown). As control for effects of heating, an α-LA solution mixed with 10 mM sodium phosphate buffer pH 8.0 without glucose was made. This sample was freeze-dried and heated the same way as described above for 10 h (labelled as A_{10}). The extinction coefficients at UV_{280} of the untreated α-LA and AG_{10} were determined to be $1.4(\pm 0.02)\cdot 10^{-4}$ $L\cdot M^{-1}\cdot cm^{-1}$ by measuring the UV_{280} of a dilution series of a protein solution with a known concentration (based on Dumas results). The value described in Uniprot (accession code P00711) is $1.4\cdot 10^{-4}$ $L\cdot M^{-1}\cdot cm^{-1}$. This means that the glycation does not have an influence on the extinction coefficient of the protein. Therefore, $1.4\cdot 10^{-4}$ $L\cdot M^{-1}\cdot cm^{-1}$ was used to calculate the concentration of the protein for all experiments at UV_{280}. All samples were made in duplicate. The standard deviations in the results of characterization of glycated α-LA represent the differences between the separately glycated duplicate samples.

Size exclusion chromatography (SEC)

SEC was performed on an ÄKTA micro system (GE Healthcare, Uppsala, Sweden) equipped with a Superdex 75 HR 10/300 column (GE Healthcare) as described previously [25]. The experiments were performed at 20 °C with 10 mM sodium phosphate buffer pH 8.0 containing 150 mM NaCl (filtered over a 0.22 μm membrane) at a flow rate of 800 μL/min. The untreated α-LA, AG_0-AG_{10} and A_{10} were dissolved in the eluent (with a protein concentration of 0.2 % (w/v) based on UV_{280}). The eluate (50 μL) was monitored at 280 nm. The average standard deviation of the duplicates for all samples was ~1 %.

Circular Dichroism (CD)

CD spectra were measured using a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan) as described elsewhere [26]. The native and glycated proteins were dissolved in 10 mM sodium phosphate buffer at pH 8.0 at protein concentrations of 0.01 % [w/v] and 0.1 % [w/v] (based on UV_{280}) for far and near UV-CD, respectively. The measurements were performed at 37 °C in quartz cuvettes with an optical path length of 1 mm. The secondary structure of the samples was measured using far UV-CD recorded from 190 to 260 nm. The tertiary structure of the samples

was measured using near UV-CD recorded from 250 to 350 nm. All spectra reported are the average of ten spectra.

o-Phtaldialdehyde method (OPA)

The amount of free amino groups was determined using the OPA method [25]. The native and glycated proteins were dissolved ($7 \cdot 10^{-4}$ M) in 2 % [w/v] sodium dodecyl sulphate solution. 5 µL of the protein solution was added to 300 µL of OPA reagent solution and incubated for 10 min. The experiments were performed in triplicate with a 96 well plate using Tecan Infinite F500 (Tecan, Männedorf, Switzerland) to measure UV_{340} absorption of the samples. To calculate the number of free amino groups, a calibration curve was made using leucine. The protein concentration of each sample was determined by UV_{280}. The average degree of glycation (DG_{av}) was defined as the average percentage of blocked amino groups per protein, which was calculated from equation 1 [26],

$$(1) \quad DG_{av,OPA}\left[\%\right] = \frac{(\#NH_{2,u} - \#NH_{2,g})}{\#NH_{2,u}} \times 100 \ \%$$

where $\#NH_{2,u}$ is the total number of free amino groups per untreated α-LA molecule and $\#NH_{2,g}$ is the number of free amino groups per protein molecule in the sample after glycation.

Reversed phase ultra-high performance liquid chromatography electron spray ionization time of flight mass spectrometry (RP-UHPLC-ESI-Q-TOF-MS)

The untreated and glycated α-LA were analyzed on an H class Acquity UPLC® system (Waters, Milford, MA, USA) equipped with a C4-reversed phase column (UPLC BEH C4 1.7 µm, 2.1×100 mm, Waters) coupled to an Acquity UPLC® PDA detector (Waters). Eluent A was Millipore water containing 1 % [v/v] acetonitrile (ACN) and 0.1 % [v/v] trifluoreacetic acid (TFA). Eluent B was ACN containing 0.1 % [v/v] TFA. The gradient profile was performed using the following steps: from 0 to 2 min isocratic 10 % B, from 2 to 12 min linear gradient from 10 % to 75 % B, from 12 to 15 min linear gradient from 75 % to 100 % B, from 15 to 20 min isocratic at 100 % B, then re-equilibration to the initial conditions. Flow rate was 350 µL/min. Samples were diluted into eluent A (1 mg/mL) and centrifuged (10 minutes, 14,000×g, 20 °C) before injection. The injection volume was 5 µL. The PDA detector was operated at a sampling rate of 40 points/second in the range 200-400 nm with a 1.2 nm resolution. The molecular masses of untreated α-LA, AG_0-AG_{10} and A_{10} were determined with an online Synapt G2-Si high definition mass spectrometer (Waters). The MS was calibrated with sodium iodide. The capillary voltage was set to 3 kV with the source operation in positive ion mode and the source temperature at 150 °C. The sample cone was operated at 30 V. Nitrogen was used as desolvation gas (500 °C, 800 L/h) and cone gas (200 L/h). MS were collected between m/z 150-4000 with a scan time of 0.3 seconds. Data were acquired by MassLynx software v4.1 (Waters). The mass spectra were deconvoluted and centred to bar spectra. To compare the MS data with the OPA results, the DG_{av} was calculated using equation 2,

$$(2) \quad DG_{av,MS}\left[\%\right] = \frac{\displaystyle\sum_{i=0}^{n} \frac{i \times TIC \ of \ variant \ i}{TIC \ of \ all \ variants}}{\# \ NH_{2,u}} \times 100 \ \%$$

where variant *i* indicates the protein molecule with *i* number of glucose molecules attached, TIC represents total ion count and $\#NH_{2,u}$ is the total number of free amino groups per untreated α-LA molecule.

High performance anion exchange chromatography (HPAEC)

The amounts of glucose and lactose in the samples were analysed by HPAEC using an ICS 5000 (Thermo Scientific, Sunnyvale, CA, USA) equipped with a CarboPac PA-1 column (2 mm i.d. × 250 mm) in combination with a Carbo Pac guard column (2 mm i.d. × 50 mm) and pulsed amperometric detection (PAD). The measurements were performed as described elsewhere [27].

Enzymatic hydrolysis

The untreated and glycated α-LA variants were dissolved in 10 mL Millipore water at a protein concentration of 1 mg/mL (based on UV_{280}). The protein solutions were pre-heated for 30 minutes at 37 °C and then adjusted to pH 8.0 before addition of the enzyme. During the hydrolysis (2 hours at 37°C), the pH was kept constant with 0.02 M NaOH using a pH-stat (Metrohm, Herisau, Switzerland) [25]. All enzymes were dissolved (10 mg powder/mL) in Millipore water and added to the protein solution. The overall hydrolysis rate constant k_{hydr} was obtained by fitting equation 3 to the curve of DH versus time t,

$$(3) \quad DH\left[\%\right] = DH_{max,\,fit}^{\,enzyme} - DH_{max,\,fit} \,/\left(1 + k_{hydr} \times DH_{max,\,fit}^{\,enzyme} \times t\right)$$

in which the k_{hydr} and $DH_{max,fit}^{enzyme}$ are fitting parameters for the overall hydrolysis rate and the maximum DH, respectively. It was assumed that the enzymatic hydrolysis follows the 2nd order reaction kinetics. The parameter $k_{hydr} \times DH_{max,fit}^{enzyme}$ was used to compare hydrolysis rates of the glycated proteins in order to rule out the influence of $DH_{max,fit}^{enzyme}$ on the k_{hydr}. To obtain similar value of $k_{hydr} \times DH_{max,fit}^{enzyme}$ [s^{-1}] for untreated α-LA, an enzyme to substrate ratio of 1:25, 1:38, 1:25, 1:13 and 1:23 [w/w protein] was used by adding 50 μL bovine trypsin, 30 μL porcine trypsin, 67 μL BLP, 90 μL α-chymotrypsin or 50 μL subtilisin A solution in 10 mL protein solution, respectively. The enzyme was added to the substrate solution within 5 minutes after dissolving in water in order to prevent autolysis. The experimental degree of hydrolysis (DH) was calculated using equation 4 [28],

$$(4) \quad DH\left[\%\right] = V_b \times N_b \times \frac{1}{\alpha} \times \frac{1}{m_p} \times \frac{1}{h_{tot}} \times 100 \ \%$$

where V_b [mL] is the volume of NaOH added in L; N_b [mol/L] is the normality of NaOH; α is the average degree of dissociation of the α-NH group (1/α=1.3 at 37°C and pH 8.0) [24]; m_p [g] is the

mass of protein in the solution in g; h_{tot} [mmol/g] is the number of millimoles of peptide bonds per gram of protein (8.6 mmole/g for α-LA). The h_{tot} is calculated based on the number of peptide bonds per α-LA molecule and the molecular weight reported in Uniprot (accession code P00711). A blank experiment was performed by incubating the untreated α-LA in 10 mL Millipore water at a protein concentration of 1 mg/mL (based on UV_{280}) with no enzyme for 2 hours. The added value of NaOH from the blank experiment was subtracted from all time points in each hydrolysis.

Results and discussion

Characterization of glycated α-LA

In the untreated α-LA as well as in the controls (AG_0 and A_{10}), 13±1 amino groups per protein were determined by OPA measurements (no further data shown). This is equal to the sum of the number of lysines (12) plus the N-terminal in the α-LA sequence. Upon heat treatment in the presence of glucose, the $DG_{av,OPA}$ as determined by the OPA method, increased from 0 % (AG_0) to 53 % (AG_{10}) (**Figure 2.2**).
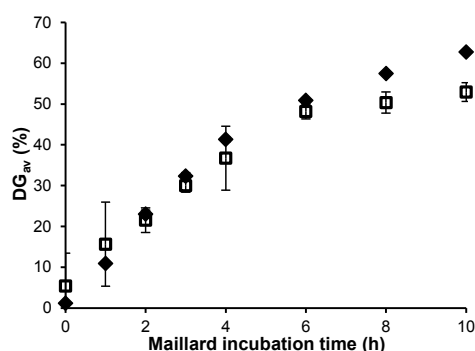

Figure 2.2: Average degrees of glycation of α-lactalbumin heated with D-glucose at 50 °C with 65 % relative humidity over 0-10 h Maillard incubation time. (□) OPA results ($DG_{av,OPA}$) and (♦) MS results ($DG_{av,MS}$).

The degree of glycation was also quantified using the mass spectra ($DG_{av,MS}$) (**Figure 2.3**). For the blank sample containing glucose and protein but without heating (AG_0), the mass spectrum showed that ~20 % of α-LA molecules had one attached glucose molecule (resulting in a molecular mass of 14,340 Da) (**Figure 2.3**). This corresponds to an average degree of glycation ($DG_{av,MS}$) of 1.5 % (**Figure 2.2**). The $DG_{av,MS}$ calculated using MS increased from 1.5 % (AG_0) to 63.0 % (AG_{10}) (**Figure 2.2**). Since OPA and MS results showed similar values for glycation and MS method had lower standard deviations, MS results were further used to present the DG_{av}.
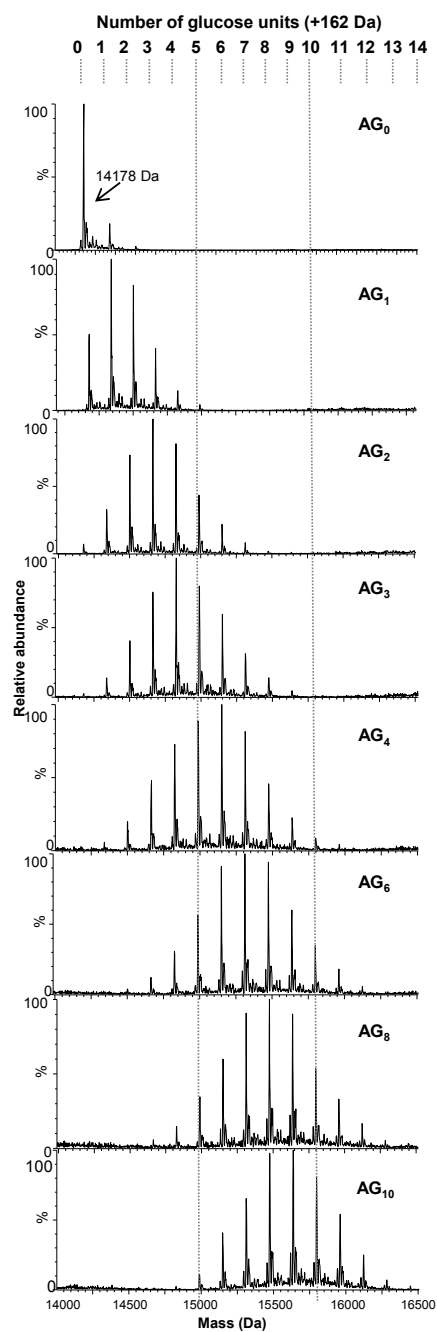
Figure 2.3: Deconvoluted mass spectra of α-lactalbumin (indicated by arrow) heated with D-glucose at 50 °C with 65 % relative humidity over 0-10 h.

In addition, the MS method visualised the distribution of masses of each sample, providing the information on the heterogeneity of the sample. It is important to note that even the mass distribution does not represent the complete heterogeneity of the sample because there are also variants of glycated proteins with the same molecular mass. These variants have the same number of glucose moieties attached but at different locations within the protein. The number of possible molecules with a certain number of glycated residues can be calculated from equation 5:

$$(5) \quad N = \frac{\#GS!}{(\#GS - \#glycated\ sites)! \times \#glycated\ sites!}$$

where N is the number of possible molecules with the same number of glucose attached, GS represents the glycation sites and n! represents the n factorial (i.e. n!=1×2×3×…×n). For α-LA as many as 3,432 possible variants can be formed when 7 out of 14 possible glycation sites are modified. As a consequence, for the case of AG$_8$, in which 11 variants of α-LA-glucose conjugates were identified by MS with 3 to 13 glucoses attached, there were actually 16,277 ($\sum_{i=3}^{13} \frac{14!}{(14-i)! \times i!}$

=16,277, i=#glycated sites) possible modified protein molecules.

Using CD, it was shown that at 37 °C the secondary and tertiary structure of all samples (AG$_0$-AG$_{10}$ and A$_{10}$) were similar to those of the untreated α-LA (**Figures 2.4A** and **2.4B**, respectively).
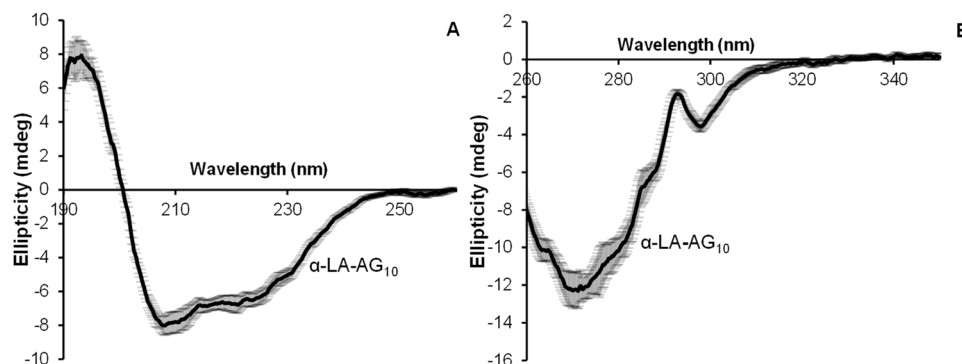

Figure 2.4: (A) Far UV-CD spectra and (B) near UV-CD of untreated α-lactalbumin and α-lactalbumin heated with D-glucose at 50 °C with 65 % relative humidity over 0-10 h, since there was no significant difference between the spectra, the line is the average of spectra over all samples. Error bars are shown to illustrate the standard deviation.

In addition, it was confirmed by SEC that in all samples (untreated α-LA, AG$_0$-AG$_{10}$ and A$_{10}$), the majority (≥90 %) of α-LA was present as monomeric proteins (based on the UV$_{280}$ peak area of the monomer as percentage of the total UV$_{280}$ peak area) (**Figure 2.5**). This means that the changes observed in hydrolysis can indeed be attributed to the modification of protein primary structure, rather than other effects of the modification or treatment.
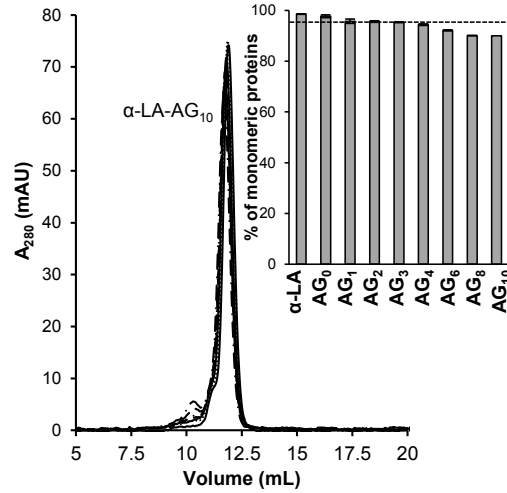
Figure 2.5: Size exclusion elution patterns of untreated α-lactalbumin and α-lactalbumin heated with D-glucose over 0-10 h and the inset shows the percentage of monomeric proteins. The dashed line (------) in the inset represents the precentage of monomeric protein in A_{10}.

Effect of glycation on enzymatic protein hydrolysis

Enzymatic hydrolysis of untreated α-LA by all enzymes

For each enzyme and substrate combination, the theoretical maximal DH that can be reached during hydrolysis was calculated from equation 6.

$$(6) \quad DH_{max,theo}^{enzyme}\left[\%\right] = \frac{\#\ cleavage\ sites}{\#\ peptide\ bonds} \times 100\ \%$$

For trypsin (specific for lysine and arginine), BLP (specific for glutamic and aspartic acids) and α-chymotrypsin (specific for leucine, phenylalanine and tyrosine) hydrolysis of α-LA, the theoretical maximal degrees of hydrolysis are 10.7, 16.4 and 17.2 %, respectively (**Figure 2.1**). For subtilisin A, the maximal theoretical degree of hydrolysis is 100 % since it is an a-specific enzyme. The experimental DH values of untreated α-LA hydrolysed by bovine trypsin, BLP and α-chymotrypsin are 3.6, 11.2 and 10.2 %, respectively, which are much lower than the $DH_{max,theo}^{enzyme}$ values. It is concluded that not all cleavage sites were efficiently hydrolysed by these enzymes. Based on the fitting parameter $DH_{max,untreated}^{enzyme}$ of untreated α-LA (derived from equation 3), the percentage of cleavage sites that were efficiently used by the enzyme, or in other words the percentage of high selectivity sites (HSSs) of the enzyme, was estimated from equation 7.

$$(7) \quad HSS_{protein}^{enzyme}\left[\%\right] = \frac{DH_{max,untreated}^{enzyme}}{DH_{max,theo}^{enzyme}} \times 100\ \%$$

For porcine trypsin (PT), the fitting parameter $DH_{max,untreated}^{PT}$ of untreated α-LA is 10.7 % (derived from equation 3), which is the same as the $DH_{max,theo}^{PT}$ (10.7 %) (equation 6). Therefore, the

$HSS_{\alpha-LA}^{PT}$ [%] is ~100 %. The $DH_{max, untreated}^{BT}$ of α-LA by bovine trypsin (BT) is one third of the $DH_{max, untreated}^{PT}$ by porcine trypsin. A higher HSS for porcine trypsin compared to bovine trypsin was expected because porcine trypsin has been reported to have less missed cleavages during protein hydrolysis than bovine trypsin [29], which means that porcine trypsin can hydrolyse the cleavage sites more efficiently.

Theoretical dependency of protein hydrolysis with modifications on the primary structure

Based on the above-mentioned results from the model made with α-LA and glucose, we derived equation 8 to describe the theoretical dependence of the degree of hydrolysis on the modifications of protein primary structure:

$$(8) \quad DH_{max}^{enzyme} [\%] = DH_{max, untreated}^{enzyme} - \frac{m \times n \times q}{\# \ peptide \ bonds}$$

where m (0≤m≤100 %) represents the percentage of cleavage sites that have a modification on at least one of the binding site positions (P4 to P4'). The value of m depends on the enzyme specificity and the primary sequence of the substrate (**Figure 2.1**). Taking one cleavage site D63 (underlined) for BLP as an example, WCK<u>D</u>DQNP (P4 to P4') is the binding site of D63. K62 is the glycation site that is on one of the binding site positions (P2) of D63. Parameter n (0≤n≤1) represents the binding site dependency coefficient. The value of n=0 when the interaction between the enzyme and the cleavage site is not influenced by modifications on the binding site; n=1 when the enzyme cannot hydrolyse cleavage sites with modifications on the binding site. Parameter q is the number of high selectivity sites that can be influenced by the modifications. In the case of glycation, the way in which q depends on the average degree of glycation (DG$_{av}$) can take three forms:

A) The glycation takes place equally at all cleavage sites (EQ) (equation 9);

$$(9) \quad q_{EQ} = DG_{av} \times (\#GS - 1) \times HSS_{protein}^{enzyme} / 100$$

B) The glycation takes place first at the high selectivity sites (HSS) (equation 10);

$$(10) \quad q_{HSS} = \begin{cases} DG_{av} \times (\#GS - 1) / 100 & DG_{av} < HSS_{protein}^{enzyme} \\ HSS_{protein}^{enzyme} \times (\#GS - 1) / 100 & DG_{av} \geq HSS_{protein}^{enzyme} \end{cases}$$

C) The glycation takes place first at the low selectivity sites (LSS) (equation 11);

$$(11) \quad q_{LSS} = \begin{cases} 0 & DG_{av} < 100 - HSS_{protein}^{enzyme} \\ DG_{av} \times (\#GS - 1) / 100 & DG_{av} \geq 100 - HSS_{protein}^{enzyme} \end{cases}$$

where #GS represents the number of glycation sites. The glycation of N-terminal amino group is corrected in all cases because the glycation on the N-terminal amino groups does not influence the extent of hydrolysis.

Effect of glycation on hydrolysis catalysed by non-lysine/arginine specific enzymes

For BLP and subtilisin A hydrolyses, the experimental DH after 2 h hydrolysis for the untreated and glycated samples were 11.2±0.2 % and 28.0±0.5 %, respectively (**Figure 2.6A**). The fitting parameter $DH_{max,fit}^{BLP}$ and $DH_{max,fit}^{subtilisin}$ of all samples were 11.7±0.2 % and 30.0±0.7 %, respectively (**Figure 2.7A**). These results showed that the extent of hydrolysis of BLP and subtilisin A is independent of glycation. This was expected because these two enzymes are reported to be not specific for lysine and arginine. In addition, the above-mentioned results further indicate that there were no other changes in the protein structure except glycation. For BLP hydrolysis, it showed that even with 12 out of 20 cleavage sites (m=60 % in equation 8) (**Figure 2.1**), the hydrolysis was not influenced by glycation. In the case of subtilisin A, m=100 % since subtilisin A is an a-specific enzyme. Similar as for BLP, the hydrolysis of subtilisin A was completely independent of the primary structural modification of the protein. It is concluded that BLP and subtilisin A can hydrolyse cleavage sites with modified residues on the binding site. Thus, the binding site dependency coefficient n equals to 0 in the equation 8 for the BLP and subtilisin A hydrolyses.

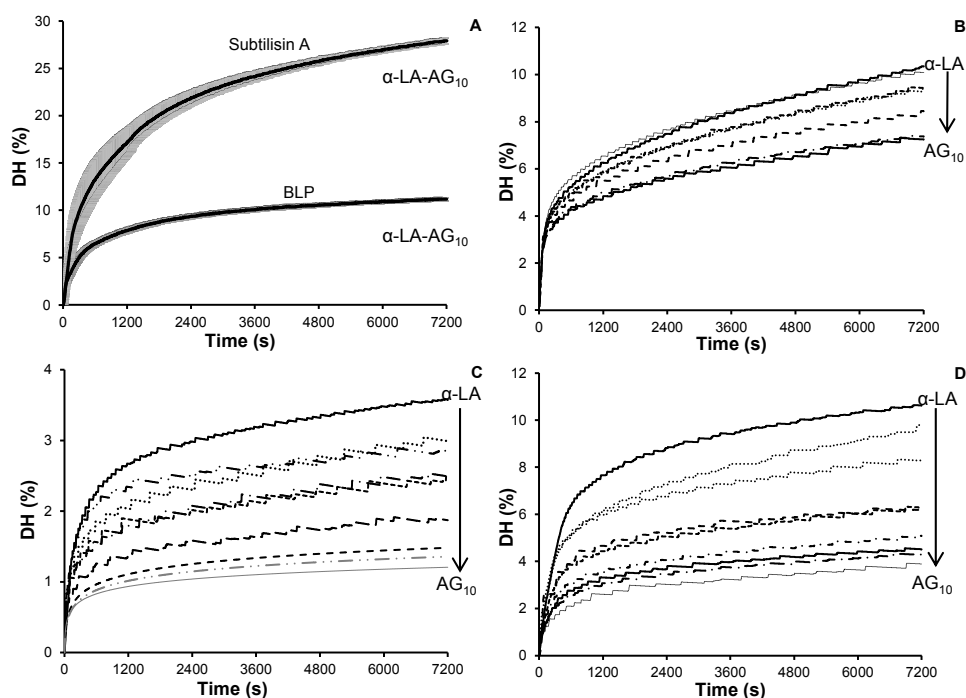**Figure 2.6:** (A) BLP and subtilisin A, (B) α-chymotrypsin, (C) bovine and (D) porcine tryptic hydrolysis curves of untreated α-lactalbumin and α-lactalbumin heated with glucose over 0-10 h. In panel (A), error bars are shown to illustrate the standard deviations of hydrolysis curves of all samples. The averaged standard errors in panel (A) for subtilisin A and BLP were 4.2 and 2.3 %, respectively.

For α-chymotrypsin (CT) hydrolysis, the experimental DH after 2 h hydrolysis decreased from 10.2 to 6.9 % (32 %) with DG_{av} increasing from 0-63 % (**Figure 2.6B**). The fitting parameter $DH^{CT}_{max,fit}$ also decreased by 32 % (from 10.1 % to 6.9 %) with increasing DG_{av} (**Figures 2.7A**). Even though the specificity of α-chymotrypsin is not lysine and arginine, the extent of hydrolysis was negatively influenced by glycation. Because there are no other changes in protein structure except glycation, the binding site dependency coefficient *n* of α-chymotrypsin in equation 8 is not 0. There are 12 out of 21 (*m*=57 %) cleavage sites of α-chymotrypsin that have glycation sites on the binding site (**Figure 2.1**). If the binding site dependency coefficient *n* is equal to 1, the theoretical scenario of $DH^{CT}_{max,fit}$ on DG_{av} was calculated using equations 8 to 11 (**Figure 2.7A**). The dependency of fitting parameter $DH^{CT}_{max,fit}$ of the chymotrypsin hydrolyses on DG_{av} follows the theoretical scenario that the glycation takes place to all sites equally on a protein. These results showed that once the cleavage sites of α-chymotrypsin have undergone modification on the binding site, the enzyme does not hydrolyse the cleavage sites anymore.
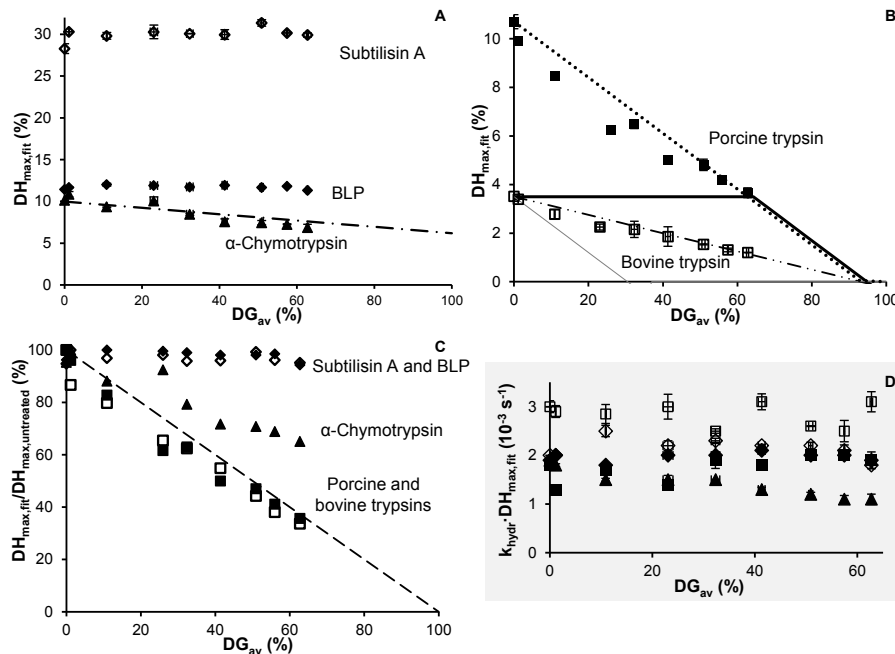

Figure 2.7: DG_{av} *versus* $DH^{enzyme}_{max,fit}$ for untreated α-lactalbumin and α-lactalbumin heated with D-glucose over 0-10 h hydrolysed by (A) non-lysine/arginine specific enzymes: (♦) BLP, (▲) α-chymotrypsin and (◊) subtilisin A, or by (B) lysine/arginine specific enzymes: (□) bovine trypsin and (■) porcine trypsin. (C) indicates the DG_{av} *versus* the relative $DH^{enzyme}_{max,fit}$ and (D) indicates the DG_{av} *versus* the overall hydrolysis rate constant ($k_{hydr}·DH_{max,fit}$). Lines indicate: (------) $DH^{enzyme}_{max,EQ}$ / $DH^{enzyme}_{max,untreated}$ [%] for bovine and porcine trypsins, (—·) $DH^{CT}_{max,EQ}$ [%], (······) $DH^{PT}_{max,EQ}$ [%] and (—··) $DH^{BT}_{max,EQ}$ [%]. For bovine trypsin, (—) $DH^{BT}_{max,LSS}$ [%] and (—) $DH^{BT}_{max,HSS}$ [%] illustrate the theoretical scenarios where the glycation occurs firstly at low and high selectivity sites, respectively.

Effect of glycation on hydrolysis catalysed by lysine/arginine specific enzymes

In contrast to hydrolysis catalysed by non-lysine/arginine specific enzymes, the obtained experimental DH values after 2 h hydrolysis using bovine trypsin decreased from 3.6±0.1 % to 1.2±0.1 % (65 % decrease) with DG_{av} increasing from 0 to 63 % (**Figure 2.6C**). The fitting parameter $DH_{max,fit}^{BT}$ (derived from equation 3) decreased linearly (R^2=0.97) from 3.5±0.1 % to 1.2±0.1 % (66 % decrease) with increasing DG_{av} (**Figures 2.6C** and **2.7B**). For porcine trypsin, the obtained experimental DH after 2 h hydrolysis decreased from 10.2±0.5 % to 3.6±0.3 % (65 % decrease) with DG_{av} increasing from 0 to 63 % (**Figures 2.6D**). The fitting parameter DH_{max}^{PT} (derived from equation 3) also decreased linearly (R^2=0.96) from 10.7±0.3 % to 3.7±0.2 % (65 % decrease) with increasing DG_{av} (**Figures 2.7B**). Because the bovine and porcine trypsins are specific for lysine and arginine, all modified lysines are automatically part of the binding site (m=100 % in equation 8). The theoretical scenario of $DH_{max,fit}^{BT/PT}$ on DG_{av} was calculated using equation 8 to 11, using the binding site dependency coefficient n=1 (**Figure 2.7B**). The dependence of fitting parameter $DH_{max,fit}^{BT/PT}$ of the bovine and porcine tryptic hydrolysis on DG_{av} follows the theoretical scenario when the glycation takes place equally on all glycation sites, which showed the same behaviour as α-chymotrypsin hydrolysis. These results indicate that all amino groups on the protein had equal probability to be glycated. Because all the cleavage sites are hydrolysed efficiently in the case of porcine tryptic hydrolysis ($HSS_{\alpha-LA}^{PT}$ =100 %), the decrease of $DH_{max,fit}^{PT}$ of porcine trypsin matches exactly with the increase of the DG_{av} (**Figure 7B**). The above-mentioned results of bovine and porcine trypsins showed that lysine/arginine specific enzymes cannot hydrolyse glycated cleavage sites (m=100 % and n=1 in equation 8). The values of fitting parameters $DH_{max,fit}$, HSS, m and n for hydrolysis of α-LA by all tested enzymes were summarized in **Table 2.1**.

Table 2.1: Fitted maximal degrees of hydrolysis ($DH_{max,fit}$), percentages of high selectivity sites (HSS), percentages of cleavage sites that have a modification on at least one of the binding site positions (m) and binding site dependency coefficients (n) for hydrolysis of α-lactalbumin by all enzymes from the fits of equations 3, 7 and 8, respectively.

Enzyme	DH$_{max,fit}$ (%)	HSS (%)	m (%)	n
Bovine trypsin	3.5	33	100	1
Porcine trypsin	10.7	100	100	1
BLP	11.7	71	60	0
Subtilisin A	30.0	30	100	0
α-Chymotrypsin	10.1	59	57	1

Comparison of the effect of glycation on hydrolysis catalysed by various enzymes

To compare the influence of glycation among all enzymes, the $DH_{max,fit}^{enzyme}$ values were normalised as relative degree of hydrolysis ($DH_{max,fit}^{enzyme} / DH_{max,untreated}^{enzyme}$) (**Figure 2.7C**), to allow fair comparison of the effects of glycation on hydrolysis among different enzymes. It is clear that the bovine and porcine tryptic hydrolyses have the exact same dependency on glycation. Therefore, the only differences between the two enzymes are the number of high selectivity sites ($HSS_{substrate}^{enzyme}$). Both the bovine and porcine tryptic hydrolysis follow the theoretical scenario when m=100 % and n=1

in equation 8. This is the maximal influence of primary structure modification of a protein on the enzymatic hydrolysis. For all three other enzymes, since the $0 \leq m \leq 100$ % and $0 \leq n \leq 1$, it was expected that the effect of glycation on the hydrolysis would be smaller than the effect on hydrolysis by bovine and porcine trypsins. For BLP and subtilisin A, since the hydrolyses are independent of glycation and $m=60$ % and 100 %, respectively, the n for both enzymes needs to be 0. This means that for all proteins, BLP and subtilisin hydrolyses would be independent of glycation. For α-chymotrypsin, $n=1$ but m can be from 0 to 100 % for different substrates, which means that the effect of glycation on α-chymotrypsin hydrolysis depends on the protein amino acid sequence. For both lysine/arginine specific and non-lysine/arginine specific enzymes, the overall hydrolysis rate constants ($k_{hydr} \cdot DH_{max,fit}$) were independent of glycation (**Figure 2.7D**). It is concluded that only the extent of hydrolysis was influenced by glycation, but not the rate of the hydrolysis.

Conclusion

The modifications of protein primary structure by Maillard induced glycation can quantitatively affect the extent of protein enzymatic hydrolysis, even in the absence of changes in secondary, tertiary or quaternary structure of the proteins. The effect of glycation on hydrolysis depended on the proteases used. Lysine/arginine specific proteases, e.g. bovine and porcine trypsins, cannot hydrolyse glycated cleavage sites. For non-lysine/arginine specific proteases, e.g. BLP and subtilisin A, glycation does not affect the extent of hydrolysis, while α-chymotrypsin cannot hydrolyse cleavage sites with glycated binding sites. Since not all cleavage sites are used by the enzymes, the extent of the effects depends on the enzyme selectivity towards cleavage sites (for trypsin) or cleavage sites near glycation sites (for α-chymotrypsin). Combining the results of all proteases, an equation was derived to describe the effect of modification of protein primary structure on the extent of hydrolysis based on the enzyme specificity, selectivity and binding site sensitivity. With these outcomes, an estimation can be obtained on the effect of the modification (glycation) of protein primary structure on the extent of hydrolysis during *in vitro*, e.g. production of commercial hydrolysates and *in vivo* hydrolysis, e.g. digestion of food proteins.

Acknowledgement

References

[1] M. Darewicz, J. Dziuba, The effect of glycosylation on emulsifying and structural properties of bovine β-casein, Food/Nahrung, 45 (2001) 15-20.

[2] P.A. Wierenga, L. Van Norél, E.S. Basheva, Reconsidering the importance of interfacial properties in foam stability, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 344 (2009) 72-78.

[3] R. ter Haar, H.A. Schols, H. Gruppen, Effect of saccharide structure and size on the degree of substitution and product dispersity of α-lactalbumin glycated via the Maillard reaction, Journal of Agricultural and Food Chemistry, 59 (2011) 9378-9385.

[4] F. Chevalier, J.M. Chobert, D. Mollé, T. Haertlé, Maillard glycation of β-lactoglobulin with several sugars: Comparative study of the properties of the obtained polymers and of the substituted sites, Lait, 81 (2001) 651-666.

[5] K.M. Biemel, O. Reihl, J. Conrad, M.O. Lederer, Formation pathways for lysine-arginine cross-links derived from hexoses and pentoses by Maillard processes: Unraveling the structure of a pentosidine precursor, Journal of Biological Chemistry, 276 (2001) 23405-23412.

[6] R. Fritsch, H. Hoffmann, H. Klostermeyer, Formation of lysinoalanine during heat treatment of milk, Zeitschrift für Lebensmittel-Untersuchung und Forschung, 176 (1983) 341-345.

[7] F.H. Böttger, M.R. Etzel, J.A. Lucey, *In vitro* infant digestion of whey protein–dextran glycates, Food Digestion, 4 (2013) 76-84.

[8] A.S. Pereyra Gonzáles, G.B. Naranjo, L.S. Malec, M.S. Vigo, Available lysine, protein digestibility and lactulose in commercial infant formulas, International Dairy Journal, 13 (2003) 95-99.

[9] Y. Wada, B. Lönnerdal, Effects of different industrial heating processes of milk on site-specific protein modifications and their relationship to *in vitro* and *in vivo* digestibility, Journal of Agricultural and Food Chemistry, 62 (2014) 4175-4185.

[10] M.S. Pinto, J. Léonil, G. Henry, C. Cauty, A.F. Carvalho, S. Bouhallab, Heating and glycation of β-lactoglobulin and β-casein: Aggregation and *in vitro* digestion, Food Research International, 55 (2014) 70-76.

[11] H. Bhatt, A. Cucheval, C. Coker, H. Patel, A. Carr, R. Bennett, Effect of lactosylation on plasmin-induced hydrolysis of β-casein, International Dairy Journal, 38 (2014) 213-218.

[12] F. Morgan, J. Léonil, D. Mollé, S. Bouhallab, Nonenzymatic lactosylation of bovine β-lactoglobulin under mild heat treatment leads to structural heterogeneity of the glycoforms, Biochemical and Biophysical Research Communications, 236 (1997) 413-417.

[13] A. Lapolla, D. Fedele, R. Reitano, N.C. Aricò, R. Seraglia, P. Traldi, E. Marotta, R. Tonani, Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides, Journal of the American Society for Mass Spectrometry, 15 (2004) 496-509.

[14] F.J. Moreno, J.E. Quintanilla-López, R. Lebrón-Aguilar, A. Olano, M.L. Sanz, Mass spectrometric characterization of glycated β-lactoglobulin peptides derived from galacto-oligosaccharides surviving the *in vitro* gastrointestinal digestion, Journal of the American Society for Mass Spectrometry, 19 (2008) 927-937.

[15] S. Carulli, C.D. Calvano, F. Palmisano, M. Pischetsrieder, MALDI-TOF MS characterization of glycation products of whey proteins in a glucose/galactose model system and lactose-free milk, Journal of Agricultural and Food Chemistry, 59 (2011) 1793-1803.

[16] B.M. Sanborn, G.E. Hein, The interaction of trypsin with neutral substrates and modifiers, Biochemistry, 7 (1968) 3616-3624.

[17] I. Schechter, A. Berger, On the size of the active site in proteases. I. Papain, Biochemical and Biophysical Research Communications, 27 (1967) 157-162.

[18] M.M. Vorob'ev, M. Dalgalarrondo, J.M. Chobert, T. Haertlé, Kinetics of β-casein hydrolysis by wild-type and engineered trypsin, Biopolymers, 54 (2000) 355-364.

[19] S.C. Cheison, M.Y. Lai, E. Leeb, U. Kulozik, Hydrolysis of β-lactoglobulin by trypsin under acidic pH and analysis of the hydrolysates with MALDI-TOF-MS/MS, Food Chemistry, 125 (2011) 1241-1248.

[20] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Introducing enzyme selectivity: A quantitative parameter to describe enzymatic protein hydrolysis, Analytical and Bioanalytical Chemistry, 406 (2014) 5827-5841.

[21] C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Determination of the influence of the pH of hydrolysis on enzyme selectivity of *Bacillus licheniformis* protease towards whey protein isolate, International Dairy Journal, 44 (2015) 44-53.

[22] W.H. Heijnis, P.A. Wierenga, W.J.H. Van Berkel, H. Gruppen, Directing the oligomer size distribution of peroxidase-mediated cross-linked bovine α-lactalbumin, Journal of Agricultural and Food Chemistry, 58 (2010) 5692-5697.

[23] O. Akpinar, M.H. Penner, Peptidase activity assays using protein substrates, Current protocols in food analytical chemistry, John Wiley & Sons, Corvallis, OR, USA, 2001, Protocol. Unit C2.2.4.

[24] C.I. Butré, P.A. Wierenga, H. Gruppen, Influence of water availability on the enzymatic hydrolysis of proteins, Process Biochemistry, 49 (2014) 1903-1912.

[25] C.I. Butré, P.A. Wierenga, H. Gruppen, Effects of ionic strength on the enzymatic hydrolysis of diluted and concentrated whey protein isolate, Journal of Agricultural and Food Chemistry, 60 (2012) 5644-5651.

[26] R.J.B.M. Delahaije, P.A. Wierenga, M.L.F. Giuseppin, H. Gruppen, Improved emulsion stability by succinylation of patatin is caused by partial unfolding rather than charge effects, Journal of Colloid and Interface Science, 430 (2014) 69-77.

[27] M.P. Van Gool, I. Vancsó, H.A. Schols, K. Toth, G. Szakacs, H. Gruppen, Screening for distinct xylan degrading enzymes in complex shake flask fermentation supernatants, Bioresource Technology, 102 (2011) 6039-6047.

[28] J. Adler-Nissen, Enzymic hydrolysis of food proteins, Elsevier Applied Science Publishers, London, UK, 1986, pp. 122-123.

[29] S.J. Walmsley, P.A. Rudnick, Y. Liang, Q. Dong, S.E. Stein, A.I. Nesvizhskii, Comprehensive analysis of protein digestion using six trypsins reveals the origin of trypsin as a significant source of variability in proteomics, Journal of Proteome Research, 12 (2013) 5666-5680.

CHAPTER 3

Towards predicting protein hydrolysis by bovine trypsin

The extent of protein enzymatic hydrolysis is considered to be mostly determined by protease specificity and the number of cleavage sites (CSs) on the substrate. However, this theoretical maximum is typically not reached. The limited hydrolysis of certain CSs may be due to the differences in enzyme preference resulting from neighbouring amino acids (AAs) of the CS (secondary specificity). This study aims to find the link between enzyme secondary specificity and the relative hydrolysis rate constants (selectivity) of individual CSs in a protein, to better predict the experimental maximum degree of hydrolysis. Bovine tryptic hydrolysis of α-lactalbumin and β-casein showed that ≥50 % of the CSs were inefficiently hydrolysed. The selectivity depended on the number of charged AAs at P2 and P2' positions of a CS. Trypsin efficiently hydrolyses CSs with neutral AAs at these two positions. The selectivity towards 12 out of 18 (67 %) CSs in β-lactoglobulin was correctly predicted. The predicted maximum degree of hydrolysis is within ~13 % error of the experimental value, which is ~5 times better than the prediction based only on enzyme specificity. This work helps to estimate the extent of hydrolysis and the peptide formation of bovine tryptic hydrolysis of other substrates.

Introduction

Typically, the final composition of enzymatic protein hydrolysates is considered to be almost entirely dependent on the specificity of the protease that is used. Protease specificity is defined as the type(s) of amino acid (AA) next to which the enzyme can hydrolyse the peptide bond (the cleavage site), either at the carboxylic or amino side. Although not often discussed, it is known that some proteases do not hydrolyse all cleavage sites (CSs) in a protein efficiently [1-3]. It has been suggested that this preference of the enzyme towards CSs was caused by the difference in the neighbouring AAs of the CS (also called secondary specificity) [4]. Recently, another indication of the secondary specificity was obtained by the quantification of relative hydrolysis rate constants for all CSs in a protein, referred to as the enzyme "*selectivity*" [3]. The current work aims to investigate the link between neighbouring AAs of the CS and the measured selectivity. Bovine trypsin (in further text referred to as trypsin), is specific for peptide bonds on the carboxylic side of lysine and arginine, and it has been reported to hydrolyse CSs in a protein at different rates [1, 2, 5]. This work will allow the formulation of rules to estimate the extent of tryptic hydrolysis and the hydrolysate composition at the end of hydrolysis by predicting the enzyme selectivity.

The influence of the neighbouring AAs of the CS on the hydrolysis is assumed to be due to the interactions between the subsite site (S4-S4') of the enzyme and the binding site (P4-P4') of the substrate [4]. In this concept, the P4-P4' positions refer to the AAs on the N- to C-terminal sides of the CS, respectively, and P1 identifies the AA for which the enzyme is specific [4]. Over the years, this concept has been used to analyse the observed CSs that were not hydrolysed by trypsin, referred to as the *missed cleavage* [6]. Recently, for bovine trypsin, the obtained binding site sequences of missed CSs were summarised in a review article [7], which were included in **Table 3.1**. In addition to the absolute missed cleavages, the binding site sequences that caused moderate inhibition for porcine trypsin based on the hydrolysis of synthetic peptides [8], were also mentioned in **Table 3.1**. Similar observations of moderate inhibition were reported according to the different hydrolysis rate constants for synthetic peptides with varying binding site sequences [9-12]. However, since in these studies the sources of the trypsin (e.g. bovine or porcine) were either not mentioned or from insects, their conclusions were not included in **Table 3.1**. The missed cleavage rules for bovine trypsin suggested that the enzyme activity is hindered by the presence of charged AAs on the binding site sequence of a CS. These rules have been used to predict missed cleavage behaviour of trypsin in synthetic peptides and in quantitative proteomics [8, 13, 14]. Also, they are used by PeptideCutter at ExPASy website (http://web.expasy.org/peptide_cutter) to generate possible peptides of a protein after tryptic hydrolysis (**Table 3.1**). Despite all these efforts, these rules have not been used to predict tryptic hydrolysis of a protein due to some unsolved issues. Firstly, it is unknown whether the tryptic selectivity towards CSs in a protein is dominated by protein primary sequence, or by protein secondary, tertiary and quaternary structure. Secondly, there are no clear rules to differentiate CSs that are hydrolysed at different rates. Recently, Fernández and Riera thoroughly studied the peptide formation kinetics in tryptic

hydrolysis of β-lactoglobulin [2]. It was observed that K47-48 was a missed CS [2], due to the proline on P1' position of the binding site sequence. The hindering effect of proline at the P1' position was reported previously [6, 15, 16]. But for the hydrolysed CSs, the difference in hydrolysis rate constants between the fastest and the slowest CSs was up to 300 times [2], which cannot be explained by the missed cleavage rules (**Table 3.1**). Therefore, it is necessary to find the link between the binding site sequence of the CS and trypsin selectivity towards that CS, which would help the prediction of the extent of tryptic hydrolysis of a protein and the kinetics of the peptide formation.

Table 3.1: Binding site sequences of cleavage sites that are not hydrolysed by bovine trypsin and only slowly hydrolysed by porcine trypsin reported in previous studies. HSS, ISS and LSS represent high, intermediate and low selectivity sites, respectively.

	P4	P3	P2	P1	P1'	P2'	P3'	Reference	Experimental	Prediction
			C	K	D			6, ExPASy	1 missed cleavage	Confirmed
			D	K	D			6, ExPASy	N/A	
			C	K	H			6, ExPASy	N/A	
			C	K	Y			6, ExPASy	N/A	
			C	R	K			6, ExPASy	N/A	
Complete			R	R	H			6, ExPASy	N/A	
missed			R	R	R			6, ExPASy	N/A	
cleavage				K/R	P			6, 15, 16, ExPASy	1 missed cleavage	Confirmed
(bovine)				K/R	K/R			6, 7, ExPASy	2HSS, 2LSS	Rejected
				K/R	D/E			6, 7	5HSS, 1ISS, 3LSS	Rejected
				K/R		D/E	D/E	6, 7	N/A	
	D/E	D/E		K/R				6, 7	N/A	
		D/E		K/R		D/E		6, 7	N/A	
			D/E	K/R				7, 8	IHSS, 4ISS, 4LSS	Rejected
		D/E	T	K/R				7, 8	N/A	
Moderate				K/R	T	D/E		8	N/A	
inhibition			E	K/R				8	1ISS, 1LSS	Confirmed
(porcine)				K/R	D			8	2HSS, 1LSS	Rejected

In this study, the trypsin selectivities towards CSs in two proteins with limited tertiary structure, i.e. apo α-lactalbumin and β-casein, were determined. The molecular properties of the binding site sequence of the CSs were analysed and correlated with the trypsin selectivity, in order to formulate rules that can explain and predict trypsin selectivity towards CSs in a protein. Based on these rules, trypsin selectivity towards CSs in another protein (β-lactoglobulin) was predicted and the correctness of the prediction was checked with the experimental data.

Materials and methods

Materials

α-Lactalbumin (α-LA) was obtained from Davisco Foods International Inc. (Le Sueur, MN, USA). Based on results from circular dichroism [17], ~72 % of the α-LA was in apo form and the rest was in holo form [18]. β-Casein (β-cas, C6905), β-lactoglobulin (β-LG, L0130), bovine trypsin (EC

3.4.21.4, T1426) and aprotinin (A6279) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Trypsin was treated with tosylphenylalanylchloromethylketone (TPCK) to inactivate any chymotrypsin activity present (≤0.1 N-benzoyl-L-tyrosineethylester (BTEE) units/mg protein). Trypsin is essentially salt-free freeze-dried powder, specified to have ≥10,000 benzoyl-L-arginineethylester (BAEE) units/mg protein. Aprotinin, present in 0.9 % sodium chloride and 0.9 % benzylalcohol solution, was determined to have a concentration of 2.3 mg/mL calculated using the UV_{214} peak area in reversed phase ultra-high performance liquid chromatography [3]. More information about the purity and properties of the used proteins and trypsin is provided in **Table 3.2**. All other chemicals were of analytical grade and obtained from Sigma or Merck.

Methods
Enzymatic hydrolysis of the proteins
The enzymatic hydrolysis of proteins was performed similarly as described elsewhere [3]. The α-LA, β-cas or β-LG were dissolved in 10 mL Millipore water at a concentration of 1 % [w powder/v]. Each protein solution was adjusted to pH 8.0 and equilibrated for 0.5 h at 37 °C. Trypsin was dissolved (10 mg powder/mL) in Millipore water and added to the equilibrated protein solution to reach an enzyme to substrate ratio of 1:100 [w/w] for all three proteins. The hydrolysis took place for 2 h using a pH-stat with 0.2 M NaOH to keep the pH constant (Metrohm, Herisau, Switzerland). The degree of hydrolysis measured using pH-stat (DH_{stat}) was calculated from equation 1 [19],

$$(1) \quad DH_{stat} \ [\%] = V_b \times N_b \times \frac{1}{\alpha} \times \frac{1}{m_p} \times \frac{1}{h_{tot}} \times 100 \ \%$$

where V_b [mL] is the volume of NaOH added; N_b [mol/L] is the normality of NaOH; α is the average degree of dissociation of the α-NH group (1/α=1.3 at 37 °C and pH 8) [20]; m_p [g] is the mass of protein in the solution (taking into account the protein content in the powder based on Dumas results); h_{tot} [mmol/g] is the number of millimoles of peptide bonds per gram of protein (**Table 3.2**). During the hydrolysis, 200 µL samples were taken at 6 DH points for each hydrolysis. Samples were taken at DH 0, 1, 2, 3, 4 and 4.7 % for α-LA, at DH 0, 1, 2, 3, 4 and 4.3 % for β-cas, and at DH 0, 1.5, 3, 4.5, 6 and 7.9 % for β-LG. To have a sufficient and fast inactivation of the trypsin, 3 µL aprotinin was directly added to each sample, yielding a molar ratio of trypsin to aprotinin of 1:1.5. The inactivation was confirmed using pH-stat by adding 150 µL aprotinin to 10 mL protein solution immediately after the trypsin was added. Once the aprotinin was added, the pH of the protein solution remained constant, showing that the hydrolysis was successfully inhibited (data not shown). As a control, the α-LA, β-cas and β-LG were incubated in the pH-stat in 10 mL Millipore water at a protein concentration of 1 % [w powder/v] in the absence of the enzyme for 2 h. The added volume of NaOH (due to the consumption of CO_2) from each control experiment was used to correct the added volume of NaOH during hydrolysis.

Table 3.2: The nitrogen to protein conversion factor (N-factor), the protein content and purity, the molar weight, the number of peptide bonds per gram of protein (h_{tot}), the number of amino acids (#AA) and the molar extinction coefficient at 214 nm (ε_{214}) of the bovine trypsin and three proteins (α-lactalbumin, β-casein and β-lactoglobulin) used, and the number of cleavage sites (#CS), the theoretical ($DH_{max,theo}$) and the fitted ($DH_{max,fit}$) maximum degree of hydrolysis, the number of high selectivity sites (#HSS) and the number of unique peptides of bovine tryptic hydrolysis of the three proteins.

Protein	Uniprot Code[1]	N-factor[1] [g of protein/g N]	Protein content [w/w]	Purity[3] [%]	Molar weight[1] [Da]	h_{tot}[4] [mmol/g]	#CS[1]	#AA[1]	$DH_{max,theo}$[5] [%]	$DH_{max,fit}$[2] [%]	#HSS	ε_{214} [M^{-1}·cm^{-1}]	#unique peptides[8]
α-LA	P00711	6.25	93 %	90	14,186	8.6	13	123	10.7	5.1	5	300,461	23
β-cas	P02666	6.39	90 %	90	23,983[7]	8.7	15	209	7.2	4.3	7	424,047	34
β-LG	P02754	6.29	96 %	100	18,367 (A[6]) 18,281 (B[6])	8.8	18	162	11.2	8.3	7	296,850 293,443	51
Trypsin	P00760	5.97	80 %	100	23,305	-	-	-	-	-	-	-	-

α-LA: α-lactalbumin; β-cas: β-casein ; β-LG: β-lactoglobulin

[1] From Uniprot (http://www.uniprot.org)

[2] From fitting the experimental data to the equation of 2nd order reaction equation

[3] The protein purity is defined as the percentage of the UV$_{280}$ peak area of the target protein in the total UV$_{280}$ area from RP-UPLC.

[4] h_{tot} is the number of millimoles of peptide bonds per gram of protein calculated using the amino acid sequence and molar weight of the protein reported in Uniprot.

[5] The $DH_{max,theo}$ refers to the theoretical maximum DH, calculated by dividing the number of cleavage sites by the total number of peptide bonds in the protein.

[6] The A and B refer to the variant A and B in the bovine β-lactoglobulin, respectively.

[7] The molar weight was calculated with 5 phosphoserines on the β-cas molecule as shown in RP-UPLC-MS.

[8] #unique peptides refers to the number of unique peptides found in the UPLC chromatograms of the tryptic hydrolysates of these proteins.

Both 1st and 2nd order reaction equations were used to fit the hydrolysis curve. The R^2 of the fit (R^2=1-SS$_{residual}$/SS$_{total}$, SS means the sum of squared errors) was higher for the fit with 2nd order reaction equation than for the fit with 1st order reaction equation. Therefore, the equation of 2nd order reaction was used to fit the hydrolysis curve. The fitted maximum degree of hydrolysis (DH$_{max,fit}$) and the overall hydrolysis rate constant (k_{DH}^{hydr} [s^{-1}]=k$_{hydr}$·DH$_{max}$) were obtained by fitting equation 2 with experimental data [18].

$$(2)\ DH[\%] = DH_{max,\,fit} - DH_{max,\,fit}\,/(1 + k_{hydr} \times DH_{max,\,fit} \times t\,)$$

All hydrolyses were performed in duplicate, and samples were taken at same DH points in each individual hydrolysis experiment. Samples taken from each duplicate hydrolysis were analysed as described below. The peptide analysis (the following 5 sections) was performed and adapted from the method developed previously by Butré and co-workers [3].

Reversed phase ultra-high performance liquid chromatography (RP-UPLC)

The hydrolysates were analyzed on an H class Acquity UPLC® system (Waters, Milford, MA, USA) equipped with a BEH C18 column (1.7 µm, 2.1×100 mm, Waters) coupled to an Acquity UPLC® PDA detector (Waters). Eluent A was 1 % [v/v] acetonitrile (ACN) containing 0.1 % [v/v] trifluoreacetic acid (TFA) in Millipore water and eluent B was 100 % ACN containing 0.1 % [v/v] TFA. To reduce disulphide bridges, samples were incubated for 2 h with 100 mM ditiothreitol (DTT) in 50 mM Tris-HCl buffer at pH 8.0 at a protein concentration of 0.5 % [w/v]. Samples were then further diluted in eluent A to a final concentration of 0.1 % [w/v] and centrifuged (10 min, 14,000×g, 20 °C) before injection (4 µL). The elution gradient was the same as reported previously [3].

Electron spray ionization time of flight mass spectrometry (ESI-Q-TOF-MS)

The mass spectra were collected with an online SYNAPT G2-Si high definition mass spectrometry (Waters) coupled to the RP-UPLC. The system was calibrated with sodium iodide. The capillary voltage was set to 3 kV with the source operation in positive ion mode and the source temperature at 150 °C. The sample cone was operated at 35 V. Nitrogen was used as desolvation gas (500 °C, 800 L/h) and cone gas (200 L/h). MS and MS/MS (MSe method) data were collected between m/z 200-3000 with a scan time of 0.3 seconds. Online lock mass data (angiotensin II, [M+2H]$^{2+}$: 523.7751) were collected and the correction was applied during the measurements. The data were manually analysed using MassLynx software v4.1 (Waters).

Peptide identification and quantification

The mass tolerance for the accepted annotation was within 50 ppm difference between the theoretical and measured mass. Methionine oxidation (+16 Da) in α-LA and phosphorylation (+80 Da per phosphoserine) in β-cas were considered as modifications during data analysis. The peptides were annotated by MS and confirmed -when possible- by identifying the b and y

fragments in the MS/MS spectra. To quantify the concentration of each peptide, equation 3 was used [3]:

$$(3) \quad C_{peptide} \left[\mu M \right] = \frac{A_{214} \cdot Q}{\varepsilon_{214} \cdot l \cdot V_{inj} \cdot k_{cell}}$$

where $C_{peptide}$ [µM] is the concentration of peptide, A_{214} [µAU·min] is the UV peak area at 214 nm, V_{inj} [µL] is the volume of sample injected, Q [µL·min^{-1}] is the flow rate and l [cm] is the path length of the UV cell, which is 1 cm according to the manufacturer. The molar extinction coefficient ε_{214} [M^{-1}·cm^{-1}] at 214 nm of each peptide was calculated as described previously [21]. Due to the multiple reflection by the coating of the cell, the effective path length of the light through the cell is not the same as the length of the detector cell as specified by the producer. This issue has also been discussed in earlier studies [22-24]. To correct for this effect, the cell constant of the UV detector (k_{cell}), was determined using a series of standard solutions of β-LG, β-cas and angiotensin II, using the approach as described elsewhere [25]. The ratio between the measured and expected peak areas was then taken as the value of k_{cell}. For the UV detector used in this study, the k_{cell} was determined to be 0.78. The linear region of the peak area (A_{214}) in the UV detector ranged from 5×10^1 to 6×10^5 µAU·min. In the hydrolysates, peaks with area ≤5×10^1 µAU·min were not included in the quantification. In all hydrolysates, on average 96±6 % of the expected total A_{214} was found in the chromatograms and of the total A_{214}, 98±3 % was assigned to annotated peptides.

Quality of the annotation and quantification of peptide analysis

Three parameters, i.e. amino acid (equation 4), peptide (equation 5) and molar (equation 6) sequence coverages, were used to check the quality of peptide analysis of each hydrolysate [3].

$$(4) \quad Amino \; acid \; sequence \; coverage \left[\% \right] = \frac{\# \, unique \; annotated \; AA}{\# \, AA_{protein}} \times 100 \; \%$$

$$(5) \quad Peptide \; sequence \; coverage \left[\% \right] = \frac{\# \, AA(\, annotated \; peptides \,)}{\# \, AA(\, annotated \; peptides \,) + \# \, AA(\, missing \; peptides \,)} \times 100 \; \%$$

$$(6) \quad Molar \; sequence \; coverage \left[\% \right] = \left(1 - \frac{\sqrt{\frac{\sum \left(C_n - C_0 \right)^2}{\left(\# \, AA_{protein} - 1 \right)}}}{C_0} \right) \times 100 \; \%$$

where C_n [µM] is the concentration of each individual amino acid (n) in the protein sequence obtained from quantification of the peptides; C_0 [µM] is the initial injected concentration of the protein and #AA$_{protein}$ is the number of amino acids in the sequence of the parental protein [3]. Missing peptides are the peptides that are not quantified in the hydrolysate but should be

present given the annotated peptides [3]. The injected initial concentrations C_0 of the three proteins were 59, 34 and 49 µM for α-LA, β-cas and β-LG, respectively (calculated by taking into account the protein content and protein purity). The experimental C_0 for α-LA, β-cas and β-LG measured in sample DH 0 % were 58±2, 32±1, 49±1 µM, respectively (equation 3). The average values for the amino acid, peptide and molar sequence coverages were the average of all hydrolysates in the duplicate experiments.

