

# Amalyse modifications during production and application

**Nielsen, Connie**

*Publication date:*
2012

*Document Version*
Publisher's PDF, also known as Version of record

# Amylase modifications during production and application

Industrial Ph.D. Thesis

July 2012

**Connie Nielsen**

Enzyme and Protein Chemistry (EPC)
Department of Systems Biology
Technical University of Denmark (DTU)

Detergent Research Center
Novozymes A/S

Supervisors:

Science Manager Svend Kaasgaard, Global Assay Development, Novozymes A/S, DK

Senior Scientist Carsten Sönksen, Protein Technology, Novozymes A/S, DK

Prof. Birte Svensson, EPC, Department of Systems Biology, DTU, DK

Assoc. Prof. Susanne Jacobsen, EPC, Department of Systems Biology, DTU, DK

Prof. Ole Nørregaard Jensen, Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK


# Preface

The following PhD thesis entitled "Amylase modifications during production and application" describes the most important results achieved during my Ph.D. study. The project was financed by Novozymes A/S and the Danish Agency for Science Technology and Innovation under the industrial Ph.D. program. The work was carried out at the Detergent Research Center, Novozymes A/S and at Enzyme and Protein Chemistry, Department of Systems Biology at the Technical University of Denmark.

Part of the work is presented in the following two manuscripts (Appendix II and III):

C. Nielsen, S. G. Kaasgaard, C. Andersen, C. P. Sönksen, S. Jacobsen, B. Svensson. Characterisation of chemical modifications on *Bacillus licheniformis* α-amylase and the implications on detergent wash performance. Manuscript in preparation.

Connie Nielsen, Svend G. Kaasgaard, Susanne Jacobsen, Birte Svensson, Carsten P. Sönksen. Identification and Quantification of Non-Enzymatically Glycated Peptides by Electron Transfer Dissociation with Isobaric Tags. Manuscript in preparation.

Some of the work performed during this project is not presented in this PhD thesis as it is part of a patent application filed on June 30$^{th}$ 2011.

**Supervisors**

Science Manager, Ph.D. Svend Kaasgaard, Global Assay Development, Novozymes A/S, Bagsværd, Denmark.

Senior Scientist, Ph.D. Carsten Sönksen, Protein Technology, Novozymes A/S, Bagsværd, Denmark.

Professor, Ph.D. Birte Svensson, Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark.

Associate Professor, Ph.D. Susanne Jacobsen, Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark.

Professor, Ph.D. Ole Nørregaard Jensen, Department of Biochemistry and Molecular Biology, University of Southern Denmark.

## Acknowledgement

First, I wish to thank my supervisors Svend Kaasgaard, Carsten Sönksen, Birte Svensson and Susanne Jacobsen for their support, scientific advice, fruitful discussions and their excellent guidance throughout this project. A special thank to my primary supervisor Svend for his patient and supportive mentoring as well as for his hard work with legal and patent making this Ph.D. possible. I also wish to thank my third party supervisor Ole Nørregaard Jensen for signing on as supervisor.

Novozymes A/S and the Danish Agency for Science Technology and Innovation are gratefully acknowledged for funding this Ph.D. project.

Thanks to Hanne Tynell and Jan Danielsen from Novozymes for their assistance and helpfulness in the laboratory. Hanne is also thanked for always being there to listen when I needed it and many good conversations.

Thanks to Liv Spångner Christiansen from Novozymes for many great runs keeping me energised and optimistic during the time course of this project and also to her and Mie Østergaard Hunderup for their assistance with the mini wash experiments. Henrik Lund from Novozymes is thanked for his help with the DSC and for the conversations on the troubles of writing a Ph.D.-thesis. Jens Øbro from Novozymes is thanked for his camaraderie and constant advising on how to best get though a PhD project.

A special thanks to Jan Juul de Jong, Peter Riis Eriksen and Clive Phipps Walter for their invaluable help with the mass spectrometry, as well as a thanks to the rest of the MS-Edman group at Novozymes for their helpfulness and interest in my work.

I wish to thank Carsten Andersen from Novozymes for providing the model protein BASI, for expression of the BLA variants and for sharing his vast knowledge of the amylase structure. Thanks to Lene Bjørg Cesar from Novozymes for her help and interest in the statistical aspects of this project. Thanks to Anne Blicher from Enzymes and Protein Chemistry, DTU Systems Biology for performing the amino acid analysis and to Jesper Hardt at Ramcon and Samuel Fox at Beckman Coulter for performing capillary isoelectric focusing.

I thank Karen Marie Jakobsen from Enzymes and Protein Chemistry, DTU Systems Biology and Lene Priem Pedersen from Novozymes for guidance and assistance with all the administrative issues.

I thank all the members of Enzymes and Protein Chemistry, DTU Systems Biology, the Detergent Research Center DK and the MS-Edman group at Novozymes for providing a wonderful work environment.

Last but not least a special thanks to my family and friends for their love and support.

# Table of Contents

## Summary

α-Amylases are a family of enzymes which catalyse the hydrolysis of α-D-(1,4) glucosidic linkages in starch and related compounds. α-Amylases are some of the most important industrial enzymes and they are used in a number of different industries one of these being the detergent industry. The commercial α-amylases are produced by fermentations of genetically modified bacteria, where the native gene has been manipulated to code for an enzyme with improved performance characteristics. Despite the genome only contains copies of the same α-amylase gene, multiple different α-amylase forms varying in pI value can be identified when analysing the final enzyme product, hereby indicating that the protein is modified.

The purpose of the present Ph.D. project is to identify differences between the multiple forms of α-amylase derived from the same gene and determine if or how these various modifications affect the activity and wash performance. The focus of the present project is on α-amylases from *Bacillus* species used in the detergent industry. The thesis begins with an introduction to *Bacillus* α-amylases, modifications which could cause the multiple forms varying in pI value and mass spectrometry (MS) methods used for the characterisation of these modifications. This is followed by a chapter concerning the methods used, followed by a chapter discussing the results obtained during the project.

Studies of the multiple forms of α-amylases revealed that the multiple forms occur during the fermentation, where all the forms were observed already early on in the process. Characterisation of the wash performance of the multiple forms from three different *Bacillus* amylases showed tendencies of decreasing wash performance with decreasing pI values. A more thorough characterisation of pools of *Bacillus licheniformis* α-amylase (BLA) varying in pI revealed that the pools containing the forms with the lowest pI values had slightly lower catalytic efficiency on maltoheptaose and amylose.

Identification of the modifications responsible for the multiple forms of BLA was performed by tandem mass spectrometry (MS/MS). One of the modifications suspected to be responsible for the multiple forms, i.e. glycations, can be difficult to identify by the fragmentation methods normally used in MS/MS, such as higher-energy C-trap dissociation (HCD) as they tend to cleave of the modification. Therefore, HCD was compared with the soft fragmentations method electron transfer dissociation (ETD), which should be better at identifying these modifications because it leaves the modifications intact. Both methods were able to identify glycation sites on a model protein, though ETD identified more sites and provided higher sequence coverage for the glycated peptides.

MS/MS analysis by HCD and ETD of the multiple forms of BLA identified deamidations, glycations and carbamylations as the modifications responsible for the forms varying in pI. The modified sites were mainly located on the surface of BLA with a few deamidations in the area around the active site. The modifications were in some cases observed on amino acids involved in stability, activity and substrate binding. Therefore, some of the multiple forms of BLA will most likely differ in stability, activity and performance due to these modifications.

Finally the influence of these modified sites was evaluated by characterisation of BLA variants with single site specific mutations. The BLA variants with mutations mimicking deamidations in the active site and in the substrate binding cleft were found to influence the kinetic parameters ($k_{cat}$ and $K_M$) in different ways. The wash performance was also influenced by these mutations though these effects were mainly positive. Modified Lys and Arg residues suspected to be involved in substrate binding of longer substrates or secondary binding sites were evaluated by site specific mutations to amino acids with neutral side chains of approximately the same size. Hereby, evaluating the effect of removing possible carbohydrate interaction sites. This did not have any effect on the kinetic parameters but it generally resulted in increased wash performances.

# Dansk resumé

α-Amylaser tilhører en familie af enzymer, der katalyserer hydrolysen af α-D-(1,4) glykosidbindinger i stivelse og lignende forbindelser. α-Amylaser er et af de vigtigste industrielle enzymer i dag og det benyttes i mange forskellige industrier, som for eksempel detergent industrien. Industrielle amylaser produceres ved fermentering af genetisk modificerede bakterier, hvor det oprindelige amylase-gen er blevet manipuleret til at kode for et enzym med forbedret ydeevne. På trods af at den benyttede ekspressions-celle kun indeholder kopier af dette ene amylase-gen identificeres mange forskellige α-amylaser med forskellige pI værdier i det endelige produkt. Dette indikerer at enzymet bliver modificeret.

Formålet med dette Ph.D. projekt er at identificere forskellene mellem de mange amylaseformer og undersøge om disse observerede modifikationer har en indflydelse på enzymets aktivitet og vaskeegenskaber. Fokus i dette projekt er på *Bacillus* α-amylaser, der benyttes i detergent industrien.

Ph.D.-afhandlingen begynder med en introduktion af *Bacillus* α-amylaser, de modifikationer der kan forårsage de observerede pI ændringer og de massespektrometriske metoder, der benyttes i projektet til identifikation af disse modifikationer. Derefter følger et metode afsnit og til sidst præsenteres og diskuteres de opnåede resultater.

Prøver fra fermenteringen viste, at de mange amylaseformer dannes allerede tidligt i løbet af denne proces. Karakterisering af de mange amylaseformer fra tre forskellige *Bacillus* amylaser viste at vaske egenskaberne blev forringet for formerne med de lavere pI værdier. Fraktioner indeholdende *Bacillus licheniformis* α-amylase (BLA) former med forskellige pI niveauer blev karakteriseret grundigere. Dette viste at fraktionerne indeholdende BLA formerne med de laveste pI værdier også havde lavere katalytisk effektivitet.

Modifikationer, der forårsager de forskellige BLA former, blev identificeret ved brug af tandem-massespektrometri (MS/MS). Glykeringer er en af de modifikationer der forventes at være med til at forårsage de forskellige former. Disse modifikationer kan være svære at identificere ved brug af de gængse fragmenteringsmetoder i MS/MS såsom higher-energy C-trap dissociation (HCD), da de kan kløve modifikationen af under fragmenteringen. Derfor blev HCD sammenlignet med en mindre hård fragmenteringsmetode kaldet electron transfer dissociation (ETD), som vil være bedre til at identificere glykeringer, da den ikke skulle kløve labile modifikationer af. Begge fragmenteringsmetoder viste sig brugbare til identifikation af glykeringer, men ETD identificerede flere glykeringssites og gav generelt en bedre sekvensdækning for de glykerede peptider.

MS/MS analyser af de forskellige BLA former ved brug af både HCD og ETD fragmentering identificerede modifikationerne deamideringer, glykeringer og carbamyleringer. Disse modifikationer vil alle forårsage de fald i pI-værdier, som blev observeret for amylaserne. Modifikationerne blev kun observeret på aminosyrer på BLA's overflade på nær nogle enkelte deamideringer, der var i området lige omkring det aktive site. Nogle af de modificerede

aminosyrer er involverede i enzymets stabilitet, aktivitet og substratbinding. Nogle af modifikationerne har derfor sandsynligvis en indflydelse på disse faktorer.

Indflydelsen af disse modifikationer på BLA blev undersøgt for nogle specifikke sites. Dette blev gjort ved karakterisering af BLA varianter med site specifikke mutationer. Kinetiske analyser af BLA varianter med mutationer, der ligner deamideringer i det aktive site og i substrat bindingskløften, viste at disse enkelt mutationer havde betydning for de kinetiske parametre $k_{cat}$ og $K_M$, bestemt på amylose og maltoheptaose. Disse BLA varianter viste også generelt forbedrede vaskeegenskaber. Modifikationer på Lys- og Arg-sidekæderne, der kunne være involveret i binding af længere substrater eller en del af sekundære bindings sites, blev også analyseret ved site specifikke mutationer til aminosyrer med neutrale sidekæder af nogenlunde samme størrelse. Derved kunne effekten af at fjerne mulige glykerings sites vurderes. Disse mutationer havde ingen indflydelse på de kinetiske parametre bestemt på substraterne amylose og maltoheptaose, men de havde derimod en generel positiv effekt på BLA's vaskeegenskaber.

## Confidentiality

The work presented in this thesis is based on an industrial PhD project in collaboration with Novozymes A/S and as a consequence there were certain constraints due to protection of intellectual property rights and confidential knowhow. These matters have been taken into consideration in some aspects of the project and the thesis such as the protein selected for thorough characterisation as well as the description of structure and sequence information.

## List of abbreviations

| | |
|---|---|
| AA560 | Alkaline *Bacillus* sp. α-amylase |
| AGC | Automatic gain control |
| AGE | Advanced glycation end product |
| BAN | *Bacillus amyloliquefaciens* α-amylase |
| BASI | Barley α-amylase/subtilisin inhibitor |
| BHA | *Bacillus halmapalus* α-amylase |
| BLA | *Bacillus licheniformis* α-amylase |
| CID | Collision induced dissociation |
| cIEF | Capillary isoelectric focusing |
| CV | Column volume |
| DP | Degree of polymerisation |
| DSC | Differential scanning calorimetry |
| ECD | Electron capture dissociation |
| ESI | Electrospray ionisation |
| ETD | Electron transfer dissociation |
| FT-ICR | Fourier transform ion cyclotron resonance |
| $G_7$ | Maltoheptaose |
| $G_7$-pNP | Maltoheptaose *p*-nitrophenol |
| HCD | Higher energy C-trap dissociation |
| HPLC | High pressure liquid chromatography |
| IEF | Isoelectric focusing |
| iTRAQ | isotope tags for relative and absolute quantification |
| $k_{cat}$ | Catalytic constant |
| $k_{cat}/ K_M$ | Catalytic efficiency |
| $K_M$ | Michaelis-Menten constant |
| LC | Liquid chromatography |
| LIT | Linear ion trap |
| MALDI | Matrix assisted laser desorption ionisation |
| MG-HI | Methylglyoxal-derived hydroimidazolone |
| Mw | Molecular weight |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| m/z | Mass per charge |
| pI | Isoelectric point |
| PTM | Post-translational modification |
| TMT | Tandem mass tags |
| $T_{max}$ | Temperature where half of the proteins are denatured |

# 1 Introduction

α-Amylases are among the most important industrial enzymes and within the detergent industry they are the second most used class of enzymes. In the detergent industry α-amylases are used to remove starch containing stains from clothes. The industrial α-amylases can become modified during the production resulting in multiple forms varying in pI values. The effect of these modifications on the activity of the α-amylase is unknown.

It is the purpose of this section to introduce *Bacillus* α-amylases, the modifications which could create the multiple forms varying in pI value and the theory behind the mass spectrometry (MS) methods used in the present project for the characterisation of these modifications.

## 1.1 α-Amylase

α-Amylases are a family of enzymes which catalyse the hydrolysis of α-D-(1,4) glucosidic linkages in starch and related compounds hereby releasing maltooligosaccharides and glucose in the α-anomeric configuration, hence their name (Machius *et al.*, 1998; Vihinen *et al.*, 1989). α-Amylases are among the most important industrial enzymes with widespread use in various industries. The enzymes are used for starch liquefaction and saccharification in the starch and fuel alcohol industries, for starch stain removal in the detergent industry as well as in industries such as baking and brewing (de Souza *et al.*, 2010; Kirk *et al.*, 2002; Nielsen *et al.*, 2000).

Starch is made of glucose residues and serves as an important energy reserve in plants. The major components of starch are the two polysaccharides amylose and amylopectin, Figure 1. Amylose is essentially a linear α-glucan consisting of approximately 99% α-(1→4) and 1% α-(1→6) bonds. This means that a certain degree of branching does occur with between nine and twenty branch points per molecule. The number of glucose residues per molecule, also referred to as degree of polymerisation (DP), range between 650-5000. Amylopectin is a bigger polysaccharide with more branching than amylose containing around 95% α-(1→4) and 5% α-(1→6) bonds. The branch chains of amylopectin are relatively short with 20 – 25 glucose units (Gaillard *et al.*, 1987; Tester *et al.*, 2005).

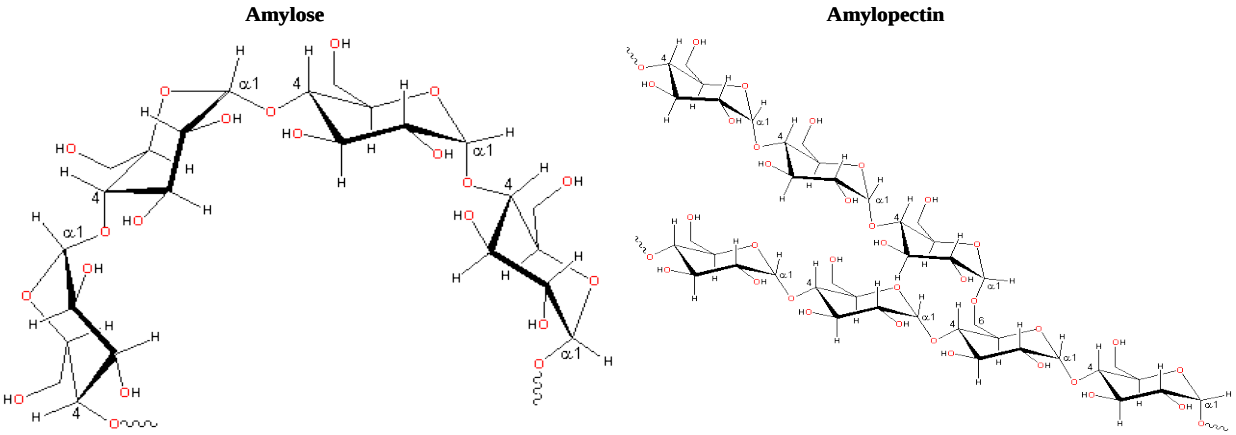



**Figure 1.** The structure of segments of amylose and amylopectin (Chaplin 2012).

The 3D-structures of α-amylases are closely related despite even large differences in their primary structure (Janecek 1997). In spite of these differences four highly conserved regions have been found by comparing amino acid sequences of different *Bacillus* α-amylases. These four regions are all related to the catalytic and substrate binding sites of the enzyme (Kuriki *et al.*, 1999; Nakajima *et al.*, 1986; Svensson 1988). The structure of α-amylases is divided into the three main domains: A, B and C. Domain A consists of an N-terminal α/β barrel which forms the core of the enzyme. This domain also contains the three catalytic site residues, two Asp and a Glu. The other two domains, B and C, are located on each side of domain A. Domain B is a loop domain and varies most between α-amylase structures (Janecek *et al.*, 1997). It connects β-strand 3 with α-helix 3 of domain A thus forming a long loop. Domain A and domain B form the substrate binding site. Domain C is a greek key motif comprising the C-terminal part of the sequence (Janecek 1997; Nielsen *et al.*, 2000; Svensson 1994). The different α-amylases can contain a variety of metal ions, but almost all have at least one bound calcium ion. The conserved $Ca^{2+}$ is interfaced between domain A and B, hereby stabilising the structure around the active site making the ion important for proper enzyme activity (Nielsen *et al.*, 2000; Valle *et al.*, 1959).

The substrate binds in the substrate binding cleft located between domain A and domain B. In different α-amylases the cleft can bind between four and ten glucose units where each unit is bound by the amino acid residues constituting the binding subsite for that glucose unit (MacGregor *et al.*, 2001; Nielsen *et al.*, 2000). The subsites are named according to the location of the site to the cleavage bond with negative nomenclature on the non-reducing side of the cleavage bond and positive on the other, so the glucose units on each side of the bond to be cleaved are in subsites -1 and +1, respectively (Davies *et al.*, 1997). There can be two to three subsites towards the substrate reducing end from the cleavage bond (+1 to +3) whereas the number of subsites towards the non reducing end of the substrate can vary between 2 and 7 (MacGregor *et al.*, 2001; Nielsen *et al.*, 2000).

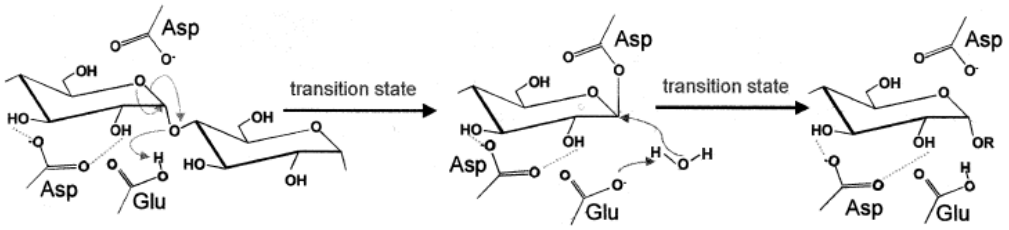



**Figure 2.** The catalytic mechanism for α-amylases. Glu protonates the glucosidic oxygen followed by the first Asp residue acting as catalytic nucleophile attacking C1 of the glucosidic residue, hereby forming the first transition state leading to a covalent intermediate. The protonated glucosidic unit leaves the active site and water molecules enter the active site. Activation of the water molecule results in the second transition state where the covalent bond between C1 and Asp is cleaved. The second Asp residue in the active site stabilises the reaction intermediates by hydrogen bonding (van der Maarel *et al.*, 2002).

The catalytic mechanism for α-amylases, Figure 2, is proposed to first apply the Glu acting as a proton donor in the protonation of the glycosidic oxygen. The first Asp residue then acts as a catalytic nucleophile attacking C1 of the glucose residue in subsite -1, hereby forming the first transition state leading to a covalent intermediate. The protonated glucose unit in subsite +1

leaves the active site and a water molecule is proposed to move into this site and be activated by the deprotonated Glu. The water molecule then hydrolyses the covalent bond formed between the oxygen of the Asp residue and the C1 of the sugar residue, to complete the catalytic mechanism. The role of the third conserved residue, the second Asp in the catalytic mechanism is presumed to be involved in the stabilisation of the reaction intermediate through hydrogen bonds to –OH groups on the substrate which makes it very important for the enzyme activity (Kuriki *et al.*, 1999; Nielsen *et al.*, 2000; van der Maarel *et al.*, 2002).

Structural studies of amylases have revealed that non-catalytic binding of carbohydrates also can occur in secondary binding sites (SBS) (Brzozowski *et al.*, 2000; Dauter *et al.*, 1999; Lyhne-Iversen *et al.*, 2006; Ragunath *et al.*, 2008; Robert *et al.*, 2003). These SBS are generally located on the surface in a certain distance from the substrate binding cleft and the catalytic site. The amino acid residues involved in SBS are not conserved among different α-amylases. However, they are often observed in the same regions on the surface of α-amylases. SBS generally contain amino acids with aromatic side chains such as Trp and Tyr which are able to hydrophobically interact with carbohydrates. The exact role of these sites is not fully elucidated and since the sites can be located in all three domains of the α-amylases it has been proposed that different sites play different roles such as cooperation with the active site in degradation of different substrate, as initial substrate recognition sites or to retaining reaction product for further processing by other proteins or for import to bacterium/cell (Cuyvers *et al.*, 2011; Janecek *et al.*, 1997; Nielsen *et al.*, 2008; Nielsen *et al.*, 2009).

### *1.1.1* Bacillus *amylases*

*Bacillus* α-amylases are widely used in industrial processes as they generally are thermo stable and/or stable under alkaline conditions, the latter being an important property in the detergent industry. The main focus of this project is on the amylases from *Bacillus* sp. working with the following: *Bacillus halmapalus* α-amylase (BHA), alkaline *Bacillus* sp. α-amylase (AA560), *Bacillus licheniformis* α-amylase (BLA) and *Bacillus amyloliquefaciens* α-amylase (BAN) where the structural analysis will be performed on BLA. The amino acid sequence of these *Bacillus* amylases differs but their overall tertiary structures are very similar.

*B. licheniformis* produces at least two α-amylases with different amino acid sequences and action patterns where BLA is very thermostable despite the bacterium itself is mesophilic (Kim *et al.,* 1992; Yuuki *et al.,* 1985). BLA consists of 483 amino acids with the molecular weight (Mw) of 55239 Da, is stable between pH 6 and 11 and has a temperature optimum between 60 and 105ºC depending on pH (Fitter *et al.,* 2001; Tomazic *et al.,* 1988b; Vihinen *et al.,* 1989; Yuuki *et al.,* 1985).

The structure of BLA follows the general features for α-amylases with domain A as the most conserved containing the active site and the conserved calcium binding site, Figure 3. Domain B is the least conserved domain and it is established as a protrusion from domain A. The main characteristics of domain B in BLA are two long β-strands which wind around each other in a double helical manner. The B domain also contains a distinct loop which wind around a $Ca^{2+}$-$Na^{+}$- $Ca^{2+}$ metal triad. This triad is quite unique as the conserved calcium binding site is a part of

this metal binding motif. The three metal ions mainly situated in domain B are linearly arranged and involved in 20 interactions with the surrounding amino acid residues and some water molecules. BLA also contains a third Ca-ion interfaced between domain A and C. This Ca-ion is believed to be a stabilising extra anchor point between the two domains. There is also found indications of a chloride binding site in BLA, but this has not been confirmed by 3D structure analysis (Machius *et al.*, 1995; Machius *et al.*, 1998).

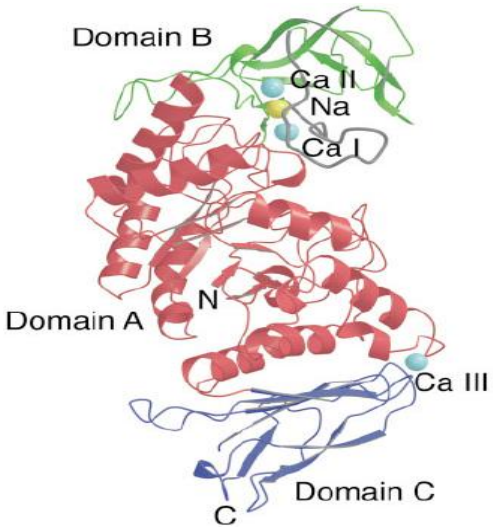



**Figure 3.** The structure of BLA. Red colour shows domain A, green domain B and blue domain C. The grey colour is the region of domain B which in Ca depleted structure is disordered. Calcium ions are shown in cyan and sodium in yellow (Machius *et al.*, 1998).

The importance of calcium has been investigated by denaturing BLA with and without $Ca^{2+}$-ions showing a much higher resistance to denaturation in the presence of calcium (Nazmi *et al.*, 2006). The importance of the metal triad for stability has also been identified by mutating amino acids in this site, which resulted in loss of stability (Declerck *et al.*, 2000; Declerck *et al.*, 2003). The role of the third $Ca^{2+}$-ion in the stability is not elucidated. The triad, however, cannot completely account for the high thermostability since other *Bacillus* α-amylases are believed to contain the same triad along with the third $Ca^{2+}$-ion, without exhibiting the high thermostability. Thus, other factors may also contribute to the stability of BLA (Declerck *et al.*, 2000; Machius *et al.*, 1998).

The three catalytic site residues in BLA are Asp231, Glu261 and Asp328 (Machius *et al.*, 1995) where Asp231 acts as the catalytic nucleophile while Asp 328 is the stabilising Asp residue. BLA contains several substrate binding sites on the surface for interaction with the polymeric substrate. One of these binding sites is the active site cleft including the catalytic site; it consists of at least eight subsites each binding a glucosyl residue of the substrate (Kandra *et al.*, 2002; Machius *et al.*, 1998).

Chemical modifications of Lys residues on the surface of BLA have in different studies resulted in changes in kinetic parameters as well as in the substrate specificity. The importance of specific Lys residues were, however, not elucidated (Habibi *et al.*, 2004; Kandra *et al.*, 2002).

### *1.1.2 Production of industrial amylases*

Industrial enzymes are produced by fermentation of genetically modified organisms such as bacteria or fungi. Multiple gene copies of the genetically optimised enzyme with improved performance are inserted into the organism's genome to increase expression of the gene and hereby the enzyme yield.

The amylases for detergent are generally produced by fed-batch fermentation of *Bacillus* strains. The organism used for the production of the amylases are first grown in an inoculation tank and is then transferred to the fermentation tank where sterilised nutrients are added to the fermentor during the growth of the biomass. The nutrients are based on renewable raw materials such as partially hydrolysed corn starch, sugars and soy grit. Additionally inorganic salts are added. The pH, temperature, oxygen consumption and $CO_2$ formation are carefully controlled during the fermentation to optimise the production process (Olsen *et al.*, 2008).

The recovery of the amylases from the fermentation medium consists of several steps where the first include flocculation. The enzymes are then separated from the biomass and other insoluble materials by drum filtration followed by a fine centrifugation or filtration step removing the small amounts of sludge remaining after the drum filtration. At last the enzyme solution is concentrated by ultrafiltration (Olsen *et al.*, 2008).

### *1.1.3 α-Amylases in detergent*

In the detergent industry α-amylases are the second most used enzyme after proteases. They are used to remove starch containing stains from for instance pasta, fruit, chocolate and gravy. Starch is also commonly used in processed foods as a thickening agent. Gelatinised starch adheres to the surface of textiles where it can bind other stain components. Detergent amylases are able to degrade the gelatinised starch into soluble dextrins and oligosaccharides, hereby removing the starch stain and the stain components bound to it (Eriksen 1996; Kirk *et al.*, 2002; Olsen *et al.*, 2008).

There are two main criteria for enzymes in detergent which is their stability in the detergent with all the different components and their activity at the wash conditions. The chemical properties of detergents can vary considerable though detergents can be divided into two main groups, liquid and powder. These two detergent types differ in wash pH as the former has a pH around pH 8 and the latter a pH around 10. The trend within laundry wash in Europe has moved toward lower temperatures to save energy with 30 – 40°C being standard wash temperatures. Though, the goal is to lower the temperature even further. Some detergent components have a negative effect on the enzymes stability and activity by interacting with the protein structure or the active site (Eriksen 1996; Kirk *et al.*, 2002; Olsen *et al.*, 2008). Especially detergents ingredients such as chelators and surfactants influence the stability (Lund *et al.*, 2012). Therefore, a good detergent amylase is stable and active at neutral to alkaline pH, in the presence of other detergent ingredients and in the temperature range 20 – 60°C.

## 1.2 Post-translational modifications

Post-translational modifications (PTM) on proteins can occur either as enzymatically induced reactions or as chemical reactions. During the production of recombinant proteins these can undergo non-enzymatic PTM. These reactions can influence the enzyme stability and activity when they occur on amino acid residues involved in these factors. Some of the most prevalent non-enzymatic PTMs are deamidation, oxidation and glycations and these will be described in the following sections (Jenkins *et al.*, 2008; Manning *et al.*, 2010).

### *1.2.1 Glycation*

Glycation is a non-enzymatic reaction between primary amines on proteins and the carbonyl group of a reducing sugar also known as a Maillard reaction (Maillard 1912). Since the reaction occurs at amino groups, the N-terminal ends (α-amino group) of proteins along with lysine (ε-amino groups) and arginine (guanidino group) side chains are prone to the reaction. The glycation forms a ketoamine also known as the Amadori compound which can undergo a series of rearrangement reactions to form a group of more reactive carbonyl compounds termed advanced glycation end products (AGEs), Figure 4 (Baynes *et al.*, 1989; Brownlee *et al.*, 1988).

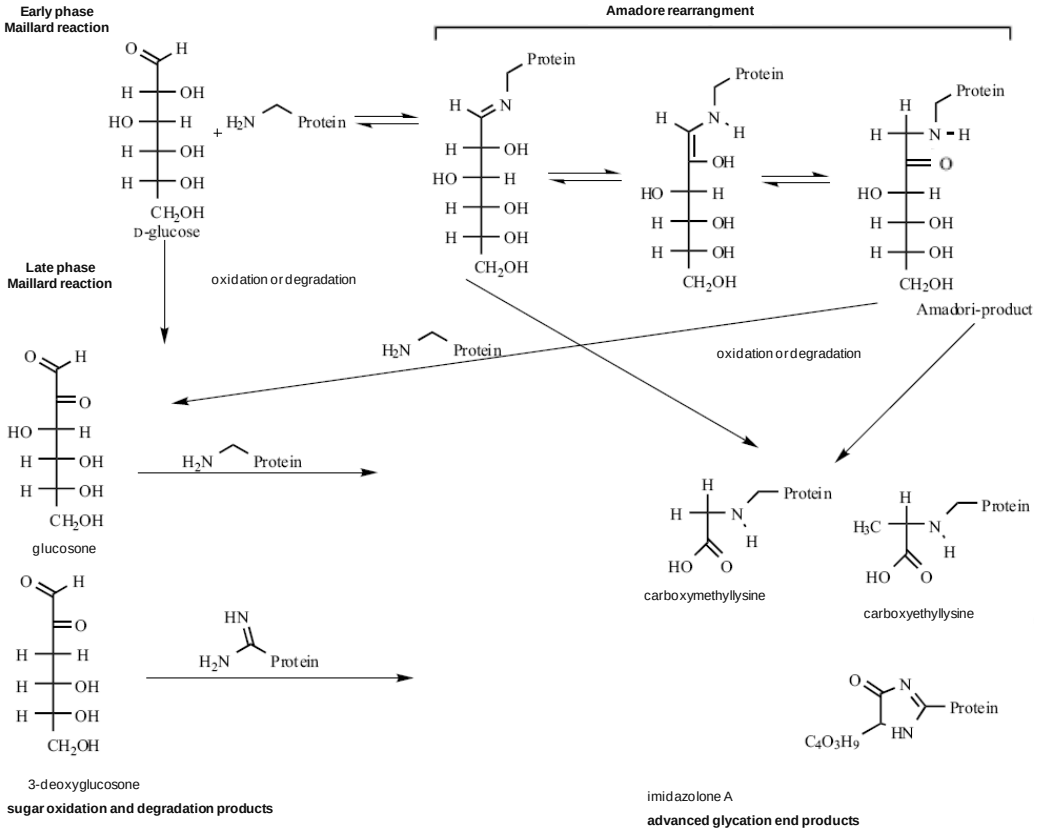



**Figure 4.** The reaction pathway for the Maillard reaction. An amino group on a protein reacts with the reducing sugar and undergoes Amadori rearrangement forming an Amadori product. This can then undergo several different reactions leading to the formation of different advanced glycation end products (AGEs) (Kislinger *et al.*, 2004).

The Maillard reaction can be divided into three stages where the first stage is the formation of a covalent bond between the amino group and the aldehyde or keto group of the reducing sugar

resulting in a Schiff base, Figure 4. Rearrangements then form the early glycation Amadori product (ketoamine). In the second stage the Amadori product can be degraded into a variety of reactive carbonyl compounds most often deoxyglucosones formed through dehydration. In the last stage the reactive carbonyl compounds leads to the formation of AGEs. These are formed by reaction of the carbonyl compounds with free amino groups and through rearrangements. So far several glycation products have been identified, but from model reactions several other structures have been proposed (Horvat *et al.*, 2004; Kislinger *et al.*, 2004; Martins *et al.*, 2000). These AGEs include 3-deoxyglucosone (144 Da), carboxymethyl (58 Da), carboxyethyl (72 Da), imidazolone A (144 Da) and methylglyoxal-derived hydroimidazolone (MG-HI, 54 Da) as some of the most described (Ames 2008; Barnaby *et al.*, 2011; Kislinger *et al.*, 2004; Kislinger *et al.*, 2005).

The Maillard reactions play an important role in food chemistry as it is a source of colour and flavour in processed food but the reactions are also suspected to be involved in the development of a number of different diseases (Priego-Capote *et al.*, 2009; Horvat *et al.*, 2004; Martins *et al.*, 2000). AGEs are believed to be involved in the development of diabetic complications (Ahmed 2005; Thorpe *et al.*, 1996) and they are suspected to play a role in other diseases for instance neurodegenerative diseases such as Alzheimer's and Parkinson's disease (Munch *et al.*, 1998; Nicolls 2004). However, glycation also occur during the production and storage of recombinant proteins. When the proteins are secreted to the cell culture medium during the fermentation they are exposed to reducing sugars. The proteins can in some formulations also be exposed to reactive sugars. This can result in conformational changes, instability, changes in biological properties of proteins and loss of enzyme activity (Bruins *et al.*, 2003; Chobert *et al.*, 2006; Fischer *et al.*, 2008; Li *et al.*, 1996; Quan *et al.*, 2008).

The Maillard reaction is highly dependent on temperature, pH, reactants (type of sugars and amine) and the concentration of the reducing sugars. The initial glycation is favoured by alkaline conditions and the rate of reaction increases with temperature. The reactions occurring in the second and third stages of the Maillard reaction also seem to be influenced by pH. Different products will form dependent on whether the conditions are acidic, neutral or alkaline (Martins *et al.*, 2000; Quan *et al.*, 2008). There are indications that properties of nearby amino acids influence which Lys residues undergo glycation (Baynes *et al.*, 1989; Johansen *et al.*, 2006).

### *1.2.2 Deamidation*

Deamidation is one of the most prevalent non-enzymatic reactions and it can have an influence on protein stability and activity. The amino acid most prone to deamidation is asparagine resulting in the formation of either isoaspartate or aspartate. Deamidation occurs in two steps, Figure 5, where the first is a nucleophilic attack by the α-amino group of Asn forming a cyclic imide. The imide intermediate then undergoes hydrolysis forming either isoaspartate or aspartate. The formation of either of these two causes a change in pI for the protein as the amino acid side chain changes from neutral to acidic (Aswad *et al.*, 2000; Manning *et al.*, 1989; Powell *et al.*, 2007).

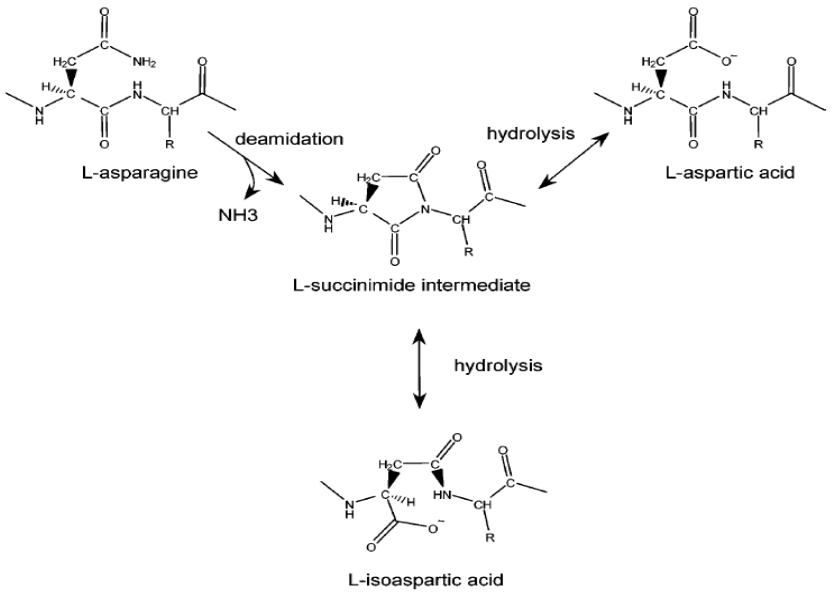



**Figure 5.** Deamidation of asparagine is caused by the nucleophilic attack by the α-amino group, releasing ammonia. A succinimide intermediate is formed which then hydrolyses to either isoaspartate or aspartate, where isoaspartate formation dominates (Chelius *et al.*, 2005).

Glutamine residues can also undergo deamidation but this reaction is one hundred times slower than for asparagines. Gln is less susceptible since it is incapable of forming the same 5-membered cyclic imide intermediate. It can produce a six membered ring structure but this is less stable than the five membered counterpart (Joshi *et al.*, 2002; Manning *et al.*, 2010; Patel *et al.*, 1990).

Deamidation can occur at neutral and alkaline conditions and the reaction is accelerated by increasing temperature and pH. Deamidations can also occur at acidic conditions, however here the side chains are hydrolysed directly to the corresponding acid not forming the cyclic intermediate (Patel *et al.*, 1990; Stratton *et al.*, 2001). Asparagines on the surface of proteins are more exposed to deamidations. The amino acid sequence also influences the rate of deamidation as amino acids with smaller side chains after Asn residues increase the reaction rate, due to low steric hindrance leaving the amino group open for attack (Vieille *et al.*, 2001). Amino acids which can perform hydrogen bond-donation also seem to promote the reaction. Therefore, it has been observed that Asn-Gly is the most reactive sequence but also Asn followed by Ser or His are very reactive sites (Aswad *et al.*, 2000; Manning *et al.*, 2010; Robinson *et al.*, 2001; Robinson 2002). The secondary structure of α-helices and β-sheets tend to protect Asn residues against deamidation (Xie *et al.*, 1999).

### *1.2.3 Carbamylation and other PTMs*

Another PTM occurring with the primary amines on proteins is carbamylation. This spontaneous non-enzymatic reaction modifies Lys and Arg residues as well as the N-terminal on proteins. Urea can spontaneously form cyanate of which the reactive form isocyanic acid can irreversibly modify non-protonated amino groups, Figure 6. This reaction can change the charge and

function of amino acids and it has been found to alter enzyme activity and protein stability (Albert *et al.*, 2011; Harding *et al.*, 1989; Kraus *et al.*, 2001).

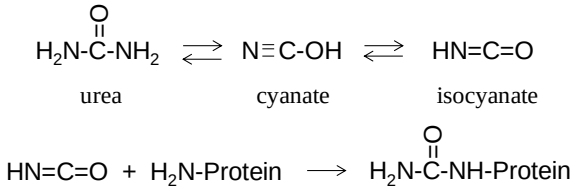



**Figure 6.** The spontaneous formation of cyanate and isocyanate from urea and the following carbamylation of proteins by isocyanate.

Acetylation is the addition of acetyl ($CH_3CO$) to amino groups of Lys residues and the N-terminal of proteins. This reaction is enzymatically controlled within cells and the modification is important for controlling the life span of proteins within the cell as non acetylated proteins quickly become degraded by intracellular proteases (Lodish *et al.*, 2004). This modification is therefore not the most likely modification to occur on *Bacillus* amylases as these are extracellular enzymes.

The side chains of Met, Cys, His, Trp and Tyr can undergo oxidation at any stage of the protein production and purification and this can lead to protein degradation. The oxidation rate is determined by the protein structure as a side chain buried in the structure can be partially protected, but also pH and temperature are important factors, though Met oxidation appears to be pH independent (Jenkins *et al.*, 2008; Manning *et al.*, 2010).

Protein truncation is the shortening of the protein by proteases resulting in a new N-terminal or C-terminal residue. Truncation may also arise from mutations leading to premature termination of mRNA translation. In this project protein truncation also includes the removal of 1 – 3 amino acids from the C-terminal end by carboxypeptidases. These enzymes are produced during the fermentation and released either as extracellular enzymes or due to cell lysis. This modification may affect the 3D structure and the activity of the enzyme. The pI value of the protein molecules are affected if some of the lost amino acids contribute to the pI of the full-length protein.

## 1.3 Mass spectrometry

MS is a method capable of determining mass per charge (m/z) from which the molecular weight of most molecules from small metal ions to proteins can be determined. MS is an invaluable tool within biochemistry as this highly sensitive method is able to identify proteins and their amino acid sequence in relatively short time. The main focus of this report is on the MS methods used to determine PTMs on proteins and the methods used to perform relative quantification.

### *1.3.1 Methods for protein characterisation*

Protein identification is possible with tandem mass spectrometry (MS/MS). In this method the protein is first digested into peptides by specific proteases such as the endoproteinases trypsin, Asp-N and Lys-C. The amino acid sequence of the peptides are then determined with MS/MS (Koomen *et al.*, 2005; Mann *et al.*, 2001; Pappin *et al.*, 1993).

In MS/MS peptides are fragmented into smaller amino acid sequences by collision with an inert gas or by reaction with electrons. The fragmentation occurs in the peptide backbone producing fragment ions which are designated, according to the Roepstorff-Folhmann-Biermann nomenclature, after the peptide bond which is fragmented, Figure 7 (Biemann 1992; Roepstorff *et al.*, 1984). For the most common fragmentation techniques the amide bonds are broken producing what is called b and y ions depending on which part of the broken peptide retains the charge (Barrett *et al.*, 2005; Mikesh *et al.*, 2006).

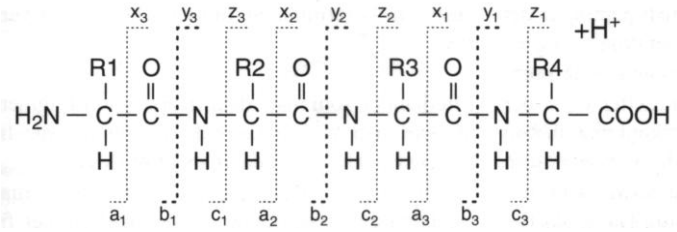



**Figure 7.** The different kind of ions obtained by MS/MS. The peptide is fragmented in the backbone. When the fragmentation occur in a peptide bond either y or b ions are obtained depending on which end is charged. If other bonds in the peptide is broken a and x ions or c and z ions are obtained (Hjernø *et al.*, 2005).

The amino acid sequence of the peptides is determined by calculating the differences between peaks and finding those correlating to amino acid masses (Roepstorff *et al.*, 1984). However, as MS/MS spectra often also contains other than b and y ions (a and x, c and z) and they can be very complex and difficult to analyse. Therefore, protein databases are most often used to analyse MS/MS spectra.

The datasets obtained from MS/MS are generally quite large and the individual spectrum can as mentioned previously be very complex so protein and peptide identifications are obtained with search engines applying search algorithms such as MASCOT or SEQUEST to search protein databases. The search engines identify proteins from the MS/MS data by comparing the masses of the peptides and the peptide fragments with theoretical masses or spectra of protein sequences in the applied database. Various algorithms and statistical tools are used to evaluate the

significance of the identifications and provide a probability score for the identification (Cramer *et al.*, 2005; Steen *et al.*, 2004).

### *1.3.2 Mass spectrometer*

A mass spectrometer consists of an inlet for the sample, an ion source, mass analyser, detector and a data processing system. The system works by first converting the sample ions into gas phase. This is done by an ion source, from here the ions are accelerated into the analyser where they are separated according to m/z. The detector then collects the signals from the ions and a data processing system converts these signals into a mass spectrum (Gross 2011; Hoffmann *et al.*, 2008a).

In MS of proteins and peptides the ion source used has to be able to perform soft ionisation to avoid excessive fragmentation of these molecules during ionisation. Two types of ion sources are able to do this, the MALDI (Matrix-Assisted Laser Desorption Ionisation) and the ESI (Electrospray Ionisation) (Domon *et al.*, 2006; Koomen *et al.*, 2005). In this project the mass spectrometers used are equipped with ESI ion source.

In the ESI ion source developed by Fenn an co-workers (1989) a liquid sample is passed through a metallic capillary tube with a strong electrical field under atmospheric pressure. This causes an accumulation of charge on the surface of the liquid which then will disperse into small highly ionised droplets. The solvent are removed by exposing the droplets to a counter current of heated inert gas most often nitrogen. In this way the molecules become charged and are hereafter led into the analyser. In ESI the peptides can be doubly or multiply charged (though they are mostly doubly charged) which makes the ESI spectrum complex (Gross 2011; Hoffmann *et al.*, 2008b).

ESI can be run in a positive and a negative mode. In positive ESI mode the positively charged ions are detected and in negative ESI mode negative ions are detected. In positive mode the positive charge on droplets is generated by removing the negative charges through electrochemical discharge of negative ions on the metal wall of the capillary tube. In negative ESI mode the positive charges are removed by the same principles (Bruins 1998).

As the ESI method is based on a liquid sample this ion source can be couple to liquid chromatography (LC) and many different types of mass analysers. This makes it possible to separate complex mixtures in the liquid phase before the ESI-MS analysis (Gross 2011; Hoffmann *et al.*, 2008b). The development of the nanoESI has made it possible to analyse much smaller volumes and hereby smaller amounts of analyst (Wilm *et al.*, 1996).

#### *1.3.2.1 Orbitrap and ion trap analysers*

The linear (or quadrupole) ion trap (LIT) analyser is able to perform both MS and MS/MS in the same analyser, as it is able to scan, select and fragment peptides. The LIT separates in time as it traps the gaseous ions between trapping plates using radiofrequency voltages. Changes in amplitude or frequencies make it possible to select specific ions for fragmentation or push ions out of the trap for detection. These analysers are sensitive, quick, have a high trapping capacity

and operate with low kinetic energy which results in higher MS/MS efficiency but it generally runs at lower resolution and mass accuracy (Glish *et al.*, 2008; Gross 2011; Hager 2002).

Another trap analyser is the orbitrap which incorporates another concept where the ions are trapped in an electrostatic field. The gaseous ions are moving in spirals around a central electrode. The orbitrap is not able to perform MS/MS, as it needs a proper ion injection and ultra high vacuum to perform properly. Therefore a radiofrequency voltage C-trap is used to accumulate, store and thermalize ions for injection into the orbitrap by a low pressure of nitrogen. The orbitrap operates at high resolution with a high mass accuracy but the kinetic energy is high which can result in reduced MS/MS efficiency (Gross 2011; Makarov *et al.*, 2006; Makarov *et al.*, 2010).

The LIT and orbitrap analysers can be combined in hybrid MS instruments such as the LTQ-Orbitrap instruments hereby combining the advantages of the two analysers such as the mass accuracy of the orbitrap with the speed and high MS/MS efficiency of the LIT. In the hybrid instrument ions can be moved after accumulation between the two analysers through a C-trap which moves packets of ions between the two traps (Glish *et al.*, 2008; Gross 2011; Makarov *et al.*, 2006). For this project the two hybrid instruments LTQ Orbitrap XL and LTQ Orbitrap Velos (both from Thermo Scientific) were available for sample analysis. The LTQ Orbitrap Velos has higher isolation and fragmentation efficiency compared to the XL as well as better resolving power and mass accuracy. The LTQ orbitrap XL was, however, upgraded with the fragmentation method electron transfer dissociation (see section 1.3.2.2).

#### *1.3.2.2 Fragmentation*

Numerous different fragmentation methods exist, however as only orbitrap hybrid instruments was used in this project the focus of this section is on fragmentation methods used in these instruments.

The most common method used for peptide fragmentation is low-energy collision induced dissociation (CID) (Mikesh *et al.*, 2006). In CID the peptides selected for MS/MS are fragmented by interaction with a collision gas most often nitrogen or argon. As peptides are most vulnerable in their amide bonds these are generally the ones broken in the collision producing b and y ions, but other bonds can also be broken. CID is most effective on short low charged peptides (Syka *et al.*, 2004; Wysocki *et al.*, 2000). CID is sometimes not hard enough to achieve sufficient fragmentation in the LIT of the hybrid instruments so a harsher CID can be obtained in the C-trap by raising the radiofrequency voltage. This method termed Higher-energy C-trap Dissociation (HCD) applies the nitrogen already present in the C-trap as collision gas. The most optimal HCD fragmentation with a wide m/z range is obtained by attaching a collision cell at the end of the C-trap (Olsen *et al.*, 2007; Scherl *et al.*, 2008). In the fragmentation of peptides with labile PTMs CID and HCD often cleave off the modification leaving the peptide backbone intact. This can result in a MS/MS spectrum devoid of consecutive peptide backbone cleavage product ions and as a consequence in poor peptide sequence assignment (Mann *et al.*, 2003; Mirgorodskaya *et al.*, 2001).

The fragmentation method Electron Capture Dissociation (ECD) (Zubarev *et al.*, 1998) is a soft fragmentation method where low energy electrons react with peptide cations in the magnetic field of Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers. This induces random breakage of the peptide backbone producing mainly c and z ions, Figure 7, while leaving labile modifications intact. However, ECD is only compatible with FT-ICR the most expensive MS instrument (Boyne *et al.*, 2006; Kjeldsen *et al.*, 2003; Zubarev 2004).

Electron transfer dissociation (ETD) is an alternative soft fragmentation method that can be used in combination with trap analysers. In ETD an electron from a radical anion, typically fluoranthene, is transferred to a multiple charged peptide hereby causing cleavage of the $C_\alpha$ – N bond in the peptide backbone producing mainly c- and z-ions but y-ions can also be produced. This soft fragmentation method is very suitable for characterisation of peptides containing PTM as it leaves labile modifications intact allowing for direct mapping of peptide modifications (Chi *et al.*, 2007; Hogan *et al.*, 2005; Kjeldsen *et al.*, 2007; Mikesh *et al.*, 2006; Syka *et al.*, 2004). However, ETD has limited application on doubly charged peptides and is most effective on triply or above charged ions, it also performs best on ions below 850 m/z (Good *et al.*, 2007; Swaney *et al.*, 2007).

### *1.3.3 Identification of post translational modifications by MS*

PTMs on proteins can be identified by determining the intact mass of the protein and compare this to the theoretical mass hereby determining the extent and mass of the PTM. The modified amino acids can be identified by MS/MS on the digested protein. In theory all PTMs can be identified by MS providing they lead to a reduction or increase in mass. However, PTM determination is not always as simple since the modified protein can comprise a very small fraction of the sample and therefore be difficult to detect, the modifications can be labile and easily lost in the fragmentation or the modification might also arise during sample preparation (Barnes *et al.*, 2007; Kamath *et al.*, 2011; Larsen *et al.*, 2006).

#### *1.3.3.1 Glycations*

Glycations can be studied by gel electrophoresis, chromatography and various assays but these methods do not provide information about the specific glycation site. MS by either ESI or MALDI are also used to study glycated peptides (Kislinger *et al.*, 2005; Lapolla *et al.*, 2005; Priego-Capote *et al.*, 2009). MS/MS is able to determine the specific glycation site on peptides. However, MS/MS of glycated peptides is difficult as the most common fragmentation techniques results in extensive patterns of neutral loss and loss of labile modifications such as glycations (Kjeldsen *et al.*, 2007; Lapolla *et al.*, 2004; Mikesh *et al.*, 2006; Priego-Capote *et al.*, 2009). A LC-MS/MS method has been developed on a quadropole-TOF (time of flight) instrument where the peptides are subjected to a weak fragmentation which results in specific neutral losses from glycated peptides (162 Da). Peptides experiencing this neutral loss are selected for further stronger fragmentation producing b- and y-ions hereby providing peptide sequence information (Gadgil *et al.*, 2007). Similar methods have been developed for ion trap instruments where product ions

showing neutral loss in $MS^2$ are selected and further fragmented in $MS^3$, hereby providing further sequence information (Zhang *et al.*, 2008b).

The soft fragmentation method ETD leaves labile modifications intact allowing for direct mapping of peptide modifications which makes it very suitable for characterisation of glycated peptides (Chi *et al.*, 2007; Hogan *et al.*, 2005; Kjeldsen *et al.*, 2007; Mikesh *et al.*, 2006; Syka *et al.*, 2004). As the fragmentation leaves the glycation intact the backbone fragmentation and peptide identification is greatly improved. It has been found that ETD fragmentation produces almost complete series of c- and z-ions without any neutral losses and it outcompetes CID in the identification of glycated peptides (Zhang *et al.*, 2007a; Zhang *et al.*, 2007b). As mentioned previously the method does, however, also have some limitations, as it is less sensitive than CID, most effective on triply or above charged ions and on ions below 850 m/z (Good *et al.*, 2007; Swaney *et al.*, 2007).

#### *1.3.3.2 Quantification of glycations*

Quantification has become an important tool in MS based proteomics and this has resulted in numerous different techniques for either absolute or relative quantification (Aebersold *et al.*, 2003; Bantscheff *et al.*, 2007; Ong *et al.*, 2005). However, only a few methods exist for relative quantification of glycations by MS and these are generally based on addition of internal standards of the corresponding isotopically labelled glycated peptides (Delatour *et al.*, 2009; Priego-Capote *et al.*, 2010).

Isobaric stable isotope tagging is a quantitative method allowing for relative quantifications of proteins or peptides between two or more samples (up to 8) by use of MS/MS. The isobaric tags are made from light and heavy isotopes, which mean that all the tags have the same total mass as well as the same physiochemical properties. The tags consist of a reporter group, a balance group and an amino reactive end which reacts with the amino groups of the N-terminus of peptides as well as the epsilon amino groups of lysine. Each sample is digested and labelled whereafter all samples are mixed and the differentially labelled peptides are selected for MS/MS as a single precursor ion, Figure 8A. Upon fragmentation the specific reporter ions for each sample are released and relative quantification is performed by determining the ratios between the reporter ion intensities. The MS/MS spectrum of the peptide is not altered by the tags and can therefore also be used to identify the peptide sequence. One of the advantages of this method is the low level of noise generally obtained in MS/MS (Dayon *et al.*, 2008; Ross *et al.*, 2004; Thompson *et al.*, 2003; Zhang *et al.*, 2005). This method does also have drawbacks for instance it relies on very high labelling efficiency for reliable quantification (Bantscheff *et al.*, 2007). The reporter ions have to exceed a certain intensity to perform consistent quantification. For instance the reporter ion intensities have to exceed 1000 and 100,000 for ETD and HCD respectively to achieve average errors below 10% for the quantifications. The reporter ion intensities can be improved by increasing the collision energy, however this results in reduced formation of suitable ions for peptide identification (Dayon *et al.*, 2010; Phanstiel *et al.*, 2008).

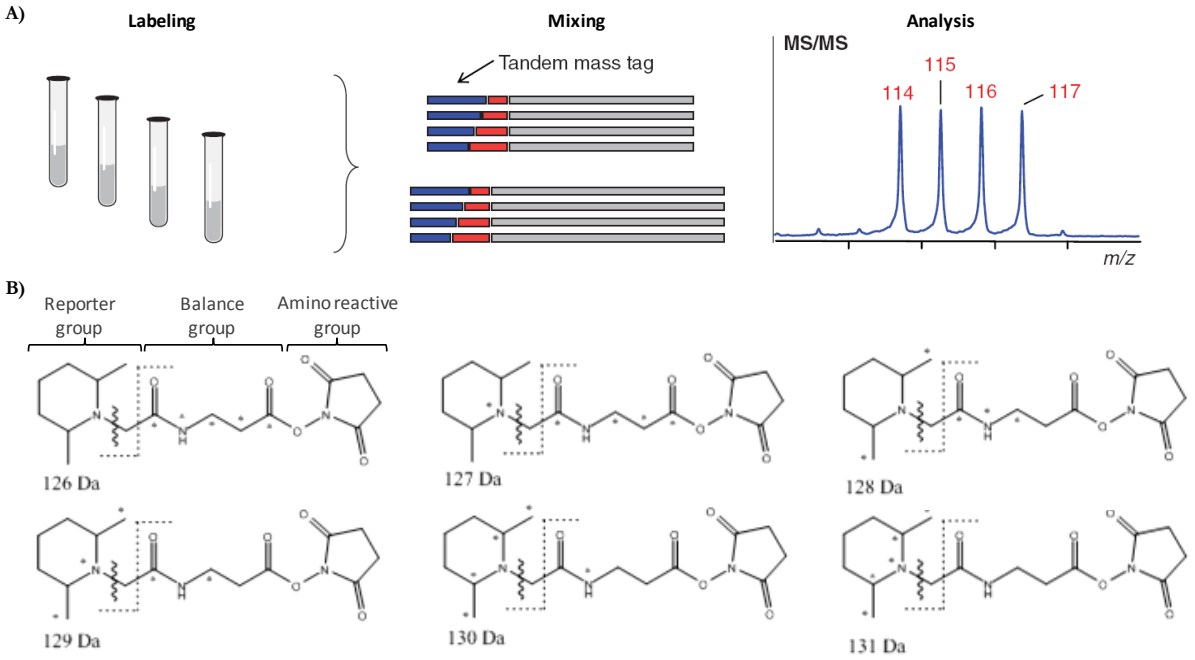



**Figure 8.** A) Quantification with isobaric stable isotope tags. The samples for quantification are prepared separately and labelled with individual tags. The labelled samples are mixed and the quantification is performed in MS/MS mode (Domon *et al.*, 2006). B) Chemical structure of the TMT 6-plex tags. The functional groups of the tags are illustrated on tag 126. Stars indicate isotope positions, the fragmentation sites are illustrated with dotted lines for CID and corrugated lines for ETD (Thermo Scientific 2012).

Two types of commercially available isobaric stable isotope tags exist; tandem mass tags (TMT) (Thompson *et al.*, 2003) allowing quantification of up to 6 samples and isotope tags for relative and absolute quantification (iTRAQ) (Ross *et al.*, 2004) quantifying up to 8 samples. A study of the commercially available tags revealed that they all perform equally in terms of quantitative precision and accuracy, but that the iTRAQ 4-plex identifies more proteins and peptides than TMT 6-plex and that iTRAQ 8-plex has the lowest identification rate (Pichler *et al.*, 2010).

It has been shown previously that ETD can be used for relative quantification with isobaric tags. However, the reporter ions produced by ETD differ from those produced by CID and similar fragmentation techniques due to the different fragmentation mechanism, producing c- and z-ions. This has resulted in some of the isobaric tags producing the same reporter ions so for instance the TMT 6-plex tags could previously only quantify up to four samples. However, very recently the position of some of the heavy isotopes have been changed for the TMT 6-plex tags so each tag now generates a unique reporter ion in ETD fragmentation (114, 115, 116, 117, 118 and 119 from the 126-131 TMT tags, respectively), Figure 8B (Han *et al.*, 2008; Phanstiel *et al.*, 2009; Phanstiel *et al.*, 2008; Thermo Scientific 2012; Viner *et al.*, 2009).

#### *1.3.3.3 Deamidations*

Deamidation of Asn and Gln creates a mass increase of 1 Da, which can be site specifically determined by MS/MS. Other methods for deamidation detection also exist such as enzymatically based detection assays, antibodies or Edman degradation. These methods do, however, have limitations such as being site unspecific, requiring highly specific antibodies or

requiring relatively high quantities of protein. Therefore, MS has become the preferred method for determination of deamidations as the method is fast, accurate, requires relatively small amounts of sample and is site specific (Li *et al.*, 2008; Powell *et al.*, 2007; Yang *et al.*, 2010).

Characterisation of deamidations by MS does, however, also pose some problems as the slightly increased temperatures and alkaline conditions during sample preparation accelerate deamidations (Chelius *et al.*, 2005). The rate of deamidations is also increased as deamidation rates are much higher on peptides than proteins as secondary and tertiary structure can protect against the reaction (Kosky *et al.*, 1999; Li *et al.*, 2008). Optimisation of the digest procedure has been proposed to minimize the deamidations occurring during sample preparation (Ren *et al.*, 2009). It has, however, also been observed that by incubation of protein in $H_2{}^{18}O$ the deamidations occurring resulted in a mass increase of 3 Da (Terashima *et al.*, 2007). So by performing the digest in $H_2{}^{18}O$ it is possible to differentiate between deamidations present prior to sample preparation (1 Da mass increase) from the ones occurring during sample preparation as these will result in a mass shift of 3 Da. The mass difference of 2 Da between the deamidations is due to the incorporation of $^{18}O$ from $H_2{}^{18}O$ (Gaza-Bulseco *et al.*, 2008; Li *et al.*, 2008). Trypsin digestion in $H_2{}^{18}O$ also results in the incorporation of one $^{18}O$ into the C-terminus resulting in a mass increase of 2 Da. Though, incorporation of two $^{18}O$ can also occur as the peptides can act as a pseudo substrate for some proteases such as trypsin, hereby incorporating a second $^{18}O$ in the C-terminus (Schnolzer *et al.*, 1996). Incorporation of two $^{18}O$ does create more complex isotopic patterns in the MS spectrum. Though by performing the sample preparation in $H_2{}^{18}O$ it is possible to differentiate between deamidations occurring during sample preparation from those occurring prior to sample preparation (Gaza-Bulseco *et al.*, 2008; Li *et al.*, 2008).

## 1.4 Objectives

α-Amylases are among the most important industrial enzymes as they are used in a number of very different industries. The commercial α-amylases are produced by fermentations of genetically modified bacteria, where both the native gene has been manipulated to code for an enzyme with improved performance characteristics and a number of gene copies have been inserted in the genome of the cell to allow expression in high yields. Although the genome only contains copies of the same α-amylase gene, multiple different α-amylase forms can be identified when analysing the final enzyme product, hereby indicating that the protein is modified.

It is the aim of this PhD project to identify the differences between the multiple forms of α-amylase derived from the same gene and determine if or how these various modifications affect the stability and performance during application. The main focus will be on α-amylases from various *Bacillus* species used in the detergent industry.

It is to be determined where in the production (fermentation and recovery) the different α-amylase forms are generated for a few selected α-amylases. The multiple forms will be isolated and the effect on wash performance will be determined for the α-amylases. The characterisation will be more thorough for one of the amylases (from *B. licheniformis*) also determining kinetic parameters on different substrates.

The nature of the modifications on the α-amylase is identified along with the site of modification by MS. The conditions during production are favourable for modifications such as deamidations and glycations. Glycations, especially in low amounts, can be difficult to detect by MS. Therefore, MS methods for determination of these modifications are set up using a glycated model protein in this case barley α-amylase/subtilisin inhibitor (BASI).

The influence of specific modifications at specific sites is evaluated by studying structural models of the α-amylases. The hypotheses for the most interesting sites are examined by characterisation of prepared enzyme variants to prove the functional roles of the specific amino acids in the α-amylases, and thus validate the role of specific modified sites in the function and stability of the enzyme.

# 2 Experimental work

## 2.1 α-Amylase

BLA, BHA, BAN and AA560 were provided by Novozymes. BLA, BHA and AA560 are expressed in recombinant *B. licheniformis* while BAN is expressed in *B. amyloliquefaciens*. The fermentation is performed as fed batch with soy protein and various carbohydrates such as degraded starch, maltose, or sucrose as substrates. The recovery occurs under alkaline conditions with pH above 9. It consists of flocculation, filtration and concentration by ultrafiltration, and hereafter the amylases are stabilised in 30% 1,2-propane diol at pH 5.5. The amylases for this project were all fermented and recovered in 2009.

## 2.2 Barley α-amylase/subtilisin inhibitor

The BASI used in these experiments was expressed in *Pichia pastoris* fermented and recovered by Novozymes in June 2009 (Micheelsen *et al.*, 2008).

## 2.3 Isoelectric focusing (IEF)

Different IEF gels were studied to find the gels with the best separation and reproducibility. The gels CleanGel IEF (GE Healthcare), Ampholine™ PAGplate pH 3.5–9.5 (out of production, GE Healthcare), Novex IEF gel (Invitrogen) and FocusGel (Gelcompany) were tested. The different gels were run according to the manufacturers' protocols. The FocusGels gave the best separation with clear bands and the results were reproducible, therefore these gels (Focusgel 4–6, 3–7 and 3–10) were used with Amersham Isoelectric Focusing Calibration Kit (GE Healthcare) broad range (pH 3–10) or low range (pH 2.5–6.5) as standards.

## 2.4 SDS-PAGE

NuPAGE 4–12% Bis-Tris gradient gel (Invitrogen) was used for SDS-PAGE. Samples were prepared and the gel ran according to the manufacturer's protocol using the described solutions from Invitrogen. For standards Precision Plus Protein Standard Kaleidoscope (BIO-RAD) was used.

## 2.5 Separation of multiple forms of BLA

For the separation of the multiple forms of BLA several different chromatography techniques were investigated; chromatofocusing, boronate affinity chromatography, cyclodextrin affinity chromatography and ion exchange chromatography. The technique which resulted in the best separation was anion exchange chromatography eluting with pH gradient followed by NaCl elution. All separation was performed on an Äkta-FPLC chromatographic system (Amersham Biosciences).

BLA was loaded on a Resource Q column 6 ml (GE healthcare) through a 2 ml loop with flow 6 ml/min washing with 1 column volume (CV) wash buffer (20 mM Tris, 1 mM $CaCl_2$, pH 8.6) and eluting with 0–100% buffer B1 (33 times diluted polybyffer 96 (GE Healthcare), 1 mM

$CaCl_2$, pH 5.6) for 110 min followed by 0–60% buffer B2 (33 times diluted polybuffer 96, 0.5 M NaCl, 1 mM $CaCl_2$, pH 5.6) for 25 min. The BLA sample loaded to the column was approximately 27 mg/ml in wash buffer. Buffer change was performed on a NAP-25 column Sephadex G-25 DNA grade (GE Healthcare) according to manufacturers' protocol. Eluted fractions were analysed by IEF gel and similar fractions were pooled, concentrated and buffer changed to 20 mM HEPES, 1 mM $CaCl_2$, pH 7.0 on Vivaspin 20 10,000 MWCO PES (Sartorius Stedim).

### 2.6 Separation of multiple forms of AA560

Separation was performed on an Äkta-FPLC chromatographic system (Amersham Biosciences). AA560 was loaded on a Resource Q column 1 ml (GE healthcare) through a 2 ml loop with flow 1 ml/min washing with 6 CV wash buffer (20 mM ethanolamine pH 9.9) and eluting with 0–80% buffer B1 (11 times diluted polybyffer 96 (GE Healthcare), 1 mM $CaCl_2$, pH 7.3) for 90 min followed by 100% buffer B2 (11 times diluted polybyffer 96, 0.5 M NaCl, 1 mM $CaCl_2$, pH 7.3) for 20 min. Approximately 10 mg AA560 in wash buffer was loaded to the column where the buffer change was performed on a NAP-25 column sephadex G-25 DNA grade (GE Healthcare) according to the manufacturer's protocol. Eluted fractions were analysed by IEF and similar fractions were pooled, concentrated and buffer changed to 20 mM HEPES 1 mM $CaCl_2$ pH 7.0 on Vivaspin 20 10,000 MWCO PES (Sartorius Stedim).