Determination of the enzyme selectivity
The concentrations of all peptides originating from hydrolysis of the same peptide bond ($C_{i,t}$) were summed up (equation 7):

$$(7) \quad C_{i,t}[\mu M] = \sum \left\{ C_{peptide}[x-y]_t \mid i = x - 1 \cup i = y \right\}$$

$C_{i,t}$ [µM] is the concentration of cleavage site products formed at each time point t. i equals the sum of the concentrations of all peptides of sequence [x-y], for which i=(x-1) or i=y. As another check for the quality of the peptide analysis, the $C_{i,t}$ was used to calculate the degree of hydrolysis (DH_{MS}) of each sample (equation 8),

$$(8) \quad DH_{MS}[\%] = \frac{\sum C_{i,t}}{2 \times \# \ peptide \ bonds \times C_0} \times 100 \ \%$$

where the #peptide bonds represents the number of peptide bonds. For each CS, the apparent hydrolysis rate constant ($k_{i,app}$ [s⁻¹]) was then calculated by fitting equation 9 (1st order reaction equation) and equation 10 (2nd order reaction equation) to the experimental data,

$$(9) \quad C_{i,t}^f[\mu M] = 2 \left(C_0^f - C_0^f \times e^{-k_{i,app}^f \times t} \right)$$

$$(10) \quad C_{i,t}^s[\mu M] = 2 \left(C_0^s - C_0^s \big/ \left(1 + k_{i,app}^s \times C_0^s \times t \right) \right)$$

For 1st order reaction kinetics, the experimental C_0 values of each protein were used as C_0^f [µM] while $k_{i,app}^f$ [s⁻¹] was the fitting parameter. For 2nd order reaction kinetics, both C_0^s [µM] and $k_{i,app}^s$ [s⁻¹·µM⁻¹_protein] were fitting parameters. The superscripts (f and s) used for C_0 and $k_{i,app}$ indicate the fitting parameters obtained from using the 1st and 2nd order reaction equations, respectively. Fitting the C_0^s allowed to take into account the fact that for some CSs the $C_{i,t}$ reached a plateau at a lower value than C_0. The reaction rate constants (k_i^f and k_i^s) of selective hydrolysis were calculated using equation 11:

$$(11) \quad k_i = \frac{k_{i,app}}{m_E}$$

where m_E [mg] is the mass of the enzyme added for the hydrolysis, which was 0.8 mg in all hydrolyses in this study. For 1st order reaction kinetics, the selectivity[f] [%] towards a given CS was calculated by dividing the hydrolysis rate constant k_i^f [s^{-1}·mg$^{-1}_{enzyme}$] of each individual CS by the sum of k_i^f of all CSs (equation 12). For 2nd order reaction kinetics, the selectivity[s] (%) towards a given CS was also calculated by dividing the hydrolysis rate constant $k_i^s \times C_0^s$ (simplified as $k_{i,c}^s$ [s^{-1}·mg$^{-1}_{enzyme}$]) of each CS by the sum of $k_{i,c}^s$ of all CSs (equation 13).

$$(12) \quad selectivity^f \ [\%] = \frac{k_i^f}{\sum k_i^f} \times 100 \ \%$$

$$(13) \quad selectivity^s \ [\%] = \frac{k_i^s \times C_0^s}{\sum \left(k_i^s \times C_0^s \right)} \times 100 \ \%$$

Analysis of the binding site sequence of the cleavage sites

To analyse the role of binding site sequence on trypsin selectivity, the CSs in the two proteins were divided into three categories, i.e. high, intermediate and low selectivity sites, based on the trypsin selectivity. The Matlab method k-means clustering was used to divide the categories. The k-means clustering method is a partitioning method, which uses the squared Euclidean metric to determine the distances and the k-means++ algorithm for the cluster centre initialization. It is used to find clusters that divide data points with minimised total sum of distances.

It was previously stated that the protease has a subsite with 8 positions (S4-S4') that interact with the substrate AAs near the CS (the binding site P4-P4')[4]. The selectivity of the enzyme towards different CSs may be due to the interaction between the subsite site (S4-S4') of the enzyme and the binding site (P4-P4') of the substrate. Therefore, the molecular properties of the binding site positions (P4-P4') were analysed. The AAs on the binding site positions of CSs were evaluated based on their molecular weight, hydrophobicity and charge, using a similar strategy as Biro and co-workers[26]. The hydrophobicity scale published by Eisenberg et al. was used, with values of the hydrophobic moments ranging from the most hydrophilic (-1.8, Arg) to the most hydrophobic (0.73, Ile)[27]. For the analysis of charge at pH 8, the charge of arginine and lysine (positive) and glutamic and aspartic acids (negative) side-chain residues were used. Histidine was considered to be a neutral amino acid because only 0.9 % of histidine residues are charged at pH 8. In addition, cysteine was highlighted, because the sulfhydryl group is charged at the used pH. The sulfhydryl group of cysteine could be free in the intact protein, or could be released as a result of disulphide bridge shuffling[28]. If the sulfhydryl group of cysteine is involved in a disulphide bridge, it will not carry a charge, but it may still affect the hydrolysis due to steric hindrance for enzyme-substrate binding. To identify regularities of the binding site positions in the 3 categories (HSS, ISS and LSS), the standard errors of the molecular weight, hydrophobic moment and charge of the AA at each binding site position were analysed.

Calculating the possible number of peptides and DH value of tryptic hydrolysis

For a protein with N cleavage sites, the theoretical maximum number of possible peptides ($NPP_{max,theo}$) was calculated using equation 14 [29],

(14) $$NPP_{max,\,theo} = (N+1)(N+2)/2$$

where N is the number of CSs in the protein sequence. Based on the number of CSs for trypsin in α-LA, β-cas and β-LG (**Table 3.1**), the $NPP_{max,theo}$ of each protein is 105, 136 and 190, respectively. However, in reality, only a few CSs were efficiently used by the enzyme, referred as the high selectivity sites (HSSs). The predicted number of possible peptides based on selectivity ($NPP_{max,HSS}$) was calculated by replacing N in equation 14 by the number of HSSs (#HSS). The predicted maximum degree of hydrolysis ($DH_{max,HSS}$) was calculated using equation 15.

(15) $$DH_{max,\,HSS}\,[\%] = \frac{\#\,HSS}{\#\,peptide\,\,bonds} \times 100\,\,\%$$

The #peptide bonds and #HSS represent the number of peptide bonds and the number of high selectivity sites, respectively. The $NPP_{max,HSS}$ and $DH_{max,HSS}$ were compared with the experimental number of peptides obtained from manually annotation and the $DH_{max,exp}$, respectively.

Results and discussion

General description of the hydrolysis of α-LA and β-cas

Trypsin hydrolysed β-cas faster than α-LA, as indicated by the overall hydrolysis rate constants (k_{DH}^{hydr}), which were $9.6(\pm0.4)\cdot10^{-3}$ and $1.6(\pm0.2)\cdot10^{-3}$ s^{-1} (equation 2), respectively. Another difference was observed when plotting the percentage of remaining intact protein against DH_{stat} (**Figure 4.1B**). The amount of intact protein against DH illustrates the affinity of the enzyme towards intact proteins compared to its affinity towards intermediate peptides [30], according to the Linderstrøm-Lang theory. According to this theory, and assuming that the affinity of enzyme towards peptides is comparable for all peptides, the affinity of the enzyme towards the intact proteins can be determined from the initial slope of the curve of the amount of remaining intact protein against DH_{stat}. However, since the hydrolysis of α-LA and β-cas did not reach the same $DH_{max,fit}$, the DH_{stat} was normalised to $DH_{stat}/DH_{max,fit}$ [%]. From the initial slopes of the two curves, the intact β-cas was shown to be more susceptible to tryptic hydrolysis than the intact α-LA (**Figure 3.1B**). Already at DH 1 %, all intact β-cas molecules were hydrolysed (**Figure 3.1B**). In this case, a DH of 1 % means that on average there were ~2 peptide bonds hydrolysed in 208 peptide bonds (in β-cas). It seems unlikely that this was due to random hydrolysis of any CSs. Rather, it suggests that there were 1 or 2 peptide bond(s) that was/were much more susceptible to hydrolysis in the intact protein than the other bonds. Based on the initial slope of the curve, the enzyme affinity towards intact α-LA was low compared to β-cas. Moreover, there was ~12 % intact α-LA not hydrolysed at the end of the hydrolysis. Using pure apo- and holo α-LA (prepared

using EDTA and dialysis, or by addition of 10 mM $CaCl_2$), it was shown that both forms were easily hydrolysed by bovine trypsin (data not shown). The remaining α-LA may be due to interactions with formed peptides, since it has been reported previously that peptides in the hydrolysates could bind non-covalently to the intact proteins [31].
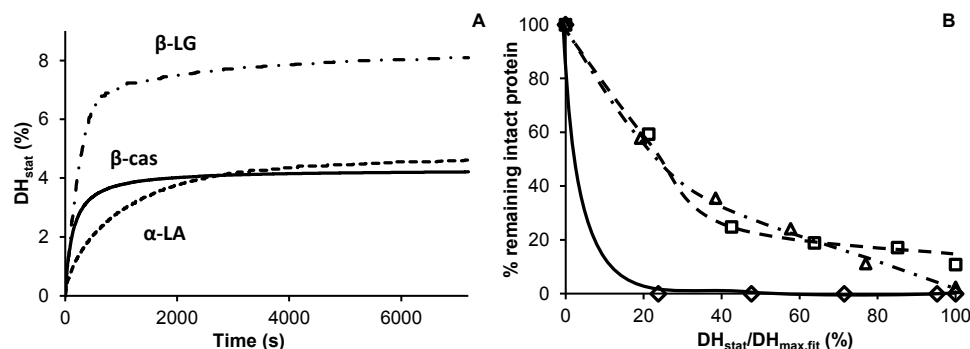

Figure 3.1: (A) Hydrolysis curves (DH_{stat} *versus* time) of (---) α-LA, (—) β-cas and (—·) β-LG, and (B) percentage of remaining intact (□) α-LA, (◊) β-cas and (△) β-LG as a function of $DH_{stat}/DH_{max,fit}$ for the hydrolysis of 1 % [w/v] protein by trypsin.

In addition to the differences in hydrolysis rate, there was also a difference in the maximal DH reached for the two proteins. For α-LA and β-cas, the experimental degree of hydrolysis ($DH_{max,exp}$) reached 4.7±0.2 % and 4.3±0.1 %, respectively (**Figure 3.1A**). The fitted $DH_{max,fit}$ values were 5.1 % and 4.3 % for α-LA and β-cas (equation 2), corresponding to ~48 % and ~60 % of the theoretical DH maximum ($DH_{max,theo}$), respectively (**Table 3.2**). Based on $DH_{max,fit}$ of α-LA and β-cas, 6 (out of 13) and 9 (out of 15) CSs, respectively, were estimated to be efficiently hydrolysed by trypsin (equation 15). The fact that the $DH_{max,fit}$ of these two proteins did not reach the $DH_{max,theo}$ agrees with the previous studies where it was shown that bovine trypsin could not hydrolyse all CSs in a protein to the full extent [1, 2].

Determination of selectivity towards cleavage sites in α-LA and β-cas
Identification and quantification of the peptides

In the tryptic hydrolysates of α-LA and β-cas, 23 and 34 unique peptides were identified and quantified, respectively (**Tables 3.2, Annexes 3.1** and **3.2**). All 23 peptides identified in α-LA hydrolysates were the result of specific cleavage by trypsin. In β-cas hydrolysates, 4 a-specific CSs, i.e. P9-10, Q38-39, H134-135 and F190-191, were cleaved to a very low extent. The resulting peptides were only identified in the final hydrolysate, which means the hydrolysis rates of these sites were extremely low. These sites were not included in further discussion about the enzyme selectivity. From the amounts of unique peptides annotated in the hydrolysates, the estimated numbers of high selectivity sites were 5 and 7 for α-LA and β-cas, respectively (equation 13). They were close to the estimations made based on the $DH_{max,fit}$ values (6 and 9 for α-LA and β-cas, respectively).
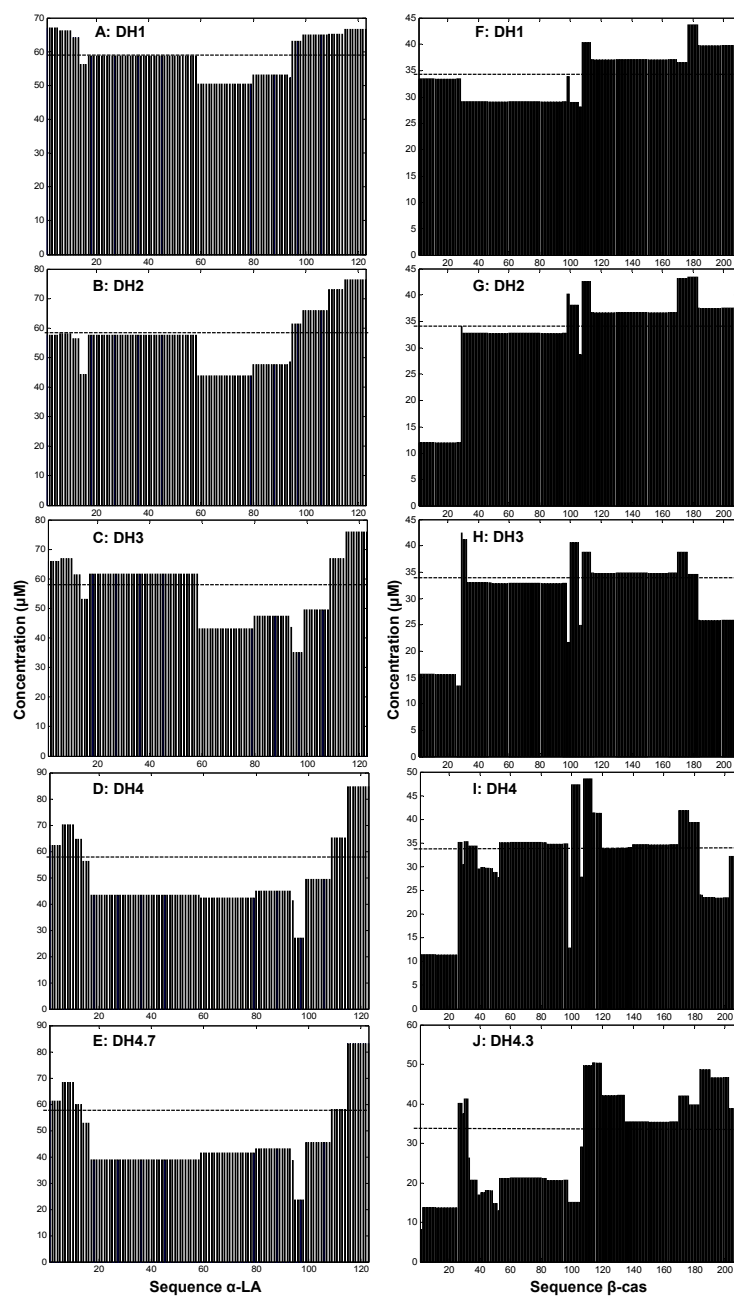
Figure 3.2: Molar sequence coverages of tryptic hydrolysates of (A-E) α-LA with a DH of 1, 2, 3, 4 and 4.7 %, and of (F-J) β-cas with a DH of 1, 2, 3, 4 and 4.3 %. The dotted line (······) indicates the initial injected protein concentration.

The AA sequence coverages (equation 3) for all hydrolysates of α-LA and β-cas were 100 %, meaning that all AA in the protein sequences were found at least once in the hydrolysates. The completeness of the peptide annotation and quantification were further verified by calculating the peptide (equation 5) and molar (equation 6) sequence coverages. On average, the peptide sequence coverages were 96±3 % and 83±10 % for α-LA and β-cas hydrolysates, respectively. The average molar sequence coverages were 80±9 % and 74±8 % for α-LA and β-cas hydrolysates, respectively (**Figure 3.2**). For β-cas, the relatively low molar sequence coverages in two parts of the sequence, i.e. [1-25] and [32-48], might be due to the fact that these peptides contain phosphoserines. Yet it has been reported that phosphorylated peptides are typically not well ionised [32, 33]. Based on data obtained from previous studies [3, 34, 35], the thresholds of acceptance for the average peptide and molar sequence coverages were 80 % and 70 %, respectively. From the average peptide and molar sequence coverages in this study (**Figure 3.2**), the quality of the peptide analysis was either better than or similar to previous studies [3, 34, 35].

Peptide release kinetics

The completeness of annotation and quantification was further illustrated by the analysis of the kinetics of peptide formation (**Annex 3.4**). Peptides that contain CSs, e.g. α-LA[1-10], were formed quickly in the start of the hydrolysis and were later hydrolysed into smaller peptides (**Figure 3.3A**). The degradation of α-LA[1-10] was quantitiatvely related to the accumulation of α-LA[1-5] and α-LA[6-10]. However, neither the concentration of α-LA[1-5] nor the concentration of α-LA[6-10] reached the initial injected concentration (C_0) of α-LA (59 μM), which indicated that cleavage site K5-6 were not cleaved completely by trypsin. The concentration of cleaved K5-6 and uncleaved remaining K5-6 in the final hydrolysate added up to obtain the C_0 of α-LA.
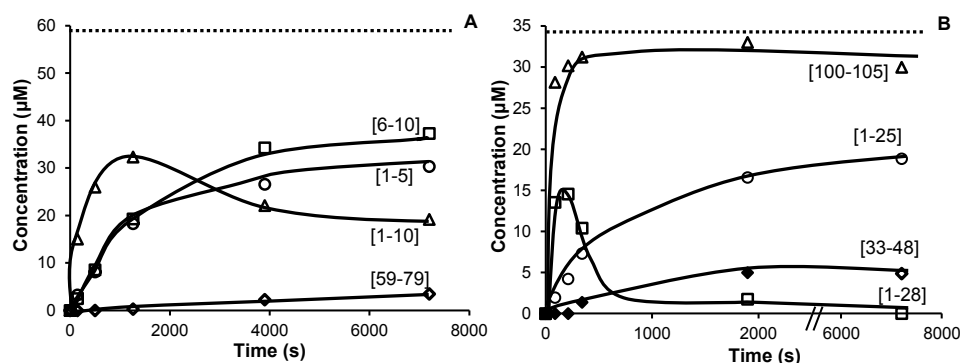
Figure 3.3: Concentrations of (A) α-LA peptides (△) [1-10], (□) [6-10], (○) [1-5] and (◇) [59-79], and (B) β-cas peptides (△) [100-105], (□) [1-28], (○) [1-25] and (◇) [33-48] as a function of time. The dotted lines (······) represent the injected concentration of each protein.

For β-cas[1-28], a fast appearance, followed by subsequent hydrolysis was observed (**Figure 3.3B**). Similar to α-LA[1-5] and α-LA[6-10], the concentration of the resulting peptide β-cas[1-25]

also reached only a plateau at 20.5 µM, which was only ~60 % of the expected injected concentration of β-cas (C_0=34 µM). However, the final peptide β-cas[100-105] formed rapidly in the start of the hydrolysis and the final concentration of β-cas[100-105] reached the C_0 of β-cas, suggesting that K99-100 and K105-106 were quickly and completely hydrolysed.

Cleavage site products formation kinetics

Based on the peptide concentrations, the sum of CS products of each CS in α-LA and β-cas were determined to obtain information on the hydrolysis of each CS (**Figures 3.4** and **Annex 3.5**). There were 2 out of 13 (K62-63 and K122-123) CSs in α-LA and 1 out of 15 (R1-2) CSs in β-cas that remained completely not hydrolysed after tryptic hydrolysis (results not shown). For the hydrolysed CSs, differences were found in the hydrolysis of individual CSs with respect to the hydrolysis rates as well as the final concentration of CS products. For example, the formation of CS products of K5-6 in α-LA was slow and the final concentration was lower than the C_0. In contrast with K5-6, the concentration of R10-11 reached the C_0. For β-cas, trypsin had a high hydrolysis rate towards K28-29 and the final concentration of the CS products also reached the C_0. However, trypsin had an extremely low hydrolysis rate towards K29-30 and it was only cleaved to very low extent. This is due to the fact that the fast cleavage of K28-29 resulted in the incompleteness of the binding site of K29-30, which led to low selectivity.
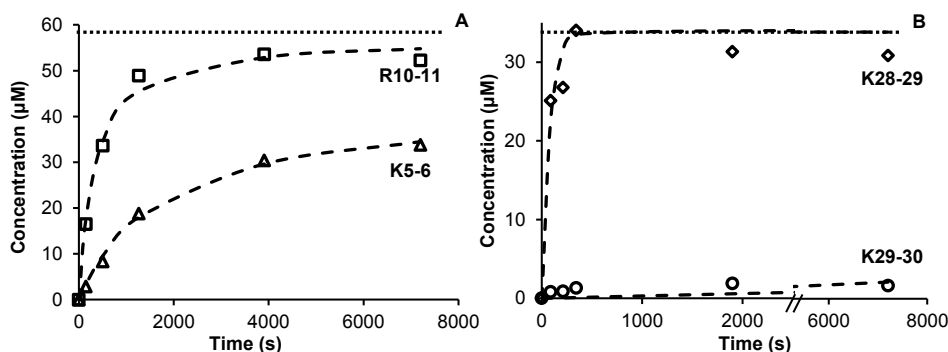
Figure 3.4: Concentrations of cleavage site products (△) K5-6 and (□) R10-11 in (A) α-LA, and (◇) K28-29 and (○) K29-30 in (B) β-cas as a function of time. The dashed lines (------) represent the fit using 2nd order reaction equation and the dotted lines (······) represent the injected concentration of each protein.

Based on the formation of CS products, the DH values determined by peptide analysis (DH$_{MS}$) of each hydrolysate were obtained (equation 8). The average standard error between the DH$_{MS}$ and the DH$_{stat}$ was ~10 % (**Figure 3.5**)**.** The overall hydrolysis rate constants k_{hydr}^{DH} were calculated from DH$_{MS}$ (using equation 2), which were 1.8($\pm$0.1)·10^{-3} and 1.1($\pm$0.1)·10^{-2} s^{-1} for α-LA and β-cas, respectively. There were ~9 % standard error between k_{hydr}^{DH} determined by the peptide analysis and by the pH-stat titration. The good agreement of the DH values and k_{hydr}^{DH} between the peptide analysis and the pH-stat data further confirmed the completeness of the peptide analysis.

Figure 3.5: Comparison of DH determined based on peptide analysis (DH$_{MS}$) and based on pH-stat titration (DH$_{stat}$), for (□) α-LA, (◊) β-cas and (△) β-LG. The dotted li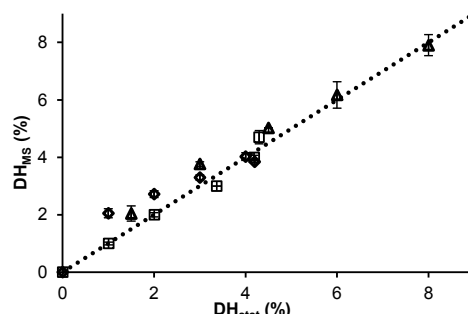ne (······) indicates the function y=x, illustrating the theoretical scenario when DH calculated using the two methods give the same results.

For the CSs that were hydrolysed, both 1st and 2nd order reaction kinetics were used to determine the hydrolysis rate constants. The R^2 for the fit of C$_{i,t}$ against time (R^2=1-SS$_{residual}$/SS$_{total}$, SS means the sum of squared errors) using 1st order reaction equation averaged over all CSs were 0.94±0.10 and 0.90±0.10 for α-LA and β-cas, respectively. For 2nd order reaction kinetics, the averaged R^2 were 0.96±0.08 and 0.90±0.14 for α-LA and β-cas, respectively. Both models were used to determine the trypsin selectivity towards CSs in α-LA and β-cas.

Trypsin selectivity

No clear differences were found between the selectivity calculated using 1st and 2nd order reaction equations (**Figure 3.6**). This shows that the way of fitting the data did not change the conclusions of the selectivity. The advantage of 2nd order reaction kinetics was that the variations in final concentration of CS products were taken into account. Therefore, the selectivity calculated using 2nd order reaction kinetics (selectivity^s) was used for further analysis. The sum of $k_{i,c}^s$ of all CSs in α-LA and β-cas were 1.1·10^{-2} and 2.1·10^{-1} s^{-1}·mg$^{-1}_{enzyme}$, respectively. The selectivity^s among CSs differed from 0-27 % for α-LA and 0-32 % for β-cas. Trypsin had an extremely high selectivity towards K169-170 in β-cas (32 %). In addition, the concentration of the CS products of K169-170 reached C$_0$ at DH 1 %, indicating that the K169-170 in all β-cas molecules were hydrolysed before the DH reached 1 %. This result agreed with the observation that no intact β-cas was found in the hydrolysate of DH 1 % (**Figure 3.1B**). Previously, it was observed that during protein enzymatic hydrolysis some CSs may only become accessible after cleavage of other bonds in the protein [36]. This phenomenon was referred to as the demasking effect. In this work, such behaviour was observed for hydrolysis of 1 out of 28 CSs, i.e. R202-203 in β-cas. The concentration of the CS products only started to increase after DH 3 %, but it reached ~30 μM after 2 h hydrolysis (C$_0$=34 μM). This shows that once it is demasked, the peptide bond is hydrolysed efficiently by trypsin.

Based on the selectivity[s], the CSs in each protein were divided into 3 categories by the Matlab k-means clustering method. From the clustering analysis of the CSs in α-LA, trypsin has high (HSS), intermediate (ISS) and low (LSS) selectivity towards 5, 4 and 4 sites, respectively. In β-cas, K169-170 was taken out of the clustering analysis but classified as HSS, since it was hydrolysed much faster than the other sites. Initially, CSs K107-108 and K183-184 were categorised by clustering analysis as ISSs. However, since they were completely hydrolysed before DH_{stat} reached $DH_{max,fit}$, they were categorised as HSS. There were 7 HSSs in β-cas, and the other 8 sites were clustered as 5 ISSs and 3 LSSs. For both proteins, the #HSS estimated from the number of unique peptides as well as from DH_{max} were similar to the experimental #HSS.
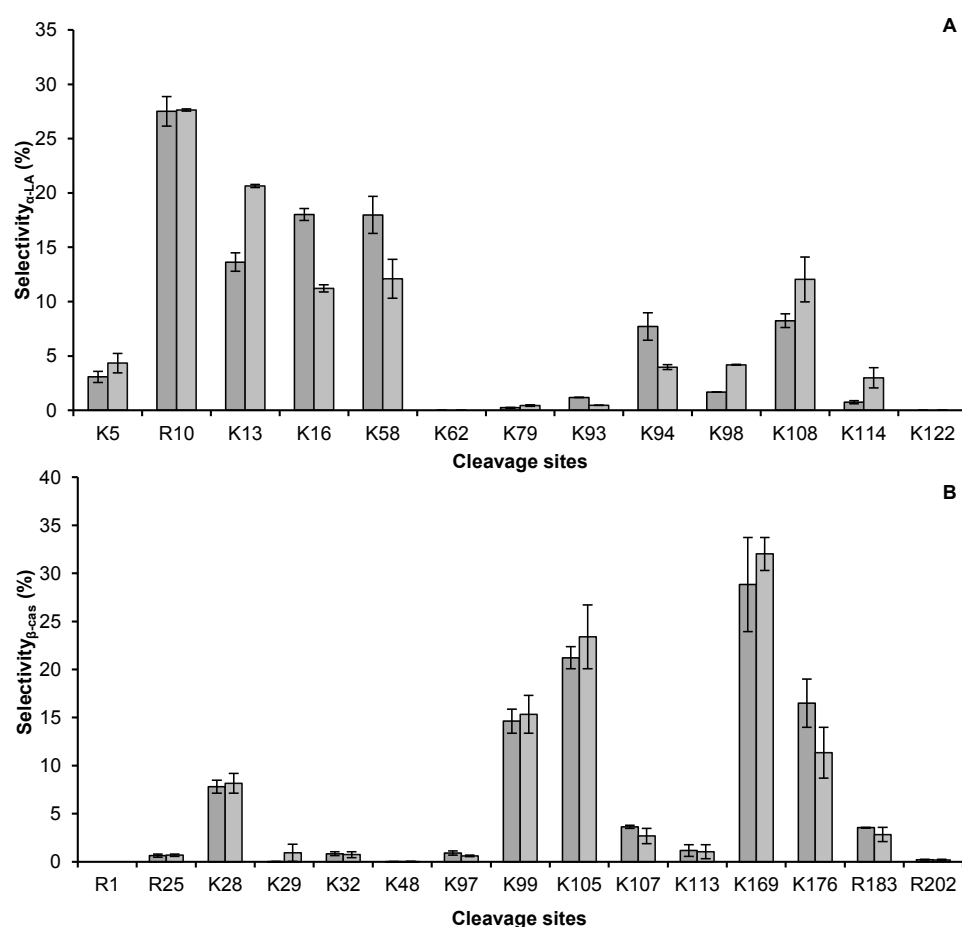

Figure 3.6: Trypsin selectivity towards cleavage sites in (A) α-LA and (B) β-cas. (■) represents selectivity[f] (selectivity calculated using 1st order reaction equation) and (■) represents selectivity[s] (selectivity calculated using 2nd order reaction equation).

Links between binding site sequence and selectivity

The selectivities of trypsin towards all CSs in the two proteins were evaluated by the missed cleavage rules of bovine trypsin proposed in the literature (**Table 3.1**). Out of the 15 binding site sequences that have been reported to cause missed cleavage of bovine trypsin towards those CSs, 11 were not applicable because the binding site sequences did not occur in the proteins used. For the applicable 4 rules, only 1 missed cleavage rule (C<u>K</u>D, the CS is underlined) was confirmed by a missed cleavage of K62-63 in α-LA. For the 3 other rules, HSSs were found, while the rules suggested that trypsin was completely inhibited towards them. In addition, 2 out of 3 sequences for moderate inhibition of porcine trypsin were present in the substrate sequences. One rule (<u>K/R</u>D) was rejected, since 2 HSSs were found in this work. The other rule (E<u>K/R</u>) was confirmed to have moderate inhibition of bovine trypsin because 1 ISS and 1 LSS were found. Lastly, it was stated that trypsin hydrolysed arginine faster than lysine in β-cas [1]. However, in the current study, trypsin only had a higher selectivity towards 1 arginine in α-LA than the lysines, but this was not the case for the 4 arginines in β-cas. It must be noted, however, that there were differences in the experimental conditions, as well as in the data analysis and interpretation between both studies. Most importantly, in the cited study, the hydrolysis rate constant of each CS was calculated only based on the release kinetics of a single peptide. Since in this work there was no clear support for the stated difference in hydrolysis of lysine *versus* arginine, it was not included in the revised set of rules. The original set of rules from the literature (**Table 3.1**) was used to predict the selectivity of trypsin towards CSs in α-LA and β-cas (**Table 3.3**). The CSs with binding site sequences that were not mentioned in literature were considered to be HSSs. This prediction was only correct for 10 (6 in α-LA and 4 in β-cas) out of 28 CSs in the two proteins. From the low correctness of the prediction based on the rules from literature, it is necessary to improve the rules to obtain better prediction.

Table 3.3: The percentage of correctly predicted cleavage sites of trypsin selectivity based on rules from literature (**Table 3.1**) and based on rules from this work (**Table 3.5**).

Protein	Prediction based on literature		% of correctly predicted CSs	Prediction based on this work		% of correctly predicted CSs
	Correct	Incorrect		Correct	Incorrect	
α-LA	6	7	46 %	10	3	77 %
β-cas	4	11	33 %	9	6	60 %
β-LG	6	12	33 %	12	6	67 %
In total	16	30	35 %	31	15	67 %

To identify rules whether there was a systematic influence of molecular properties of AAs on specific positions on the binding site, the molecular weight, hydrophobicity and charge of the AA on each binding site position were evaluated. There was no correlation observed for the molecular weight or the hydrophobic moment of AAs and the enzyme selectivity. For both properties, the standard errors were similar for the 3 categories (HSS, ISS and LSS). For molecular weight, the standard error range from 10-20 % , which is close to the standard error over all 20 protein AAs (22 %). For hydrophobicity, the absolute values of standard errors ranged from 50

and 1600 %, and the absolute standard error over all 20 protein AAs is ~420 %. By analysing the charge of the AA in the binding site sequence, it was found that of the HSS, 11 out of 12 of the binding site sequences did not have charged AA at the P2 and P2' positions (**Table 3.4** and **Figure 3.7**).

Table 3.4: Molecular properties of the binding sites (from P4 to P4') surrounding the trypsin cleavage sites in α-LA and β-cas. Amino acids with a charged side chain (KRDE) are marked with ▮ and cysteines are marked with ▮. "-" means there is no amino acid present. The ⬭ emphasized the P2 and P2' positions of the binding site.

Cleavage sites	Selectivity^s (%)	$k^s_{i,c}$ ($s^{-1} \cdot mg^{-1}_{enzyme}$)	Binding sites							
			P4	P3	P2	P1	P1'	P2'	P3'	P4'
High selectivity										
α-LA R10	27.6±0.1	3.0E-03	E	V	F	R	E	L	K	D
α-LA K13	20.6±0.1	2.3E-03	R	E	L	K	D	L	K	G
α-LA K58	12.1±1.7	1.3E-03	I	N	N	K	I	W	C	K
α-LA K108	12.0±2.0	1.3E-03	L	A	H	K	A	L	C	S
α-LA K16	11.2±0.3	1.2E-03	K	D	L	K	G	Y	G	G
β-cas K169	32.0±1.7	6.6E-02	S	Q	S	K	V	L	P	V
β-cas K105	23.4±3.3	4.8E-02	M	A	P	K	H	K	E	M
β-cas K99	15.3±1.9	3.1E-02	S	K	V	K	E	A	M	A
β-cas K176	11.4±2.6	2.4E-02	V	P	Q	K	A	V	P	Y
β-cas K28	8.2±1.0	1.7E-02	R	I	N	K	K	I	E	K
β-cas R183	2.8±0.7	6.0E-03	Y	P	Q	R	D	M	P	I
β-cas K107	2.7±0.7	5.7E-03	P	K	H	K	E	M	P	F
Intermediate selectivity										
α-LA K5	4.3±0.8	4.8E-04	Q	L	T	K	C	E	V	F
α-LA K98	4.2±0.0	4.6E-04	I	L	D	K	V	G	I	N
α-LA K94	4.0±0.2	4.4E-04	C	V	K	K	I	L	D	K
α-LA K114	3.0±0.9	3.3E-04	C	S	E	K	L	D	Q	W
β-cas K113	1.0±0.7	2.3E-03	P	F	P	K	Y	P	V	E
β-cas K29	0.9±0.8	1.8E-03	I	N	K	K	I	E	K	F
β-cas K32	0.7±0.3	1.6E-03	K	I	E	K	F	Q	S*	E
β-cas R25	0.7±0.1	1.4E-03	S	I	T	R	I	N	K	K
β-cas K97	0.6±0.1	1.3E-03	G	V	S	K	V	K	E	A
Low/zero selectivity										
α-LA K93	0.5±0.0	5.1E-05	M	C	V	K	K	I	L	D
α-LA K79	0.4±0.1	4.7E-05	S	C	D	K	F	L	D	D
α-LA K62	0±0	0	I	W	C	K	D	D	Q	N
α-LA K122	0±0	0	L	C	E	K	L	-	-	-
β-cas R202#	0.2±0.1	3.9E-04	G	P	V	R	G	P	F	P
β-cas K48	0.02±0.01	1.5E-05	L	Q	D	K	I	H	P	F
β-cas R1	0±0	0	-	-	-	R	E	L	E	E

represents masked cleavage site in the intact protein
* represents phosphoserine

It was a strong indication that trypsin can hydrolyse such CSs unhindered. There were 5 HSSs with glutamic (E) or aspartic acids (D) on P1' position. They were reported in the literature as

complete missed cleavage as mentioned in **Table 3.1** [6, 7]. This means that charged AA in P1' position do not hinder tryptic hydrolysis. For ISSs, there were 2 out of 9 CSs with no charged AA both at the P2 and P2' positions, but 5 CSs had one charged AA at either the P2 or P2' position. Therefore, we proposed that trypsin was moderately inhibited towards CSs with one charged AA at P2 or P2' positions. For LSSs, it was clear that trypsin cannot hydrolyse CSs when the P2 to P2' positions of a CS are the C- or N-terminal AA of a protein (K122-123 in α-LA and R1-2 in β-cas). This observation also partially explained why the concentration of the CS products for certain CSs did not reach the C_0, e.g. in β-cas, K28-29 was efficiently hydrolysed while K29-30 was barely hydrolysed (**Figure 3.4**). For other positions in the binding site sequence (P4, P3, P3' or P4'), no rules were found by correlating the charge of the AA with the selectivity.
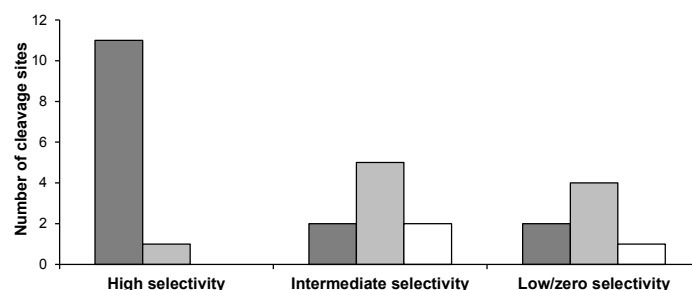

Figure 3.7: Numbers of cleavage sites with 0 (■), 1 (■) and 2 (□) charged amino acids on the P2 and P2' positions of the binding site sequences of cleavage sites catogorised as high, intermediate and low/zero selectivity sites in α-LA and β-cas, based on **Table 3.4**.

Using the revised and extended rules (**Table 3.5**), the trypsin selectivity towards 77 % and 60 % of the CSs in α-LA and β-cas were predicted correctly, respectively, whereas the original rules from literature had 46 % and 33 % correctness (**Table 3.3**). The ~30 % improvement of the prediction by the revised set of rules indicated that these rules can better describe the tryptic protein hydrolysis than the previous rules. Therefore, the revised set of rules was used to predict trypsin selectivity towards CSs in β-LG.

Table 3.5: The revised rules for trypsin selectivity towards cleavage sites in a protein. ■ indicates charged amino acids (KRDE) and ■ indicates neutral amino acids (All others). Cysteine is marked with ■. The letters in the columns represent the abbreviation of the amino acids. "-" means there is no amino acid present.

Category	Binding site			
	P2	P1	P1'	P2'
High selectivity	All others	K/R		All others
Intermediate selectivity	KRDE	K/R		All others
	All others	K/R		KRDE
	KRDE	K/R		KRDE
Low/zero selectivity	-	K/R	-	-
		K/R	P	
	C	K/R		

Prediction of tryptic hydrolysis of β-LG

From literature, proline was shown to be an inhibiting factor for trypsin when it was at the P1′ position of the binding site sequence [6, 15, 16]. Although it was not applicable for CSs in α-LA and β-cas, it was verified as a valid rule by the missed cleavage of K47 in β-LG (**Table 3.6**). Based on the revised rules, it was predicted that out of 18 CSs in β-LG there would be 10 HSSs, 5 ISSs and 3 LSSs (**Table 3.6**). Using the number of HSS and the rules revised in this paper for trypsin selectivity towards CSs, the release of 66 possible peptide (NPP$_{max,HSS}$) was expected (equation 14) and the predicted maximum degree of hydrolysis (DH$_{max,HSS}$) was 6.2 % (equation 15).

The maximal DH determined experimentally (DH$_{max,exp}$) of β-LG tryptic hydrolysis reached 7.9±0.1 % (**Figure 3.1A**). The fitted DH$_{max,fit}$ was 8.3 %, which was ~75 % of the DH$_{max,theo}$ (**Table 3.2**), and only fitted DH$_{max,fit}$ was ~2 % higher than DH$_{max,pre}$. This small difference was attributed to the small contribution of ISSs to DH$_{max}$. The overall hydrolysis rate k_{hydr}^{DH} was 6.5(±0.2)·10^{-3} s^{-1}. The decrease of intact β-LG as a function of DH$_{stat}$/DH$_{max}$ showed that intact β-LG was as susceptible to tryptic hydrolysis as intact α-LA (**Figure 3.1B**). In total, 51 unique peptides were generated from β-LG hydrolysed by trypsin (**Table 3.2** and **Annex 3.3**), which was much closer to the NPP$_{max,HSS}$ than the theoretical number of possible peptides (NPP$_{max,theo}$) based on the trypsin specificity (190). The AA sequence coverages were 100 % for all hydrolysates. The average peptide and molar sequence coverages were 93±5 % and 84±2 %, respectively, which were both within acceptable ranges and similar to the values obtained for α-LA and β-cas hydrolysates. The DH$_{MS}$ calculated from the peptide analysis and DH$_{stat}$ obtained from pH-stat titration had an average standard error of 9 % (**Figure 3.5**). The k_{hydr}^{DH} was 6.6(±0.1)·10^{-3} s^{-1}, and the standard error between the k_{hydr}^{DH} calculated from the pH-stat titration and the peptide analysis was ~1 %. All these evaluations confirmed that the peptide analysis of β-LG hydrolysates were of similar quality as the analyses of α-LA and β-cas, which allows the correct determination of trypsin selectivity towards CSs in β-LG.

After determining trypsin selectivity towards CSs in β-LG, 7, 6, and 5 (out of 18) CSs were categorised as HSSs, ISSs and LSSs, respectively (**Table 3.6**). In general, the selectivity of each CS in β-LG was in agreement with previously reported data from Fernández and Riera [2] (**Table 3.6**). Due to the different models to fit the CS products formation kinetics, the absolute values of hydrolysis rate constants of each CS were in different scales. The trypsin selectivity towards CSs in β-LG in this work was compared to the selectivity derived from previous results by Fernández and Riera [2]. For 11 out of 18 (61 %) CSs, the determined selectivity from the two studies were similar [2]. The reason why trypsin selectivity towards the rest 7 CSs were different between the two studies, might be that different substrate concentrations (10 g/L in this study and 45 g/L in the study of Fernández and Riera) and enzyme to substrate ratios (1:100 [w/w] in this study and 1:650 [w/w] in the study of Fernández and Riera) were used [2].

Table 3.6: Comparison of experimental and predicted selectivity of trypsin towards cleavage sites in β-LG. The experimental (from this study and from Fernández and Riera [2]) and predicted (based on **Table 3.5**) high, intermediate and low/zero selectivity categories are separated by ▪, ▪ and ▪, respectively. The charged amino acids (KRDE) are marked with ▪ and cysteine is marked with ▪. The ▭ emphasized the P2 and P2' positions of the binding site.

Cleavage sites	Experimental selectivity^s [%]	$k_{i,c}^s$ [s^{-1}·mg$^{-1}_{enzyme}$]	Binding site								Prediction based on Table 3.5	Experimental data based on REF2
			P4	P3	P2	P1	P1'	P2'	P3'	P4'		
K8	13.7±2.0	1.01E-02	Q	T	M	K	G	L	D	I	High	High
R148	11.0±1.2	8.13E-03	M	H	I	R	L	S	F	N	High	High
K69	10.1±0.4	7.50E-03	C	A	Q	K	K	I	I	A	High	Low
R40	9.9±0.2	7.31E-03	A	P	L	R	V	Y	V	E	High	High
K141	9.4±0.9	6.95E-03	K	A	L	K	A	L	P	M	High	High
K75	9.1±1.8	6.72E-03	I	A	E	K	T	K	Y	K	Low	High
K14	7.4±1.5	5.49E-03	D	I	Q	K	V	A	G	T	High	High
K70	6.1±0.9	4.54E-03	A	Q	K	K	I	I	A	E	Intermediate	High
K138	5.3±0.6	3.73E-03	K	F	D	K	A	L	P	M	Intermediate	High
R124	5.0±0.4	3.69E-03	C	L	V	R	T	P	E	V	High	Intermediate
K91	3.8±1.3	2.80E-03	N	E	N	K	V	L	V	L	High	Intermediate
K101	3.6±0.9	2.64E-03	D	Y	K	K	Y	L	L	F	Intermediate	Intermediate
K83	2.9±0.8	2.16E-03	A	V	F	K	I	D	A	L	Intermediate	Intermediate
K135	1.6±0.00	1.15E-03	A	L	E	K	F	D	K	A	Low	Low
K77	1.2±0.7	9.01E-04	E	K	T	K	I	P	A	V	High	High
K60	0.2±0.0	1.44E-04	L	L	Q	K	W	E	N	G	Intermediate	Intermediate
K100	0.1±0.1	1.09E-04	T	D	Y	K	K	Y	L	L	High	Intermediate
K47	0±0	0	E	E	L	K	P	T	P	E	Low	Low

The experimental selectivity of each CS was compared with the predicted selectivity. Out of the 7 experimental HSSs, 6 sites only contained neutral AA at the P2 and P2' positions on the binding site sequence. This confirmed that this revised rule was valid. There were 4 out of 6 ISSs and 2 out of 5 LSSs predicted correctly. In total, 12 out of 18 (67 %) CSs were correctly predicted using the revised rules, which was significantly higher than the prediction based on the missed cleavage rules from literature (33 %) (**Tables 3.3** and **3.6**). There were 3 (out of 6) incorrectly predicted CSs (HSS-K75, LSS-K77, LSS-K100) that have another lysine on the binding site sequence. The cleavage of the quickly hydrolysed CS compromised the completeness of the binding site of the other CS, which led to the incorrect prediction. On average, using the revised rules, ~67 % of the CSs were correctly predicted for all three proteins, meaning that the dominant factor of trypsin selectivity in a protein is the primary structure. The average $k_{i,c}^s$ of the HSSs in β-cas was higher than the average $k_{i,c}^s$ of the HSSs in other two proteins. This observation may be due to the differences in protein secondary, tertiary and quaternary structure among the proteins. Since α-LA is in a molten globule and β-cas is a random coil protein, it is difficult to evaluate whether the changes in tertiary structure result in different selectivity of trypsin towards CSs. This indicated that protein secondary structure might play a role on the hydrolysis rate differences of the HSSs.

To obtain better prediction of the maximum degree of hydrolysis based on the primary sequence of the protein, the $k_{i,c}^s$ of the HSSs, ISSs and LSSs were further analysed and compared. Within one protein, the average hydrolysis rate constants $k_{i,c}^s$ of HSSs, ISSs and LSSs were significantly different from each other. Over the three proteins, the average $k_{i,c}^s$ of ISSs and LSSs were ~24 % and ~3 % of the average $k_{i,c}^s$ of HSSs, respectively. Therefore, the contributions to the $DH_{max,exp}$ by ISSs and LSSs were also included in the prediction of DH_{max}, where equation 15 was revised to equation 16.

$$(16) \quad DH_{max,\,pre}\,[\%]= \frac{\# HSS\ +\ 0.24\ \times \# ISS\ +\ 0.03\ \times \# LSS}{\# peptide\ \ bonds}\times 100\ \ \%$$

Using equation 16, the predicted $DH_{max,pre}$ values of α-LA, β-cas and β-LG were 5.7 %, and 4.8 % and 7.0 %, respectively. The $DH_{max,theo}$ calculated based on the enzyme specificity had an average standard error ($|DH_{max,pre/theo}-DH_{max,exp}|/DH_{max,exp}\times 100$ %) with the experimental values of ~70 %. The $DH_{max,pre}$ predicted based on the enzyme selectivity solely had average standard error with the experimental values of ~13 %. This 5-time decrease in the standard error means that using the estimated number of HSSs, LSSs and ISSs, the prediction of $DH_{max,exp}$ has been improved.

By extending the knowledge about protein enzymatic hydrolysis from the enzyme specificity to the enzyme selectivity, the estimation of $DH_{max,exp}$ becomes feasible. In addition, more insight is obtained about changes in hydrolysate composition during hydrolysis. This will help the food and agricultural industry to understand and estimate the *in vitro* protein digestion and the production of commercial peptides. A possible application is the production of inhibitory peptides for angiotensin-I converting enzyme. Such studies typically compare the inhibitory activity of hydrolysates formed during hydrolysis of different proteins by different enzymes [37-40]. Understanding the enzyme selectivity towards CSs in these proteins will allow a better understanding of results from such studies.

Conclusion

In this work, the selectivity of bovine trypsin towards cleavage sites in 3 proteins was determined from the relative hydrolysis rate constant of each cleavage site. A correlation between molecular properties of binding sites surrounding the cleavage sites and trypsin selectivity towards these cleavage sites was found. This correlation was used to propose a set of rules for bovine trypsin selectivity, based on the results from the hydrolysis of α-lactalbumin and β-casein. For bovine trypsin to hydrolyse a cleavage site efficiently, neutral amino acids are preferred on the P2 and P2' positions of the binding site of that cleavage site. When one of the two positions is a charged amino acid, ~24 % of the efficient hydrolysis rate of trypsin will be obtained. When both positions are charged amino acids, only ~3 % of the efficient hydrolysis rate is obtained. Using these rules, trypsin selectivity towards 2/3 of the cleavage sites in β-lactoglobulin (as well as α-lactalbumin and β-casein) were correctly predicted. This means that the primary structure of the protein has

the dominant influence on the bovine trypsin selectivity. Based on protein primary structure, the predicted maximum degree of hydrolysis of bovine trypsin is with ~13 % standard error of the experimental value, which is ~5 times better than the prediction based on enzyme specificity alone. The approach and the outcome of this work can be further applied to understand the *in vitro* protein digestion as well as the commercial peptide production.

Acknowledgement

References

[1] M.M. Vorob'ev, M. Dalgalarrondo, J.M. Chobert, T. Haertlé, Kinetics of β-casein hydrolysis by wild-type and engineered trypsin, Biopolymers, 54 (2000) 355-364.

[2] A. Fernández, F. Riera, β-Lactoglobulin tryptic digestion: A model approach for peptide release, Biochemical Engineering Journal, 70 (2013) 88-96.

[3] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Introducing enzyme selectivity: A quantitative parameter to describe enzymatic protein hydrolysis, Analytical and Bioanalytical Chemistry, 406 (2014) 5827-5841.

[4] I. Schechter, A. Berger, On the size of the active site in proteases. I. Papain, Biochemical and Biophysical Research Communications, 27 (1967) 157-162.

[5] J. Tauzin, L. Miclo, S. Roth, D. Mollé, J.L. Gaillard, Tryptic hydrolysis of bovine $α_{S2}$-casein: Identification and release kinetics of peptides, International Dairy Journal, 13 (2003) 15-27.

[6] B. Keil, Specificity of proteolysis, Springer-Verlag Berlin Heidelberg, New York, USA, 1992, pp. 66-69.

[7] E. Vandermarliere, M. Mueller, L. Martens, Getting intimate with trypsin, the leading protease in proteomics, Mass Spectrometry Reviews, 32 (2013) 453-465.

[8] T. Šlechtová, M. Gilar, K. Kalíková, E. Tesařová, Insight into trypsin miscleavage: comparison of kinetic constants of problematic peptide sequences, Analytical Chemistry, 87 (2015) 7636-7643.

[9] S.R. Marana, A.R. Lopes, L. Juliano, M.A. Juliano, C. Ferreira, W.R. Terra, Subsites of trypsin active site favor catalysis or substrate binding, Biochemical and Biophysical Research Communications, 290 (2002) 494-497.

[10] A.G. Mikhailova, V.V. Likhareva, B.V. Vaskovsky, S.K. Garanin, L.V. Onoprienko, I.A. Prudchenko, L.D. Chikin, L.D. Rumsh, Study of secondary specificity of enteropeptidase in comparison with trypsin, Biochemistry (Moscow), 69 (2004) 909-917.

[11] P.D. Gershon, Cleaved and missed sites for trypsin, Lys-C, and Lys-N can be predicted with high confidence on the basis of sequence context, Journal of Proteome Research, 13 (2014) 702-709.

[12] A.R. Lopes, M.A. Juliano, S.R. Marana, L. Juliano, W.R. Terra, Substrate specificity of insect trypsins and the role of their subsites in catalysis, Insect Biochemistry and Molecular Biology, 36 (2006) 130-140.

[13] C. Lawless, S.J. Hubbard, Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics, OMICS A Journal of Integrative Biology, 16 (2012) 449-456.

[14] J.A. Siepen, E.-J. Keevil, D. Knight, S.J. Hubbard, Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics, Journal of Proteome Research, 6 (2007) 399-408.

[15] J. Rodriguez, N. Gupta, R.D. Smith, P.A. Pevzner, Does trypsin cut before proline?, Journal of Proteome Research, 7 (2008) 300-305.

[16] J.V. Olsen, S.E. Ong, M. Mann, Trypsin cleaves exclusively C-terminal to arginine and lysine residues, Molecular and Cellular Proteomics, 3 (2004) 608-614.

[17] W.H. Heijnis, P.A. Wierenga, W.J.H. Van Berkel, H. Gruppen, Directing the oligomer size distribution of peroxidase-mediated cross-linked bovine α-lactalbumin, Journal of Agricultural and Food Chemistry, 58 (2010) 5692-5697.

[18] Y. Deng, P.A. Wierenga, H.A. Schols, S. Sforza, H. Gruppen, Effect of Maillard induced glycation on protein hydrolysis by lysine/arginine and non-lysine/arginine specific proteases, Food Hydrocolloids, 69 (2017) 210-219.

[19] J. Adler-Nissen, Enzymic hydrolysis of food proteins, Elsevier Applied Science Publishers, London, UK, 1986, pp. 122-123.

[20] C.I. Butré, P.A. Wierenga, H. Gruppen, Influence of water availability on the enzymatic hydrolysis of proteins, Process Biochemistry, 49 (2014) 1903-1912.

[21] B.J.H. Kuipers, H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis, Journal of Agricultural and Food Chemistry, 55 (2007) 5445-5451.

[22] G.J.M. Bruin, G. Stegeman, A.C. Van Asten, X. Xu, J.C. Kraak, H. Poppe, Optimization and evaluation of the performance of arrangements for UV detection in high-resolution separations using fused-silica capillaries, Journal of Chromatography A, 559 (1991) 163-181.

[23] J.P. Chervet, R.E.J. Van Soest, M. Ursem, Z-shaped flow cell for UV detection in capillary electrophoresis, Journal of Chromatography A, 543 (1991) 439-449.

[24] S.E. Moring, R.T. Reel, R.E.J. van Soest, Optical improvements of a Z-shaped cell for high-sensitivity UV absorbance detection in capillary electrophoresis, Analytical Chemistry, 65 (1993) 3454-3459.

[25] H.A. Kosters, P.A. Wierenga, R. de Vries, H. Gruppen, Characteristics and effects of specific peptides on heat-induced aggregation of β-lactoglobulin, Biomacromolecules, 12 (2011) 2159-2170.

[26] J.C. Biro, Amino acid size, charge, hydropathy indices and matrices for protein structure analysis, Theoretical Biology & Medical Modelling, 3 (2006) 15-15.

[27] D. Eisenberg, R.M. Weiss, T.C. Terwilliger, W. Wilcox, Hydrophobic moments and protein structure, Faraday Symposia of the Chemical Society, 17 (1982) 109-120.

[28] P.W.J.R. Caessens, W.F. Daamen, H. Gruppen, S. Visser, A.G.J. Voragen, β-Lactoglobulin hydrolysis. 2. Peptide identification, SH/SS exchange, and functional properties of hydrolysate fractions formed by the action of plasmin, 47 (1999) 2980-2990.

[29] C.I. Butré, Introducing enzyme selectivity as a quantitative parameter to describe the effects of substrate concentration on protein hydrolysis, PhD thesis Wageningen University & Research, Wageningen, the Netherlands, 2014.

[30] K. Linderstrøm-Lang, The initial phases of the enzymatic degradation of proteins, Bulletin de la Société de chimie biologique, 35 (1953) 100-116.

[31] H.A. Kosters, P.A. Wierenga, H. Gruppen, SELDI-TOF-MS as a rapid tool to study food related protein-peptide interactions, Food Hydrocolloids, 24 (2010) 667-673.

[32] Y. Gao, Y. Wang, A method to determine the ionization efficiency change of peptides caused by phosphorylation, Journal of the American Society for Mass Spectrometry, 18 (2007) 1973-1976.

[33] F.A. Solari, M. Dell'Aica, A. Sickmann, R.P. Zahedi, Why phosphoproteomics is still a challenge, Molecular BioSystems, 11 (2015) 1487-1493.

[34] C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Determination of the influence of the pH of hydrolysis on enzyme selectivity of *Bacillus licheniformis* protease towards whey protein isolate, International Dairy Journal, 44 (2015) 44-53.

[35] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Determination of the influence of substrate concentration on enzyme selectivity using whey protein isolate and *Bacillus licheniformis* protease, Journal of Agricultural and Food Chemistry, 62 (2014) 10230-10239.

[36] M.M. Vorob'ev, E.A. Paskonova, S.V. Vitt, V.M. Belikov, Kinetic description of proteolysis. Part 2. Substrate regulation of peptide bond demasking and hydrolysis. Liquid chromatography of hydrolyzates, Die Nahrung, 30 (1986) 995-1001.

[37] W.M.Y. Lo, E.C.Y. Li-Chan, Angiotensin I converting enzyme inhibitory peptides from *in vitro* pepsin-pancreatin digestion of soy protein, Journal of Agricultural and Food Chemistry, 53 (2005) 3369-3376.

[38] B.A. Murray, R.J. FitzGerald, Angiotensin converting enzyme inhibitory peptides derived from food proteins: Biochemistry, bioactivity and production, Current Pharmaceutical Design, 13 (2007) 773-791.

[39] J.O. Onuh, A.T. Girgih, R.E. Aluko, M. Aliani, Inhibitions of renin and angiotensin converting enzyme activities by enzymatic chicken skin protein hydrolysates, Food Research International, 53 (2013) 260-267.

[40] S. Rudolph, D. Lunow, S. Kaiser, T. Henle, Identification and quantification of ACE-inhibiting peptides in enzymatic hydrolysates of plant proteins, Food Chemistry, 224 (2017) 19-25.