### 2.7 Separation of multiple forms of BHA

Separation was performed, on an Äkta-FPLC chromatographic system (Amersham Biosciences). BHA was loaded on a Resource Q column 1 ml (GE healthcare) through a 2 ml loop with flow 1 ml/min washing with 5 CV wash buffer (20 mM Bis-Tris, 1 mM $CaCl_2$, pH 7.0) and eluting with 0–100% buffer B1 (20 times diluted polybyffer 74 (GE Healthcare), 2 mM $CaCl_2$, pH 5.0) for 125 min followed by 100% buffer B2 (20 times dilute polybyffer 74, 0.5 M NaCl, 2 mM $CaCl_2$, pH 5.0) for 15 min. Approximately 14 mg BHA in wash buffer was loaded to the column where the buffer change was performed on a NAP-25 column Sephadex G-25 DNA grade (GE Healthcare) according to manufactures protocol. Eluted fractions were analysed by IEF and similar fractions were pooled, concentrated and buffer changed to 20 mM HEPES, 1 mM $CaCl_2$, pH 7.0 on Vivaspin 20 10,000 MWCO PES (Sartorius Stedim).

### 2.8 Capillary isoelectric focusing (cIEF)

CIEF was performed by Samuel Fox (application support specialist) at Beckman Coulter (High Wycombe, UK) on a PA800 Plus (Beckman Coulter). Samples were prepared and run following the chemical mobilization method according to the application guide. Samples were analysed as duplicates. For analysis samples undergo buffer exchange to 20 mM Tris pH 8.0 with 30 kDa MWCO filters with a final concentration of approximately 10 mg/ml. cIEF were performed in pI range 3–10 with three standards at pI values 4.1, 9.5 and 10.

### 2.9 Amino acid analysis

Amino acid analysis was performed by Anne Blicher, Enzyme Protein and Chemistry, Department of Systems Biology, DTU. Samples were hydrolysed in 6 M HCl, oxidised and derivatized with orto-phtalaldehyde and injected for ion exchange chromatography. Each amino acid was quantified and identified according to 1 nmol of standard amino acid mixture (Barkholt *et al.*, 1989; van Wandelen *et al.*, 1997). Double determination was performed for all samples.

### 2.10 Laundry wash

Wash experiments was performed at pH 7.5 and 10 in 50 mM $Na_2CO_3$ buffer at 30 or 50°C in duplicates. Each wash was performed in 20 ml buffer under magnetic stir with 3 CS-28 (coloured rice starch cotton, Center For Testmaterials BV) patches of 2 cm diameter. The wash ran 20 min with enzyme concentration of 0.15 mg/l (determined by A280) and afterwards the patches were rinsed 5 min in cold tap water and dried over night. The colour remission was measured at 460 nm in a Macbeth COLOR-EYE 7000 spectrophotometer (Macbeth) subtracting the blank value.

### 2.11 Differential scanning caliometry

Differential scanning calorimetry (DSC) measurements were carried out using a Microcal VP-capillary DSC (GE Healthcare). Scan rate of 1.5°C $min^{-1}$ was used to a final temperature of 130°C. Samples for analysis were approximately 1 mg/ml in 25 mM HEPES, 1 mM $CaCl_2$, pH 7.0 and 400 µl sample including a reference with the HEPES buffer was added to a 96 well plate. The plate was placed in the thermostated auto sampler compartment kept at 10°C. $T_{max}$ was determined as the temperature at the peak maximum of the transition from the folded to unfolded state. For data analysis the Origin software package (MicroCal, GE Healthcare) was used for baseline subtraction and graph presentations.

### 2.12 Maltoheptaose *p*-nitrophenol ($G_7$-pNP) assay

The amylase kit AMYL α-amylase liquid (Roche) was used for kinetic determinations on maltoheptaose (G7) substrate. This kit contains an α-glucosidase solution R1 and a substrate solution R2: 22 mM 4,6-ethylidene-(G7)-pNP. The R2 solution was diluted with 52.5 mM HEPES pH 7.0 to the following final substrate concentrations in the assay; 3, 2, 1, 0.5, 0.25, 0.125, 0.08, 0.075, 0.0625, 0.05, 0.025 mM. The substrate solutions were inserted in the KONELAB ARENA 30 (Thermo Electron Corporation) as samples. The R1 solution and enzyme solution (3–18 nM in 52.5 mM HEPES pH 7) were inserted as Reagent 1 and 2, respectively. The experiment was run on the KONELAB with a program set at 30ºC where 147 µl Reagent 1 was mixed with 83 µl sample and incubated 5 min. Hereafter, 20 µl Reagent 2 was added and further incubated 3 min. After this the absorbance was measured 7 times over 120 sec at 405 nm. pNP standard (0, 0.030, 0.075, 0.151, 0.226, 0.301, 0.602 mM in HEPES buffer) was made from *p-n*itrophenol standard solution 10 mM (Sigma) and mixed as for the substrate solutions and measured. $K_M$ and $k_{cat}$ were obtained by fitting to the Michaelis-Menten equation with the program CurveExpert 1.3 (http://www.curveexpert.net/). Significant differences

between the forms are evaluated by a 95% confidence interval. The standard deviation for $k_{cat}/K_M$ was calculated according to the formula: $\sigma_{\frac{A}{B}} = \frac{A}{B}\sqrt{(\frac{\sigma_A}{A})^2 + (\frac{\sigma_B}{B})^2}$ where $\sigma$ is the standard deviation and A and B are average values. As the $K_M$ and $k_{cat}$ values are determined from the same assays they may not be independent variable and a term reducing the standard deviation should be used when the two factors are dependent. This can be complicated therefore the stated formula was used though it might overestimate the uncertainty somewhat.

### 2.13 Amylose DP440 assay

The initial rate of formation of reducing power was determined at ten different amylose DP440 (Amylose type III: from potato, essentially free of amylopectin, Sigma) concentrations; 2.5, 2, 1.5, 1, 0.75, 0.5, 0.4, 0.3, 0.2, 0.1 mg/ml in 20 mM HEPES pH 7, 5 mM $CaCl_2$, 4.44% v/v Dimethyl sulfoxide (DMSO, minimum 99,5% GC, Sigma), 0.005% w/v bovine serum albumin (BSA, Fraction V minimum 96% lyophilized powder, Sigma). Each substrate solution 990 µl was equilibrated 10 min at 30°C. With 15 sec intervals 110 µl enzyme (3–5 nM in 20 mM HEPES pH 7, 5 mM $CaCl_2$, 0.005% w/v BSA) was added to each substrate solution. With 2.5 min intervals 200 µl aliquots was removed from each solution and mixed with a stop solution containing 300 µl $H_2O$ and 500 µl reducing sugar assay reagent (RSAR). RSAR was a 1:1 mixture of reagent A and B. Reagent A was made by dissolving 27.4 g $Na_2CO_3$ and 12.0 g $NaHCO_2$ *ad* 400 ml $H_2O$, hereafter 971 mg bicinchonic acid disodium salt hydrate (Sigma) was added and $H_2O$ *ad* 500 ml. Reagent B was 620 mg $CuSO_4$ (pentahydrate, JT Baker) and 630 mg L-serine (ReagentPlus™ ≥99%, Sigma) dissolved in 500 ml $H_2O$. Maltose (monohydrate) standards (0, 0.0158, 0.0211, 0.0317, 0.0528 mM) were mixed 1:1 with RSAR and these were along with the samples incubated 30 min at 80°C, and then quickly cooled on ice. The samples were centrifuged 5 min at 14000 g, transferred to a microtiter plate and absorbance was read at 540 nm. $K_M$ and $k_{cat}$ were obtained as above.

### 2.14 Forced glycation of proteins

The experiment was performed by incubating a solution of 2 mg/ml protein in either 5% w/v glucose, maltose, maltotriose, maltoheptaose or 2 mg/ml soluble potato starch in a 25 mM $NaHCO_3$, 0.1 mM $CaCl_2$, 0.01% w/w Triton x-100, pH 9.5 buffer at 40°C. Negative control was also made with 2 mg/ml protein in 25 mM $NaHCO_3$, 0.1 mM $CaCl_2$, 0.01% w/w Triton x-100, pH 9.5. The samples were incubated up to 17 days and aliquots were withdrawn at different time intervals and stored at -20°C.

### 2.15 In-gel digestion with $H_2{}^{18}O$ from IEF gels

The BLA sample and the partly separated pools of BLA PA–PD run on Focusgel 3–7 (as described in section 2.3) loading 60 µg of each twice. Bands were cut out of the gel and destained in 50% ethanol for 2x30 min in thermomixer at 12000 rpm. The gel pieces were dehydrated in 100% acetonitrile for 15 min, supernatant was removed and the samples dried. All the following

solutions were prepared with water-$^{18}O$, 97atom% $^{18}O$ (Sigma-Aldrich). To each gel pieces 15 µl 0.025 µg/µl trypsin (Trypsin Sequencing grade, Roche Diagnostics) or 15 µl 0.0027 µg/µl Asp-N (Endoproteinase Asp-N Sequencing grade, Roche Diagnostics) in 50 mM $NH_4HCO_3$ is added and stored for 45 min at 5°C. The remaining protease solution is removed and 25 µl 50 mM $NH_4HCO_3$ is added to each gel piece and incubate at 37°C over night. After digesting the gel pieces are spun down and 50 µl 70% acetonitrile 0.1% triflouroacetic acid is added. After 15 min the supernatants are transferred to a clean tube. This step is repeated twice. The supernatants are dried and dissolved in 80 µl 5% formic acid.

### 2.16 In solution digestion

The proteolytic digestions were performed on 100 µg protein. BASI samples were added 350 µl 12.5 mM DTT in 8 M UREA 50 mM Tris pH 8.0 and incubated 1 h 42°C. Then 50 µl 20 µg/µl iodoacetamide was added and incubated 1 hr at room temperature in the dark. Before digestion 500 µl 50 mM $NH_4HCO_3$ pH 7.8 were added. BLA samples either incubated 6 h or over night in cold acetone (preincubated at -20°C) at 4°C, centrifuged 10 min at 8000 g and dried pellet dissolved in 250 µl 50 mM $NH_4HCO_3$ pH 7.8 or incubated 20 min in 25 mM EDTA before digestion.

For digestion 15 µl 0.1 µg/µl trypsin in 50 mM $NH_4HCO_3$ pH 7.8 were added to the samples and incubated at 37°C for 5 hr or over night. The samples were purified on EMPORE C18 standard density, solid phase extraction cartridge (3M) according to manufactures protocol, dried and dissolved in 5% formic acid.

### 2.17 Intact protein mass spectrometry

Intact proteins were analysed on a micrOTOF II (Bruker Daltonics) coupled to an Agilent LC 1000 series (Agilent). LC was performed on a MassPREP On-Line Desalting Cartridge 2.1x10 mm (Waters Corporation) eluting with an acetonitrile step gradient in 0.05% triflouroacetic acid with flow 400 µl/min. For each analysis 100 µg protein was used. Data analysis was performed with DataAnalysis vers. 3.4 (Bruker Daltonics).

### 2.18 Tandem mass spectrometry

The LC-MS/MS experiments were performed on either a LTQ Orbitrap Velos (Thermo Scientific) for samples fragmented with HCD or on a LTQ Orbitrap XL (Thermo Scientific) fragmenting with HCD or ETD. Both instruments have a nano-ESI source (Nanospray Flexion Source, Thermo Scientific) couple to an EASY-nLC II (Thermo Scientific). The peptides were separated on Thermo Scientific EASY-Column (C18) 2 cm x 100 µm i.d. column followed by a Thermo Scientific EASY-Column (C18) 10 cm x 75 µm i.d. column, eluting from 5–50% acetonitrile in 0.1% formic acid over 60 min at a flow rate of 250 nl/min.

Experiments on the Orbitrap Velos were only conducted in the orbitrap with automatic gain control (AGC) target of 1,000,000 with 30,000 resolution and a maximum ion injection time of 10 ms in full MS. For MS/MS AGC target was 50,000 with injection time of 100 ms and

resolution of 7500. The HCD normalized collision energy was 40% and the 10 most intense precursor ions in the range 350–1750 m/z were selected for MS/MS.

For analysis on the Orbitrap XL $MS^1$ was performed in the orbitrap mass analyser with a resolution of 60,000 (at 400 m/z) for ETD and of 30,000 for HCD. AGC target was 500,000 with a maximum ion injection time of 300 ms for the orbitrap in full MS. AGC target for MS/MS was 40,000 for the ion trap and 100,000 for the orbitrap. Normalized collision energy of 45 were used for HCD and fragment ions were detected in the orbitrap with 7500 resolution with a maximum injection time of 300 ms using 2 microscans selecting precursor ions in the range 300–2000 m/z. ETD reactions were performed for 85 ms detecting fragment ions in the ion trap with a maximum injection time of 200 ms using 3 microscans selecting precursor ions in the range 350–1200 m/z. For both HCD and ETD the four most abundant precursor ions were selected, rejecting charge state 1 for HCD and charge states 1 and 2 for ETD.

The data analysis was performed with Mascot Distiller (Matrix Science) and Mascot Deamon (Matrix Science) for raw data processing and automation of the searches with the Mascot search engine. The searches were conducted against a custom made database containing 14,100 sequences with decoy option enabled. The searches were conducted with trypsin or Asp-N as enzyme allowing 5 missed cleavages with 10 ppm precursor mass tolerance, 0.5 Da fragments mass tolerance for ETD and a mass tolerance of 0.02 Da for HCD. Static modification carbamidomethyl (C) was set for searches on BASI and no static modifications were selected for searches on BLA. The following variable modifications were chosen: Deamidation (NQ), Hex (KR), 3-deoxyglucosone (KR), carboxymethyl (KR), Carboxyethyl (KR), MG-H1 (KR), Oxidation (M), Carbamyl (KR), Acetyl (K) and Deamidated: 18O(1) (NQ), Label: 18O(1)(C-term), Label: 18O(2)(C-term) for samples digested with $H_2{}^{18}O$.

### 2.19 Relative quantification of glycation sites

The TMTsixplex isobaric mass tagging kit was purchased from Thermo Scientific (Rockford). Samples were digested, reduced, alkylated and labelled with TMT tags according to manufacturer protocol. Glycated BASI samples (see section 2.14) were mixed 1:1 for BASI incubated 4 and 8 days. The BLA pools (PA-PD) were mixed 1:1:1:1 after labelling. Labelled samples were purified on Solid phase extraction cartridge (C18 octadecyl standard density, 3M) according to manufacturer's protocol. The samples were dried and diluted in 5% formic acid.

All experiments were performed on a LTQ Orbitrap XL mass spectrometer (Thermo Scientific) with nano-ESI source couple to an EASY-nLC II (Thermo Scientific). The peptides were separated on Thermo Scientific EASY-Column (C18) 2 cm x 100 μm i.d. column followed by a Thermo Scientific EASY-Column (C18) 10 cm x 75 μm i.d. column, eluting from 5–50% acetonitrile in 0.1% formic acid over 60 min at a flow rate of 250 nL/min. $MS^1$ was performed in the orbitrap mass analyser with a resolution of 60,000 (at 400 m/z), AGC target was 500,000 with a maximum ion injection time of 300 ms. AGC target for MS/MS was 40,000 for the ion trap and 100,000 for the orbitrap. Normalized collision energy of 45 and 55% were used for HCD and fragment ions were detected in the orbitrap with 7500 resolution with a maximum injection

time of 300 ms using 2 microscans selecting precursor ions in the range 300 – 2000 m/z. ETD reactions were performed for 85 – 100 ms detecting fragment ions in the ion trap with a maximum injection time of 100 ms using 1 microscan selecting precursor ions in the range 350 – 1200 m/z. For both HCD and ETD the four most abundant precursors ions were selected, rejecting charge state 1 for HCD and charge state 1 and 2 for ETD.

The data analysis was performed using Proteome Discoverer 1.3 (Thermo Scientific) for spectra processing with SEQUEST or Mascot (Matrix Science) as search engines for protein/peptide identification and quantification. Mascot Distiller (Matrix Science) was also used for raw data processing and Mascot Daemon (Matrix Science) was used for search automation with Mascot server as search engine for protein/peptide identification and quantification.

For data analysis in Proteome Discoverer 1.3 SEQUEST searches were conducted against a custom made against databases containing only BASI or BLA. The searches were conducted against tryptic peptides allowing 5 missed cleavages with 10 ppm precursor mass tolerance, 0.5 Da fragments mass tolerance for ETD and 0.02 Da mass tolerance for HCD. Static modification TMT-6plex of N-terminus and carbamidomethyl (C) was set and the following variable modifications were chosen: Deamidation (NQ), Hex (KR), 3-deoxyglucosone (KR), carboxymethyl (KR), Carboxyethyl (KR), MG-H1 (KR), Oxidation (M). The quantitation was set to TMT-6plex. The false discovery rates were set below 0.05. Both Mascot Searches was conducted with the same parameters as described for SEQUEST against a custom made database containing 14100 sequences (mainly enzyme sequences from different organisms) also searching against a decoy database.

For TMT quantification normalization was not used, auto outlier removal was chosen and peptide threshold was set to above homology level. Quantification data was corrected for TMT isotope impurities. The peptide ratios and standard deviations were obtained from triple determinations of each labelled sample mix.

### 2.20 Expression of BLA mutants

BLA mutants and BLA for comparison were produced by Novozymes in October 2011. The constructs were made by site directed mutagenesis according to the principles described by others (Deng *et al.*, 1992). BLA and its variants was expressed in recombinant *B. subtilis*, fermented in 500 ml shaking flasks at 270 rpm and 37°C for 96 h in a medium consisting of 100 mg/ml sucrose, 40 mg/ml soya flour, 10 mg/ml $Na_2HPO_4 \cdot 12H_2O$, 0.01% v/v Dowfax 63N10, 5 mg/ml $CaCO_3$.

### 2.21 Purification of BLA variants

The fermented samples were centrifuged at 5400 g for 50 min and filtered through 0.2 µm PES filters. MOPS, $CaCl_2$ and $(NH_4)_2SO_4$ were added to concentrations of 20 mM, 1 mM and 1 M respectively, and pH adjusted to 7.5. Samples were loaded to a 50 ml phenyl sepharose column (Phenyl sepharose 6 fast flow packed in a XK 26/20 column, GE Healthcare) with flow 8 ml/min and washed with 2 CV washing buffer (20 mM MOPS, 1 mM $CaCl_2$, 1 M $(NH_4)_2SO_4$, pH

7.5) eluting from 100–0% $(NH_4)_2SO_4$. Selected fractions were tested on SDS-PAGE. Eluted amylase fractions were dialysed (MWCO 12–14,000) against 10 mM Tris, 1 mM $CaCl_2$, pH 8.9. The dialysed sample (conductivity ≤ 2.0 ms/cm) were loaded to a 50 ml Source 30Q (GE healthcare) packed in a XK 26/20 column with flow 8 ml/min washed with 1 CV washing buffer (20 mM Tris, 1 mM $CaCl_2$, pH 8.9) eluting with 0–60% elution buffer (1M NaCl, 20 mM Tris, 1 mM $CaCl_2$, pH 8.9) over 110 min. Eluted fractions were tested on SDS-PAGE. Fractions only containing BLA were pooled, concentrated and buffer changed to 20 mM HEPES, 1 mM $CaCl_2$, pH 7.0 on Vivaspin 20 10,000 MWCO PES (Sartorius Stedim).

### 2.22 Mini wash

Wash performance for BLA variants were determined on a mini wash robot (Novozymes) on coloured rice starch cotton CS-28 (Center For Testmaterials BV). CS-28 strips (5 cm wide) mounted on racks which during the wash are moving up and down into the wash beakers (50 times per min). Washes were performed at 40°C for 30 min at enzyme concentrations 0, 0.015, 0.05, 0.25, 0.5 and 1.0 mg /L followed by 5 min rinsing in cold water. Each wash beaker contain 60 ml of either 0.8 g/l liquid detergent pH 8.5 (20% linear alkylbenzene sulfonate (LAS, Nacconol 90G), 10% alkyl ether sulfate (STEOL CS-370E), 10% alcohol ethoxylate (AEO, Bio-soft N25-7, Stepan Company), 4% oleic acid, 0.5% Ethanol, 1.5% SXS (sodium xylene sulfonate), 7% 1,2-propanediol, 8% tri-Na-citrat, 0.5% triethanolamine, 1.5% boric acid, pH 8.5) or powder detergent pH 10.0 of 1.715 g/L AP powder model detergent X without nonionic surfactant (17% LAS, 20% Sodium carbonate, 12% Sodium disilicate, 15% Zeolite A, 1% PCA, 34% Sodium sulfate) and 0.035 g/L AEO, pH 10.0, both detergents prepared with water hardness of 15°dH (Ca/Mg 4:1). Double determinations were performed for each wash. The CS-28 strips dried O/N and measured twice at 460 nm on TIDAS (Zeiss spectrophotometer). All percentages stated are in w/w.

# 3 Results and discussion

## 3.1 Multiple forms of *Bacillus* α-amylases

The four α-amylases BAN, BHA, BLA and AA560 were examined for the presence of multiple forms. Samples from different time points during the fermentation and from the recovery were analysed by IEF to determine where in the process the multiple forms arise. The multiple forms of some of the amylases were partially separated. The wash performance of the partially separated forms was evaluated in a simple laundry wash assay, to determine if the modifications generating the different forms influence the wash performance. The distribution of the forms was also analysed by capillary IEF (cIEF). Based on these results a more thorough evaluation of the influence of the modifications was performed on one of the amylases.

### *3.1.1 α-Amylase modifications during production*

The recovered α-amylases all appears to be pure and uniform on SDS-PAGE but on IEF gel multiple bands occur for all of them, Figure 9. This reveals that all the amylases become modified during the fermentation, recovery or formulation resulting in multiple forms varying in pI values.

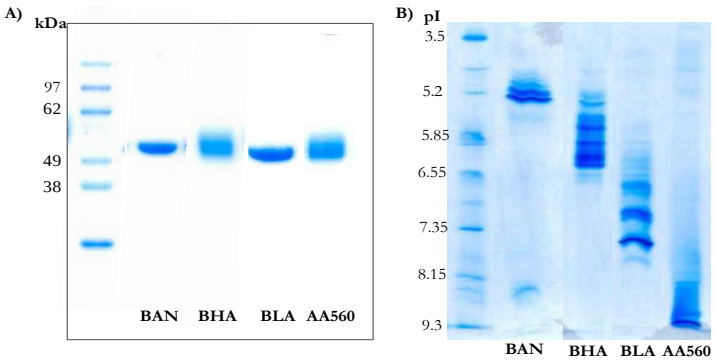



**Figure 9.** The α-amylases BAN, BHA, BLA and AA560 on A) SDS-PAGE B) IEF gel.

There appear to be fewer forms of BAN than of the other amylases. AA560 appears to be the most complex of the amylases as it merely creates a smear on the IEF gel whereas for the other several stronger bonds are observed. The decrease in pI value is most likely due to modification of positively charged amino acids or the addition of a charged modification to the amino acids. Therefore it is most likely the positively charged Lys and Arg residues which are modified. The $pK_a$ values of Lys and Arg side chains are approximately 10.8 and 12.5, respectively (Berg *et al.*, 2007). Neutral modifications on these side chains will remove the charge and hereby lower the pI value of the protein. It can also stem from modifications such as deamidation where the uncharged Asn and Gln residues are converted into the negatively charged Asp and Glu. This conversion results in a decrease in pI value as these neutral amino acids side chains will change to charged side chains with $pK_a$ values of approximately 4.1 (Berg *et al.*, 2007).

The observed difference in modification level of the amylases could be caused by variations in the number of Asn, Gln, Lys and Arg residues on the amylases. Though, the change can also

reflect structural differences as some structural conformations can protect amino acid residues from modifications. The number of Lys and Arg residues in the four amylases is more or less the same differing only by a few residues as BHA has 44 while AA560, BLA and BAN have 48, 49 and 50 of these residues, respectively. The number of Gln and Asn residues differ more between the amylases as BAN and BLA have 46 and 45 of these residues, respectively, whereas both BHA and AA560 have 62 residues. This could explain why BHA and AA560 appear to be more heterogeneous and hence more extensively modified than BAN and BLA. BLA, however, also appears to be more modified than BAN even though they almost contain the same number of Gln and Asn residues. This could indicate that the amino acid composition does not entirely account for the differences in the modification level. The structure of the amylases and the exposure of reactive amino acid residues most likely also account for some of the differences in the modification level.

To determine where in the production process the multiple forms are created, samples were withdrawn at different time points during the fermentation and from different stages of the recovery. The samples were withdrawn from the production of BLA, BAN and AA560. They were analysed on IEF gel and SDS-PAGE.

Samples of BLA were withdrawn during the fermentation, after drum filtration and ultrafiltration in the recovery and were analysed by IEF gel, Figure 10. This revealed that the multiple forms of BLA were present already in the fermentation after 20 hours. It also seems that all the forms are present already after 20 hours of fermentation and no new forms appear to be generated in the downstream processes. However, the samples are very complex and it is therefore not possible to exclude that a few more forms of BLA arise specifically during the purification.

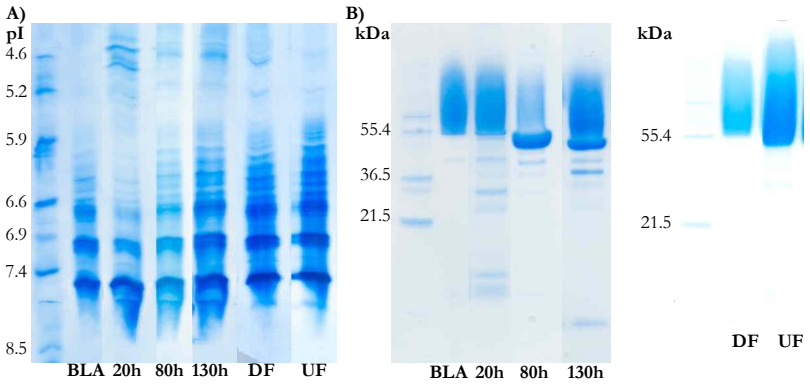



**Figure 10.** Samples of BLA withdrawn from the fermentation after 20, 80 and 130 hours and after drum filtration (DF) and ultra filtration (UF) on A) IEF gel B) SDS-PAGE.

The samples withdrawn from the production of BAN and AA560 also show that the multiple forms arise during the fermentation, Figure 11. No new forms seem to appear during recovery of the amylases, as also observed for BLA (data not shown).

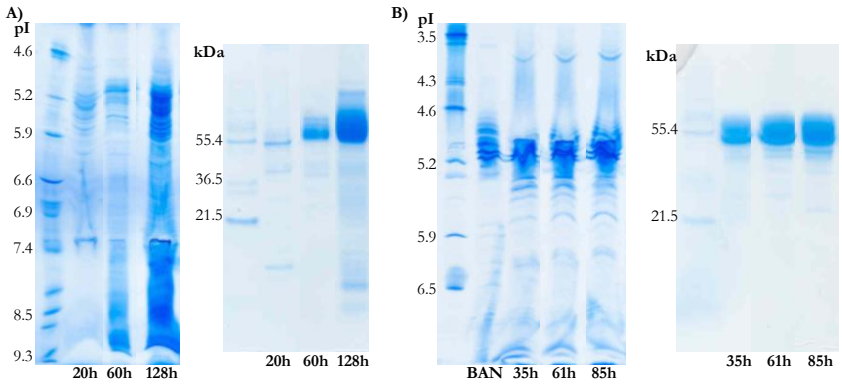



**Figure 11.** Samples of AA560 and BAN withdrawn at different time points during fermentation tested on IEF gel and SDS-PAGE A) AA560 B) BAN.

The pI characterisations of the four different *Bacillus* amylases revealed that they all contain multiple forms varying in pI values. The number of forms varies between the amylases indicating that some of the amylases are more prone to modification than others. This could be due to the differences in the number of reactive amino acids, the number of exposed amino acids for instance on the surface of the protein or other structural differences. The multiple forms of the amylases were found to primarily arise during the fermentation.

### *3.1.2 Effect of the multiple forms of amylases*

The multiple forms of amylase were observed in all the four amylases. It was attempted to separate the multiple forms of the amylases by different chromatographic methods. However, due to the high similarity of the forms it was not possible to separate single forms. The multiple forms of the amylases were therefore only partly separated into pools varying in pI range. The performance of the pools of partly separated amylases were analysed to determine whether the forms differ in wash performance. The focus of this project is on detergent amylases therefore only the detergent amylases BLA, BHA and AA560 were explored further. The performance of the amylases was analysed in some simple laundry wash assays on rice starch cotton patches. The washes were performed in $NaHCO_3$ buffer at 50°C as all the amylases are relatively active and stable at this temperature. The washes were carried out at pH 7.5 and 10.3 as this is the approximate pH in wash with liquid and powder detergent, respectively.

The starch on the cotton patches are coloured pink, so the more of starch is remove from the patch the lighter the colour of the patch becomes. The same amount of enzyme protein (in mg) was added to the laundry wash assays for each of the pools. The wash performance on the patches was determined spectrophotometrically by measuring the remission of the patches. Remission is the reflection of light by a material. In the spectrophotometer the cotton patches are subjected to light from a light source. The light reflecting from the patch is then measured at an appropriate wavelength. The wash performance of the amylase is determined in delta remission units. The delta remission unit is the difference between the absorbance of the remission of a blank (a cotton patch washed without enzyme) and the sample. The higher the delta remission the better is the wash performance of the enzyme.

The multiple forms of BHA were separated by ion exchange chromatography into six pools varying in pI range, Figure 12A. The wash performance of these pools decrease for the pools with the lowest pI range, Figure 12B. There is only a slight decrease at pH 7.5 which is barely significant, but at pH 10.3 the decrease is more pronounced. PN2 has the highest wash performance at pH 10 being slightly higher than the wash performance for PN1 which is suspected to containing the least modified forms. Thus it appears that a lower amount of modifications might have a positive effect on wash performance. PN3 has the same wash performance as PN1 at both wash conditions. The remaining pools containing the forms with the lowest pI values decrease in wash performance especially at pH 10. The forms with the lowest pI values are probably also the most modified. This indicates that a few modifications of BHA have a small positive or no effect on the wash performance at 50°C. However, the more modified BHA becomes the bigger the loss in wash performance, especially at pH 10.3. The same tendencies were observed in similar experiments on AA560 (data not shown).

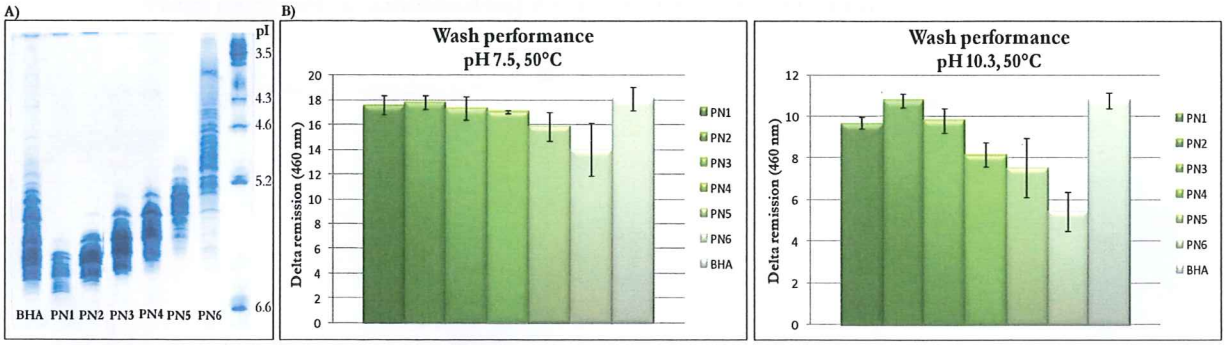



**Figure 12.** Wash performance for the partly separated forms of BHA in $NaHCO_3$ buffer. The wash performances were determined by double determinations and the 95% confidence interval is marked for each sample. A) The partly separated forms of BHA on an IEF gel B) Wash performance of the pools of BHA at pH 7.5 and 10.3 on rice starch cotton patches.

The multiple forms of BLA were also separated by ion exchange chromatography into eight pools varying in pI range, Figure 13A. The wash performance for these pools does not reveal changes in performance for the different pI ranges at pH 7.5, Figure 13B. At pH 10.3 the results indicate that the wash performance of the more modified forms slightly decreases, Figure 13B. The modifications seem to have less effect on the wash performance of BLA than they do on BHA and AA560. This could be due to the thermostability of BLA and the fact that it is relatively stable under alkaline conditions. It is possible that modifications have a larger effect on the wash performance under more stressed conditions such as higher temperature or in a detergent, but this has not been further investigated.

The wash performance results indicated that the forms with the lowest pI values have a lower wash performance. Therefore, two of the amylases were analysed by cIEF to get an approximate distribution of the multiple forms. The cIEF was performed by Beckman Coulter as a demonstration of the equipment, therefore only very few samples could be analysed. BLA and AA560 were chosen for cIEF as BLA appears to be the least modified of the three detergent amylases and AA560 appears to be the most complex of the three on IEF gel.

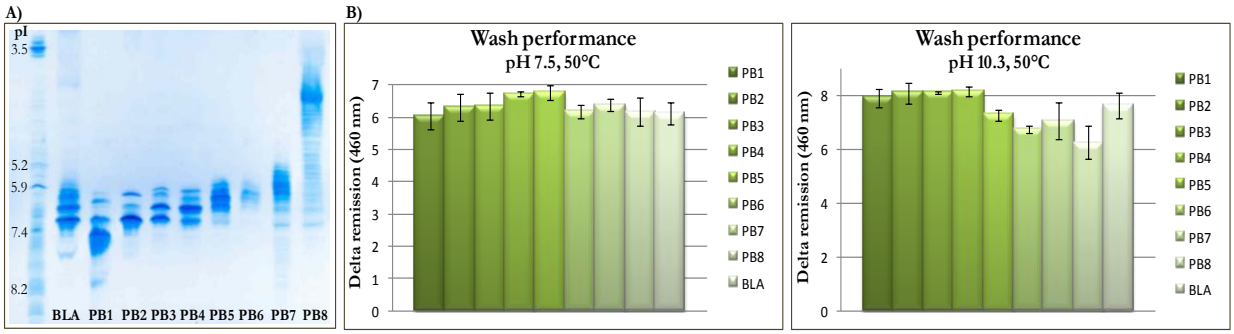



**Figure 13.** Wash performance for the partly separated forms of BLA in $NaHCO_3$ buffer. The wash performances were determined by double determinations and the 95% confidence interval is marked for each sample. A) The partly separated forms of BLA on an IEF gel B) Wash performance of the pools of BLA at pH 7.5 and 10.3 on rice starch cotton patches.

The cIEF for BLA revealed that the main band observed in the IEF gel has a pI value of 6.6 and that this corresponds to approximately 40% of the sample, Figure 14A. The BLA forms with the lowest pI values correspond to 12 – 30% of the sample. This indicates that at least 70% of BLA is of the forms which did not seem to lose wash performance in the wash assay.

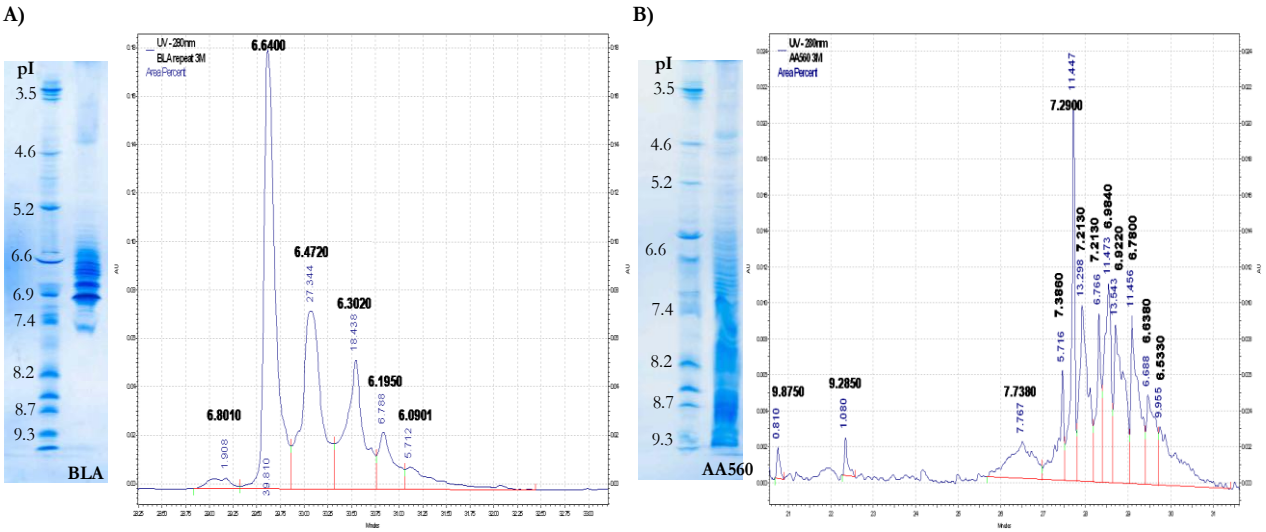



**Figure 14.** Results from cIEF of BLA and AA560. Numbers in blue are the percentage of the form in the sample and numbers in black bold are the determined pI value. A) IEF gel and cIEF results for BLA B) IEF gel and cIEF results for AA560.

The cIEF results for AA560 revealed a much more complex sample covering a much broader pI range, Figure 14B, as was also expected from the IEF results. Two smaller peaks occur with pI values of 9.9 and 9.3 only corresponding to approximately 2% of the sample. The next peak has a pI value of 7.7 and it is followed by multiple other peaks with decreasing pI. The highest amount of one of the peaks represents approximately 14% of the sample. These results indicate that AA560 contains a larger portion of modified forms with relatively low pI values which most likely have decreased wash performance.

The cIEF results for BLA and AA560 revealed that AA560 is much more complex than BLA, similar to the observations made from IEF gels. AA560 also appears to contain a larger portion of the forms with lower pI values which have lower wash performance at 50°C, especially at pH

10.3. The overall wash performance of AA560 could therefore be more affected by the multiple forms than BLA, as it appears that a smaller amount of this sample is the forms with very low pI values having the decreased wash performance.

The results indicate that the modifications responsible for the multiple forms can result in loss of wash performance. It is highly likely that the more modified the amylases become the higher the performance loss, especially under less favourable conditions such as around pH 10.

### *3.1.3 Characterisation of multiple forms of BLA*

In the previous section it was described that the modifications on the amylases could have a negative effect on the wash performance, especially the forms with the lowest pI values appears to be affected. The smaller decreases in pI values did not seem to affect the wash performance negatively. BLA was selected for further characterisation of the activity and stability of the multiple forms. This amylase was selected as its structure and its structure function relationship is well described in the literature.

The multiple forms of BLA were separated into four pools of forms varying in pI value. As mentioned previously (section 2.5) several different chromatography techniques were investigated where the best separation was obtained by anion exchange chromatography eluting with pH gradient followed by a NaCl gradient. This did not result in a complete separation of the forms so pools of forms varying in pI range was characterised. The pool PA contain the forms with the highest pI values and the pI values of the following pools PB-PD decreases, so the forms in PD have the lowest pI values. IEF gel of the pools and the chromatogram from the separation highlighting the selected fractions are illustrated in Figure 15. The concentration of the pools was determined by amino acid analysis.

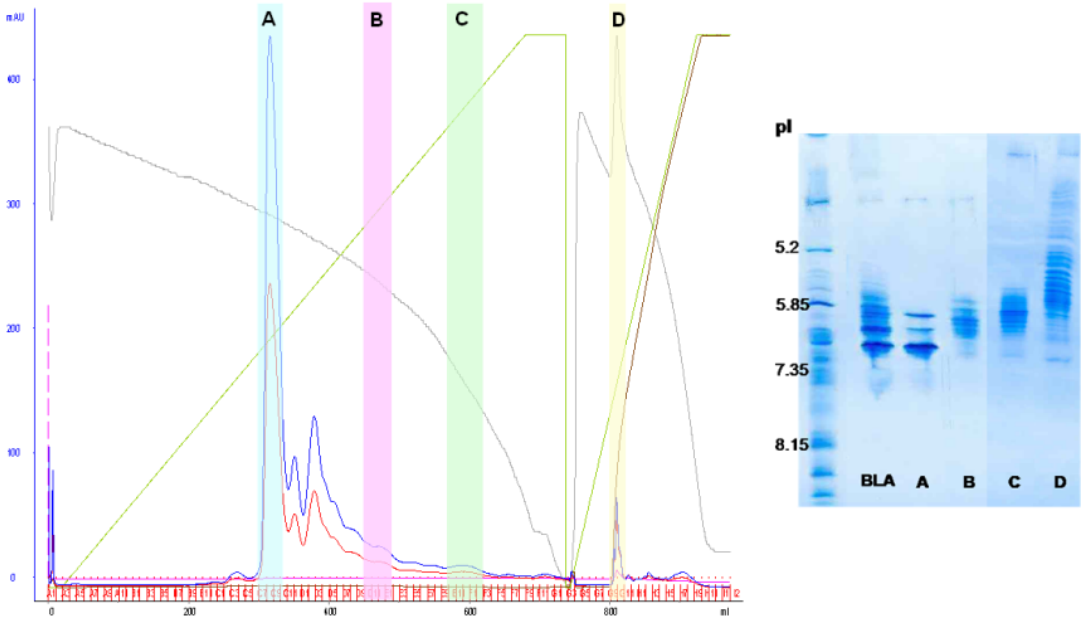



**Figure 15.** The chromatogram from the separation of the multiple forms of BLA with an IEF gel of the four pools of BLA. In the chromatogram the fractions collected for each pool is highlighted with colours. In the chromatogram the blue line is UV280, the red line is UV260, the green line is the gradient, the gray line is pH and the brown is conductivity.

The stability of the pools was analysed by DSC in buffer at pH 7. This provided $T_{max}$ values between 104.2 – 104.4°C (Appendix I Figure 1). The four pools do not differ in stability as they all show the same $T_{max}$ value. Hereby, revealing that the thermostability of the enzyme is not affected by the modification occurring in the different pools. It is, however, possible that some of the modifications occurring have a negative effect on the stability. Not all the BLA forms are present in the pools so it is possible that forms with lower $T_{max}$ is not included in the pools. It is also possible that a form with a modification affecting $T_{max}$ is such a small part of the pools that it does not affect the overall $T_{max}$. Thus, indicating that the effect of a single destabilising modification is small. Overall it appears that the modifications in general do not affect the thermostability of BLA. Studies show that modifying Lys residues decreased the thermostability of BLA (Khajeh *et al.*, 2001a; Khajeh *et al.*, 2001b; Tomazic *et al.*, 1988b). The DSC results for the four pools did not indicate that the modifications present in these pools influence the stability negatively. The present data therefore suggests that the modifications occurring are not of Lys residues, or that the modified Lys residues or other modifications occurring during the fermentation does not influence the enzymes stability.

The kinetic parameters $k_{cat}$, $K_M$ and $k_{cat}/K_M$ were determined for the four pools on maltoheptaose and on amylose at 30°C. The conditions for these experiments were chosen to correspond to the conditions normally used in an European laundry wash, where the temperature is around 30 – 40°C and the pH is around pH 8 and 10 dependent on the type of detergent.

The kinetic parameters with maltoheptaose as substrate, Table 1, were only determined at pH 7 as the substrate is unstable at alkaline pH. The $K_M$ values do not reveal any differences between the pools but smaller variations occur for $k_{cat}$. The smaller variations in the catalytic constant ($k_{cat}$) reveal that PB and PC have slightly higher $k_{cat}$ values than PA and PD. The catalytic efficiencies ($k_{cat}/K_M$) of the pools do not reveal any significant changes between the pools. The results show tendencies of a decreasing catalytic efficiency with decreasing pI values though the differences are not significant. Therefore it cannot be concluded that the modifications resulting in decreasing pI values influences the catalytic efficiency of BLA on the G7 substrate at pH 7, 30°C.

**Table 1.** The kinetic parameters $k_{cat}$, $K_M$ and $k_{cat}/K_M$ for the four BLA pools, PA-PD, varying in pI range. The kinetic parameters were determined at 30°C pH 7.0 on a maltoheptaose substrate.

| | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ ($s^{-1}$ $mM^{-1}$) |
|---|---|---|---|
| PA | 74 ± 7 | 0.011 ± 0.002 | 7117 (± 1364) |
| PB | 92 ± 6 | 0.013 ± 0.002 | 6968 (± 1108) |
| PC | 85 ± 4 | 0.014 ± 0.002 | 6336 (± 919) |
| PD | 68 ± 3 | 0.013 ± 0.002 | 5072 (± 621) |

The kinetic parameters for the four pools on amylose were determined at pH 7.5 and 10.3, Table 2. The parameters for pH 7.5 reveals the same tendencies as observed on the short substrate. There are no significant differences in $K_M$ for the pools. However, smaller differences occur for the $k_{cat}$ values as PD appears to be lower than the other pools. The catalytic efficiency reveals the same tendencies as for the short substrate. The decrease in pI values result in

decreasing catalytic efficiency though again the results do not differ significantly between the pools.

The kinetic parameters for amylose at pH 10.2 are considerably lower than at pH 7.5 revealing that the enzyme is not very active at low temperature and alkaline pH. The tendencies at this pH also differ from the previous observations as substrate catalysis is the same for the four pools. The $K_M$ values, however, differ slightly as $K_M$ for PA and PD appears to be a bit better than for PB and PC. This also appears to be the case for the catalytic efficiencies where PA has the highest efficiency followed by PD and the PB and PC with the lowest efficiencies. It appears that the modifications have a negative effect on the catalytic efficiency at pH 10 where the least modified sample, PA, is better than the other three. The loss in catalytic efficiency does not seem to be dependent on level of modification as the sample which appears to contain the most modified BLA forms, PD, has the second best catalytic efficiency. It is therefore likely that part of the loss in catalytic efficiency at pH 10 is due to specific modifications in specific sites.

**Table 2.** The kinetic parameters $k_{cat}$, $K_M$ and $k_{cat}/K_M$ for the BLA pools varying in pI range determined at 30°C pH 7.5 and 10.2 on amylose DP440.

| | pH 7.5 | | | pH 10.2 | | |
|---|---|---|---|---|---|---|
| | $k_{cat}$ ($s^{-1}$) | $K_M$ (mg/ml) | $k_{cat}/K_M$ ($s^{-1}$ $mg^{-1}$ ml) | $k_{cat}$ ($s^{-1}$) | $K_M$ (mg/ml) | $k_{cat}/K_M$ ($s^{-1}$ $mg^{-1}$ ml) |
| PA | 301 ± 21 | 0.192 ± 0.050 | 1625 (± 424) | 34 ± 4 | 0.224 ± 0.015 | 152 (± 20) |
| PB | 313 ± 10 | 0.208 ± 0.020 | 1515 (± 152) | 31 ± 3 | 0.333 ± 0.044 | 94 (± 15) |
| PC | 279 ± 12 | 0.229 ± 0.060 | 1283 (± 314) | 28 ± 5 | 0.376 ± 0.068 | 74 (± 19) |
| PD | 234 ± 20 | 0.213 ± 0.004 | 1101 (± 96) | 29 ± 3 | 0.225 ± 0.028 | 129 (± 20) |

The wash performance of the four pools was also determined in buffer at 30°C, pH 7.5 and 10.3 on rice starch cotton patches, Figure 16. These results correlate with the determinations on amylose and show that BLA generally have lower activity around pH 10 than around pH 7.5 at 30°C. The wash performance for PA and PB appears to be slightly higher than for PC and PD. PC appears to have the lowest wash performance. The wash results obtained at 30°C are considerable lower than the ones observed for the forms of BLA at 50°C, Figure 13. This difference could be due to the enzyme being more active at higher temperatures or that the structure of the substrate differs at the two temperatures so it is more accessible for the enzyme at higher temperatures.

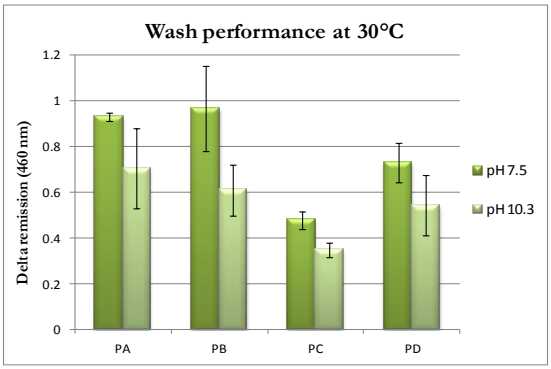



**Figure 16.** Wash performance of the partly separated pools of BLA, PA-PD, in $NaHCO_3$ buffer pH 7.5 and 10.3 at 30°C on rice starch cotton patches.

The characterisation of the BLA pools varying in pI range revealed that the modifications can have an effect on the activity of the enzyme. The kinetic parameters and the wash performance assay indicated that the pools containing the forms suspected to contain the most modifications generally showed slightly lower catalytic efficiency and wash performance. The modifications do, however, not seem to have much effect on the thermostability of BLA.

Identifying the specific modifications and the modification sites could provide more insight into which sites are prone to modifications. Amino acid substitutions at these sites could offer further understanding of the importance of these sites and the influence of the modifications in these sites on the overall activity of the amylases.

## 3.2 Study of glycation sites on a model protein

A MS method for the characterisation of glycation sites was setup with barley α-amylase/subtilisin inhibitor (BASI) as model protein. As mentioned previously glycations can be difficult to identify by the conventional fragmentation methods such as CID and HCD as they cleave of the modification which can result in poor peptide sequence assignment (Mann *et al.*, 2003). The fragmentation method ETD should leave labile modifications such as glycations intact allowing for direct mapping of the modified site (Chi *et al.*, 2007; Syka *et al.*, 2004). Glycations is expected to be among the modifications responsible for the multiple forms of the amylases. Therefore, a MS methods fragmenting with ETD was setup for the characterisation of glycation sites. The ETD method was compare to HCD to determine if ETD is better at identifying glycation sites. BASI was chosen as model protein as it is a smaller protein (20 kDa) with a neutral pI value (theoretical pI 6.5) like BLA and it contains multiple Lys (7) and Arg (12) residues (Svendsen *et al.*, 1986; Vallée *et al.*, 1998). As glycations occur on Lys and Arg residues BASI is a good model protein for the characterisation of this modification as it contains many of these residues and it, therefore is possible to force several glycations on it.

It could be interesting to compare the amount of glycations in the four pools of partly separated BLA varying in pI range (pools PA-PD from section 3.1.3). Therefore, a method for relative quantification of glycations and the advanced glycation end products (AGEs) was setup using isobaric stable isotope tagging with ETD and HCD comparing the two fragmentation methods ability to quantify glycation sites with this quantification method.

### *3.2.1 Identification of glycation sites by HCD and ETD*

BASI was glycated by incubating with glucose at 40°C pH 9.5 for four and eight days. The occurrence of the forced glycations on BASI was evaluated by full-length MS. Before the incubation with glucose the homogeneity of the BASI sample was evaluated by full-length MS (Appendix I Figure 2). This revealed a peak at 19875.5 m/z corresponding to unmodified BASI (theoretical Mw 19875 Da). The main peak was followed by a few smaller shoulders corresponding to smaller modifications. The BASI sample was relatively homogeneous and therefore not extensively modified so it could be used for the glucose incubation experiments.

Full-length MS of BASI incubated four days with glucose (BASIG4), Figure 17A, and eight days with glucose (BASIG8), Figure 17B both resulted in a pattern of peaks with a mass difference of approximately 162 Da corresponding to glycations. In both spectra BASI molecules with up to four glycations were detected. The MS spectrum for BASIG8 is much more complex than for BASIG4 and it could contain BASI molecules with 5 – 6 glycations. It is, however, not possible to identify peaks corresponding specifically to this.

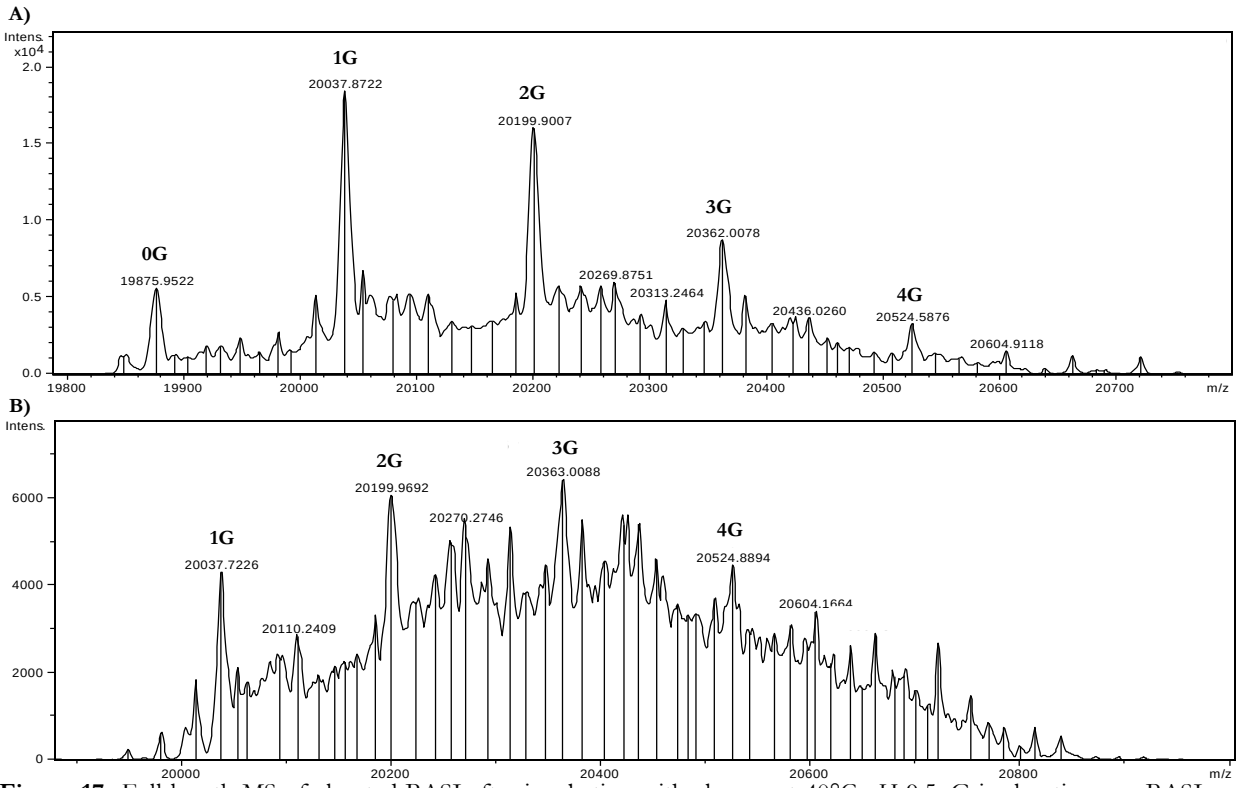



**Figure 17.** Full-length MS of glycated BASI after incubation with glucose at 40°C pH 9.5. G is glycations on BASI. A) BASI incubated 4 days with glucose. B) BASI incubated 8 days with glucose.

The BASI samples were digested by trypsin and analysed by LC-MS/MS fragmenting with either ETD or HCD to identify the glycation sites. The MS/MS data was searched using the Mascot search engine.

Mascot determines the theoretical data for each of the sequences in the database and compares these to the experimental data. All calculated data correlating with the experimental data within a given mass tolerance counts as a match and each match is assigned a statistical weight. Peptide matches obtained from the MS/MS data are assigned a score which is based on the calculated probability that the match between the experimental data and the database sequence is random (Matrix Science 2012; Perkins *et al.*, 1999). This score has to exceed a significance threshold which for this project was set to 5%. Thus a match is significant if it is expected to occur at random with a frequency of less than 5%. Peptide hits of which the scores exceed this significance threshold are then reported to exceed the "identity" threshold. However, some MS/MS spectra might be of poor quality, for instance if the signal to noise ratio is low. This may lead to a correct sequence match with an insignificant peptide score. Therefore, Mascot characterises the distribution of random scores and determines a lower relative threshold, referred to as the "homology" threshold. The peptide scores are accompanied by an "expect" value which is the number of matches with equal or higher score that are expected to occur by chance. Mascot also provides a score for each protein hit which is derived from the peptide scores. The protein score is the sum of the highest peptide scores for each distinct sequence match where a small correction is performed to reduce contributions from low scoring random matches (Matrix Science 2012; Perkins *et al.*, 1999).

A MS/MS method was setup with ETD for identification of glycation sites. This method was then compared to MS/MS fragmenting with HCD to evaluate the fragmentation methods ability to identify glycation sites. In the experiments MS was performed in the orbitrap analyser and so was the HCD fragmentation, whereas MS/MS with ETD was analysed in the ion trap. ETD should be most effective on peptides with three or more charges (Zhang *et al.*, 2008a) and ions below 850 m/z (Good *et al.*, 2007; Swaney *et al.*, 2007). Therefore, the most intense ions in MS with charge 3 or above within the range 350 – 1200 m/z were selected for ETD fragmentation. A quick comparison also revealed that these settings provided better results than settings selecting charge state 2 or above in the MS range 350 – 2000 m/z. The reaction time for ETD fragmentation was also optimised on glycated BASI as this setting differs between studies performed by others (Good *et al.*, 2007; Zhang *et al.*, 2008a). ETD was tested with reaction times in the range 85 – 160 ms. The highest sequence coverage for BASI and the highest number of identified glycation sites were obtained with reaction times of 85 and 100 ms. Reaction time of 85 and 100 ms both resulted in a protein hit for BASI with 92% sequence coverage and approximately the same protein score. However, three more glycation sites were identified with activation time of 85 ms and the MS/MS search results for the glycated peptides had slightly better sequence coverage at this activation time. Therefore, activation time of 85 ms was used for the identification of glycation sites with ETD fragmentation.

The sample was also analysed by MS/MS fragmenting with HCD. This resulted in one hit for BASI with sequence coverage of 95%. The overall coverage and score for BASI was slightly better for the fragmentation with HCD compared to ETD which was expected as HCD is a very efficient fragmentation method.

Comparison of the results for identified glycation sites by the two fragmentation methods reveals the identification of more glycation sites by ETD, Table 3. There were found 7 glycation sites with ETD and in total 12 combinations of the 7 modification sites with various glycations. For the HCD fragmentation 5 glycation sites and 7 of the sites and glycation combinations were observed. The sites identified by HCD were also identified by ETD so for the identification of glycation sites ETD outperforms HCD.

The full-length MS of BASIG4 and BASIG8, Figure 17, identified BASI molecules with up to four glycations and for BASIG8 it indicated that five or six glycations might occur. The MS/MS result identifies seven glycation sites though this method cannot determine if all the sites are modified on some of the BASI molecules.

**Table 3.** Glycations observed by LC-MS/MS fragmenting with ETD or HCD on trypsin digested BASI incubated 4 days with glucose. The results were obtained by Mascot searches. Hex: Glycation with glucose, Cbe: Carboxyethyl, Cbm: Carboxymethyl, MG-HI: methylglyoxal-derived hydroimidazolone, () is hits exceeding the homology threshold but not the identity threshold, Covered is whether or not the glycation site is directly mapped by the MS/MS results, M: missed cleavages.

| Site | Modification | Hits | Score | Expect | Covered | Peptide | M | Sequence |
|---|---|---|---|---|---|---|---|---|
| **ETD** | | | | | | | | |
| **R27** | Cbm | 1 | 56 | 4.8E-5 | + | 16 – 41 | 1 | R.ADANYYVLSANRAHGGGLTMAPGHGR.H |
| **K72** | Hex | 7 (1) | 61 | 1.7E-5 | + | 62 – 75 | 1 | R.ITPYGVAPSDKIIR.L |
| **K72** | Cbm | (3) | 23 | 0.033 | + | 62 – 75 | 1 | R.ITPYGVAPSDKIIR.L |
| **K115** | Hex | (5) | 40 | 0.0025 | + | 108 – 122 | 1 | R.HVITGPVKDPSPSGR.E |
| **K130** | Hex | 1 (2) | 54 | 1.2E-4 | + | 128 – 140 | 1 | R.IEKYSGAEVHEYK.L |
| **R155** | MG-HI | 1 | 50 | 1.1E-4 | + | 141 – 158 | 1 | K.LMSCGDWCQDLGVFRDLK.G |
| **R155** | Cbm | 1 (1) | 50 | 0.0024 | + | 141 – 158 | 1 | K.LMSCGDWCQDLGVFRDLK.G |
| **R155** | Cbe | 1 (1) | 56 | 5.9E-5 | + | 141 – 158 | 1 | K.LMSCGDWCQDLGVFRDLK.G |
| **K158** | Hex | 4 (5) | 69 | 5.2E-6 | + | 156 – 176 | 1 | R.DLKGGAWFLGATEPYHVVVFK.K |
| **K158** | Cbm | 5 | 65 | 5.1E-6 | + | 156 – 176 | 1 | R.DLKGGAWFLGATEPYHVVVFK.K |
| **K176** | Hex | 8 | 72 | 7.0E-6 | + | 159 - 177 | 1 | K.GGAWFLGATEPYHVVVFKK.A |
| **K176** | Cbm | (1) | 37 | 7.3E-4 | + | 159 - 177 | 1 | K.GGAWFLGATEPYHVVVFKK.A |
| **HCD** | | | | | | | | |
| **R27** | Cbm | 1 | 65 | 2.9E-6 | + | 16 – 41 | 1 | R.ADANYYVLSANRAHGGGLTMAPGHGR.H |
| **K72** | Cbm | (5) | 38 | 3.3E-4 | + | 62 – 75 | 1 | R.ITPYGVAPSDKIIR.L |
| **K115** | Hex | (4) | 18 | 0.046 | - | 108 – 122 | 1 | R.HVITGPVKDPSPSGR.E |
| **K115** | Cbm | (2) | 24 | 0.0098 | - | 108 – 122 | 1 | R.HVITGPVKDPSPSGR.E |
| **K158** | Hex | (9) | 43 | 1.6E-4 | - | 156 – 176 | 1 | R.DLKGGAWFLGATEPYHVVVFK.K |
| **K158** | Cbm | 3 | 62 | 6.0E-6 | + | 156 – 176 | 1 | R.DLKGGAWFLGATEPYHVVVFK.K |
| **K176** | Hex | (5) | 24 | 0.01 | - | 159 - 177 | 1 | K.GGAWFLGATEPYHVVVFKK.A |

The score and expect value for the observed sites were not necessarily better for ETD. Though as expected the sequence coverage for the glycated peptides and the coverage of the glycation sites were generally better for ETD than for HCD as illustrated in, Figure 18. MS/MS spectra and Mascot peptide ion data for the two fragmentations in Figure 18 are shown in Appendix I Figure 3 and 4. MS/MS spectra and Mascot peptide ion data for the two fragmentations of the peptide H108 – R122 with glycation on K115 are also illustrated in Appendix I Figure 5 and 6. For this peptide the mapping of the glycation site by HCD is not very good and the peptide scores are quite low especially for HCD.

It was unexpected that some of the glycation sites are covered by the HCD fragmentation, Figure 18. ETD did generally resulted in better coverage by the glycated ions in the MS/MS spectrum. This indicates that HCD not always cleaves off labile modifications such as glycations and that it sometime is possible to obtain good sequence coverage for glycated peptide by HCD.

All the peptides identified to be glycated contain one missed cleavage. This is because trypsin is unable to cleave after glycated Lys and Arg residues. In Mascot hits where the glycated site is not directly mapped, Mascot assigns the glycation site to the Lys or Arg residue in the C-terminal end of the peptide instead of assigning it to the Lys or Arg residue at the missed cleavage site. Therefore, one should be aware that for peptide hits with low sequence coverage Mascot tend to assign the glycation to the wrong Lys or Arg residue.

The MS/MS analysis of glycation sites by HCD and ETD revealed that ETD was better at identifying glycation sites whereas HCD provided a higher overall protein coverage. Therefore, it

would be advantageous to use both methods for the identification of modification sites on amylases to achieve the best possible results.

| HCD | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| **b** | | | | | | ● | ● | ● | | | | | | | | | | | | | |
| **$b^{++}$** | | | | | ● | ● | ● | ● | ● | ● | | | | | | | | | | | |
| **Sequence** | **D** | **L** | **K** | **G** | **G** | **A** | **W** | **F** | **L** | **G** | **A** | **T** | **E** | **P** | **Y** | **H** | **V** | **V** | **V** | **F** | **K** |
| **$y^{++}$** | | | | | | | ● | ● | ● | ● | ● | ● | ● | ● | | ● | | | | | |
| **Y** | | | | | | | | | | | | | | ● | | ● | ● | ● | ● | ● | ● |
| | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| **ETD** | | | | | | | | | | | | | | | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| **c** | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | | ● | ● | | | | | | |
| **$c^{++}$** | | | | | | | | | | | | ● | | | | ● | | ● | ● | ● | |
| **Sequence** | **D** | **L** | **K** | **G** | **G** | **A** | **W** | **F** | **L** | **G** | **A** | **T** | **E** | **P** | **Y** | **H** | **V** | **V** | **V** | **F** | **K** |
| **$z+1^{++}$** | | | ● | ● | ● | ● | | ● | ● | | ● | | ● | | | | | | ● | | |
| **z+1** | | | | | | | | | | | | | ● | | | ● | ● | ● | ● | ● | ● |
| **Y** | | | | | | | | ● | ● | ● | | | ● | | | | ● | | | ● | ● |
| | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

**Figure 18.** Sequence coverage obtained by HCD and ETD for the peptide D156 – K176 with the AGE carboxymethyl on K158. Red dots show the ions found in the Mascot search of the MS/MS spectrum. The ion score and expect value for the HCD results were 62 and 6.0E-6, respectively and for the ETD results they were 65 and 5.1E-6, respectively. The K with the green background in the sequence is the glycated residue. It is observed that even though b ions with the glycation are found in the MS/MS spectra for HCD the data for ETD still provides a more complete coverage of the sequence series with the glycation.

### *3.2.2 Method setup for relative quantification of glycation sites by HCD and ETD*

A method for relative quantification of glycation sites was set up so the glycations in the four partially separated BLA pools (PA – PD) could be quantified between the pools. The method was setup on BASIG4 and BASIG8 using isobaric stable isotope tags for relative quantification by LC-MS/MS fragmenting with HCD or ETD. Isobaric stable isotope tags were selected for the quantification as the method can perform relative quantification between four samples in the same run hereby reducing the error rate. The method is also easy to start up and use though on the downside it is quite expensive.

The commercially available isobaric tags the tandem mass tags (TMT) 6-plex was selected as they are able to produce 6 unique reporter ions in ETD. The other commercially available tags (iTRAQ 4-plex) only produces 3 unique reporter ions by ETD fragmentation from the four unique tags due to the different reporter ion formation caused by ETD, described previously (Phanstiel *et al.*, 2008).

The intensities of the reporter ions are important for the quantification as reporter ions with low intensity exhibit poor quantifications providing high error rates. As mentioned previously the reporter ions have to exceed intensities of 1,000 in ETD and 100,000 in HCD to achieve average errors below 10% (Phanstiel *et al.*, 2008). Increasing the collision energy should improve the reporter ion intensities so the fragmentation methods were optimised to achieve the best quantification results while still obtaining good sequence coverage for the peptides.

The quantification was optimised on glycated BASI labelled with TMT tags. The quantification with HCD was performed with normalizing collision energy of 45% and 55%. This resulted in

sequence coverage of 95% for both collision energies. The two methods both identified glycation sites though a few more were identified with 55% normalised collision energy and the hits also appeared to be slightly better for most of the observed modifications with this method, Table 4. The reporter ion intensities are high enough to be used for quantification for all the glycated peptides identified. The intensities of the reporter ions are, however, higher with 55% normalised collision energy and some of the intensities are close to the lower limit when analysed with 45% normalised collision energy. Therefore, 55% collision energy was selected for further quantification experiments with HCD fragmentation.

**Table 4.** Glycations sites observed on TMT labelled glycated BASI by LC-MS/MS fragmenting with HCD fragmentation at normalised collision energy of 45% and 55%. All the hits provided TMT reporter ions with intensities high enough for quantification. The results were obtained using Mascot. Hex: Glycation with glucose, Cbm: Carboxymethyl, () is the number of hits only exceeding the homology threshold in Mascot, Cov: whether or not the glycation site is directly mapped by the MS/MS results, M: missed cleavages.

| | | 45% Normalised collision energy | | | | | 55% Normalised collision energy | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Site** | **Modification** | **Score** | **Expect** | **Hits** | **Cov.** | **M** | **Score** | **Expect** | **Hits** | **Cov** | **M** |
| **R27** | Cbm | 65 | 2.9E-6 | 1 | + | 1 | - | - | - | - | - |
| **K72** | Cbm | 38 | 3.3E-4 | (5) | + | 1 | 30 | 0.0027 | (1) | + | 1 |
| **K72** | Hex | - | - | - | - | | 31 | 0.0062 | (8) | + | 1 |
| **K115** | Cbm | 24 | 9.8E-3 | (2) | - | 1 | 43 | 5.9E-4 | 1 | + | 1 |
| **K115** | Hex | 18 | 0.046 | (4) | - | 1 | 31 | 0.0021 | (3) | - | 1 |
| **K130** | Hex | - | - | - | - | | 20 | 0.017 | (2) | - | 1 |
| **K158** | Hex | 43 | 1.6E-4 | (9) | - | 1 | 59 | 4.9E-6 | 4 (7) | - | 1 |
| **K158** | Cbm | 62 | 6E-6 | 3 | + | 1 | 48 | 3.7E-5 | 1 (2) | + | 1 |
| **K176** | Hex | 24 | 0.01 | (5) | + | 1 | 57 | 9.3E-6 | 3 (5) | - | 1 |

The ETD reaction time can influence the performance of the TMT quantification (Viner *et al.*, 2009). The reporter ion intensities on ETD fragmentation were evaluated with reaction times of 85 and 100 ms. These two times were selected as they previously provided the best result for the identification of glycation sites and these has by others been reported to provide the best quantification results (Phanstiel *et al.*, 2008; Viner *et al.*, 2009). The results for the two ETD methods on TMT labelled glycated BASI revealed that the reaction time of 85 ms provided slightly better results for the identification of glycation sites, Table 5. The reporter ion intensities were high enough to be used for quantification on five of the glycated peptides in both methods. The 85 ms reaction time method did, however, provide more hits with reporter ion intensities high enough for quantification for each of the quantified glycated peptides. Therefore, 85 ms reaction time was selected for further quantification experiments with ETD.