Annex 3.1: List of peptides derived from α-LA hydrolysed by bovine trypsin. (* indicates methionine oxidation)

Peptides	Sequence	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	Charge state	b fragments identified	y fragments identified
[1-5]	EQLTK	617.3187	617.3384	-32	1	1,2,3,4	1,2,3,4
[1-10]	EQLTKCEVFR	1251.6004	1251.6281	-22	2	2	1,2,3,4,5,6,7,8,
[1-13]	EQLTKCEVFRELK	1621.8086	1621.8497	-25	2	1,2,3,4,5,6,7,8,10,12	1,2,3,4,6,7,8,9,10,11,
[1-16]	EQLTKCEVFRELKDLK	1978.082	1978.0557	13	2	1,2,3,6,7,9,10,13,14,15	1,2,3,4,5,9,10,11,12,13,14
[1-58]	EQLTKCEVFRELKDLKGYGGVSLPEWVCTTFHTSGYDTQA IVQNNDSTEYGLFQINNK	6612.9752	6613.1845	-32	4	9	3,4,5,6,8,13,16
[6-10]	(K)CEVFR	652.2826	652.3003	-27	1		1,3
[11-13]	(R)ELK	388.2226	388.2322	-25	1	1,3	1,2
[11-16]	(R)ELKDLK	744.4124	744.4381	-35	2	1,2,3,4,5	1,2,3,4,5
[14-16]	(K)DLK	374.2068	374.2165	-26	1		1,2
[17-58]	(K)GYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTEYGLFQINNK	4652.988	4653.1394	-33	3		
[17-123]	(K)GYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTEYGLFQINNKIWCKDDQNPH SSNICNISCDKFLDDDLTDDIMCVKKILDKVGINYWLAHKALCSEKLDQWLCEKL	12224	12224.81	-66	8		
[59-79]	(K)IWCKDDQNPHSSNICNISCDK	2418.9744	2419.0355	-25	3	1, 2,3,6,8,14,15,17,19,20	1,2,3,4,6,7,10,11,13,14, 15,16,17,18,19,20
[59-93]	(K)IWCKDDQNPHSSNICNISCDKFLDDDLTDDIM*CVK	4042.6134	4042.7515	-34	3	4,26	4,7,11,17,18
[59-94]	(K)IWCKDDQNPHSSNICNISCDKFLDDDLTDDIM*CVKK	4170.7155	4170.8465	-31	3	2,5,6,9,13,14,16,20,22,26,29,30	2,5,6,7,9,12,15,31,34
[80-93]	(K)FLDDDLTDDIM*CVK	4189.7192	4189.72	0	4	7	3,10,11,12
[94-98]	(K)ILDK	615.3781	615.3955	-28	1	1,2,3,4	1,2,3,4
[94-108]	(K)ILDKVGINYWLAHK	1796.9442	1797.0301	-48	2	1,4,6,8,9,11,14	2,3,4,6,7,9,12,13,14
[94-123]	(K)ILDKVGINYWLAHKALCSE KLDQWLCEKL	3556.7919	3556.8938	-29	3	1,5,6,8	2,3,4,5,6,7,8,9,10,11,12 ,13
[95-108]	(K)ILDKVGINYWLAHK	1668.8958	1668.9951	-24	2	2,3,4,5,6,7,8,9,11,12,13	2,3,4,5,6,7,8,9,10,11,12 ,13
[95-123]	(K)ILDKVGINYWLAHKALCSEKLDQWLCEKL	3428.7249	3428.7988	-22	3	3,4,5,6,7,9	1,2,3,4,5,6,7,8,9
[99-108]	(K)VGINYWLAHK	1199.603	1199.6451	-35	2		1,2,3,4,5,6,7,8,9,10,11, 12,13,14
[109-123]	(K)ALCSEKLDQWLCEKL	1777.8234	1777.8742	-29	2	2,3,4,5,6,7,8,9,10,11,12,13,14	1,2,3,4,5,6,7,8,9,10,11, 12,13,14
[115-123]	(K)LDQWLCEKL	1146.5306	1146.5743	-38	2	2,4	2,3,4,5,6,7,8

Annex 3.2: List of peptides derived from β-cas hydrolysed by bovine trypsin. (* indicates phosphoserine)

Peptides	Sequence	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	Charge state	b fragments identified	y fragments identified
[1-9]	RELEELNVP	1097.6232	1097.57	48	2		
[1-25]	RELEELNVPGEIVESLSSSEESITR	2801.2532	2801.39	-49	2	4,5,6,7,8,12	2,4,5,7,12
[1-28]	RELEELNVPGEIVES*LS*S*S*EESITRINK	3156.487	3156.61	-39	3	8,10	7,8
[1-29]	RELEELNVPGEIVES*LS*S*S*EESITRINKK	3284.5735	3284.71	-42	3	8,10	7,8
[1-97]	RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPP FLQPEVMGVSK	10915	10915.6114	-56	8		
[1-99]	RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPP FLQPEVMGVSKVK	11149	11149.63	-57	8		
[26-28]	(R)INK	373.2413	373.23	30	1		2
[29-32]	(K)IEK	516.3382	516.33	16	1		2,3
[29-97]	(K)IEKFQS*EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK	7781	7781.87	-9	6	1,3,4	
[29-99]	(K)IEKFQS*EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVK	8009	8009.17	-21	6		
[30-38]	(K)IEKFQSEEQ	1136.5778	1136.54	33	2	1	1,3
[33-48]	(K)FQS*EEQQQTEDELQDK	1980.8384	1980.85	-6	2		3,4,6,7,8
[33-97]	(K)FQS*EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK	6513.0708	6512.48	91	4		
[49-105]	(K)IHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPK	6174.2492	6174.33	-13	4		
[98-105]	(K)VKEAMAPK	872.486	872.479	8	2	6	1,2,3,4,5,6,7
[98-134]	(K)VKEAMAPKHKEMPPKYPVEPFTESQSLTLTDVENLH	4266.981	4267.13	-35	3	1,2,3,4	1,2,3,4,5
[100-105]	(K)EAMAPK	645.3217	645.316	9	1	2,	2,4,5,6,7
[100-107]	(K)EAMAPKHK	910.4834	910.47	15	2	3,4,6	1,2,4,5,7
[106-113]	(K)HKEMPPK	1012.5278	1012.52	8	2		
[106-169]	(K)HKEMPPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK	7357	7357.55	-75	6		
[106-209]	(K)HKEMPPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKAVPYPQR DMPIQAFLLYQEPVLGPVRGPFPIIV	11824	11823.96	3	8		
[108-113]	(K)EMPFPK	747.3618	747.363	-2	1		2
[108-169]	(K)EMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK	7092	7092.24	-34	5		
[114-169]	(K)YPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK	6362	6362.34	-53	4		
[170-176]	(K)VLPVPQK	779.4889	779.491	-3	1	5	2,3,4,5,6
[170-183]	(K)VLPVPQKAVPYPQR	1590.9338	1590.92	9	2	2,9	1,4,5,7,8,9,10,12
[170-209]	(K)AVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	4484.4729	4484.42	12	3	3	1,2,4,5
[177-183]	(K)AVPYPQR	829.4421	829.445	-3	1		
[177-209]	(K)AVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	3722.0016	3722.46	-123	3		
[184-190]	(R)DMPIQAF	820.3878	820.379	11	1	4,5,	3,
[184-202]	(R)DMPIQAFLLYQEPVLGPVR	2185.2002	2185.16	18	2		
[184-209]	(R)DMPIQAFLLYQEPVLGPVRGPFPIIV	2908.5834	2908.59	-2	3	3,4,5,7,8	2,3,4,5,6,7,8,9,10
[191-202]	(F)LLYQEPVLGPVR	1382.8226	1382.79	24	2	2,3,4,5,6,7	
[203-209]	(R)GPFPIIV	741.4525	741.44	17	1		

Annex 3.3 List of peptides derived from β-LG hydrolysed by bovine trypsin. (A indicates peptides only in variant A and B indicates only in variant B)

Peptides	Sequence	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	Charge state	b fragments identified	y fragments identified
[1-8]	LIVTQTMK	932.5597	932.537	24	1	1,2,4,6,7,8	1,2,3,4,5,6,7
[9-14]	(K)GLDIQK	672.3961	672.381	22	1	2,3,4,5	1,2,3,4
[9-20]	(K)GLDIQKVAGTWY	1349.7098	1349.7	7	2	2	
[9-40]	(K)GLDIQKVAGTWYSLAMAASDISLLDAQSAPLR	6890.6044	6890.85	-36	4		
[9-70]	(K)GLDIQKVAGTWYSLAMAASDISLLDAQSAPLRVYVEELKPTPEGDLEILLQKWENDECAQKK	10915	10915.6114	-56	8		
[15-20]	(K)VAGTWY	695.3435	695.328	22	1	2,3,4,5	2,3,4
[15-40]	(K)VAGTWYSLAMAASDISLLDAQSAPLR	2706.414	2706.37	16	2		
[21-40]	(Y)SLAMAASDISLLDAQSAPLR	2029.0856	2029.05	18	2		3,4,5,6,7,8,9,11,12,1 3,14,15,16,17,
[22-39]	(S)LAMAASDISLLDAQSAPL	1785.8932	1785.92	-15	2		
[22-40]	(S)LAMAASDISLLDAQSAPLR	1942.0622	1942.02	22	2	7,8,9,15	1,3,4,5,6,7,8,9,10,11, 3,4,5,6,7
[41-60]	(R)VYVEELKPTPEGDLEILLQK	2312.2908	2312.25	18	2	2	
[41-69]A	(R)VYVEELKPTPEGDLEILLQKWENDECAQK	3415.7466	3415.69	17	3	2,3,4,5,6,7,9,12	1,2,3,4,5,6,7,
[41-70]A	(R)VYVEELKPTPEGDLEILLQKWENDECAQKK	3543.8592	3543.78	22	3	2,3,4,5,6,7,9,12	1,2,3,4,5,6,7,
[47-55]	(L)KPTPEGDLE	984.5876	984.476	113	2		
[52-64]B	(E)GDLEILLQKWENG	1513.8046	1513.78	16	2		
[59-77]A	(L)QKWENDECAQKKIIAEKTK	2289.156	2289.18	-10	3	1,2	1,6
[61-69]B	(K)WENGECAQK	1063.445	1063.44	5	2	3	1,2,7,8
[61-70]B	(K)WENGECAQKK	1191.5578	1191.53	23	2	2,3,4,5	1,2,3,4
[70-75]	(K)KIIAEK	700.4643	700.448	23	1	1,3,4,5,6	1,2,3,4,5,
[71-75]	(K)IIAEK	572.3641	572.353	19	1	2,3,4,5	1,2,3,4
[71-83]	(K)IIAEKTKIPAVFK	1456.915	1456.9	10	2	2,6,8,11,12	2,3,5,6,8
[71-84]	(K)IIAEKTKIPAVFKI	1569.8736	1569.99	-74	2		
[71-91]	(K)IIAEKTKIPAVFKIDALNENK	2354.3896	2354.36	13	2	2,3,4,5,6,7,8,9,10, 11	1,2,3,4,5,6,7,8,9,10, 1,12,13,14
[71-92]	(K)IIAEKTKIPAVFKIDALNENKV	2453.2698	2453.43	-65	3		
[71-162]A	(K)IIAEKTKIPAVFKIDALNENKVLVLDTDYKKYLLFCMENSAEPEQSLVCQCLVRTPEVDDEALE KFDKALKALPMHIRLSFNPTQLEEQCHI	10579	10579.36	-34	5		
[76-83]	(K)TKIPAVFK	902.5718	902.559	14	2	2,3,5,6	1,2,3,4,5,6,
[76-91]	(K)TKIPAVFKIDALNENK	1800.035	1800.01	14	2	1,2,3,5,6,7,8,10,11, 12,14,15	1,2,3,4,5,6,7,8,9,10,1 1,12,13,14
[76-101]	(K)TKIPAVFKIDALNENKVLVLDTDYKK	2974.7451	2974.67	25	3	2,3,6,8,	3,5,6,7,8,9,
[76-162]	(K)TKIPAVFKIDALNENKVLVLDTDYKKYLLFCMENSAEPEQSLVCQCLVRTPEVDDEALEKFDK ALKALPMHIRLSFNPTQLEEQCHI	10025	10024.37	63	5		
[78-91]	(K)IPAVFKIDALNENK	1570.9012	1570.87	20	2	2,3,4,5,6,7,8,9,11	1,2,3,4,5,6,7,8,9,10,1 1,12,13
[81-83]	(A)VFK	392.2532	392.242	29	1		1,2
[84-91]	(K)IDALNENK	915.4768	915.466	12	2	1,2,3,5,7,8	1,2,3,4,5,6,7
[84-100]	(K)IDALNENKVLVLDTDYK	1962.0532	1962.03	12	2	2,3,4,5,8,9,10,11,1 2,13,	1,2,3,4,5,6,7,8,9,11,13, 15
[84-101]	(K)IDALNENKVLVLDTDYKK	2090.1536	2090.13	11	2	2,3,4,5,6,7,8,9,10, 11,12,13	1,2,3,4,5,6,7,8,10,

Annex 3.3: Continued.

Peptides	Sequence	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	Charge state	b fragments identified	y fragments identified
[86-100]	(D)ALNENKVLVLDTDYK	1733.9444	1733.92	14	2	2,4	3,6,7
[92-100]	(K)VLVLDTDYK	1064.5932	1064.58	12	2	2,3,	1,2,3,4,5,6,7,8
[92-101]	(K)VLVLDTDYKK	1192.6878	1192.67	15	2		
[93-102]	(V)LVLDTDYKKY	1256.6593	1256.67	-9	1		
[101-124]B	(K)KYLLFCMENSAEPEQSLACQCLVR	2774.3343	2774.29	16	3		
[102-124]A	(K)YLLFCMENSAEPEQSLACQCLVR	2646.259	2646.2	22	2		
[102-138]A	(K)YLLFCMENSAEPEQSLVCQCLVRTPEVDDEALEKFDK	4293.0519	4293.89	-17	3		
[106-124]	(F)CMENSAEPEQSLVCQCLVR	2137.975	2137.93	21	2		
[125-135]	(R)TPEVDDEALEK	1244.5904	1244.58	8	2		
[125-138]	(R)TPEVDDEALEKFDK	1634.7894	1634.77	12	2	1,2,3,4,5,6,7,9,10,11,12,13	1,2,3,4,5,6,8,9,10,11,12
[125-141]	(R)TPEVDDEALEKFDKALK	1947.0162	1946.98	19	2	1,2,3,4,5,6,7,8,9,11,13	1,2,3,4,5,6,7,9,10,11,12,13
[136-138]	(K)FDK	408.2093	408.201	20	1	2,3	1,2
[139-141]	(K)ALK	330.2348	330.227	24	1		1
[139-152]	(K)ALKALPMHIRLSFN	1609.874	1609.91	-22	2		
[142-148]	(K)ALPMHIR	836.4875	836.469	22	1		
[146-148]	(M)HIR	424.2578	424.255	7	1	1,2	1,2
[149-162]	(R)LSFNPTQLEEQCHI	1657.8088	1657.78	17	2	3,4,5,6,	2,3,4,5,

Annex 3.4: List of peptide concentrations at all DH values in the 3 proteins.

α-LA peptides	Concentration (μM)				
	DH 1	DH 2	DH 3	DH 4	DH 4.7
[1-5]	3.3	8.1	18.3	26.6	30.3
[1-10]	15.0	26.0	32.4	22.1	19.1
[1-13]	6.1	4.5	2.8	2.0	1.2
[1-16]	1.7	1.0	0.3	-	-
[1-58]	1.6	0.8	-	-	-
[6-10]	2.5	8.5	19.3	34.3	37.3
[11-13]	9.9	25.3	39.0	47.1	45.0
[11-16]	5.6	7.5	7.2	3.7	3.2
[14-16]	8.1	17.8	33.3	40.8	39.0
[17-58]	8.3	13.5	29.9	28.6	27.8
[17-123]	9.6	17.5	10.1	0.8	0.4
[59-79]	-	-	0.3	2.3	3.4
[59-93]	0.4	2.2	5.9	9.2	6.4
[59-94]	1.2	4.8	11.4	16.1	15.6
[80-93]	2.6	3.8	4.7	5.0	5.1
[94-98]	-	4.6	8.0	10.4	10.9
[94-108]	0.3	0.6	0.6	0.2	0.1
[94-123]	2.0	1.7	-	-	-
[95-108]	1.0	2.6	3.4	1.9	1.3
[95-123]	10.9	10.3	-	-	-
[99-108]	1.8	9.2	22.3	32.8	32.7
[109-123]	3.4	18.9	38.6	50.5	46.8
[115-123]	1.4	3.3	9.1	19.5	25.0

β-cas peptides	Concentration (μM)				
	DH 1	DH 2	DH 3	DH 4	DH 4.3
[1-9]	-	-	-	-	0.5
[1-25]	-	-	2.3	9.6	7.6
[1-28]	4.4	9.4	11.4	1.9	0.6
[1-29]	-	1.3	1.3	-	-
[1-97]	1.6	1.3	0.8	-	-
[1-99]	27.5	-	-	-	-
[26-28]	-	-	-	33.3	33.9
[29-32]	-	-	9.1	30.4	31.8
[29-97]	-	-	18.9	-	-
[29-99]	-	31.5	12.4	-	-
[30-38]	-	-	-	4.9	3.5
[33-48]	-	-	1.0	2.7	3.2
[33-97]	-	-	-	27.5	12.4
[49-105]	-	-	0.7	-	-
[98-105]	6.4	8.7	8.5	5.4	-
[98-134]	-	-	-	-	6.7
[100-105]	19.6	28.5	31.3	34.4	-
[100-107]	2.9	0.8	-	-	-
[106-113]	2.1	4.0	9.1	20.4	13.9
[106-169]	19.5	18.5	18.0	-	-
[106-209]	3.6	5.4	-	-	-
[108-113]	1.1	1.9	4.9	20.7	31.4
[108-169]	14.0	12.8	12.5	-	-
[114-169]	-	-	14.0	34.0	35.4
[170-176]	17.7	29.0	38.7	41.1	42.0
[170-183]	2.6	2.5	-	-	-
[170-209]	12.6	6.2	-	-	-
[177-183]	4.9	12.6	30.6	38.7	39.7
[177-209]	20.0	16.8	4.0	-	-
[184-190]	-	-	-	-	5.3
[184-202]	-	-	-	-	44.7
[184-209]	3.6	9.2	21.9	23.4	10.2
[191-202]	-	-	-	-	6.2
[203-209]	-	-	-	8.8	28.6

β-LG peptides	Concentration (μM)				
	DH 1.5	DH 3	DH 4.5	DH 6	DH 7.9
[1-8]	24.4	36.0	43.7	49.6	52.0
[9-14]	16.3	29.6	40.0	48.5	54.9
[9-20]	-	-	-	-	0.3
[9-40]	-	16.3	9.9	-	-
[9-70]	9.6	-	-	-	-
[15-20]	-	0.7	1.4	2.8	30.3
[15-40]	15.4	17.2	19.3	31.5	14.1
]21-40]	-	-	-	3.7	37.3
[22-39]	-	-	0.5	0.8	1.9
[22-40]	-	-	-	0.3	0.4
[41-60]	-	-	1.5	4.4	30.6
[41-69]A	11.4	19.0	14.9	14.9	8.1
[41-70]A	9.4	19.6	24.2	25.7	12.0
[47-55]	-	-	-	-	0.5
[52-64]B	-	-	-	-	0.3
[59-77]A	-	-	-	0.2	-
[61-69]B	-	-	-	1.0	10.5
[61-70]B	-	0.6	1.2	2.0	15.9
[70-75]	3.7	4.6	8.9	13.8	18.1
[71-75]	14.1	24.3	34.3	42.8	50.9
[71-83]	0.5	0.5	0.4	-	-
[71-84]	-	-	-	0.3	-
[71-91]	0.8	2.0	1.8	-	-
[71-92]	-	-	0.6	0.7	-
[71-162]A	2.0	-	-	-	-
[76-83]	5.2	12.9	25.5	47.3	64.8
[76-91]	2.0	5.0	5.6	4.0	0.8
[76-101]	0.8	1.4	1.8	-	-
[76-162]	2.1	-	-	-	-
[78-91]	-	-	-	2.1	-
[81-83]	-	-	-	-	1.2
[84-91]	2.4	6.1	13.1	28.0	48.2
[84-100]	-	0.9	1.0	1.1	-
[84-101]	0.9	2.0	2.6	1.7	1.0
[86-100]	-	-	0.6	0.5	-
[92-100]	1.4	3.0	5.5	10.0	30.5
[92-101]	3.0	8.0	15.5	26.8	16.6
[93-102]	-	-	-	0.6	0.7
[101-124]B	-	-	2.7	2.1	2.1
[102-124]A	-	16.8	11.4	35.4	31.3
[102-138]A	8.1	12.9	17.0	-	-
[106-124]	-	-	-	0.7	0.8
[125-135]	-	-	0.7	1.7	8.7
[125-138]	2.4	8.4	17.2	29.8	35.3
[125-141]	1.5	2.6	2.7	1.7	0.4
[136-138]	-	-	1.3	1.9	9.1
[139-141]	8.3	18.2	27.0	38.4	46.2
[139-152]	-	-	-	-	6.2
[142-148]	20.9	35.5	45.7	53.3	50.6
[146-148]	-	-	-	-	6.3
[149-162]	22.1	34.3	42.5	49.3	54.8

"-" indicates the peptide was not found in the hydrolysate.

* The average standard error of the concentration of the peptides was ~15 %.

Annex 3.5: Concentrations of cleavage site products in (A) α-LA and (B) β-cas as a function of time. The dashed lines (------) represent the fit using 2nd order reaction equation.
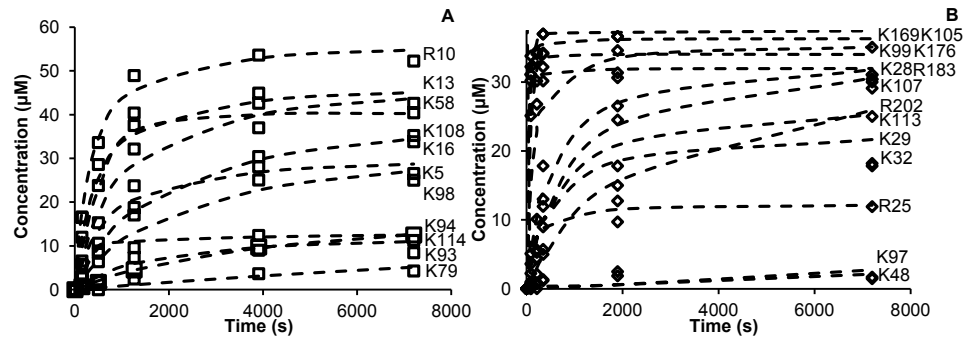
CHAPTER 4

Comparison of protein hydrolysis catalysed by bovine, porcine or human trypsins

Bovine and porcine trypsins specifically hydrolyse peptide bonds on the carboxylic side of lysine and arginine. Knowing this specificity and the protein amino acid sequence allows to estimate the theoretical maximum degree of hydrolysis ($DH_{max,theo}$) that can be reached. The $DH_{max,theo}$ is the same for enzymes with the same specificity. However, for α-lactalbumin hydrolysis by porcine trypsin, the experimental maximum degree of hydrolysis ($DH_{max,exp}$) was similar to the $DH_{max,theo}$, while by bovine trypsin, the $DH_{max,exp}$ was much lower. The low $DH_{max,exp}$ of bovine trypsin was explained by its secondary specificity, resulting in different relative hydrolysis rate constants (selectivity) towards cleavage sites (CSs) in a protein. This study aims to reveal if the differences in $DH_{max,exp}$ of hydrolysis by bovine, porcine or human trypsins can be explained based on their enzyme secondary specificity (selectivity). For both α-lactalbumin and β-casein, the differences between the $DH_{max,exp}$ of porcine tryptic hydrolysis and $DH_{max,theo}$ were small, whereas large differences between the $DH_{max,exp}$ of bovine or human tryptic hydrolysis and $DH_{max,theo}$ were found. This was explained by the fact that porcine trypsin efficiently hydrolysed ~78 % of the CSs, while bovine and human trypsins only efficiently hydrolysed ~47 % and ~53 % of the CSs, respectively. The prediction of DH_{max} based on enzyme selectivity was 4 times closer to the $DH_{max,exp}$ than the estimates based on enzyme specificity ($DH_{max,theo}$). Preliminary relations between the binding site sequence and the selectivity of each trypsin were proposed, which would allow the estimation of $DH_{max,exp}$ of tryptic hydrolysis of other substrate proteins.

Introduction

The *in vivo* protein digestion in the intestines is often simulated by the *in vitro* hydrolysis using pancreatin, in which the main proteases are trypsin and chymotrypsin. Trypsins specifically hydrolyse peptide bonds on the carboxylic side of lysine and arginine. In principle, when the specificity of the enzyme is known, as well as the amino acid (AA) sequence of the protein, the theoretical maximum degree of hydrolysis ($DH_{max,theo}$) can be calculated. In the case of tryptic hydrolysis of α-lactalbumin, the $DH_{max,theo}$ is 10.7 % (**Table 4.2**). When this protein was hydrolysed by porcine trypsin, the experimental maximum degree of hydrolysis ($DH_{max,exp}$) (~10.7 %) was indeed similar to $DH_{max,theo}$ [1]. Using bovine trypsin, surprisingly, the $DH_{max,exp}$ was only ~3.6 % [1]. Peptide analysis showed that the lower $DH_{max,exp}$ reached by bovine trypsin was because 8 out of 14 cleavage sites (CSs) on the protein were hydrolysed by the enzyme only at a low rate (**Chapter 3**). These low hydrolysis rate constants were proposed to be due to the hindrance of bovine trypsin by the charged AAs surrounding the CSs [2] (**Table 4.1**). The differences in hydrolysis rate constants of CSs in a protein are quantitatively described by enzyme "*selectivity*" [3]. The current study aims to investigate the hydrolysis by bovine, porcine and human trypsins, and to study if the differences in $DH_{max,exp}$ could be explained based on the enzyme selectivity. A detailed understanding of the differences in enzyme selectivity of the trypsins will allow a better understanding of the reported *in vitro* digestibility of a protein.

In commercial trypsins extracted from bovine and porcine pancreas, two variants of trypsin, i.e. cationic (trypsin-1) and anionic (trypsin-2) trypsins, are present. The trypsin-1 is the dominant variant. Human trypsin consists of three variants of trypsin. The content of trypsin-1 (cationic), trypsin-2 (anionic) and trypsin-3 (mesotrypsin) were found to be 60 %, 30 % and 10 %, respectively [4]. Unless stated otherwise, in this study, bovine and porcine trypsins are further referred to the trypsin-1, and human trypsin is further referred to the trypsin-2. Using bovine trypsin, the experimental maximum degree of hydrolysis ($DH_{max,exp}$) of α-lactalbumin, β-casein and β-lactoglobulin were reported to be 5.1 % [2], 4.3 % [2] and 8.0 % [2, 5], respectively. These values were only 48 %, 60 % and 75 % of the theoretical maximum degree of hydrolysis ($DH_{max,theo}$) of each protein. The peptide analysis of α-lactalbumin hydrolysates showed that 2 out of 13 CSs were not hydrolysed (**Chapter 3**). For the hydrolysed CSs, the hydrolysis rate constants of bovine trypsin towards the individual CSs varied from $4.7 \cdot 10^{-5}$ to $3.0 \cdot 10^{-3}$ $s^{-1} \cdot mg^{-1}_{enzyme}$, which were equivalent to selectivity values (relative hydrolysis rate constants) ranging from 0.4 to 27.6 % (**Chapter 3**). The 2 CSs that were not hydrolysed were referred to as the missed cleavages. The occurrence of missed cleavages was also shown in bovine tryptic hydrolysates of human serum albumin hydrolysates, where ~100 peptides with missed CSs were identified [6]. The occurrence of missed cleavages in bovine tryptic hydrolysis has been attributed to the presence of charged AAs surrounding the CSs [7, 8]. When a protease binds to a substrate, part of the enzyme (the subsite of the enzyme, S4-S4') interacts with part of the substrate AA sequence (the binding site of the CS, P4-P4') [9]. The P4-P4' positions are numbered from the N- to C-terminal sides of the CS, and P1 identifies the AA for which the enzyme is specific [9]. Recently, this concept was applied to

describe bovine trypsin selectivity towards CSs in α-lactalbumin, β-casein and β-lactoglobulin (**Chapter 3**). To obtain a better understanding of bovine tryptic hydrolysis, the relation between experimental selectivity and molecular properties of the binding site positions was investigated (**Chapter 3**). Based on the relation, CSs could be differentiated by the charge of AAs on the P2 and P2' positons of the binding site sequence into 3 categories (**Table 4.1**). The 3 categories were CS that were 1) hydrolysed efficiently, when neutral AAs are on the two positions (high selectivity sites); 2) hydrolysed with an intermediate hydrolysis rate, when a charged AA is on one of the two positions (intermediate selectivity sites); 3) hydrolysed slowly, when charged AAs are on both positions (low/zero selectivity sites) (**Chapter 3**). The type of binding site sequences of CSs in the 3 categories was taken as a set of rules to predict the $DH_{max,exp}$ of proteins (**Table 4.1**). Using these rules (based on selectivity), the predicted maximum degree of hydrolysis was found to be ~5 times closer to the $DH_{max,exp}$ than the $DH_{max,theo}$ (the prediction based on specificity) (**Chapter 3**).

Table 4.1: Selectivity rules for bovine trypsin [2] and the number of correctly predicted cleavage sites (CSs) of the total CSs in α-lactalbumin and β-casein for bovine, porcine and human trypsins based on the rules*. ▆ indicates charged amino acids (KRDE) and ▆ indicates neutral amino acids (All others). The letters in the columns represent the abbreviation of the amino acids. "-" means there is no amino acid present.

Category	Binding site				Number of correctly predicted CSs/total CSs		
	P2	P1	P1'	P2'	Bovine trypsin	Porcine trypsin	Human trypsin
High selectivity	All others	K/R		All others	11/12	14/21	9/14
Intermediate selectivity	KRDE	K/R		All others	5/9	0/2	2/3
	All others	K/R		KRDE			
Low/zero selectivity	KRDE	K/R		KRDE	1/7	2/5	3/11
	-	K/R	-	-	2/7		
		K/R	P				
	C	K/R			1/7		

Despite the same specificity, trypsins from different sources might not have the same selectivity. For example, in a porcine tryptic hydrolysate of human serum albumin, the number of peptides that contained missed CSs was 60 % lower than in a bovine tryptic hydrolysate [6]. Moreover, the $DH_{max,exp}$ of porcine tryptic hydrolysis of α-lactalbumin was equal to the $DH_{max,theo}$ [1]. Therefore, porcine trypsin is not expected to be largely influenced by the molecular properties of the binding site sequences. Different peptide profiles were also found between the hydrolysates of β-lactoglobulin taken after 6 h of hydrolysis by bovine or ovine trypsins [10]. These observations indicated that the reported *in vitro* protein digestibility and the hydrolysate composition of one substrate protein determined by hydrolysis using different trypsins could have various outcome.

In this study, the hydrolysis by porcine or human trypsins of individual CSs in two proteins (α-lactalbumin and β-casein) were analysed, and compared with previous results of bovine trypsin (**Chapter 3**). The determined hydrolysis rate constants of the three trypsins of CSs in the two proteins were used to explain the differences in the $DH_{max,exp}$. The dependence of selectivity on the molecular properties of binding site sequences was proposed for each trypsin.

* To avoid repetition, the right part of the table (number of correctly predicted CSs/total CSs) was from experimental data, which will be discussed in page 92.

Materials and methods

Materials

α-Lactalbumin (α-LA) was obtained from Davisco Foods International Inc. (Le Sueur, MN, USA). Based on results from circular dichroism, ~72 % of the α-LA was in apo form and the rest was in holo form [1]. β-Casein (β-cas, C6905), bovine trypsin (BT, T1426), porcine trypsin (PT, T0303) and aprotinin (A6279) were all purchased from Sigma-Aldrich (St.Louis, MO, USA). Recombinant human trypsin-2 (further referred to as human trypsin, abbreviated as HT) was purchased from Shanghai Yaxin Biotechnology Co., Ltd (Shanghai, China). The molecular properties, including the protein content and purity, of the two proteins, i.e. α-LA and β-cas, and the three trypsins, i.e. BT, PT and HT, were listed in **Table 4.2**. To check the composition of the enzymes, all three trypsins were analysed using ultra-performance liquid chromatography coupled with mass spectrometry (UPLC-MS). There was no chymotrypsin peak found in the chromatograms of any trypsins. The activities of bovine, porcine and human trypsins were specified to be ≥10,000, 13,000-20,000 and ≥7500 benzoyl-L-arginine-ethyl-ester units/mg protein by the manufacturers, respectively. Aprotinin was used to inhibit each trypsin hydrolysis, as described below. Aprotinin present in 0.9 % sodium chloride and 0.9 % benzylalcohol solution, was determined to have a concentration of 2.3 mg/mL based on the peak area at 214 nm (A_{214}) using RP-UPLC-UV [2]. All other chemicals were of analytical grade and purchased from Sigma or Merck.

Methods

Enzymatic hydrolysis of the proteins

The enzymatic hydrolysis of proteins was performed as described previously [3]. The α-LA or β-cas solutions (1 % [w powder/v], 10 mL in Millipore) were adjusted to pH 8.0 and equilibrated for 0.5 h at 37 °C in a pH-stat device (Metrohm, Herisau, Switzerland). Trypsin solution (100 μL, 10 mg/mL in Millipore) was added to the equilibrated protein solution to reach an enzyme to substrate ratio of 1:100 [w/w]. The hydrolysis was performed at 37 °C for 3 h using 0.2 M NaOH to keep the pH constant. The degree of hydrolysis (DH_{stat}) was calculated using equation 1 [11],

$$(1) \quad DH_{stat}\ [\%] = V_b \times N_b \times \frac{1}{\alpha} \times \frac{1}{m_p} \times \frac{1}{h_{tot}} \times 100\ \%$$

where V_b [mL] is the volume of NaOH added; N_b [mol/L] is the normality of NaOH; α is the average degree of dissociation of the α-NH groups ($1/\alpha$=1.3 at 37 °C and pH 8) [12]; m_p [g] is the mass of protein in the solution (taking into account the protein content in the powder based on Dumas results); h_{tot} [mmol/g] is the total number of millimoles of peptide bonds per gram of protein substrate (**Table 4.2**).

Table 4.2: The nitrogen to protein conversion factor (N-factor), the protein content and purity, the molar weight, the number of peptide bonds per gram of protein (h_{tot}), the number of amino acids (#AA) and the molar extinction coefficient at 214 nm (ε_{214}) of the three trypsins (bovine, porcine and human trypsins) and two proteins (α-lactalbumin and β-casein) used, and the number of cleavage sites (#CS), the theoretical maximum degree of hydrolysis ($DH_{max,theo}$) and the number of unique peptides annotated in each hydrolysate.

Protein	Uniprot code.[1]	N-factor[1] [g of protein/g N]	Protein content [w/w]	Purity[2] [%]	Molar weight[3] [Da]	h_{tot}[4] [mmol/g]	#CS[1]	#AA[1]	$DH_{max,theo}$ [%]	ε_{214} [M⁻¹·cm⁻¹]	#α-LA[6] peptides	#β-cas[7] peptides
α-LA	P00711	6.25	93 %	90	14,186	8.6	13	123	10.7	300,461	-	-
β-cas	P02666	6.39	90 %	90	23,983[6]	8.7	15	209	7.2	424,047	-	-
BT	P00760	5.97	80 %	100	23,305	-	-	223	-	-	23	34
PT	P00761	5.84	86 %	100	23,475	-	-	223	-	-	30	36
HT	P07478	5.98	77 %	100	24,041	-	-	224	-	-	30	40

α-LA: α-lactalbumin; β-cas: β-casein; BT: bovine trypsin; PT: porcine trypsin; HT: human trypsin

[1] From Uniprot (http://www.uniprot.org)

[2] The protein purity is defined as the percentage of UV₂₈₀ peak area of the target protein out of total UV₂₈₀ area from RP-UPLC.

[3] The molar weight was calculated with 5 phosphoserines on the β-cas molecule as shown in RP-UPLC-MS.

[4] h_{tot} is the number of millimoles of peptide bonds per gram of protein calculated using the amino acid sequence and molar weight of the protein reported in Uniprot.

[6,7] #α-LA and β-cas peptides stand for the unique peptides identified in the hydrolysates of α-LA and β-cas, respectively.

For each hydrolysis, at set DH values, 200 µL sample was taken. The DH at which samples were taken depended on the $DH_{max,exp}$ reached after 3 h of hydrolysis. During α-LA hydrolysis, samples were taken at DH 0, 1.5, 3, 4.5, 6, 7.5, 9 and 10 % for porcine tryptic hydrolysis and at DH 0, 1, 2, 3, 4, 5 and 6.3 % for human trypsin hydrolysis. During β-cas hydrolysis, samples were taken at DH 0, 1.5, 3, 4.5, 6, 6.7 % for porcine tryptic hydrolysis and at DH 0, 1, 2, 3, 4, 5 and 5.2 % for human tryptic hydrolysis. In a previous study, the same experiments were performed using bovine trypsin (**Chapter 3**). In that study, samples were taken at DH 0, 1, 2, 3, 4 and 4.7 % for α-LA and at DH 0, 1, 2, 3, 4 and 4.3 % for β-cas. To inactivate each trypsin, 3 µL aprotinin with a concentration of 2.3 mg/mL was directly added to each sample taken during hydrolysis, resulting in a molar ratio of trypsin to aprotinin of 1:1.5. The inactivation was confirmed for each trypsin using the pH-stat. For this, 150 µL aprotinin was added to a 10 mL protein solution (10 mg/mL) immediately after trypsin was added. For each enzyme, the pH of the solution remained constant after the addition of aprotinin, proving that all enzymes were successfully inhibited (data not shown). Due to the dissolution of CO_2 in protein solutions during hydrolysis, the consumption of NaOH needs to be corrected. As a control experiment, 1 % [w powder/v] of α-LA and β-cas solutions were incubated in a pH-stat in the absence of trypsin for 3 h. The added volume of NaOH from each control experiment was subtracted from the added volume of NaOH at all time points during hydrolysis. Both 1st and 2nd order reaction equations were used to fit the experimental DH_{stat} curves. Using 2nd order reaction equation, the R^2 of the fit against time (R^2=1-$SS_{residual}/SS_{total}$, SS means the sum of squared errors) was higher than using 1st order reaction equation. Thus, the 2nd order reaction equation (equation 2) was used to fit the experimental data,

$$(2) \quad DH_{stat}[\%] = DH_{max,fit} - DH_{max,fit} /(1 + k_{hydr} \times DH_{max,fit} \times t)$$

where the maximum degree of hydrolysis ($DH_{max,fit}$) and the overall hydrolysis rate constant (k_{DH}^{hydr} [s^{-1}]=$k_{hydr} \cdot DH_{max,fit}$) were fitting parameters. All hydrolyses were performed in duplicate, and samples were taken at the same DH points during each individual experiment. All hydrolysates were analysed by adapting the method as developed previously by Butré and co-workers [3].

Estimation of the number of efficiently hydrolysed cleavage sites

Based on the fitting parameter $DH_{max,fit}$ (derived from equation 2), the percentage of CSs that were efficiently hydrolysed by the enzyme, or in other words the percentage of high selectivity sites (HSSs) of the enzyme, was estimated from equation (3) [1].

$$(3) \quad HSS_{protein}^{enzyme}(\%) = \frac{DH_{max,fit}^{enzyme,protein}}{DH_{max,theo}^{enzyme,protein}} \times 100 \ \%$$

The superscripts and subscript of the parameters refer to the enzyme-substrate combinations.

Reversed phase ultra-high performance liquid chromatography (RP-UPLC)

The hydrolysates were analysed on an H class Acquity UPLC® system (Waters, Milford, MA, USA) equipped with a BEH C18 column (1.7 µm, 2.1×100 mm, Waters), which was connected to an Acquity UPLC® PDA detector (Waters). The disulphide bridges were reduced by incubating the samples at a protein concentration of 0.5 % for 2 h with 100 mM ditiothreitol (DTT) in 50 mM Tris-HCl buffer at pH 8.0. The reduced samples were further diluted to a protein concentration of 0.1 % [w/v] and centrifuged (10 min, 14,000×g, 20 °C) before injection (4 µL).

Electron spray ionization time of flight mass spectrometry (ESI-Q-TOF-MS)

The MS and MS/MS (MSe) data of the peptides were collected with an online SYNAPT G2-Si high definition mass spectrometry (Waters) coupled to the RP-UPLC system. The MS system was calibrated with sodium iodide. Online lock mass data (angiotensin II, [M+2H]$^{2+}$: 523.7751) was acquired [2]. Based on the differences between measured and theoretical lock mass, corrections were applied on the m/z of peptides during data collection. The data were analysed manually using MassLynx software v4.1 (Waters) and UNIFI software v1.8 (Waters).

Estimation of the number of possible peptides

For a protein with N cleavage sites, the theoretical number of possible peptides (NPP$_{max,theo}$) was calculated based on enzyme specificity using equation 4 [13],

(4) $NPP_{max,theo} = (N+1)(N+2)/2$

where N is the number of CSs in the protein sequence. Experimentally, the enzyme only efficiently hydrolysed a few -high selectivity- CSs (HSSs). Therefore, a corrected number of possible peptides (NPP$_{max,HSS}$) based on enzyme selectivity was calculated by replacing N in the equation 4 by the number of HSS (#HSS) estimated using equation 3.

Peptide identification and quantification

The mass tolerance between the theoretical mass and the measured mass for the accepted annotation was set at 100 ppm. Two AA modifications, i.e. methionine oxidation (+16 Da) in α-LA and the phosphorylation (+80 Da per phosphoserine) in β-cas, were taken into account. The peptides were annotated based on the MS spectra and were confirmed by identifying the b and y fragments in the MS/MS spectra when possible. In the case of co-elution, the concentration of each of the co-eluting peptides was calculated by using the intensity of total ion count to divide the UV area at 214 nm. The concentration of each peptide was calculated using equation 5,

(5) $C_{peptide} \ [\mu M] = \dfrac{A_{214} \cdot Q}{\varepsilon_{214} \cdot l \cdot V_{inj} \cdot k_{cell}}$

where C$_{peptide}$ [µM] is the peptide concentration, A$_{214}$ [µAU·min] is the UV area at 214 nm, V$_{inj}$ [µL] is the injected volume of the sample, Q [µL·min^{-1}] is the flow rate and l [cm] is the path

length of the UV cell, which is 1 cm according to the manufacturer. ε_{214} [M^{-1}·cm^{-1}] is the molar extinction coefficient at 214 nm, which was calculated as described previously [14]. Due to the multiple reflection by the coating of the UV cell, the effective path length of the light through the cell is not the same as the length specified by the producer. This issue was also reported in earlier studies [15-17]. To correct for this effect, the cell constant of the UV detector (k_{cell}), was determined using a series of standard solutions made by β-lactoglobulin, β-cas and angiotensin II, and the approach as described elsewhere [18]. The ratio between the measured and expected A_{214} was taken as the k_{cell} value. For the UV cell used in this study, the k_{cell} was reported to be 0.78 [2]. The linear region of the peak area (A_{214}) in the UV detector ranges from 5×10^1 to 6×10^5 µAU·min [2]. Therefore, peptide quantification was performed for peptides with $A_{214} \geq 5\times10^1$ µAU·min. In all hydrolysates, the total peak area in the chromatograms was 95±6 % of the expected value, and of the total A_{214}, 97±3 % was assigned to annotated peptides.

Enzyme affinity towards intact proteins

The percentage of remaining intact protein in the hydrolysates was plotted against hydrolysis time. The initial hydrolysis rate constants of the hydrolysis of intact protein ($k_{intact}^{enzyme,protein}$ [s^{-1}]) were calculated from the initial slope of the curve (the linear slope of the first 3 points as a function of time). According to the Linderstrøm-Lang theory, the amount of intact protein as a function of DH illustrates the enzyme affinity towards intact proteins relative to the enzyme affinity towards intermediate peptides [19]. Based on this theory, the percentage of intact protein in the hydrolysates was also plotted against DH$_{stat}$ [%]. Because the hydrolyses by the three trypsins did not reach the same DH$_{max,fit}$, the DH$_{stat}$ was normalised by dividing by DH$_{max,fit}$.

Quality check of the annotation and quantification of the peptide analysis

The quality of peptide analysis was analysed using three parameters, i.e. amino acid, peptide and molar sequence coverages, which were calculated by equations 6, 7 and 8, respectively [3]. These checks were performed for all hydrolysates. For each enzyme-substrate combination, the average value for each parameter was calculated as the average of that parameter for the duplicate experiments at all DH values.

$$(6) \quad Amino \ acid \ sequence \ coverage \ [\%] = \frac{\# unique \ annotated \ AA}{\# AA_{protein}} \times 100 \ \%$$

$$(7) \quad Peptide \ sequence \ coverage \ [\%] = \frac{\# AA(annotated \ peptides)}{\# AA(annotated \ peptides) + \# AA(missing \ peptides)} \times 100 \ \%$$

$$(8) \quad Molar \ sequence \ coverage \ [\%] = \left(1 - \frac{\sqrt{\dfrac{\sum (C_n - C_0)^2}{(\# AA_{protein} - 1)}}}{C_0} \right) \times 100 \ \%$$

C_n [μM] represents the concentration of each individual AA (*n*) in the protein sequence; C_0 [μM] is the initial injected protein concentration and #AA$_{protein}$ is the number of AA in the sequence of the parental protein [3]. The injected initial concentrations (C_0) were 59 μM and 34 μM for α-LA and β-cas, respectively (calculated based on the protein content and purity listed in **Table 4.2**). In the samples at DH 0 % of α-LA and β-cas, the experimental C_0 were measured to be 58±2 and 33±1 μM, respectively, using equation 5. The missing peptides are the peptides that were not found but should be present given the annotated peptides [3]. To further check the quality of peptide analysis, for each hydrolysate, the degree of hydrolysis based on peptide analysis (DH$_{MS}$) was calculated using equation 9,

$$(9) \quad DH_{MS} \left[\% \right] = \frac{\sum C_{i,t}}{2 \times \# \ peptide \ bonds \ \times C_0} \times 100 \ \%$$

where #peptide bonds stands for the number of peptide bonds and the $C_{i,t}$ is the concentration of the cleavage site products, calculated using equation 10.

Determination of the hydrolysis rate constants

For each CS, the concentration of cleavage site products ($C_{i,t}$) that originate from the hydrolysis of that peptide bond was calculated using equation 10.

$$(10) \quad C_{i,t} \left[\mu M \right] = \sum \left\{ C_{peptide} \left[x - y \right]_t \middle| i = x - 1 \cup i = y \right\}$$

where $C_{i,t}$ [μM] refers to the concentration of CS products formed at each time point *t*. *i* equals the sum of the concentrations of all peptides of sequence [x-y], for which *i*=(x-1) or *i*=y. The apparent hydrolysis rate constant was obtained by fitting the equation of 2nd order reaction equation (equation 11) to the experimental data [2],

$$(11) \quad C_{i,t}^{enzyme,protein} \left[\mu M \right] = 2 \left(C_0^{enzyme,protein} - C_0^{enzyme,protein} \middle/ \left(1 + k_{i,app}^{enzyme,protein} \times t \times C_0^{enzyme,protein} \right) \right)$$

where both $C_0^{enzyme,protein}$ [μM] and $k_{i,app}^{enzyme,protein}$ [s^{-1}·μM^{-1}$_{protein}$] are fitting parameters, and the superscripts refer to the enzyme-substrate combination. The apparent hydrolysis rate constant $k_{i,app}^{enzyme,protein} \times C_0^{enzyme,protein}$ is referred to as $k_{i,app,c}^{enzyme,protein}$ [s^{-1}]. The hydrolysis rate constant $k_{i,c}^{enzyme,protein}$ [s^{-1}·mg^{-1}$_{protein}$] of each selective hydrolysis was calculated using equation 12,

$$(12) \quad k_{i,c}^{enzyme,protein} = \frac{k_{i,app,c}^{enzyme,protein}}{m_E}$$

where m_E [mg] is the mass of the enzyme added for the hydrolysis. The $C_{i,t}$ was also calculated based on the concentration of remaining non-hydrolysed CSs ($C_{remain,t}$) ($C_{i,t}$=C_0-$C_{remain,t}$) in the hydrolysates. Hydrolysis rate constants calculated based on the two methods were compared.

Enzyme selectivity

The selectivity [%] of enzymes towards CSs present in the two substrate proteins was calculated by dividing the hydrolysis rate constant $k_{i,c}^{enzyme,protein}$ of each CS by the sum of hydrolysis rate constants for all CSs for that enzyme-substrate combination (equation 13) [2].

$$(13) \quad Selectivity^{enzyme, protein} \left[\%\right] = \frac{k_i^{enzyme, protein} \times C_0^{enzyme, protein}}{\sum \left(k_i^{enzyme, protein} \times C_0^{enzyme, protein}\right)} \times 100 \ \%$$

In a recent work for bovine trypsin, the high (HSSs), intermediate (ISSs) and low/zero (LSSs) selectivity sites in α-LA and β-cas were clustered using the Matlab method k-means clustering [2]. As a partitioning method, the k-means clustering uses the squared Euclidean metric to determine the distances and the k-means++ algorithm for the cluster centre initialization. It is used to find clusters that divide data points with minimised total sum of distances. To compare hydrolysis rate constants of porcine and human trypsins towards CSs with those of bovine trypsin, for each substrate, the lowest hydrolysis rate constant of bovine trypsin towards CSs in each category (HSS, ISS and LSS) was set as the thresholds. Using these thresholds, the number of HSSs, ISSs and LSSs of porcine and human trypsins towards CSs in α-LA and β-cas were determined.

Prediction of maximum degree of hydrolysis

Previously, it was found that for bovine trypsin, the average $k_{i,c}^{BT,protein}$ of ISS and LSS were ~24 % and ~3 % of the average $k_{i,c}^{BT,protein}$ of HSSs, respectively [2]. It means that within the time of the hydrolysis experiment when most HSSs were hydrolysed, ~24 % and ~3 % of the ISSs and LSSs were hydrolysed, respectively. Hence, the predicted maximum degree of hydrolysis (DH$_{max,pre}$) based on enzyme selectivity can be calculated using equation 14 [2].

$$(14) \quad DH_{max, pre} \left[\%\right] = \frac{\# HSS + 0.24 \times \# ISS + 0.03 \times \# LSS}{\# peptide \ bonds} \times 100 \ \%$$

where #HSS, #ISS and #LSS stand for the number of high, intermediate and low/zero selectivity sites, respectively. Since the #HSS, #ISS and #LSS of porcine and human tryptic hydrolysis of α-LA and β-cas were obtained based on the thresholds set from bovine trypsin selectivity, their DH$_{max,pre}$ were also predicted using equation 14.

Relation between binding site sequence and enzyme selectivity towards CSs

The correlation between the binding site sequence and the hydrolysis rate constant was analysed using the information for all CSs. For this, the standard errors of the molecular weight, hydrophobic moment and charge of the AA at each binding site position were analysed, using the same approach as reported previously [2]. The scale of hydrophobic moments ranged from most hydrophilic AA (arginine, -1.8), to the most hydrophobic AA (isoleucine, 0.73) [20]. For the analysis at pH 8, arginine and lysine were positively charged, which were the CSs, and glutamic and aspartic acids were negatively charged. Histidine was not included as a charged AA because only 0.9 % of histidine was charged at pH 8. In addition, cysteine was considered as a charged

AA, because the sulfhydryl group was charged at pH 8. The sulfhydryl group of cysteine could be free in the intact proteins, or could be released as a result of disulphide bridge shuffling [21]. If the sulfhydryl group of cysteine was involved in a disulphide bridge, it would not carry a charge, but it may still affect the hydrolysis due to steric hindrance for enzyme-substrate binding.

Results and discussion

General description of α-LA and β-cas hydrolyses by the three trypsins

The hydrolysis curves (DH_{stat} *versus* time) of α-LA by bovine, porcine and human trypsins were fitted using equation 2. The overall hydrolysis rate constants ($k_{DH}^{BT/PT/HT, \ \alpha-LA}$) of hydrolysis of α-LA by bovine ($1.6(\pm0.2)\cdot10^{-3}$ s^{-1} (**Chapter 3**)) or porcine ($1.7(\pm0.0)\cdot10^{-3}$ s^{-1}) trypsins were slightly higher than by human trypsin ($1.1(\pm0.3)\cdot10^{-3}$ s^{-1}) . Larger differences were found in the $DH_{max,exp}$ of hydrolysis by the three trypsins. The $DH_{max,fit}$ values derived from equation 2 of hydrolysis by the three trypsins were compared. For porcine trypsin, the $DH_{max,fit}$ of α-LA hydrolysis (10.0 ± 0.3 %) (**Figure 4.1A**) was similar to the $DH_{max,theo}$ (10.7 %) (**Table 4.2**). This suggested that porcine trypsin efficiently cleaved ~94 % of the CSs in α-LA, suggesting that around 12 out of 13 CSs were high selectivity sites (HSSs, equation 3). The $DH_{max,fit}$ of α-LA hydrolysed by human or bovine trypsins were 6.8 ± 0.1 % and 5.1 ± 0.2 % (**Chapter 3**), respectively, which were much lower than the $DH_{max,theo}$. Based on the $DH_{max,fit}$, for bovine or human tryptic hydrolysis, only 48 % (6 out of 13) and 63 % (8 out of 13) of the CSs in α-LA, respectively, were estimated to be HSSs.

Similarly as for α-LA hydrolysis, the overall hydrolysis rate constants ($k_{DH}^{BT/PT/HT, \ \beta-cas}$) of β-cas by bovine ($9.6(\pm0.4)\cdot10^{-3}$ s^{-1} (**Chapter 3**)) or porcine ($7.1(\pm0.7)\cdot10^{-3}$ s^{-1}) trypsins were slightly higher than by human trypsin ($6.4(\pm0.2)\cdot10^{-3}$ s^{-1}). For all enzymes, the hydrolysis rate constants of β-cas hydrolysis were ~5 times higher than those of α-LA hydrolysis. The $DH_{max,fit}$ of β-cas by bovine, porcine or human trypsins were 4.3 ± 0.1 % (**Chapter 3**), 6.7 ± 0.3 % and 5.3 ± 0.1 %, respectively (**Figure 4.1B**). Similarly as for α-LA hydrolysis, there was a difference in the number of estimated HSSs for bovine, porcine or human trypsins, which were 60 % (9 out of 15), 93 % (14 out of 15) and 74 % (11 out of 15) of the CSs in β-cas, respectively.
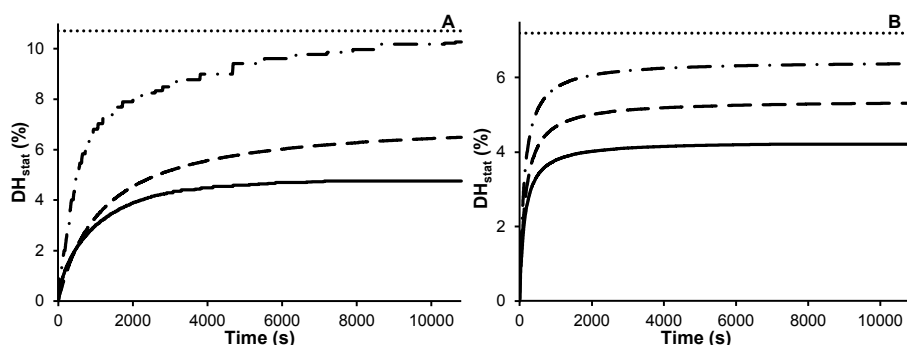
Figure 4.1: Hydrolysis curves (DH_{stat} *versus* time) of hydrolysis of (A) α-LA and (B) β-cas by (—) bovine, (—·) porcine or (---) human trypsins. The dotted lines (······) indicate the $DH_{max,theo}$.

For α-LA hydrolysis, similar initial hydrolysis rate constants were found for bovine, porcine and human trypsins ($k_{intact}^{BT/PT/HT\alpha-LA}$): 0.14, 0.18 and 0.10 s^{-1}, respectively (**Figure 4.2A**). The percentage of remaining intact protein in the hydrolysates as a function of $DH_{stat}/DH_{max,fit}$ indicated that the affinities of three enzymes towards intact α-LA were also similar (**Figure 4.2B**). The $k_{intact}^{BT/PT/HT\beta-cas}$ and enzyme affinities towards the intact β-cas for the three enzymes were also similar (**Figures 4.2C** and **4.2D**). Between hydrolyses of the two proteins, some differences were observed. Firstly, the $k_{intact}^{BT/PT/HT\beta-cas}$ of intact β-cas hydrolysed by bovine, porcine or human trypsins were 1.2, 2.5, 1.5 s^{-1}, respectively, which were 8-14 times higher than the $k_{intact}^{BT/PT/HT\alpha-LA}$ of intact α-LA. Secondly, at DH_{stat} 1 %, all intact β-cas molecules were hydrolysed by the three trypsins (**Figure 4.2D**). Since a DH_{stat} of 1 % means on average that ~2 peptide bonds were hydrolysed in 208 peptide bonds (in β-cas), this suggests that there were 1 or 2 peptide bond(s) that was/were much more susceptible to hydrolysis in intact β-cas than other peptide bonds. This was not observed in α-LA hydrolysis by any of the three trypsins.

Since no difference in the enzyme affinity towards intact proteins was found in the hydrolysis by the three trypsins, the difference in $DH_{max,exp}$ was most likely due to the variations in the amounts of HSSs of the hydrolysis. This was analysed in detail based on the quantitative peptide analysis.

Figure 4.2: Percentage of remaining intact (A, B) α-LA and (C, D) β-cas in the hydrolysates during hydrolysis by (◇) bovine, (□) porcine or (△) human trypsins as a function of (A, C) time and (B, D) $DH_{stat}/DH_{max,fit}$. The lines in panel (B) and (D) were used to guide the eye. The error bars were smaller than the marker size.

Peptide analysis of α-LA hydrolysis by the three trypsins

Peptide annotation of α-LA hydrolysis by the three trypsins

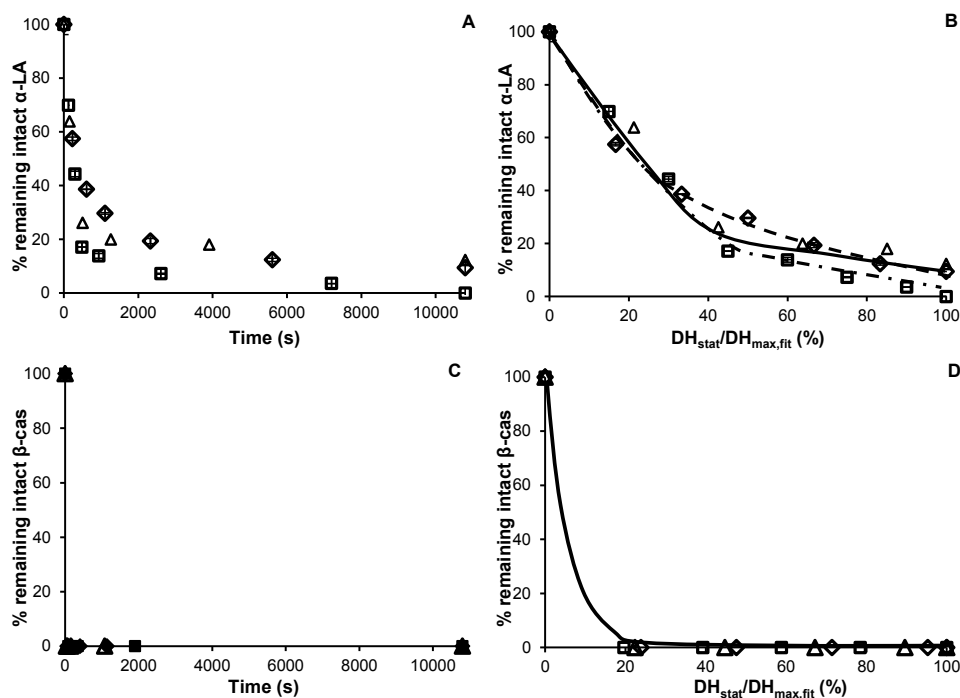Based on the estimated number of HSSs, the predicted maximum number of peptides (NPP$_{max,HSS}$) was 28, 91 and 45 for bovine, porcine and human tryptic hydrolysates, respectively (equation 4). Experimentally, 23 (**Chapter 3**), 30 and 31 unique peptides were annotated in bovine, porcine and human tryptic hydrolysates of α-LA, respectively (**Table 4.2** and **Annex 4.1**). The numbers of annotated peptides were lower than the NPP$_{max,HSS}$, particularly for porcine trypsin. This could be because a large number of intermediate peptides that should be formed during the hydrolysis only accumulated to very low concentrations (only 5 % of the expected A$_{214}$ was not annotated), e.g. below detection or quantification limit. The quality of peptide annotation and quantification was analysed to identify how much of the total hydrolysate was represented by the quantified peptides.

Quality check of the peptide analysis of α-LA by the three trypsins

The quality of peptide analysis was considered to be sufficient, based on the analysis by four parameters: the average AA, peptide and molar sequence coverages, and the DH calculated from the peptide analysis (DH$_{MS}$). The AA sequence coverages of all samples were 100 % (**Table 4.3**). The average peptide sequence coverages for bovine, porcine and human tryptic hydrolysates were comparable: 96±3 % (**Chapter 3**), 90±2 % and 95±5 %, respectively. The average molar sequence coverages for bovine, porcine and human tryptic hydrolysates were 80±9 % (**Chapter 3**), 70±11 % and 72±7 %, respectively (**Table 4.3** and **Annex 4.3**). The sequence coverages in this work were higher than the thresholds set based on the data obtained from previous studies [3, 22, 23], i.e. 80 % and 70 % for the average peptide and molar sequence coverages, respectively [2].

Table 4.3: The amino acid, peptide and molar sequence coverages from the analysis of the bovine, porcine or human tryptic hydrolysates of α-LA and β-cas.

Protein / Enzyme	Average sequence coverage [%]					
	Amino acid		Peptide		Molar	
	α-LA	β-cas	α-LA	β-cas	α-LA	β-cas
Bovine	100±0	100±0	96±3	83±10	80±9	74±8
Porcine	100±0	99.5±1	90±2	87±8	70±11	72±7
Human	100±0	100±0	95±5	91±4	78±8	81±8

A further confirmation was obtained by comparing the DH$_{MS}$ with the DH values taken from the pH-stat titration (DH$_{stat}$). Averaged over all hydrolysates, the standard error between DH$_{stat}$ and DH$_{MS}$ was ~7.3 % (**Figure 4.3A**). For the 3 porcine tryptic hydrolysates towards the end of the hydrolysis, the DH$_{MS}$ were smaller than the DH$_{stat}$, which was also reflected by the low molar sequence coverages (≤60 %) (**Annex 4.3**). This only had a minor effect on the fitting parameters, because for the fitting of data with the equation of 2nd order reaction, the results were mostly determined by the initial 5 data points. Those 5 hydrolysates had high molar sequence coverages (≥70%).
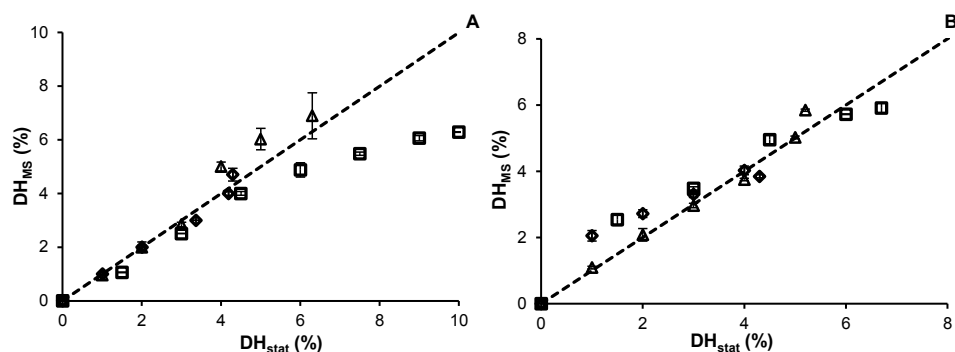
Figure 4.3: Comparison of DH calculated based on peptide analysis (DH_{MS}) and pH-stat titration (DH_{stat}) for (A) α-LA and (B) β-cas hydrolysis by (◇) bovine, (□) porcine or (△) human trypsins. The dotted line (······) indicates the function y=x. The error bars are typically smaller than the marker size.

In bovine tryptic hydrolysates, only specific cleavages were found (**Chapter 3**), while in porcine and human tryptic hydrolysates, 3 and 1 a-specific cleavage(s) were identified, respectively. Such a-specific cleavages were previously also found in the later stage of hydrolysis of whey protein isolate by *Bacillus licheniformis* protease (BLP) [24]. These sites were observed to be hydrolysed only in certain specific peptides. Later on, these peptides were synthesized and incubated in the absence and presence of BLP [24]. Even without the presence of enzyme, the a-specific cleavages still occurred on those unstable peptide bonds [24]. This autolytic degradation of intrinsically unstable peptide bonds were referred to as the "*spontaneous cleavage*" [24]. In this study, the peptides that formed after a-specific cleavages were found only at the end of hydrolysis, and these sites were cleaved to low extents (≤6 µM). Therefore, they were not included as CSs.

Peptide release kinetics of α-LA hydrolysis by the three trypsins

In all hydrolyses, the peptide concentrations were followed during the hydrolysis (**Annex 4.5**). The differences between the hydrolyses by the three enzymes were illustrated for the peptides α-LA[14-16] and α-LA[1-10] (**Figures 4.4A** and **4.4B**). During bovine and porcine tryptic hydrolysis, α-LA[14-16] was released rapidly. Because α-LA[14-16] was a final peptide, i.e. a peptide that did not contain CSs, the concentration only increased. In porcine tryptic hydrolysis, the concentration of α-LA[14-16] in the final hydrolysate reached the injected concentration of α-LA (C_0= 59 µM) (**Figure 4.4A**). This illustrated that CSs K13-14 and K16-17 were efficiently hydrolysed by porcine trypsin. During bovine tryptic hydrolysis, the concentration of peptide α-LA[14-16] reached ~42 µM at the end of hydrolysis. During human tryptic hydrolysis, peptide α-LA[14-16] was released slowly and the concentration reached ~34 µM at the end.