Some of the ETD spectra for the glycated peptides contained the structural ions necessary for peptide identification but did not contain any reporter ions. The lack of reporter ions in otherwise good ETD spectra is also observed in other studies as well as a higher percent of the scans producing low reporter ion intensities. It is proposed that ETD exhibits lower reporter ion intensities due to ETD cleaving randomly along the peptide backbone hereby not preferring the chemical modifications (Phanstiel *et al.*, 2008; Viner *et al.*, 2009).

**Table 5.** Glycations sites observed on TMT labelled glycated BASI using LC-MS/MS fragmenting with ETD at activation time of 85 and 100 ms. The results were obtained by Mascot. Hex: Glycation with glucose, Cbe: Carboxyethyl, Cbm: Carboxymethyl, MG-HI: methylglyoxal-derived hydroimidazolone, () is the number of hits only exceeding the homology threshold, Cov: whether or not the glycation site is directly mapped by the MS/MS results, M: missed cleavages, Quan: hits with reporter ion intensities high enough for quantification.

| | | 85 ms activation time | | | | | | 100 ms activation time | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Site** | **Modification** | **Score** | **Expect** | **Hits** | **Cov** | **M** | **Quan** | **Score** | **Expect** | **Hits** | **Cov** | **M** | **Quan** |
| **R27** | Cbm | 56 | 4.8E-5 | 1 | + | 1 | - | - | - | - | - | - | - |
| **K72** | Cbm | 19 | 0.023 | (3) | + | 1 | - | 26 | 0.01 | (2) | + | 1 | - |
| **K72** | Hex | 61 | 1.7E-5 | 7 (1) | + | 1 | - | 56 | 3.4E-5 | 5 (4) | + | 1 | - |
| **K115** | Hex | 40 | 0.0025 | (5) | + | 1 | 4 | 48 | 2.2E-4 | 1 (2) | + | 1 | 1 |
| **K130** | Hex | 54 | 1.2E-4 | 1 (2) | + | 1 | 3 | 54 | 2.1E-4 | 1 | + | 1 | 1 |
| **K140** | Hex | - | - | - | - | - | - | 29 | 0.045 | (2) | - | 1 | - |
| **R155** | Cbe | 56 | 5.9E-5 | 1 (1) | + | 1 | - | 36 | 0.012 | (2) | + | 1 | - |
| **R155** | Cbm | 50 | 0.0024 | 1 (1) | + | 1 | - | 42 | 8.8E-4 | (1) | + | 1 | - |
| **R155** | MG-HI | 50 | 1.1E-4 | 1 | + | 1 | - | - | - | - | - | - | - |
| **K158** | Hex | 69 | 5.2E-6 | 4 (5) | + | 1 | 4 | 74 | 1.7E-6 | 6 (4) | + | 1 | 5 |
| **K158** | Cbm | 65 | 5.1E-6 | 5 | + | 1 | 3 | 69 | 2E-6 | 3 | + | 1 | 3 |
| **K176** | Hex | 72 | 7E-6 | 8 | + | 1 | 5 | 70 | 1.3E-5 | 5 (2) | + | 1 | 3 |
| **K176** | Cbm | 37 | 7.3E-4 | (1) | + | 1 | - | 28 | 0.028 | (1) | + | 1 | - |

### *3.2.3 Relative quantification of glycation sites on BASI by HCD and ETD*

Relative quantification of glycation sites was performed on BASIG4 and BASIG8 to evaluate the quantification method. BASIG4 and BASIG8 were labelled with TMT126 and TMT128, respectively, and analysed by LC-MS/MS fragmenting by HCD and ETD. The data were then analysed in three different ways: with Mascot Distiller for data file processing searching with the Mascot server and with the program Proteome Discoverer for data file processing using either SEQUEST or Mascot as search engines. The three different data analysis/search methods were conducted to determine whether one outperformed the others or if the results differed in such a way that they should be combined to achieve the best results.

Proteome Discoverer is a program which extracts the relevant MS/MS spectra from the raw data file so it can then be searched in selected search engines. The SEQUEST search engine applies an algorithm that correlates the experimental MS/MS spectra to theoretical spectra of predicted fragment ions derived from the protein database selected for the search. The peptide hits are then scored according to the correlation. The score is generated from a scoring system based on correlations to the main ion series as well as water and ammonia loss (Sadygov *et al.*, 2004).

The overall search results for the three quantifications of glycation sites on the HCD and ETD data reveal that the three methods identify approximately the same number of individual glycations, Table 6. The SEQUEST search did, however, overall quantify a higher number of glycated peptides though this did not result in quantification of more specific glycated peptides than the other two methods for either of the fragmentations.

**Table 6.** Overall comparison of the glycation sites quantified by Mascot with Mascot Distiller for data processing (Mascot), by Proteome Discoverer (PD) using Mascot and by Proteome Discoverer using SEQUEST. The comparison was on data acquired by HCD and ETD fragmentation. Glycated sites are the number of glycated Arg and Lys residues quantified. Unique glycation sites are the number of glycation sites including differences in the specific glycations.

| | Quantification of glycations by HCD | | | Quantification of glycations by ETD | | |
|---|---|---|---|---|---|---|
| | **Glycated sites** | **Unique glycation sites** | **Total glycated peptides** | **Glycated sites** | **Unique glycation sites** | **Total glycated peptides** |
| **Mascot** | 11 | 21 | 132 | 6 | 10 | 77 |
| **PD Mascot** | 11 | 22 | 132 | 6 | 10 | 79 |
| **PD SEQUEST** | 12 | 20 | 153 | 7 | 10 | 94 |

The quantification results for the glycations are very similar between the three different quantification programs on the ETD generated data, Table 7. The relative standard deviation is in general lowest for the quantitative results obtained by Mascot searches with Mascot Distiller.

The search in SEQUEST also assigned the glycation site to the C-terminal Lys or Arg residue for glycated peptides with low sequence coverage as was observed previously for Mascot.

**Table 7.** Quantification of glycation sites between BASI incubated with glucose four (BASIG4) and eight days (BASIG8) obtained by ETD fragmentation where BASIG4 was labelled with TMT126 and BASIG8 with TMT128. The results were obtained by three different MS search programs. Hex: Glycation with glucose, Cbm: Carboxymethyl, MG-HI: methylglyoxal-derived hydroimidazolone, PD: Proteome Discoverer, SD: Standard deviation, RSD: Relative standard deviation, Hits: The number of times the glycated peptides has been quantified, 128/126: relative quantification results between BASIG8 and BASIG4.

| | ETD Mascot | | | | ETD PD Mascot | | | | ETD PD SEQUEST | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Site modified** | **128/126** | **SD** | **RSD** | **Hits** | **128/126** | **SD** | **RSD** | **Hits** | **128/126** | **SD** | **RSD** | **Hits** |
| R27 MG-HI | 1.59 | 0.34 | 0.21 | 4 | 1.58 | - | - | 1 | 1.44 | 0.19 | 0.13 | 2 |
| R61 MG-HI | - | - | - | - | - | - | - | - | 1.43 | - | - | 1 |
| K72 Cbm | - | - | - | - | - | - | - | - | 1.33 | - | - | 1 |
| K115 Hex | 2.17 | 0.29 | 0.13 | 8 | 1.95 | 0.43 | 0.22 | 17 | 1.92 | 0.42 | 0.22 | 34 |
| K115 Cbm | 1.73 | 0.15 | 0.09 | 2 | 1.72 | - | - | 1 | 1.45 | 0.39 | 0.27 | 2 |
| K130 Hex | 1.23 | 0.13 | 0.10 | 8 | 1.20 | 0.22 | 0.19 | 8 | 1.20 | 0.22 | 0.19 | 8 |
| K140 Hex | 1.43 | - | - | 1 | 1.03 | 0.01 | 0.01 | 2 | - | - | - | - |
| K158 Cbm | 1.73 | 0.34 | 0.20 | 14 | 1.66 | 0.32 | 0.19 | 12 | 1.66 | 0.32 | 0.19 | 12 |
| K158 Hex | 1.36 | 0.21 | 0.15 | 18 | 1.49 | 0.30 | 0.20 | 18 | 1.47 | 0.29 | 0.20 | 19 |
| K158 + K176 Hex | 2.83 | 0.50 | 0.18 | 2 | 2.87 | 0.51 | 0.18 | 3 | - | - | - | - |
| K176 Hex | 1.88 | 0.29 | 0.16 | 16 | 1.84 | 0.42 | 0.23 | 13 | 1.86 | 0.46 | 0.25 | 11 |
| K176 Cbm | 1.82 | 0.15 | 0.08 | 4 | 1.62 | 0.29 | 0.18 | 4 | 1.62 | 0.29 | 0.18 | 4 |

The quantification results generated by HCD fragmentation are also very similar between the three different quantification methods, Table 8. There are smaller differences between the results generated by the programs but none of the quantification methods appear to be considerably better than the others. The relative standard deviations for the quantifications obtained by the HCD fragmentation are generally lower than the ones obtained by ETD fragmentation.

The quantification with the three different programs revealed that for the MS/MS data derived by HCD fragmentation the quantitative results for glycation sites were very similar. The quantification results for the data obtained by ETD fragmentation also revealed similar results for the various methods though the relative standard deviations were lower for the quantification results generated by Mascot Daemon. This could indicate that the peak picking is slightly more precise on the low intensity ions in Mascot Distiller than it is in Proteome Discoverer. Therefore,

Mascot Distiller searching in the Mascot server was used for the following quantification experiments.

**Table 8.** Quantification of glycation sites between BASI incubated with glucose four (BASIG4) and eight days (BASIG8) obtained by HCD fragmentation where BASIG4 was labelled with TMT126 and BASIG8 with TMT128. The results were obtained by three different search programs. Hex: Glycation with glucose, Cbe: Carboxyethyl, Cbm: Carboxymethyl, MG-HI: methylglyoxal-derived hydroimidazolone, PD: Proteome Discoverer, SD: Standard deviation, RSD: Relative standard deviation, Hits: The number of times the glycated peptides has been quantified, 128/126: relative quantification results between BASIG8 and BASIG4.

| | HCD Mascot | | | | HCD PD Mascot | | | | HCD PD Sequest | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Site modified** | **128/126** | **SD** | **RSD** | **Hits** | **128/126** | **SD** | **RSD** | **Hits** | **128/126** | **SD** | **RSD** | **Hits** |
| R27 MG-HI | 1.92 | 0.28 | 0.14 | 7 | 1.85 | 0.25 | 0.13 | 9 | 2.12 | 0.11 | 0.05 | 3 |
| R27 Cbe | 2.03 | 0.01 | 0.05 | 2 | 1.89 | 0.15 | 0.08 | 5 | 1.94 | 0.01 | 0.01 | 2 |
| R27 Cbm | 1.69 | - | - | 1 | 1.66 | - | - | 1 | - | - | - | - |
| K72 Cbm | 1.82 | 0.10 | 0.06 | 6 | 1.92 | 0.17 | 0.09 | 10 | 1.98 | 0.15 | 0.07 | 19 |
| K72 Hex | 1.91 | 0.07 | 0.04 | 16 | 1.98 | 0.05 | 0.02 | 4 | 1.95 | 0.08 | 0.04 | 41 |
| R81 MG-HI | 2.74 | 0.08 | 0.03 | 2 | 2.69 | 0.06 | 0.02 | 3 | 2.69 | 0.06 | 0.02 | 3 |
| R81 Cbm | 3.63 | 0.08 | 0.02 | 3 | 3.50 | 0.15 | 0.04 | 5 | 3.44 | 0.18 | 0.05 | 3 |
| R81 Cbe | 2.67 | - | - | 1 | 2.69 | 0.05 | 0.02 | 4 | 2.69 | 0.05 | 0.02 | 4 |
| R106 Cbe | 1.61 | - | - | 1 | 1.59 | - | - | 1 | 1.50 | - | - | 1 |
| R107 Cbm | - | - | - | - | 2.46 | 0.25 | 0.10 | 2 | 2.38 | 0.22 | 0.09 | 3 |
| K115 Hex | 2.17 | 0.14 | 0.06 | 18 | 2.19 | 0.12 | 0.06 | 17 | 2.17 | 0.13 | 0.06 | 20 |
| K115 Cbm | 1.88 | 0.03 | 0.02 | 3 | 1.95 | 0.24 | 0.12 | 4 | 1.95 | 0.24 | 0.12 | 4 |
| R127 MG-HI | 2.09 | - | - | 1 | - | - | - | - | 2.08 | 0.03 | 0.01 | 2 |
| K130 Hex | 1.44 | 0.03 | 0.02 | 3 | 1.43 | 0.04 | 0.03 | 5 | 1.43 | 0.04 | 0.03 | 5 |
| K130 Cbm | 1.26 | - | - | 1 | 1.25 | 0.00 | 0.00 | 2 | 1.25 | 0.00 | 0.00 | 2 |
| K140 Hex | 1.81 | 0.09 | 0.05 | 20 | 1.80 | 0.07 | 0.04 | 9 | 1.79 | 0.07 | 0.04 | 7 |
| R155 MG-HI | 1.61 | 0.01 | 0.01 | 3 | 1.58 | 0.04 | 0.02 | 4 | 1.58 | 0.04 | 0.02 | 4 |
| R155 Cbe | 1.56 | 0.01 | 0.01 | 2 | 1.58 | 0.00 | 0.00 | 3 | - | - | - | - |
| R155 Cbm | - | - | - | - | 1.85 | 0.04 | 0.02 | 3 | 1.83 | 0.03 | 0.02 | 2 |
| K158 Cbm | 1.82 | 0.06 | 0.03 | 17 | 1.78 | 0.06 | 0.03 | 16 | 1.78 | 0.06 | 0.03 | 16 |
| K158 Hex | 1.47 | 0.05 | 0.03 | 10 | 1.45 | 0.04 | 0.03 | 11 | - | - | - | - |
| K176 Hex | 1.95 | 0.08 | 0.04 | 4 | 1.97 | 0.06 | 0.03 | 6 | 1.94 | 0.04 | 0.02 | 4 |
| K176 Cbm | 2.28 | 0.06 | 0.02 | 6 | 2.24 | 0.07 | 0.03 | 8 | 2.24 | 0.07 | 0.03 | 8 |

The quantification ratios between glycations in BASIG4 and BASIG8 reveal that for all the quantified glycation sites the BASIG8 sample contains larger amounts of the glycations than the BASIG4 sample as would be expected as the sample incubated longer time with glucose.

To evaluate the relative quantification of glycation sites with HCD and ETD using TMT a second sample of TMT labelled BASIG4 and BASIG8 was made with the 127 and 130 tags, respectively.

The comparison of ETD and HCD for relative quantification of glycation sites was performed on glycated peptides which were quantified by both HCD and ETD in both of the labelled samples. Seven glycated peptides were quantified by both fragmentation methods in the two labelled samples, Figure 19. From these results it appears that the relative quantification obtained by HCD fragmentation generally is slightly higher than the ETD results of the same sample. For some of the glycated peptides it appeared that ETD and HCD provided comparable quantification results whereas the peptide with K176 glycated with carboxymethyl and the one containing glycated K130 the quantifications appeared to differ significantly between the two fragmentation methods. The quantification results also indicated that for some of the glycated peptides the different labels can influence the quantification results.

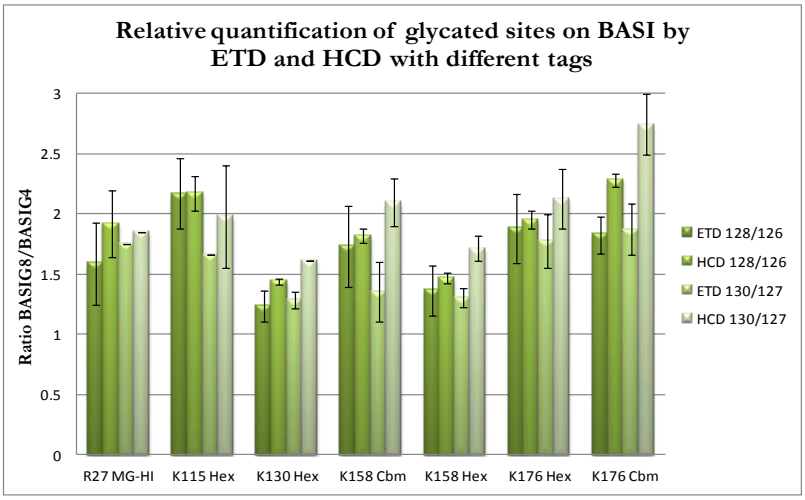



**Figure 19.** Relative quantification of glycated peptides between BASIG4:BASIG8 labelled with both TMT126:TMT128 and TMT127:TMT130, respectively. The quantifications were determined by ETD and HCD fragmentation. The glycated amino acid is stated with the identified glycation. Hex: Glycation with glucose, Cbm: Carboxymethyl. The error bars indicate standard deviation. Columns without error bars are based on single quantification determinations.

Comparison of the quantifications by ETD and HCD for the seven peptides, Figure 20, indicates that for most of the peptides the quantifications obtained by ETD and HCD are similar. Though, for the glycation of K176 with carboxymethyl the quantification obtained by HCD and ETD seems to differ. The comparison also indicates that the same is the case for the peptide with glycated K130.

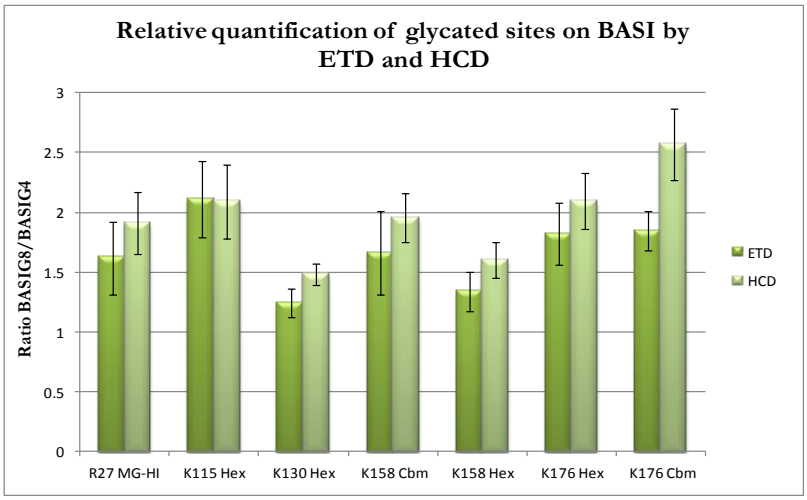



**Figure 20.** Relative quantification of TMT labelled glycated peptides using ETD and HCD. BASIG4 and BASIG8 were mixed 1:1. The data is based on two samples labelled with 127/130 and 126/128 for bASIG4/BASIG8, respectively. The glycated amino acid is stated with the observed glycation. Hex: Glycation with glucose, Cbm: Carboxymethyl. The error bars indicate standard deviation.

A more thorough statistical analysis was performed on quantification results for the peptides containing the following glycated residues: K115, K158, K176 glycated with glucose as well as K158 and K176 modified with carboxymethyl. These glycated peptides were chosen for the statistical analysis as there for all were quantification data on more than one quantification determination for each of the label combinations (130/127 and 128/126) with both

fragmentation methods. This analysis was performed to determine if and how much the three factors the peptide, the label selection or the fragmentation method influence the quantifications. The significance of these factors was assessed using a 3-factor full factorial ANOVA model where outlier detection was used to remove six determinations from the entire dataset.

The statistical analysis revealed that the performance of the fragmentation methods depends on the analysed glycated peptide. HCD and ETD perform similarly on the peptides containing K115 and K176 glycated with glucose but the methods are giving significantly different results on the peptides containing K158 glycated with glucose, K158 glycated with carboxymethyl and K176 glycated with carboxymethyl. The analysis also agreed with the observations from Figure 19 and Figure 20 previously stating that the ratios obtained by HCD are higher than the ones obtained by ETD. The fragmentation methods dependence on the glycated peptides for their performance could at least partly be caused by the extensive pattern of neutral loss which HCD fragmentation of glycated can result in. This neutral loss pattern could interfere with the reporter ions providing variations in the ratio calculations.

The statistical analysis also revealed that the performance of the labels can depend on the glycated peptides. This was observed for the peptide with glycated K115 and the peptide containing K176 glycated with carboxymethyl where the 128/126 and the 130/127 ratios differed significantly. The glycated peptides influence on the performance of the labels could be caused by the neutral losses occurring in HCD described previously. It is also possible that the fragmentation of some peptides forms peptide ions close to or with the same mass as the reporter ions.

In conclusion both ETD and HCD fragmentation can be used for the identification of glycation sites by MS/MS, however ETD identifies more glycation sites than HCD. So ETD is a very efficient fragmentation method for the identification of glycation site. If one also have to identify other non labile modifications or want the highest sequence coverage ETD should be combined with HCD. Therefore, for the characterisation of modifications on BLA both fragmentation methods will be used to achieve high sequence coverage and to identify as many glycations as possible.

Relative quantification by TMT labelling was easy to use. Relative quantification of glycation sites was performed with both ETD and HCD. Even though ETD identified more glycation sites HCD was better at acquiring high reporter ion intensities and it was thus able to quantify more glycation sites than ETD. Different software was evaluated for the quantification and they performed very similar though the relative standard deviation was slightly lower when Mascot Distiller was used to transform the raw MS/MS data for the further search and quantification compared to Proteome Discoverer.

It is, however, important to note that the glycated peptides can influence the performance of the fragmentation methods and the TMT labels in the quantification experiments. It is therefore important in the relative quantification of specifically glycated peptides to use different labels and or label combinations and if possible different fragmentation techniques to evaluate the significance of the quantifications. The method will later on be used for quantification of

glycations between the four partly separated pools of BLA. Here different label combinations will be used as well as the both ETD and HCD for the quantification.

It could be interesting to perform a larger study with more TMT labels and more glycation sites and specific glycations (AGEs) for the evaluation of the influence of the peptide, the specific glycations, the tags and the fragmentation techniques on the relative quantification. This would also provide better understanding of factors to consider in the experimental design. This was, however, not possible in the present project due to time limitations.

## 3.3 Analysis of BLA modifications

The modifications on BLA were investigated using different strategies to determine the extent of the modifications and whether specific modification sites are likely to influence the stability or activity of the enzyme. Intact mass analysis by LC-MS of the partially separated pools of BLA forms varying in pI range was used to determine micro heterogeneity within the pools as well as the differences between the pools. The aim was to determine which modifications are responsible for the multiple forms. For further identification of the specific modifications and the modified sites, in-gel digest followed by LC-MS/MS with both HCD and ETD fragmentation were performed. The method set up in the previous section for relative quantification of glycation sites was used to perform relative quantification of glycation sites between the four pools of partly separated forms of BLA varying in pI range. The glycation patterns produced on BLA were investigated using different carbohydrates in order to assess the susceptibility of BLA to glycation caused by glucose and maltodextrins.

### *3.3.1 Identification of BLA modifications*

The four BLA pools PA – PD (from section 3.1.3) were subjected to intact protein mass analysis to determine the overall extent of modifications for the four pools.

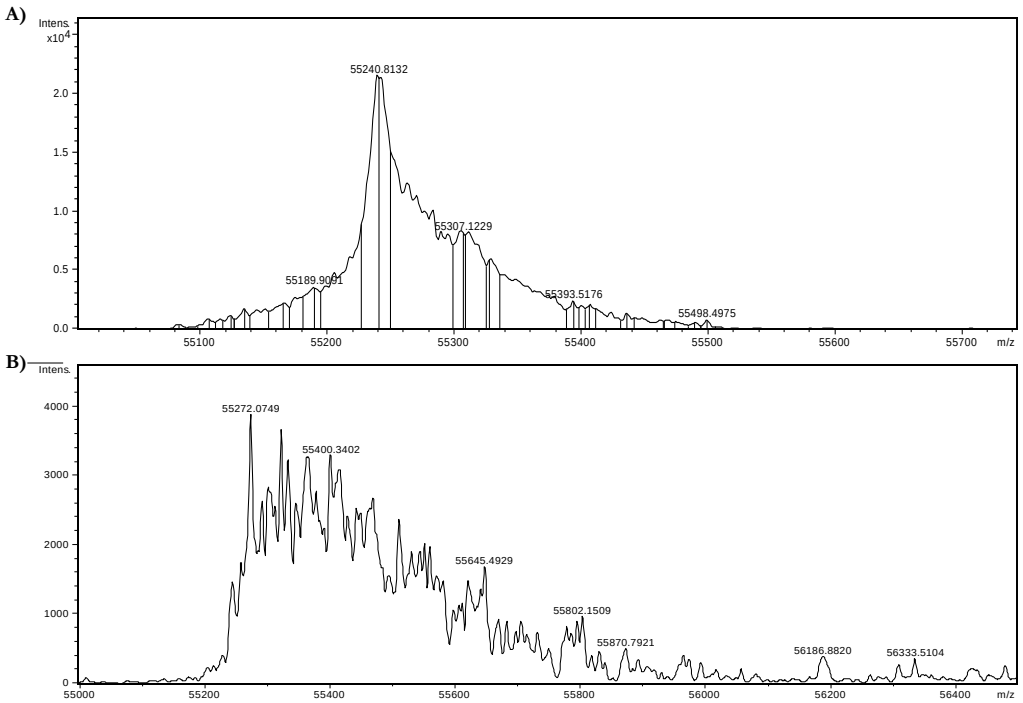



**Figure 21.** Full-length MS by LC-ESI on two of the BLA pools varying in pI range. A) Full-length MS of partly separated BLA pool PA with the main peak corresponding to unmodified BLA. B) Full-length MS of the BLA pool PB is more complex with multiple peaks being BLA species with various modifications.

The intact mass determination for pool PA resulted in a broad peak in the mass spectrum corresponding to the mass of BLA, Figure 21A. These results indicate that the PA sample contain unmodified BLA (theoretical Mw 55239 Da). Though, the resolution of this method is not high enough to separate mass differences of a few Da, so the main peak could also be

deamidated BLA. The peak is broad with smaller shoulders which indicate the sample contains different forms of BLA modified by smaller PTMs. It is not possible to determine exactly which modifications are present but it could be deamidations, oxidations, carbamylations and AGEs.

The results for PB were more complex consisting of multiple peaks, Figure 21B. The first small peak could be unmodified BLA or deamidated BLA. The second large peak corresponds to a mass increase of approximately 31 Da which could be a combination of oxidations and deamidations. So the sample contains multiple forms of modified BLA and maybe a small amount of unmodified BLA.

The signal to noise ratio for the PC and PD samples were quite low and the spectra contained multiple peaks, Figure 22. Both spectra contain the peak of approximately 55272 m/z (also observed in the spectra for PB) followed by multiple other peaks. It is not possible to assign specific modifications to each peak as they can be results of multiple modifications of the BLA. The signal to noise ratio for the PA sample was high but it decreases for the PB, PC and PD samples. As the complexity of the samples increases the signal to noise ratio decreases.

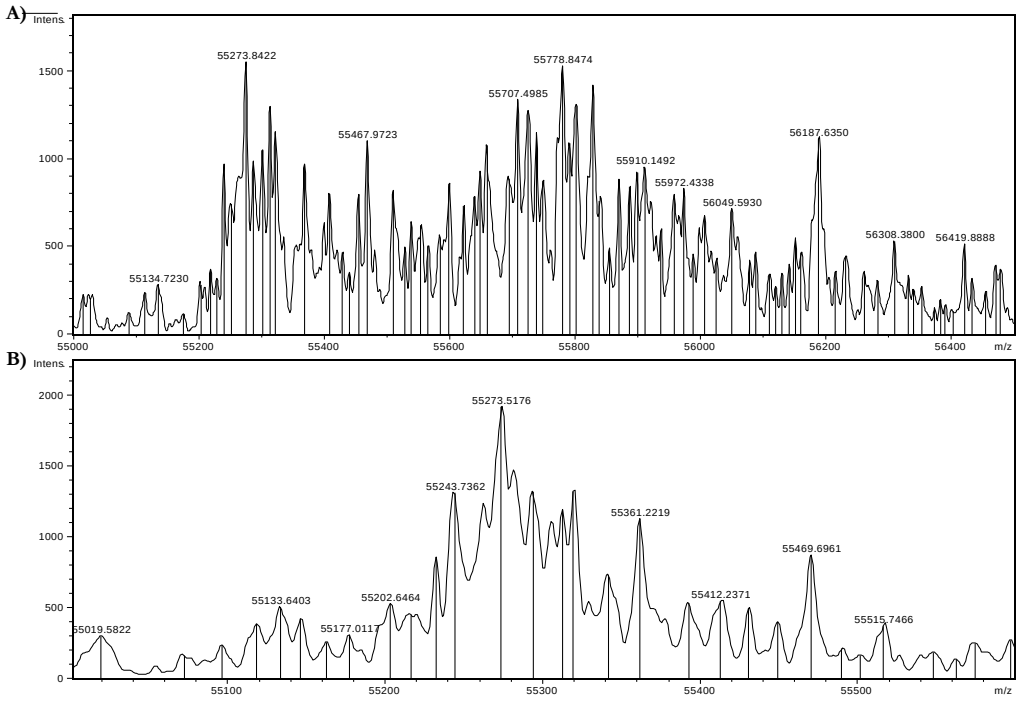



**Figure 22.** Full-length MS by LC-ESI on the two of the BLA pools, PC and PD, varying in pI range. The spectra for both PC (A) and PD (B) are very complex containing multiple peaks. Both samples appear to only contain various species of modified BLA.

In general the observations from the full-length MS analysis support the observations from the IEF gel, Figure 15 (section 3.1.3). The full-length analysis of the samples did not indicate that any of the multiple forms were caused by truncations as no decrease in mass was observed. The PA sample could contain unmodified BLA as well as deamidated forms though other modified forms of BLA also seemed to be present. The PB, PC and PD samples all appeared to only contain modified BLA though the specific modifications could not be identified from the spectra. The

analysis revealed that the PA sample was the least complex of the samples with PB, PC and PD increasing in sample complexity.

To identify the modifications responsible for the multiple forms of BLA in-gel digest was performed from IEF gel followed by LC-MS/MS fragmenting with HCD and ETD. Bands selected for digestion with trypsin and Asp-N are illustrated in Figure 23.

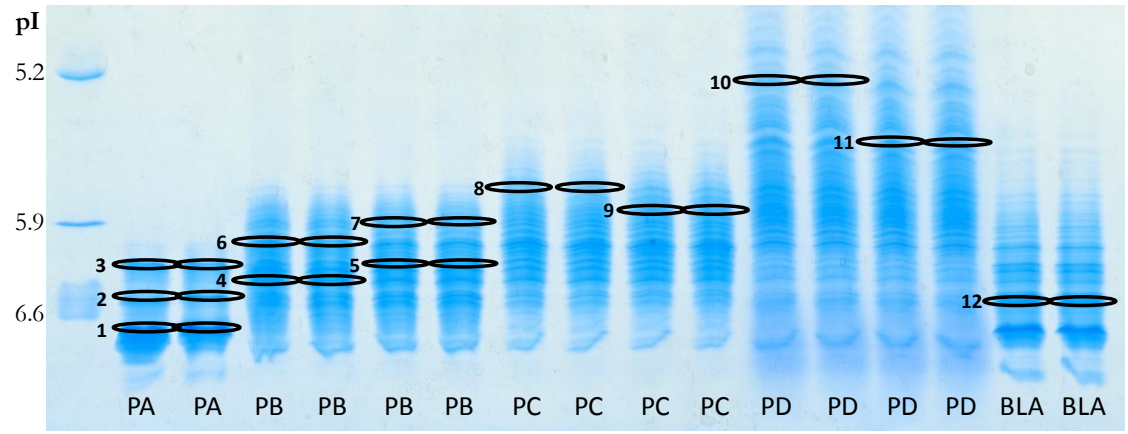



**Figure 23.** IEF of BLA and the four pools of partly separated forms of BLA. The bands selected for in-gel digest with trypsin and Asp-N followed by LC-MS/MS is marked on the gel.

The Mascot searches on the MS/MS data for the samples all gave BLA as the primary significant hit. The total sequence coverage for the selected bands ranged from 48 – 100%, where the coverage for band 1 – 7 and 12 were 90% or above, Table 9. As also observed in the previous section (section 3.2.1) the HCD results provide higher overall sequence coverage compared to ETD fragmentation. For some of the Asp-N samples data was not obtained with HCD due to instrumental problems.

**Table 9.** The sequence coverage obtained for the 12 BLA bands digested by trypsin and Asp-N and analysed with LC-MS/MS fragmenting with HCD and ETD.

| Band | Trypsin (HCD/ETD) | | Asp-N (HCD/ETD) | | Sequence |
|---|---|---|---|---|---|
| | Sequence coverage (%) | Mascot score | Sequence coverage (%) | Mascot score | Coverage (%) |
| **1** | 92/60 | 49666/4783 | -/52 | -/6433 | 94% |
| **2** | 60/52 | 9069/1663 | 80/30 | 2129/3241 | 94% |
| **3** | 52/32 | 5171/1047 | 80/38 | 2746/1124 | 92% |
| **4** | 77/60 | 18837/4148 | 86/44 | 2810/3144 | 94% |
| **5** | 83/43 | 23862/1940 | 75/41 | 1700/4195 | 90% |
| **6** | 78/43 | 20372/2125 | 73/30 | 2388/2917 | 91% |
| **7** | 54/35 | 6875/856 | 79/35 | 1438/663 | 92% |
| **8** | 30/26 | 1033/471 | 70/12 | 672/272 | 80% |
| **9** | 55/29 | 7242/756 | -/33 | -/1824 | 68% |
| **10** | 36/12 | 1167/204 | -/18 | -/728 | 48% |
| **11** | 73/36 | 8108/1683 | -/38 | -/1694 | 77% |
| **12** | 98/57 | 20332/4762 | 78/33 | 1605/3951 | 100% |

The modifications oxidations (Met), deamidations, carbamylations and glycations as well as the advanced glycation end products (AGEs) 3-deoxyglucosone, carboxyethyl, carboxymethyl and MG-HI were identified by Mascot searches for the 12 samples. The modified residues were

mainly located on the surface of BLA, though the modifications were not found in all the possible modification sites on the surface. This indicates that some of the residues on the surface can be favoured for the modifications or that some are protected against modifications possibly by interactions with other amino acid residues or by amino acids in the close vicinity creating an environment unfavourable for the observed modifications. It is also possible that the modifications at these sites were not detected in these experiments or modifications of these sites are present in some of the BLA forms which were not analysed.