Peptide α-LA[1-10] is an intermediate peptide, of which the concentration can decrease after the hydrolysis of CS K5-6 (**Figure 4.4B**). For hydrolysis by all trypsins, α-LA[1-10] formed rapidly in the beginning. However, for the hydrolysis by different trypsins, the concentration reached different values before it started to decrease. In final porcine and human tryptic hydrolysates,

α-LA[1-10] was not found, which means that K5-6 was fully hydrolysed. Unlike hydrolysis by these two trypsins, the concentration of α-LA[1-10] in the final bovine tryptic hydrolysate was ~10 μM. This means that K5-6 was not fully hydrolysed by bovine trypsin. This was also seen by the final cleavage site products of K5-6, which will be discussed in the next section.
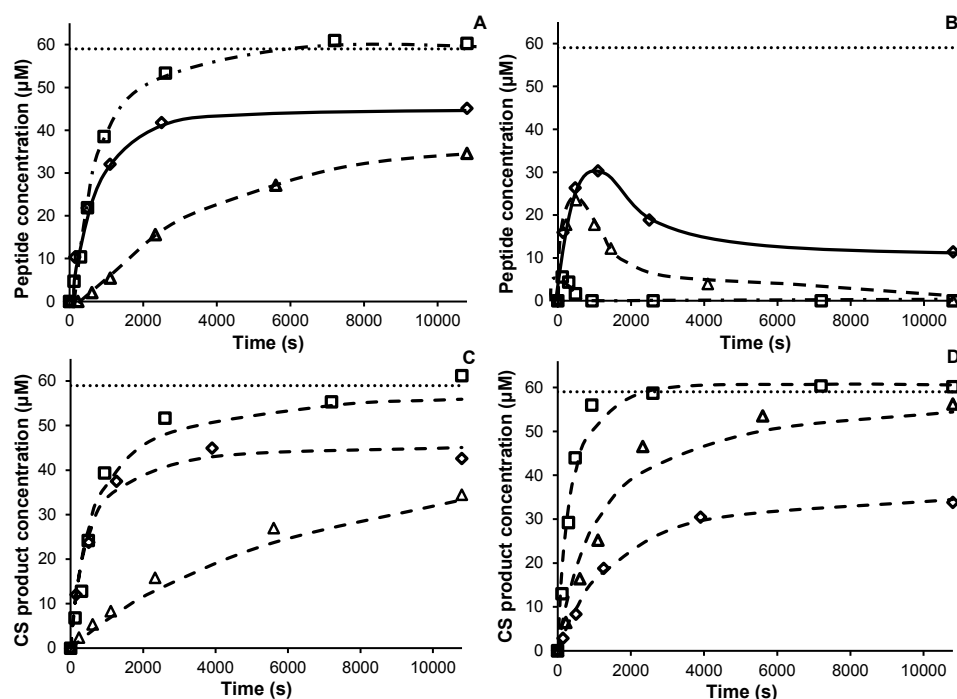
Figure 4.4: Concentrations of peptides (A) α-LA[14-16] and (B) α-LA[1-10], cleavage site products (C) α-LA K13-14 and (D) α-LA R10-11 formed during hydrolysis by (◊) bovine, (□) porcine or (△) human trypsins. The dotted lines (······) represent the initial injected concentration of α-LA. In panels (C) and (D), the dashed lines (------) represent the fits using the 2nd order reaction equation. The markers represent averaged values of duplicate hydrolysates.

Hydrolysis rate constants of CSs in α-LA by the three trypsins

Based on the peptide concentrations at each DH, the concentration of cleavage site products of each CS during α-LA hydrolysis was determined. There were clear differences in the hydrolysis rate of each trypsin of individual CSs, as shown for instance for CS K13-14 (**Figure 4.4C**). This CS was rapidly hydrolysed by bovine and porcine trypsins, but not by human trypsin. CS K13-14 was fully hydrolysed by porcine trypsin as indicated by the final CS products concentration, as well as by the peptide release kinetics of α-LA[14-16]. For CS K5-6, the concentration of the CS products increased rapidly for hydrolysis by all trypsins (**Figure 4.4D**), which was expected based on the peptide release kinetics of peptide α-LA[1-10] discussed above. Still, while for porcine and

human tryptic hydrolyses, the concentrations of CS products at the end of hydrolysis reached 59 and 55 μM, respectively, for bovine tryptic hydrolysis, it only reached 34 μM.

During porcine tryptic hydrolyses, most CSs reached C_0, meaning that they were fully hydrolysed. This was not the case for bovine and human trypsins. For some CSs, the hydrolysis rate constants of bovine and human trypsins were low, meaning that they were not hydrolysed yet within the experiment time. In other cases, the molar sequence coverages of some part of the sequence were not complete, resulting in a low concentration of CS products. In addition, for bovine and human tryptic hydrolysis, there was ~12 % intact α-LA not hydrolysed at the end of hydrolysis. This means that the remaining non-hydrolysed CSs were in the remaining intact α-LA.

From the CS products formation kinetics, hydrolysis rate constants ($k_{i,c}^{BT/PT/HT,\alpha\text{-}LA}$) of the three enzymes towards CSs in α-LA were determined (**Table 4.4**). Differences between the enzymes were seen for instance by the hydrolysis kinetics of CS K62-63. This CS was efficiently hydrolysed by porcine trypsin ($1.7 \cdot 10^{-3}$ s$^{-1} \cdot$mg$^{-1}_{enzyme}$). However, it was not hydrolysed by bovine trypsin at all, and was only slowly hydrolysed by human trypsin ($0.07 \cdot 10^{-3}$ s$^{-1} \cdot$mg$^{-1}_{enzyme}$). Overall, the total hydrolysis rate constants of porcine trypsin for CSs in α-LA ($k_{i,c}^{PT,\alpha\text{-}LA}$) were ~2 times higher than those of bovine and human trypsins. The $k_{i,c}^{BT,\alpha\text{-}LA}$ for bovine trypsin towards CSs in α-LA were equally distributed among the range of 10^{-5}, 10^{-4} and 10^{-3} s$^{-1} \cdot$mg$^{-1}_{enzyme}$. The similar range of $k_{i,c}^{BT,\alpha\text{-}LA}$ was found for $k_{i,c}^{HT,\alpha\text{-}LA}$ of human trypsin. Unlike these two trypsins, for porcine trypsin, 10 out of 13 CS had similar hydrolysis rate constants, which were all ≥10^{-3} s$^{-1} \cdot$mg$^{-1}_{enzyme}$.

Table 4.4: Individual and total hydrolysis rate constants (±16 % standard error) of all cleavage sites (CSs) in α-LA during hydrolysis by bovine, porcine or human trypsins based on the CS products concentrations. CSs classified as high, intermediate and low selectivity sites were highlighted with ▢, ▢ and ▢, respectively.

	Hydrolysis rate constants $k_{i,c}^{enzyme,\ \alpha\text{-}LA}$ [10^{-3} s$^{-1} \cdot$mg$^{-1}_{enzyme}$]		
Enzyme CS	Bovine trypsin	Porcine trypsin	Human trypsin
K5	0.5(±0.1)	3.4(±0.1)	1.3(±0.2)
R10	3.0(±0.0)	2.8(±0.2)	2.7(±0.1)
K13	2.3(±0.0)	2.4(±0.2)	0.2(±0.0)
K16	1.2(±0.0)	3.4(±0.1)	1.3(±0.2)
K58	1.3(±0.2)	1.8(±0.3)	2.2(±0.5)
K62	0.0(±0.0)	1.7(±0.1)	0.07(±0.00)
K79	0.05(±0.01)	0.01(±0.00)	0.3(±0.1)
K93	0.05(±0.01)	2.1(±0.0)	0.3(±0.1)
K94	0.4(±0.0)	2.3(±0.0)	1.6(±0.5)
K98	0.5(±0.0)	1.5(±0.2)	0.9(±0.1)
K108	1.3(±0.0)	3.9(±0.4)	1.2(±0.0)
K114	0.3(±0.1)	0.004(±0.000)	0.004(±0.000)
K122	0.0(±0.0)	0.002(±0.000)	0.002(±0.000)
Total	10.9	25.3	12.1

For porcine and human trypsins, the hydrolysis rate constants of CSs in α-LA were also calculated based on the concentration of remaining non-hydrolysed CS ($C_{remain,t}$) in the hydrolysates. For CSs in the part of AA sequence where the molecular sequence coverages were high, the hydrolysis rate constants calculated based on the two methods have an average standard error of 26 %. This was 10 % higher than the average standard error of hydrolysates in duplicate experiments (16 %). The hydrolysis rate constants of CSs calculated using both methods were used to further calculate the enzyme selectivity.

Using the same approach as the peptide analysis of α-LA hydrolysates, the hydrolysis kinetics, peptide release kinetics and hydrolysis rate constants of the three trypsins towards CSs in β-cas were analysed.

Peptide analysis of β-cas hydrolysis by the three trypsins
Peptide annotation of the β-*cas hydrolysis by the three trypsins*
Based on the estimated number of HSSs, the predicted maximum number of peptides ($NPP_{max,pre}$) was 55, 120 and 78 for bovine, porcine and human tryptic hydrolysates of β-cas, respectively. Experimentally, 34 (**Chapter 3**), 35 and 40 unique peptides were annotated in bovine, porcine and human tryptic hydrolysates, respectively (**Table 4.2** and **Annex 4.2**). Similarly as for α-LA hydrolysis, the annotated numbers of peptides were lower than the predicted numbers.

Quality check of the peptide analysis of β-*cas by the three trypsins*
Overall, for all β-cas hydrolysates, the average AA, peptide and molar sequence coverages were in the same range as those for α-LA (**Table 4.3** and **Annex 4.4**). Only porcine tryptic hydrolysates at DH 6.7 % had an AA sequence coverage less than 100 %. This was because this hydrolysate mainly contained final peptides, where dipeptides β-cas[98-99] and β-cas[106-107], were not found in the chromatograms. Because dipeptides have low molar extinction coefficients, the loss of UV signal was not visible from the total UV signal. The quality of annotation and quantification was further confirmed by the agreement between the calculated DH_{MS} of each β-cas hydrolysate and DH_{stat} (**Figure 4.3B**). Averaged over all samples, the standard error between the DH_{stat} and DH_{MS} was ~7.7 %. In bovine, porcine and human hydrolysates of β-cas, peptides with 4, 5 and 4 a-specific cleavages were found, respectively. Due to low concentrations of CS products for these sites (≤8 µM), they were not included as CS.

Peptide release kinetics during β-*cas hydrolysis by the three trypsins*
The differences in peptide release kinetics during hydrolysis of β-cas by the three trypsins were compared (**Annex 4.6**). This was illustrated for the peptides β-cas[108-113] and β-cas[183-209] (**Figures 4.5A** and **4.5B**). As expected, for porcine tryptic hydrolysis, the final peptide β-cas[108-113] formed rapidly in the beginning and reached ~35 µM after 1000 s, which was close to the initial injected concentration (34 µM) (**Figure 4.5A**). This indicated that CSs K108-109 and K114-115 were fully hydrolysed by porcine trypsin. During bovine tryptic hydrolysis, β-cas[108-113]

was released gradually and the concentration reached ~23 µM. For hydrolysis by human trypsin, β-cas[108-113] was not formed in the first ~3500 s, but was so rapidly accumulated after this point that it still reached the initial injected concentration (34 µM) before the end of the experiment. Previously, such delay in the release kinetics of peptides was attributed to the "demasking" phenomonon. Demasking describes that during protein enzymatic hydrolysis some CSs may only become accessible after cleavage of other peptide bonds in the protein [25].

The concentration of the intermediate peptide β-cas[183-209], increased rapidly from t=0 and started to decrease at DH 2 % in porcine tryptic hydrolysis, and was completely degraded at ~1000 s (**Figure 4.5B**). For both bovine and human tryptic hydrolyses, β-cas[183-209] formed gradually during the hydrolysis and started to decrease at ~400 and ~800 s, respectively. There was ~5 µM of β-cas[183-209] remained in the final bovine tryptic hydrolysate, whereas for the other two trypsins, it was completely degraded.
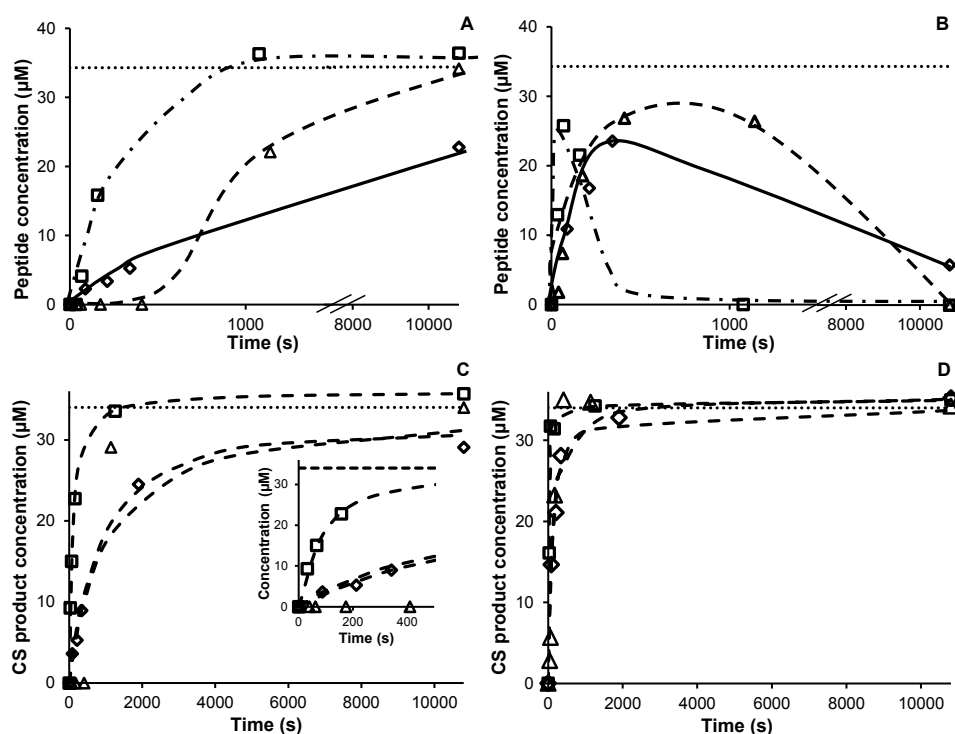
Figure 4.5: Concentrations of peptides (A) β-cas[108-113], (B) β-cas[184-209], cleavage site products (C) β-cas K113-114, (D) β-cas K183-184 formed during hydrolysis by (◊) bovine, (□) porcine and (△) human trypsins. The dashed lines (------) represent the fits using the equation of 2nd order reaction and the dotted lines (······) represent the injected concentration of each protein. The inset in panel (C) shows the zoomed curve of the cleavage site products formation of β-cas K113-114 from 0-500 s. The markers are the averaged values of duplicate hydrolysates.

Hydrolysis rate constants of CSs in β-cas by the three trypsins

During porcine tryptic hydrolysis, the concentration of CS products of K113-114 increased rapidly at the start, reaching the initial injected concentration (34 µM) within ~1500 s (**Figure 4.5C**). The hydrolysis of K113-114 by bovine trypsin was slower than by porcine trypsin. By bovine trypsin, the concentration of the CS products only reached ~27 µM at the end of hydrolysis. For human trypsin, as indicated by the peptide release kinetics of β-cas[108-113], the cleavage of K113-114 was masked in intact β-cas. Indeed, no CS products of K113-114 were formed in the beginning of hydrolysis. However, after DH 4 %, the concentration increased rapidly and reached ~33 µM (C_0=34 µM). This indicated that human trypsin could hydrolyse K113-114 once it was demasked. This also explained why the fit using the 2nd order reaction equation was not good for this CS. Clear evidence of demasking effects were observed in only 4 cases (R202-203 for the three enzymes and K113-114 for human trypsin) in all hydrolyses. The equation for the consecutive reaction, as describe previously [26], was used to calculate the hydrolysis rate constants of the CSs. For CS K183-184, all trypsins hydrolysed the CS efficiently, illustrated by the fact that all the CS products in the end of hydrolysis reached C_0 (**Figures 4.5D**). However, porcine trypsin hydrolysed the K183-184 faster than the other two trypsins.

The hydrolysis rate constants ($k_{i,c}^{BT/PT/HT,\beta-cas}$) of all CSs in β-cas by the three enzymes were determined (**Table 4.5**). Similarly as for α-LA hydrolysis, the total hydrolysis rate constants ($k_{i,c}^{PT,\beta-cas}$) of porcine trypsin towards CSs in β-cas were higher than bovine and human trypsins.

Table 4.5: Individual and total hydrolysis rate constants (±15 % standard error) of all cleavage sites (CSs) in β-cas during hydrolysis by bovine, porcine or human trypsins based on the CS products concentrations. CSs classified as high, intermediate and low selectivity sites were highlighted with ▢, ▢ and ▢, respectively.

Enzyme / CS	Hydrolysis rate constants $k_{i,c}^{enzyme,\beta-cas}$ $[10^{-3}\ s^{-1}\cdot mg^{-1}_{enzyme}]$		
	Bovine trypsin	Porcine trypsin	Human trypsin
R1	0.0(±0.0)	0.0(±0.0)	0.0(±0.0)
R25	1.4(±0.4)	14.1(±1.3)	64.4(±7.6)
K28	16.7(±0.9)	114.8(±21.9)	56.2(±2.6)
K29	1.8(±0.1)	4.4(±0.3)	0.1(±0.0)
K32	1.6(±0.7)	8.0(±0.14)	0.9(±0.1)
K48	0.002(±0.000)	0.2(±0.0)	0.7(±0.1)
K97	1.3(±0.3)	53.1(±3.7)	53.7(±2.6)
K99	31.4(±1.9)	17.5(±0.8)	0.5(±0.0)
K105	47.8(±3.5)	92.6(±20.1)	12.1(±0.3)
K107	5.7(±2.0)	5.9(±1.8)	0.7(±0.2)
K113	2.3(±0.2)	10.1(±0.7)	0.1(±0.0)
K169	66.4(±8.2)	191.5(±13.1)	19.4(±0.7)
K176	23.8(±7.1)	98.7(±27.8)	13.9(±1.1)
R183	6.0(±1.9)	37.8(±2.8)	10.2(±0.1)
R202	0.4(±0.2)	2.5(±0.3)	0.04(±0.0)
Total	207.6	651.2	232.9

89

In the final hydrolysate of β-cas during hydrolysis by porcine trypsin, mainly final peptides were found, indicating that almost all CSs were fully hydrolysed. This means that the differences in $k_{i,c}^{PT,\beta-cas}$ were not from the differences in $C_0^{PT,\beta-cas}$ but almost entirely from the differences in $k_{i,app}^{PT,\beta-cas}$. None of the three enzymes hydrolysed the CS R1-2. For the hydrolysed CSs, bovine and human trypsins were both more selective than porcine trypsin, but the preferred CSs were not the same. As shown in **Figure 4.2B**, in all hydrolyses, there were 1 or 2 CSs highly preferred by each trypsin, resulting in almost complete hydrolysis of intact β-cas at DH$_{stat}$ 1 %. This CS highly preferred by bovine trypsin was reported to be K169-170 (**Chapter 3**). In this work, K169-170 was also found to be highly preferred by porcine trypsin, indicated by the fact that at DH$_{stat}$ 1 % the CS products of K169-170 reached C$_0$. Unlike bovine and porcine trypsins, the highly preferred CS by human trypsin was R25-26. For porcine and human trypsins, the hydrolysis rate constants of CSs were also calculated based on the concentration of remaining non-hydrolysed CSs in the hydrolysates. For CSs that were in the part of the AA sequence with high molar sequence coverages, the hydrolysis rate constants of the CSs determined based on both methods had an average standard error of 28 %.

It should be noted that even for the HSSs, the hydrolysis rate constants of the three trypsins towards CSs in β-cas were higher than those of CSs in α-LA. This might be due to the differences in hydrolysis rate constants of intact proteins (k_{intact}) during the hydrolysis. The k_{intact} values of the three trypsins of β-cas were 8-14 times higher than their k_{intact} values of α-LA. This was also reflected on the total hydrolysis rate constants of the two proteins. This then suggests that the measured hydrolysis rate constant of CSs are the product of the k_{intact} and the hydrolysis rate constants of the CS in the peptides.

Enzyme selectivity and the prediction of maximum degree of hydrolysis

Based on the hydrolysis rate constants calculated by the concentrations of CS products and the remaining non-hydrolysed CSs ($C_{remain,t}$) in the hydrolysates, the selectivities of bovine, porcine and human trypsin towards CSs in α-LA and β-cas were calculated using equation 13. Calculated based on the two methods, for the CSs in the part of AA sequence where the molecular sequence coverages were high, the selectivity values had an average standard error of 26 %. The selectivity values calculated based on the concentrations of CS products were used for further comparison between the selectivity of each trypsin (**Figures 4.6A** and **4.6C**). The selectivity values of porcine trypsin towards most CSs in α-LA were similar, while the range of selectivity values for bovine and human trypsin selectivity was much larger. This means that bovine and human trypsins were more selective than porcine trypsin towards CSs in α-LA (**Figure 4.6A**). For β-cas hydrolysis, the total hydrolysis rate constant of porcine tryptic hydrolysis was 3 times higher than the bovine and human tryptic hydrolysis (**Figure 4.6D**). Although relatively the ranges of selectivity values were similar for hydrolysis by the three trypsins, porcine trypsin efficiently hydrolysed more CSs than the other two trypsins.
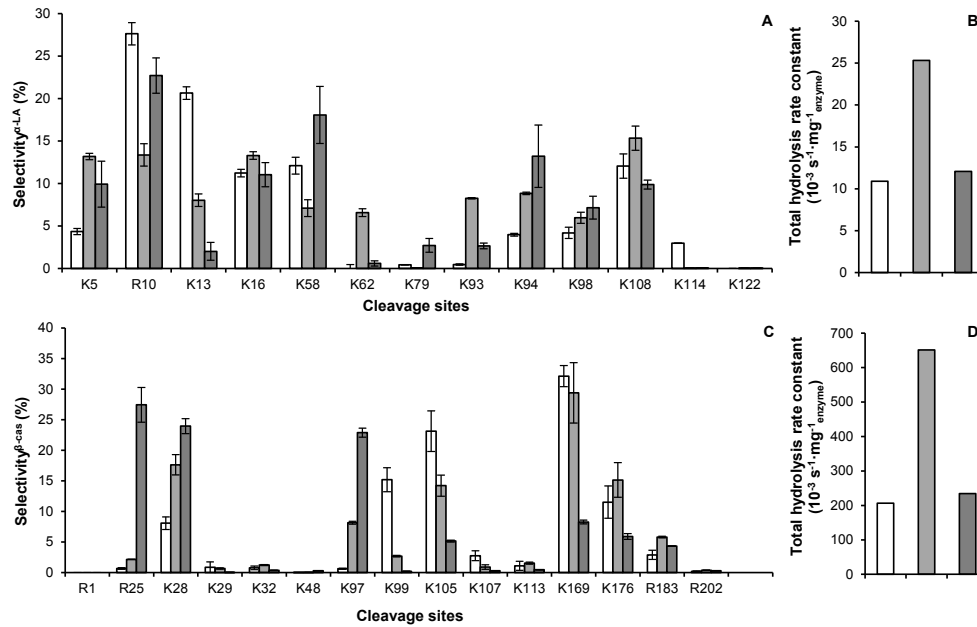
Figure 4.6: (A, C) Bovine (□), porcine (■) and human (■) trypsin selectivity towards cleavage sites calculated based on the concentration of the CS products and (B, D) the total hydrolysis rate constants ($\sum \left(k_i \times C_0 \right)$) in (A, B) α-LA and (C, D) β-cas.

Based on the threshold $k_{i,c}^{BT/PT/HT, \alpha\text{-}LA}$ values from the results of bovine trypsin selectivity, the high (HSSs), intermediate (ISSs) and low/zero (LSSs) selectivity sites of porcine and human trypsins towards CSs in α-LA and β-cas were categorised (**Table 4.6**). For CSs in α-LA, the amount of HSSs for each trypsin was close to but slightly lower than the prediction based on DH$_{max,fit}$, which was 6, 12 and 8 for bovine, porcine and human trypsins, respectively. This was also observed for β-cas hydrolysis. That lower number of HSSs obtained experimentally was due to the fact that ISSs also contributed to the DH$_{max,fit}$.

Table 4.6: Number of cleavage sites in α-LA and β-cas towards which bovine, porcine and human trypsins have high, intermediate or low/zero selectivity.

Categories / Enzyme	α-LA			β-cas		
	Bovine trypsin	Porcine trypsin	Human trypsin	Bovine trypsin	Porcine trypsin	Human trypsin
High selectivity sites	5	10	6	7	11	7
Intermediate selectivity sites	4	-	3	5	2	1
Low/zero selectivity sites	4	3	4	3	2	7

Based on the numbers of HSSs, ISSs and LSSs, the predicted maximum degree of hydrolysis (DH$_{max,pre}$) based on enzyme selectivity was calculated using equation 14, and was compared with the DH$_{max,exp}$ (**Figure 4.7**). The average standard error (|DH$_{max,pre/theo}$-DH$_{max,exp}$|/DH$_{max,exp}$×100 %)

between the predicted $DH_{max,pre}$ based on selectivity and the experimental $DH_{max,exp}$ was ~14 %, which was ~4 times more accurate than the prediction based on enzyme specificity ($DH_{max,theo}$).
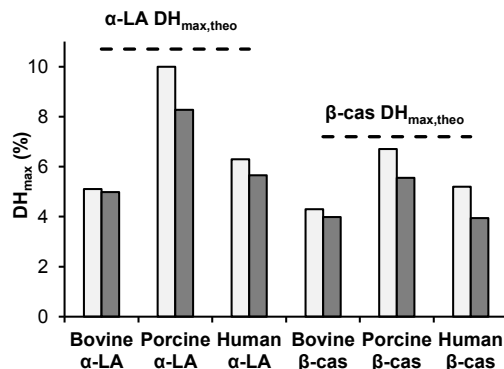
Figure 4.7: Experimental maximum degree of hydrolysis ($DH_{max,exp}$, marked with ▪) and the predicted maximum degree of hydrolysis based on enzyme selectivity ($DH_{max,pre}$, marked with ■) of α-LA and β-cas hydrolysis by bovine, porcine or human trypsins. The dashed lines (------) represent the theoretical maximum degree of hydrolysis based on trypsin specificity ($DH_{max,theo}$) of α-LA and β-cas.

Based on coherent outcome of α-LA and β-cas hydrolysis by bovine, porcine or human trypsins, it is concluded that the differences in $DH_{max,exp}$ of hydrolysis by the three trypsins can be explained by their differences in enzyme selectivity. Based on the percentage of HSSs of the total CSs, for the two substrates, bovine, porcine and human trypsin efficiently hydrolysed 43 %, 75 % and 50 %, respectively.

Link between binding site sequence and selectivity of porcine and human trypsins

Using the rules of bovine trypsin selectivity, 57 % (16 out of 28) and 50 % (14 out of 28) of CSs for porcine and human trypsin selectivity, respectively, were correctly predicted (**Table 4.1**). To improve the prediction, the Link between the binding site sequence and the selectivity of porcine and human trypsins was studied. The selectivity of porcine and human trypsin towards CSs, was not related to the hydrophobicity and molecular weight of the AAs on the binding site sequence. For both trypsins, the standard errors of the molecular weights of AAs in all positions in the binding site sequence in the 3 categories (HSS, ISS and LSS) were similar (10-22 %), and close to the standard error over all 20 protein AAs (22 %). For hydrophobicity, the absolute values of standard errors for hydrophobic moments of all binding site positions in the 3 categories were between 95 and 8220 %, and the absolute standard error over all 20 protein AA is ~420 %. There was a relation between selectivity of porcine trypsin, and charge of the AAs on the P2 position. All 5 LSSs had a negative charged AA on that position (**Table 4.7**). Out of the 21 HSSs, only 2 CSs (α-LA K62 and K98) had a negative charged AA on P2 position (if the cysteine was taken as a charged AA). For all other binding site positions, no clear differences were found between the HSSs and LSSs. This rule, i.e. porcine trypsin efficiently hydrolyses CSs with no negatively charged AAs on P2 position (**Table 4.9**), explained the porcine trypsin selectivity towards of 24 out of 28

(~86 %) CSs in α-LA and β-cas. Positively charged AAs did not seem to hinder the porcine trypsin selectivity since they were not present in the binding site sequences of the LSSs, but were mostly present in the binding site sequences of the HSSs.

Table 4.7: Porcine trypsin hydrolysis rate constants and the amino acid sequences of binding site positions (from P4 to P4') of cleavage sites in α-LA and β-cas. Amino acids with a negatively charged side chain (DE) are marked with ▮. Amino acids with a positively charged side chain (KR) are marked with ▮ except when they are on P1. Cysteines are marked with ▮. "-" means there is no amino acid present. The ⬭ emphasized the P2 position of the binding site.

Cleavage sites	$k_{i,c}^{PT}$ (s^{-1}·mg$^{-1}_{enzyme}$)	Binding sites							
		P4	P3	P2	P1	P1'	P2'	P3'	P4'
High selectivity									
α-LA K108	3.9E-3	L	A	H	K	A	L	C	S
α-LA K5	3.4E-3	Q	L	T	K	C	E	V	F
α-LA K16	3.4E-3	K	D	L	K	G	Y	G	G
α-LA R10	2.8E-3	E	V	F	R	E	L	K	D
α-LA K13	2.4E-3	R	E	L	K	D	L	K	G
α-LA K94	2.3E-3	C	V	K	K	I	L	D	K
α-LA K93	2.1E-3	M	C	V	K	K	I	L	D
α-LA K58	1.8E-3	I	N	N	K	I	W	C	K
α-LA K62	1.7E-3	I	W	C	K	D	D	Q	N
α-LA K98	1.5E-3	I	L	D	K	V	G	I	N
β-cas K169	1.9E-1	S	Q	S	K	V	L	P	V
β-cas K28	1.1E-1	R	I	N	K	K	I	E	K
β-cas K176	9.9E-2	V	P	Q	K	A	V	P	Y
β-cas K105	9.3E-2	M	A	P	K	H	K	E	M
β-cas K97	5.3E-2	G	V	S	K	V	K	E	A
β-cas R183	3.8E-2	Y	P	Q	R	D	M	P	I
β-cas K99	1.7E-2	S	K	V	K	E	A	M	A
β-cas R25	1.4E-2	S	I	T	R	I	N	K	K
β-cas K113	1.0E-2	P	F	P	K	Y	P	V	E
β-cas K32	7.9E-3	K	I	E	K	F	Q	S*	E
β-cas K107	5.9E-3	P	K	H	K	E	M	P	F
Intermediate selectivity									
β-cas K29	4.4E-3	I	N	K	K	I	E	K	F
β-cas R202#	2.5E-3	G	P	V	R	G	P	F	P
Low/zero selectivity									
α-LA K79	1.4E-5	S	C	D	K	F	L	D	D
α-LA K114	3.6E-6	C	S	E	K	L	D	Q	W
α-LA K122	2.2E-6	L	C	E	K	L	-	-	-
β-cas K48	2E-4	L	Q	D	K	I	H	P	F
β-cas R1	0	-	-	-	R	E	L	E	E

represents masked cleavage site in intact protein
* represents phosphoserine

For human trypsin, it was found that 11 out of 15 ISSs and LSSs had a negative charged AA on either P2 or P1' positions of the binding site sequences (**Table 4.8**). For HSSs, only 3 out of 13 CSs had negative charged AAs on either P2 or P1' positions of the binding site sequence (if cysteine was taken as a charged AA). This means that human trypsin selectivity towards 21 out of 28 (~82 %) CS can be explained by the hindrance of negatively charged AAs on either P2 or P1' positions

(**Table 4.9**). No clear observation was found on the effect of positively charged AAs on the human trypsin selectivity. That porcine trypsin was only hindered by a charged AA on 1 position of the binding site sequence explained why it reached higher $DH_{max,exp}$ than the other two trypsins.

Table 4.8: <u>Human trypsin</u> hydrolysis rate constants and the amino acid sequences of binding site positions (from P4 to P4') of cleavage sites in α-LA and β-cas. Amino acids with a negatively charged side chain (DE) are marked with ▨. Amino acids with a positively charged side chain (KR) are marked with ▨ except when they are on P1. Cysteines are marked with ▨. "-" means there is no amino acid present. The ▱ emphasized the P2 and P1' positions of the binding site.

Cleavage sites	$k_{i,c}^{HT}$ $(s^{-1} \cdot mg^{-1}_{enzyme})$	Binding sites							
		P4	P3	P2	P1	P1'	P2'	P3'	P4'
High selectivity									
α-LA R10	2.7E-3	E	V	F	R	E	L	K	D
α-LA K58	2.2E-3	I	N	N	K	I	W	C	K
α-LA K94	1.6E-3	C	V	K	K	I	L	D	K
α-LA K5	1.3E-3	Q	L	T	K	C	E	V	F
α-LA K16	1.3E-3	K	D	L	K	G	Y	G	G
α-LA K108	1.2E-3	L	A	H	K	A	L	C	S
β-cas R25	6.4E-2	S	I	T	R	I	N	K	K
β-cas K28	5.6E-2	R	I	N	K	K	I	E	K
β-cas K97	5.3E-2	G	V	S	K	V	K	E	A
β-cas K169	1.9E-2	S	Q	S	K	V	L	P	V
β-cas K176	1.4E-2	V	P	Q	K	A	V	P	Y
β-cas K105	1.2E-2	M	A	P	K	H	K	E	M
β-cas R183	1.0E-2	Y	P	Q	R	D	M	P	I
Intermediate selectivity									
α-LA K98	8.5E-4	I	L	D	K	V	G	I	N
α-LA K79	3.2E-4	S	C	D	K	F	L	D	D
α-LA K93	3.2E-4	M	C	V	K	K	I	L	D
β-cas K32	9.0E-3	K	I	E	K	F	Q	S*	E
Low/zero selectivity									
α-LA K13	2.3E-4	R	E	L	K	D	L	K	G
α-LA K62	6.9E-5	I	W	C	K	D	D	Q	N
α-LA K114	3.6E-6	C	S	E	K	L	D	Q	W
α-LA K122	2.2E-6	L	C	E	K	L	-	-	-
β-cas K107	7.4E-4	P	K	H	K	E	M	P	F
β-cas K48	7.0E-4	L	Q	D	K	I	H	P	F
β-cas K113#	8.0E-5	P	F	P	K	Y	P	V	E
β-cas K29	7.1E-5	I	N	K	K	I	E	K	F
β-cas K99	5.1E-5	S	K	V	K	E	A	M	A
β-cas R202#	3.8E-5	G	P	V	R	G	P	F	P
β-cas R1	0	-	-	-	R	E	L	E	E

represents masked cleavage sites in intact protein
* represents phosphoserine

Table 4.9: Revised rules for porcine and human trypsin selectivity towards cleavage sites (CSs) in a protein, and the number of correctly predicted CSs of the total CSs in α-LA and β-cas based on the rules. ■ indicates negatively charged amino acids (DE) and ■ indicates all other amino acids. The letters in the columns represent the abbreviation of the amino acids.

Categories	Porcine trypsin		Number of correctly predicted CSs/total CSs
	Binding sites		
	P2	**P1**	
High selectivity	All others	K/R	19/21
Intermediate or low/zero selectivity	DE	K/R	5/7
		Total	**24/28**

Categories	Human trypsin			Number of correctly predicted CSs/total CSs
	Binding sites			
	P2	**P1**	**P1'**	
High selectivity	All others	K/R	All others	10/14
Intermediate or Low/zero selectivity	DE	K/R	All others	6/15
	All others	K/R	DE	4/15
	DE	K/R	DE	1/15
			Total	**21/28**

*Due to limited number of CSs were categorised as intermediate selectivity sites, the intermediate and low/zero selectivity sites were combined.

Conclusion

For two proteins (α-lactalbumin and β-casein), large differences were found in the $DH_{max,exp}$ of two proteins by bovine, porcine or human trypsin. The difference of the $DH_{max,exp}$ of hydrolysis by porcine trypsin and the theoretical maximum was small, whereas the $DH_{max,exp}$ of bovine or human tryptic hydrolysis were much lower than the theoretical maximum. These differences were explained by the differences between the relative hydrolysis rate constants of the three trypsins towards cleavage sites in both proteins (enzyme selectivity). Porcine trypsin hydrolysed most cleavage sites in both proteins efficiently, while bovine and human trypsins only hydrolysed half of the cleavage sites efficiently. Knowing the enzyme selectivity, the prediction of $DH_{max,pre}$ was ~4 times better than the estimation based only on the enzyme specificity. Preliminary relations between the amino acid sequence of the binding site positions and the selectivity of each trypsin were proposed, which would allow estimations of the $DH_{max,exp}$ of hydrolysis of other substrate proteins by these trypsins. This work can help to explain the differences in the reported *in vitro* digestibility of proteins, and to define expectations of the outcome of future experiments using these trypsins.

Acknowledgement

References

[1] Y. Deng, P.A. Wierenga, H.A. Schols, S. Sforza, H. Gruppen, Effect of Maillard induced glycation on protein hydrolysis by lysine/arginine and non-lysine/arginine specific proteases, Food Hydrocolloids, 69 (2017) 210-219.

[2] Y. Deng, F. van der Veer, S. Sforza, H. Gruppen, P.A. Wierenga, Towards predicting protein hydrolysis by bovine trypsin, Process Biochemistry, 65 (2018) 81-92.

[3] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Introducing enzyme selectivity: A quantitative parameter to describe enzymatic protein hydrolysis, Analytical and Bioanalytical Chemistry, 406 (2014) 5827-5841.

[4] H. Rinderknecht, I.G. Renner, C. Carmack, Trypsinogen variants in pancreatic juice of healthy volunteers, chronic alcoholics, and patients with pancreatitis and cancer of the pancreas, Gut, 20 (1979) 886-891.

[5] A. Fernández, F. Riera, β-Lactoglobulin tryptic digestion: A model approach for peptide release, Biochemical Engineering Journal, 70 (2013) 88-96.

[6] S.J. Walmsley, P.A. Rudnick, Y. Liang, Q. Dong, S.E. Stein, A.I. Nesvizhskii, Comprehensive analysis of protein digestion using six trypsins reveals the origin of trypsin as a significant source of variability in proteomics, Journal of Proteome Research, 12 (2013) 5666-5680.

[7] B. Keil, Specificity of proteolysis, Springer-Verlag Berlin Heidelberg, New York, USA, 1992, pp. 66-69.

[8] E. Vandermarliere, M. Mueller, L. Martens, Getting intimate with trypsin, the leading protease in proteomics, Mass Spectrometry Reviews, 32 (2013) 453-465.

[9] I. Schechter, A. Berger, On the size of the active site in proteases. I. Papain, Biochemical and Biophysical Research Communications, 27 (1967) 157-162.

[10] K. Dallas Johnson, A. Clark, S. Marshall, A functional comparison of ovine and porcine trypsins, Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 131 (2002) 423-431.

[11] J. Adler-Nissen, Enzymic hydrolysis of food proteins, Elsevier Applied Science Publishers, London, UK, 1986, pp. 122-123.

[12] C.I. Butré, P.A. Wierenga, H. Gruppen, Influence of water availability on the enzymatic hydrolysis of proteins, Process Biochemistry, 49 (2014) 1903-1912.

[13] C.I. Butré, Introducing enzyme selectivity as a quantitative parameter to describe the effects of substrate concentration on protein hydrolysis, PhD thesis Wageningen University & Research, Wageningen, the Netherlands, 2014.

[14] B.J.H. Kuipers, H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis, Journal of Agricultural and Food Chemistry, 55 (2007) 5445-5451.

[15] G.J.M. Bruin, G. Stegeman, A.C. Van Asten, X. Xu, J.C. Kraak, H. Poppe, Optimization and evaluation of the performance of arrangements for UV detection in high-resolution separations using fused-silica capillaries, Journal of Chromatography A, 559 (1991) 163-181.

[16] J.P. Chervet, R.E.J. Van Soest, M. Ursem, Z-shaped flow cell for UV detection in capillary electrophoresis, Journal of Chromatography A, 543 (1991) 439-449.

[17] S.E. Moring, R.T. Reel, R.E.J. van Soest, Optical improvements of a Z-shaped cell for high-sensitivity UV absorbance detection in capillary electrophoresis, Analytical Chemistry, 65 (1993) 3454-3459.

[18] H.A. Kosters, P.A. Wierenga, R. de Vries, H. Gruppen, Characteristics and effects of specific peptides on heat-induced aggregation of β-lactoglobulin, Biomacromolecules, 12 (2011) 2159-2170.

[19] K. Linderstrøm-Lang, The initial phases of the enzymatic degradation of proteins, Bulletin de la Société de chimie biologique, 35 (1953) 100-116.

[20] D. Eisenberg, R.M. Weiss, T.C. Terwilliger, W. Wilcox, Hydrophobic moments and protein structure, Faraday Symposia of the Chemical Society, 17 (1982) 109-120.

[21] P.W.J.R. Caessens, W.F. Daamen, H. Gruppen, S. Visser, A.G.J. Voragen, β-Lactoglobulin hydrolysis. 2. Peptide identification, SH/SS exchange, and functional properties of hydrolysate fractions formed by the action of plasmin, 47 (1999) 2980-2990.

[22] C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Determination of the influence of the pH of hydrolysis on enzyme selectivity of *Bacillus licheniformis* protease towards whey protein isolate, International Dairy Journal, 44 (2015) 44-53.

[23] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Determination of the influence of substrate concentration on enzyme selectivity using whey protein isolate and *Bacillus licheniformis* protease, Journal of Agricultural and Food Chemistry, 62 (2014) 10230-10239.

[24] C.I. Butré, S. Buhler, S. Sforza, H. Gruppen, P.A. Wierenga, Spontaneous, non-enzymatic breakdown of peptides during enzymatic protein hydrolysis, Biochimica et Biophysica Acta - Proteins and Proteomics, 1854 (2015) 987-994.

[25] M.M. Vorob'ev, E.A. Paskonova, S.V. Vitt, V.M. Belikov, Kinetic description of proteolysis. Part 2. Substrate regulation of peptide bond demasking and hydrolysis. Liquid chromatography of hydrolyzates, Die Nahrung, 30 (1986) 995-1001.

[26] M.M. Vorob'ev, C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Demasking kinetics of peptide bond cleavage for whey protein isolate hydrolysed by *Bacillus licheniformis* protease, Journal of Molecular Catalysis B: Enzymatic, 133 (2016) S426-S431.

Annex 4.1: List of peptides derived from α-LA hydrolysed by the three trypsins. (* indicates methionine oxidation)

Peptides	Sequence	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	b fragments identified	y fragments identified	Bovine	Porcine	Human
[1-5]	EQLTK	617.3187	617.3384	-32		1,2,3,4	✓	✓	✓
[1-10]	EQLTKCEVFR	1251.6004	1251.6281	-22	2	1,2,3,4,5,6,7,8,	✓	✓	✓
[1-13]	EQLTKCEVFRELK	1621.8086	1621.8497	-25	1,2,3,4,5,6,7,8,10,12	1,2,3,4,6,7,8,9,10,11,	✓	✓	✓
[1-16]	EQLTKCEVFRELKDLK	1978.082	1978.0557	13	1,2,3,6,7,9,10,13,14,15	1,2,3,4,5,9,10,11,12,13,14	✓	✓	✓
[1-58]	EQLTKCEVFRELKDLKGYGGVSLPEWVCTTFHTSGYDTQA IVQNNDSTEYGLFQINNK	6612.9752	6613.1845	-32	9	3,4,5,6,8,13,16	✓		✓
[6-10]	(K)CEVFR	652.2826	652.3003	-27	2,3	1,3,4	✓	✓	✓
[6-13]	(K)CEVFRELK	1022.538	1022.52	18	2,3	1,2,3,4	✓	✓	✓
[11-13]	(R)ELK	388.2226	388.2322	-25	1,3	1,2	✓		✓
[11-16]	(R)ELKDLK	744.4124	744.4381	-35	1,2,3,4,5	1,2,3,4,5	✓	✓	✓
[11-62]	(R)ELKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTEYGLFQINNKIWCK	5909.7084	5909.83	-21					✓
[11-123]	(R)ELKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTEYGLFQINNKIWCKDDQNPHSSNIC NISCDKFLDDDLTDDIMCVKKILDKVGINYWLAHKALCSEKLDQWLCEKL	12951.00	12951.68	-53					✓
[14-16]	(K)DLK	374.2068	374.2165	-26		1,2	✓		✓
[14-58]	(K)DLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTEYGLFQINNK	5009.277	5009.35	-15			✓	✓	✓
[17-58]	(K)GYGGVSLPEWVCTTFHTSGY DTQAIVQNNDSTEYGLFQINNK	4652.988	4653.1394	-33			✓	✓	✓
[17-123]	(K)GYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTEYGLFQINNKIWCKDDQNPHSSNICNISCDKF LDDDLTDDIMCVKKILDKVGINYWLAHKALCSEKLDQWLCEKL	12224.001	12224.81	-66					✓
[54-58]	(F)QINNK	615.342	615.344	-3	1,2	1,2,3		✓	✓
[59-68]	(K)IWCKDDQNPH	1254.522	1254.55	-22	2, 1,	1,2,3		✓	✓
[59-79]	(K)IWCKDDQNPHSSNICNISCDK	2418.9744	2419.0355	-25	2,3,6,8,14,15,17,19,20	1,2,3,4,6,7,10,11,13,14,15, 16,17,18,19,20	✓		✓
[59-93]	(K)IWCKDDQNPHSSNICNISCDKFLDDDLTDDIM*CVK	4042.6134	4042.7515	-34	4,26	4,7,11,17,18	✓		✓
[59-94]	(K)IWCKDDQNPHSSNICNISCDKFLDDDLTDDIM*CVKK	4170.7155	4170.8465	-31	2,5,6,9,13,14,16,20,22, 26,29,30	2,5,6,7,9,12,15,31,34	✓		✓
[59-123]	(K)IWCKDDQNPHSSNICNISCDKFLDDDLTDDIMCVKKILDKVGINYWLAHKALCSEKLDQWLCEKL	7587.00	7586.79	28					
[63-79]	(K)DDQNPHSSNICNISCDK	1888.746	1888.77	-13	2,4,6,7	2,6,7			
[63-93]	(K)DDQNPHSSNICNISCDKFLDDDLTDDIMCVK	3512.421	3512.48	-17	4				
[63-108]	(K)DDQNPHSSNICNISCDKFLDDDLTDDIMCVKKILDKVGINYWLAHKALCSE	5307.447	5307.498	-10				✓	✓
[63-113]	(K)DDQNPHSSNICNISCDKFLDDDLTDDIMCVKKILDKVGINYWLAHKALCSEKLDQWL	5798.442	5798.58	-24				✓	✓
[63-123]	(K)DDQNPHSSNICNISCDKFLDDDLTDDIMCVKKILDKVGINYWLAHKALCSEKLDQWLCEKL	7067.9748	7067.37	86				✓	✓
[69-79]	(H)SSNICNISCDK	1182.488	1182.5	-10		2,3,6,7,8,9		✓	✓
[80-93]	(K)FLDDDLTDDIM*CVK	4189.7192	4189.72	0	7	3,10,11,12		✓	✓
[80-94]	(K)FLDDDLTDDIM*CVKK	1768.828	1768.82	5		12	✓		
[80-119]	(K)FLDDDLTDDIMCVKKILDKVGINYWLAHKALCSEKLDQWL	4707.009	4707.38	-79				✓	✓
[94-98]	(K)ILDK	615.3781	615.3955	-28	1,2,3,4	1,2,3,4	✓	✓	✓
[94-108]	(K)ILDKVGINYWLAHK	1796.9442	1797.0301	-48	1,4,6,8,9,11,14	2,3,4,6,7,9,12,13,14	✓	✓	✓
[94-123]	(K)ILDKVGINYWLAHKALCSEKLDQWLCEKL	3556.7919	3556.8938	-29	1,5,6,8	3	✓	✓	✓
[95-98]	(K)ILDK	487.30663	487.301	12			✓	✓	✓
[95-108]	(K)ILDKVGINYWLAHK	1668.8958	1668.9351	-24	2,3,4,5,6,7,8,9,11,12,13	2,3,4,5,6,7,8,9,10,11,12,13	✓	✓	✓
[95-123]	(K)ILDKVGINYWLAHKALCSEKLDQWLCEKL	3428.7249	3428.7988	-22	1,2		✓	✓	✓
[99-103]	(K)VGINY	564.29676	564.291	10	3,4,5,6,7,9	1,2,3,4,5,6,7,8,9	✓	✓	✓
[99-108]	(K)VGINYWLAHK	1199.603	1199.6451	-35	2,3,4,5,6,7,8,9,10,11,12, 13,14	1,2,3,4,5,6,7,8,9,10,11,12, 13,14	✓	✓	✓
[109-123]	(K)ALCSEKLDQWLCEKL	1777.8234	1777.8742	-29	13,14	13,14	✓	✓	✓
[115-123]	(K)LDQWLCEKL	1146.5306	1146.5743	-38	2,4	2,3,4,5,6,7,8	✓	✓	✓

Annex 4.2: List of peptides derived from β-cas hydrolysed by the three trypsins. (* indicates phosphoserine)

Peptides	Sequence	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	b fragments identified	y fragments identified	Bovine	Porcine	Human
[1-9]	RELEELNVP	1097.6232	1097.57	48			✓		
[1-25]	RELEELNVPGEIVESLSSSEESITR	2801.2532	2801.39	-49	4,5,6,7,8,12	2,4,5,7,12	✓	✓	✓
[1-28]	RELEELNVPGEIVES*LS*S*S*EESITRINK	3156.487	3156.61	-39	8,10	7,8	✓	✓	✓
[1-29]	RELEELNVPGEIVES*LS*S*S*EESITRINKK	3284.5735	3284.71	-42	8,10	7,8	✓	✓	✓
[1-97]	RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPP LTQTPVVVPPFLQPEVMGVSK	10915	10915	0			✓		
[1-99]	RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPP LTQTPVVVPPFLQPEVMGVSKVK	11149	11149	0					
[26-28]	(R)INK	373.2413	373.23	30		2		✓	✓
[26-29]	(R)INKK	501.347	501.328	38		1,2		✓	✓
[26-32]	(R)INKKIEK	871.634	871.549	98		2,3,4,5,6	✓	✓	✓
[26-209]	(R)INKKIEKFQSEEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK VKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQHQPLPPTVMFPPQSVLSLSQSK VLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	20878	20878	48					✓
[29-32]	(K)IEK	516.3382	516.33	16	1,3,4	2,3	✓	✓	✓
[29-48]	(K)IEKFQSEEQQQTEDELQDK	2559.33	2559.17	63	8	2,3,4,5,6,8	✓	✓	✓
[29-97]	(K)IEKFQS*EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK	7781	7781	0			✓	✓	✓
[29-99]	(K)IEKFQS*EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVK	8009	8009.17	-21					
[29-209]	(K)IEKFQSEEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKE AMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQHQPLPPTVMFPPQSVLSLSQSKVLP VPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	20524	20523	49					✓
[30-32]	(K)IEK	388.275	388.232	111		1,2			✓
[30-38]	(K)IEKFQSEEQ	1136.5778	1136.54	33	1	1,3	✓		
[30-48]	(K)IEKFQSEEQQQTEDELQDK	2431.202	2431.08	50	5	2,3,4,6,7,8		✓	✓
[33-48]	(K)FQS*EEQQQTEDELQDK	1980.8384	1980.85	-6		3,4,6,7,8	✓	✓	✓
[33-97]	(K)FQS*EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK	6513.0	6512.5	77			✓		
[33-99]	(K)FQS*EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVK	7359.22	7358.69	72			✓		
[49-56]	(K)IHPFAQTQ	940.49	940.477	20	2,4,5,6	6		✓	✓
[49-68]	(K)IHPFAQTQSLVYPFPGPIPN	2222.364	2222.15	96		3,4,5,6,7,8		✓	✓
[49-97]	(K)IHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK	5316.216	5315.85	69					
[49-105]	(K)IHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPK	6174.2492	6174.33	-13				✓	
[49-119]	(K)IHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPF	7897.456	7897.17	36	5,7	2,6,8,9,10			✓
[53-68]	(F)AQTQSLVYPFPGPIPN	1728.024	1727.89	78					✓
[69-97]	(N)SLPQNIPPLTQTPVVVPPFLQPEVMGVSK	3111.993	3111.7	94					✓
[95-141]	(G)VSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQ	5356.209	5355.83	71					✓
[98-105]	(K)VKEAMAPK	872.486	872.479	8	6	1,2,3,4,5,6,7	✓	✓	✓
[98-113]	(K)VKEAMAPKHKEMPFP	1867.06	1866.98	43	2	1,2,3,4,5,6,	✓	✓	✓
[98-134]	(K)VKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLH	4266.981	4267.13	-35			✓		
[98-209]	(K)VKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQHQPLPPTVMFPPQSVLSLSQ SKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	12680	12679.02	77					✓
[100-105]	(K)EAMAPK	645.3217	645.316	9	1,2,3,4	1,2,3,4,5	✓	✓	✓
[100-107]	(K)EAMAPKHK	910.4834	910.47	15	2	2,4,5,6,7	✓	✓	✓
[100-209]	(K)EAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQHQPLPPTVMFPPQSVLSLSQSK VLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	12452	12451.72	22					✓
[106-113]	(K)HKEMPFPK	1012.5278	1012.52	8	3,4,6	1,2,4,5,7	✓	✓	✓
[106-169]	(K)HKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQHQPLPPTVMFPPQSVLSLSQSK	7357	7357.55	-75			✓	✓	✓

Annex 4.2: Continued.

Peptides	Sequence	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	b fragments identified	y fragments identified	Bovine	Porcine	Human
[106-209]	(K)HKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	11824	11823.96	3					✓
[108-113]	(K)EMPFPK	747.3618	747.363	-2			✓	✓	✓
[108-169]	(K)EMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK	7092	7092.24	-34		2	✓	✓	✓
[114-163]	(K)YPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL	5728.428	5727.92	89					
[114-169]	(K)YPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK	6362	6362.34	-53			✓	✓	✓
[142-169]	(Q)SWMHQPHQPLPPTVMFPPQSVLSLSQSK	3183.831	3183.6	73		2,3,4,5,6,7	✓	✓	✓
[164-169]	(L)SLSQSK	648.408	648.344	99	2,3	2,3,4	✓	✓	✓
[170-176]	(K)VLPVPQK	779.4889	779.491	-3	5	2,3,4,5,6	✓	✓	✓
[170-183]	(K)VLPVPQKAVPYPQR	1590.9338	1590.92	9	2,9	1,4,5,7,8,9,10,12	✓	✓	✓
[170-209]	(K)AVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	4484.4729	4484.42	12			✓	✓	✓
[177-183]	(K)AVPYPQR	829.4421	829.445	-3	3	1,2,4,5	✓	✓	✓
[177-209]	(K)AVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV	3722.002	3722.46	-123			✓	✓	✓
[184-190]	(R)DMPIQAF	820.3878	820.379	11	4,5,	3,	✓	✓	✓
[184-202]	(R)DMPIQAFLLYQEPVLGPVR	2185.2002	2185.16	18			✓	✓	✓
[184-209]	(R)DMPIQAFLLYQEPVLGPVRGPFPIIV	2908.5834	2908.59	-2			✓	✓	✓
[191-202]	(F)LLYQEPVLGPVR	1382.8226	1382.79	24	3,4,5,7,8	2,3,4,5,6,7,8,9,10	✓	✓	✓
[203-209]	(R)GPFPIIV	741.4525	741.44	17	2,3,4,5,6,7		✓		✓

Annex 4.3: Molar sequence coverages of α-LA hydrolysates catalysed by (A) bovine, (B) porcine or (C) human trypsins at all DH values. The dotted line (······) indicates the initial protein concentration.
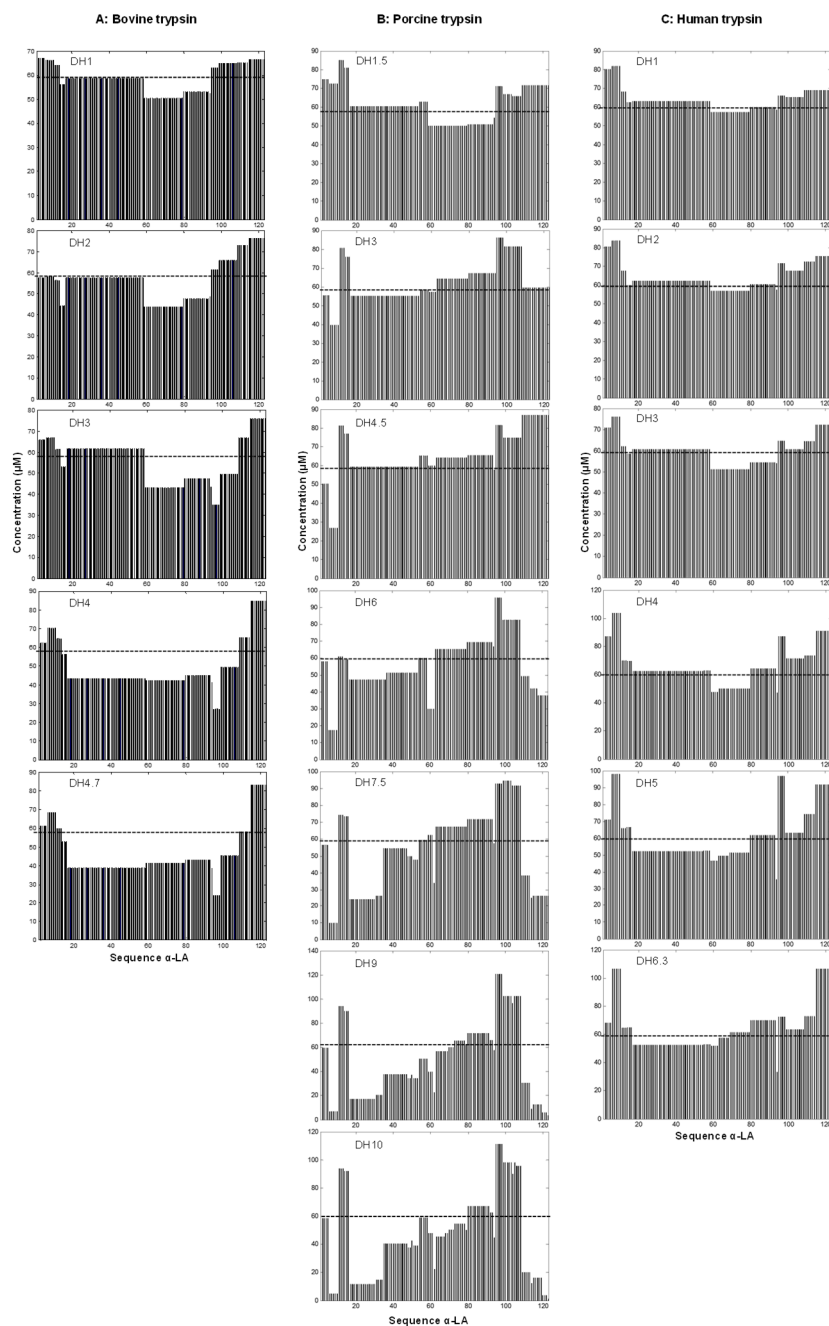
Annex 4.4: Molar sequence coverages of β-cas hydrolysates catalysed by (A) bovine, (B) porcine or (C) human trypsins at all DH values. The dotted line (······) indicates the initial protein concentration.
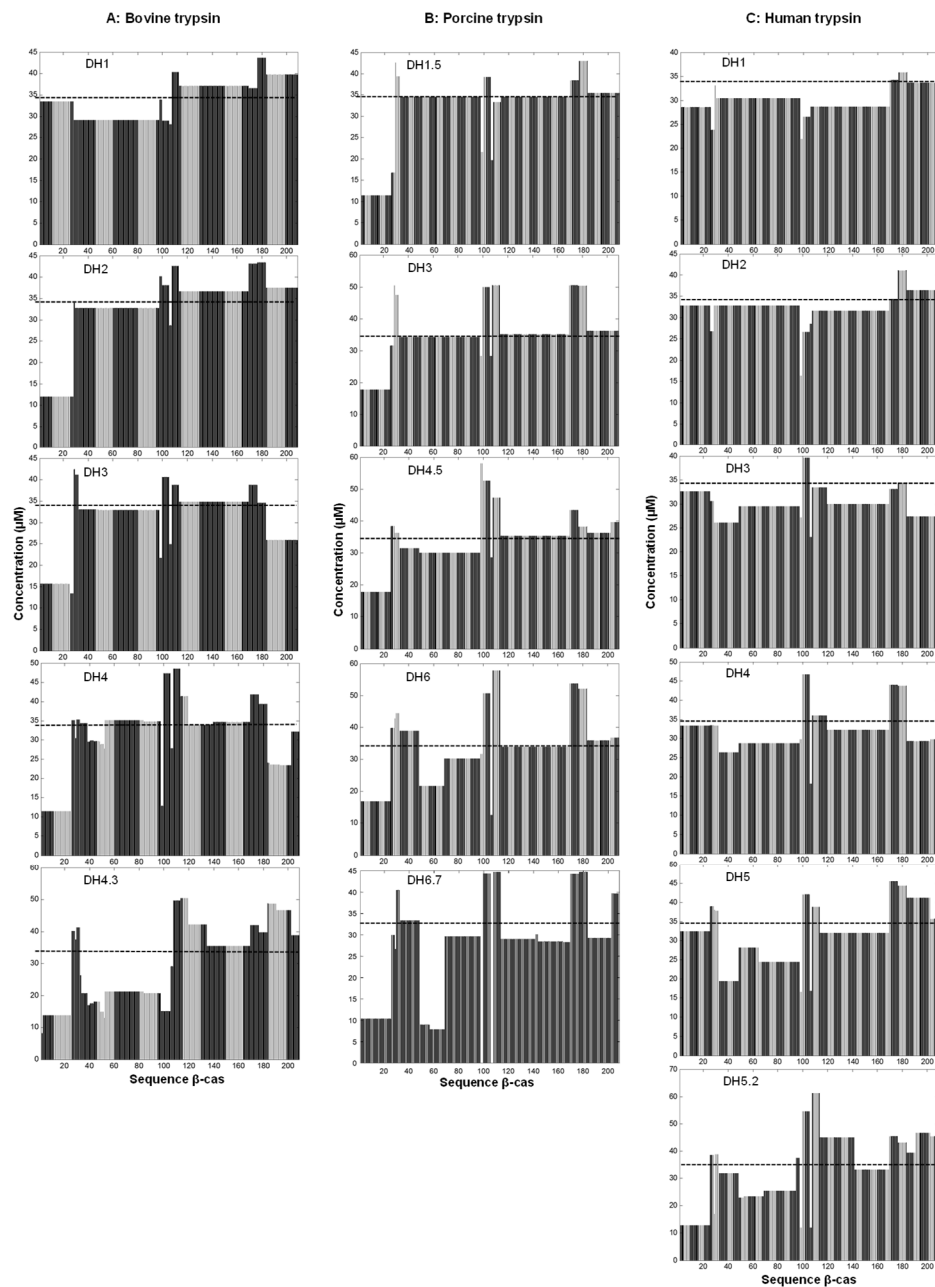
Annex 4.5: List of peptide concentrations at all DH values in α-LA hydrolysed by porcine or human trypsins

Porcine peptides	Concentration (µM)							Human peptides	Concentration (µM)					
	DH1.5	DH3	DH4.5	DH6	DH7.5	DH9	DH10		DH1	DH2	DH3	DH4	DH5	DH6.3
[1-5]	7.8	22.4	37.2	51.2	53.2	57.6	59.0	[1-5]	6.6	13.6	22.7	46.6	49.4	57.3
[1-10]	5.3	3.0	1.5	-	-	-	-	[1-10]	17.9	22.0	16.7	10.6	2.0	-
[1-13]	0.5	0.5	-	-	-	-	-	[1-13]	0.6	0.5	-	-	-	-
[1-16]	0.6	-	-	-	-	-	-	[1-16]	0.6	-	-	-	-	-
[1-58]	14.1	-	-	-	-	-	-	[1-58]	6.8	5.1	-	-	-	-
[1-123]	42.3	26.8	10.4	8.4	4.4	2.2	-	[1-123]	33.3	22.4	17.2	11.2	7.2	5.5
[6-10]	3.6	6.8	9.0	8.6	4.7	4.8	4.6	[6-10]	7.2	16.1	27.0	61.0	69.8	83.6
[6-13]	1.6	-	-	-	-	-	-	[11-13]	4.1	6.9	9.4	18.2	33.0	45.4
[11-13]	6.3	14.7	29.3	45.8	54.8	63.7	66.5	[11-16]	9.6	17.9	23.1	27.9	15.0	11.7
[11-16]	15.6	24.2	22.6	25.3	24.3	23.3	28.4	[11-123]	0.4	-	-	-	-	-
[11-62]	-	10.5	19.0	-	-	-	-	[14-16]	-	2.0	5.2	17.5	33.1	42.9
[14-16]	4.4	10.4	24.7	45.2	54.3	60.4	61.9	[14-58]	-	-	1.2	-	-	-
[17-58]	-	13.9	26.6	40.1	14.6	13.0	11.3	[17-58]	9.0	20.7	29.5	36.7	36.4	40.4
[54-58]	2.7	3.3	6.2	9.2	11.8	17.0	20.4	[17-123]	0.6	1.3	-	-	-	-
[59-93]	1.6	9.9	18.4	18.6	29.3	16.8	18.2	[59-68]	-	-	-	-	3.4	4.5
[59-94]	1.7	6.1	6.1	0.6	-	-	-	[59-79]	0.4	1.7	4.8	17.6	27.0	28.7
[63-108]	-	17.6	25.4	23.5	12.3	12.9	13.8	[59-93]	1.9	5.5	5.2	3.9	3.1	4.5
[63-113]	-	-	-	7.6	13.9	15.6	8.5	[59-94]	4.6	10.9	12.3	4.8	0.5	0.2
[63-123]	-	-	-	4.2	8.4	-	-	[59-123]	3.1	1.7	-	-	-	-
[80-93]	-	1.3	1.4	2.6	3.8	4.4	5.2	[63-79]	-	-	-	1.5	4.3	6.4
[80-94]	0.7	1.1	-	-	-	-	-	[63-93]	-	-	-	0.5	-	-
[80-119]	-	-	-	-	-	6.7	9.2	[69-79]	-	-	-	-	6.7	9.8
[94-98]	1.8	5.9	10.7	13.7	11.4	5.4	-	[80-93]	2.6	3.4	4.6	22.4	40.1	49.0
[94-108]	3.3	5.6	5.4	5.7	8.2	6.4	4.5	[80-94]	-	1.0	2.9	6.6	3.2	3.3
[95-98]	4.6	8.1	14.9	32.1	44.6	61.8	65.2	[94-98]	1.5	4.4	8.4	14.8	13.7	15.9
[95-108]	1.0	-	-	-	-	-	-	[94-108]	1.6	2.7	1.6	-	-	-
[95-123]	5.6	13.0	17.1	8.4	3.5	4.6	4.9	[95-98]	-	5.4	13.1	41.9	54.3	57.6
[99-103]	3.0	3.5	-	-	-	-	-	[95-108]	2.9	8.9	11.5	4.0	0.5	0.3
[99-108]	-	-	-	-	3.1	6.2	8.2	[95-123]	6.3	6.3	-	-	-	-
[109-123]	2.1	9.4	19.3	36.1	43.3	43.3	43.0	[99-108]	1.1	6.6	18.2	41.1	45.5	51.5
[115-123]	-	-	-	-	-	5.9	8.2	[109-123]	2.7	8.2	12.6	16.9	16.5	14.0
								[115-123]	-	0.5	1.5	3.2	3.2	5.0

"-" indicates the peptide was not found in the hydrolysate.
* The average standard error of the concentration of the peptides was ~15 %.

Annex 4.6: List of peptide concentrations at all DH values in β-cas hydrolysed by porcine or human trypsins

Porcine peptides	Concentration (µM)					Human peptides	Concentration (µM)					
	DH1.5	DH3	DH4.5	DH6	DH6.7		DH1	DH2	DH3	DH4	DH5	DH5.2
[1-25]	2.7	6.1	15.1	16.9	12.2	[1-25]	24.2	33.1	32.7	32.8	29.3	13.6
[1-28]	10.4	11.6	2.7	-	-	[1-28]	1.7	-	-	-	-	-
[1-29]	2.8	2.1	-	-	-	[1-29]	2.5	-	-	-	-	-
[26-28]	2.6	11.1	31.6	29.4	30.1	[26-28]	15.4	24.7	31.0	32.5	32.6	34.6
[26-29]	-	-	-	3.7	-	[26-29]	-	-	-	-	2.5	3.7
[26-32]	1.9	3.3	-	-	-	[26-209]	5.0	-	-	-	-	-
[29-32]	3.2	7.6	30.9	32.1	32.1	[29-32]	-	-	1.9	8.9	30.4	12.5
[29-97]	29.2	29.1	-	-	-	[29-48]	-	-	-	1.0	1.3	0.9
[30-32]	-	-	-	3.4	12.9	[29-97]	22.2	33.0	25.2	24.6	-	-
[30-48]	-	-	-	1.8	0.5	[29-209]	3.3	-	-	-	-	-
[33-48]	-	-	1.4	13.6	32.9	[30-32]	-	-	-	-	-	24.8
[33-99]	-	-	24.9	13.1	-	[30-48]	-	-	-	-	2.5	1.2
[49-56]	-	-	-	-	1.1	[33-48]	-	-	-	0.8	13.8	30.2
[49-68]	-	-	-	3.8	8.1	[49-68]	-	-	-	-	-	1.8
[69-97]	-	-	-	12.5	30.6	[49-97]	-	-	-	-	23.9	21.3
[98-105]	20.0	24.1	23.8	11.4	-	[49-119]	-	-	3.8	3.1	-	-
[98-113]	-	-	0.6	-	-	[53-68]	-	-	-	-	-	0.3
[100-105]	14.0	18.5	21.5	31.3	44.6	[69-97]	-	-	-	-	-	3.9
[100-107]	1.8	-	-	-	-	[95-141]	-	-	-	-	-	11.4
[106-113]	4.6	9.1	24.4	11.5	-	[98-105]	2.0	9.4	23.9	24.8	15.8	-
[106-169]	11.5	14.5	-	-	-	[98-209]	11.2	4.4	-	-	-	-
[108-113]	-	4.1	16.5	37.2	44.7	[100-105]	-	4.0	13.1	17.0	26.2	43.1
[108-169]	13.0	15.0	-	-	-	[100-107]	-	2.5	-	-	-	-
[114-163]	-	-	-	-	7.0	[100-209]	5.0	2.1	-	-	-	-
[114-169]	-	-	28.2	29.2	20.2	[106-113]	-	-	-	-	11.4	-
[142-169]	-	-	-	-	1.3	[106-169]	-	16.4	17.2	14.5	8.3	-
[164-169]	-	-	-	-	6.3	[106-209]	4.0	2.7	1.7	-	-	-
[170-176]	22.0	37.7	38.2	45.4	44.7	[108-113]	-	-	-	-	24.3	49.7
[170-183]	8.2	5.6	-	-	-	[108-169]	-	5.8	11.1	16.9	-	-
[170-209]	7.6	-	-	-	-	[114-169]	-	-	-	-	25.4	30.5
[177-183]	14.5	32.1	33.8	44.0	45.0	[142-169]	-	-	-	-	-	0.5
[177-209]	11.0	3.9	-	-	-	[170-176]	2.0	7.5	27.0	43.3	45.6	46.0
[184-202]	-	-	12.0	34.1	30.6	[170-183]	2.0	3.3	2.6	-	-	-
[184-209]	12.2	26.9	18.0	-	-	[170-209]	2.2	12.8	3.4	-	-	-
[203-209]	-	-	12.2	31.4	39.2	[177-183]	1.9	6.3	23.8	41.8	44.4	43.5
						[177-209]	2.0	7.5	4.7	1.0	-	-
						[184-202]	-	-	-	5.0	14.3	34.7
						[184-209]	1.8	5.7	19.0	29.7	25.4	-
						[191-202]	-	-	-	-	-	8.1
						[203-209]	-	-	-	1.4	10.4	41.0

"-" indicates the peptide was not found in the hydrolysate.
* The average standard error of the concentration of the peptides was ~15 %.

CHAPTER 5

Influence of substrate concentration on the extent of enzymatic protein hydrolysis

One difference in the published methods for *in vitro* protein digestion is the used substrate concentration. This may affect the outcome due to the fact that the experimental maximum degree of hydrolysis ($DH_{max,exp}$) by several enzymes, e.g. *Bacillus licheniformis* protease (BLP), was influenced by substrate concentrations used. This study aims to identify the reason of the influence of substrate concentration on hydrolysis by digestive proteases. Apo α-lactalbumin (0.05 %-10 %) was hydrolysed by bovine, porcine and human trypsins, bovine α-chymotrypsin or BLP at a constant enzyme to substrate ratio (E:S). Hydrolysis by human trypsin was not sensitive to changes in substrate concentration. The $DH_{max,exp}$ of hydrolysis by bovine and porcine trypsins, or bovine α-chymotrypsin increased with an increasing substrate concentration, while the $DH_{max,exp}$ of BLP hydrolysis decreased. Particularly, the $DH_{max,exp}$ of 0.1 % α-LA hydrolysis by bovine trypsin was 5 times lower than that of 1 %-10 % hydrolysis. Besides changes in $DH_{max,exp}$, also a change in hydrolysis mechanism, described by the percentage of remaining intact protein as a function of DH, was observed for hydrolysis with different substrate concentrations. This suggested that the concentration of individual peptides formed during hydrolysis changed with substrate concentrations. Assuming that some of these peptides could act as protease inhibitors, the observed decrease in $DH_{max,exp}$ could be explained. This hypothesis was tested and confirmed by performing the hydrolysis at a higher E:S. Increasing the E:S from 1:100 to 1:25 for 0.1 % α-LA hydrolysis by bovine trypsin indeed led to a similar $DH_{max,exp}$ reached by 1 % hydrolysis at an E:S of 1:100.

Introduction

One of the differences in published methods for *in vitro* protein digestion is the substrate concentration used, which ranged from 1 to 625 g/L [1-4]. The substrate concentration was previously shown to largely influence the experimental maximum degree of hydrolysis ($DH_{max,exp}$) by several proteases [5-9]. For example, the $DH_{max,exp}$ of whey protein isolate (WPI) hydrolysed by Alcalase increased by 50 % when the substrate concentration decreased from 30 % to 1 % at a constant enzyme to substrate ratio (E:S) [6]. This influence of substrate concentration was not expected because the amount of cleavage sites (CSs) per enzyme stayed constant. This study aims to identify the reason for the influence of substrate concentration on the extent of protein hydrolysis by digestive proteases.

The extent of protein hydrolysis, as described by the theoretical maximum degree of hydrolysis ($DH_{max,theo}$), can in first approximation be estimated from the specificity of the protease used and the amino acid (AA) sequence of the protein. In many cases, however, the measured $DH_{max,exp}$ was lower than the $DH_{max,theo}$ [10-12]. For instance, the $DH_{max,exp}$ of whey protein isolate (WPI, mainly contained 74 % β-lactoglobulin) hydrolysis by *Bacillus licheniformis* protease (BLP) ranged from 13 % to 6 % with an increasing substrate concentration from 0.1 % to 10 % [5]. The $DH_{max,exp}$ of hydrolysis with all concentrations were lower than the $DH_{max,theo}$ (16 %). In the BLP hydrolysis of 1 % WPI, the hydrolysis rate constants of CSs in β-lactoglobulin were found to vary from 0 to $1.1 \cdot 10^{-6}$ s^{-1}·mg$^{-1}_{enzyme}$ [13]. The relative hydrolysis rate constants of CSs in a protein were previously referred to as the enzyme selectivity [13]. Because BLP had a low selectivity towards half of the CSs in WPI [13], the $DH_{max,exp}$ did not reach the $DH_{max,theo}$. However, this cannot explain the decrease in $DH_{max,exp}$ with an increasing substrate concentration. This influence of substrate concentration on $DH_{max,exp}$ was also observed for WPI hydrolysis by Alcalase [6], and for wheat gluten hydrolysis by Flavourzyme [14]. This indicated that besides enzyme selectivity, there was another mechanistic change that affected the $DH_{max,exp}$. An indication for this mechanistic change was found in the WPI hydrolysis by Alcalase, where the plot of the percentage of remaining intact proteins against DH was different for the different substrate concentrations [6]. At the same DH, the hydrolysates with 1 % WPI contained less intact proteins than the hydrolysates produced by hydrolysis with higher substrate concentrations [6]. According to the Linderstrøm-Lang theory, the amount of remaining intact proteins as a function of DH illustrates the enzyme affinity towards intact proteins compared to its affinity towards intermediate peptides [15]. Two extreme scenarios, zipper and one-by-one, were distinguished in the Linderstrøm-Lang theory [15] (**Figure 5.1**). In the one-by-one mechanism, the enzyme affinity towards intact proteins is very low compared to its affinity towards intermediate peptides. In the zipper mechanism, the enzyme affinity towards intact proteins is the same as its affinity towards peptides [15]. Due to this difference in enzyme affinity, in the one-by-one mechanism at each point during hydrolysis there are remaining intact proteins, and final peptides, i.e. peptides that do not contain CSs. However, in the zipper mechanism at most points during hydrolysis, there is a gradual decrease of peptide length.
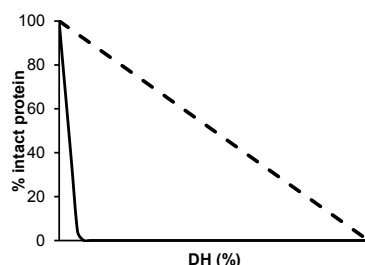
Figure 5.1: Theoretical scenarios of hydrolysis with (—) zipper and (——) one-by-one mechanisms, based on the Linderstrøm-Lang theory [15].

Since the enzyme affinities towards intact proteins were different in Alcalase hydrolysis with various substrate concentrations [6], the compositions of the hydrolysates during hydrolysis were also different. Changes in hydrolysate composition may affect the $DH_{max,exp}$, if the formation of certain peptides during the hydrolysis would inhibit the protease. Previously, the hydrolysates from many proteins were observed to have inhibitory activity for proteases, e.g. angiotensin-I converting enzyme (ACE). For instance, in bovine casein hydrolysates, 30 peptides were found to have inhibitory activity for ACE [16, 17]. The concentration of peptides needed to achieve 50 % inhibition (IC_{50}) varied from 2 to >$1·10^3$ µM for ~0.15 µM of ACE [16, 17]. This means that the molar ratio of inhibitory peptides to fully inhibit ACE varied from 26:1 to >1,300:1 [16, 17]. Changing the pH, temperature and E:S of the whey protein hydrolysis was found to change the ACE inhibitory activity of the hydrolysates [18]. This change in ACE inhibition can perhaps be the result of the different inhibiting peptides in the hydrolysates due to the different hydrolysate compositions. Besides ACE, inhibitory peptides for dipeptidyl peptidase IV (DPP-IV) have also been found in many protein hydrolysates. For example, changing the temperature, time and E:S of the milk protein isolate hydrolysis by Neutrase, the IC_{50} of the hydrolysates varied from 0.55 to 2.14 mg/mL for 3 nM of DPP-IV [19]. In addition, a hydrolysate of *Mytilus edulis* proteins by trypsin was found to have ~2 times more inhibitory activity for thrombin than the hydrolysates produced by pepsin or Neutrase [20]. If the peptides released during a hydrolysis have inhibitory activity for the enzyme used in that hydrolysis, adding fresh enzyme to the hydrolysate should increase the $DH_{max,exp}$. It was found that when the Alcalase hydrolysis of chick pea protein slowed down, adding fresh Alcalase indeed doubled the $DH_{max,exp}$ values [21]. Previously, bovine trypsin-1, porcine trypsin-1 and human trypsin-2 were found to have different enzyme selectivity towards CSs in two proteins, leading to large differences in $DH_{max,exp}$ (**Chapter 4**). Hydrolysis by these enzymes was shown to release different peptides at different moments. The question is whether these trypsins would have different sensitivity to inhibitory peptides.

In this study, the effects of substrate concentration on hydrolysis by digestive enzymes, i.e. bovine trypsin-1, porcine trypsin-1, human trypsin-2 or bovine α-chymotrypsin, were compared. BLP hydrolysis was also performed and compared with the reported data of WPI hydrolysis. To understand the differences in $DH_{max,exp}$ of hydrolysis with different substrate concentrations, for

each hydrolysis, the enzyme affinity towards intact proteins was determined, and correlated with the $DH_{max,exp}$.

Materials and methods

Materials

α-Lactalbumin was obtained from Davisco Foods International Inc. (Le Sueur, MN, USA). Based on the results from circular dichroism, ~72 % of the α-lactalbumin was in the apo form [10]. Bovine trypsin-1 (further referred to as bovine trypsin, BT, T1426) and porcine trypsin-1 (further referred to as porcine trypsin, PT, T0303), bovine α-chymotrypsin (further referred to as α-chymotrypsin, CT, C3142) and aprotinin (A6279) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Recombinant human trypsin-2 (further referred to as human trypsin, HT) was purchased from Shanghai Yaxin Biotechnology Co., Ltd (Shanghai, China). *Bacillus licheniformis* protease (BLP) (batch NS-37005) was provided by Novozymes (Bagsvaerd, Denmark). The properties of the 5 enzymes, including the protein content and purity, were determined (**Table 5.1**). All enzymes were essentially salt-free and freeze-dried. The three trypsins did not contain α-chymotrypsin, and the α-chymotrypsin did not contain trypsin, as determined by reversed phase ultra-high performance liquid chromatography coupled with mass spectrometry (RP-UPLC-MS). According to the specification from the manufacturers, the enzyme activities of bovine, porcine and human trypsins were ≥10,000, 13,000-20,000 and ≥7500 benzoyl-L-arginine-ethyl-ester units/mg protein, respectively. The α-chymotrypsin was specified to have ≥40 N-benzoyl-L-tyrosine-ethyl-ester units/mg protein. Since the BLP was partly insoluble, it was purified as described previously [22]. A suspension of BLP was prepared and then centrifuged (10 min, 4000×*g*, 25 °C), after which the supernatant was subsequently dialysed (cut-off 12-14 kDa) against 150 mM NaCl solution and subsequently demineralized water. The freeze-dried BLP had an activity of 3.9 AU mg^{-1} min^{-1} as determined by the azocasein assay [22]. Aprotinin, present in 0.9 % sodium chloride and 0.9 % benzylalcohol solution, was measured to have a concentration of 2.3 mg/mL based on the UV_{214} area using RP-UPLC [23]. All other chemicals were of analytical grade and purchased from Sigma or Merck.

Table 5.1: The protein content, purity, the number of cleavage sites (#CS), specificity and the fitted ($DH_{max,fit}$) and theoretical maximum degree of hydrolysis ($DH_{max,theo}$) of α-lactalbumin for the 5 enzymes.

Enzyme	Uniprot Code [1]	N-factor [g of protein/g N]	Protein content [w/w]	Purity[2] [%]	Specificity	#CS[1]	$DH_{max,theo}$ [%]	DH_{max}[3] [%]
BT	P00760	5.97	80 %	100	KR	13	10.7	7.0
PT	P00761	5.84	86 %	100	KR	13	10.7	10.8
HT	P07478	5.98	77 %	100	KR	13	10.7	8.2
CT	P00766	5.99	86 %	100	FLY[4]	21	17.2	9.3
BLP	P80057	5.93	60 %	100	DE	20	16.4	7.5

[1] From Uniprot (http://www.uniprot.org)

[2] The purity is defined as the percentage of UV_{280} peak area of the target protein out of total UV_{280} area from RP-UPLC.

[3] The DH_{max} was the highest $DH_{max,fit}$ of hydrolysis with different substrate concentrations at an E:S of 1:100.

[4] The specificity of α-chymotrypsin was taken from the three most frequently hydrolysed AAs in MEROPS database [24].

BT: bovine trypsin; PT: porcine trypsin; HT; human trypsin; CT: α-chymotrypsin; BLP: *Bacillus licheniformis* protease

Methods

Preparation of apo α-LA

The preparation of apo α-lactalbumin was performed using the method described previously (**Chapter 4**). Ethylenediaminetetraacetic acid (EDTA) was used to chelate the calcium ions attached to the molecules. EDTA was added to 1 L of 10 % [w/v] protein solution to obtain a protein to EDTA ratio of 1:5 [mol/mol]. The solution was stirred overnight at 4 °C, and filtered over a UFP-10-C-5 membrane (cut-off 10 kDa, AG Technologies Corp., Tokyo, Japan) with 7 L of Millipore water. Then, the protein powder was freeze dried and stored at -20 °C. The prepared apo α-lactalbumin was further referred to as α-LA. The protein content of the α-LA powder was ~90 % based on the Dumas result (N×6.25, based on the AA composition of α-LA as described in Uniprot accession code P00711, www.uniprot.org). Of the total protein content, ~90 % was α-LA, based on the proportion of α-LA UV_{214} peak area of the total UV_{214} area using RP-UPLC [25]. The other proteins present were β-lactoglobulin and bovine serum albumin.

Enzymatic hydrolysis of the proteins

The enzymatic hydrolysis of α-LA by each enzyme was performed as described previously [13]. The protein was dissolved in Millipore water at a concentration of 10 % [w powder/v]. The solution was stirred at 4 °C overnight and diluted to different substrate concentrations (0.05 %-5 %). The protein solution was adjusted to pH 8.0 and was equilibrated for 0.5 h at 37 °C in a pH-stat device (Metrohm, Herisau, Switzerland). Two sets of hydrolysis experiments were performed at an E:S of either 1:100 or 1:25 [w/w] at 37 °C. For hydrolysis at an initial E:S of 1:25, the hydrolysis by all enzymes was performed with a substrate concentration of 0.1 % [w/v powder], using 0.02 M NaOH to keep the pH constant for 2 h. No samples were taken in these experiments. For the hydrolysis at an E:S of 1:100, hydrolysis by these enzymes was performed with substrate concentrations of 0.1, 0.5, 1.0, 5.0 and 10.0 % [w/v powder], using 0.02, 0.1, 0.2, 1.0 and 2.0 M NaOH to keep the pH constant for 3 h, respectively. Hydrolysis by porcine and human trypsins was also performed with a substrate concentration of 0.05 % [w/v powder] using 0.01 M NaOH to keep the pH constant. Porcine tryptic hydrolysis was also performed with 2 % [w/v powder] α-LA using 0.4 M NaOH to keep the pH constant. During these experiments, 200 μL sample at 5 to 8 DH points was taken. In addition, for 0.1 % α-LA hydrolysis by bovine trypsin, after 3 h of hydrolysis, 40 μL of bovine trypsin solution (10 mg/mL powder) was added to the hydrolysate to reach a final E:S of 1:25. The hydrolysis was monitored for another 3 h and samples were taken in between. The degree of hydrolysis was calculated using equation 1 [26],

(1) $$DH_{exp}^{enzyme, conc.} \left[\% \right] = V_b \times N_b \times \frac{1}{\alpha} \times \frac{1}{m_p} \times \frac{1}{h_{tot}} \times 100 \ \%$$

where V_b [mL] is the volume of NaOH added; N_b [mol/L] is the normality of NaOH; α is the average degree of dissociation of the α-NH group ($1/\alpha$=1.3 at 37 °C and pH 8) [27]; m_p [g] is the mass of protein in the solution (taking into account the protein content of α-LA based on Dumas results); h_{tot} [mmol/g] is the total number of millimoles of peptide bonds per gram of protein substrate

(8.6 mmol/g for α-LA). The consumption of NaOH due to the dissolution of CO_2 in the protein solutions during the hydrolysis needs to be corrected. Therefore, as control experiments, 0.05 %-10 % [w powder/v] of α-LA solutions were incubated in a pH-stat in the absence of enzyme for 3 h. The added volume of NaOH in each control experiment was subtracted from the added volume of NaOH at all time points during hydrolysis. The 0.1 % and 0.5 % hydrolysis by α-chymotrypsin at an E:S of 1:100 were performed in triplicate. The average standard error in the $DH_{max,exp}$ among triplicate hydrolyses was ~6 %. To inactivate the three trypsins, aprotinin (2.3 mg/mL) was directly added to each sample to reach a final molar ratio of trypsin to aprotinin of 1:1.5. For samples taken during α-chymotrypsin hydrolysis, aprotinin was directly added to reach a molar ratio of α-chymotrypsin to aprotinin of 1:3. To confirm the inactivation of the trypsins and α-chymotrypsin, 150 or 300 µL aprotinin was added to a 10 mL α-LA solution immediately after each trypsin or α-chymotrypsin was added, respectively. The pH of the protein solution with each enzyme remained constant after the addition of aprotinin, proving that all hydrolyses were successfully inhibited (data not shown). The BLP was inactivated by adjusting the pH of the solutions to pH 2 with 5 M HCl, as described previously [27]. After ±10 minutes, the pH was re-adjusted to pH 8 with NaOH before storage at -20 °C [27]. The maximum degree of hydrolysis ($DH_{max,fit}^{enzyme,\ conc.}$) and the overall hydrolysis rate constant ($k_{DH}^{enzyme,\ conc.}$ [s^{-1}] = $k_{hydr}^{enzyme,\ conc.}$ × $DH_{max,fit}^{enzyme,\ conc.}$) were determined by fitting equation 2 (2nd order reaction equation) to experimental DH curves, as described previously [10],

$$(2)\quad DH_{exp}^{enzyme,\ conc.}\ [\%] = DH_{max,\ fit}^{enzyme,\ conc.} - DH_{max,\ fit}^{enzyme,\ conc.}\ /(1 + k_{hydr}^{enzyme,\ conc.} \times DH_{max,\ fit}^{enzyme,\ conc.} \times t)$$

where the superscript refers to the enzyme-substrate concentration combination. For each enzyme, the highest $DH_{max,fit}^{enzyme,conc.}$ of the hydrolysis reached (for any of the concentrations) at an E:S of 1:100 was taken as the DH_{max}^{enzyme} of the hydrolysis by that enzyme. Based on the DH_{max}^{enzyme}, the percentage of efficiently hydrolysed CSs in α-LA by each enzyme (also referred to as the high selectivity sites, abbreviated as HSSs) could be estimated using equation 3 [10].

$$(3)\quad HSS_{protein}^{enzyme}\ [\%] = \frac{DH_{max}^{enzyme}}{DH_{max,\ theo}^{enzyme}} \times 100\ \%$$

Test for trypsin activity under hydrolysis conditions

The activity of the trypsins during the 3 h hydrolysis was determined by measuring the release of p-nitroaniline from the Nα-Benzoyl-L-arginine-4-nitroanilide hydrochloride (BAPNA) during incubation with trypsins. The release of p-nitroaniline was measured in a 96 well plate using a plate reader Tecan Infinite F500 (Tecan, Männedorf, Switzerland). For each trypsin, enzyme solutions with concentrations of 1.0 mg/mL (trypsin concentration in 10 % α-LA hydrolysis) and 0.01 mg/mL (trypsin concentration in 0.1 % α-LA hydrolysis) were incubated at 37 °C at pH 8.0 for 3 h. Samples were taken at 0, 1, 2, and 3 h of incubation. Samples (100 µL) were added to 200 µL of 0.025 g/L BAPNA. The amount of absorbance of p-nitroaniline, released during the tryptic hydrolysis of BAPNA, was measured at 405 nm every 3 min for 4 h, in duplicate.

Reversed phase ultra-high performance liquid chromatography (RP-UPLC)

The hydrolysates were analysed on an H class Acquity UPLC® system (Waters, Milford, MA, USA) equipped with a BEH C18 column (1.7 μm, 2.1×100 mm, Waters). An Acquity UPLC® PDA detector (Waters) attached to the UPLC system. The eluents and elution gradient were the same as described previously [13]. Samples were incubated for 2 h with 100 mM ditiothreitol (DTT) in 50 mM Tris-HCl buffer at pH 8.0 to reduce the disulfide bridges. For samples taken during hydrolysis with each substrate concentration, the amount of DTT added was to a molar ratio of 5:1 to the amount of cysteine in the samples. The reduced samples were further diluted to 0.05 % [w/v] and centrifuged (10 min, 14,000×g, 20 °C) before injection (4 μL).

Electron spray ionisation time of flight mass spectrometry (ESI-Q-TOF-MS)

The mass spectra of the peptides were collected using an online SYNAPT G2-Si high definition mass spectrometry (Waters) coupled to the UPLC system. Sodium iodide was used for the MS calibration. The MS and MS/MS (MSe) data were collected according to methods developed previously [13]. The data were analysed using the MassLynx software v4.1 (Waters).

Determination of the concentration of the intact α-LA in hydrolysates

Intact α-LA was identified using the MS spectra. The concentration of intact α-LA in each hydrolysate was calculated based on the UV_{214} using equation 4 [13]:

$$(4) \quad C_{protein} \left[\mu M \right] = \frac{A_{214} \cdot Q}{\varepsilon_{214} \cdot l \cdot V_{inj} \cdot k_{cell}}$$

where C [μM] is the protein concentration, A_{214} [μAU·min] is the UV area at 214 nm, V_{inj} [μL] is the injected volume, Q [μL·min^{-1}] is the flow rate and l [cm] is the path length of the UV cell, which is 1 cm according to the manufacturer. ε_{214} [M^{-1}·cm^{-1}] is the molar extinction coefficient at 214 nm, which was 300,461 M^{-1}·cm^{-1} for α-LA calculated as described previously [28]. Because of the multiple reflection by the coating of the UV cell, the effective path length of the light through the cell is not the same as the length specified by the producer. To correct for this effect, the cell constant of the UV detector (k_{cell}), was determined using a series of standard solutions made by β-lactoglobulin, β-casein and angiotensin II, and the approach as described elsewhere [29]. The ratio between the measured and the expected A_{214} was the value of k_{cell}. The k_{cell} was determined to be 0.78 for the UV cell used in this study [23]. For hydrolysis by all enzymes with all substrate concentrations, the average concentration of intact α-LA in samples at DH 0 % was 95±9 % of the expected value (calculated based on the injected protein content, protein purity and amount of inhibitor added).

Determination of the enzyme affinity towards intact proteins

Based on the Linderstrøm-Lang theory, the percentage of intact protein against DH indicates the enzyme affinity towards intact proteins relative to its affinity towards intermediate peptides. Because the hydrolysis by the 5 enzymes did not reach the same DH_{max}^{enzyme}, the DH was normalised

using DH/ DH_{max}^{enzyme} [%]. For hydrolysis by each enzyme, the percentages of remaining intact protein as a function of DH/ DH_{max}^{enzyme} for the theoretical scenarios of zipper and one-by-one hydrolysis mechanisms were calculated using equation 5 and 6, respectively.

$$(5) \quad \%intact_{zipper}\left[\%\right]=100-\#HSS^{enzyme}\times\frac{DH}{DH_{max}^{enzyme}}$$

$$(6) \quad \%intact_{one-by-one}\left[\%\right]=100-\frac{DH}{DH_{max}^{enzyme}}$$

#HSS[enzyme] refers to the number of HSSs of each enzyme towards CSs in the protein. #HSS[enzyme] was calculated by multiplying the number of CSs by the percentage of HSSs estimated based on DH_{max}^{enzyme} using equation 3. Assuming that the enzyme affinities towards all peptides were comparable, the enzyme affinity toward intact proteins ($0<P^{intact}\leq 1$) was calculated from the initial slope ($IS_{conc.}$) of the curve, i.e. the percentage of intact protein *versus* DH/ DH_{max}^{enzyme}, using equation 7.

$$(7) \quad P^{intact}=\frac{IS_{one-by-one}-IS_{conc.}}{IS_{one-by-one}-IS_{zipper}}$$

$IS_{conc.}$ is the initial slope of the curve of hydrolysis with each substrate concentration. $IS_{one-by-one}$ and IS_{zipper} refer to the initial slopes of the curves of theoretical scenarios of zipper and one-by-one hydrolysis mechanisms obtained from equations 5 and 6, respectively. For each enzyme, the P^{intact} of the hydrolysis that reached DH_{max}^{enzyme} was set as the reference P^{intact}, referred to as the P_{max}^{intact} [%]. Because the hydrolysis by each enzyme also had a different P_{max}^{intact}, the P^{intact} was also normalised using P^{intact}/ P_{max}^{intact} [%]. Therefore, the changes in hydrolysis mechanism by different substrate concentrations were shown by plotting P^{intact} / P_{max}^{intact} *versus* $DH_{max,fit}^{enzyme,conc.}$ / DH_{max}^{enzyme}.

Results and discussion

Check for trypsin activity under hydrolysis conditions

Enzyme activity was determined using BAPNA for all trypsins after incubating the enzyme alone (in 1.0 mg/mL or 0.01 mg/mL) for 3 h. The activity after incubation was found to be the same as for the fresh enzyme (data not shown). This indicated that autolysis of the trypsins did not occur during hydrolysis under these conditions, and the enzyme remained active during the hydrolysis.

Influence of substrate concentration on the extent of hydrolysis

The hydrolysis curves (DH *versus* time) for all enzymes with different substrate concentrations were compared (**Figure 5.2**). Only for porcine tryptic hydrolysis, the hydrolysates of 10 % and 5 % α-LA formed a gel during the hydrolysis, which hindered the sampling of the hydrolysates. In addition, the hydrolysate of 2 % α-LA by porcine trypsin became turbid during the hydrolysis, indicating the formation of aggregates. For the effect of substrate concentration on $DH_{max,exp}$, the responses of the 5 enzymes were clearly different. Hydrolysis by human trypsin was not affected

by changing substrate concentrations (**Figure 5.2C**). The $DH_{max,exp}$ of hydrolysis by bovine and porcine trypsins, or α-chymotrypsin increased with an increasing substrate concentration (**Figures 5.2A, 5.2B** and **5.2D**). The $DH_{max,exp}$ hydrolysis by BLP decreased with an increasing concentration (**Figures 5.2E**).
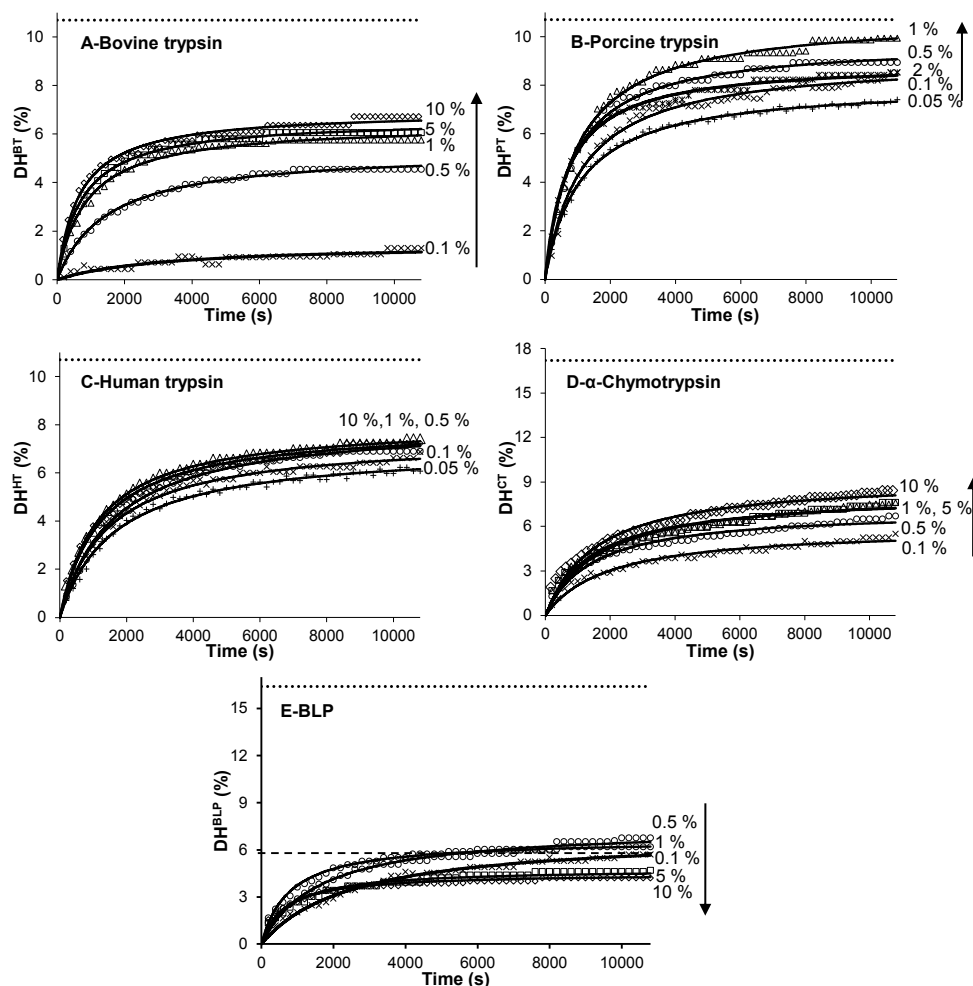

Figure 5.2: Hydrolysis curves (DH *versus* time) of α-LA by (A) bovine, (B) porcine and (C) human trypsins, (D) bovine α-chymotrypsin or (E) *Bacillus licheniformis* protease (BLP) with substrate concentrations of (◇) 10 %, (□) 5 %, (✳) 2%, (△) 1%, (○) 0.5 %, (×) 0.1 % or (+) 0.05 %. The solid lines (—) indicate the fit using the 2nd order reaction equation. The dotted lines (······) refer to the theoretical maximum degree of hydrolysis ($DH_{max,exp}$) of α-LA by each enzyme. The dashed line (------) in panel (E) indicates the $DH_{max,theo}$ of BLP hydrolysis when only CSs after glutamic acids were considered as CSs.

The $DH_{max,fit}$ values of all hydrolyses were obtained from fitting the hydrolysis curves with 2nd order reaction equation (**Figure 5.3A**). For human tryptic hydrolysis, the $DH_{max,fit}$ of hydrolysis

with all substrate concentrations were similar (7.8±0.5 %), showing that human tryptic hydrolysis was not sensitive to substrate concentrations in the range tested. For bovine tryptic hydrolysis, the $DH_{max,fit}$ increased drastically from 1.5 % to 7.0 % with an increasing substrate concentration (0.1 %-10 %). For porcine tryptic hydrolysis with an increasing substrate concentration (0.05 %-1 %), the $DH_{max,fit}$ increased from 8.1 % to 10.8 %. The $DH_{max,fit}$ of α-chymotrypsin hydrolysis increased from 5.9 % to 9.3 % with an increasing substrate concentration (0.1 %-10 %). For BLP hydrolysis, the $DH_{max,fit}$ decreased (7.5 %-4.4 %) with an increasing substrate concentration (0.1 %-10 %). This decrease was similar to previous results of the BLP hydrolysis of WPI [5]. These results confirmed that the influence of substrate concentration on $DH_{max,exp}$ is not unique to a substrate. Both for bovine trypsin and BLP, the transition is around 1 % substrate concentration, but the $DH_{max,fit}$ of bovine tryptic hydrolysis decreased for substrate concentrations ≤1 %, while the BLP hydrolysis decreased for substrate concentrations ≥1 % (**Figure 5.3A**).
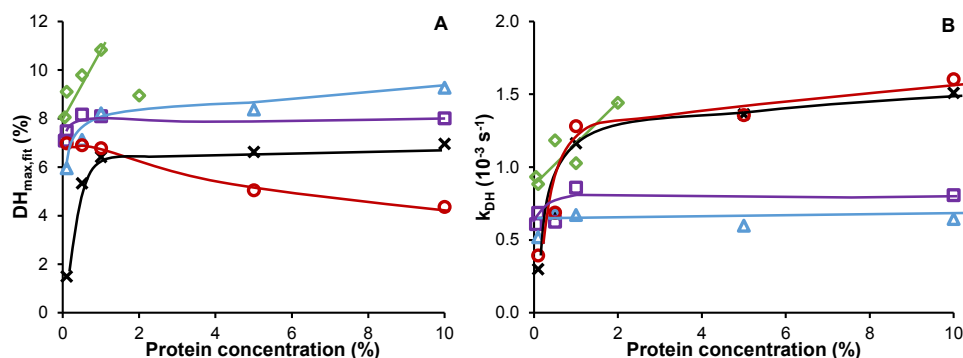
Figure 5.3: (A) Fitted maximum degrees of hydrolysis ($DH_{max,fit}$) and (B) the overall hydrolysis rate constants (k_{DH}) for α-LA hydrolysis by (×) bovine, (◇) porcine and (□) human trypsins, (△) α-chymotrypsin or (○) *Bacillus licheniformis* protease, as a function of substrate concentration [%]. The porcine tryptic hydrolysis with 2 % substrate concentration showed turbidity during hydrolysis.

The highest $DH_{max,fit}^{enzyme,conc.}$ values reached by each enzyme in this set of experiments were taken as the DH_{max}^{enzyme} values. The hydrolysis that had this DH_{max}^{enzyme} was assumed not to be inhibited by the formed peptides, and the DH_{max}^{enzyme} was assumed to be purely determined by enzyme selectivity. For bovine, porcine and human trypsins, based on their specificity, the $DH_{max,theo}$ is 10.7 % (**Table 5.1**). For hydrolysis by the three trypsins, only the DH_{max}^{PT} of porcine tryptic hydrolysis (10.8 %) was similar to the $DH_{max,theo}$. The $DH_{max}^{BT/PT}$ of bovine or porcine tryptic hydrolysis were 7.0 % and 8.2 %, respectively, corresponding to ~65 % and ~77 % of the $DH_{max,theo}$. The different DH_{max}^{enzyme} reached by the three trypsins were previously explained by their differences in the enzyme selectivity (**Chapter 4**). Both bovine and human trypsins were reported to be hindered by the charged AAs surrounding the CSs, while porcine trypsin was less sensitive to charged AAs surrounding the CSs (**Chapter 4**). The DH_{max}^{CT} of α-chymotrypsin hydrolysis was 9.3 %, corresponding to ~54 % of the $DH_{max,theo}$ (17.2 %). The DH_{max}^{BLP} of BLP hydrolysis (7.5 %) was ~46 % of the $DH_{max,theo}$ (16.4 %). Previously for β-lactoglobulin hydrolysis by BLP, it was observed that

BLP hydrolysed the CSs with a hydrolysis rate constant varying from 0 to $6\cdot10^{-3}$ $s^{-1}\cdot mg^{-1}{}_{enzyme}$ [13]. This could explain why the $DH_{max,exp}$ of that hydrolysis was only 50 % of the $DH_{max,theo}$ [13]. It was found that although the CSs after glutamic acids were hydrolysed with various rates, on average, BLP hydrolysed CSs after glutamic acids ~1,000 times faster than CSs after aspartic acids [13]. If only CSs after glutamic acids were considered as CSs in α-LA, then the $DH_{max,theo}$ would be 5.7 %. This means in the case of α-LA hydrolysis by BLP, some CSs after aspartic acids were also cleaved.

For the overall hydrolysis rate constants ($k_{max,fit}^{BT/PT/HT/CT,conc.}$), the effect of substrate concentration on hydrolysis by the three trypsins and α-chymotrypsin was similar as for the $DH_{max,fit}^{BT/PT/HT/CT,conc.}$ (**Figure 5.3B**). Only for hydrolysis by BLP, the $DH_{max,fit}^{BLP,conc.}$ decreased with an increasing substrate concentration, while the $k_{DH}^{BLP,conc}$ increased with an increasing concentration. This was in contrast with previously reported data, where the overall hydrolysis rate constant of whey protein isolate hydrolysis by BLP decreased with an increasing substrate concentration [5]. It is not clear what the differences were between the two experiments.

Based on the DH_{max}^{enzyme} of hydrolysis by these enzymes, the numbers of efficiently hydrolysed CSs were estimated to be 8, 13, 10, 11 and 9 for hydrolysis by bovine, porcine and human trypsins, α-chymotrypsin and BLP, respectively. These estimated numbers were used to calculate the enzyme affinity towards intact proteins of the hydrolysis having a zipper mechanism by each enzyme.

Enzyme affinity towards intact proteins

The enzyme affinities towards intact proteins (P^{intact}) were analysed by plotting the percentage of remaining intact α-LA as a function of DH/ DH_{max}^{enzyme} [%] (**Figure 5.4** and **Table 5.2**). For hydrolysis by α-chymotrypsin, in all hydrolysates, no intact proteins were found (**Figure 5.4D**), indicating that the P^{intact} of the hydrolysis was extremely high (P^{intact}≥0.9). For porcine and human tryptic hydrolysis, the P^{intact} of hydrolysis was overall not affected by the changes in substrate concentration (**Figures 5.4B** and **5.4C**). Only for the hydrolysis of 0.05 % α-LA, it seems that the P^{intact} was closer to the zipper mechanism. Larger differences in P^{intact} were observed in bovine trypsin and BLP hydrolyses. For bovine tryptic hydrolysis at 0.1 % α-LA, the experimental curve was almost similar to the curve of the one-by-one scenario (P^{intact}≈0). This means that the affinity of bovine trypsin towards intact proteins in 0.1 % α-LA hydrolysis was extremely low (P^{intact}=0.1). With an increasing substrate concentration, the P^{intact} of bovine tryptic hydrolysis increased (**Figure 5.4A**). For BLP hydrolysis, the P^{intact} decreased with an increasing substrate concentration (**Figure 5.4E**), which was the opposite of the bovine tryptic hydrolyses. However, because the $DH_{max,exp}$ of the two hydrolyses were also oppositely influenced by an increasing substrate concentration, the P^{intact} followed the same trend/correlation with the $DH_{max,fit}$.
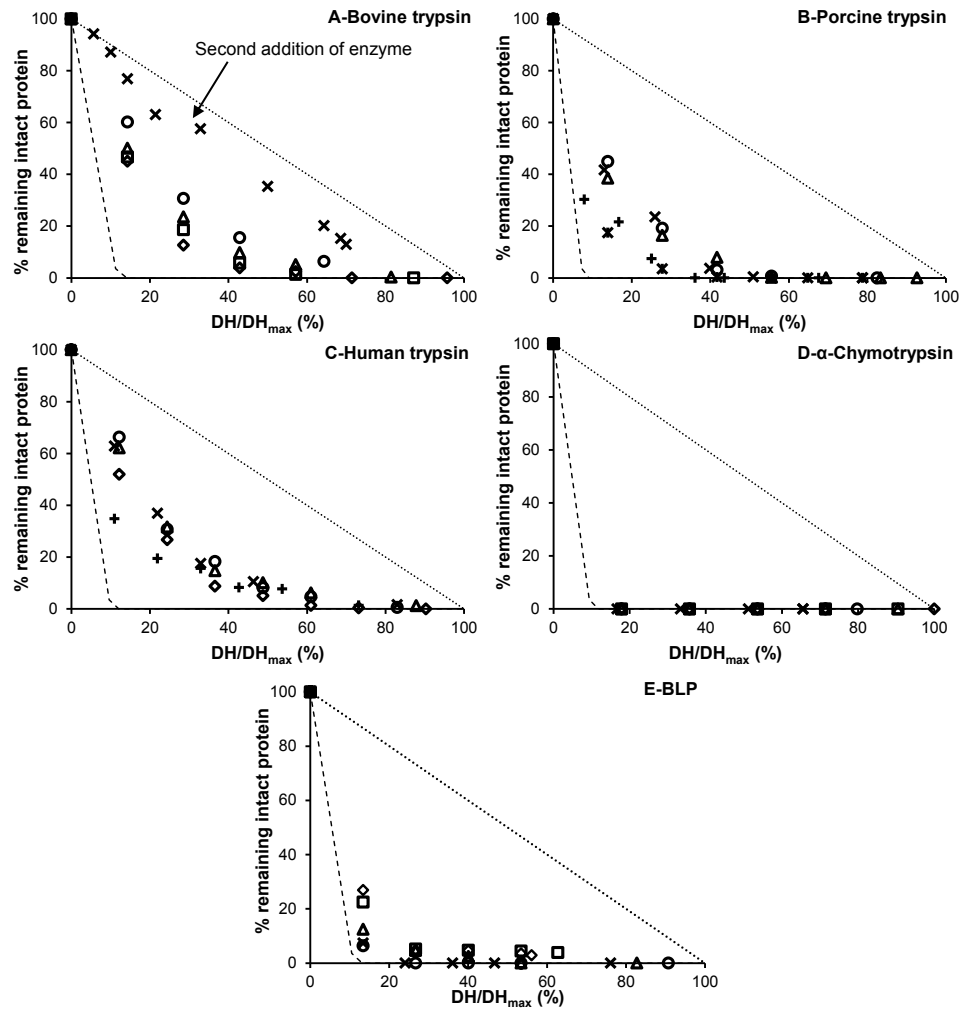
Figure 5.4: The percentage of remaining intact protein in α-LA hydrolysates during hydrolysis by (A) bovine, (B) porcine and (C) human trypsins, (D) bovine α-chymotrypsin or (E) *Bacillus licheniformis* protease (BLP) with substrate concentrations of (◇) 10 %, (□) 5 %, (✳) 2 %, (△) 1 %, (○) 0.5 %, (×) 0.1 % or (+) 0.05 %, as a function of DH/DH$_{max}$ [%]. The dotted lines (·····) and dashed lines (------) indicate the theoretical scenarios of one-by-one and zipper hydrolysis mechanisms, respectively. In panel (A), the second addition of bovine trypsin to the hydrolysate of (×) 0.1 % substrate concentration as a function of DH/DH$_{max}$ was indicated.

Table 5.2: Fitted maximum degree of hydrolysis ($DH_{max,fit}$) and enzyme affinity towards intact proteins (P^{intact}) of α-LA hydrolysis by bovine, porcine and human tryptins, α-chymotrypsin, or BLP with substrate concentrations of 0.05 % to 10 %.

Conc.[%]	Bovine trypsin		Porcine trypsin		Human trypsin		α-Chymotrypsin		BLP	
	$DH_{max,fit}$ [%]	P^{intact}	$DH_{max,fit}$ [%]	P^{intact}	$DH_{max,fit}$ [%]	P^{intact}	$DH_{max,fit}$ [%]	P^{intact}	$DH_{max,fit}$ [%]	P^{intact}
0.05	-	-	8.1	0.20	7.1	0.25	-	-	-	-
0.1	1.5	0.10	9.1	0.16	7.5	0.20	5.9	≥0.9	7.0	0.73
0.5	5.3	0.19	9.8	0.16	8.2	0.20	7.1	≥0.9	7.5	0.74
1	6.4	0.22	10.8	0.16	8.1	0.20	8.2	≥0.9	6.7	0.68
2	-	-	8.9*	0.20	-	-	-	-	-	-
5	6.6	0.24	*	-	-	-	8.4	≥0.9	4.8	0.59
10	7.0	0.26	*	-	8.0	0.20	9.3	≥0.9	4.4	0.55

Not determined values were represented by "-";

BLP: *Bacillus licheniformis* protease

* Porcine tryptic hydrolysates of 10 % and 5 % α-LA formed a gel during hydrolysis and the hydrolysate of 2 % was turbid.

To compare the correlation between the P^{intact} and $DH_{max,fit}$ of hydrolysis by different enzymes (**Table 5.2**), both parameters were normalised. The α-chymotrypsin hydrolysis was excluded from this analysis since no intact proteins were found in any hydrolysates. The normalised P^{intact} (P^{intact} / P_{max}^{intact}) of hydrolysis by the 4 enzymes were plotted as a function of the normalised $DH_{max,fit}$ ($DH_{max,fit}^{enzyme,conc.}$ / DH_{max}^{enzyme}) (**Figure 5.5**). For bovine trypsin and BLP hydrolyses, a linear correlation (R^2=0.93) was found between the normalised $DH_{max,fit}$ and the normalised P^{intact} (**Figure 5.5A**). This suggested that more peptides with inhibitory activity for the enzyme were formed in hydrolysis with a low P^{intact}.
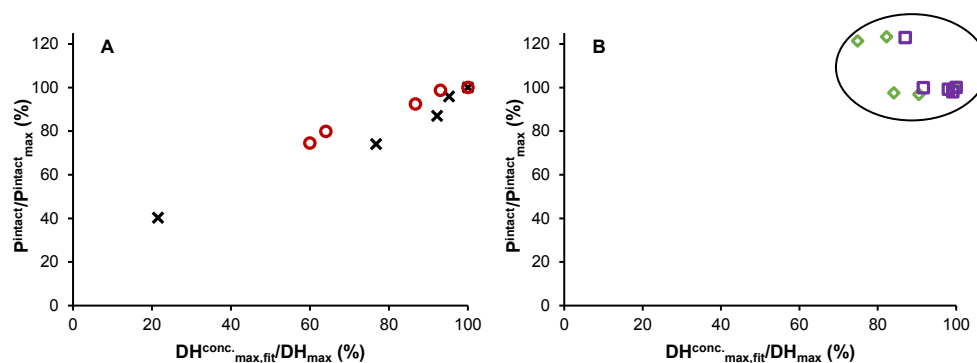
Figure 5.5: Normalised P^{intact} (P^{intact} / P_{max}^{intact} [%]) *versus* normalised $DH_{max,fit}$ ($DH_{max,fit}^{enzyme,conc.}$ / DH_{max}^{enzyme} [%]) of hydrolysis by (A) (×) bovine trypsin or (○) BLP and by (B) (◇) porcine and (□) human trypsins with different substrate concentrations.

This means that besides the enzyme secondary specificity, also the P^{intact}, through effects that are not fully understood at this moment, quantitatively determines the $DH_{max,exp}$ of a hydrolysis. The possible explanation of this effect could be the formation of inhibitory peptides for the protease during the hydrolysis.

Indications of the existence of inhibitory peptides

To test the effect of the formation of inhibitory peptides on $DH_{max,exp}$, hydrolysis by the 5 enzymes were performed with a substrate concentration of 0.1 % and at an E:S of 1:25. For 0.1% α-LA hydrolysis by bovine trypsin, at an E:S of 1:25, a $DH_{max,fit}$ of 4.0 % was reached (**Figure 5.6A**). This value was higher than the $DH_{max,fit}$ of 0.1 % α-LA hydrolysis by bovine trypsin at an E:S of 1:100. When additional enzyme (to a final E:S of 1:25) was added to the 0.1 % α-LA hydrolysate at an E:S of 1:100 (at 3 h), the hydrolysis continued to reach a final $DH_{max,fit}$ of 4.4 %. This value was similar to the $DH_{max,fit}$ reached at an E:S of 1:25, and a lot closer to the DH_{max} (7.0 %) of hydrolysis at an E:S of 1:100 (**Table 5.2**). This means that the remaining CSs were still available for the fresh enzyme. These observations proved that the low $DH_{max,exp}$ of the 0.1 % α-LA hydrolysis at an E:S of 1:100 was not caused by the problem of the substrate but of the enzyme. To test if the second addition of bovine trypsin changed the hydrolysis mechanism, the percentage of remaining intact α-LA in the hydrolysates was plotted against DH/DH_{max} (**Figure 5.4A**). The additional hydrolysis had the same P^{intact} as the previous hydrolysis, indicating that the hydrolysis might still produce inhibitory peptides. However, the amount of active enzyme in the system was more than the amount of inhibitory peptides, resulting in a higher $DH_{max,exp}$. This means that changes in the hydrolysis mechanism was induced by the changes in substrate concentrations but not by the changes in E:S. If bovine trypsin in 0.1 % hydrolysis at an E:S of 1:100 was considered to be completely inhibited but hydrolysis at an E:S of 1:25 was completely not, the inhibited amount of bovine trypsin was estimated to be 0.43-1.72 µM. In the 0.1 % α-LA hydrolysate, the starting protein concentration was ~57 µM (taking into account the protein purity and protein content based on the Dumas results). This indicated that perhaps the binding between inhibitory peptides and bovine trypsin was not one peptide per enzyme. This was similar to the inhibitory peptides for ACE, where the molar ratio of inhibitory peptides derived from bovine caseins to fully inhibit ACE varied from 26:1 to >$1.3 \cdot 10^3$:1 [16, 17].

For porcine or human tryptic hydrolysis, the $DH_{max,fit}$ of 0.1 % α-LA hydrolysis at an E:S of 1:25 reached that of 10 % and 1 % α-LA hydrolysis at an E:S of 1:100 ($DH_{max}^{HT/PT}$, i.e. the DH_{max} assumed to be reached without enzyme inhibition), respectively (**Figures 5.6B** and **5.6C**). This suggested that the enzymes were indeed inhibited by the peptides formed during the hydrolysis at an E:S of 1:100. For hydrolysis by α-chymotrypsin or BLP at an E:S of 1:25, the $DH_{max,fit}$ were even 2 times higher than the $DH_{max}^{CT/BLP}$ of hydrolysis at an E:S of 1:100 (**Figures 5.6D** and **5.6E**). This indicated that for these two enzymes, the enzymes in all hydrolysis at an E:S of 1:100 were inhibited. For BLP, the hydrolysis at an E:S of 1:25 had a $DH_{max,fit}$ of 11.2 %, which was ~68 % of the $DH_{max,theo}$ (**Table 5.1**). It means that ~14 CSs in α-LA were estimated to be HSSs. This indicated that besides the 7 CSs after glutamic acids in α-LA, also 7 CSs after aspartic acids were efficiently hydrolysed by BLP. In conclusion, to overcome the low $DH_{max,exp}$ values reached due to the inhibitory peptides, adding fresh enzymes or having more enzymes initially in the system could be sufficient. Using these conditions, there was still enough amount of active enzyme in the hydrolysate, which can still hydrolyse the HSSs, to reach a higher $DH_{max,exp}$.

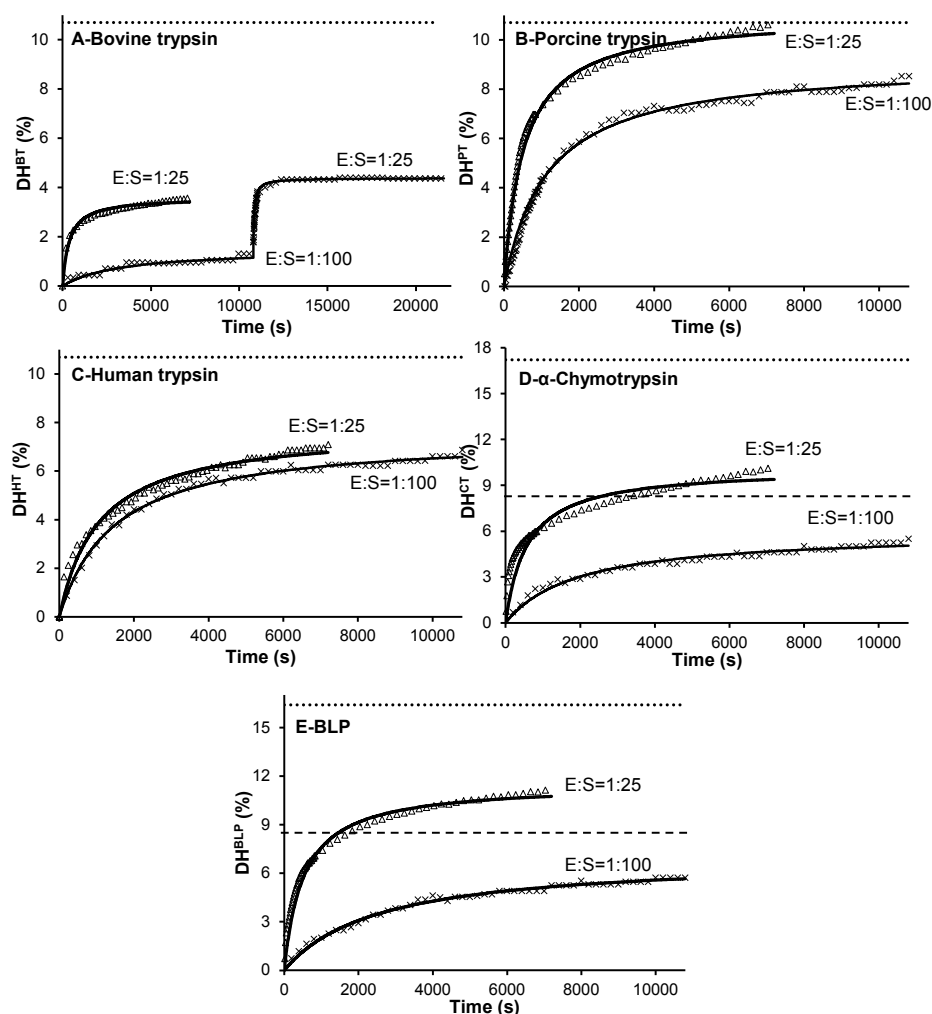**Figure 5.6:** Hydrolysis curves (DH *versus* time) of 0.1 % α-LA hydrolysed by (A) bovine, (B) porcine and (C) human trypsins, (D) α-chymotrypsin or (E) *Bacillus licheniformis* protease (BLP) at an E:S of (×) 1:25 or (△) 1:100. In panel (A), the hydrolysis curve from 3-6 h was the hydrolysis of the 0.1 % α-LA hydrolysate by the second addition of bovine trypsin (final E:S=1:25). The solid lines (—) indicate the fit using the 2nd order reaction equation. The dotted lines (······) refer to the theoretical maximum degree of hydrolysis ($DH_{max,theo}$) of α-LA by each enzyme. The dashed lines (------) in panel (D) and (E) refer to the experimental maximum degree of hydrolysis at an E:S of 1:100.

Conclusions

Large differences were found on the influence of substrate concentration on hydrolysis by bovine, porcine and human trypsins, bovine α-chymotrypsin or *Bacillus licheniformis* protease

(BLP). Hydrolysis by human trypsin was not sensitive to changes in substrate concentrations. The $DH_{max,exp}$ of hydrolysis by bovine and porcine trypsins, or bovine α-chymotrypsin, increased with an increasing substrate concentration, while the $DH_{max,exp}$ of BLP hydrolysis decreased with an increasing concentration. The change in the $DH_{max,exp}$ of hydrolysis seems to relate to a shift from one-by-one to zipper hydrolysis mechanism, which was proposed to be due to the formation of inhibitory peptides. For some enzymes (e.g. bovine trypsin), this was a result of an increasing substrate concentration, whereas for other enzymes (e.g. BLP) this was a result of a decreasing concentration. Due to different mechanisms, hydrolysates produced by hydrolysis with different substrate concentrations had different peptide profiles. Assuming some peptides formed during hydrolysis could inhibit the protease, the influence of substrate concentration was explained. An excess of enzyme, e.g. by using a higher enzyme to substrate ratio, could overcome the effect of inhibitory peptides. The outcome of this work can be used to compare the reported digestibility of proteins, where various substrate concentrations were used to perform the experiments.