Met oxidations and deamidations were observed in all the analysed bands. Especially band nr 1 contained many identified deamidations. There were only observed few Lys and Arg modifications in bands nr 1 and 2. Modifications of Lys and Arg residues were mainly observed in the samples from bands with lower pI values. The data suggest that what appears to be one band on the IEF gel can contain forms modified in different sites. For instance in band 1, which is the band with the highest pI value, multiple deamidation sites were observed. If all these deamidations are present on the same BLA molecule the pI value would be expected to be lower. It is likely that the band contains multiple BLA forms with one or a few modification at different sites. One could thus speculate that one band can contains different BLA forms with approximately the same degree of modifications but at different sites, and that the lower the pI value the higher the degree of modification.

#### *3.3.1.1 Oxidations*

There were identified Met oxidations in all the digested samples, Table 10. Oxidations of M304 and M366 were observed in all the bands except nr 10 and 6, respectively. Oxidations of M8, M15, M256 and M438 were also observed. All the oxidised Met residues are located in domain A except M438 which is located in domain C.

**Table 10.** The Met residues found to be oxidised in the 12 digested BLA samples analysed by LC-MS/MS fragmenting with HCD and ETD. The results were obtained using Mascot. T: trypsin digested, A: Asp-N digested, () is hits exceeding the homology threshold, Covered is whether or not the glycation site is directly mapped by the MS/MS results.

| Site | Hits | Score | Expect | Covered | Identity in band | Homology in band | ID |
|---|---|---|---|---|---|---|---|
| **M8** | 17 (1) | 78 | 1.7e-8 | + | 1, 5, 6, 12 | | $HCD_{AT}$ |
| **M15** | 18 (3) | 87 | 2e-9 | + | 1, 3, 5, 6, 12 | | $HCD_{AT}$ |
| **M256** | 14 (3) | 97 | 2e-10 | + | 1, 4, 12 | 2, 4, 6 | $HCD_{AT}$ |
| **M304** | 65 (18) | 97 | 5.8e-9 | + | 1 – 7, 9, 11, 12 | 8 | $ETD_{AT}/HCD_{AT}$ |
| **M366** | 242 | 70 | 9.7e-8 | + | 1 – 5, 7 – 12 | | $ETD_T$/ $HCD_T$ |
| **M438** | 7 (18) | 60 | 6.2e-6 | + | 5, 11 | 2, 3, 6 – 10 | $ETD_A/HCD_{AT}$ |

Studies show that oxidation of Met can decrease the stability and activity of amylases (Hagihara *et al.*, 2003; Lin *et al.*, 2003). Therefore, some of the oxidations might influence the activity and stability of BLA. Oxidation of M197 and the equivalent position in other *Bacillus* α-amylases is reported to be involved in oxidative inactivation (Hagihara *et al.*, 2003; Hatada *et al.*, 2006; Lin *et al.*, 2003). This position was not found to be oxidised but the sequence coverage for this area was not very good. It is therefore possible that it occurs. Mutations of M15 provide increased stability

and activity (Shaw *et al.*, 1999). Therefore, the oxidation observed at M15 could have an effect on activity and/or stability of these BLA forms.

The main focus of the present project was on the modifications responsible for the pI changes and since Met oxidation does not provide changes in the charge of the residue, these modifications will not be discussed further.

#### *3.3.1.2 Deamidations*

As mentioned previously deamidations can occur during the protein digestion. Therefore, the protein digest was performed in $H_2{}^{18}O$ to differentiate between deamidations present prior to sample preparation (1 Da mass increase) and the ones occurring during sample preparation (3 Da mass increase).

There were identified 15 deamidation sites and additionally 6 sites were indicated by Mascot with scores between the homology and the identity thresholds, Table 11. The deamidations were mainly observed in samples 1 – 4 and 12. The observed deamidations were located in all three BLA domains, though mainly on the surface of BLA, Figure 24. The deamidations at Q9, N326 and Q360 are located in the area around the active site in domain A.

**Table 11.** The observed deamidation site in the 12 digested bands analysed by LC-MS/MS fragmenting with HCD and ETD. The results were obtained using Mascot. T: trypsin digested, A: Asp-N digested, () is hits only exceeding the homology threshold, Covered is whether or not the glycation site is directly mapped by the MS/MS results.

| Site | Hits | Score | Expect | Covered | Identity in band | Homology in band | ID |
|---|---|---|---|---|---|---|---|
| **Q9** | 8 | 61 | 7.8E-7 | + | 1 | | $HCD_T$ |
| **N27** | 2 (8) | 36 | 0.0016 | + | 6 | 1, 2, 3, 5, 12 | $ETD_A$ |
| **Q51** | 1 | 46 | 2.3E-5 | - | 2 | | $HCD_A$ |
| **N172** | 1 | 39 | 0.021 | + | 9 | | $ETD_A$ |
| **N246** | (1) | 37 | 0.07 | + | | 5 | $ETD_T$ |
| **Q264** | 2 (1) | 52 | 5.8E-6 | + | 12 | 1 | $HCD_T$ |
| **N265** | (8) | 39 | 0.00038 | - | | 2, 4, 5, 6, 12 | $HCD_{AT}$ |
| **N272** | 1 | 26 | 0.0026 | - | | 2 | $HCD_A$ |
| **N278** | 1 | 78 | 7.8E-5 | + | 1 | | $ETD_T$ |
| **N280** | (1) | 43 | 0.028 | + | | 1 | $ETD_T$ |
| **Q291** | 3 (1) | 72 | 0.00034 | + | 1, 4 | | $ETD_T$ |
| **Q298** | 4 (2) | 78 | 6.8E-7 | + | 1 | | $ETD_T$ |
| **N309** | 7 (1) | 42 | 6.5E-5 | + | 3, 4, 11 | | $ETD_T/HCD_T$ |
| **N326** | 7 (4) | 92 | 7.8E-8 | + | 1, 2, 9, 12 | 7 | $ETD_A$ |
| **Q360** | 2 | 41 | 8.4E-5 | + | 2 | | $HCD_T$ |
| **Q393** | 1 | 79 | 8.7E-5 | + | 4 | | $ETD_T$ |
| **Q399** | 4 (3) | 77 | 7.3E-5 | + | 12 | 1 | $ETD_T/HCD_T$ |
| **N421** | 1 (2) | 67 | 0.0006 | + | 4 | | $ETD_A$ |
| **Q443** | (5) | 39 | 0.00014 | - | | 2, 3, 7, 9 | $HCD_T$ |
| **N444** | (6) | 39 | 0.00014 | + | | 2, 6, 7, 9 | $ETD_A/HCD_T$ |
| **N463** | 1 (4) | 49 | 1.2E-5 | - | 6 | 3, 5, 7 | $HCD_A$ |

Deamidations are suspected to have a negative effect on the thermostability of BLA (Tomazic *et al.*, 1988a; Tomazic *et al.*, 1988b). However, as some of the observed deamidation sites are located in the area around the active site, these modifications might also influence the enzyme activity.

The residue Q9 is located close to the active site and is proposed to play a role in the substrate binding (Machius *et al.*, 1995). A deamidation in this site could influence substrate binding and

possibly the enzyme's activity. This deamidation was observed in sample 1 with good search results as the score was 61 and expect value of 7.8E-7.

Q360 is located close to the active site and a deamidation in this site might influence the activity and possibly the substrate binding. The search results for this deamidation observed in sample 2 gave a score 41 which is above the identity threshold and expect value of 8.4E-5.

The last deamidation site in the active site area was at N326. This site is located very close to the active site as H327 is proposed to be involved in substrate binding in subsite -1 (Machius *et al.*, 1998). The search results for this site were very good with a score of 92 and expect value of 7.8E-8. The deamidation in this site is likely to influence substrate binding and catalytic activity.

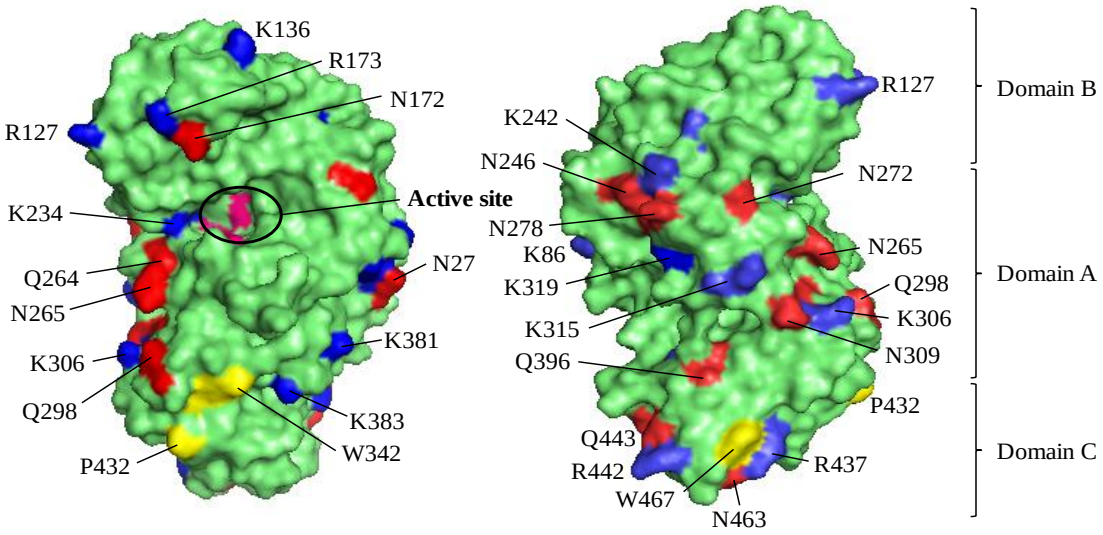



**Figure 24.** Surface of BLA (PDB 1BLI) with deamidation sites shown in red, modified Lys and Arg residues shown in blue (modified by glycations and/or carbamylations) and secondary binding sites shown in yellow. The residue R437 are shown as a modification site but it is also a residue in a secondary surface binding site. (The PyMOL molecular Graphics System, Version 1.3, Schrödinger, LCC)

Deamidation at Q264 was observed in digest samples 12 and 1 with a max score of 52 and expect value of 5.8E-6 exceeding the identity threshold. The site is located at the enzyme surface in the substrate binding site and is proposed to be involved in substrate binding in subsite +3 (Machius *et al.*, 1998). The deamidation in Q264 will therefore very likely affect the substrate binding. The results also indicated that N265 could be deamidated. The search results for this modification only exceeded the homology threshold with a max score of 39 and expect value of 3.8E-4. The sequence coverage of this peptide was quite low and the deamidation site was not covered, though it was observed in several of the digested samples. As the site is located on the surface next to the identified deamidation site Q264 it is, however, possible that N265 also become deamidated. As N265 furthermore is located right beside Q264 in the substrate binding cleft, a deamidation in this site could also influence the substrate binding. The double mutation Q264S-N265Y is shown to significantly increase the activity towards maltoheptosidic substrate and starch (Declerck *et al.*, 2003). Thus, indicating that changes, such as deamidations, in these sites can influence the activity of BLA.

The deamidation at N172 is the only one of the observed deamidation sites located in domain B. This site is relatively close to the active site and situated in the area between domains A and B containing the Ca-Na-Ca triade, which is proposed to play a vital role in the activity and thermostabillity of BLA. Studies show that N172 influences the thermostability of BLA (Declerck *et al.*, 2000; Declerck *et al.*, 2003; Machius *et al.*, 1998). Domain B is also speculated to contain part of an extended substrate binding site from the active site, however, the sites involved are not fully elucidated. The area in which N172 is located is suspected to be involved in this extended substrate binding site (Machius *et al.*, 1998). The deamidation on N172 was observed in sample 9 with expect value 0.021 and score 39 just exceeding the identity threshold.

The site N272 is located in domain A in the area between domains A and B containing the Ca-Na-Ca triade, where it is expected to be involved in the residues stabilising the structure for the metal binding sites (Machius *et al.*, 1998). This deamidation was only found in sample 2 and the results was not very strong exceeding the homology threshold with score 26 and expects 0.026 not mapping the deamidation site in the sequence coverage. Though, a deamidation in this site could affect stability and possibly activity.

The four observed deamidation sites in domain C (Q399, Q443, N444 and N463) are not suspected to influence the Ca binding site in this region, Figure 3, which involves the residues G300, Y302, H406, D407 and D430. Domain C also contains secondary binding sites and N463 is located in the vicinity of one of these sites, so it might influence the function of this site. The deamidation at N463 exceeded the identity threshold with a score of 49 and expect value 1.2E-5.

#### *3.3.1.3 Modification of Lys and Arg residues*

Several Lys and Arg residues were found to be modified by glycations (including AGEs) and/or carbamylations, Table 12. The modified Lys and Arg residues were all located at the surface of BLA, Figure 24. The role of Lys residues on the surface of BLA has been studied by introducing carboxylic groups on these residues by chemical modifications with reagents such as citraconic anhydride (Habibi *et al.*, 2004; Khajeh *et al.*, 2001a; Khajeh *et al.*, 2001b). These studies identify approximately 10 modified Lys residues on BLA, though the exact Lys residues involved are not identified. These modifications affect the thermostability, kinetic activity and substrate specificity (Habibi *et al.*, 2004; Khajeh *et al.*, 2001a; Khajeh *et al.*, 2001b). It is therefore possible that some of the modified Lys residues identified by LC-MS/MS in the present study influence the activity and stability.

One of the studies by Khajeh and co-workers (2001a) identifies 10±1 modified Lys residues on BLA. This corresponds to the number of modified Lys residues identified on the surface of BLA in the present study where 11 out of the 27 Lys residues were found to be modified to some extent. Additionally 8 modified Arg residues were found from a total of 22. As some of the modified Lys and Arg residues observed are located close to the active site, in areas involved in stability as well as close to suspected secondary binding sites, some of the modifications are likely to influence activity, substrate binding and stability of BLA.

Sucrose was used as substrate during the fermentation of BLA. This carbohydrate is not a reducing sugar so it cannot form glycations. However, sucrose can be hydrolysed during sterilisation into fructose and glucose (Petersen *et al.,* 1999). So the glycations could be formed by these sucrose hydrolysis products formed during sterilisation of the substrate.

Urea was not used for the fermentation of the BLA characterised in the present project so exactly how the carbamylations is formed is unknown. It could be that some of the substrates are degraded during the sterilisation forming urea or isocyanate as a by product. Studies at Novozymes have shown that there is a correlation between the amount of ammonium in the fermentation broth and the level of carbamylation (Andersen 2012a).

**Table 12.** Identified and indicated modifications on Lys and Arg residues in the 12 digested bands analysed by LC-MS/MS fragmenting with HCD and ETD. The results were obtained using Mascot. Hex: Glycation with hexose, 3DG: 3-Deoxyglucosone, Cbm: Carboxymethyl, Cbe: Carboxyethyl, Car: Carbamylation, MG-HI: methylglyoxal-derived hydroimidazolone, T: trypsin digested, A: Asp-N digested, () is hits exceeding the homology threshold but not the identity threshold, Covered is whether or not the glycation site is directly mapped by the MS/MS results. Light green shadings for glycations and AGEs.

| Site | Modification | Hits | Score | Expect | Covered | Identity in band | Homology in band | ID |
|---|---|---|---|---|---|---|---|---|
| **R23** | Cbm/Cbe | (4) | 32 | 0.0031 | + | | 1, 6 | $ETD_A$ |
| **R24** | MGHI | (1) | 15 | 0.041 | + | | 5 | $HCD_A$ |
| **K80** | Car | 3 (1) | 45 | 0.0009 | + | 4, 6 | 12 | $ETD_T$ |
| **K88** | Cbe | (1) | 28 | 0.0091 | + | | 6 | $ETD_T$ |
| **R127** | MGHI | 1 (4) | 38 | 0.00015 | + | 5 | 9, 12 | $ETD_T/HCD_T$ |
| **K136** | Cbm | (1) | 23 | 0.025 | + | | 6 | $ETD_T$ |
| **K154** | Cbe | 1 (2) | 43 | 0.0012 | + | 5 | | $ETD_A$ |
| **K154** | Car | (2) | 24 | 0.0067 | - | | 5, 6 | $HCD_A$ |
| **R173** | Car | 1 | 39 | 0.021 | + | 9 | | $ETD_A$ |
| **R173** | 3DG | 7 (7) | 34 | 0.017 | + | 4, 5, 6, 12 | | $ETD_T$ |
| **K234** | Cbm | 4 | 53 | 0.00023 | + | 12 | | $ETD_A$ |
| **K234** | Car | 10 (2) | 48 | 0.00023 | + | 4, 6, 7, 9, 11 | | $ETD_A/HCD_A$ |
| **K237** | Car | 47 (4) | 83 | 6.8e-6 | + | 4 – 9, 11, 12 | | $ETD_{AT}/HCD_A$ |
| **K237** | Cbm/Cbe | 9 (3) | 39 | 0.0011 | + | 6, 7 | 5, 8, 9 | $ETD_{AT}/ETD_T$ |
| **R242** | Cbe/MGHI | 11 (3) | 71 | 2.5e-6 | + | 4, 5, 6, 7, 9 | 8 | $ETD_A/HCD_A$ |
| **R242** | Car | (1) | 29 | 0.015 | + | | 11 | $ETD_T$ |
| **K306** | Car | 52 (7) | 80 | 3.9e-7 | + | 4 – 9, 11, 12 | | $ETD_{AT}/HCD_{AT}$ |
| **K306** | Cbm | 5 (7) | 57 | 0.00033 | + | 5, 12 | 6 | $ETD_T/HCD_T$ |
| **K315** | Car | 34 (3) | 86 | 1.1e-7 | + | 4 – 7, 9, 11, 12 | | $ETD_{AT}/HCD_T$ |
| **K315** | Hex/Cbm | 4 (2) | 50 | 0.0009 | + | 4 – 6, 12 | | $ETD_T$ |
| **K319** | Car | 1 | 35 | 0.037 | + | 4 | | $HCD_A$ |
| **K381** | Cbm | (1) | 29 | 0.0011 | + | | 12 | $HCD_T$ |
| **K383** | Car | 7 (2) | 39 | 0.0035 | + | 6, 11 | 9 | $HCD_T$ |
| **K383** | Cbe | (1) | 20 | 0.011 | + | | 12 | $HCD_T$ |
| **K389** | Cbm/Cbe | 5 (2) | 49 | 0.00057 | + | 4, 9 | 5, 11 | $ETD_T$ |
| **K389** | Car | 1 | 34 | 0.0061 | + | 5 | | $HCD_T$ |
| **K392** | Cbe | 1 | 79 | 8.7e-5 | + | 4 | | $ETD_T$ |
| **R413** | Cbe | 4 (1) | 79 | 0.00012 | + | 5, 6, 12 | | $ETD_T$ |
| **R437** | MGHI | 1 (6) | 30 | 0.001 | + | 11 | 4 – 6, 9, 12 | $ETD_T/HCD_T$ |
| **R442** | MGHI/Cbe/Hex | 1 (8) | 54 | 0.00014 | + | 7 | 2, 6, 9 | $ETD_A/HCD_T$ |
| **R442** | Car | 2 (3) | 37 | 0.0085 | + | 11, 12 | 2, 9, 10 | $ETD_A$ |

Structural studies of BHA reveal three secondary binding sites; a maltose binding site and two glucose binding sites. The maltose binding in BHA involves P434, D432 and W347. The first glucose binding site consist of W439 and W469 and the second glucose binding site is at Y363 (Lyhne-Iversen *et al.*, 2006). The equivalence to these sites can be found on BLA. From the

findings in BHA it can be deduced that the residues W342, P432 and D430 in BLA may compose a maltose binding site, Figure 24. The site K383 is located very close to W342 and this site was found as a likely modification site for both carbamylation and glycation with scores just exceeding the identity threshold for carbamylations. Modification of K383 could influence the secondary binding site.

The two residues R437 and W467 in BLA correspond to the first of the glucose binding sites, Figure 24, observed in BHA. Therefore, the possible glycation observed on R437 is located right in a possible secondary binding site. The score and value expect for this identification is not very high with score 30 and expect value 0.001, just exceeding the identity threshold. The modification is observed in half of the analysed samples. A glycation in this site will most likely interfere with this secondary binding site.

The site K234 was found to be both glycated and carbamylated though carbamylation were most prevalent with score exceeding the identity hreshold. This site is located right in the active site and is proposed to form part of subsite +2 in the substrate binding site (Machius *et al.*, 1995; Machius *et al.*, 1998; Nielsen *et al.*, 2000). Modifications in this site are therefore very likely to influence substrate binding and activity.

K237 is also found to be both glycated and carbamylated again with the carbamylation as the most prevalent. The scores for the carbamylations were very good with score of 83 and expect value of 6.8E-6. The site is proposed to form a salt bridge with D204 connecting domains A and B in the metal ion triade structure. The site is therefore proposed to play an important role in the stability of BLA (Machius *et al.*, 1995; Machius *et al.*, 2003). Mutations in this site support this theory as they result in considerable destabilisation of BLA (Declerck *et al.*, 2000; Declerck *et al.*, 2003). So the modifications in this site could interrupt the salt bridge hereby destabilising BLA.

Both carbamylations and glycations were found on K306 with very good scores for the carbamylations. The site is located in what could be the elongation of the substrate binding which may be involved in binding of longer substrates. Modifications in this site could therefore possibly affect the binding of longer substrates.

Four regions in BLA have been proposed to be of particular importance for BLAs thermostability, one of these regions consists of residues T112 – H142 in domain B, more precisely the part of the area involved in the domains A and B contact and the metal ion triad (Nielsen *et al.*, 2000). In this region the two sites R127 and K136 appear to be glycated. The glycation of K136 was only indicated as the score only exceeded the homology threshold whereas the glycation in R127 was identified with score 38 and expect value 0.00015. R127 forms a salt bridge with D121 which contributes to the thermostability of BLA (Declerck *et al.*, 2000). A modification of R127 is therefore anticipated to influence the thermostability of BLA. The specific influence of a modification on K136 is more uncertain as it merely is located in an area which is important for the stability. The specific site might not in itself be important.

Glycation and carbamylation were determined in the site R173. The scores were not particularly high for these identifications though for both modifications the scores are above the indicated identity threshold. R173 is located right next to N172 on the surface of domain B and as

mentioned previously N172 influences the thermostability of BLA (Declerck *et al.*, 2000; Declerck *et al.*, 2003; Machius *et al.*, 1998) and this area in domain B could be involved in substrate binding (Machius *et al.*, 1998). It is therefore possible that the modifications on R173 affect the stability and/or the substrate binding.

In conclusion multiple modification sites were identified on BLA. A few Met oxidations occurred. Multiple deamidations of Asn and Gln residues as well as glycations and carbamylations of Lys and Arg residues were identified which all can be responsible for the creation of the multiple forms of varying pI values. The majority of modifications were observed on the surface of BLA though a few deamidations were present inside the structure in the area around the active site. Both deamidations and modifications of Lys residues have in other studies shown to have a negative effect on the stability of BLA. Several residues related to the stability of BLA were also found to be modified especially the modification sites R127, N172, K237 and N272 are likely to be important for the enzymes thermostability. However, there were also observed multiple modification sites in the active site area likely to influence activity and substrate binding. The modifications in the sites Q9, Q360 and N326 located close to the active site are likely to result in changes in the enzyme activity. The modifications on K234 and Q264 which are part of subsites in the substrate binding likely influences the substrate binding and possible hereby also the activity. The sites N265 and N172 are also located in close proximity to the substrate binding sites and K306 possibly in the elongation of the substrate binding. So the modifications in these sites are likely to affect substrate binding. BLA also contains secondary binding sites and the modification sites K383 and R437 are located so close to such sites that they could influence these secondary binding sites most likely modifications such as glycations would decrease the efficiency of the site or completely inactivate the binding site. As the exact role of these site are not fully understood the effect of the modifications in these sites are not known. Though if the sites function as initial substrate recognition sites or substrate binding for longer substrates glycations at the secondary binding sites could decrease the efficiency of the amylase.

Overall several modifications and multiple modification sites were identified or indicated. Many of the modifications were found in sites which are important for BLAs stability, activity and substrate interaction. These modifications can therefore have a large effect on the function of BLA, especially for the BLA molecules with multiple modifications. Further investigations of these sites can provide important knowledge concerning the importance of these sites and the effect of the modifications. Site specific mutations of the modification sites can provide this information and this will be done in section 3.4, where the effect of the modifications in some of the modification sites are analysed.

### *3.3.2 Relative quantification of glycation sites in BLA pools PA – PD*

Relative quantification of glycation sites were performed on the four BLA pools. Two samples were prepared where the BLA pools were digested with trypsin, each pool was labelled with a TMT (isobaric stable isotope tag) and the samples were mixed. The pools were labelled with

different tags in the two samples. The two TMT labelled samples were analysed by LC-MS/MS fragmenting with ETD and HCD.

The data for the HCD fragmentations resulted in sequence coverage on BLA up to 90% with Mascot score of 11031. The Mascot search of the HCD data did, however, not reveal any glycation sites.

The results from the ETD fragmentation also resulted in BLA as primary significant hit with sequence coverage of 97% and Mascot score of 8682. Four glycation sites were observed by the ETD fragmentation, Table 13, all the glycations were the AGE MG-HI. The reporter ions for the glycated peptides were, however, far too low to perform relative quantification between the four pools.

Of the four observed glycation sites only R242 was found in the previous investigations. The four sites were all located on the surface of BLA. The site K436 is located right next to the secondary binding site containing of R437 and W467. Glycation on K436 could influence the function of this carbohydrate binding site.

**Table 13.** The four BLA pools PA – PD were digested by trypsin labelled with TMT, mixed and analysed by LC-MS/MS fragmenting with ETD. The following glycations were observed by Mascot searches of the MS/MS data. Quan: the quantification results obtained for the glycated peptides, MG-HI: methylglyoxal-derived hydroimidazolone, () is hits exceeding the homology threshold, Cov: is whether or not the glycation site is directly mapped by the MS/MS results.

| Site | Modification | Hits | Score | Expect | Cov | Quan | Peptide | Sequence |
|---|---|---|---|---|---|---|---|---|
| **K76** | MG-HI | 1 (3) | 56 | 3.4E-5 | + | – | 75 – 88 | R.TKYGTKGELQSAIK.S |
| **R242** | MG-HI | (1) | 29 | 0.013 | + | – | 238 – 249 | K.FSFLRDWVNHVR.E |
| **R375** | MG-HI | 1 (8) | 39 | 0.0023 | + | – | 371 – 381 | K.GDSQREIPALK.H |
| **K436** | MG-HI | (1) | 25 | 0.047 | – | – | 414 – 437 | R.EGDSSVANSGLAALITDGPGGAKR.M |

Overall the ETD fragmentation of the samples provided poor reporter ion intensities with only a few exceeding the ion intensity threshold of 1,000, which is required to get an error below 10%. The HCD fragmentation provided better results though several of the reporter ions were below the threshold and the intensities in general were lower than in the previous experiments on BASI. In the quantification experiments on BASI (section 3.2.3) only a few reporter ions were below the intensity threshold for each sample. This indicates that increasing the number of tags in a sample decreases the reporter ion intensities. It is possible that further optimisation of the MS/MS method could improve the reporter ion intensities. The quantification could possibly also be improved by quantifying two pools at a time as this would increase the amount of each reporter ion in the sample. As described previously ETD fragments randomly along the peptide backbone without preference for chemical modifications, therefore only applying two TMT in a sample is likely to increase the intensities of the ions as they are in higher concentrations. This would most likely also result in identification of more glycation sites as they most likely are such a small part of the sample that they are not detected in the LC-MS/MS analysis when all the four pools are mixed. It has also been proposed that the charge state influences the formation of reporter ions by ETD so selecting precursors of charge state four or above could improve the ETD quantification of TMT labelled peptides (Viner *et al.*, 2009). Of the four glycated peptides

identified by ETD only the peptide glycated in R242 had charge state four. The other three glycated peptides had charge state 3, though the quantification results were not better for the peptide with charge state four.

In conclusion, the glycation sites in the four BLA pools varying in pI range were not quantified. The quantification method needs to be further optimised to perform the relative quantification. This might be done by only quantifying two samples at a time as this most likely will increase the amount of glycated peptides and the amount of each reporter ion in the quantification sample. The relative quantification by ETD might also be optimised by selecting peptides with charge state four or above as this has been proposed to increase the reporter ion intensities.

### *3.3.3 Forced glycations of BLA*

The partly purified pool of BLA with the fewest forms in it (PA) was incubated with glucose, maltose, maltotriose, maltoheptaose, soluble starch and in a blank without hexoses, Figure 25.

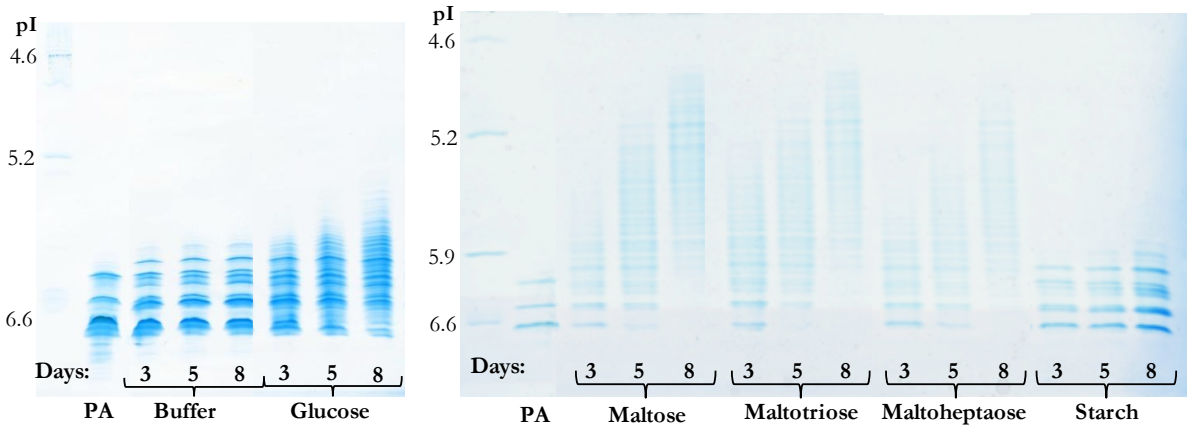



**Figure 25.** IEF gels of PA incubated at 40°C pH 9.5 up to eight days in buffer (blank) and in different carbohydrate solutions.

In the absence of reducing sugars (i.e. in the blank) a few extra forms of BLA with lower pI values appeared over time. These extra forms are most likely created by deamidations. Incubation with glucose, maltose, maltotriose and maltoheptaose resulted in multiple bands appearing over time and the forms with the highest pI values disappearing. Maltose, maltotriose and maltoheptaose appeared to be more reactive than glucose as incubation with these rapidly resulted in multiple forms with very low pI values and the initial forms completely disappearing. Incubation with starch resulted in the same pattern as the incubation with buffer, so glycation does not seem to occur in this incubation.

The forced glycations revealed that glycation also occur with longer reducing sugars such as maltoheptaose but at some point the longer chains become less reactive as starch does not seem to perform glycation or very slow if it does. This is unexpected as it would be expected that the amylase degraded some of the starch to smaller and more reactive carbohydrates and therefore some glycations could be observed on the IEF gel.

It also appeared that glucose was less reactive than maltose and the longer maltosaccharides. A study of the reactivity of 15 monosaccharides with amino groups shows that glucose is the least

reactive. This decreased reactivity is dependent on the extent to which the carbohydrate exists in the open conformation (Bunn *et al.*, 1981). The incubation experiment indicated that glucose also is the least reactive of the shorter carbohydrates (up to 7 glucose units). Thus, indicating that the reactive glucose units in maltose to maltoheptaose tend to be more in the open conformation than glucose itself. Therefore, it is probably advantageous to use glucose rather than the larger carbohydrates and dextrins as carbon source in the fermentation to reduce the glycation rate.

The glucose and maltose incubated PA samples were analysed by LC-MS to determine the extent of glycation at different time points. For MS analysis of glucose incubated samples, Figure 26, the sample incubated three days showed a peak corresponding to unmodified BLA at 55236 m/z followed by three larger peaks. The larger peaks correspond to the addition of glucose (+162 Da). In between the peaks corresponding to glycated BLA smaller peaks appear which could be AGEs or other smaller modifications. The presence of 4 – 5 glycations can be observed in the spectrum for the sample. The accuracy of the MS analysis on the instrument used can vary up to 5 Da depending on protein size and sample complexity. This explains the inaccuracies observed in the m/z determinations.

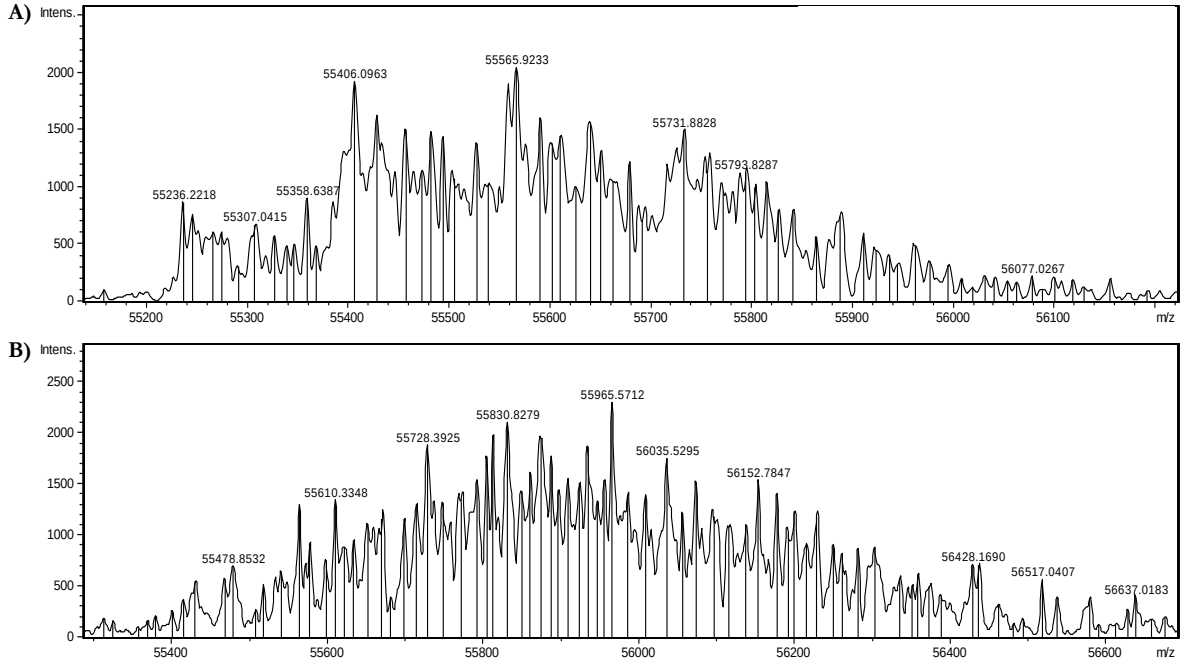



**Figure 26.** Full-length MS of PA incubated with glucose at 40°C pH 9.5. A) Three days incubation. B) five days incubation.