Acknowledgement

References

[1] S. Boisen, J.A. Fernández, Prediction of the total tract digestibility of energy in feedstuffs and pig diets by *in vitro* analyses, Animal Feed Science and Technology, 68 (1997) 277-286.

[2] D. Dupont, G. Mandalari, D. Molle, J. Jardin, J. Léonil, R.M. Faulks, M.S.J. Wickham, E.N.C. Mills, A.R. Mackie, Comparative resistance of food proteins to adult and infant *in vitro* digestion models, Molecular Nutrition and Food Research, 54 (2010) 767-780.

[3] G. Mandalari, K. Adel-Patient, V. Barkholt, C. Baro, L. Bennett, M. Bublin, S. Gaier, G. Graser, G.S. Ladics, D. Mierzejewska, E. Vassilopoulou, Y.M. Vissers, L. Zuidmeer, N.M. Rigby, L.J. Salt, M. Defernez, F. Mulholland, A.R. Mackie, M.S.J. Wickham, E.N.C. Mills, *In vitro* digestibility of β-casein and β-lactoglobulin under simulated human gastric and duodenal conditions: A multi-laboratory evaluation, Regulatory Toxicology and Pharmacology, 55 (2009) 372-381.

[4] M. Minekus, M. Alminger, P. Alvito, S. Ballance, T. Bohn, C. Bourlieu, F. Carriere, R. Boutrou, M. Corredig, D. Dupont, C. Dufour, L. Egger, M. Golding, S. Karakaya, B. Kirkhus, S. Le Feunteun, U. Lesmes, A. Macierzanka, A. Mackie, S. Marze, D.J. McClements, O. Menard, I. Recio, C.N. Santos, R.P. Singh, G.E. Vegarud, M.S.J. Wickham, W. Weitschies, A. Brodkorb, A standardised static *in vitro* digestion method suitable for food - an international consensus, Food & Function, 5 (2014) 1113-1124.

[5] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Determination of the influence of substrate concentration on enzyme selectivity using whey protein isolate and *Bacillus licheniformis* protease, Journal of Agricultural and Food Chemistry, 62 (2014) 10230-10239.

[6] C.I. Butré, P.A. Wierenga, H. Gruppen, Effects of ionic strength on the enzymatic hydrolysis of diluted and concentrated whey protein isolate, Journal of Agricultural and Food Chemistry, 60 (2012) 5644-5651.

[7] N.A. Hardt, A.J. van der Goot, R.M. Boom, Influence of high solid concentrations on enzymatic wheat gluten hydrolysis and resulting functional properties, Journal of Cereal Science, 57 (2013) 531-536.

[8] M.C. Márquez, M.A. Vázquez, Modeling of enzymatic protein hydrolysis, Process Biochemistry, 35 (1999) 111-117.

[9] W. Qi, Z. He, Enzymatic hydrolysis of protein: Mechanism and kinetic model, Frontiers of Chemistry in China, 1 (2006) 308-314.

[10] Y. Deng, P.A. Wierenga, H.A. Schols, S. Sforza, H. Gruppen, Effect of Maillard induced glycation on protein hydrolysis by lysine/arginine and non-lysine/arginine specific proteases, Food Hydrocolloids, 69 (2017) 210-219.

[11] A. Fernández, F. Riera, β-Lactoglobulin tryptic digestion: A model approach for peptide release, Biochemical Engineering Journal, 70 (2013) 88-96.

[12] S.C. Cheison, E. Leeb, J. Toro-Sierra, U. Kulozik, Influence of hydrolysis temperature and pH on the selective hydrolysis of whey proteins by trypsin and potential recovery of native alpha-lactalbumin, International Dairy Journal, 21 (2011) 166-171.

[13] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Introducing enzyme selectivity: A quantitative parameter to describe enzymatic protein hydrolysis, Analytical and Bioanalytical Chemistry, 406 (2014) 5827-5841.

[14] N.A. Hardt, A.E.M. Janssen, R.M. Boom, A.J. van der Goot, Factors impeding enzymatic wheat gluten hydrolysis at high solid concentrations, Biotechnology and Bioengineering, 111 (2014) 1304-1312.

[15] K. Linderstrøm-Lang, The initial phases of the enzymatic degradation of proteins, Bulletin de la Société de chimie biologique, 35 (1953) 100-116.

[16] S. Maruyama, H. Mitachi, H. Tanaka, N. Tomizuka, H. Suzuki, Studies on the active site and antihypertensive activity of angiotensin I-converting enzyme inhibitors derived from casein, Agricultural and Biological Chemistry, 51 (1987) 1581-1586.

[17] S. Maruyama, H. Mitachi, J. Awaya, M. Kurono, N. Tomizuka, H. Suzuki, Angiotensin I-converting enzyme inhibitory activity of the C-terminal hexapeptide of α_{s1}-casein, Agricultural and Biological Chemistry, 51 (1987) 2557-2561.

[18] Y. Guo, D. Pan, M. Tanokura, Optimisation of hydrolysis conditions for the production of the angiotensin-I converting enzyme (ACE) inhibitory peptides from whey protein using response surface methodology, Food Chemistry, 114 (2009) 328-333.

[19] A.B. Nongonierma, C. Mazzocchi, S. Paolella, R.J. FitzGerald, Release of dipeptidyl peptidase IV (DPP-IV) inhibitory peptides from milk protein isolate (MPI) during enzymatic hydrolysis, Food Research International, 94 (2017) 79-89.

[20] L. Feng, M. Tu, M. Qiao, F. Fan, H. Chen, W. Song, M. Du, Thrombin inhibitory peptides derived from *Mytilus edulis* proteins: Identification, molecular docking and in silico prediction of toxicity, European Food Research and Technology, *Article in press, DOI: 10.1007/s00217-017-2946-7*.

[21] M.C. Marquez Moreno, V. Fernandez Cuadrado, Enzymic hydrolysis of vegetable proteins: Mechanism and kinetics, Process Biochemistry, 28 (1993) 481-490.

[22] O. Akpinar, M.H. Penner, Peptidase activity assays using protein substrates, Current protocols in food analytical chemistry, John Wiley & Sons, Corvallis, OR, USA, 2001, pp. Unit C2.2.4.

[23] Y. Deng, F. van der Veer, S. Sforza, H. Gruppen, P.A. Wierenga, Towards predicting protein hydrolysis by bovine trypsin, Process Biochemistry, 65 (2018) 81-92.

[24] N.D. Rawlings, M. Waller, A.J. Barrett, A. Bateman, MEROPS: The database of proteolytic enzymes, their substrates and inhibitors, Nucleic Acids Research, 42 (2014) D503-D509.

[25] C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Determination of the influence of the pH of hydrolysis on enzyme selectivity of *Bacillus licheniformis* protease towards whey protein isolate, International Dairy Journal, 44 (2015) 44-53.

[26] J. Adler-Nissen, Enzymic hydrolysis of food proteins, Elsevier Applied Science Publishers, London, UK, 1986, pp. 122-123.

[27] C.I. Butré, P.A. Wierenga, H. Gruppen, Influence of water availability on the enzymatic hydrolysis of proteins, Process Biochemistry, 49 (2014) 1903-1912.

[28] B.J.H. Kuipers, H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis, Journal of Agricultural and Food Chemistry, 55 (2007) 5445-5451.

[29] H.A. Kosters, P.A. Wierenga, R. de Vries, H. Gruppen, Characteristics and effects of specific peptides on heat-induced aggregation of β-lactoglobulin, Biomacromolecules, 12 (2011) 2159-2170.

CHAPTER 6

Understanding glycation kinetics of individual peptides in protein hydrolysates

During industrial processing of protein hydrolysates, peptides can be glycated through the Maillard reaction if carbohydrates are present. In hydrolysates, peptides with different lengths, type (α/ε-) and number of amino groups are present. Based on previous studies with free amino acids (AA), synthetic peptides or intact proteins, these molecular properties might influence the glycation kinetics of individual peptides. In this study, the glycation kinetics of individual peptides in protein hydrolysates was followed. The aim was to identify the effects of peptide properties and hydrolysate composition on the glycation kinetics of individual peptides. α-Lactalbumin was hydrolysed by *Bacillus licheniformis* protease (to degrees of hydrolysis of 2, 4, 6 and 8 %), and glycated with D-glucose (0-8 h, 50 °C). The hydrolysates contained, in addition to the remaining intact protein, 24 unique peptides, ranging from 2-123 AAs with 0-12 lysine(s). The glycation rate constant (k_g) and the maximum average degree of glycation ($DG\_P_{av,max}$) of peptides were independent of the hydrolysate composition. The k_g values showed 3 populations, but could not be correlated with the peptide length or the number of amino groups per peptide. Surprisingly, the $DG\_P_{av,max}$ of peptides did not reach 100 %. The maximum degrees of glycation of α- and ε-amino groups were 13 % and 60 %, respectively, which explained why the $DG\_P_{av,max}$ ranged from 9 % to 59 % for peptides with 0-12 lysine(s). This work showed that the extent of glycation of individual peptides in hydrolysates can be predicted based on the AA sequences of the peptides.

Introduction

The reaction kinetics of the Maillard reaction have typically been studied using pure peptides and proteins. However, protein hydrolysates are used as ingredients in many food products, e.g. hypoallergenic infant formula. During industrial processing (or storage) in the presence of reducing carbohydrates, the peptides can become glycated through the Maillard reaction. The Maillard reaction was suggested to occur more in hydrolysates. It was shown that the amount of advanced Maillard products in a glycated whey protein hydrolysate was found 6 times more than in the glycated intact protein [1]. This might be because in contrast with systems with one type of molecules, such as intact proteins, in a hydrolysate, peptides with different amino acid (AA) sequences are present. According to reported results of synthetic peptides, the peptide length [2], and the type (α/ε-) and number of amino groups [3, 4] could influence the glycation kinetics. It is unclear how much these peptide properties and the hydrolysate composition influence the glycation kinetics of peptides in hydrolysates. To answer this, this study aimed to follow the glycation kinetics of individual peptides in protein hydrolysates.

In literature, it was suggested that the Maillard reaction occurs more in hydrolysates than in samples containing only intact proteins. For instance, for whey protein hydrolysate with a degree of hydrolysis of 9.3 %, the formed advanced Maillard products were 6 times more than in glycated whey protein isolate [1]. However, in that study the Maillard reaction we evaluated by the formation of brown colour, so the conclusions were based on the extent to which the secondary reactions occurred. It was unclear if the difference was already in the initial stage of the reaction, i.e. the glycation, where the reducing carbohydrates react with free amino groups of the proteins or peptides. Studies with synthetic peptides showed differences between the chemical reactivity of α- and ε-amino groups. In the reaction with triazinyl dye, the reaction rate constant of the ε-amino group on lysine was ~10 times higher than that of the α-amino group [5]. For the Maillard reaction, most studies did not focus on the initial stage of the reaction (glycation), but rather on the Maillard browning [3, 6-8]. For instance, the browning intensities (absorbance at 420 nm) of lysine and 8 other AAs during 0-12 h of heating with D-glucose were measured [7]. The reaction rate constant of ε-amino group, estimated from the absorbance at 420 nm *versus* the heating time fitted with 1st order reaction equation, was ~2.5 times higher than that of the α-amino group [7]. The fitted maximum absorbances at 420 nm of the two types of amino groups were similar [7].

For synthetic peptides that do not contain lysines, the glycation of α-amino groups was reported to depend on the type of N-terminal AA and the peptide length. For instance, the glycation rate constants of dipeptides with histidine as the N-terminal AA were 1-4 times (depending on the C-terminal AA) higher than that of dipeptides with the other 7 AAs tested as the N-terminal AA, which all had similar rate constants (~5 s^{-1}M^{-1}) [9]. The effect of peptide length was proposed to be due to the increase in pKa of the α-amino group, caused by the carboxyl group of the

C-terminal AA [2]. This means less α-amino groups were unprotonated at the pH during glycation (pH 7.8) [2]. The rate constants of di- and tri-glycine, calculated from 1st order reaction equation of the percentages of unbound glucose *versus* the heating time, were ~2 times higher than that of glycine [2]. Since protonation is a very fast equilibrium, the shift in pKa is not expected to affect the extent of the reaction. However, the final percentage of unbound glucose in glycine solution was ~25 % higher than that of in di- or tri-glycine solutions [2]. The influence of this shift in pKa on glycation was also discussed in synthetic peptides that contained lysine [10]. A shift in pKa was observed for lysines in 78 proteins, ranging from 5.7-12.1 [11]. This shift in pKa was attributed to the environment of the lysine residues, i.e. surrounding AAs. For example, the extent of glycation for synthetic peptides that had aspartic acid as a surrounding AA (44 or 60 % glycated, depending on the position of aspartic acid) was lower than that of the peptide surrounded by neutral AAs (76 % glycated) [10]. Besides the charge of surrounding AAs, the hydrophobicity was also reported to affect the glycation [12]. For dipeptides with lysine as the N-terminal AA, if the C-terminal AA was hydrophobic, the dipeptides were fully glycated after 30 minutes [12]. When other AAs were the C-terminal AA, 30-90 % of the peptides remained non-glycated [12].

To identify the effects of peptide properties on the glycation of peptides in protein hydrolysates, in this study, the glycation of individual peptides in hydrolysates was followed. Hydrolysates that contained peptides with different properties, i.e. peptide length, type (α/ε-) and number of amino groups on the peptide, and different relative abundances, were prepared and glycated under identical conditions.

Materials and methods

Materials

α-Lactalbumin was obtained from Davisco Foods International Inc. (Le Sueur, MN, USA). ~72 % of it was in apo form and the rest was in holo form, based on results from circular dichroism [13]. *Bacillus licheniformis* protease (BLP) (batch NS-37005) was obtained from Novozymes (Bagsvaerd, Denmark). The BLP was partly insoluble, it was purified as described previously [14]. The suspension of BLP was centrifuged (10 min, 4000×g, 25 °C) and the supernatant was dialyzed (cut-off 12-14 kDa) against 150 mM NaCl solution, then against demineralized water and then freeze-dried. The freeze-dried powder had a protein content of ~60 % [w/w] based on the Dumas results (N×5.93, based on the AA composition of the protein as described in Uniprot accession code P80057, www.uniprot.org). The enzyme activity was 3.9 AU mg^{-1}·min^{-1} determined using the azocasein assay [15]. The purity of BLP was 100 % based on the UV$_{280}$ peak area determined using the PDA detector attached to the ultra-performance liquid chromatography (UPLC) system and was 92 % based on the UV$_{214}$ peak area (A$_{214}$). In the UV$_{214}$ chromatogram, two peaks were found, of which the main peak (78 % of total A$_{214}$) was identified to be the BLP (23.6 kDa) and the minor peak (14 % of total A$_{214}$) was the pro-peptide (6.9 kDa). All other chemicals were of analytical grade and purchased from Sigma or Merck.

Methods

Preparation of apo α-LA

Ethylenediaminetetraacetic acid (EDTA) was used to chelate the calcium ions attached to the α-lactalbumin molecules. EDTA was added to 1 L of 10 % [w/v] protein solution at an α-lactalbumin to EDTA ratio of 1:5 [mol/mol], and the solution was stirred overnight at 4 °C. To remove EDTA, the solution was ultra-filtered with 7 L of Millipore water over a UFP-10-C-5 membrane (cut-off 10 kDa, AG Technologies Corp., Tokyo, Japan). The prepared apo α-lactalbumin (further referred to as α-LA) was freeze dried and stored at -20 °C before further analysis. The protein content was ~90 % [w/w] based on Dumas method (N×6.25, based on the AA composition as described in Uniprot accession code P00711). Of the total protein content, ~90 % was α-LA, based on the A_{214} determined using the PDA detector attached to the UPLC system, and the remaining 10 % consisted of β-lactoglobulin and bovine serum albumin.

Production of α-LA hydrolysates with different degrees of hydrolysis

The enzymatic hydrolysis was performed as described elsewhere [16]. The α-LA was dissolved in 20 mL Millipore water at a concentration of 1 % [w powder/v]. The solution was adjusted to pH 8.0 and equilibrated for 0.5 h at 37 °C in a pH-stat device (Metrohm, Herisau, Switzerland). The BLP was dissolved (10 mg powder/mL) in Millipore water and added to the equilibrated protein solution to reach an enzyme/substrate ratio of 1:100 [w/w] [16]. The hydrolysis was performed at 37 °C using a pH-stat device with 0.2 M NaOH to keep the pH constant. The degree of hydrolysis (DH_{stat}) was calculated using the equation published previously [17]. The hydrolysis was stopped at DH 2, 4, 6 and 8 % (labelled DH2-DH8) by adjusting the pH to 2 using 5 M HCl [16]. The pH of the hydrolysate was re-adjusted to 8.0 after ≥10 minutes of inactivation. As a control experiment, 1 % α-LA solution was incubated in a pH-stat in the absence of BLP. The added volume of NaOH from the control experiment was subtracted from the added volume of NaOH at all time points during hydrolysis, to compensate the consumption of NaOH due to the dissolution of CO_2. The samples were stored at -20 °C.

Glycation of intact and hydrolysed α-LA

The glycation of intact and hydrolysed α-LA was performed as described previously [13]. To a 5 mL protein/hydrolysate solution at 1 % [w powder/v], 0.05 mL of a 1 M sodium phosphate buffer (pH 8.0) was added to reach a molarity of phosphate buffer of 10 mM, with a negligible change in protein concentration. D-glucose was dissolved in 10 mM sodium phosphate buffer (pH 8.0) at concentrations of 3.5, 3.8, 4.4, 4.9 and 5.5 mg/mL. 5 mL of each D-glucose solution was mixed with 5 mL of the hydrolysate solution (1 % [w powder/v]) with a DH_{stat} of 0, 2, 4, 6 and 8 %, respectively, in order to reach a molar ratio of free amino groups to reducing ends of 1:2. The concentration of free amino groups in each hydrolysate was calculated based on the number of lysine residues (12 in α-LA) and N-termini per original protein molecule. The number of N-termini equals to $1+DH_{stat}/100×122$ (the number of peptide bonds in α-LA). The mixtures were freeze dried, followed by incubation at 50 °C under 65 % relative humidity for 0, 2, 4, 6 and 8 h in a

humidity control chamber HCP108 (Memmert, Schwabach, Germany). The glycated samples were labelled $DH2\_G_{0-8}$-$DH8\_G_{0-8}$. The non-hydrolysed α-LA (DH0) was glycated for 0-24 h ($DH0\_G_{0-24}$) and the reaction was performed in duplicate. As a control, the hydrolysates were also incubated without D-glucose ($DH0_8$-$DH8_8$) at 50 °C under 65 % relative humidity for 8 h. All samples were freeze-dried and stored at -20 °C.

o-Phtaldialdehyde method (OPA)

The total number of free amino groups of the samples ($DH0\_G_{0-8}$-$DH8\_G_{0-8}$) was determined using the OPA method [18]. The samples were dissolved at a concentration of 0.5 % [w/v] in a 2 % [w/v] SDS solution. 5 μL of sample solution was added to 300 μL of OPA reagent solution and incubated for 10 minutes. The experiments were performed in a 96 well plate using Tecan Infinite F500 (Tecan, Männedorf, Switzerland) to obtain the UV_{340} absorption of the samples. The measurements were performed in triplicate. A calibration curve was made with leucine. The protein concentration of each sample was determined by the Dumas method. The total degree of glycation (DG_T_{OPA}) of the hydrolysates was calculated using equation 1, adapted from an equation published previously [19],

$$(1) \quad DG\_T_{OPA}\left[\%\right] = \frac{(\#NH_{2,u} - \#NH_{2,g})}{\#NH_{2,u}} \times 100 \ \%$$

where $\#NH_{2,u}$ and $\#NH_{2,g}$ were the number of free amino groups per original protein molecule in the untreated and glycated hydrolysates, respectively, calculated from the concentration of free amino group, corrected for the concentration of total proteinaceous matter.

Dumas

The protein content was determined using a Flash EA 1112 NC Analyzer (Thermo Fisher Scientific Inc, Waltham, MA, USA). The nitrogen-protein conversion factor of 6.25 for non-hydrolysed α-LA was used, calculated from the AA composition of the protein (Uniprot accession code P00711). Nitrogen-protein conversion factors of 6.27, 6.29, 6.31 and 6.33 were used for hydrolysates with a DH_{stat} of 2, 4, 6 and 8 %, respectively, based on the number of water molecules introduced to the peptides during hydrolysis.

Reversed phase ultra-high performance liquid chromatography (RP-UPLC)

The samples were analysed on an H class Acquity UPLC® system (Waters, Milford, MA, USA) equipped with a BEH C18 column (1.7 μm, 2.1×100 mm, Waters). An Acquity PDA® detector (Waters) was attached to the UPLC system. The eluents and elution gradient were the same as described previously [16]. Samples were incubated for 2 h with 100 mM ditiothreitol (DTT) in 50 mM Tris-HCl buffer at pH 8.0 to reduce the disulphide bridges, and diluted to a protein concentration of 0.1 % [w/v powder] and centrifuged (10 min, 14,000×g, 20 °C) before injection (4 μL).

Electron spray ionization time of flight mass spectrometry (ESI-Q-TOF-MS)

The mass spectra of the samples were measured using an online SYNAPT G2-Si high definition mass spectrometry (Waters) coupled to the RP-UPLC system. Sodium iodide was used for the MS calibration. The MS and MS/MS (MSe) data were collected using the methods developed previously [16]. Online lock mass (angiotensin II, $[M+2H]^{2+}$: 523.7751) was acquired. Based on the differences in the measured and theoretical lock mass, corrections were applied on m/z of the peptides during measurements. The data were analysed manually using MassLynx software v4.1 (Waters).

Peptide identification and quantification

The peptides were manually annotated based on MS and -when possible- confirmed by identifying the b and y fragments in MS/MS spectra. The mass tolerance between the theoretical and measured mass for the accepted annotation was set at 100 ppm. Glycated peptides were identified by the extra masses of n dehydrated glucose(s) ($+n \times 162$ Da) on top of the mass of the original peptide. The concentration of each peptide ($C_{peptide}$) was calculated using equation 2 [16]:

$$(2) \quad C_{peptide} \; [\mu M] = \frac{A_{214} \cdot Q}{\varepsilon_{214} \cdot l \cdot V_{inj} \cdot k_{cell}}$$

All hydrolysates were injected at an equal concentration [g/L]. To compare samples, $C_{peptide}$ was re-calculated using a correction for the changed protein content in the glycated hydrolysates, based on the Dumas results. A_{214} [μAU·min] is the UV peak area at 214 nm, V_{inj} [μL] is the injected volume of the sample, Q [μL·min^{-1}] is the flow rate and l [cm] is the path length of the UV cell, which is 1 cm according to the manufacturer. ε_{214} [M^{-1}·cm^{-1}] is the molar extinction coefficient at 214 nm, calculated using an equation described previously [20]. The molar extinction coefficient of glycated α-LA was reported to be the same as the untreated α-LA, proven by determining the UV absorbances of a dilution series of non-glycated and glycated α-LA solutions with known concentrations (based on the Dumas results) [13]. Therefore, the molar extinction coefficients of the glycated and non-glycated peptides were assumed to be the same. Due to the multiple reflection by the coating of the cell, the effective path length of the light through the cell is not the same as the length specified by the producer. To correct for this effect, the cell constant of the UV detector (k_{cell}) was determined, using a series of standard solutions made by β-lactoglobulin, β-casein and angiotensin II, and the approach described elsewhere [21]. The ratio between the measured and expected A_{214} was taken as the value of k_{cell}. For the UV cell used in this work, the k_{cell} was determined to be 0.78 [22]. The linear region of the A_{214} in the PDA detector ranges from 5×10^1 to 6×10^5 μAU·min. Peptide quantification was done for peptides with a $A_{214} \geq 5 \times 10^1$ μAU·min. For all hydrolysates, on average 95±10 % of the expected total A_{214} was found. Of the total A_{214}, on average 94±3 % in the samples was assigned to annotated peptides. When a peptide was glycated, it sometimes co-eluted with the non-glycated peptide. In the case of co-elution, the concentration of each co-eluting peptide was calculated using the intensity of the total ion count to divide the A_{214}.

Quality check of the annotation and quantification of the peptide analysis

The quality of the peptide analysis was checked by determining the amino acid (AA), peptide and molar sequence coverages, calculated by equation 3, 4 and 5, respectively [16].

$$(3) \quad Amino \ acid \ sequence \ coverage \ [\%] = \frac{\# unique \ annotated \ AA}{\# AA_{protein}} \times 100 \ \%$$

$$(4) \quad Peptide \ sequence \ coverage \ [\%] = \frac{\# AA \ (annotated \ peptides)}{\# AA \ (annotated \ peptides) + \# AA \ (missing \ peptides)} \times 100 \ \%$$

$$(5) \quad Molar \ sequence \ coverage \ [\%] = \left(1 - \frac{\sqrt{\frac{\sum (C_n - C_0)^2}{(\# AA_{protein} - 1)}}}{C_0} \right) \times 100 \ \%$$

C_n [μM] is the concentration of each individual AA (n) in the protein sequence obtained from quantification of the peptides; C_0 [μM] is the initial injected protein concentration and $\#AA_{peptide}$ is the number of AAs in the sequence of the protein [16]. The initial injected protein concentrations were determined using the Dumas method. The missing peptides are the peptides that were not quantified but should be present given the annotated peptides [16]. The three parameters were determined for all non-glycated and glycated hydrolysates. To further check the quality of the analysis, the degree of hydrolysis based on peptide analysis (DH_{MS}) of each hydrolysate was calculated using equation 6,

$$(6) \quad DH_{MS} \ [\%] = \frac{\sum C_{i,t}}{2 \times \# \ peptide \ bonds \times C_0} \times 100 \ \%$$

where #peptide bonds is the number of peptide bonds and $C_{i,t}$ [μM] is the concentration of cleavage site products formed. i equals the sum of concentrations of all peptides of sequence [x-y], for which $i=(x-1)$ or $i=y$. The $C_{i,t}$ can be calculated using equation 7.

$$(7) \quad C_{i,t} [\mu M] = \sum \left\{ C_{peptide}[x-y]_t \middle| i = x - 1 \cup i = y \right\}$$

The average degree of glycation of a peptide

For each peptide in the hydrolysates, the average degree of glycation (DG_P_{av}) was calculated as the percentage of average number of glycated amino groups in total number of amino groups of that peptide (equation 8) [13]. In other words, DG_P_{av} is the weighted average of concentrations of all variants of the peptide (glycated and non-glycated).

$$(8) \quad DG\_P_{av} [\%] = \frac{\sum_{i=0}^{n} \frac{i \times C_{peptide, \ i}}{C_{peptide, \ tot}}}{\# NH_{2,u}} \times 100 \ \%$$

$C_{peptide,i}$ is the concentration of the peptide variant with *i* glucose molecules attached, $C_{peptide,tot}$ is the sum of concentration of glycated and non-glycated variants and $\#NH_{2,u}$ is the total number of amino groups of the untreated peptides, which is the sum of the number of lysine residues and the N-terminal AA.

The total degree of glycation of a hydrolysate based on peptide analysis

Based on peptide analysis, the total DG (DG_T$_{MS}$) of a hydrolysate was determined using the DG_P$_{av}$ and $C_{peptide,tot}$ in the hydrolysates (equation 9), and compared with the total DG determined using the OPA method (DG_T$_{OPA}$).

$$(9) \quad DG\_T_{MS}[\%] = \frac{\sum \left(C_{peptide,tot} \times DG\_P_{av} \times \# NH_{2,u} \right)}{\sum \left(C_{peptide,tot} \times \# NH_{2,u} \right)} \times 100\%$$

where $\#NH_{2,u}$ is the total number of amino groups of the untreated peptides.

The maximum average degree of glycation and the glycation rate constant of a peptide

The maximum average degree of glycation (DG_P$_{av,max}$) and glycation rate constant (k_g) during Maillard incubation were determined by fitting the glycation curve (DG_P$_{av}$ *versus* time) of each peptide with the 1st order reaction equation (equation 10),

$$(10) \quad DG\_P_{av}[\%] = DG\_P_{av,max} \times \left(1 - e^{-k_g \cdot t} \right) + DG\_P_{av,0}$$

where both DG_P$_{av,max}$ [%] and k_g [s^{-1}] were fitting parameters. DG_P$_{av,0}$ [%] is the DG$_{av}$ of the non-heated samples (DH0_G$_0$-DH8_G$_0$), which was set at 0 because the MS intensities of the glycated peptides in these samples were within the noise of the MS spectra.

Non-targeted clustering of the DG_P$_{av,max}$ and k_g

The DG_P$_{av,max}$ and k_g of peptides were fitted with a normal distribution curve. The structure of the residuals of the fit was used to identify the number of populations in the data set. Once the number of populations (k) was identified, the Matlab method k-means clustering was used to categorise the (k) populations. The k-means clustering is a partitioning method, which uses the squared Euclidean metric to determine the distances and the k-means++ algorithm for the cluster centre initialization. It is used to group data points into populations with minimized total sum of distances.

The maximum degree of glycation of α- and ε-amino groups

The maximum degree of glycation of α-amino groups (DG$_{max,α-NH2}$) was estimated based on the average value of DG_P$_{av,max}$ of the peptides that did not contain lysine. Based on the DG$_{av,max}$ of

each peptide, the maximum degree of glycation of ε-amino group ($DG_{max,\varepsilon\text{-}NH2}$) in each peptide was calculated based on equation 11.

$$(11) \quad DG\_P_{av,max}\left[\%\right] = DG_{max,\alpha\text{-}NH_2} \times \frac{1}{\#GS} + DG_{max,\varepsilon\text{-}NH_2} \times \frac{\#lysine}{\#GS}$$

The #lysine and #GS represent the number of lysine and glycation site (the total number of lysine and N-terminal AA), respectively. The heterogeneity of the glycation process was checked by determining the $DH_{max,\varepsilon\text{-}NH2}$ of the glycated variants of the peptides, calculated by dividing the $DH_{max,\varepsilon\text{-}NH2}$ by the percentage of glycated variants ($DG_{max,\varepsilon\text{-}NH2}$)/(1-%non-glycated peptide).

Results and discussion

Analysis of the non-glycated hydrolysates
The preparation of the α-LA hydrolysates was reproducible, indicated by the similarity in the hydrolysis curves (DH_{stat} *versus* time) for the four hydrolysates prepared (**Figure 6.1**).
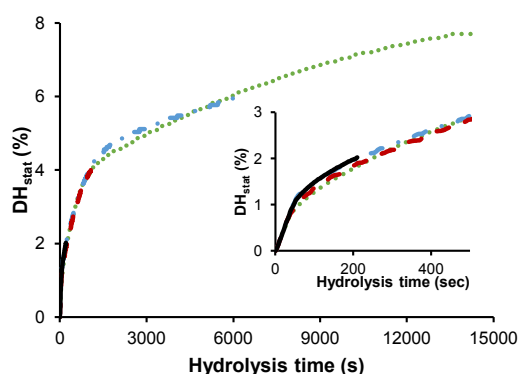
Figure 6.1: Hydrolysis curves (DH_{stat} *versus* time) of hydrolyses for the preparation of α-LA hydrolysates with a DH_{stat} of 2 (—), 4 (---), 6 (—·) and 8 (⋯) %. The inset shows the zoomed hydrolysis curves from 0-500 sec.

In non-glycated hydrolysates $DH2\_G_0$, $DH4\_G_0$, $DH6\_G_0$ and $DH8\_G_0$, 17, 20, 17 and 13 peptides were identified and quantified, respectively (**Table 6.1** and **Annex 6.1**). In total, in addition to the remaining intact protein, 24 unique peptides were annotated, of which 23 were found in more than one hydrolysate. The relative abundance (in mole/mole %) of these peptides ranged from 0.1 % to 24 % of the total amount of peptides (**Table 6.1**). Out of these 25 peptides (including the intact protein), 5 peptides did not contain any lysine. The length of these peptides ranged from 4 to 24 AAs. For the peptides that contained 1 or 2 lysine(s), the peptide lengths ranged from 2-8 and from 10-42 AAs, respectively.

Table 6.1: Number of lysines per peptide, peptide length and the concentration of peptides identified and quantified in non-glycated α-LA hydrolysates with a DH_{stat} of 2, 4, 6 and 8 %.

Peptides	#lysines	Length	Concentration of peptides in hydrolysates [µM]			
			DH2_G_0 (17)	DH4_G_0 (20)	DH6_G_0 (17)	DH8_G_0 (13)
[8-11]	0	4	4.0	23.0	24.1	23.0
[117-121]	0	5	0.2	1.3	3.2	9.8
[38-46]	0	9	-	-	3.5	8.6
[38-49]	0	12	-	2.5	8.4	7.6
[26-49]	0	24	0.1	1.1	1.3	-
[122-123]	1	2	8.6	27.7	23.0	20.2
[12-14]	1	3	-	-	2.9	7.8
[1-7]	1	7	10.9	27.1	21.5	19.0
[8-14]	1	7		0.6	1.0	-
[114-121]	1	8	3.7	19.4	9.3	0.2
[114-123]	2	10	3.2	0.9	0.6	-
[12-25]	2	14	0.4	4.6	29.3	37.1
[57-78]	2	22	-	-	-	0.5
[12-37]	2	26	-	0.3	3.7	0.2
[12-46]	2	35	-	0.6	1.7	-
[12-49]	2	38	5.8	18.4	1.2	-
[8-49]	2	42	5.8	1.1	-	-
[1-49]	3	49	10.4	0.3	-	-
[79-113]	5	35	-	-	1.5	4.7
[50-113]	6	64	3.7	19.7	21.4	15.3
[50-121]	7	72	1.7	0.5	-	-
[50-123]	8	74	12.1	19.7	-	-
[8-113]	8	106	1.6	0.7	-	-
[8-121]	9	114	0.9	1.9	-	-
[1-123]	12	123	2.3	-	-	-

"-" indicates that the peptide were not found in the hydrolysate.

The quality of peptide analysis was evaluated using the average AA, peptide and molar sequence coverages, and the DH values determined based on peptide analysis (DH_{MS}). For all hydrolysates, the AA sequence coverage was 100 % (**Table 6.2**). The peptide sequence coverages of the hydrolysates decreased from 99 % to 89 % with increasing DH_{stat} values, and the molar sequence coverages ranged from 85 % to 58 %. The average peptide and molar sequence coverages were 94 % and 72 %, respectively. Based on data reported previously [16, 23, 24], the thresholds of acceptance for the average peptide and molar sequence coverages were set at 80 % and 70 %, respectively [22]. Both average coverages in this study were above the thresholds.

Table 6.2: The amino acid (AA), peptide and molar sequence coverages of the non-glycated and glycated α-LA hydrolysates with different DH values (DH2_G_{0-8}-DH8_G_{0-8}). The average values and standard deviations were calculated from the hydrolysates with the same DH_{stat} but different Maillard incubation time.

Samples	Average sequence coverage [%]			Samples	Average sequence coverage [%]		
	Amino acid	Peptide	Molar		Amino acid	Peptide	Molar
DH2_G_0	100	99	80	DH2_G_{0-8}	100±0	99±0	85±6
DH4_G_0	100	94	85	DH4_G_{0-8}	100±0	93±0	77±7
DH6_G_0	100	92	67	DH6_G_{0-8}	100±0	91±2	70±4
DH8_G_0	100	89	58	DH8_G_{0-8}	100±0	89±0	62±3
Average	100	94±4	72±12	**Average**	100±0	93±4	73±10

To further check the quality of peptide analysis, the DH values analysed with peptide analysis (DH_{MS}) of each hydrolysate was calculated, and compared with the DH_{stat} based on pH-stat titration (**Figure 6.2**). Averaged over all samples, the standard error between DH_{stat} and DH_{MS} was ~14 %. This standard error was similar to previously reported results (**Chapters 3** and **4**). Based on these quality checks, the peptide analysis for non-glycated hydrolysates was of a high quality, which allows further analysis on the glycation on each peptide.
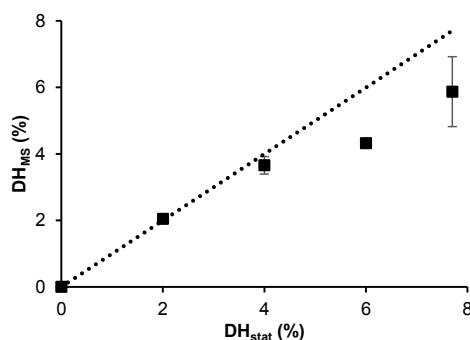
Figure 6.2: Comparison of DH determined based on peptide analysis (DH_{MS}) amd pH-stat titration (DH_{stat}) for α-LA hydrolysates. Each data point was the average of hydrolysates with the same DH_{stat} but with different Maillard incubation time (0-8 h). The dotted line (······) indicates the function y=x.

Analysis of the glycated α-LA hydrolysates

In the control samples $DH0_8$-$DH8_8$ (hydrolysates heated without D-glucose for 8 h), no glycated or lactosylated protein/peptides were found, showing that the non-glycated samples did not contain any glucose or lactose. For non-hydrolysed α-LA, DG_{MS} increased from 0-85 % after incubation from 0-24 h (**Figure 6.3A**). The fitted maximum average degree of glycation ($DG_{av,max}$) was only 86 % (using equation 10). If the fitting of the curve was done using a fixed $DG_{av,max}$ of 100 %, the sum of squared residuals was ~4 times higher than when $DG\_P_{av,max}$ was fitted. That the $DG_{av,max}$ was lower than 100 % was not due to the occurrence of protein cross-linking. This was shown by the fact that no protein aggregates were observed in the profile of size exclusion chromatography, and no unknown masses were found in the deconvoluted MS spectra (data not shown). In addition, the concentration of $DH0\_G_{24}$, injected into the UPLC-UV system was ~100 % of the expected concentration (based on the Dumas result), meaning that the obtained $DG_{av,max}$ was representative of the whole sample. The amount of added D-glucose was 2 times higher than the amount of free amino groups, which means that the low $DG_{av,max}$ was not caused by the lack of reactants. This is a strong indication that the glycation process either stopped, or reached an equilibrium while there were still free amino groups and reducing end sugars in the system. A plateau value for the DG_{MS} was previously also observed during the glycation of 3 different proteins with 7 different carbohydrates until 48 h of glycation [25]. Due to the long incubation time, the plateau value of $DG_{av,max}$ was more defined. Values of $DG_{av,max}$ in that study ranged from 6 % to 92 % for different protein-carbohydrate combinations [25].
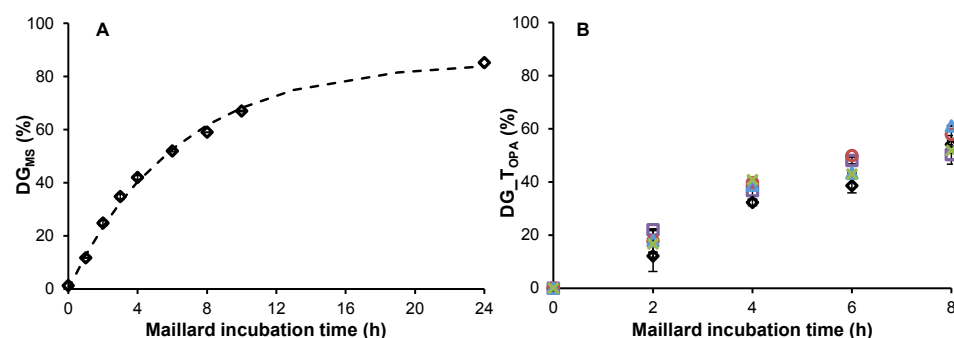
Figure 6.3: (A) Average degree of glycation of the non-hydrolysed α-LA as a function of Maillard incubation time. (B) Total degree of glycation analysed with OPA method (DG_T$_{OPA}$) *versus* Maillard incubation time for glycated α-LA hydrolysates with a DH$_{stat}$ of 0 (□), 2 (◊), 4 (○), 6 (△) and 8 (×) %. The error bars were typically smaller than the size of the marker.

Upon heating in the presence of D-glucose for 0-8 h, the total degree of glycation determined using the OPA method (DG_T$_{OPA}$) of all the glycated hydrolysates increased from 0 to ~55 % (**Figure 6.3B**). These values were similar to the results of the non-hydrolysed α-LA. This indicated that the created free α-amino groups by hydrolysis did not have a major contribution to the total glycation, which implies that the glycation of ε-amino group dominates the total degree of glycation.

Similarly as for the non-glycated hydrolysates, the quality checks of peptide analysis for glycated hydrolysates were based on the average AA, peptide, molar sequence coverages and the DH$_{MS}$ *versus* DH$_{stat}$. In addition, the concentrations of individual peptides as a function of the Maillard incubation time were compared. For the hydrolysates with the same DH$_{stat}$, the AA, peptide and molar coverages of non-glycated and glycated α-LA hydrolysates were averaged (**Table 6.2**). All hydrolysates had an AA sequence coverage of 100 %. The average peptide sequence coverages were 99±0, 93±0, 91±2 and 89±0 % for hydrolysates with a DH$_{stat}$ of 2, 4, 6 and 8 %, respectively, and the average molar sequence coverages were 85±6, 77±7, 70±4 and 62±3 % (**Annex 6.2**). Averaged over all hydrolysates, the peptide and molar sequence coverages were 93±4 and 73±10 %, respectively, which were above the thresholds for acceptance. The DH$_{MS}$ of each hydrolysate was comparable with the DH$_{stat}$ (**Figure 6.2**). Overall, there was only a minor variation in the quantified concentration of peptides in the samples with increasing Maillard incubation time, as illustrated for DH2_G$_{0-8}$ (**Table 6.3**). For samples DH2_G$_{0-8}$, the concentrations of peptides over the Maillard incubation time had an average standard error of ~17 %. When calculated over all hydrolysates, the average standard error of the concentration of peptides in hydrolysates with the same DH$_{stat}$ was ~20 % (**Annex 6.3**). This is similar to the standard error in the peptide concentrations determined in duplicate hydrolysates (~15 %) (**Chapters 3** and **4**), confirming that the quality of the peptide analysis was acceptable.

Table 6.3: The total and average concentrations of peptides (glycated and non-glycated) in hydrolysates with a DH_{stat} of 2 % ($DH2\_G_{0-8}$).

Peptides	Concentration of peptides in hydrolysates [µM]						
	$DH2\_G_0$	$DH2\_G_2$	$DH2\_G_4$	$DH2\_G_6$	$DH2\_G_8$	Average	Stderr [%]
[8-11]	4.0	4.2	4.8	5.4	4.9	4.7±0.6	12
[117-121]	0.2	0.2	0.2	0.3	0.3	0.2±0.0	21
[26-49]	0.1	0.1	0.2	0.1	0.1	0.1±0.0	35
[122-123]	8.6	12.0	14.0	12.1	15.3	12.4±2.5	20
[1-7]	10.9	9.3	10.6	9.1	10.4	10.1±0.8	8
[114-121]	3.7	3.3	3.7	2.9	3.0	3.3±0.4	11
[114-123]	3.2	2.6	3.4	2.4	2.8	2.9±0.4	15
[12-25]	0.4	0.4	0.4	0.3	0.5	0.4±0.1	22
[12-49]	5.8	6.4	7.2	6.1	5.2	6.1±0.8	12
[8-49]	5.8	6.6	7.6	7.1	5.7	6.5±0.8	13
[1-49]	10.4	9.8	10.5	11.6	15.7	11.6±2.4	20
[50-113]	3.7	4.1	5.0	4.0	3.0	4.0±0.7	19
[50-121]	1.7	1.4	1.6	1.9	1.5	1.6±0.2	12
[50-123]	12.1	15.0	16.6	15.2	18.0	15.4±2.2	14
[8-113]	1.4	1.5	1.5	1.8	1.2	1.5±0.2	14
[8-121]	0.9	1.1	1.1	1.2	0.9	1.0±0.1	12
[1-123]	2.3	1.2	1.2	1.8	2.1	1.7±0.5	28
						Average	17

[#] Typical standard error of peptides in duplicate hydrolysates is ~15 %.

In the glycated hydrolysates ($DH2\_G_{2-8}$-$DH8\_G_{2-8}$), glycated variants of all peptides were found. The glycated variants of each peptide eluted either slightly before, or at the same moment as that original peptide. The shift in elution time (**Figure 6.4A**) and mass spectra (**Figure 6.4B**) of the glycated peptides was illustrated for peptide α-LA[1-7]. This peptide contains 1 lysine, which means that the glycation can occur on the ε-amino group of that lysine and/or on the α-amino group of the N-terminal AA. After 2 h incubation, 17 % of this peptide had 1 glucose unit attached (m/z of 1012.4). The intensity ratio between the glycated and the original peptide increased over the Maillard incubation time. After 6 h incubation, the peak with an m/z of 1174.5 started to appear, which illustrated that there were 2 glucose units attached on the peptide molecule.
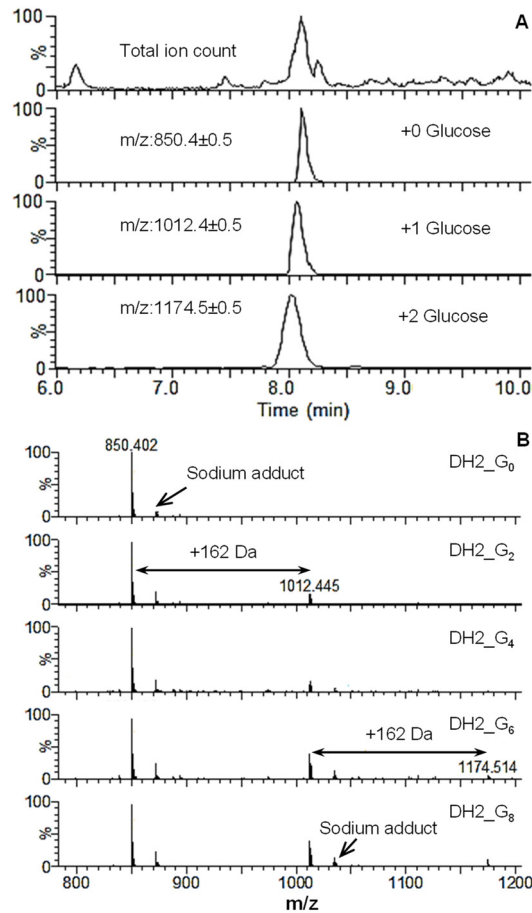
Figure 6.4: (A) Total ion count and (B) mass spectra of non-glycated and glycated peptide α-LA[1-7] analysed in the RP-UPLC chromatograms of DG2_G_{0-8}, using extracted ions at m/z of 850.4±0.5 (+0 glucose) (100 % MS intensity is $3.6 \cdot 10^4$), 1012.4±0.5 (+1 glucose) (100 % MS intensity is $3.0 \cdot 10^3$) and 1174.5±0.5 (+2 glucose) (100 % MS intensity is $1.4 \cdot 10^3$).

Using the DG_P_{av} values, the total degrees of glycation of hydrolysates (DG_T_{MS}) were calculated, and compared with the total degrees of glycation analysed using the OPA method (DG_T_{OPA}) (**Figure 6.5**). Averaged over all samples, the standard error between the DG_T determined using the two methods was ~13 %.
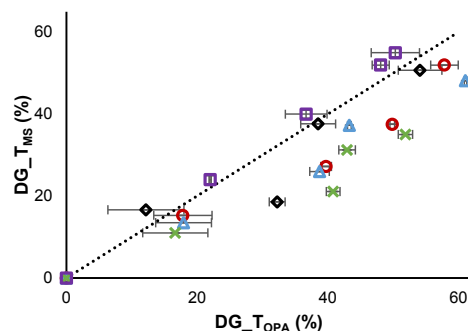
Figure 6.5: (A) Total degrees of glycation analysed with peptide analysis (DG_T_{MS}) *versus* with OPA method (DG_T_{OPA}) for hydrolysates with DH$_{stat}$ values of 0 (□), 2 (◇), 4 (○), 6 (△) and 8 (×) %. The dotted line (······) indicates the function y=x.

The maximum average degree of glycation (DG_$P_{av,max}$) and the glycation rate constant (k_g) for individual peptides in all hydrolysates were determined (**Table 6.4**). To illustrate the glycation of individual peptides in the hydrolysates, the glycation curves (DG_P_{av} *versus* time) of peptides α-LA[38-46], α-LA[114-121] and α-LA[12-49] were taken as examples because they contained 0, 1 and 2 lysine(s), respectively (**Figure 6.6A**). The DG_P_{av} values of the peptides increased with the number of lysine residues in the peptide. For peptides that were present in multiple hydrolysates (e.g. α-LA[50-113]), the glycation curves in all hydrolysates were compared (**Figure 6.6B**). Despite their different concentrations in different hydrolysates, ranging from 4.0 to 23.0 µM (**Annex 6.3**), the changes in DG_P_{av} *versus* glycation time were similar. The average standard errors for DG_$P_{av,max}$ and k_g of the same peptides in different hydrolysates were both 14 %. From this, it was concluded that the glycation of peptides in hydrolysates was independent of the hydrolysate composition. Therefore, if peptides were present in several hydrolysates, the DG_P_{av} values obtained in all hydrolysates were combined to perform the fit for DG_$P_{av,max}$ and k_g (**Table 6.4**).
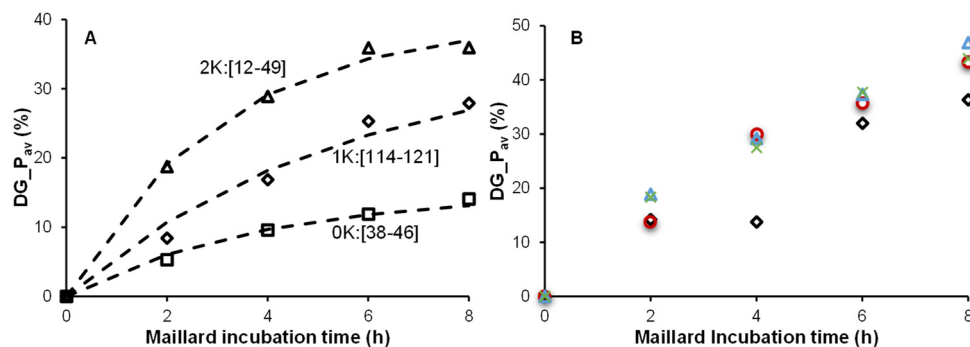
Figure 6.6: (A) Glycation curves (DG_P_{av} *versus* time) of peptide (□) α-LA[38-46] (0K: contains 0 lysine), (◇) α-LA[114-121] (1K: contains 1 lysine), (△) α-LA[12-49] (2K: contains 2 lysines) in the samples DH6_G_{0-8}. The dashed lines (------) represent the fits using 1st order reaction equation. (B) Glycation curves of peptide α-LA[50-113] in hydrolysates with DH$_{stat}$ values of 2 (◇), 4 (○), 6 (△) and 8 (×) %.

Table 6.4: The number of lysines (#Lys) per peptide, peptide length, percentage of non-glycated peptide, glycation rate constant (k_g), maximum average degree of glycation (DG_$P_{av,max}$), average number of glucose units (#glc) per peptide, maximum degree of glycation on α- (DG$_{max,α-NH2}$) and ε-amino (DG$_{max,ε-NH2}$) groups of peptides and DG$_{max,ε-NH2}$ of glycated variant of peptides, in glycated α-LA hydrolysates with a DH$_{stat}$ of 2, 4, 6 and 8 %.

Peptides	#Lys	Length	k_g [10^{-1} s^{-1}] DH2	DH4	DH6	DH8	Average	Fitted with all points	DG_$P_{av,max}$ [%] DH2	DH4	DH6	DH8	Average	Fitted with all points	Avr. #glc per peptide	% non-glycated peptide	DG$_{max,α-NH2}$ [%]	DG$_{max,ε-NH2}$ [%]	DG$_{max,ε-NH2}$ of glycated variant [%]
[8-11]*	0	4	2.4	2.0	2.0	1.8	2.0±0.1	2.5	22.0	13.8	11.5	13.8	15.3±4.6	13.0	0.13	89	13.0	-	-
[117-121]	0	5	n.d.	4.3	5.1	3.7	4.4±0.7	2.7	n.d.	17.0	12.1	14.0	14.4±2.5	15.9	0.16	83	15.9	-	-
[38-46]	0	9			2.6	2.8	2.7±0.1	2.3			15.0	12.0	13.5±2.1	14.5	0.15	87	14.5	-	-
[38-49]	0	12		1.3	1.7	1.5	1.5±0.2	1.6		11.8	18.9	10.4	13.6±4.5	11.5	0.12	80	11.5	-	-
[26-49]	0	24	n.d.	1.4	1.4		1.4±0.0	1.1	n.d.	5.7	10.0		7.9±3.0	9.1	0.09	92	9.1	-	-
[122-123]	1	2	6.1	7.3	5.1	5.8	6.1±0.9	6.0	33.7	38.2	37.7	35.7	36.3±2.1	36.4	0.73	36	12.8	60.2	94.1
[12-14]	1	3			2.2	3.1	2.6±0.7	2.6			28.6	30.4	29.5±1.3	26.9	0.54	53	12.8	46.2	98.3
[1-7]	1	7	3.5	3.6	2.6	2.8	3.1±0.5	2.8	20.9	21.0	27.0	21.1	22.5±3.0	23.0	0.46	64	12.8	32.2	89.4
[8-14]*	1	7		2.5	2.3		2.4±0.2	2.5		31.2	27.6		29.4±2.5	30.5	0.61	24	12.8	46.2	60.8
[114-121]	1	8	1.9	2.1	1.8	n.d.	2.0±0.1	1.9	22.9	32.0	35.0	n.d.	30.9±6.3	33.2	0.66	50	12.8	49.2	98.4
[114-123]	2	10	2.6	2.9	2.2		2.6±0.4	2.3	31.6	45.0	40.0		42.5±6.8	42.0	1.26	25	12.8	57.6	76.8
[12-25]	2	14	1.3	0.7	1.2	1.0	1.0±0.3	0.9	34.5	42.8	35.0	32.0	36.1±4.7	36.1	1.08	18	12.8	47.6	58.0
[57-78]	2	22				1.8	1.8±0.0	1.8				38.6	38.6±0.0	38.6	1.16	31	12.8	51.6	74.8
[12-37]	2	26		2.8	2.1	2.3	2.4±0.4	2.0		39.9	46.8	42.8	43.2±3.5	45.9	1.38	24	12.8	58.1	76.4
[12-46]	2	35		n.d.	4.5		4.5±0.0	4.5		n.d.	35.2		35.2±0.0	35.2	1.06	26	12.8	46.6	63.0
[12-49]	2	38	3.1	2.8	3.3		3.1±0.2	2.6	36.5	38.2	39.9		38.2±1.7	39.7	1.19	31	12.8	51.1	74.1
[8-49]*	2	42	3.0	3.5			3.3±0.4	2.3	34.0	32.7			33.3±0.9	36.4	1.09	37	12.8	43.6	69.2
[1-49]*	3	49	2.5	3.0			2.8±0.3	2.3	40.6	42.0			41.3±1.0	45.1	1.80	21	12.8	50.7	64.2
[79-113]	5	35			1.8	1.5	1.6±0.2	1.5			53.8	58.6	56.2±3.4	56.0	3.36	7	12.8	64.8	69.7
[50-113]	6	64	1.8	1.4	1.5	1.7	1.6±0.2	1.7	48.1	64.2	64.4	58.9	58.9±7.7	56.8	3.98	3	12.8	66.5	68.6
[50-121]	7	72	1.6	3.1			2.3±1.1	2.6	63.0	59.5			61.3±2.5	58.1	4.65	0	12.8	68.2	68.2
[50-123]	8	74	2.4	2.1			2.2±0.2	2.2	45.0	55.0			50.0±7.0	54.0	4.86	2	12.8	54.7	54.7
[8-113]*	8	106	1.7	1.7			1.7±0.0	2.0	48.7	62.6			55.6±9.9	51.9	4.67	2	12.8	60.9	62.1
[8-121]*	9	114	1.4	2.3			1.8±0.6	1.9	62.7	49.5			56.1±9.3	53.0	5.30	3	12.8	60.9	62.8
[1-123]*	12	123	2.9				2.9±0.0	2.9	57.5				57.5±0.0	57.5	7.48	0	12.8	61.3	61.3
Average																		53.9±9.0	72.2±13.2

* For these peptides, arginine was not taken into account in the #NH$_2$.
n.d. indicates the glycation kinetics was not determined due to a low amount of the peptide.

To identify whether there were clusters of populations of the k_g and DG_$P_{av,max}$ values, the data sets were fitted with a normal distribution curve. For k_g and DG_$P_{av,max}$ values, there were 3 and 5 populations identified, respectively, based on the structure of the residuals of the fit (data not shown). Hence, the data was categorised by the Matlab method k-means clustering with k (the number of clusters) as 3 and 5 for k_g and DG_$P_{av,max}$, respectively (**Table 6.5**). For the k_g values, the 3 clusters have average values of 1.6(±0.4), 2.5(±0.2) and 5.3(±1.1) s^{-1}, respectively. For the DG_$P_{av,max}$ values, the 5 clusters have average values of 12.8(±2.6) %, 26.8(±3.8) %, 36.5(±2.1) %, 44.3(±2.1) % and 55.3(±2.4) %, respectively. For the k_g values, the clusters did not seem to be correlated with the number of lysines on the peptide nor the peptide length. For the DG_$P_{av,max}$ values, clearly, the clustering had a quite good agreement with the number of lysines on the peptide. For peptides α-LA[122-123] and α-LA[12-46], the k_g values were 6.0·10^{-1} s^{-1} and 4.5·10^{-1} s^{-1} respectively. They were 2-3 times higher than the k_g values of the other 23 peptides (**Table 6.4**). For peptide α-LA[122-123], this can be caused by the fact that it is a dipeptide (KL) with leucine as the C-terminal AA and lysine as the N-terminal AA. In literature, peptide KL was reported to have the highest glycation rate among 11 dipeptides with different AAs as the C-terminal AA [12]. For peptide α-LA[12-46], no explanation was found for the high k_g value.

Table 6.5: The clusters of peptides based on the glycation rate constant (k_g) and the maximum average degree of glycation (DG_$P_{av,max}$).

Peptides	#Lys	Length	k_g [10^{-1} s^{-1}]	Cluster #k_g	DG_$P_{av,max}$ [%]	Cluster of #DG_$P_{av,max}$
[8-11]	0	4	2.5	1	13.0	1
[117-121]	0	5	2.7	1	15.9	1
[38-46]	0	9	2.3	1	14.5	1
[38-49]	0	12	1.6	2	11.5	1
[26-49]	0	24	1.1	2	9.1	1
[122-123]	1	2	6.0	3	36.4	3
[12-14]	1	3	2.6	1	26.9	2
[1-7]	1	7	2.8	1	23.0	2
[8-14]	1	7	2.5	1	30.5	2
[114-121]	1	8	1.9	2	33.2	3
[114-123]	2	10	2.3	1	42.0	4
[12-25]	2	14	0.9	2	36.1	3
[57-78]	2	22	1.8	2	38.6	3
[12-37]	2	26	2.0	2	45.9	4
[12-46]	2	35	4.5	3	35.2	3
[12-49]	2	38	2.6	1	39.7	3
[8-49]	2	42	2.3	1	36.4	3
[1-49]	3	49	2.3	1	45.1	4
[79-113]	5	35	1.5	2	56.0	5
[50-113]	6	64	1.7	2	56.8	5
[50-121]	7	72	2.6	1	58.1	5
[50-123]	8	74	2.2	1	54.0	5
[8-113]	8	106	2.0	2	51.9	5
[8-121]	9	114	1.9	2	53.0	5
[1-123]	12	123	2.9	1	57.5	5

For the other 23 peptides, as suggested by the results of the clustering, the k_g was independent of the number of lysines on the peptide (**Figure 6.7A**) and the peptide length (**Figure 6.7B**).
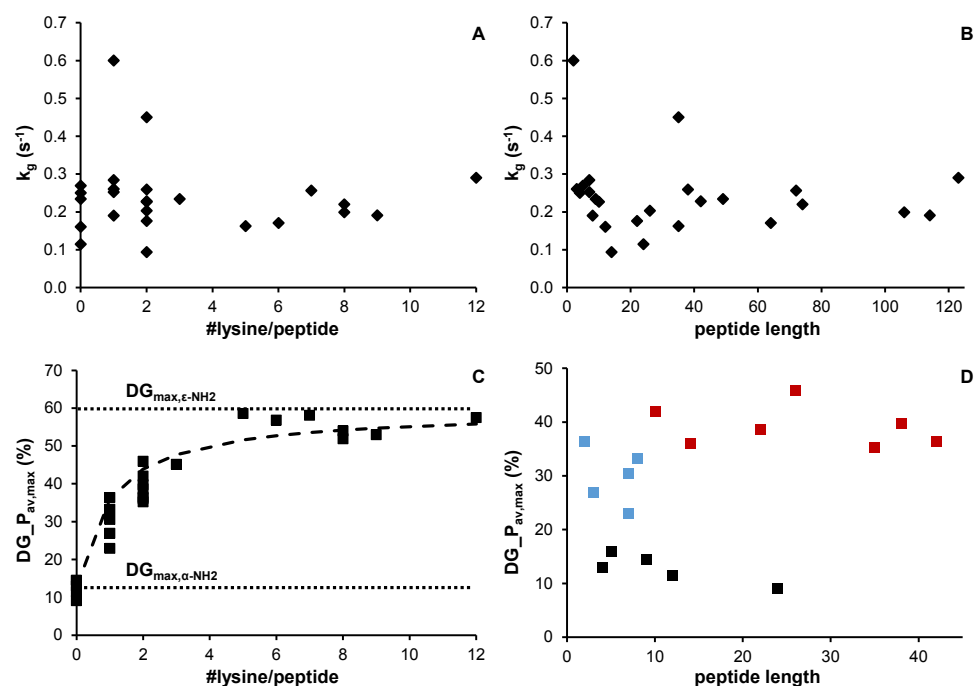
Figure 6.7: (A, B) Glycation rate constant (k_g) and (B, D) DG_$P_{av,max}$ of individual peptides in α-LA hydrolysates (DH2_G_{0-8}-DH8_G_{0-8}) as a function of (A, C) the number of lysines per peptide (#lysine/peptide) and (B, D) the peptide length. In panel (C), the dashed line (------) represents the fit using equation 11. The dotted lines (······) represent the maximum degree of glycation on α- ($DG_{max,α-NH2}$) and ε-amino ($DG_{max,ε-NH2}$) groups. (D) DG_$P_{av,max}$ *versus* peptide length of peptides that contain (■) 0, (■) 1 or (■) 2 lysine(s).