For the sample incubated five days in glucose, no or very small amounts of unmodified BLA appeared to be present, Figure 26B. The spectrum was more complex with multiple peaks and no specific pattern occurred. This indicates that multiple different modifications occur on BLA, most likely combinations of glycations, AGEs, oxidations and deamidations. The peak at 56428 m/z could, however, indicate that some of the BLA molecules are glycated up to seven times. Due to the complexity of the spectrum it is difficult to obtain more information from the spectrum.

Three days incubation with maltose resulted in a much more complex MS spectrum, Figure 27, than three days incubation with glucose. The spectrum revealed that the sample only contain low amounts of the unmodified BLA which also correspond to the observation on the IEF gel. The peak at 56573 m/z could be BLA with four glycations by maltose. So the glycation level does not seem to be much different from the glucose incubation which also indicated up to four glycations. It just appears that more BLA becomes modified by maltose. These observations again correspond with the observations from the IEF gel that maltose is more reactive than glucose. MS on samples incubated five and eight days in maltose did not provide interpretable spectra possibly due to the complexity of these samples.

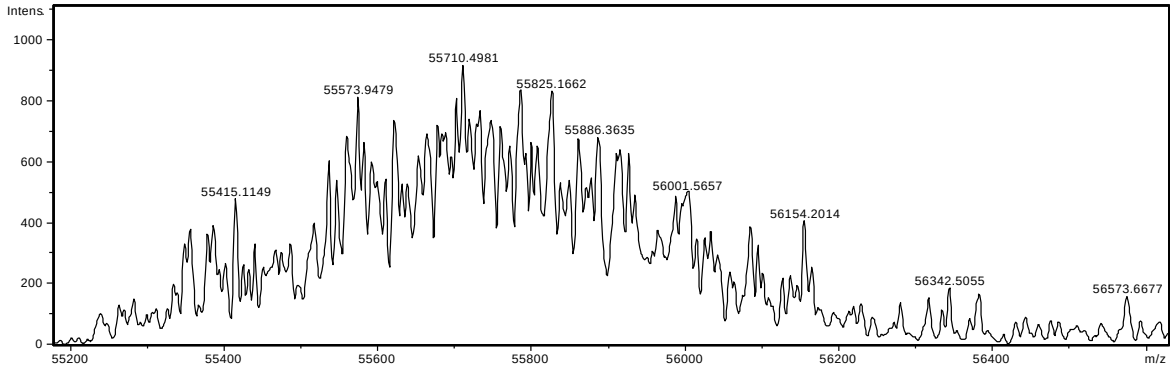



**Figure 27.** Full-length MS of PA incubated three days with maltose at 40°C pH 9.5.

The samples incubated in glucose three and eight days along with the sample incubated 3 days with maltose were selected for trypsin digest and LC-MS/MS fragmenting with ETD to determine the glycation sites, Table 14. The Mascot search results for all three samples resulted in significant hits for BLA with sequence coverage of 66% (score 1852), 60% (score 5642) and 62% (score 3065) for PA incubated in glucose three and eight days and in maltose three days, respectively.

In the sample incubated three days with glucose three glycation sites were identified. There were identified six glycation sites in the sample incubated eight days with glucose and additionally 5 sites were observed with scores only exceeding the homology threshold. These observations support the previous observations that the sample becomes more modified with eight days incubation in glucose. In the sample incubated eight days with glucose the AGE carboxyethyl was also found. LC-MS/MS of the sample incubated three days with maltose only revealed one glycation site. This site was also identified in the glucose incubated samples.

The trypsin digests found 11 glycation sites where two of these had not been observed in the MS/MS analysis of the multiple forms of BLA. The sites R93 and K370 were only identified with scores above the homology threshold in the searches. The sites are located on the surface of BLA in domain A away from the active site, the metal ion triad and substrate binding site.

**Table 14.** Glycations found by LC-MS/MS on trypsin digested PA sample incubated three and eight days with glucose and three days with maltose. The results were obtained by Mascot searches. Hex: Glycation with glucose, Hex(2): Glycation with maltose, Cbe: Carboxyethyl, () is hits only exceeding the homology threshold, Covered is whether or not the glycation site is directly mapped by the MS/MS results.

| Site | Modification | Hits | Score | Expect | Covered | Peptide | Sequence |
|---|---|---|---|---|---|---|---|
| **PA glucose 3 days** | | | | | | | |
| **K88** | Hex | 3 | 61 | 1.7E-5 | + | 81 – 93 | **K.GELQSAIKSLHSR.D** |
| **K381** | Hex | 1 | 21 | 0.033 | + | 376 – 389 | **R.EIPALKHKIEPILK.A** |
| **K392** | Hex | 2 | 68 | 1E-6 | + | 392 – 413 | **R.KQYAYGAQHDYFDHHDIVGWTR.E** |
| **PA glucose 8 days** | | | | | | | |
| **K80** | Hex | (2) | 40 | 0.0036 | + | 77 – 88 | **K.YGTKGELQSAIK.S** |
| **K88** | Hex | 2 | 88 | 3.4E-8 | + | 77 – 93 | **K.YGTKGELQSAIKSLHSR.D** |
| **K88** | Hex | 6 (1) | 69 | 1.5E-5 | + | 81 – 93 | **K.GELQSAIKSLHSR.D** |
| **R93** | Cbe | (2) | 34 | 0.0026 | + | 89 – 106 | **K.SLHSRDINVYGDVVINHK.G** |
| **K306** | Hex | (2) | 31 | 0.0057 | + | 306 – 319 | **R.KLLNGTVVSKHPLK.S** |
| **K315** | Hex | (1) | 24 | 0.018 | + | 306 – 319 | **R.KLLNGTVVSKHPLK.S** |
| **K370** | Hex | (1) | 24 | 0.042 | + | 355 – 375 | **R.ESGYPQVFYGDMYGTKGDSQR.E** |
| **K381** | Hex | 2 | 42 | 0.0025 | + | 376 – 389 | **R.EIPALKHKIEPILK.A** |
| **K383** | Hex | 1 | 30 | 0.054 | + | 382 – 389 | **K.HKIEPILK.A** |
| **K389** | Hex | 1 | 33 | 0.036 | + | 384 – 391 | **K.IEPILKAR.K** |
| **K392** | Hex | 7 | 102 | 3.9E-9 | + | 392 – 413 | **R.KQYAYGAQHDYFDHHDIVGWTR.E** |
| **PA maltose 3 days** | | | | | | | |
| **K88** | Hex(2) | 3 (1) | 54 | 0.00021 | + | 81 – 93 | **K.GELQSAIKSLHSR.D** |

Overall the forced glycations revealed that maltose, maltotriose and maltoheptaose are more reactive than glucose in the formation of glycations. Therefore, glucose is to prefer as carbon source during fermentation compared to longer reducing sugars or dextrins, since it reduces the overall glycation rate. Two more glycation sites were identified by peptide mapping and MS/MS of glycated BLA. The two sites R93 and K370 are as the other identified glycation sites located on the surface of the enzyme but they are not suspected to influence the amylase activity or stability.

### 3.4 Characterisation of BLA variants

On the basis of the modifications and modification sites identified in the previous section new BLA variants were produced to determine the possible impact of some of the modifications on BLA. A lot of the studies performed on the BLA structure attempted to elucidate the structural features related to the thermostability (Declerck *et al.*, 1997; Declerck *et al.*, 2000; Declerck *et al.*, 2003). The focus of the present project is on BLA as a detergent enzyme. The temperatures used in wash today are relatively low with 30 – 40°C as common wash temperatures in Europe. Therefore, the focus is on the activity and wash performance rather than thermostability. Time did not allow for characterisation of variants by site specific mutation at all of the identified modification sites. The modification sites chosen for mutagenesis were therefore selected based on their possible impact on the active site or on secondary binding sites. The effect of the mutations was characterised by determining the kinetic parameters $k_{cat}$, $K_M$, $k_{cat}/K_M$ and the wash performance. The pI values of the variants were analysed by IEF and the stability was tested by DSC.

The variants produced and the possible significance of the selected sites are listed below:

**Q9E** – is located behind the active site and may affect activity, pH profile and substrate binding.

**N172D** – is located close to the metal ion triad in domain B and may affect substrate binding as well as thermostability.

**Q264E** – is located in the substrate binding site and is involved in substrate binding in subsite +3, thus possibly affecting substrate binding and thus activity.

**N265D** – is located in the substrate binding cleft and may therefore affect substrate binding.

**Q360E** – is located behind the active site and may affect activity, substrate binding and pH profile.

**K306M** – may be located in an elongation of the substrate binding cleft and perhaps affecting binding of longer substrates.

**K383M** – is located close to a secondary binding site and may affect substrate binding.

**R437L** – is located right in a secondary binding site and may affect substrate binding.

The possible deamidation sites selected were changed to the corresponding acid to mimic the deamidation (Q→E, N→D), hereby mimicking the influence of the specific deamidation. For the selected glycation sites Lys was substituted with Met and Arg with Leu to evaluate the effect of removing the glucose reactive site. Met and Leu were chosen because of their neutral side chain and as being close in side chain size to Lys. The variant R437M was also produced, but the expression of this variant was very low and it could therefore not be characterised.

The pI values of the BLA variants were determined on an EIF gel, Figure 28. All the mutations result in substitutions to an amino acid with a lower pI value, therefore one should expect a

decrease in pI value for all the variants. This is, however, not the case. It appears that the same amino acid substitutions at different parts of the molecule can result in different changes in pI value or no change at all. The Q264E, N265D and Q360E all result in the same decrease in pI value. The pI decrease for Q9E is slightly lower than for the other three, whereas the substitution N172D does not give a change in pI value. The decrease in pI value for K383M and R437L appears to be the same as for Q9E; for K306M the pI value is unaffected. The changes observed for some of the single mutations seem to correspond to bands observed in the BLA sample. Several microenvironmental factors can influence the pKa values of the amino acid residues in proteins such as electrostatic and hydrophobic effects, hydrogen bond formation and helix dipole interactions (Li *et al.*, 2005; Nielsen *et al.*, 2001). This can explain why the same amino acid residue substitution or chemical modification at different positions of the amylase results in different shifts in pI.

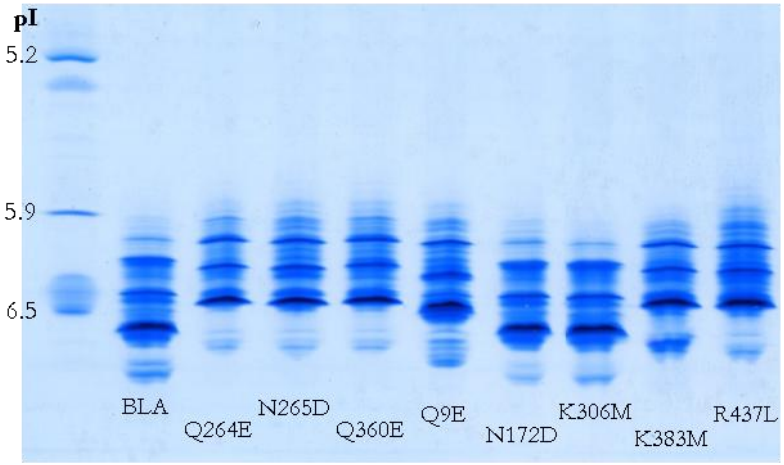



**Figure 28.** IEF gel of the 8 BLA variants.

The observed changes in pI values correlate with the hypothesis made in the previous section that one band can contain multiple BLA forms with one or a few modification at different sites, as well as with the hypothesis that a band at lower pI value has more modifications on the forms present in that band.

The stability of the variants was determined by DSC, Table 15, (DSC results Appendix I Figure 7). BLA is a very stable enzyme as the DSC results also show with a $T_{max}$ value of 103.2°C. The same is evident for the variants as only $T_{max}$ values of Q9E, N172D and Q264E are slightly lower by 2.6, 4.3 and 4°C, respectively. A drop of ≥3 – 4°C is generally considered as being significant (Lund 2011).

**Table 15.** $T_{max}$ values for the BLA variants determined by DSC.

| | BLA | Q9E | N172D | Q264E | N265D | Q360E | K306M | K383M | R437L |
|---|---|---|---|---|---|---|---|---|---|
| **$T_{max}$ (°C)** | 103.2 | 100.6 | 98.9 | 99.2 | 102.9 | 102.7 | 103.4 | 103.2 | 103.1 |

As previously mentioned N172 is suspected to be involved in the thermostability of the enzyme since it is located in domain B in the area interacting with the Ca-Na-Ca triade. The introduction of a deamidation in this site seems to have a small negative effect on the stability of the enzyme.

Q9 and Q264 are located in the active site area and changes in these sites were mainly expected to influence the substrate binding and activity of the enzyme. The influence of a deamidation in Q9 on the stability is almost insignificant. Declerck and co-workers (2003) showed that the double mutation Q264S-N265Y increases the stability and has increased activity. This agrees with the DSC results which revealed that Q264 can influence stability. By contrast the N265D mutation did not seem to affect the stability. The effect of the Q264N mutation was however the opposite of the observations made by Declerck and co-workers (2003) for the Q264S-N265Y mutant. This may indicate that deamidation at Q264 (i.e. Q264E) has a negative effect on the stability. The mutation Q264S, substituting the possible deamidation site with another polar uncharged amino acid, increases the stability.

The kinetic parameters $k_{cat}$, $K_M$ and $k_{cat}/K_M$ were determined for the BLA variants on a short G7 substrate, Table 16, and on a long amylose DP440 substrate, Table 17. Alterations in the kinetic parameters on G7 substrates occur for the Q264E, Q360E and Q9E variants. Q264E markedly decreased in $K_M$ which resulted in an increase in $k_{cat}/K_M$. Both $k_{cat}$ and $K_M$ increased for Q360 which meant that $k_{cat}/K_M$ overall was unaffected by the mutation. The Q9E mutation had a negative effect on $K_M$ as this value almost doubled which decreased $k_{cat}/K_M$. The $k_{cat}$ value was slightly higher for both N265D and R437L but the variation was quite high for R437L so this difference was not a significant.

**Table 16.** The kinetic parameters $k_{cat}$, $K_M$ and $k_{cat}/K_M$ for the BLA variants and BLA determined at 30°C pH 7.0 on maltoheptaose substrate.

| Sample | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ ($s^{-1}$ $mM^{-1}$) |
|---|---|---|---|
| BLA | 65 ± 12 | 0.011 ± 0.003 | 5922 (± 1779) |
| Q9E | 63 ± 14 | 0.027 ± 0.008 | 2394 (± 847) |
| N172D | 70 ± 9 | 0.013 ± 0.005 | 5462 (± 1926) |
| Q264E | 60 ± 5 | 0.005 ± 0.001 | 13013 (± 2887) |
| N265D | 83 ± 5 | 0.011 ± 0.002 | 7277 (± 1584) |
| Q360E | 155 ± 24 | 0.026 ± 0.005 | 5996 (± 1388) |
| K306M | 70 ± 4 | 0.010 ± 0.001 | 7335 (± 1059) |
| K383M | 69 ± 10 | 0.010 ± 0.002 | 7133 (± 2044) |
| R437L | 81 ± 18 | 0.010 ± 0.003 | 8153 (± 3051) |

The kinetic parameters obtained on amylose revealed some of the same tendencies as the determinations on the G7 substrate. The changes were primarily observed for Q264E, Q360E and Q9E. For Q264E and Q360E the changes were more or less the same as for the G7 substrate though the decrease in $K_M$ for Q264E was not as pronounced on amylose as on the G7 substrate. The changes in kinetic parameters were, however, different for Q9E. On amylose $k_{cat}$ decreased, whereas on the G7 substrate $K_M$ increased. The different changes observed on the different substrates for Q9E did, however result in a decrease in the catalytic efficiency for both

substrates. R437L had a higher $k_{cat}$ value on amylose, but the variation for this determination was quite high. It is therefore difficult to determine if the difference is significant.

**Table 17.** The kinetic parameters $k_{cat}$, $K_M$ and $k_{cat}/K_M$ for the BLA variants and BLA determined at 30°C pH 7.0 on amylose DP440.

| Sample | $k_{cat}$ ($s^{-1}$) | $K_M$ (mg/ml) | $k_{cat}/K_M$ ($s^{-1}$ $mg^{-1}$ ml) |
|---|---|---|---|
| BLA | 200 ± 8 | 0.140 ± 0.03 | 1455 (± 286) |
| Q9E | 117 ± 12 | 0.112 ± 0.04 | 1099 (± 363) |
| N172D | 201 ± 6 | 0.129 ± 0.01 | 1555 (± 132) |
| Q264E | 228 ± 15 | 0.089 ± 0.02 | 2632 (± 584) |
| N265D | 232 ± 28 | 0.128 ± 0.03 | 1852 (± 483) |
| Q360E | 416 ± 17 | 0.261 ± 0.06 | 1630 (± 363) |
| K306M | 213 ± 38 | 0.125 ± 0.04 | 1740 (± 602) |
| K383M | 241 ± 33 | 0.139 ± 0.02 | 1743 (± 349) |
| R437L | 307 ± 67 | 0.137 ± 0.02 | 2249 (± 707) |

Overall the kinetic studies revealed that the mutation Q264E located in subsite +3 in the substrate binding site resulted in higher affinity for both substrates so deamidation in this site could improve the substrate binding and herby the catalytic efficiency at 30°C pH 7. These data correlates with other studies stating that position 264 is involved in substrate binding (Declerck *et al.*, 2003; Machius *et al.*, 1998).

The Q360E modification appeared to increase both $k_{cat}$ and $K_M$ so a deamidation at Q360 may result in lower substrate binding, but higher rate of substrate catalysis, though the modification did not appear to influence the overall catalytic efficiency at these conditions.

The mutation Q9E decreased the catalytic efficiency strongly indicating that a deamidation in this site has a negative effect on the efficiency of the enzyme. The modification appeared to give lower substrate affinity towards the smaller substrate and lower substrate turnover number on long substrate.

Removing the reducing sugar reactive sites did not have much influence on the kinetic parameters. Although removing the glucose reactive site at R437 showed a small positive effect on $k_{cat}$ and the catalytic efficiency but the variations on these results were quite large.

The wash performance of the variants were analysed on rice starch cotton patches at 40°C in a liquid and a powder detergent. European wash conditions were chosen for the variants which include running the wash at 40°C. The powder detergent results in pH 10 in the wash and the liquid detergent in pH 8. The difference between the wash results has to be more than 1-1.5 delta remission units to determine a difference in wash performance.

The wash performance for the BLA variants in liquid detergent revealed that all the variants had improved wash performance, Figure 29. The wash performance of Q360E was approximately 1 delta remission higher than BLA, therefore it could not be concluded that this form had an increased wash performance. The N265D variant stood out from the rest with the best wash performance. The remaining variants had a better wash performance than BLA and Q360E but not as good a performance as N265D. The fact that more or less all the variants except Q360E increased in wash performance at these conditions was unexpected. Though the wash results for

the partly separated forms of amylases (section 3.1.2) did indicate that few modifications could increase the wash performance.

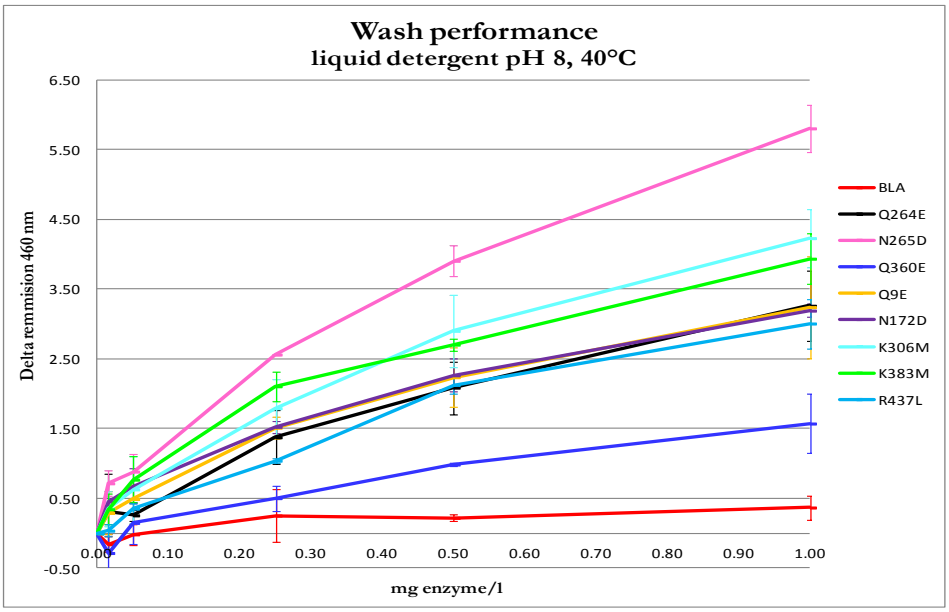



**Figure 29.** The wash performance of the BLA variants on rice starch cotton patches in a liquid detergent. Standard deviation for the determinations is shown.

The wash performances in the powder detergent, Figure 30, were in general higher than the performances observed in liquid detergent. Again N265D had the highest wash performance. The wash performance of K306M, Q360E, Q264E and K383M was better than of BLA, but not as good as N265D. The variants N172D and R437L had the same wash performance as BLA under these conditions, whereas Q9E had a lower wash performance.

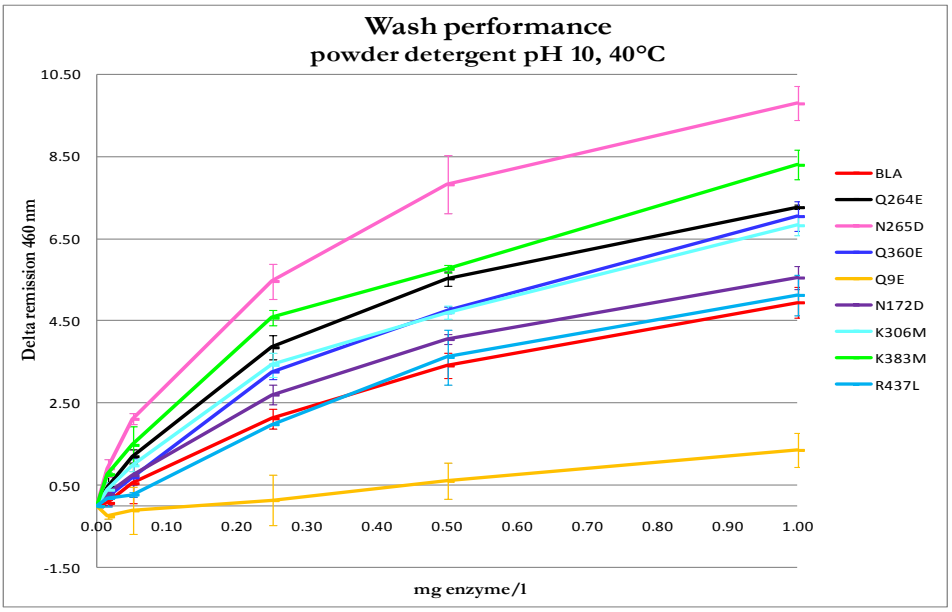



**Figure 30.** The wash performance of the BLA variants on rice starch cotton patches in a powder detergent. Standard deviation for the determinations is shown.

In general there was no strong correlation between the kinetic parameters obtained on G7 and amylose with the wash performance data. This shows that the structural factors influencing the kinetic parameters in the assay are not necessarily the same as the ones affecting the wash performance.

The increased wash performance observed for most of the single mutations can be caused by a number of different factors. It has been proposed that decreasing the thermostability of BLA might increase the activity at lower temperatures. This is suspected to be due to the rigidity of thermostable enzyme as they generally have the highest activity at elevated temperatures but the activity decreases with temperature as the structure becomes more rigid (Radestock *et al.*, 2011). There was, however, not observed a loss in stability by DSC corresponding to increased activity for the BLA variants. It could also be that the mutations on the surface somehow influence the enzyme interaction with the substrate structure on the textile hereby increasing the wash performance. For the mutation of Lys and Arg residues it is possible the increased wash performance is due to the removal of the reducing sugar reactive site. Changing the pKa values of amino acid residues around the active site has also been proposed by others to be able to influence the pH activity profile of enzymes (Nielsen *et al.*, 2001). Thus the changes around the active site could influence the pH activity profile.

The Q264E mutation resulted in a great improvement on the catalytic efficiency compared to BLA and the other variants but in the wash performance the Q264E variant was quite average. The increased catalytic efficiency was caused by improved substrate binding. The crystal structure of a BLA variant (not from this project) with acarbose has been solved by Novozymes. This structure is not published (Andersen 2012b). In the crystal structure of the active site of BLA in complex with acarbose Q264 is oriented away from the substrate, Figure 31. In theory the carbonyl or amide groups on Gln should be able to make hydrogen bonds with hydroxyl groups on the substrate. The Gln side chain is very flexible in this structure so it is possible that hydrogen bonding does occur. The Q264E mutation increases the substrate binding. This strongly suggests that Glu is hydrogen bonding with hydroxyl groups on the substrate, whereas Gln does not. If Gln does form a hydrogen bond to the substrate, the bonds formed by Glu appear to be considerably stronger resulting in increased substrate binding. The improved substrate binding also is accompanied by a positive effect on the wash performance at pH 8 and 10.

The kinetic parameters for the N265D variant did not show an effect of this mutation on the enzyme but it outperformed all the other variants in wash performance in both liquid and powder detergent. N265 is located right next to residue 264 in which Q264E has shown to interact with the substrate. The crystal structure of BLA in complex with acarbose, Figure 31, reveals that N265 is located close to the substrate and it should be able to form hydrogen bonds with this. Since Q264E increased substrate binding it would be expected that the N265D mutation had a similar effect especially on a longer substrate as it looks like a longer substrate could bind with that side chain. This is, however, not confirmed by the $K_M$ values for the G7 substrate and amylose. The substantial increase in wash performance for N265D does show that

the mutation has a positive effect on the performance. This indicates that the mutation has improved substrate binding to the rice starch or the starch structure on the textile.

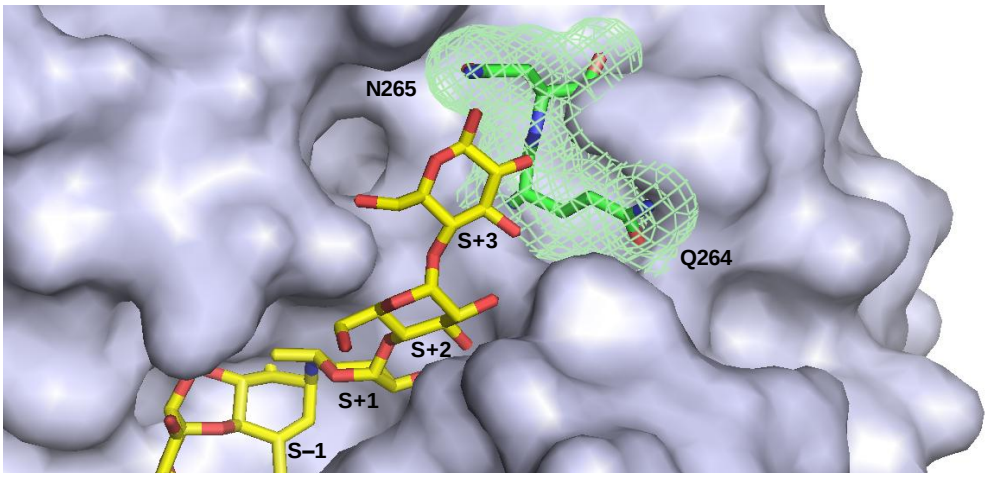



**Figure 31.** The active site of BLA in complex with acarbose showing the possible substrate binding residues of Q264 and N265 (Unpublished structure of BLA with acarbose from Novozymes). Acarbose shown in yellow with the bond cleaved in the corresponding substrate, in the active site shown in blue. The subsites +1, +2, +3 and -1 in the substrate binding site are shown as S+1, S+2, S+3 and S–1, respectively. Oxygen atoms are shown in red and nitrogen in blue. Pymol was used for molecular rendering (The PyMOL molecular Graphics System, Version 1.3, Schrödinger, LCC).

The kinetic parameters for Q9E revealed a lower catalytic efficiency on both substrates. Q9 is located right behind H327 which can interact with the substrate in subsite -1, Figure 32. This region is therefore very important for both substrate binding and catalytic activity. Q9 is most likely able to hydrogen bond with H327, though when the substrate binds in the active site H327 also hydrogen bond to one of the hydroxyl groups in the substrate. The Q9E mutation most likely results in salt bridge formation between the carboxy group on Q9E with the imidazole group in H327. This bond is stronger than the hydrogen bond so the interaction between H327 and the substrate will be weakened or even non-existing. This can explain the decreased substrate binding for the G7 substrate resulting in lower catalytic efficiency. The mutation does not have a negative effect on the substrate affinity for the longer substrate probably because this binding is stabilised by more binding sites on BLA than the short substrate. The mutation does, however, result in a decreased substrate turnover. The possible loss of hydrogen bonding between H327 and amylose could result in a different conformation of the substrate in the active site, hereby lowering the catalytic efficiency.

Despite the decrease in catalytic efficiency the wash performance of Q9E increased in liquid detergent. The wash performance in powder detergent was, however, considerably lower than of BLA and all the other variants. This indicates that the site influences the pH profile of the amylase as the performance is good at pH 8 but decreases significantly at pH 10. This also correlates with other studies showing that pKa values of amino acid residues around the catalytic site can influence the pH activity profile of an enzyme (Nielsen *et al.*, 2001). The differences in wash performance could, however, also be due to some in-wash stability problems with the powder detergent possibly caused by one or several of the detergent components. Overall the

different assays reveal that the Q9 site has an influence on the substrate catalysis influencing substrate affinity and substrate turnover as well as possibly the pH activity profile.

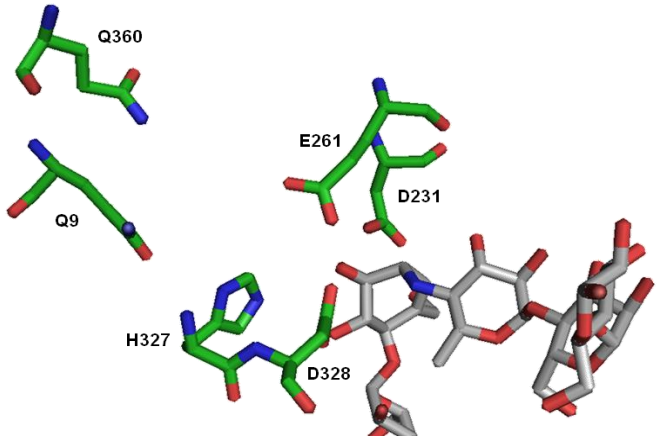



**Figure 32.** Close-up of acarbose in the active site showing the three active site residues D231, E261 and D328 along with the residues Q9, Q360 and H327. Acarbose is shown in grey with the bond to be cleaved in a corresponding substrate in the active site shown in blue. Oxygen atoms are red and nitrogen blue. (The PyMOL molecular Graphics System, Version 1.3, Schrödinger, LCC)

Q360 is located very close to Q9 in the area around the active site, Figure 32. Q360 is surrounded by several hydrophilic amino acid residues. It is therefore possible that this site is involved in the overall packing of the surrounding area as it could be loosening up the packing of this area. The Q360E mutation which is more polar at neutral pH could further loosen up the packing, thus affecting the area around the active site. The position of Q360 is approximately 5.9 Å from the active site residue E261. The Q360E mutation is therefore likely to influence the pKa of E261. The kinetic parameters revealed that the Q360E mutation improved the substrate turnover but lowered $k_{cat}$ thus not influencing the overall catalytic efficiency. This could indicate that the Q360E mutation gives rise to some slightly larger conformational changes in the area that changes the general substrate binding and catalytic turnover. The Q360E mutation does not influence the wash performance in liquid detergent but the performance is slightly improved in powder detergent. This could indicate that the site influences the pH activity profile of the amylase as no change in performance occurs at pH 8 but performance increases at pH 10.

The N172D mutation did not have any effect on the kinetic parameters. The crystal structure of the active site in complex with acarbose, Figure 33, reveals that N172 is relatively close to the substrate binding, but approximately 5 Å from the substrate which is too far to form hydrogen bonds. Therefore, it has no effect on the substrate binding. The wash performance in powder detergent was the same as for BLA though the performance improved in the liquid detergent. The $T_{max}$ determinations showed that the mutation resulted in a small decrease in the stability of BLA. As previously mentioned it has been hypothesised that for thermostabile enzymes decreasing the stability might increase activity at lower temperatures. It is also possible that the mutation improves interaction with substrate on the fabric at pH 8 hereby improving wash performance. Though, it could also be other parameters responsible for the change such as the different components in the detergents.

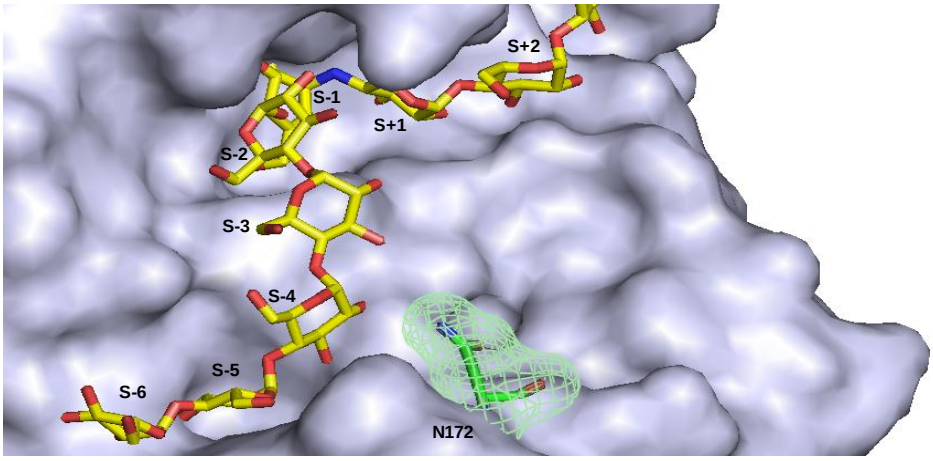



**Figure 33.** BLA in complex with acarbose in the active site showing the residue N172 in green. Extended acarbose is shown in yellow with the bond to be cleaved in a corresponding substrate in the active site shown in blue. The subsites in the binding site are indicated by S and the site number. Oxygen atoms are red and nitrogen blue. (The PyMOL molecular Graphics System, Version 1.3, Schrödinger, LCC)

The position of K306 in relation to the substrate binding site is illustrated in Figure 34A. It appears that this site could be located in an area in which longer substrates bind. Though, the K306M mutation did not have any effect on the kinetic parameters thus indicating that K306 is not important for interaction with amylose. The mutation does, however, increase wash performance. The K → M mutation on the enzyme surface would be expected to have a lower binding affinity toward the substrate on the textile as the positively charged Lys can be involved in these interactions. The amylose used in the kinetic assay was gelatinised and the substrate might have a different structure on the textile which could explain the different results. The improved wash performance could be caused by improved interaction with the substrate on the fabric or by lover unproductive binding of the enzyme to other surface such as the textile.