Similarly as for the intact protein, the DG_$P_{av,max}$ did not reach 100 % for all peptides. For peptides that did not contain lysines, i.e. peptides α-LA[8-11], α-LA[117-121], α-LA[38-46], α-LA[38-49] and α-LA[26-49], the DG_$P_{av,max}$ values were comparable: 12.8±2.6 % (**Table 6.4**). Because the variation in DG_$P_{av,max}$ of these peptides was small, no clear effect of the type of N-terminal AA was observed. Since these peptides only have an α-amino group as the glycation site, it means that the maximum degree of glycation for α-amino group ($DG_{max,α-NH2}$) was 12.8 %. It should be noted that peptide α-LA[8-11] contains an arginine (R10) within its AA sequence. The fact that the DG_$P_{av,max}$ of α-LA[8-11] was the same as the other peptides which only contained α-amino group as the glycation site, suggests that arginine did not react during the glycation process. In previous studies with free AA and intact proteins, the glycation of arginine was identified [7, 26]. The reason that different results were obtained between the current and previous studies could be the differences in the glycation conditions used.

For the lysine containing peptides, the DG_$P_{av,max}$ ranged from 23.0 % to 58.1 % (**Table 6.4**). The maximum degree of glycation of ε-amino groups (DG$_{max,ε-NH2}$) in each peptide was calculated using equation 11. It was found that the DG$_{max,ε-NH2}$ averaged over peptides that contained ε-amino groups was 53.9±9.0 %. To check whether the DG$_{max,ε-NH2}$ did not reach 100 % was because of the heterogeneity of the samples, the DG$_{max,ε-NH2}$ of each peptide was also calculated after correction for the amount of remaining non-glycated peptides (**Table 6.4**). The percentage of glycated ε-amino groups of the glycated peptides was 72.2±13.2 %. Since it was also lower than 100 %, it was concluded that all lysines in the hydrolysates were glycated to similar extents, but the glycation process either stopped before all available amino groups reacted or reached an equilibrium. This also indicated that the glycation reactivity of each lysine in α-LA was not significantly different from each other. This indication was previously suggested in another article, where the bovine tryptic hydrolysis of the glycated α-LA followed the theoretical scenario in which all lysines were glycated to the same extent [13]. These data also showed that on average only 12.8 % of the α-amino groups and 53.8 % of the ε-amino groups were glycated. Since the DG$_{max,ε-NH2}$ was much higher than the DG$_{max,α-NH2,}$ the DG_$P_{av,max}$ increased with the number of lysines on the peptide (#lysine/peptide) (**Figure 6.7C**). This is because for the peptides that contain more lysines, the low glycation reactivity of α-amino group would have a smaller effect on the final DG_$P_{av,max}$. This means that the number of lysines per peptides dominates the DG_$P_{av,max}$ of the peptide. The DG$_{max,ε-NH2}$ was also obtained by fitting equation 11 to the experimental data of all peptides (**Figure 6.7C**). The R^2 of the fit was 0.91. The DG$_{max,ε-NH2}$ was fitted to be 60.0 %, which was close to the average DG$_{max,ε-NH2}$ (53.9 %) obtained from individual peptides. Since DG$_{max,ε-NH2}$ was ~5 times higher than DG$_{max,α-NH2}$, it was concluded that the number of lysine of the peptide is the dominant property that determines the glycation of individual peptides in a hydrolysate.

Hydrolysates contain peptides with different numbers of lysines, meaning that the average DG_$P_{av,max}$ of these peptides would be lower than the intact proteins. However, this does not mean that the total degrees of glycation of the hydrolysates were lower than that of the intact proteins. Actually, in both hydrolysates and intact proteins, all lysines were glycated to ~60.0 % and all N-termini were glycated to ~12.8 %. Still, for peptides with the same number (0, 1 and 2) of lysines, there were variations in the DG_$P_{av,max}$ values. This variation was not correlated with peptide length (**Figure 6.7D**). This means that the glycation of peptides in a hydrolysate under the experimental condition used in this study was independent of the peptide length.

Conclusions

In this study, using a quantitative peptide analysis, the glycation kinetics of individual peptides in protein hydrolysates was followed. The extent of glycation for all peptides did not reach 100 %, meaning that the glycation process either stopped before all amino groups reacted or reached an equilibrium. The rate and extent of glycation of individual peptides in hydrolysates were

independent of the hydrolysate composition. For a peptide in a hydrolysate, the number of lysines on the peptide is the dominant factor that determines the extent of glycation. This is because the glycation reactivity of ε-amino group (lysine) is ~5 times higher than the reactivity of α-amino group (N-terminal amino acid). The glycation rate constants were independent of the peptide length or the number of amino groups on the peptide. The outcome of this work can be used to predict the extent of glycation of peptides in hydrolysates based on the amino acid sequences of the peptides.

Acknowledgement

References

[1] E.M. Mulcahy, C.W. Park, M. Drake, D.M. Mulvihill, J.A. O'Mahony, Improvement of the functional properties of whey protein hydrolysate by conjugation with maltodextrin, International Dairy Journal, 60 (2016) 47-54.

[2] J.-S. Kim, Y.-S. Lee, Study of Maillard reaction products derived from aqueous model systems with different peptide chain lengths, Food Chemistry, 116 (2009) 846-853.

[3] G. Münch, D. Schicktanz, A. Behme, M. Gerlach, P. Riederer, D. Palm, R. Schinzel, Amino acid specificity of glycation and protein-AGE crosslinking reactivities determined with a dipeptide SPOT library, Nature Biotechnology, 17 (1999) 1006-1010.

[4] I. Penndorf, D. Biedermann, S.V. Maurer, T. Henle, Studies on N-terminal glycation of peptides in hypoallergenic infant formulas: Quantification of α-N-(2-Furoylmethyl) amino acids, Journal of Agricultural and Food Chemistry, 55 (2007) 723-727.

[5] D.J. Cowley, Triazinylaniline derivatives as fluorescence probes. Part 4. Kinetics and selectivity in the reactions of N,N-diethyl-4-(dichloro-1,3,5-triazinyl)aniline with amines, amino acids and proteins relevant to fluorescence labelling, Journal of the Chemical Society, Perkin Transactions 2, (1996) 2431-2439.

[6] E.H. Ajandouz, A. Puigserver, Nonenzymatic browning reaction of essential amino acids: Effect of pH on caramelization and Maillard reaction kinetics, Journal of Agricultural and Food Chemistry, 47 (1999) 1786-1793.

[7] E.-J. Kwak, S.-I. Lim, The effect of sugar, amino acid, metal ion, and NaCl on model Maillard reaction under pH control, Amino Acids, 27 (2004) 85-90.

[8] L. Lamberts, I. Rombouts, J.A. Delcour, Study of nonenzymic browning in α-amino acid and γ-aminobutyric acid/sugar model systems, Food Chemistry, 111 (2008) 738-744.

[9] Y. Bai, H. Ueno, J.M. Manning, Some factors that influence the nonenzymatic glycation of peptides and polypeptides by glyceraldehyde, Journal of Protein Chemistry, 8 (1989) 299-315.

[10] J. Venkatraman, K. Aggarwal, P. Balaram, Helical peptide models for protein glycation: Proximity effects in catalysis of the Amadori rearrangement, Chemistry and Biology, 8 (2001) 611-625.

[11] G.R. Grimsley, J.M. Scholtz, C.N. Pace, A summary of the measured pK values of the ionizable groups in folded proteins, Protein Science, 18 (2009) 247-251.

[12] C. Mennella, M. Visciano, A. Napolitano, M.D. Del Castillo, V. Fogliano, Glycation of lysine-containing dipeptides, Journal of Peptide Science, 12 (2006) 291-296.

[13] Y. Deng, P.A. Wierenga, H.A. Schols, S. Sforza, H. Gruppen, Effect of Maillard induced glycation on protein hydrolysis by lysine/arginine and non-lysine/arginine specific proteases, Food Hydrocolloids, 69 (2017) 210-219.

[14] C.I. Butré, P.A. Wierenga, H. Gruppen, Influence of water availability on the enzymatic hydrolysis of proteins, Process Biochemistry, 49 (2014) 1903-1912.

[15] O. Akpinar, M.H. Penner, Peptidase activity assays using protein substrates, Current protocols in food analytical chemistry, John Wiley & Sons, Corvallis, OR, USA, 2001, pp. Unit C2.2.4.

[16] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Introducing enzyme selectivity: A quantitative parameter to describe enzymatic protein hydrolysis, Analytical and Bioanalytical Chemistry, 406 (2014) 5827-5841.

[17] J. Adler-Nissen, Enzymic hydrolysis of food proteins, Elsevier Applied Science Publishers, London, UK, 1986, pp. 122-123.

[18] C.I. Butré, P.A. Wierenga, H. Gruppen, Effects of ionic strength on the enzymatic hydrolysis of diluted and concentrated whey protein isolate, Journal of Agricultural and Food Chemistry, 60 (2012) 5644-5651.

[19] R.J.B.M. Delahaije, P.A. Wierenga, M.L.F. Giuseppin, H. Gruppen, Improved emulsion stability by succinylation of patatin is caused by partial unfolding rather than charge effects, Journal of Colloid and Interface Science, 430 (2014) 69-77.

[20] B.J.H. Kuipers, H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis, Journal of Agricultural and Food Chemistry, 55 (2007) 5445-5451.

[21] H.A. Kosters, P.A. Wierenga, R. de Vries, H. Gruppen, Characteristics and effects of specific peptides on heat-induced aggregation of β-lactoglobulin, Biomacromolecules, 12 (2011) 2159-2170.

[22] Y. Deng, F. van der Veer, S. Sforza, H. Gruppen, P.A. Wierenga, Towards predicting protein hydrolysis by bovine trypsin, Process Biochemistry, 65 (2018) 81-92.

[23] C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Determination of the influence of the pH of hydrolysis on enzyme selectivity of *Bacillus licheniformis* protease towards whey protein isolate, International Dairy Journal, 44 (2015) 44-53.

[24] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Determination of the influence of substrate concentration on enzyme selectivity using whey protein isolate and *Bacillus licheniformis* protease, Journal of Agricultural and Food Chemistry, 62 (2014) 10230-10239.

[25] H.B. Cardoso, P.A. Wierenga, H. Gruppen, H.A. Schols, Maillard induced glycation behaviour of individual milk proteins, Food Chemistry, 252 (2018) 311-317.

[26] F. Fenaille, F. Morgan, V. Parisod, J.-C. Tabet, P.A. Guy, Solid-state glycation of β-lactoglobulin by lactose and galactose: Localization of the modified amino acids using mass spectrometric techniques, Journal of Mass Spectrometry, 39 (2004) 16-28.

Annex 6.1: List of non-glycated and glycated peptides derived from α-LA hydrolysed by *Bacillus licheniformis* protease.

Peptides	Sequence	+ glucose units	Observed mass (Da)	Theoretical mass (Da)	Error (ppm)	b fragments identified	y fragments identified
[1-7]	EQLTKCE	1,2	849.402	849.390	14	2,3,4,5,6	2,3,4,5,6
[1-49]	EQLTKCEVFRELKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQN NDSTE	1,2,3,4	5535.663	5535.624	7	3,8	1,2,3,5,6,7,8,9,10,33(2+),35(2+)
[1-123]	EQLTKCEVFRELKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQN NDSTEYGLFQINNKIWCKDDQNPHSSNICNISCDKFLDDDLTDDI MCVKKILDKVGINYWLAHKALCSEKLDQWLCEKL	1 to 11	Deconvoluted mass: 14186	Avg mass: 14186	0		
[8-11]	(E)VFRE	1	549.295	549.291	7	2,3	2,3
[8-14]	(E)VFRELKD	1,2	905.499	905.497	2	1,2	3
[8-49]	(E)VFRELKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTE	1,2,3,4	4704.264	4704.244	4		1,2,3,4,5,6,7,8,9,10,33(2+),35(2+)
[8-113]	(E)VFRELKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTE YGLFQINNKIWCKDDQNPHSSNICNISCDKFLDDDLTDDIMCVK KILDKVGINYWLAHKALCSE	1 to 6	Deconvoluted mass: 12096	Avg mass: 12096.61	0		
[8-121]	(E)VFRELKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTE YGLFQINNKIWCKDDQNPHSSNICNISCDKFLDDDLTDDIMCVK KILDKVGINYWLAHKALCSEKLDQWLCE	1 to 8	Deconvoluted mass: 13112	Avg mass: 13112	0		
[12-14]	(E)LKD	1	374.228	374.216	32	2	3
[12-25]	(E)LKDLKGYGGVSLPE	1,2,3	1474.808	1474.803	3	3,5,8,9,10,11,12	3,10,15
[12-37]	(E)LKDLKGYGGVSLPEWVCTTFHTSGYD	1,2	2872.378	2872.374	1	2,3,4,5,7,8,9,11,12	2,3,4,5,6,8,10,11,12
[12-46]	(E)LKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNND	1,2	3855.864	3855.841	6	3,25(2+)	2,3,4,5,6,7,21(2+)
[12-49]	(E)LKDLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNNDSTE	1,2,3	4173.006	4172.964	10	3,8,35(3+)	2,3,4,5,6,7,8,9,10,28(2+)
[26-49]	(E)WVCTTFHTSGYDTQAIVQNNDSTE	1	2716.184	2716.171	5	8,9,20(2+),23(2+)	6,8,9
[38-46]	(D)TQAIVQNND	1	1001.489	1001.478	11	1,2,3,4,5,6,7,10,11,12	1,2,6,7,8,9
[38-49]	(D)TQAIVQNNDSTE	1	1318.609	1318.600	7	1,2,3,4,5,6,7,10,11,12	1,2,6,7,8,9
[50-113]	(E)YGLFQINNKIWCKDDQNPHSSNICNISCDKFLDDDLTDDIMC VKKILDKVGINYWLAHKALCSE	1 to 6	7402.568	7402.537	4	2,3,4,5	1,7,8
[50-121]	(E)YGLFQINNKIWCKDDQNPHSSNICNISCDKFLDDDLTDDIMC VKKILDKVGINYWLAHKALCSEKLDQWLCE	1 to 7	8418.125	8418.017	13	2,3,4,5	1,2,4,5
[50-123]	(E)YGLFQINNKIWCKDDQNPHSSNICNISCDKFLDDDLTDDIMC VKKILDKVGINYWLAHKALCSEKLDQWLCEKL	1 to 8	8660.230	8659.195	120	2,3,4,5	2,3,4,5
[57-78]	(N)NKIWCKDDQNPHSSNICNISCD	1	2533.13	2533.078	21		7
[79-113]	(D)KFLDDDLTDDIMCVKKILDKVGINYWLAHKALCSE	1 to 5	4052.064	4052.046	4	1,3,4,5,6	2,6,7,8,15(2+,16(2+)
[114-121]	(E)KLDQWLCE	1	1033.496	1033.490	6	1,4,6	
[114-123]	(E)KLDQWLCEKL	1,2,3	1274.686	1274.669	13	1,2,3,4,5,6,7,8,9	2,3,4,5,6,7,8,9
[117-121]	(D)QWLCE	1	677.296	677.284	18	2,3	
[122-123]	(E)KL	1,2	259.198	259.190	31	1	2

Annex 6.2: Molar sequence coverages of non-glycated and glycated α-LA hydrolysates with different DH_{stat} and heated with D-glucose for different time. The dotted lines are the expected concentration based on Dumas results.
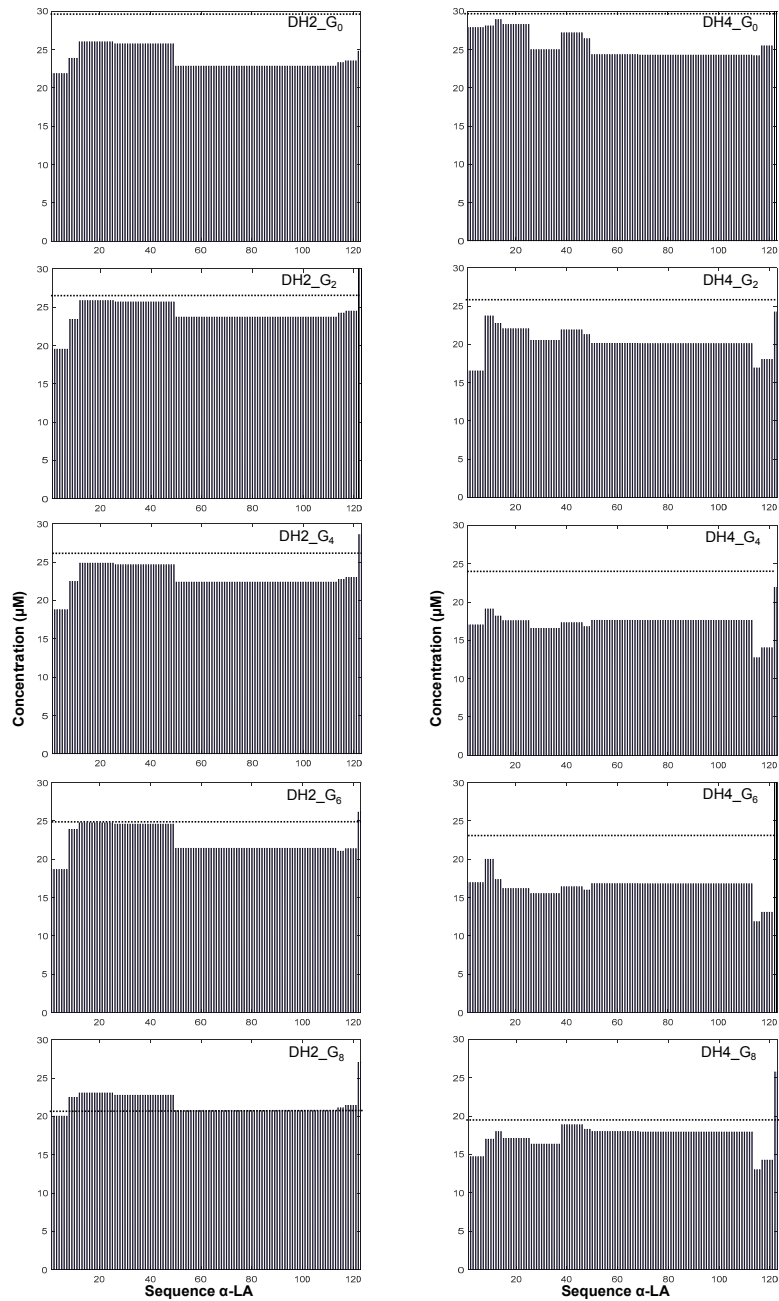
Annex 6.2: Continued.
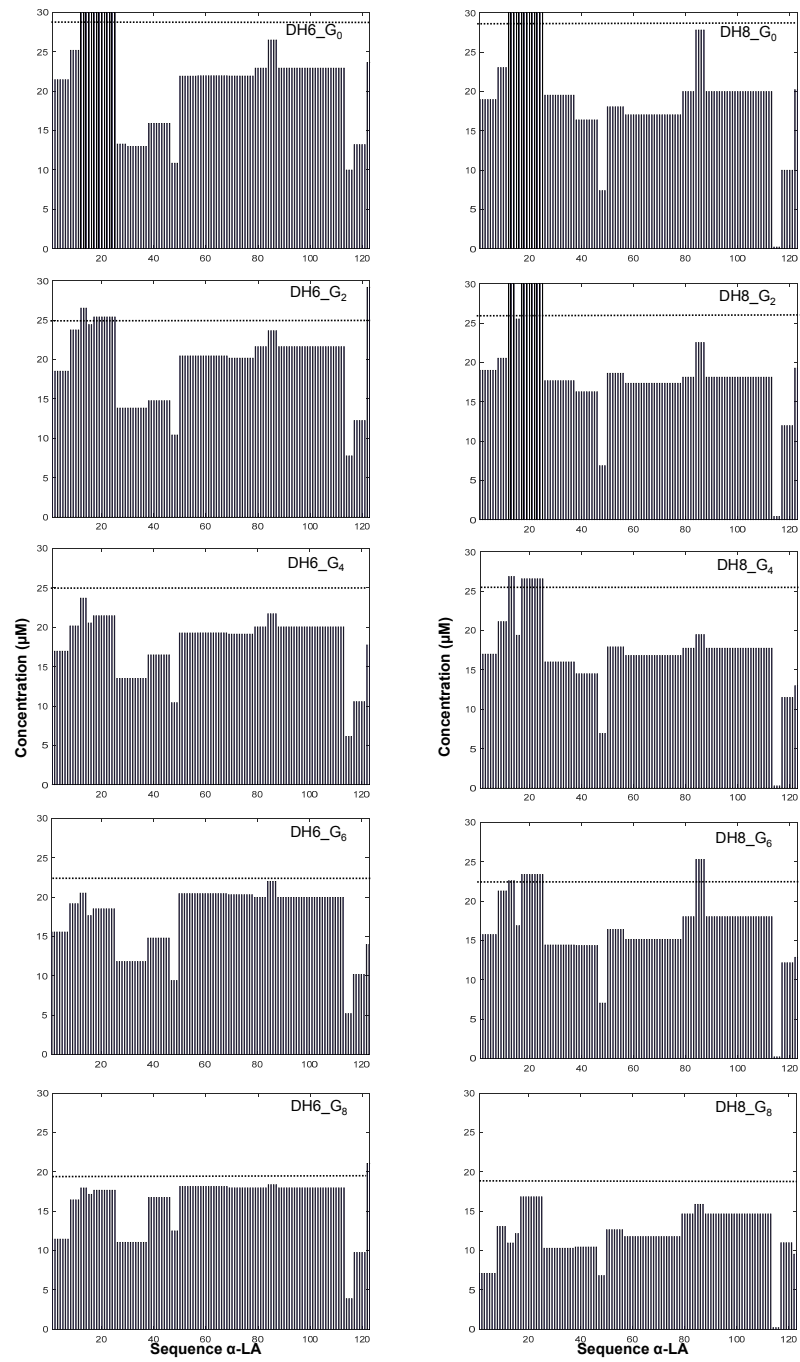
Annex 6.3: Average total concentration of peptides (glycated and non-glycated) in hydrolysates with a DH_{stat} of 2, 4, 6, and 8 % during 0-8 h Maillard incubation time.

Peptides	#lysines	Length	Concentration of peptides in hydrolysates (µM)			
			$DH2\_G_{0-8}$ (17)	$DH4\_G_{0-8}$ (20)	$DH6\_G_{0-8}$ (17)	$DH8\_G_{0-8}$ (13)
[8-11]	0	4	4.7±0.6	21.1±1.3	23.3±2.0	22.6±2.2
[117-121]	0	5	0.2±0.0	1.5±0.3	5.6±2.2	12.9±2.5
[38-46]	0	9	-	-	2.9±0.4	7.8±1.8
[38-49]	0	12	-	2.3±1.1	9.2±2.8	0.04±0.01
[26-49]	0	24	0.1±0.0	1.3±0.2	1.3±0.2	
[122-123]	1	2	12.4±2.5	30.4±5.8	24.7±7.6	16.9±3.1
[12-14]	1	3	-	-	2.8±0.4	8.3±1.0
[1-7]	1	7	10.1±0.8	21.6±3.3	19.4±1.7	17.4±3.8
[8-14]	1	7	-	1.0±0.4	1.1±0.4	-
[114-121]	1	8	3.3±0.4	14.8±3.0	7.2±1.6	0.3±0.1
[114-123]	2	10	2.9±0.4	0.8±0.1	6.3±2.1	-
[12-25]	2	14	0.4±0.1	3.1±0.9	18.8±6.3	22.9±8.6
[57-78]	2	22	-	-	-	1.0±0.2
[12-37]	2	26	-	0.3±0.1	3.2±1.1	0.7±0.3
[12-46]	2	35	-	0.6±0.2	1.7±0.9	-
[12-49]	2	38	6.1±0.8	16.5±1.8	2.3±1.0	-
[8-49]	2	42	6.5±0.8	0.5±0.1	-	-
[1-49]	3	49	11.6±2.4	0.3±0.1	-	-
[79-113]	5	35	-	-	1.6±0.4	2.1±0.2
[50-113]	6	64	4.0±0.7	19.7±1.9	23.0±3.0	16.6±0.8
[50-121]	7	72	1.6±0.2	0.4±0.1	-	-
[50-123]	8	74	15.4±2.2	1.0±0.3	-	-
[8-113]	8	106	1.5±0.2	0.5±0.1	-	-
[8-121]	9	114	1.0±0.1	1.5±0.3	-	-
[1-123]	12	123	1.7±0.5	-	-	-

"-" indicates that the peptide were not found in the hydrolysate.
* Typical standard error of peptides in duplicate hydrolysates is ~15 %.

CHAPTER 7

Industrial and societal relevance

Part of the risk analysis and technology assessment for the NanoNextNL program

This PhD research is one of the topics under project 05D02 "Ingredient isolation" within the theme "Food" of the Dutch consortium NanoNextNL. This section considers the risk analysis and technology assessment within the scope of this thesis, which is an essential part of the NanoNextNL programme. Risk analysis (RA) considers the issues of potential risks (of chemical substances) to human or environmental health. Technology assessment (TA) concerns the broader industrial and societal impacts of the technological innovations. A TA reflection towards the end of this work is relevant due to the potential impacts of technology developed in this thesis on other research fields, food and feed industries as well as the society. Commonly, food and feed industries manufacture the products via industrial processing. These processes might unintentionally influence the structure of the proteins in the products, which subsequently alter the protein functionality, e.g. protein digestibility. It is beneficial for food industries to control the digestibility of the products. Hence, it is important to know how and how much the industrial processes can influence the protein digestion. Protein digestion in human body is mainly done by digestive enzymes in the gastrointestinal tract. Mechanistically, to investigate the influence of industrial processes on protein digestibility is narrowed down to the research topic: the effect of protein structural modifications on enzymatic hydrolysis by digestive enzymes, which is extensively addressed in this thesis. By implementing the knowledge and conclusions obtained in this work, food and feed industries could -in the future- better control different processes to fine-tune the digestibility of their products. Such implementation could also lead to a better understanding of the relation between food intake and health. Although a lot of efforts have been made from different research groups towards the understanding and the prediction of protein digestibility, two challenges are described and the steps to move forward are discussed.

Y. Deng, H.A. Schols, P.A. Wierenga

Challenges in the development of controlling protein digestion

The methods used in laboratories cannot be directly applied to products from the industries. The industrial partner within the project also recognised this gap. Food products do not only contain proteins but also many other ingredients, which the matrix would interfere with the analysis. Secondly, food science and animal nutrition science have different definitions and determination methods of protein digestibility. It is a challenge for the industries to interpret reported data from different fields to improve the processing and control the digestibility of their products. These two challenges were commented in more details in the following sections.

Challenge 1: Integration of different factors that influence protein digestibility

Protein digestibility can be influenced by many factors, e.g. modifications of protein structure [1], protein concentration [2], protein aggregation [3], pH [4] and temperature [5], etc. On top of those, other ingredients in the food products might interact with proteins, which influences the protein digestibility further. Although these factors have been addressed in previous studies, it remains difficult to weigh the importance of each factor. For example, in this thesis, when 50 % of the amino groups on a protein were glycated, the $DH_{max,exp}$ of hydrolysis by bovine trypsin decreased by 50 % (**Chapter 2**). However, in **Chapter 5**, if the protein concentration decreased from 1 % to 0.1 %, the $DH_{max,exp}$ decreased by 80 %. Therefore, it is challenging to state for different circumstances, with other ingredients present, which factor has more weight on the final digestibility. This is difficult to study because some of the analytical methods developed in our laboratory cannot be directly applied to samples taken during industrial processing.

Challenge 2: Differences in terminology and determination of protein digestibility

Protein digestibility has been widely studied in both food and animal nutrition sciences. There are many differences in the terminology and the means of detection between these two fields. In food science, protein digestibility is investigated by performing the *in vitro* enzymatic protein hydrolysis to simulate the gastrointestinal digestion of proteins. One common way to express protein digestibility is the degree of hydrolysis. This information is on the level of the breakage of peptide bonds, which is the main difference between food science and other fields. Besides the degree of hydrolysis, the percentage of intact protein and the molecular weight distribution of the peptides are supplementing the information. Moreover, peptide analysis, including both identification and quantification of the peptides, is the detailed analysis of which and how much the compounds are formed at what time during the hydrolysis [6]. The molecular level of information allows to find the mechanism of how enzymes act on substrates. The focus is on the mode of action of the enzymes and the influence of the modifications of proteins on the enzymatic hydrolysis. In food science, many studies used single enzyme incubation, which is suitable for the analysis of enzyme-substrate interaction but not ideal for the simulation of the protein digestion. Besides incubation using a single enzyme, the sequential hydrolysis has been received some attention. However, many studies used diluted systems with pure and soluble

proteins. It was proven that concentrated protein systems behave differently to diluted systems due to the low free to bond water ratio [7]. Since it is *in vitro* experiment, the absorption of the peptides in the body as well as the action of the stomach muscles are not taken into account. In addition, the digestive enzymes used to perform the *in vitro* hydrolysis are also crucial to the final digestibility. Due to the high expenses of human enzymes, the digestive enzymes from different sources (e.g. porcine and bovine) are commonly chosen to simulate the human gastrointestinal digestion. However, as shown in this thesis, the extents of hydrolysis by bovine, porcine or human trypsins were quite different (**Chapter 4**). Hence, cautions need to be taken when comparing results from experiments performed with enzymes from different sources.

In the field of animal nutrition, both *in vitro* and *in vivo* experiments are commonly performed. For *in vitro* digestion studies, the digestibility is normally expressed as the solubility of the final hydrolysates [8]. It is assumed that the soluble part of the protein materials is the part that is digested. The difference between the solubility of different samples is the indication of high/low digestibility of the proteins. It should be noted that the protein solubility of the starting materials are not always the same, which brings uncertainty to the comparison. Therefore, using solubility as the only indicator to evaluate protein digestibility would easily give false negative or positive results. However, this issue has not been taken into account in many cases. In addition, during protein hydrolysis, due to the exposure of hydrophobic groups in the amino acid sequence, the solubility of the hydrolysates can vary along with the progress of hydrolysis. Hence, the protein solubility can be one indication of the extent of protein hydrolysis but should not be the only indication. For *in vivo* digestion studies, the apparent digestibility is defined as the difference between the amount of nitrogen ingested and excreted, expressed as a proportion of nitrogen ingested [9]. After the correction of endogenous nitrogen of the basal diet, the true digestibility is obtained. However, the nitrogen in the faecal sample also represents a significant proportion of non-dietary AAs (extra endogenous nitrogen), from sources such as the mucus, cells and bile, for which a correction should be made. After the correction of the basal and the extra endogenous nitrogen, the real digestibility is obtained, which reflects the absolute amounts of nitrogen and AAs that are absorbed through the gastrointestinal tract [9]. It is a global parameter to express how much nitrogen has been lost during digestion. However, it does not directly correlate with the data from the solubility experiment, which is the indicator of the *in vitro* digestion method in animal nutrition. Thus, it is important to be aware of the fact that there are possible incoherent conclusions from the *in vitro* and *in vivo* results due to differences in methods and calculations.

Because of the differences in the terminology and methods, incomparable results are expected between *in vitro*, *in vivo* data from animal sciences and *in vitro* data from food sciences. However, food industries usually collaborate with both science disciplines. These differences might cause some confusions.

Moving forward towards harmonization

For the first challenge, in order to better understand and control protein digestion, it is essential to investigate how much each factor influences the protein digestion. Systematic approaches of analysing protein structure and hydrolysis are required. The multiple factors that might influence the hydrolysis should be integrated to the model systems. Both experimental analysis and computational modelling need to achieve higher levels. A computational simulation model to perform protein hydrolysis might be useful to reduce the complexity and the duration of the experimental input. By applying these approaches, it might become possible to fine-tune the protein digestibility by using one factor to compensate the loss of digestibility caused by another factor. It will be possible for the industries to predict protein digestibility based on the processes.

For the second challenge, the misunderstanding of terminology and determination methods can be eliminated by initiating interdisciplinary dialogues between scientists in the two fields and food industries. Both regular public seminars and one-to-one consultation should be arranged. Scientists from both fields can write joint review articles to explain the differences. By addressing these differences more frontally, it would allow the industries to match the most relevant field to consult the research group and the collaboration would be more efficient. For the *in vitro* hydrolysis, recently, some attempts were made to approach harmonization. In 2014, a detailed protocol and a line-by-line guidance of a standardised *in vitro* digestion method, the harmonized INFOGEST method, were published [10]. The aim was to set up the experimental conditions that are close to physiological situations of human body, and that are practical and can be used as a basic method to various research questions. The work has been cited nearly 400 times in 3 years, proving that a standard method is in urgent needs. Although it is a step forward towards the harmonization of *in vitro* protein digestion models, it should be noted that the outcome of this method is not the real protein digestibility in human bodies. The following problems were found: 1) static model lacks of realistic enzyme to substrate ratios, pH changes, transit times and removal of released peptides; 2) the enzymes used do not act the same way as human enzymes, and the brush border enzymes that are responsible for the final digestion of peptides were not taken into account; 3) no standard methods are proposed to express the protein digestibility.

In summary, it is beneficial for food industries to understand the relation between the industrial processing and the protein digestion, in order to predict and fine-tune the protein digestibility of their products. It remains challenging to weigh the importance of each factor that affects the protein digestibility, indicating that the conclusions from model systems in laboratories were not translated to real-life scenarios. In addition, major differences of terminology and methods of protein digestibility between food science and animal nutrition science were identified, leading to misinterpretation. These challenges can be solved by integrating the factors that affect the protein hydrolysis to model systems preferably using computational simulations, and by opening interdisciplinary dialogues between the two fields and the industries.

Acknowledgement

References

[1] E. Leeb, A. Götz, T. Letzel, S.C. Cheison, U. Kulozik, Influence of denaturation and aggregation of β-lactoglobulin on its tryptic hydrolysis and the release of functional peptides, Food Chemistry, 187 (2015) 545-554.

[2] W. Qi, Z. He, Enzymatic hydrolysis of protein: Mechanism and kinetic model, 1 (2006) 308-314.

[3] I.B. O'Loughlin, B.A. Murray, P.M. Kelly, R.J. Fitzgerald, A. Brodkorb, Enzymatic hydrolysis of heat-induced aggregates of whey protein isolate, Journal of Agricultural and Food Chemistry, 60 (2012) 4895-4904.

[4] S.C. Cheison, E. Leeb, J. Toro-Sierra, U. Kulozik, Influence of hydrolysis temperature and pH on the selective hydrolysis of whey proteins by trypsin and potential recovery of native alpha-lactalbumin, International Dairy Journal, 21 (2011) 166-171.

[5] S.C. Cheison, M. Schmitt, E. Leeb, T. Letzel, U. Kulozik, Influence of temperature and degree of hydrolysis on the peptide composition of trypsin hydrolysates of β-lactoglobulin: Analysis by LC–ESI-TOF/MS, Food Chemistry, 121 (2010) 457-467.

[6] C.I. Butré, Introducing enzyme selectivity as a quantitative parameter to describe the effects of substrate concentration on protein hydrolysis, PhD thesis Wageningen University & Research, Wageningen, the Netherlands, 2014.

[7] C.I. Butré, P.A. Wierenga, H. Gruppen, Influence of water availability on the enzymatic hydrolysis of proteins, Process Biochemistry, 49 (2014) 1903-1912.

[8] S. Boisen, J.A. Fernández, Prediction of the total tract digestibility of energy in feedstuffs and pig diets by *in vitro* analyses, Animal Feed Science and Technology, 68 (1997) 277-286.

[9] A.J. Darragh, S.M. Hodgkinson, Quantifying the digestibility of dietary protein, The Journal of Nutrition, 130 (2000) 1850S-1856S.

[10] M. Minekus, M. Alminger, P. Alvito, S. Ballance, T. Bohn, C. Bourlieu, F. Carriere, R. Boutrou, M. Corredig, D. Dupont, C. Dufour, L. Egger, M. Golding, S. Karakaya, B. Kirkhus, S. Le Feunteun, U. Lesmes, A. Macierzanka, A. Mackie, S. Marze, D.J. McClements, O. Menard, I. Recio, C.N. Santos, R.P. Singh, G.E. Vegarud, M.S.J. Wickham, W. Weitschies, A. Brodkorb, A standardised static *in vitro* digestion method suitable for food - an international consensus, Food & Function, 5 (2014) 1113-1124.

CHAPTER 8

General discussion

In food industry, enzymatic hydrolysis of proteins is used to produce protein hydrolysates as ingredients in food products. Moreover, using enzymatic hydrolysis as the alternative method to simulate the *in vivo* protein digestion is often of a high interest to food and feed industries. Some studies have aimed to understand or describe the outcome of a protein hydrolysis, e.g. the hydrolysate composition and the (bio-)functionality of the total hydrolysate, as a function of the hydrolysis conditions. However, there is currently not even a method to define expectations for a simple parameter such as the final degree of hydrolysis reached. This is illustrated by the fact that the estimation of the theoretical maximum degree of hydrolysis ($DH_{max,theo}$), calculated by the content (in % of number) of cleavage sites (CSs) of a protease in a protein, did not match the experimental maximum degree of hydrolysis ($DH_{max,exp}$) (**Chapters 2-5**). The goal of this PhD project was to define expectation of the extent of protein hydrolysis based on the molecular properties of the protein and the protease, and effects of modifications on protein primary structure and changes in system conditions.

Prediction model of protein hydrolysis

By combining the results of all previous chapters, we developed a model to predict the $DH_{max,exp}$ of hydrolysis of a given protein by a specific protease (equation 1). The prediction model was based on 1) <u>the enzyme secondary specificity</u> and 2) <u>the enzyme affinity towards intact proteins</u> (P^{intact}). These in turn depend on the protein amino acid (AA) sequence, as well as <u>possible modifications</u>, and the enzyme sensitivity towards the AAs surrounding the CSs. The P^{intact} depends in principle on the protein structural stability, but was also found to be influenced by the <u>substrate concentration</u> used in the protein hydrolysis. The predicted maximum degree of hydrolysis ($DH_{max,pre}$) can be calculated using equation 1:

$$(1) \quad DH_{max,pre}[\%] = DH_{max,theo} \cdot a \cdot b - \frac{m \times n \times q}{\# \, peptide \, bonds}$$

<u>Parameters derived from literature:</u>
Information needed: the enzyme specificity and the number of individual AAs per protein.

#peptide bonds is the number of peptide bonds per substrate protein, from data reported in the Uniprot database (http://www.uniprot.org/) for most proteins.

DH_{max,theo} is the theoretical maximum degree of hydrolysis based on the enzyme specificity, which is reported in the MEROPS database (https://www.ebi.ac.uk/merops/) [1] for most of the proteases. $DH_{max,theo}$ is calculated by dividing the number of cleavage sites (#CS) in a protein (derived from the AA sequence of the protein and the enzyme specificity) by the number of peptide bonds per protein.

Parameters derived from additional experiments:

Information needed: the enzyme secondary specificity and the AA sequence of the protein.

a (0≤*a*≤100 %) is the percentage of CSs that are hydrolysed within a realistic timescale of the hydrolysis experiment. *a* is derived from the enzyme selectivity towards CSs in a protein, which is the quantitative parameter that expresses the consequence of enzyme secondary specificity (equation 2).

$$(2) \quad a \; [\%] = \frac{\#\,HSS + 0.24 \times \#\,ISS + 0.03 \times \#\,LSS}{\#\,CS} \times 100 \; \%$$

where #HSS, #ISS and #LSS refer to the numbers of high, intermediate and low selectivity sites per protein, respectively. These numbers can be derived from the AA sequence of the protein knowing the secondary specificity. The AA sequences of most proteins are reported in the Uniprot database. If the enzyme secondary specificity is unknown, #HSS, #ISS and #LSS can be determined as described in **Chapters 3** and **4**.

b (0≤*b*≤100 %) is the normalised enzyme affinity towards intact proteins ($P^{intact} / P^{intact}_{max}$, where P^{intact}_{max} is the P^{intact} of the hydrolysis that reached the highest $DH_{max,exp}$). Currently, *b* cannot be predicted but should be determined experimentally. *b* is sensitive to changes in substrate concentration used in the hydrolysis, and seems to affect the $DH_{max,exp}$ because it changes the hydrolysis mechanism (**Chapter 5**). It seems that with different mechanisms, different amounts of inhibitory peptides are formed, leading to a change in $DH_{max,exp}$. This also means that *b* depends on the molar ratio between the inhibitory peptides formed during the hydrolysis and the amount of enzyme added (**Chapter 5**).

It should be noted that changing parameter *b* should in principle not change the parameter *a*, but it might influence the measurement of *a*. If *a* is measured in a hydrolysis condition where *b* is close to zero, since the P^{intact} is too low, the variation in the enzyme selectivity towards CSs in a protein would be too small to be distinguished. Under such conditions, the enzyme secondary specificity cannot be obtained from the enzyme selectivity.

Parameters derived from the analysis of modifications:

Information needed: the possible modification sites on the protein AA sequence

m (0≤*m*≤100 %) refers to the percentage of CSs that have a possible modification site on at least one of the binding site positions (the neighbouring AAs of CSs) (**Chapter 2**).

q is the number of hydrolysed CSs within a realistic timescale of the experiment that are influenced by the modifications. In the case that the modifications occurred on the possible modification sites equally, *q* is calculated from equation (3):

(3) $q = \left(\dfrac{DM_{av}}{100} \times \#MS \right) \times a \times b$

where DM_{av} is the average degree of modification (percentage of reacted groups) of the protein and $\#MS$ is the number of possible modification sites per protein. In the case of glycation, and most other typical protein modifications, the amino groups in the protein react. This means that in equation 3, a small correction needs to be done for the possible modification of the α-amino group of the protein, leading to equation (4),

(4) $q_g = \left(\dfrac{DG_{av}}{100} \times \left(\#GS - 1 \right) \right) \times a \times b$

where DG_{av} [%] is the average degree of glycation (percentage of glycated amino groups) of the protein and $\#GS$ is the number of possible glycation sites per protein (**Chapters 2** and **6**).

n ($0 \leq n \leq 1$) refers to the binding site dependency coefficient, describing if the interaction between the enzyme and the CS is ($n=1$) or is not ($n=0$) influenced by the modifications of AAs on the binding site sequence (**Chapter 2**).

Parameter estimates for protein hydrolysis obtained in this thesis

These 5 parameters and their values for certain combinations of protein and protease, with modification through glycation, were obtained/derived from previous chapters (**Table 8.1**).

Table 8.1: The values of the 5 parameters in equation 1 (a, b, m, n and q_g) determined for the enzyme-substrate combinations used in this thesis. The white cells were determined in previous chapters. The grey cells were calculated based on the obtained conclusions and equations.

Substrate	Enzyme	a^1 [%]	b^2 [%]	m [%]	n^3	q_g^4
α-lactalbumin	Bovine trypsin-1	47	40-100	100	1	$5.64 \cdot DG_{av}$
	Porcine trypsin-1	78	97-123	100	1	$9.36 \cdot DG_{av}$
	Human trypsin-2	53	98-123	100	1	$6.36 \cdot DG_{av}$
	Bacillus licheniformis protease	71	74-100	60	0	$8.52 \cdot DG_{av}$
	Bovine α-chymotrypsin	59	100	57	1	$7.08 \cdot DG_{av}$
	Subtilisin A	30	-	100	0	$3.60 \cdot DG_{av}$
β-casein	Bovine trypsin-1	55	-	100	1	$6.05 \cdot DG_{av}$
	Porcine trypsin-1	77	-	100	1	$8.47 \cdot DG_{av}$
	Human trypsin-2	50	-	100	1	$5.50 \cdot DG_{av}$
β-lactoglobulin	Bovine trypsin-1	48	-	100	1	$7.20 \cdot DG_{av}$

Not determined values were represented by "-";

DG_{av}: average degree of glycation;

[1] a for *Bacillus licheniformis* protease, bovine α-chymotrypsin and subtilisin A were estimated from $DH_{max,exp}/DH_{max,theo}$.

[2] One value of b (obtained in 0.05 % hydrolysis) for porcine and human trypsins was higher than 100 %, but the $DH_{max,exp}$ was similar to the $DH_{max,exp}$ with a b of 100 %.

[3] n for human trypsin was predicted based on the n of bovine and porcine trypsins.

[4] q_g for bovine and porcine trypsins were calculated based on a determined in **Chapters 3** and **4** and $b=100$ %. The estimated q_g in **Chapter 2** (based on $DH_{max,exp}/DH_{max,theo}$) for bovine and porcine tryptic hydrolysis of α-lactalbumin were $4.26 \cdot DG_{av}$ and $12.00 \cdot DG_{av}$, respectively.

Application of the model for 5 different enzymes

In this thesis, the effects of glycation (parameters *m*, *n* and *q*) and substrate concentration (parameter *b*) on hydrolysis were studied separately, however, they were both included in the prediction model (equation 1). This raises the question whether these two effects are independent of each other. This was checked by hydrolysing apo α-lactalbumin (α-LA) by bovine trypsin at different substrate concentrations and enzyme to substrate ratios (E:S). By increasing the substrate concentration from 0.1 % to 1 %, the P^{intact} of the hydrolysis increased from 0.10 to 0.22 (**Figure 8.1A**). The different conditions changed the P^{intact} and thereby also the $DH_{max,exp}$ for non-glycated proteins (**Figure 8.1B**). This did not influence the effect of glycation, since for all 3 conditions, the $DH_{max,exp}$ of the glycated variants were exactly the $DH_{max,exp}$ expected from the $DH_{max,exp}$ of non-glycated protein and the linear decrease due to the increase in the average degree of glycation (DG_{av}). Therefore, it was concluded that the effects of glycation and substrate concentration were independent of each other, which means that the joint effect can be estimated by the prediction model.
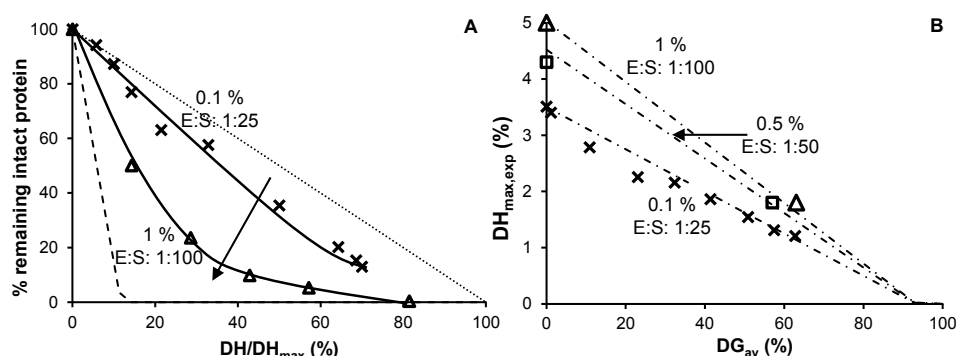
Figure 8.1: (A) Percentage of intact protein in hydrolysates during hydrolysis of (△) 1 % apo α-LA at an E:S of 1:100 and (×) 0.1 % at an E:S of 1:25 as a function of DH/DH_{max}. The dotted line (······) and dashed line (------) indicate the theoretical one-by-one and zipper mechanisms, respectively. (B) DG_{av} *versus* $DH_{max,exp}$ for the hydrolysis of (△) 1 % apo α-LA at an E:S of 1:100, (□) 0.5 % apo α-LA at an E:S of 1:50 and (×) 0.1 % apo α-LA at an E:S of 1:25. (—·) refers to the theoretical scenario that the glycation takes place equally at all CSs. The hydrolysis conditions were similar as described previously [2].

Using equation 1 and the determined values of parameters shown in **Table 8.1**, an estimation of $DH_{max,pre}$ was made as a function of the substrate concentration and average degree of glycation (DG_{av}) for hydrolysis by different enzymes, using apo α-LA as the substrate protein (**Figure 8.2**). The surface area of each graph represents the range of $DH_{max,pre}$ of hydrolysis by each enzyme under the joint effect of the substrate concentration and glycation.

This three-dimensional graph (**Figure 8.2**) could also be made for hydrolysis of other given proteins by these 5 enzymes without additional experiments. This can be done by 1) using the protein AA sequence and the enzyme specificity to determine the parameters $DH_{max,theo}$, *a* and

m, 2) taking the obtained values of parameters *b* and *n* (**Table 8.1**) and 3) calculating parameter q_g based on the values of *a* and *b*. To make this graph for hydrolysis by other enzymes, the values of *a*, *b* and *n* needs to be determined. If parameters *a* and *b* cannot be predicted (as will be discussed later), the hydrolysis experiments by the enzyme with different substrate concentrations need to be performed. To determine parameter *n*, the hydrolysis of a glycated sample (with one DG_{av} value) should be performed. All the other data points in this graph can be predicted based on equation 1. In addition, the required experiments could be performed with a small substrate protein as α-lactalbumin and the data can then be used to calculate $DH_{max,exp}$ for other substrates. Although a few experiment is required, the time needed is significantly less than determining all data points in this group experimentally.
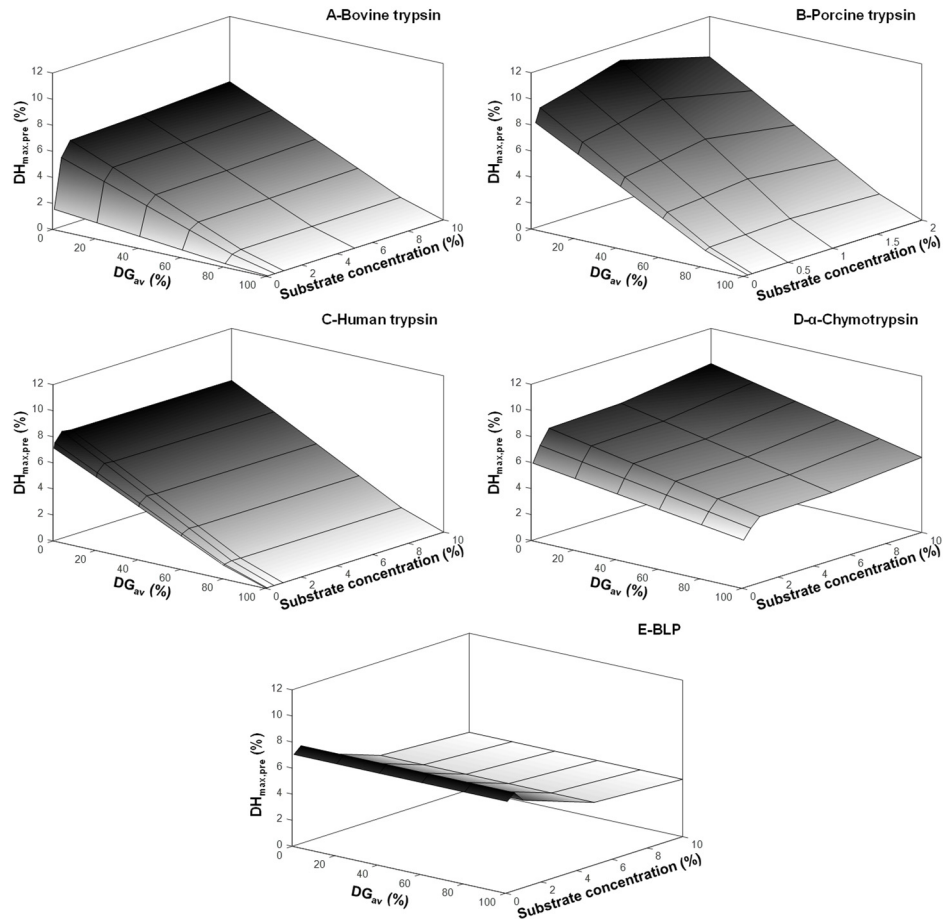
Figure 8.2: Influence of the average degree of glycation (DG_{av}) and the substrate concentration on the predicted maximum degree of hydrolysis ($DH_{max,pre}$) of apo α-LA hydrolysis by (A) bovine, (B) porcine or (C) human trypsins, (D) bovine α-chymotrypsin or (E) BLP, based on equation 1 and **Table 8.1**.

Towards future predicting the parameters in the model

To make this three-dimensional graph for hydrolysis by other enzyme-substrate combinations, the values of parameters *a*, the consequence of enzyme secondary specificity, and *b*, the indication of the amount of inhibitory peptides, need to be determined. This raises the question whether it is possible in the future to predict the values of *a* and *b*, based on the molecular properties of the substrate protein and the protease.

Towards predicting the enzyme secondary specificity

To obtain the values of parameter *a* for other enzyme-substrate combinations, one way is to determine the enzyme selectivity towards CSs in the substrate protein, as previously described in **Chapters 3** and **4**. As shown in **Chapter 3**, the rules of bovine trypsin selectivity obtained based on hydrolysis of CSs in apo α-LA and β-casein (β-cas) could be used to predict the $DH_{max,exp}$ of β-lactoglobulin (β-LG). This means that to determine the enzyme secondary specificity, it does not matter which substrate protein it is used. This suggested that the secondary specificity of an enzyme could be obtained by determining the enzyme selectivity towards CSs in 3-4 proteins with a relatively short (≤20 kDa) AA sequence and can be easily obtained with a high purity. These proteins can be ionised and quantified using the methods described in **Chapters 3** and **4**. Once the enzyme selectivity is determined, the CSs need to be grouped into several categories, in order to analyse the binding site sequences. In this thesis, three categories of CSs were used, i.e. high, intermediate and low selectivity sites. After grouping the categories, the interaction of the AAs on the binding site positions and the enzyme selectivity towards CSs needs to be correlated. Besides the analysis performed in **Chapters 3** and **4**, this can also be achieved by calculating the cleavage entropy of each subsite pocket on the enzyme [3]. Moreover, correlating the activation energy and the substrate binding energy of each subsite pocket is another possible method [4]. It can target the binding site position to which the enzyme is more sensitive. For instance, the total activation energy of 4 insect trypsins were calculated and were correlated with the substrate binding energy of each subsite pocket [5]. Some subsite pockets were found to be involved more in catalysis while others were involved in substrate binding [5]. This means that by analysing the AAs on the binding site positions that the corresponding subsite pocket is involved in binding, the enzyme secondary specificity can be found more efficiently. After identifying the secondary specificity of the protease, it can be applied to predict the $DH_{max,exp}$ of the proteins of interest.

If determining the enzyme selectivity is not an option, some studies used existing databases, e.g. PROSPER database [6], to analyse the enzyme secondary specificity based on the occurrence of cleavages [7, 8]. However, using this type of approach, only the binary rules, i.e. missed cleavage rules, can be identified, because the hydrolysis rate constants of CSs are typically not provided in the databases. Alternatively, it may be possible -in the future- to derive the enzyme secondary specificity *in silico* from the three-dimensional structure of the protease, because the secondary

specificity should come from the subsite of the enzyme. Taking the three trypsins used in **Chapter 4**, i.e. bovine trypsin-1, porcine trypsin-1 and human trypsin-2, as examples, their catalytic sites consist of the same AAs. However, by comparing their AA sequences using the ClustalX software v2.1, only 25 % and 79 % of homology were obtained between bovine trypsin-1 and human trypsin-2, and between porcine trypsin-1 and human trypsin-2, respectively. To compare the three-dimensional structure of these trypsins, the crystal structure of human trypsin-2 needed to be simulated firstly since it has not been reported. The crystal structure of human trypsin-2 was generated using the MODELLER software v9.19 (Sali Lab, San Francisco, USA) by taking the crystal structure of human trypsin-1 (PDB 2RA3 [9]) as a template. By comparing the three-dimensional structure of bovine trypsin-1 (PDB 4I8L [10]), porcine trypsin-1 (PDB 1S81 [11]) and the generated human trypsin-2, it was found that the charged AAs (KRDE) at pH 8 (the pH of the hydrolysis) surrounding the catalytic sites of the three trypsins were indeed different (**Figure 8.3**). If they are on the subsite positions, this can explain why these three trypsins had different sensitivity towards charged AAs on the binding site positions. In literature, the binding between the subsite of the enzyme and the binding site of the CS on the substrate protein has been studied using molecular dynamics simulations [12-14]. This is also referred to as the molecular docking studies, which is an approach commonly used in field of drug design.
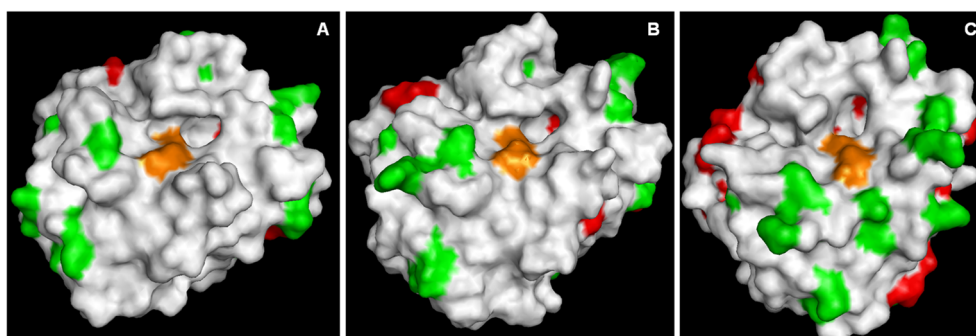
Figure 8.3: Crystal structures of (A) bovine trypsin-1 (PDB 4I8L [10]), (B) porcine trypsin-1 (PDB 1S81 [11]) and (C) human trypsin-2*. The AAs on the catalytic sites were marked in orange. The positively charged AAs (KR) were marked in green and the negatively charged AAs (DE) were marked in red.

The Molecular Operating Environment (MOE) software (Chemical Computing Ltd. Montreal, Canada) has been commonly used to find the most favourable binding interaction based on the lowest simulated docking energy. For instance, the individual interaction between two peptides of hemagglutinin ('RERRRKKRG' and 'SIQSRG'), and trypsin (source not mentioned) as well as furin were simulated using MOE [13]. In that study, 100 conformations were generated for each enzyme-peptide combination, and the conformation with the lowest docking energy was used for further energy minimizing calculation to create the final conformation [13]. Besides MOE, there are many other protein-ligand docking softwares, e.g. PRODOCK [15], VoteDock [16], etc. Although the final binding conformation is a result of personal judgment of the authors [12], it can be the first step to predict the enzyme secondary specificity. To check the correctness of the prediction,

* The crystal structure of human trypsin-2 was generated using the crystal structure of human trypsin-1 as a template, with help from Adrie Westphal.

the enzyme secondary specificity obtained from the molecular docking studies needs to be tested in synthetic peptides or in small proteins.

Towards predicting the formation of inhibitory peptides

Besides the enzyme secondary specificity, parameter *b* also needs to be obtained in order to predict the $DH_{max,exp}$. Basically, this boils down to the question if the formation of inhibitory peptides during hydrolysis is predictable. To study this, the binding between the enzyme and peptides should be identified. To study the protease-protease inhibitor interactions, X-ray is one possible method. The individual interactions between two human trypsins (type-1 and type-3) and the aprotinin (bovine pancreatic trypsin inhibitor) were studied using X-ray [9]. However, during the hydrolysis, perhaps many inhibitory peptides are formed, and it is not sure whether the binding between the inhibitory peptides to proteases is one peptide per one protease molecule. In addition, it is not certain if the binding between enzyme and inhibitory peptide is as strong as the binding between enzyme and aprotinin. This was illustrated by the fact that the molar ratio of inhibitory peptides derived from bovine caseins to fully inhibiting angiotensin-I converting enzyme varied from 26:1 to >1,300:1 [17, 18]. Moreover, in **Chapter 5**, the amount of inhibited bovine trypsin in 0.1 % apo α-LA hydrolysate was estimated to be 0.43-1.72 µM, whereas the protein concentration was ~57 µM. Therefore, X-ray might not be applicable in this case. The surface enhanced laser desorption ionization time of flight mass spectrometry (SELDI-TOF-MS) might be another option to study the non-covalent protein-peptide interaction [19]. In a β-LG hydrolysates produced from hydrolysis by BLP, 5 peptides were found to bind to β-LG on a SELDI ProteinChip [20].

Actually, inhibitory peptides for trypsins might have similar molecular structure as the activity site of the aprotinin [21] or the soybean protein trypsin inhibitor [22]. It could be a good starting point to select candidate peptides with potential inhibitory activities, based on the AA sequence of the protein. The above-mentioned molecular dynamics simulations can be used to select candidate inhibitory peptides. After simulating the bindings between the protease and the candidate inhibitory peptides, the good candidate peptides can be synthesized and tested experimentally.

Testing the boundary of the prediction model

In industrial protein ingredients, protein can be present in native, unfold, refold and aggregated states, depending on the conditions used during processing. Changes in protein structure might affect the accessibility of the CSs for the enzyme, and thereby the rate of hydrolysis as well as the $DH_{max,exp}$. The main conclusions in this thesis were made based on the results of apo α-LA and β-cas (**Table 8.1**). β-cas is a random coil protein with limited secondary and tertiary structure [23]. α-LA can occur in a folded structure (referred to as the holo α-LA) when a calcium is bound, and in a molten globule form (referred to as the apo α-LA) when it is calcium depleted [24]. The apo form still has secondary structure but no fixed tertiary structure [25]. In addition, the

denaturation temperature of apo α-LA was at ~37 °C [26], which was the hydrolysis temperature used in this thesis. This raises the question if the prediction model can be used in the hydrolysis of globular, (heat induced) unfolded and (the Maillard reaction induced) aggregated proteins.

Testing the prediction model in globular proteins

The prediction of $DH_{max,exp}$ of hydrolysis of bovine serum albumin (BSA) by bovine trypsin was obtained based on the secondary specificity proposed in **Chapter 3**. The predicted $DH_{max,pre}$ of β-LG hydrolysis by porcine or human trypsins were obtained based on their secondary specificity proposed in **Chapter 4**. These predicted values were compared with the $DH_{max,exp}$ (**Figure 8.4**). The $DH_{max,pre}$ of BSA hydrolysis by bovine trypsin was 7.7 %, which had a standard error of 18 % with the $DH_{max,exp}$ (9.4 %) (**Figure 8.4A**). The $DH_{max,pre}$ values for β-LG hydrolysis by porcine or human trypsins were 8.8 % (**Figure 8.4B**) and 7.9 % (**Figure 8.4C**), respectively. These values had an average standard error of 9 % with the $DH_{max,exp}$. The prediction based on the secondary specificity was 2-3 times better than the prediction based on the enzyme specificity ($DH_{max,theo}$). This is similar to the results of apo α-LA and β-cas reported in **Chapters 3** and **4**, which confirmed that the model can be used to predict the hydrolysis of non-modified globular proteins.
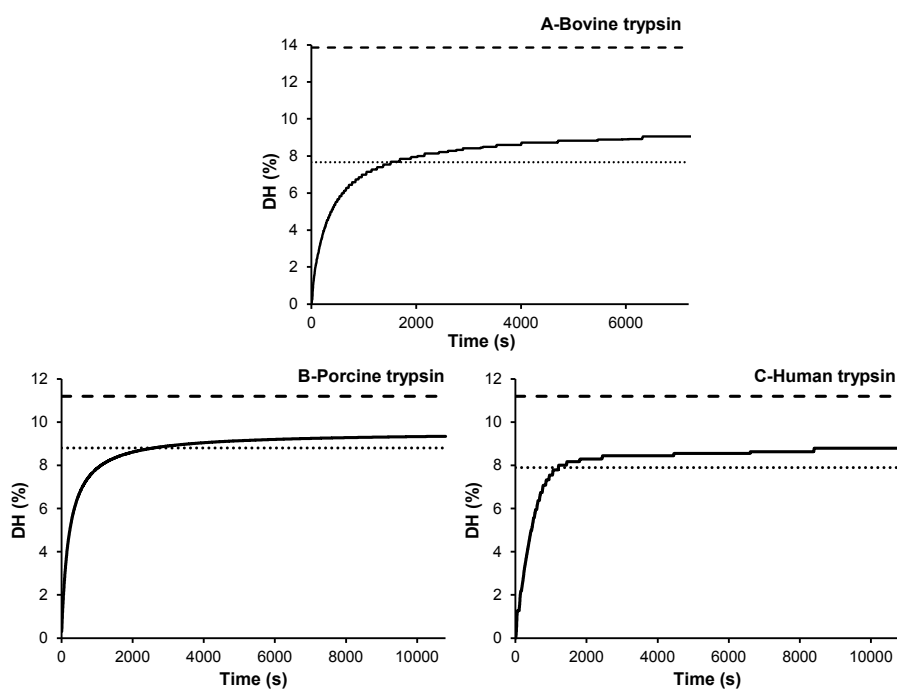
Figure 8.4: Hydrolysis curves (DH *versus* time) of (A) bovine tryptic hydrolysis of 1 % BSA, (B) porcine or (C) human tryptic hydrolysis of 1 % β-LG at an E:S of 1:100. The dashed lines (----) and dotted lines (······) indicate the $DH_{max,theo}$ and the $DH_{max,pre}$ of the hydrolyses, respectively. The hydrolysis conditions were similar as described previously [27].

To check whether the parameters of glycation (m, n and q_g) estimated in this thesis can be used to describe the hydrolysis of globular proteins, the effect of glycation on holo α-LA hydrolysis was tested. The tertiary structure of holo α-LA was much more compact than that of apo α-LA (**Figure 8.5A**). Glycated holo α-LA samples were hydrolysed by bovine trypsin (lysine/arginine specific) or BLP (non-lysine/arginine specific), and the results were compared with the results of apo α-LA reported in **Chapter 2**. Similarly as for apo α-LA, the glycation did not influence the tertiary structure of holo α-LA (**Figure 8.5A**). For hydrolysis by both enzymes, the $DH_{max,exp}$ values of non-glycated holo and apo α-LA were similar, and the relations between glycation and hydrolysis were the same as those of apo α-LA (**Figure 8.5B**). In addition, the hydrolysis rate constants of holo α-LA by both enzymes were not affected by glycation (data not shown), and were similar to those of apo α-LA (**Chapter 2**). This means that indeed the values of these parameters can be used to describe the effects of glycation on hydrolysis of globular proteins.
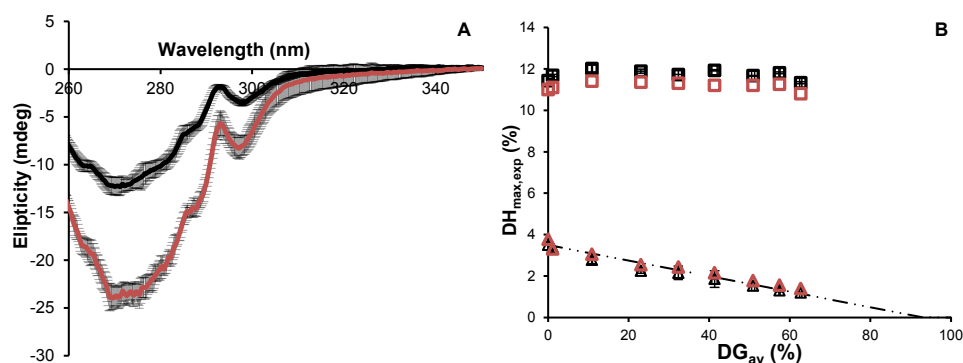
Figure 8.5: (A) Near UV-circular dichroism of (—) apo and (—) holo α-LA glycated with D-glucose at 50 °C with 65 % relative humidity for 0-10 h. The lines are the average of spectra over all samples and the error bars illustrate the standard deviations. (B) DG_{av} *versus* $DH_{max,exp}$ for hydrolysis of non-glycated and glycated (black) apo and (red) holo α-LA by (△) bovine trypsin or (□) BLP. (—·) refers the theoretical scenario that glycation takes place equally at all CSs. The glycation and hydrolysis conditions were similar as described previously [2].

To see whether the effect of substrate concentration on hydrolysis can also be found in globular proteins, the influence of substrate concentration on bovine tryptic hydrolysis of whey protein isolate (WPI) was tested. The effect of substrate concentration on the $DH_{max,exp}$ of WPI hydrolysis was the same as apo α-LA (**Figure 8.6**), indicating the same type of inhibition. This suggested that inhibitory peptides were also formed in hydrolysis of globular proteins.
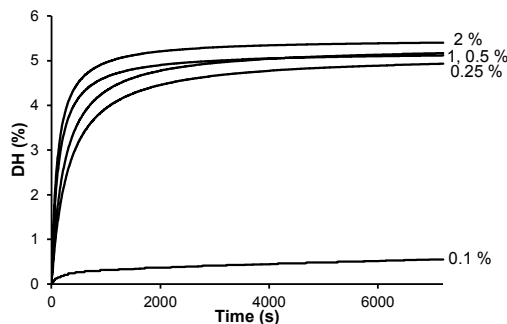
Figure 8.6: Hydrolysis curves (DH *versus* time) of bovine tryptic hydrolysis of whey protein isolate with different substrate concentrations (0.1 %-2 %) at an E:S of 1:100. The hydrolysis conditions were similar as described in **Chapter 5**.

The above-mentioned tests proved that the parameter estimates in **Table 8.1** can be used to predict hydrolysis of globular proteins. However, in the published articles of enzymatic protein hydrolysis, the hydrolysis was shown to be influenced by system conditions other than substrate concentrations, e.g. pH [28], temperature [29], etc. The question is whether the prediction model can be used to identify and explain the changes in these conditions on hydrolysis.

The $DH_{max,exp}$ of 5 % WPI hydrolysis by BLP increased from 4 % to 8 % with an increase in hydrolysis pH from 7 to 9 [30]. Firstly, at different pH values, the numbers of charged histidine residues in the substrate protein and the enzyme were different. At pH 7, 9 % of the histidine residues had a positive charge, while at pH 8 or 9, only 0.9 % or 0.09 % of the histidine residues had a positive charge, respectively. Only 2 histidine residues are present in β-LG (the main protein in WPI), but there is one histidine in the catalytic site of the protease (which contains 3 residues, i.e. serine, histidine and aspartic acid [31]). This suggested that the binding between BLP and all CSs in WPI might be different at different pH values, leading to a change in parameter *a* (% CSs hydrolysed in a realistic timescale). In addition, in the article that studied WPI hydrolysis by BLP at different pH values, it was shown that the P^{intact} (parameter *b*) increased with the pH of the hydrolysis [30]. Based on indications shown in **Chapter 5**, this means that more inhibitory peptides might be formed during hydrolysis at a lower pH value, leading to a lower $DH_{max,exp}$. Therefore, in the case of changes in pH of the BLP hydrolysis, both parameters *a* and *b* could be influenced.

For the effect of temperature, the $DH_{max,exp}$ of porcine tryptic hydrolysis of a pea protein isolate increased from 7 % to 13 % when the hydrolysis temperature increased from 35 °C to 50 °C [32]. It was reported the pKa of the side chain of lysine decreased with an increasing temperature, but the pKa of glutamic acid was independent of the temperature [33]. This means that in the case of porcine tryptic hydrolysis of pea protein isolate, the binding between the enzyme and the CSs could be different, which means that parameter *a* might be influenced. Besides parameter *a*, there was another factor changed if the hydrolysis was performed at a different temperature, which could affect the $DH_{max,exp}$. It was shown that there were a-specific peptides formed in the

bovine tryptic hydrolysis of β-LG at 40-50 °C while these peptides were not formed in the hydrolysis at a temperature lower than 40 °C [29]. The a-specific cleavage was previously found to be the autolytic degradation of certain intrinsically unstable peptides bonds, referred to as the "*spontaneous cleavage*" [34]. It seems that the spontaneous cleavages occurred more in the hydrolysis at a higher temperature, leading to a higher $DH_{max,exp}$. This could be another reason why in the previous article the $DH_{max,exp}$ of the pea protein isolate hydrolysis increased with an increasing temperature [32]. In another article, the $DH_{max,exp}$ of porcine tryptic hydrolysis of 1 % WPC at 50 °C was ~14 %, which was even higher than the $DH_{max,theo}$ (11.2 %) [35]. This is also a strong indication that the spontaneous cleavages occurred.

Testing the prediction model in heat induced unfolding of proteins

Heating induced unfolding of proteins was shown to influence the rate and extent of hydrolysis. For example, the native WPI hydrolysis reached DH 5 % after 110 min of hydrolysis by digestive-enzyme complex Corolase PP, whereas the heated WPI (10 min, at 80 °C) hydrolysis reached DH 5 % at 50 min [36]. This indicated that heating affected the rate of hydrolysis. Besides influencing the rate, the $DH_{max,exp}$ of 0.1 % heated (10 min, at 100 °C) whey protein concentrate (WPC) after a sequential hydrolysis by porcine pepsin and bovine trypsin was 15 %, whereas that of untreated WPC was 5 % [37]. At higher concentration (1 %) this difference was smaller, where the $DH_{max,exp}$ of 1 % WPC hydrolysis were 25 % and 20 % for heated and untreated WPC, respectively [37]. This effect can be understood perhaps with the information from **Chapter 5**. For bovine tryptic hydrolysis with a substrate concentration of 0.1 % the *b* (the normalised P^{intact}) was low (40 %), whereas at 1 % hydrolysis *b* was 85 %. This means if heating would result in an increase of *b*, there would be an effect on bovine tryptic hydrolysis of 0.1 % WPC (since *b* for the non-heated protein is low), but it cannot further largely increase the *b* for 1 % hydrolysis.

Testing the prediction model in aggregated proteins

The prediction model seems to be able to explain results in literature, where the protein tertiary structure or the system conditions were different from this thesis. However, these examples were still in relatively clean systems, where the samples were not heavily modified or the hydrolysis was not with multiple enzymes. This raises the question whether the approach and knowledge obtained in this thesis can be extrapolated to hydrolysis samples that are heavily modified, and are hydrolysed by multiple enzymes.

Effect of Maillard induced aggregation on a single-enzyme hydrolysis

For trypsins and α-chymotrypsin, the effects of glycation on hydrolysis were shown in **Chapter 2**. Additional experiments with pepsin hydrolysis on these glycated samples showed that pepsin hydrolysis was independent of the glycation (data not shown). These samples were incubated at low temperatures, to avoid any secondary reactions. During incubation done at 80 °C for 0-8 h, the average degrees of glycation (DG_{av}) of the samples were quite constant from 1-8 h (~85 %, measure using the OPA method), the amount of aggregated proteins increased (**Figure 8.7**). Such

aggregation, observed during the Maillard reaction, was also previously reported [38]. As a result of the secondary reactions, protein cross-links, such as lysine-arginine cross-links [39], could be formed. The heating itself might also result in protein aggregation through lysinoalanyl cross-links [40] or non-covalent aggregation.
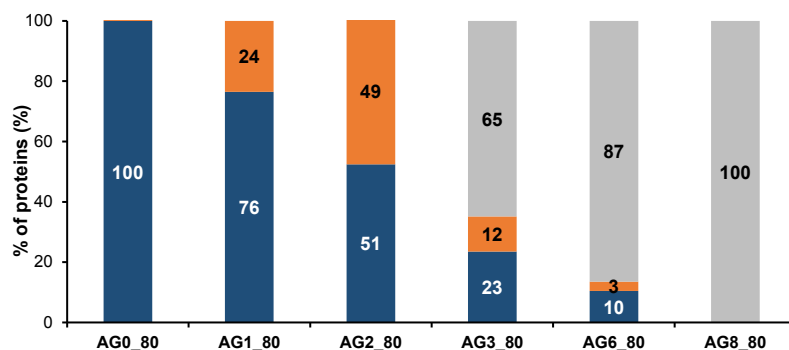

Figure 8.7: Percentage of (■) monomeric, (■) soluble aggregated and (■) insoluble aggregated proteins for apo α-LA heated with D-glucose at 80 °C with 65 % relative humidity for 0-8 h. The amount of monomeric proteins and soluble aggregates in the samples were measured using size exclusion chromatography. The amount of insoluble aggregates were determined by subtracting the total area under the peak at UV_{280} of each sample from that of the non-glycated protein. The glycation conditions were similar as described previously [2].

These samples were then hydrolysed separately by porcine pepsin, or by porcine pancreatin, which is a mixture of trypsin, α-chymotrypsin, carboxylpeptidases A and B. The $DH_{max,exp}$ of porcine pepsin hydrolysis decreased from 16 % to 0 % with an increasing amount of aggregated proteins (calculated as the sum of soluble and insoluble aggregates) in the samples (**Figure 8.8A**). This means that pepsin hydrolysis was hindered by the aggregation, which might be caused by a decrease in parameter *b* (normalised p^{intact}) due to the low accessibility of CSs induced by the protein aggregation. To separate the effects of soluble and insoluble aggregates, the aggregates should be -in the future- isolated and hydrolysed individually. This was previously done in an article that reported the hydrolysis of rapeseed meal by a mixture of trypsin, chymotrypsin and porcine intestinal peptidase [41]. In that article, the $DH_{max,exp}$ of hydrolysis of the insoluble fraction of rapeseed meal reached the same extent as that of the unseparated meal (18 %), whereas the $DH_{max,exp}$ of hydrolysis of the soluble fraction was 14 % [41].

For porcine pancreatic hydrolysis, based on equation 1 and parameters described in **Table 8.1**, the sum of the predicted $DH_{max,pre}$ of non-glycated apo α-LA hydrolysis by porcine trypsin and bovine α-chymotrypsin was 22 %. The $DH_{max,exp}$ of hydrolysis by porcine pancreatin was 29 % (**Figure 8.8B**), which means the carboxylpeptidases contributed to 7 % of the total $DH_{max,exp}$. It should be noted that in porcine pancreatin, the α-chymotrypsin was from pig, whereas the α-chymotrypsin tested in the prediction model was from cow. Since all samples had a DG_{av} of 85 %, based on equation 1, the sum of the predicted decrease in $DH_{max,pre}$ of hydrolysis by porcine trypsin and bovine α-chymotrypsin was ~11 %. This means the $DH_{max,pre}$ of these samples based

on their DG_{av} values should be ~18 %. Experimentally, averaged over all samples, the $DH_{max,exp}$ of hydrolysis by porcine pancreatin was 18.6±1.4 % (**Figure 8.8B**), which was similar to the $DH_{max,pre}$. This means that the porcine pancreatic hydrolysis was not influenced by protein aggregation. It was concluded that the influence of glycation or aggregation on the extent of hydrolysis were not compensated by the presence of other proteases.
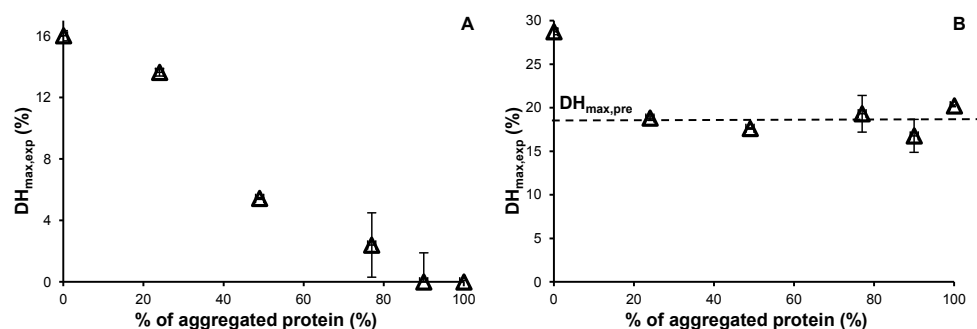

Figure 8.8: Percentage of aggregated proteins *versus* $DH_{max,exp}$ for hydrolysis of 0.1 % apo α-LA heated with D-glucose at 80 °C with 65 % relative humidity for 0-8 h by (A) porcine pepsin at pH 3 and by (B) porcine pancreatin at pH 8 at an E:S of 1:100 at 37 °C. The aggregation analysis was explained in the caption of **Figure 8.7**. The hydrolysis conditions were similar as described previously [27]. The dotted line (---) refers to the $DH_{max,pre}$ using equation 1 and **Table 8.1**.

Based on the data of pepsin hydrolysis and previously mentioned results in literature, parameter *b* seems to be strongly influenced by different heat treatments of the proteins. It seems that during heating induced unfolding of proteins, *b* was higher in the heated proteins than in the untreated proteins, while during the Maillard reaction and/or heating induced aggregation, parameter *b* decreased with heating time. This suggested that perhaps for heating induced unfolding and aggregation of proteins, *b* would reach the optimum when the proteins are unfolded but not yet aggregated (**Figure 8.9**). However, for the Maillard reaction induced aggregation, the protein was not perhaps unfolded, then *b* would only decrease with the heating time/temperature (**Figure 8.9**).
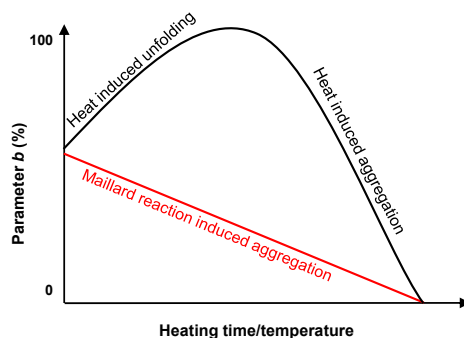
Figure 8.9: Hypothetical influences of heating time and temperature of (—) the Maillard reaction induced aggregation and (—) heat induced unfolding and aggregation on the value of parameter *b* in equation 1.

Effect of glycation and aggregation on sequential protein hydrolysis

The *in vivo* protein digestion is a sequential hydrolysis with multiple enzymes, e.g. pepsin, trypsin and α-chymotrypsin. In the existing *in vitro* digestion models, a two-step sequential hydrolysis, i.e. gastric and duodenal phase, was commonly used [42]. However, the protein digestion by the brush border enzymes, i.e. (exo-)aminopeptidases, that are responsible for the final digestion of proteins and the uptake of free AAs [43, 44], was not taken into account in these models. Thus, to better simulate the *in vivo* protein digestion, a three-step sequential hydrolysis model was developed. The three steps were:

1) gastric phase: samples were incubated with porcine pepsin (pH 3) for 1 h at 37 °C;

2) intestinal phase (lumen) phase: the pH of the pepsin hydrolysates was adjusted to 8, and the hydrolysates were incubated with porcine pancreatin for 2 h at 37 °C;

3) intestinal phase (brush border) phase: the pancreatic hydrolysates were incubated with the Flavourzyme for 1 h at pH 8 at 37 °C. The Flavourzyme contains 2 aminopeptidases [45], which were used as the source of the aminopeptidases of the brush border digestion.

The hydrolysis of apo α-LA, a glycated but not aggregated sample (AG8_50) and two glycated and aggregated samples (AG3_80 and AG6_80) were tested in the sequential hydrolysis experiment. For the sample AG8_50, the increase in $DH_{max,exp}$ during pepsin and Flavourzyme hydrolysis was minor, but for pancreatic hydrolysis, the $DH_{max,exp}$ was lower than that of apo α-LA (**Figure 8.10B**). The decrease in final $DH_{max,exp}$ after the three-step sequential hydrolysis was mostly due to the decreased $DH_{max,exp}$ in pancreatic hydrolysis (**Figure 8.10A**). The additional aggregation in samples AG3_80 and AG6_80 resulted in hindering the pepsin hydrolysis (**Figure 8.10B**), but had no influences on the pancreatin and Flavourzyme hydrolysis. For these two samples, the changes in the final $DH_{max,exp}$ compared to sample AG8_50 were therefore mostly due to the changes in pepsin hydrolysis (**Figure 8.10A**). These results showed that once any type of modifications hindered one of the digestive enzymes in the sequential hydrolysis, other enzymes could not compensate the decrease in $DH_{max,exp}$. This means that even through the absolute $DH_{max,exp}$ values of $DH_{max,exp}$ of hydrolysis by a single enzyme or by a sequential hydrolysis were different, the relative effects of glycation and aggregation were still the same. Therefore, the approaches used and the results obtained in this thesis can be extrapolated to explain the $DH_{max,exp}$ reached by a hydrolysis that the samples and the hydrolysis process are complex, i.e. heavily modified samples in a multi-step hydrolysis.
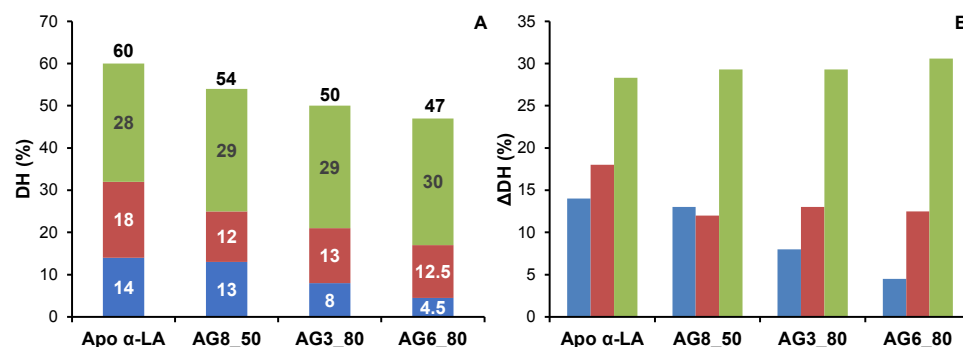
Figure 8.10: (A) Final DH and (B) ΔDH of apo α-LA, AG8_50 (heated with D-glucose at 50 °C for 8 h, glycated but not aggregated [2]), AG3_80 and AG6_80 (glycation conditions were explained in the caption of **Figure 8.7**), during hydrolysis by the three-step sequential hydrolysis. The gastric, intestinal (lumen) and intestinal (brush border) phases were marked with ■, ■ and ■, respectively. The hydrolysis conditions were similar as described previously [2].

Considering the presence of exo-peptidases, it was expected that a higher final $DH_{max,exp}$ was reached. It was then considered that the calculation of $DH_{max,exp}$ based on pH-stat titration might underestimate the $DH_{max,exp}$ of exopeptidase [46]. This is due to the fact that the pKa of the α-amino group increased by ~0.5 pH unit from polypeptide (pKa=7.5 at 37 °C) to di or tri-peptide [47]. If this was taken into account in the $DH_{max,exp}$ calculation, the final $DH_{max,exp}$ values of the sequential hydrolysis of these samples were 90-70 %. This was also confirmed by the results of the reserved phase ultra-high performance liquid chromatography (RP-UPLC). Only two peptides with masses ≤350 Da were found in the chromatograms (data not shown), indicating that the RP column could not be used to perform the peptide analysis for this type of samples. Since the final hydrolysates of the sequential hydrolysis had DH values of 90-70 %, they contained mainly dipeptides and free AAs. The dipeptides and single AAs in the hydrolysates were not found in chromatograms of RP-UPLC, due to their low hydrophobicity, resulting in a very early elution on the C18 column (the RP column used in **Chapters 3-6**) [48]. A hydrophilic interaction chromatography (HILIC) column could be used to perform the analysis [49]. To test this, a known sample (previously analysed by RP-UPLC), a porcine tryptic hydrolysate of β-cas with a DH of 6.7 % was analysed using an Acquity UPLC BEH Amide column (2.1 × 150 mm, 1.7 μm particle size), coupled to a LTQ-VelosPro mass spectrometry (MS) system (Thermo Scientific). Using the HILIC column, the elution of free lysine in the hydrolysate was found, and the lysine concentration was quantified using a calibration curve made using pure L-lysine (Sigma) (**Figure 8.11B**). The concentration of the free lysine in the hydrolysate was ~12.7 μM, which was 4.7 μM higher than the expected concentration of K29 based on the molar sequence coverage analysed using the C18 column (~8 μM) (**Figure 8.11A**). The elution of two dipeptides, β-cas[98-99] and β-cas[106-107], and a tripeptide, β-cas[26-28], on the HILIC column were found by extracting their m/z values from the full MS (**Figure 8.11C**). Based on the molar sequence coverage of this sample, these two dipeptides should be present in the sample but they were not found at all (**Figure 8.11A**). Tripeptide β-cas[26-28] eluted the

earliest using the C18 column (retention time: 2.2 minutes). The retention time of the two dipeptides using the HILIC column was similar to the tripeptide β-cas[26-28], meaning that their elution time using the C18 column could be too early for detection. These results showed that using HILIC column, the elution of single AAs and dipeptides can be found, which can be used to perform peptide analysis for the hydrolysates after the sequential hydrolysis, where the C18 column could not be used. It can also help to compensate the peptide analysis of hydrolysates by the specific enzymes performed typically using the C18 column. Further method optimization can be performed to better quantify the compounds.
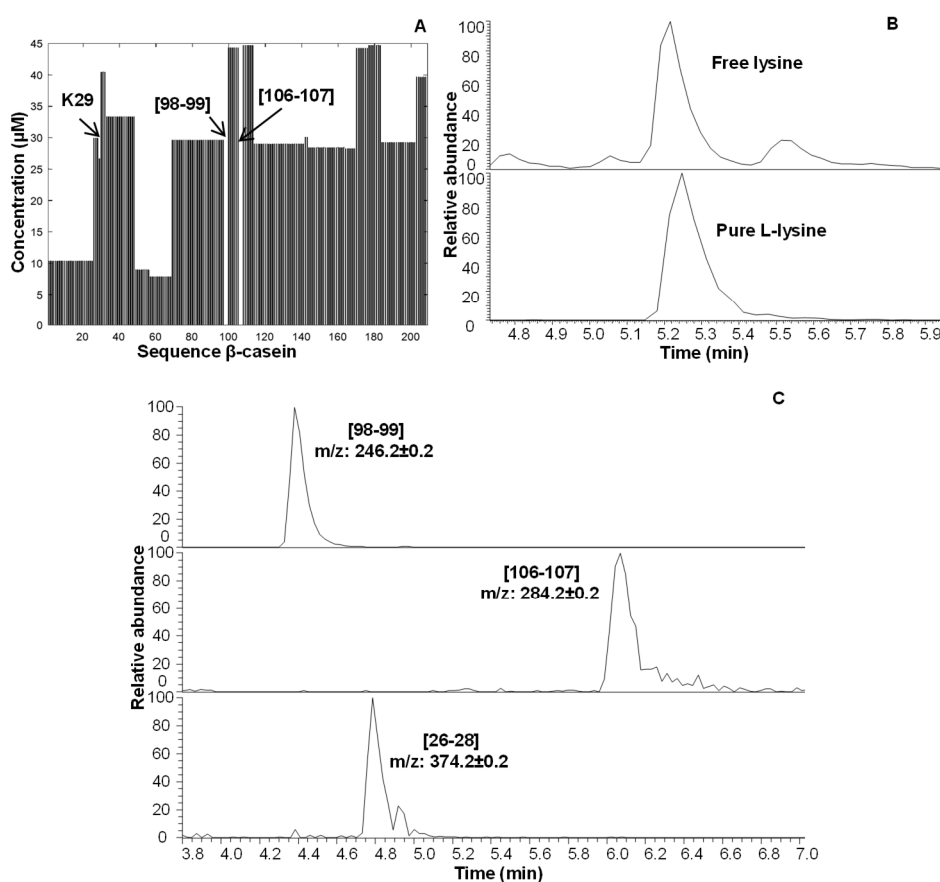
Figure 8.11: (A) Molar sequence coverage of a porcine tryptic hydrolysate of β-cas with a DH of 6.7 %, (B) selected reaction monitoring (SRM) elution profiles of free lysine in the hydrolysate and pure L-lysine, and (C) total ion count of peptides β-cas[98-99], β-cas[106-107] and β-cas[26-28] extracted from the full MS using HILIC column. The normalised collision energy was set at 35 for $^{13}C_6,^{15}N_2$-lysine (internal standard) and 30 for lysine in positive mode. The parent m/z for lysine was set at 147 and the SRM of the fragment was set at 130 (m/z tolerance: ±1). The quantification was performed using an external standard calibration curve by plotting MS peak area of the SRM fragment divided by the MS peak area of the SRM fragment of $^{13}C_6,^{15}N_2$-lysine. This method was adapted from van Rooijen, *et.al* [49].

Concluding remarks

The model proposed can be used to predict the extent of hydrolysis of any given protein by any specific protease. In addition, it was shown to predict the extent of hydrolysis of proteins that are modified on the primary structure. As shown by additional tests and comparison with data from literature, the model can be used to predict the extent of hydrolysis for complex systems, e.g. where the substrate proteins are aggregated and the hydrolysis is performed with multiple enzymes. For enzymes that were not studied in this thesis, the model can be used to set up the guidelines for the minimum amount of experiments needed for the prediction of the extent of hydrolysis. Of course there are parts of the model, e.g. the methods to determine the value of each parameter, that need to be further validated and improved. But the concept of this model is shown to be an effective tool that can be used to define expectations of the extent of enzymatic protein hydrolysis. This can -in the future- help to control and predict the protein hydrolysis and *in vitro* protein digestion.

References

[1] N.D. Rawlings, M. Waller, A.J. Barrett, A. Bateman, MEROPS: The database of proteolytic enzymes, their substrates and inhibitors, Nucleic Acids Research, 42 (2014) D503-D509.

[2] Y. Deng, P.A. Wierenga, H.A. Schols, S. Sforza, H. Gruppen, Effect of Maillard induced glycation on protein hydrolysis by lysine/arginine and non-lysine/arginine specific proteases, Food Hydrocolloids, 69 (2017) 210-219.

[3] J.E. Fuchs, S. von Grafenstein, R.G. Huber, M.A. Margreiter, G.M. Spitzer, H.G. Wallnoefer, K.R. Liedl, Cleavage entropy as quantitative measure of protease specificity, PLoS Computational Biology, 9 (2013) e1003007.

[4] S.R. Marana, A.R. Lopes, L. Juliano, M.A. Juliano, C. Ferreira, W.R. Terra, Subsites of trypsin active site favor catalysis or substrate binding, Biochemical and Biophysical Research Communications, 290 (2002) 494-497.

[5] A.R. Lopes, M.A. Juliano, S.R. Marana, L. Juliano, W.R. Terra, Substrate specificity of insect trypsins and the role of their subsites in catalysis, Insect Biochemistry and Molecular Biology, 36 (2006) 130-140.

[6] J. Song, H. Tan, A.J. Perry, T. Akutsu, G.I. Webb, J.C. Whisstock, R.N. Pike, PROSPER: An integrated feature-based tool for predicting protease substrate cleavage sites, PLoS ONE, 7 (2012) e50300.

[7] C.Y. Yen, S. Russell, A.M. Mendoza, K. Meyer-Arendt, S. Sun, K.J. Cios, N.G. Ahn, K.A. Resing, Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: Protease cleavage and SCX elution rules from data mining of MS/MS spectra, Analytical Chemistry, 78 (2006) 1071-1084.

[8] J.A. Siepen, E.-J. Keevil, D. Knight, S.J. Hubbard, Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics, Journal of Proteome Research, 6 (2007) 399-408.

[9] M.d.A. Salameh, A.S. Soares, A. Hockla, E.S. Radisky, Structural basis for accelerated cleavage of bovine pancreatic trypsin inhibitor (BPTI) by human mesotrypsin, Journal of Biological Chemistry, 283 (2008) 4115-4123.

[10] D. Liebschner, M. Dauter, A. Brzuszkiewicz, Z. Dauter, On the reproducibility of protein crystal structures: Five atomic resolution structures of trypsin, Acta Crystallographica Section D: Biological Crystallography, 69 (2013) 1447-1462.

[11] T.R. Transue, J.M. Krahn, S.A. Gabel, E.F. DeRose, R.E. London, X-ray and NMR characterization of covalent complexes of trypsin, borate, and alcohols, Biochemistry, 43 (2004) 2829-2839.

[12] B.J. Waldner, J.E. Fuchs, R.G. Huber, S. von Grafenstein, M. Schauperl, C. Kramer, K.R. Liedl, Quantitative correlation of conformational binding enthalpy with substrate specificity of serine proteases, The Journal of Physical Chemistry B, 120 (2016) 299-308.

[13] X.L. Guo, L. Li, D.Q. Wei, Y.S. Zhu, K.C. Chou, Cleavage mechanism of the H5N1 hemagglutinin by trypsin and furin, Amino Acids, 35 (2008) 375-382.

[14] M.P. Allen, Introduction to molecular dynamics simulation, Computational soft matter: from synthetic polymers to proteins, 23 (2004) 1-28.

[15] J.Y. Trosset, H.A. Scheraga, PRODOCK: Software package for protein modeling and docking, Journal of Computational Chemistry, 20 (1999) 412-427.

[16] D. Plewczynski, M. Łażniewski, M.V. Grotthuss, L. Rychlewski, K. Ginalski, VoteDock: Consensus docking method for prediction of protein–ligand interactions, Journal of Computational Chemistry, 32 (2011) 568-581.

[17] S. Maruyama, H. Mitachi, H. Tanaka, N. Tomizuka, H. Suzuki, Studies on the active site and antihypertensive activity of angiotensin I-converting enzyme inhibitors derived from casein, Agricultural and Biological Chemistry, 51 (1987) 1581-1586.

[18] S. Maruyama, H. Mitachi, J. Awaya, M. Kurono, N. Tomizuka, H. Suzuki, Angiotensin I-converting enzyme inhibitory activity of the C-terminal hexapeptide of α_{s1}-casein, Agricultural and Biological Chemistry, 51 (1987) 2557-2561.

[19] H.A. Kosters, P.A. Wierenga, H. Gruppen, SELDI-TOF-MS as a rapid tool to study food related protein–peptide interactions, Food Hydrocolloids, 24 (2010) 667-673.

[20] H.A. Kosters, P.A. Wierenga, R. de Vries, H. Gruppen, Characteristics and effects of specific peptides on heat-induced aggregation of β-lactoglobulin, Biomacromolecules, 12 (2011) 2159-2170.

[21] D.M. Blow, C.S. Wright, D. Kukla, A. Rühlmann, W. Steigemann, R. Huber, A model for the association of bovine pancreatic trypsin inhibitor with chymotrypsin and trypsin, Journal of Molecular Biology, 69 (1972) 137-144.

[22] D.M. Blow, J. Janin, R.M. Sweet, Mode of action of soybean trypsin inhibitor (Kunitz) as a model for specific protein-protein interactions, 249 (1974) 54-57.

[23] L.K. Creamer, T. Richardson, D.A.D. Parry, Secondary structure of bovine α_{s1}- and β-casein in solution, Archives of Biochemistry and Biophysics, 211 (1981) 689-696.

[24] K. Kuwajima, The molten globule state of α-lactalbumin, FASEB Journal, 10 (1996) 102-109.

[25] D.A. Dolgikh, R.I. Gilmanshin, E.V. Brazhnikov, V.E. Bychkova, G.V. Semisotnov, S.Y. Venyaminov, O.B. Ptitsyn, α-Lactalbumin: Compact state with fluctuating tertiary structure?, FEBS Letters, 136 (1981) 311-315.

[26] E.A. Permyakov, D.I. Kreimer, Effects of pH, temperature and Ca^{2+} content on the conformation of α-lactalbumin in a medium modelling physiological conditions, General physiology and biophysics, 5 (1986) 377-389.

[27] Y. Deng, F. van der Veer, S. Sforza, H. Gruppen, P.A. Wierenga, Towards predicting protein hydrolysis by bovine trypsin, Process Biochemistry, 65 (2018) 81-92.

[28] S.C. Cheison, E. Leeb, J. Toro-Sierra, U. Kulozik, Influence of hydrolysis temperature and pH on the selective hydrolysis of whey proteins by trypsin and potential recovery of native alpha-lactalbumin, International Dairy Journal, 21 (2011) 166-171.

[29] S.C. Cheison, M. Schmitt, E. Leeb, T. Letzel, U. Kulozik, Influence of temperature and degree of hydrolysis on the peptide composition of trypsin hydrolysates of β-lactoglobulin: Analysis by LC–ESI-TOF/MS, Food Chemistry, 121 (2010) 457-467.

[30] C.I. Butré, S. Sforza, P.A. Wierenga, H. Gruppen, Determination of the influence of the pH of hydrolysis on enzyme selectivity of *Bacillus licheniformis* protease towards whey protein isolate, International Dairy Journal, 44 (2015) 44-53.

[31] E.I. Mil'gotina, T.L. Voyushina, G.G. Chestukhina, Glutamyl endopeptidases: structure, function, and practical application, Russian Journal of Bioorganic Chemistry, 29 (2003) 511-522.

[32] M. Karamac, R. Amarowicz, H. Kostyra, Effect of temperature and enzyme/substrate ratio on the hydrolysis of pea protein isolates by trypsin, Czech journal of food sciences, 20 (2002) 1-6.

[33] H. Nagai, K. Kuwabara, G. Carta, Temperature dependence of the dissociation constants of several amino acids, Journal of Chemical & Engineering Data, 53 (2008) 619-627.

[34] C.I. Butré, S. Buhler, S. Sforza, H. Gruppen, P.A. Wierenga, Spontaneous, non-enzymatic breakdown of peptides during enzymatic protein hydrolysis, Biochimica et Biophysica Acta - Proteins and Proteomics, 1854 (2015) 987-994.

[35] S.M. Kamau, R.-R. Lu, The effect of enzymes and hydrolysis conditions on degree of hydrolysis and DPPH radical scavenging activity of whey protein hydrolysates, Current Research in Dairy Sciences, 3 (2011) 25-35.

[36] I.B. O'Loughlin, B.A. Murray, P.M. Kelly, R.J. Fitzgerald, A. Brodkorb, Enzymatic hydrolysis of heat-induced aggregates of whey protein isolate, Journal of Agricultural and Food Chemistry, 60 (2012) 4895-4904.

[37] S.B. Kim, K.S. Ki, M.A. Khan, W.S. Lee, H.J. Lee, B.S. Ahn, H.S. Kim, Peptic and tryptic hydrolysis of native and heated whey protein to reduce its antigenicity, Journal of Dairy Science, 90 (2007) 4043-4050.

[38] M. Corzo-Martínez, A.C. Soria, J. Belloque, M. Villamiel, F.J. Moreno, Effect of glycation on the gastrointestinal digestibility and immunoreactivity of bovine β-lactoglobulin, International Dairy Journal, 20 (2010) 742-752.

[39] K.M. Biemel, O. Reihl, J. Conrad, M.O. Lederer, Formation pathways for lysine-arginine cross-links derived from hexoses and pentoses by Maillard processes: Unraveling the structure of a pentosidine precursor, Journal of Biological Chemistry, 276 (2001) 23405-23412.

[40] R. Fritsch, H. Hoffmann, H. Klostermeyer, Formation of lysinoalanine during heat treatment of milk, Zeitschrift für Lebensmittel-Untersuchung und Forschung, 176 (1983) 341-345.

[41] S. Salazar-Villanea, E.M.A.M. Bruininx, H. Gruppen, P. Carré, A. Quinsac, A.F.B. van der Poel, Effects of toasting time on digestive hydrolysis of soluble and insoluble 00-rapeseed meal proteins, Journal of the American Oil Chemists' Society, 94 (2017) 619-630.

[42] S.J. Hur, B.O. Lim, E.A. Decker, D.J. McClements, *In vitro* human digestion models for food applications, Food Chemistry, 125 (2011) 1-12.

[43] S.A. Adibi, D.W. Mercer, Protein digestion in human intestine as reflected in luminal, mucosal, and plasma amino acid concentrations after meals, The Journal of Clinical Investigation, 52 (1973) 1586-1594.

[44] S.G. Bradley, T.M. Antalis, J.S. Bond, Proteases in the mammalian digestive system, Proteases: Structure and function, Springer Vienna, Vienna, 2013, pp. 373-393.

[45] M. Merz, T. Eisele, P. Berends, D. Appel, S. Rabe, I. Blank, T. Stressler, L. Fischer, Flavourzyme, an enzyme preparation with industrial relevance: Automated nine-step purification and partial characterization of eight enzymes, Journal of Agricultural and Food Chemistry, 63 (2015) 5682-5693.

[46] S.M. Rutherfurd, Methodology for determining degree of hydrolysis of proteins in hydrolysates: A review, Journal of AOAC International, 93 (2010) 1515-1522.

[47] J. Adler-Nissen, Enzymic hydrolysis of food proteins, Elsevier Applied Science Publishers, London, UK, 1986, pp. 135-142.

[48] H. Schlichtherle-Cerny, M. Affolter, C. Cerny, Hydrophilic interaction liquid chromatography coupled to electrospray mass spectrometry of small polar compounds in food analysis, Analytical Chemistry, 75 (2003) 2349-2354.

[49] C. van Rooijen, G. Bosch, C.I. Butré, A.F.B. van der Poel, P.A. Wierenga, L. Alexander, W.H. Hendriks, Urinary excretion of dietary Maillard reaction products in healthy adult female cats, Journal of animal science, 94 (2016) 185-195.

Summary

In food industry, enzymatic protein hydrolysis is used to produce protein hydrolysates for use as ingredients in food products, and as an alternative method to simulate *in vivo* protein digestion. However, the final degree of hydrolysis reached experimentally had many uncertainties in terms of changes in protein structure or system conditions. The goal of this PhD project was to develop a model to define expectations of the extent of protein hydrolysis based on molecular properties of the protein and the protease, and effects of modifications on protein primary structure and changes in system conditions. The current state of knowledge on enzymatic protein hydrolysis is discussed in **Chapter 1**. It was found that the estimation of the theoretical maximum degree of hydrolysis ($DH_{max,theo}$), calculated by the content (in % of number) of cleavage sites (CSs) of a protease in a protein, did not match the experimental data reported in literature. Two possible explanations, the enzyme secondary specificity and the formation of inhibitory peptides during hydrolysis, are discussed. Despite the widespread information of the two hypotheses, little attention has been given to the possible consequence each factor has on the hydrolysis process. Moreover, the effect of the Maillard reaction on protein hydrolysis was widely studied, however, no mechanistic understanding was obtained.

The influence of glycation, the initial stage of the Maillard reaction, on protein hydrolysis, is described in **Chapter 2**. Apo α-lactalbumin (α-LA) was glycated with D-glucose at 50 °C for 0-10 h, and hydrolysed by lysine/arginine specific proteases (bovine and porcine trypsins) or by non-lysine/arginine specific proteases (*Bacillus licheniformis* protease (BLP), bovine α-chymotrypsin and subtilisin A). The Maillard incubation resulted in proteins with different average degrees of glycation (DG_{av}=0-63 %) without modifications on the protein secondary, tertiary and quaternary structure. For hydrolysis by bovine or porcine trypsins, the experimental maximum degree of hydrolysis ($DH_{max,exp}$) decreased linearly by 65 % from non-glycated to maximal glycated proteins. BLP and subtilisin A hydrolyses were independent of glycation, while α-chymotrypsin cannot hydrolyse cleavage sites (CSs) with glycated neighbouring amino acids (binding site of the CS). Combining results of all enzymes, an equation was derived to describe the effect of modifications on the protein primary structure (through glycation) on the extent of hydrolysis based on the enzyme specificity, selectivity and binding site sensitivity.

It is described in **Chapter 2** that the influence of glycation on protein hydrolysis can be predicted. However, the $DH_{max,exp}$ of non-glycated proteins needed for this prediction has to be obtained experimentally. Therefore, for **Chapter 3**, the aim was to predict the $DH_{max,exp}$ of hydrolysis of a non-modified protein by bovine trypsin, based on the enzyme secondary specificity, i.e. the enzyme preference towards neighbouring amino acids (AAs) of the CSs. The consequence of the enzyme secondary specificity on protein hydrolysis was that the enzyme hydrolyses CSs with different rates. Previously, a quantitative parameter that describes the relative hydrolysis rate constants of CSs in a protein was referred to as the enzyme selectivity. To see if the enzyme selectivity can be used to obtain the enzyme secondary specificity, the link between bovine trypsin selectivity and its secondary specificity was studied. This was done by full quantitative

analysis of peptides formed during bovine tryptic hydrolysis of apo α-LA and β-casein (β-cas). Results showed that ≥50 % of the CSs were inefficiently hydrolysed. The enzyme selectivity was found to depend on the number of charged AAs at P2 and P2' positions of a CS. Bovine trypsin efficiently hydrolyses CSs with neutral AAs at these two positions. Using the obtained secondary specificity, bovine trypsin selectivity towards 67 % of CSs in β-lactoglobulin was correctly predicted. The predicted maximum degree of hydrolysis ($DH_{max,pre}$) based on selectivity is ~5 times better than the prediction based on enzyme specificity ($DH_{max,theo}$).

In **Chapter 4**, the approach as described in **Chapter 3** for bovine trypsin was used to reveal the secondary specificity of porcine and human trypsins. The aim was to explain their differences in $DH_{max,exp}$ based on their differences in secondary specificity (selectivity). In **Chapter 2**, by porcine trypsin, the $DH_{max,exp}$ of 0.1 % apo α-LA hydrolysis is described to be similar to $DH_{max,theo}$, whereas the $DH_{max,exp}$ of bovine tryptic hydrolysis is much lower. In **Chapter 4**, this observation was confirmed in hydrolysis of 1 % apo α-LA and 1 % β-cas. For both substrate proteins, large differences were also found between the $DH_{max,exp}$ of hydrolysis by human trypsins and the $DH_{max,theo}$. This was explained by the fact that porcine trypsin efficiently hydrolysed ~78 % of the CSs, whereas bovine and human trypsins only efficiently hydrolysed ~47 % and ~53 % of the CSs, respectively. The predicted $DH_{max,pre}$ based on the enzyme selectivity was ~4 times closer to $DH_{max,exp}$ than the prediction based on enzyme specificity ($DH_{max,theo}$). Preliminary secondary specificities of porcine and human trypsins were proposed. Porcine trypsin was found to be hindered by charged AAs on P2 position, and human trypsin-2 was hindered by charged AAs on P2 and P1' positions. This work showed that there are large differences in enzyme selectivity of the three trypsins, leading to large differences in $DH_{max,exp}$. The knowledge of the selectivity of these enzymes allows the estimation of $DH_{max,pre}$ of tryptic hydrolysis of other substrate proteins.

In **Chapters 3** and **4**, the $DH_{max,exp}$ of bovine, porcine and human trypsins are explained based on the enzyme selectivity. However, for several proteases, large variations in the $DH_{max,exp}$ were found to result from changes in the substrate concentration even at a constant enzyme to substrate ratio (E:S). This indicated that, next to secondary specificity, another mechanism was involved in reaching the $DH_{max,exp}$. This mechanism is described in **Chapter 5**. Apo α-LA was hydrolysed by bovine, porcine and human trypsins, bovine α-chymotrypsin or BLP at various substrate concentrations. Hydrolysis by human trypsin was not sensitive to changes in substrate concentration. The $DH_{max,exp}$ of hydrolysis by bovine and porcine trypsins, or bovine α-chymotrypsin increased with an increasing substrate concentration, while the $DH_{max,exp}$ of BLP hydrolysis decreased. Besides the change in $DH_{max,exp}$, also a change in hydrolysis mechanism, described by the percentage of remaining intact protein as a function of DH, was observed for hydrolysis with different substrate concentrations. This suggested that the concentration of individual peptides formed during hydrolysis changed with the substrate concentrations. Assuming that some of these peptides could act as protease inhibitors, the observed decrease in $DH_{max,exp}$ could be explained. This hypothesis was tested by performing the hydrolysis at a

higher E:S. Increasing the E:S from 1:100 to 1:25 for 0.1 % apo α-LA hydrolysis by bovine trypsin indeed led to a similar value of $DH_{max,exp}$ reached for hydrolysis of 1 % apo α-LA at an E:S of 1:100.

In addition to the questions about the effect of glycation on protein hydrolysis, there were also questions about the glycation of protein hydrolysates. During industrial processing of protein hydrolysates, peptides can be glycated if carbohydrates are present. In hydrolysates, peptides with different lengths, type (α/ε-) and number of amino groups are present. The effects of these molecular properties on glycation kinetics of peptides in hydrolysates are described in **Chapter 6**. Apo α-LA was hydrolysed by *Bacillus lichenformis* protease (to DH 2, 4, 6 and 8 %), and glycated with D-glucose at 50 °C for 0-8 h. The hydrolysates contained, in addition to the remaining intact protein, 24 unique peptides, ranging from 2-123 AAs with 0-12 lysine(s). The glycation rate constant (k_g) and the maximum average degree of glycation ($DG\_P_{av,max}$) of these peptides were independent of the hydrolysate composition. The k_g values showed 3 populations, but could not be correlated with the peptide length or the number of amino groups per peptide. Surprisingly, the $DG\_P_{av,max}$ of peptides did not reach 100 %. The maximum degrees of glycation of α- and ε-amino groups were 13 % and 60 %, respectively, which explained why the $DG\_P_{av,max}$ ranged from 9 % to 59 % for peptides with 0-12 lysine(s). This work showed that the extent of glycation of peptides in hydrolysates can be predicted based on the AA sequences of the peptides.

To provide a perspective for the industrial and societal relevance of topics studied in this thesis, in **Chapter 7**, the risk analysis and technology assessment, an essential part of the NanoNextNL programme, is discussed. In order to predict and fine-tune the digestibility of food products, it is beneficial for food industry to understand the relation between the industrial processing and the protein digestibility. However, it remains challenging to translate the conclusions from model systems in laboratories to real-life scenarios. In addition, differences in the terminology and the determination methods of the protein digestibility between food and animal nutrition sciences were identified, leading to misinterpretation. These challenges can be solved by integrating the factors that may influence protein hydrolysis to model systems, preferably using computational simulations, and by opening interdisciplinary dialogues between the two fields and the industries.

A step towards computational simulations of the $DH_{max,exp}$ of hydrolysis of a given protein by a specific protease is described in **Chapter 8**. We developed a model (equation 1), where the prediction of $DH_{max,exp}$ (=$DH_{max,pre}$) was based on 1) the enzyme secondary specificity (parameter *a*, **Chapters 3** and **4**), and 2) the enzyme affinity towards intact proteins (parameter *b*, **Chapter 5**). These in turn depend on the AA sequence of the protein, as well as possible modifications (parameters *m*, *n* and *q*, **Chapters 2** and **6**), and the enzyme sensitivity towards binding site positions of CSs. Parameter *b* depends in principle on the structural stability of the substrate protein, but is affected by substrate concentrations used in the hydrolysis.

$$(1) \quad DH_{max,pre}\left[\%\right] = DH_{max,theo} \cdot a \cdot b - \frac{m \times n \times q}{\# \text{ peptide bonds}}$$

#peptide bonds refers to the number of peptide bonds per protein. Based on equation 1 and the parameter estimates obtained in the thesis, an estimation of $DH_{max,pre}$ was made as a function of substrate concentration and DG_{av} for apo α-LA hydrolysis by different enzymes. Of course, in industrial ingredients, proteins are present in different structural states. The extent to which these factors can still be accounted for in the model was tested using data from literature and other experimental data. For this, the hydrolysis of globular, unfolded and aggregated proteins by a single enzyme and/or by a sequential enzyme incubation (to mimic the *in vivo* protein digestion) are described. This model is shown to be an effective tool that can be used to define expectations of the extent of enzymatic protein hydrolysis. This can -in the future- help to control and predict the protein hydrolysis and the *in vitro* protein digestion.

Acknowledgements

When I looked back these four years, I realized that I am actually truly happy about my life as a PhD student, even during those super busy periods of the PhD project. I would like to take this opportunity to express my deep appreciation for the people who have guided and accompanied me through the four years.

Firstly, I am very grateful for all the supervision from my awesome promotor and co-promotors. Harry, thank you for the extremely useful discussions we had about my articles. I have learned so much from you, to the point that I could quote your comments in my head when I corrected my reading version of the thesis myself before handing in. Besides your content-wise support, I also truly appreciate your moral support. Every time I showed uncertainty about my capability for the work (which happened several times…), you spotted it and ensured me that I can do it. Peter, you are like the light in my journey. In the middle of my chaos, there was you, who have guided me where to go and where to land. You are there for me during my MSc thesis and internship, my PhD and my post-doc. You had my trust since day one, which actually now makes me a bit insecure knowing that one day I will not be a student of you anymore. 我会好好学习. Henk, thank you so much for making time for me during those periods that I was super stressed and you were super busy. You pushed me out of my comfort zone to make a stronger conclusion for my thesis, which was a good learning moment for me. Thank you for helping me to achieve it. I am proud to be a member of the FCH family. I hope you all would also be proud of having me as a student.

Secondly, I would like to thank my other co-authors. Claire, I feel like I shared everything at work during the four years with you. Communicating with you made everything easier. Thank you so much for helping me to analyse results, to write two chapters and to do the final spelling check for the thesis. And thank you for introducing me to the cutest little Sofia, and allowing me to be a small part of her life. Hugo, you were there for me when I had problems with Synapt, when I couldn't reach the top shelf in the lab, when I couldn't carry the dialysis bottles, when I couldn't lift my boxes in Q3, when I couldn't drive (I am learning now!). Thank you for being my project partner. It is a great and fun journey. Stefano, I really appreciate your support to my work. Many thanks to your expertise and your patience. Frank, you were my first student. I was probably more nervous than you at that time. Thanks for your contribution to my project.

Well, I will now thank two persons that I have never met in my life but they are very important in these four years, Kaj Ulrik Linderstrøm-Lang and Jens Adler-Nissen. My deep appreciation for your contribution to my project was shown in all previous chapters.

My special thank goes to Jolanda. You are the secret of everybody's success in FCH. You have made everything so easy for us. I had so many problems with the contract, the residence permit, etc. Thank you for your help and patience.

I would also like to thank Jean-Paul, for your support during the last phase of my PhD. You are so sensitive to my facial expression that you comforted me every time when I was stressed.

I would also like to thank our technicians. Thank you for helping me with experiments, ordering goods and fixing equipments. Special thank goes to René, the person I go to when things looked weird and I didn't know what to do.

Many thanks to my best protein buddies, Claire, Roy, Frederik, Surender, Hugo, Emma, Alexandra, Abhishek, Maud, Vian, Evan, Yunus, Fen, Hsuan, Sergio, Lotte and of course Peter, Stefano and René. Our protein meeting is the best meeting ever. I enjoyed our discussions a lot. You guys are like a protein Wikipedia. If I have a question, at least one of you would know the answer. Besides work, I also had a lot of fun during our protein dinners, canoeing trip, escaping room experience. Special thank goes to Roy, who helped me a lot with my research. Discussing my results with you became a habit of mine. You are missed a lot in FCH!

I would like to thank my lab mates, Mark, Silvia, Sylvia, Carla, Anne, Renske, Milou, Annewieke, Wouter, Ya, Zhibin and Junfeng. As a tiny protein girl in a busy phyto lab, you need to make room for stuffs that only I would use and I always need to keep an extra step for myself…

I would also like to thank my dear officemates, Claire, Abhishek, Connie, Carla, Bianca, Maud, Moheb and Dimitris. I had a wonderful time in our office. You guys were so nice and patient to me even when my stress level in the last months was so high that I knew it affected your work…

Many thanks to my BSc and MSc students, Frank, Dimitris, Lirong, Maria, Annemiek, Jeffrey and Nicole. Thank you for your interests in my topic. This actually means a lot to me, that there are more people who are passionate about protein hydrolysis like me.

I would like to thank Lingmin, Red, Fangjie and Ya. 谢谢你们这些年的鼓励，支持和帮助。在你们每个人身上我都看到了很多我可以学习借鉴的地方。能和你们在一个组读博士是我的幸运。

During the four years, I have been in two committees, the Lab trip committee 2015 and the PhD trip committee 2016. Hugo and Jorien, we organised the lab trip together and it was such a fun activity. I could still laugh for hours about the "giraffe!" incident. Suzanne, Renske, Wouter and Bianca, our Japan PhD trip committee. We were very efficient, organised and always on the same page. Let's cherish again the moment when Harry said it was a very good trip, not an average trip. I would also like to thank Matthias, Hugo, Gijs, Donny and Thibaut for your contributions to my weirdness during the trip. I guess probably there will not be anybody I would meet in the future, who would launch me from a chair, put my shoe inside of another shoe, and carry me to film a Japanese girl band. These are all very special memories. Silvia, we know each other for almost 7 years, and in Japan we even spent 7 nights together. Thank you for being there for me.

About the author

Curriculum Vitae

Yuxi Deng was born on 2nd August, 1989 in Wuhan, Hubei province, China. Her family moved to Beijing when she was four years old. After finishing a 6-year science advance class in the middle and high school affiliated to Renmin University of China, she started her bachelor study of Food Science and Engineering at China Agricultural University in 2007. During her bachelor study, she attended the CARS International Student Forum as the representative of China in 2011 at University of Tsukuba, Japan. Her BSc degree was completed with a thesis project at the Laboratory of Functional Dairy Products on the quantification of intracellular amino acids during the fermentation of *Lactobacillus salivarius*, under the supervision of Prof. Beizhong Han. In 2011, Yuxi started her Master programme of Food Technology at Wageningen University. She finished her MSc thesis at the Laboratory of Food Chemistry on the quantification of di-tyrosine during protein cross-linking with horseradish peroxidase, under the supervision of Dr. Surender Dhayal and Dr. Peter Wierenga. Her MSc internship was performed at FrieslandCampina DMV, Wageningen, on the characterization of (physico-)chemical properties of caseinates, under the supervision of Dr. Arno Alting and Dr. Peter Wierenga. After obtaining her MSc degree in 2013, she started her PhD research at the Laboratory of Food Chemistry under the supervision of Prof. Harry Gruppen, Dr. Peter Wierenga and Prof. Henk Schols. The results of her PhD research are presented in this thesis. Currently, Yuxi is working as a postdoctoral researcher at the Laboratory of Food Chemistry.
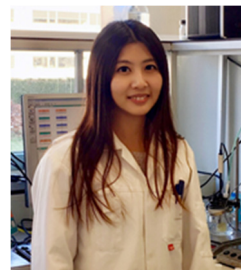
Contact: deng_yu_xi@hotmail.com

List of publication

Y. Deng, P.A. Wierenga, H.A. Schols, S. Sforza, H. Gruppen, Effect of Maillard induced glycation on protein hydrolysis by lysine/arginine and non-lysine/arginine specific proteases, Food Hydrocolloids, 69 (2017) 210-219.

Y. Deng, F. van der Veer, S. Sforza, H. Gruppen, P.A. Wierenga, Towards predicting protein hydrolysis by bovine trypsin, Process Biochemistry, 65 (2018) 81-92

Y. Deng, H. Gruppen, P.A. Wierenga, Comparison of protein hydrolysis catalysed by bovine, porcine or human trypsins, *Submitted for publication*.

Y. Deng, C.I. Butré, P.A. Wierenga, Influence of substrate concentration on the extent of enzymatic protein hydrolysis, *Submitted for publication*.

Y. Deng, C.I. Butré, P.A. Wierenga, Understanding glycation kinetics of individual peptides in protein hydrolysates, *Submitted for publication*.

H.B. Cardoso, **Y. Deng**, P.A. Wierenga, H. Gruppen, H.A. Schols, Influence of saccharide size and charge on the glycation reactivity of individual lysines in α-lactalbumin during the Maillard induced glycation, *To be submitted*.

C.I. Butré, **Y. Deng**, M. Jonathan, H. Gruppen, P.A. Wierenga, Determination of the extent of protein modification/glycation by quantification of Maillard reaction products by MS, *To be submitted*.

Overview of completed training activities

<u>Discipline specific activities</u>
Courses
Summer Course Glycosciences[1], VLAG/GBB, Wageningen, NL, 2014
Reaction kinetics in food science, VLAG, Wageningen, NL, 2014
Advanced Food Analysis[1], VLAG, Wageningen, NL, 2015
Food & Biorefinery Enzymology[1], VLAG, Wageningen, NL, 2015
Asymmetric flow field flow fractionation on-site training, Wyatt, Wagenigen, NL, 2016
Unify LC-MS on-site training, Waters/FCH, Wagenigen, NL, 2016

Conferences and meetings
Workshop "Bubbles, droplets and spheres"[1], NanoNextNL, Wageningen, NL, 2014
12th International Symposium on the Maillard Reaction[2], IMARS, Tokyo, Japan, 2015
NanoCity 2015[1], NanoNextNL, Amersfoort, NL, 2015
Maillard workshop on food and feed[2], FQD, Wageningen, NL, 2015
FrieslandCampina Maillard Symposium[2], FrieslandCampina, Wageningen, NL, 2015
NanoCity 2016[1], NanoNextNL, Amsterdam, NL, 2016

<u>General courses</u>
Intellectual property and valorisation awareness, NanoNextNL, Utrecht, NL, 2014
Scientific Writing, WGS/WUR in'to Languages, Wageningen, NL, 2015
Mobilising your scientific network, WGS, Wageningen, NL, 2015
Analytic storytelling, NanoNextNL, Utrecht, NL, 2015
Risk analysis and technology assessment (coaching), NanoNextNL, Wageningen, NL, 2016
Career Prospective, WGS/VLAG, Wageningen, NL, 2017

<u>Additional activities</u>
Preparation of research proposal, FCH, Wageningen, NL, 2014
PhD study trip to Germany, Denmark, Sweden and Finland[1,2], FCH, 2014
PhD study trip to Japan[1,2] (organizer), FCH, 2016
PhD presentations[2], FCH, Wageningen, NL, 2013-2017
BSc & MSc student presentations, FCH, Wageningen, NL, 2013-2017
NanoNextNL project meetings[1,2], NanoNextNL, NL, 2013-2016
FrieslandCampina project meetings[2], FCH/FrieslandCampina, Wageningen, NL, 2013-2017

[1] poster presentation; [2] oral presentation;

VLAG: Graduate School for Nutrition, Food Technology, Agrobiotechnology and Health Sciences;

GBB: Groningen Biomolecular Sciences and Biotechnology Institute; WUR: Wageningen University & Research

IMARS: International Maillard Reaction Society; WGS: Wageningen Graduate Schools;

FCH: Laboratory of Food Chemistry; FQD: Laboratory of Food Quality of Design; NL: The Netherlands