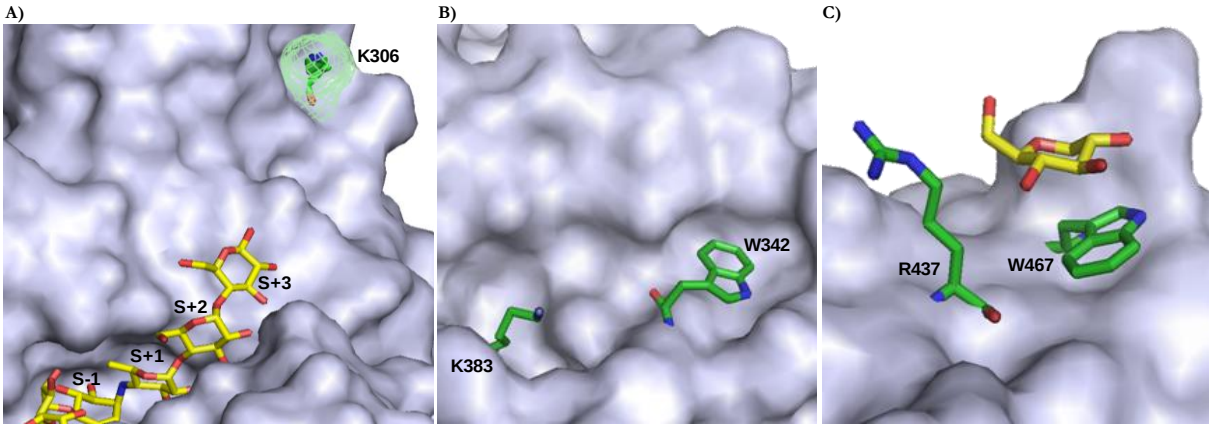



**Figure 34.** Different views of the crystal structure of BLA. Oxygen atoms are red and nitrogen blue. A) The structure of acarbose in the active site showing K306 in what could be an elongation of the substrate binding. Acarbose is shown in yellow with the bond to be cleaved in a corresponding substrate in the active site shown in blue. The subsites in the binding site are indicated by S and the site number. B) K383 and the secondary binding site W342. C) R437 as part of a secondary binding site with W467 with glucose bound. The position of R437 is estimated by modelling after the position of W439 in BHA solved with glucose in the binding site. (The PyMOL molecular Graphics System, Version 1.3, Schrödinger, LCC)

K383 is located very close to the secondary binding site at W342, Figure 34B, so K383 could be involved in carbohydrate binding in this site. The mutation K383M did not result in any changes in the kinetic parameters for BLA. It did, however, result in improved wash performance. As described above the increased wash performance could be because the K383M mutation improves interaction of the enzyme with the substrate on the fabric. It is also possible that the mutation hinder the glycation in this site which could block the secondary binding site.

R437 is a part of a secondary binding site in domain C along with W467, Figure 34C. The interaction with glucose shown in the figure is an estimated modelling of the position of R437 after the position of W439 in BHA of which the structure was solved with glucose in this binding site (Lyhne-Iversen *et al.*, 2006). From this structure it could appear that R437 is able to form hydrogen bonds to the carbohydrate. The R437L mutation may have a small positive effect on the $k_{cat}$ and hereby a slight improvement of the catalytic efficiency, but the variation of this result is very high. These results indicate that the site may be involved in substrate binding possibly as an initial substrate recognition site. The R437L mutation also has a positive effect on the wash performance in liquid detergent. The fact that the improvement only occurs in one of the detergents could indicate that the improvement is due to some interaction or lack of interaction with a detergent ingredient which is present in one of the detergents.

The analysis of the variants shows that deamidations in the area close to the active site influence the kinetic parameters on different substrates and hereby the activity and substrate binding. Deamidations a bit further from the active site do not seem to have much effect on the kinetic parameters but they can influence other factors such as stability. The Q→E and N→D variants also indicated that deamidations could affect wash performance generally positively though in the area close to the active site the modifications are also likely to influence the pH activity profile which can have a negative effect on wash performance.

The removal of possible reducing sugar reactive sites on the surface of BLA (K→M and R→L) did not seem to have any effect on the kinetic parameters though all the substitutions seemed to have a positive effect on the wash performance in liquid detergent as well as in powder detergent for the K→M mutations. This may be the result of changes in the interaction of the amylase surface with the substrate on the textile.

In general the single mutations had a positive effect on the wash performance thus indicating that some of the modifications on the surface of BLA improve wash performance. The introduction of multiple mutations in one molecule most likely has other effects on the performance. This section only focused on single modifications in selected sites likely to affect activity and substrate interaction whereas sites relevant to enzyme stability were not evaluated. Modifications on these sites might show a more complex picture of the influence of modification sites on the wash performance.

# 4 Conclusion

During fermentation, multiple forms of the α-amylases varying in pI values were formed. Studies of partially separated pools of these forms in such different alpha-amylases as BHA, AA560 and BLA revealed that the wash performance of the amylases tends to decrease with the decrease in pI value. This was especially evident at pH 10.

A more in-depth characterisation of the BLA pools varying in pI range indicated that the multiple modifications affected the enzyme negatively as the pools containing the forms with the most modifications showed slightly lower catalytic efficiency and wash performance compared to the forms with fewer modifications (i.e. forms in the higher end of the pI range).

LC-MS/MS methods with the fragmentation methods HCD and ETD were evaluated for their ability to identify glycation sites using BASI, a small protein as a model. Both methods were able to identify glycation sites though ETD performed better as it identified more glycation sites and the sequence coverage of the glycated peptides was generally higher with ETD fragmentation.

Methods for relative quantification of glycation sites using tandem mass tags (TMT) were also set up using LC-MS/MS fragmenting with HCD and ETD with BASI as model protein. Here HCD outperformed ETD as it was much better at acquiring reporter ion intensities high enough for quantification. Statistical analysis of the quantification results revealed that the glycated peptides could influence the performance of the fragmentation methods and the TMT labels. It is, therefore, important in the relative quantification of glycated peptides to perform the quantifications with different label combinations and if possible also using different fragmentation techniques to evaluate the significance of the quantifications.

The modifications responsible for the multiple forms of BLA varying in pI values were determined by LC-MS/MS by HCD and ETD. Here multiple deamidation, glycation and carbamylation sites were identified on the surface of BLA as well as a few deamidations in and around the active site. Several of the modified amino acid residues are suspected to be involved in the enzyme activity, stability and substrate binding.

Modifications were observed at the amino acid residues R127, N172, K237 and N272 which are suspected to be involved in the thermostability of BLA and it is very likely that the modifications in these sites decrease the enzyme stability.

Modifications were observed at the three residues Q9, Q360 and N326 located in the active site and on K234, Q264 and N265 which are located near the substrate binding cleft. These modifications would most likely influence the enzymes activity and substrate binding. Modifications were also observed at sites which could be involved in binding of longer substrates and in residues close to or part of BLAs secondary binding sites.

On the basis of the identified modification sites BLA variants were produced with focus on the impact of modifications in the active site area, substrate binding and on secondary carbohydrate binding sites.

Analysis of BLA variants with mutations mimicking deamidations (Q→E and N→D) revealed that deamidations in the area close to the active site influence the activity and substrate binding.

These variants generally revealed increased wash performance indicating that single deamidations in this area could have a positive influence on the wash performance.

BLA variants were also made with mutations of Arg and Lys residues on the surface of BLA suspected to be involved in binding of longer substrates and secondary binding sites (K→M and R→L). These changes did not have an effect on the kinetic parameters though the substitutions generally improved the wash performance. The improved wash performance may be due to changes in the interaction between the surface of BLA with the starch, the fabric or a combination of the two.

# 5 Perspectives

The influence of the observed modifications was determined by characterisation of BLA variants with site specific mutations in identified modification sites. These variants generally showed increased wash performance. The mechanism causing this improved effect is not fully understood wherefore these results are very interesting. Looking at the mutations in sites suspected to be involved in substrate binding or secondary binding sites one explanation for the increased performance would be that these sites are involved. This should be explored further and it could especially be interesting for the sites located on the surface of BLA away from the active site. In a wash performance assay the performance is affected by the interaction between the amylase and its substrate, the substrates structure on the textile and unproductive binding of the enzyme to other surfaces e.g. textiles. This could be investigated by studies of the BLA variants binding affinity to the substrate. This may provide further knowledge of the mechanisms involved in the wash performance of amylases which could be used for optimisation of amylases in the detergent industry.

Extending the investigation of the BLA variants exploring more variants with single modifications in identified modifications sites could also be very interesting, possibly in sites not expected to be involved in the stability, substrate binding and activity. Investigations of the wash performance of these variants might reveal more sites involved which can influence the wash performance, possibly some where the modifications occurring have a negative effect. Study of variants with multiple mutations in modification sites could provide knowledge about whether some combinations further increase the performance or have the opposite effect.

LC-MS/MS methods fragmenting with ETD and HCD were set up for relative quantification of glycated peptides with isobaric stable isotope labelling. Statistical analysis of the quantification results revealed that the glycated peptides, the tags and the fragmentation method could influence the quantification results. As mentioned previously it could therefore be interesting to perform a larger study of the quantification of glycated peptides where the influence of factors such as the glycated peptides, the specific glycations, the tags and the fragmentation methods are studied more thoroughly.

The relative quantification of glycation sites between the four pools of BLA varying in pI values did not provide any quantification results. As mentioned previously the method could be optimised by only quantifying two samples at a time. The quantification with ETD could possibly also be further optimised by only selecting precursor ions of charge state four as this might improve reporter ion intensities.

# 6 Reference List


Aebersold, R. & Mann, M. "Mass spectrometry-based proteomics." Nature, 422 (2003): 198–207.

Ahmed, N. "Advanced glycation endproducts - role in pathology of diabetic complications." Diabetes Research and Clinical Practice, 67 (2005): 3–21.

Albert, C., Mertens P. R., & Bartsch P. "Urea and atherosclerosis-evidence for a direct link involving apolipoprotein B protein modifications." International Urology and Nephrology, 43 (2011): 933–36.

Ames, J. M. "Mass spectrometry to detect the site specificity of advanced glycation/lipoxidation end-product formation on protein: some challenges and solutions." Biochemical Society Transactions, 36 (2008): 1051–1054.

Andersen, B. Science Manager, Peptide and Cell Discovery, Novozymes, Bagsvaerd (2012a), Personal Communication.

Andersen, C. Senior Science Manager, Protein Diversity, Novozymes, Bagsvaerd (2012b), Personal Communication.

Aswad, D. W., Paranandi M. V. & Schurter B. T. "Isoaspartate in peptides and proteins: formation, significance, and analysis." Journal of Pharmaceutical and Biomedical Analysis, 21 (2000): 1129–36.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. "Quantitative mass spectrometry in proteomics: a critical review." Analytical and Bioanalytical Chemistry, 389 (2007): 1017–1031.

Barkholt, V. & Jensen, A. L. "Amino acid analysis: Determination of cysteine plus half-cystine in proteins after hydrochloric acid hydrolysis with a disulfide compound as additive." Analytical Biochemistry, 177, (1989): 318-322.

Barnaby, O. S., Cerny, R. L., Clarke, W. & Hage, D. S. "Comparison of modification sites formed on human serum albumin at various stages of glycation." Clinica Chimica Acta, 412 (2011): 277–285.

Barnes, C. A. S., & Lim, A. "Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals." Mass Spectrometry Reviews, 26 (2007): 370–388.

Barrett, J., Brophy, P. M., & Hamilton, J. V. "Analysing proteomic data." International Journal for Parasitology, 35 (2005): 543–553.

Baynes, J. W., Watkins, N. G., Fisher, C. I., Hull, C. J., Patrick, J. S., Ahmed, M. U., Dunn, J. A. & Thorpe, S. R. "The Amadori product on protein: structure and reactions." Progress in Clinical and Biological Research, 304 (1989): 43–67.

Berg, J. M., Tymoczko J. L., & Stryer L. "Protein Composition and Structure." In Biochemistry 6. ed., W.H. Freeman and Company, New York (2007): Chapter 2.

Biemann, K. "Mass Spectrometry of Peptides and Proteins." Annual Review of Biochemistry, 61 (1992): 977–1010.

Boyne, M. T., Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. "Precise Characterization of Human Histones in the H2A Gene Family by Top Down Mass Spectrometry." Journal of Proteome Research, 5 (2006): 248–53.

Brownlee, M., Cerami, A., & Vlassara, H. "Advanced glycosylation end products in tissue and the biochemical basis of diabetic complications." The New England Journal of Medicine 318 (1988): 1315-1321.

Bruins, A. P. "Review: Mechanistic aspects of electrospray ionization." Journal of Chromatography A, 794 (1998): 345–357.

Bruins, M. E., Van Hellemond, E. W., Janssen, A. E. M. & Boom, R. M. "Maillard reactions and increased enzyme inactivation during oligosaccharide synthesis by a hyperthermophilic glycosidase." Biotechnology and Bioengineering, 81 (2003): 546–552.

Brzozowski, A. M., Lawson, D. M., Turkenburg, J. P., Bisgaard-Frantzen, H., Svendsen, A., Borchert, T. V., Dauter, Z., Wilson, K. S. & Davies, G. J. "Structural analysis of a chimeric bacterial alpha-amylase. High-resolution analysis of native and ligand complexes." Biochemistry, 39 (2000): 9099–9107.

Bunn, H. F. & Higgins P. J. "Reaction of Monosaccharides with Proteins: Possible Evolutionary Significance." Science, 213 (1981): 222–224.

Chaplin, M. "Starch", http://www.lsbu.ac.uk/water/hysta.html (2012), accessed 10/04/2012.

Chelius, D., Rehder, D. S., & Bondarenko, P. V. "Identification and characterization of deamidation sites in the conserved regions of human Immunoglobulin Gamma antibodies." Analytical Chemistry, 77 (2005): 6004–6011.

Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J., Syka, J. E. P., Bai, D. L., Shabanowitz, J., Burke, D. J., Troyanskaya, O. G. & Hunt, D. F. "Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry." Proceedings of the National Academy of Sciences, 104 (2007): 2193–2198.

Chobert, J. M., Gaudin, J. C., Dalgalarrondo, M., & Haertlé, T. "Impact of Maillard type glycation on properties of beta-lactoglobulin." Biotechnology Advances, 24 (2006): 629–632.

Cramer, R., Gobom, J., & Nordhoff, E. "High-throughput proteomics using matrix-assisted laser desorption/ionization mass spectrometry." Expert Review of Proteomics, 2 (2005): 407–420.

Cuyvers, S., Dornez, E., Delcour, J. A. & Courtin, C. M. "Occurrence and functional significance of secondary carbohydrate binding sites in glycoside hydrolases." Critical Reviews in Biotechnology, 32 (2011): 93–107.

Dauter, Z., Brzozowski, A. M., Christensen, S., Borchert, T. V., Beier, L., Wilson, K. S. & Davies, G. J. "X-ray Structure of Novamyl, the Five-Domain 'Maltogenic' alpha-Amylase from *Bacillus stearothermophilus*: Maltose and Acarbose Complexes at 1.7 A Resolution." Biochemistry, 38 (1999): 8385–8392.

Davies, G. J., Wilson, K. S., & Henrissat, B. "Nomenclature for sugar-binding subsites in glycosyl hydrolases." Biochemical Journal, 321 (1997): 557–559.

Dayon, L., Pasquarello, C., Hoogland, C., Sanchez, J. C. & Scherl, A. "Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags." Journal of Proteomics, 73 (2010): 769–777.

Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R. & Sanchez, J. C. "Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags." Analytical Chemistry, 80 (2008): 2921–2931.

de Souza, P. M. & Magalhaes P. D. E. "Application of Microbial alpha-Amylase in Industry - a review." Brazilian Journal of Microbiology, 41 (2010): 850–861.

Declerck, N., Machius, M., Chambert, R., Wiegand, G., Huber, R. & Gaillardin, C. "Hyperthermostable mutants of *Bacillus licheniformis* alpha-amylase: thermodynamic studies and structural interpretation." Protein Engineering, 10 (1997): 541–549.

Declerck, N., Machius, M., Joyet, P., Wiegand, G., Huber, R. & Gaillardin, C. "Hyperthermostabilization of *Bacillus licheniformis* alpha-amylase and modulation of its stability over a 50°C temperature range." Protein Engineering, 16 (2003): 287–294.

Declerck, N., Machius, M., Wiegand, G., Huber, R. & Gaillardin, C. "Probing structural determinants specifying high thermostability in *Bacillus licheniformis* alpha-amylase." Journal of Molecular Biology, 301 (2000): 1041–1057.

Delatour, T., Hegele, J., Parisod, V., Richoz, J., Maurer, S., Steven, M. & Buetler, T. "Analysis of advanced glycation endproducts in dairy products by isotope dilution liquid chromatography-electrospray tandem mass spectrometry. The particular case of carboxymethyllysine." Journal of Chromatography A, 1216 (2009): 2371–2381.

Deng, W. P. and Nickoloff, J. A. "Site-directed mutagenesis of virtually any plasmid by eliminating a unique site." Analytical Biochemistry, 200 (1992): 81–88.

Domon, B. & Aebersold, R. "Mass spectrometry and protein analysis." Science, 312 (2006): 212–217.

Eriksen, N. "Detergents." In Industrial Enzymology 2. ed. Editors Godfrey, T. & West, S., MacMillan Press Ltd, London (1996): Chapter 2.10.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. "Electrospray Ionization for Mass Spectrometry of Large Biomolecules." Science, 246 (1989): 64–71.

Fischer, S., Hoernschemeyer, J. & Mahler, H. C. "Glycation during storage and administration of monoclonal antibody formulations." European Journal of Pharmaceutics and Biopharmaceutics, 70 (2008): 42–50.

Fitter, J., Herrmann, R., Dencher, N. A., Blume, A. & Hauss, T. "Activity and stability of a thermostable alpha-amylase compared to its mesophilic homologue: Mechanisms of thermal adaptation." Biochemistry, 40 (2001): 10723–10731.

Gadgil, H. S., Bondarenko, P. V., Treuheit, M. J. & Ren, D. "Screening and Sequencing of Glycated Proteins by Neutral Loss Scan LC/MS/MS Method." Analytical Chemistry, 79 (2007): 5991–5999.

Gaillard, T. & Bowler, P. "Morphology and Composition of Starch." In Starch: Properties and Potential. Editor Gaillard, T., John Wiley & Sons, Chichesters (1987): Chapter 3.

Gaza-Bulseco, G., Li, B. Q., Bulseco, A. & Liu, H. C. "Method to Differentiate Asn Deamidation that Occurred Prior to and during Sample Preparation of a Monoclonal Antibody." Analytical Chemistry, 80 (2008): 9491–9498.

Glish, G. L. & Burinsky, D. J. "Hybrid Mass Spectrometers for Tandem Mass Spectrometry." Journal of the American Society for Mass Spectrometry, 19 (2008): 161–172.

Good, D. M., Wirtala, M., McAlister, G. C. & Coon, J. J. "Performance characteristics of electron transfer dissociation mass spectrometry." Molecular & Cellular Proteomics, 6 (2007): 1942–1951.

Gross, J. H. Mass Spectrometry : a Textbook 2. Ed. Springer, Heidelberg (2011).

Habibi, A. E., Khajeh, K. & Nemat-Gorgani, M.. "Chemical Modification of Lysine Residues in *Bacillus licheniformis* alpha-Amylase: Conversion of an Endo- to an Exo-type Enzyme." Journal of Biochemistry and Molecular Biology, 37 (2004): 642–647.

Hager, J. W. "A new linear ion trap mass spectrometer." Rapid Communications in Mass Spectrometry, 16 (2002): 512–526.

Hagihara, H., Hatada, Y., Ozawa, T., Igarashi, K., Araki, H., Ozaki, K., Kobayashi, T., Kawai,S. & Ito, S. "Oxidative Stabilization of an Alkaliphilic *Bacillus* alpha-Amylase by Replacing a Single Specific Methionine Residue by Site-directed Mutagenesis." Journal of Applied Glycoscience, 50 (2003): 367–72.

Han, H., Pappin, D. J., Ross, P. L. & McLuckey, S. A. "Electron Transfer Dissociation of iTRAQ Labeled Peptide Ions." Journal of Proteome Research, 7 (2008): 3643–3648.

Harding, J. J., Beswick, H. T., Ajiboye, R., Huby, R., Blakytny, R. & Rixon, K. C. "Non-enzymic post-translational modification of proteins in aging. A review." Mechanisms of Ageing and Development, 50 (1989): 7–16.

Hatada, Y., Masuda, N., Akita, M., Miyazaki, M., Ohta, Y., & Horikoshi, K. "Oxidatively stable maltopentaose-producing alpha-amylase from a deep-sea *Bacillus* isolate, and mechanism of its oxidative stability validated by site-directed mutagenesis." Enzyme & Microbial Technology, 39 (2006): 1333–1340.

Hjernø, K. & Højrup, P. "Massespektrometri." In Proteiner - oprensning og karakterisering 3. ed. Editors Kielberg, V. & Rasmussen, L., Gyldendal, Copenhagen (2005): 12.

Hoffmann, E. & Strobant, V. "Introduction." In Mass spectrometry principles and applications 3. ed. Wiley, Chichester (2008a).

Hoffmann, E. & Strobant, V. "Ion Sources." In Mass spectrometry principles and applications 3. ed. Wiley, Chichester (2008b): Chapter 1.

Hogan, J. M., Pitteri, S. J., Chrisman, P. A. & McLuckey, S. A. "Complementary Structural Information from a Tryptic N-Linked Glycopeptide via Electron Transfer Ion/Ion Reactions and Collision-Induced Dissociation." Journal of Proteome Research, 4 (2005): 628–632.

Horvat, S. & Jakas, A.. "Peptide and amino acid glycation: New insights into the Maillard reaction." Journal of Peptide Science, 10 (2004): 119–137.

Janecek, S. "alpha-Amylase family: Molecular biology and evolution." Progress in Biophysics and Molecular Biology, 67 (1997): 67–97.

Janecek, S., Svensson, B. & Henrissat, B. "Domain evolution in the alpha-amylase family." Journal of Molecular Evolution, 45 (1997): 322–331.

Jenkins, N., Murphy, L. & Tyther, R. "Post-translational Modifications of Recombinant Proteins: Significance for Biopharmaceuticals." Molecular Biotechnology, 39 (2008): 113–118.

Johansen, M. B., Kiemer, L. & Brunak, S. "Analysis and prediction of mammalian protein glycation." Glycobiology, 16 (2006): 844–853.

Joshi, A. B. & Kirsch, L. E. "The relative rates of glutamine and asparagine deamidation in glucagon fragment 22-29 under acidic conditions." Journal of Pharmaceutical Sciences, 91 (2002): 2332–2345.

Kamath, K. S., Vasavada, M. S., & Srivastava, S. "Proteomic databases and tools to decipher post-translational modifications." Journal of Proteomics, 75 (2011): 127–144.

Kandra, L., Gyémánt, G., Remenyik, J., Hovánszki, G. & Lipták, A. "Action pattern and subsite mapping of *Bacillus licheniformis* alpha-amylase (BLA) with modified maltooligosaccharide substrates." FEBS Letters, 518 (2002): 79–82.

Khajeh, K., Naderi-Manesh, H., Ranjbar, B., Moosavi-Movahedi, A. A. & Nemat-Gorgani, M. "Chemical modification of lysine residues in *Bacillus* alpha-amylases: effect on activity and stability." Enzyme and Microbial Technology, 28 (2001a): 543–549.

Khajeh, K., Ranjbar, B., Naderi-Manesh, H., Habibi, A. E. & Nemat-Gorgani, M. "Chemical modification of bacterial alpha-amylases: changes in tertiary structures and the effect of additional calcium." Biochimica et Biophysica ACTA - Protein Structure and Molecular Enzymology, 1548 (2001b): 229–237.

Kim, I. C., Cha, J. H., Kim, J. R., Jang, S. Y., Seo, B. C., Cheong, T. K., Lee, D. S., Choi, Y. D. & Park, K. H. "Catalytic properties of the cloned amylase from *Bacillus licheniformis*." The Journal of Biological Chemistry, 267 (1992): 22108–22114.

Kirk, O., Borchert, T. V. & Fuglsang, C. C. "Industrial enzyme applications." Current Opinion in Biotechnology, 13 (2002): 345–51.

Kislinger, T., Humeny, A., Peich, C. C., Becker, C. M. & Pischetsrieder, M. "Analysis of Protein Glycation Products by MALDI-TOF/MS." Annals of the New York Academy of Sciences, 1043 (2005): 249–259.

Kislinger, T., Humeny, A. & Pischetsrieder, M. "Analysis of protein glycation products by matrix-assisted laser desorption ionization time-of-flight mass spectrometry." Current Medicinal Chemistry, 11 (2004): 2185–2193.

Kjeldsen, F., Haselmann, K. F., Budnik, B. A., Sorensen, E. S. & Zubarev, R. A. "Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage." Analytical Chemistry, 75 (2003): 2355–2361.

Kjeldsen, F., Giessing, A. M. B., Ingrell, C. R. & Jensen, O. N. "Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry." Analytical Chemistry, 79 (2007): 9243–9252.

Koomen, J., Hawke, D. & Kobayashi, R. "Developing an understanding of proteomics: An introduction to biological mass spectrometry." Cancer Investigation, 23 (2005): 47–59.

Kosky, A. A., Razzaq, U. O., Treuheit, M. J. & Brems, D. N. "The effects of alpha-helix on the stability of Asn residues: Deamidation rates in peptides of varying helicity." Protein Science, 8 (1999): 2519–2523.

Kraus, L. M. & Kraus, A. P. Jr. "Carbamoylation of amino acids and proteins in uremia." Kidney International, 78 (2001): S102–S107.

Kuriki, T. & Imanaka, T. "The concept of the alpha-amylase family: structural similarity and common catalytic mechanism." Journal of Bioscience and Bioengineering, 87 (1999): 557–565.

Lapolla, A., Fedele, D., Reitano, R., Aricò, N. C., Seraglia, R.; Traldi, P., Marotta, E. & Tonani, R. "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.

Lapolla, A., Traldi, P. & Fedele, D. "Importance of measuring products of non-enzymatic glycation of proteins." Clinical Biochemistry, 38 (2005): 103–115.

Larsen, M. R., Trelle, M. B., Thingholm, T. E. & Jensen, O. N. "Analysis of posttranslational modifications of proteins by tandem mass spectrometry." Biotechniques, 40 (2006): 790–798.

Li, Hui, Robertson, A. D. & Jensen, J. H. "Very fast empirical prediction and rationalization of protein pKa values." Proteins, 61 (2005): 704–721.

Li, S., Patapoff, T. W., Overcashier, D., Hsu, C., Nguyen, T. H. & Borchardt, R. T."Effects of reducing sugars on the chemical stability of human relaxin in the lyophilized state." Journal of Pharmaceutical Sciences, 85 (1996): 873–877.

Li, X. J., Cournoyer, J. J., Lin, C. & O'Cormora, P. B. "Use of O-18 labels to monitor deamidation during protein and peptide sample processing." Journal of the American Society for Mass Spectrometry, 19 (2008): 855–864.

Lin, L. L., Lo, H. F., Chiang, W. Y., Hu, H. Y., Hsu, W. H. & Chang, C. T. "Replacement of Methionine 208 in a Truncated *Bacillus* sp. TS-23 alpha-Amylase with Oxidation-Resistant Leucine Enhances Its Resistance to Hydrogen Peroxide." Current Microbiology, 46 (2003): 211–216.

Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A.. Scott, M. P., Zipursky, S. L. & Darnell, J. Molecular Cell Biology 5th ed. W. H. Freeman and Company, New York (2004).

Lund, H. Research Scientist, Detergent Research Center, Novozymes, Bagsvaerd (201), Personal Communication.

Lund, H., Kaasgaard, S., Skagerlind, P., Jorgensen, L., Jørgensen, C. & Weert, M. "Protease and Amylase Stability in the Presence of Chelators Used in Laundry Detergent Applications: Correlation Between Chelator Properties and Enzyme Stability in Liquid Detergents." Journal of Surfactants & Detergents, 15 (2012): 265–276.

Lyhne-Iversen, L., Hobley, T. J., Kaasgaard, S. G., & Harris, P. "Structure of *Bacillus halmapalus* alpha-amylase crystallized with and without the substrate analogue acarbose and maltose." Acta Crystallographica: Section F, 62 (2006): 849–854.

MacGregor, E. A., Janecek, S. & Svensson, B. "Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes." Biochimica et Biophysica Acta, 1546 (2001): 1–20.

Machius, M., Declerck, N., Huber, R. & Wiegand, G. "Kinetic Stabilization of *Bacillus licheniformis* alpha-Amylase through Introduction of Hydrophobic Residues at the Surface." Journal of Biological Chemistry, 278 (2003): 11546–11553.

Machius, M., Wiegand, G., & Huber, R. "Crystal Structure of Calcium-depleted *Bacillus licheniformis* alpha-amylase at 2.2 Å Resolution." Journal of Molecular Biology, 246 (1995): 545–559.

Machius, M., Declerck, N., Huber, R. & Wiegand, G. "Activation of *Bacillus licheniformis* alpha-amylase through a disorder → order transition of the substrate-binding site mediated by a calcium-sodium-calcium metal triad." Structure, 6 (1998): 281–92.

Maillard, L. C. "The action of amino acids on sugar; The formation of melanoidin by a methodic route." Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences, 154 (1912): 66–68.

Makarov, A., Denisov, E., Kholomeev, A., Baischun, W., Lange, O., Strupat, K. & Horning, S. "Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer." Analytical Chemistry, 78 (2006): 2113–2120.

Makarov, A. & Scigelova, M. "Coupling liquid chromatography to Orbitrap mass spectrometry." Journal of Chromatography A, 1217 (2010): 3938–3945.

Mann, M., Hendrickson, R. C. & Pandey, A. "Analysis of proteins and proteomes by mass spectrometry." Annual review of biochemistry, 70 (2001): 437–473.

Mann, M. & Jensen, O. N. "Proteomic analysis of post-translational modifications." Nature Biotechnology, 21 (2003): 255–261.

Manning, M. C., Patel, K. & Borchardt, R. T. "Stability of protein pharmaceuticals." Pharmaceutical Research, 6 (1989): 903–918.

Manning, M., Chou, D., Murphy, B., Payne, R. & Katayama, D. "Stability of Protein Pharmaceuticals: An Update." Pharmaceutical Research, 27 (2010): 544–575.

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

Matrix Science, "Mascot search." http://www.matrixscience.com/search_form_select.html (2012), accessed 2012/03/29.

Micheelsen, P. O., Østergaard, P. R., Lange, L. & Skjøt, M. "High-level expression of the native barley alpha-amylase/subtilisin inhibitor in *Pichia pastoris*." Journal of Biotechnology, 133 (2008): 424–432.

Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E. P., Shabanowitz, J. & Hunt, D. F. "The utility of ETD mass spectrometry in proteomic analysis." Biochimica et Biophysica Acta - Proteins and Proteomics, 1764 (2006): 1811–1822.

Mirgorodskaya, E., Hassan, H., Clausen, H. & Roepstorff, P. "Mass Spectrometric Determination of O-Glycosylation Sites Using beta-Elimination and Partial Acid Hydrolysis." Analytical Chemistry, 73 (2001): 1263–1269.

Munch, G., Gerlach, M., Sian, J., Wong, A. & Riederer, P. "Advanced glycation end products in neurodegeneration: More than early markers of oxidative stress?" Annals of Neurology, 44 (1998): S85–S88.

Nakajima, R., Imanaka, T. & Aiba, S. "Comparison of amino acid sequences of eleven different alpha-amylases." Applied Microbiology and Biotechnology, 23 (1986): 355–360.

Nazmi, A. R., Reinisch, T. & Hinz, H. J. "Ca-binding to *Bacillus licheniformis* alpha-amylase (BLA)." Archives of Biochemistry and Biophysics, 453 (2006): 16–23.

Nicolls, M. R. "The Clinical and Biological Relationship between Type II Diabetes Mellitus and Alzheimer's Disease." Current Alzheimer Research, 1 (2004): 47–54.

Nielsen, J. E. & Borchert, T. V. "Protein engineering of bacterial alpha-amylases." Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology, 1543 (2000): 253–274.

Nielsen, J. E., Borchert, T. V. & Vriend, G. "The determinants of alpha-amylase pH−activity profiles." Protein Engineering, 14 (2001): 505–512.

Nielsen, M. M., Bozonnet, S., Seo, E. S., Motyan, J. A., Andersen, J. M., Dilokpimol, A., Hachem, M. A., Gyemant, G., Naested, H., Kandra, L., Sigurskjold, B. W. & Svensson, B. "Two Secondary Carbohydrate Binding Sites on the Surface of Barley alpha-Amylase 1 Have Distinct Functions and Display Synergy in Hydrolysis of Starch Granules." Biochemistry, 48 (2009): 7686–7697.

Nielsen, M. M., Seo, E. S., Dilokpimol, A., Andersen, J., Hachem, M. A., Naested, H., Willemoes, M., Bozonnet, S., Kandra, L., Gyemant, G., Haser, R., Aghajari, N. & Svensson, B. "Roles of multiple surface sites, long substrate binding clefts, and carbohydrate binding modules in the action of amylolytic enzymes on polysaccharide substrates." Biocatalysis and Biotransformation, 26 (2008): 59–67.

Olsen, H. S., Damhus, T., Kaasgaard, S. & Lundquist, H. Enzymes at Work 3. ed. Novozymes, Bagsvaerd (2008).

Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S. & Mann, M. "Higher-energy C-trap dissociation for peptide modification analysis." Nature Methods, 4 (2007): 709–712.

Ong, S. E. & Mann, M. "Mass spectrometry-based proteomics turns quantitative." Nature Chemical Biology, 1 (2005): 252–262.

Pappin, D. J., Hojrup, P. & Bleasby, A. J. "Rapid identification of proteins by peptide-mass fingerprinting." Current Biology, 3 (1993): 327–332.

Patel, K. & Borchardt, R. T. "Chemical Pathways of Peptide Degradation. II. Kinetics of Deamidation of an Asparaginyl Residue in a Model Hexapeptide." Pharmaceutical Research, 7 (1990): 703–711.

Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J. S. "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis, 20 (1999): 3551–3567.

Petersen, K. K., Hansen, J. & Krogstrup, P. "Significance of different carbon sources and sterilization methods on callus induction and plant regeneration of *Miscanthus x ogiformis* Honda 'Giganteus'." Plant Cell Tissue and Orgen Culture, 58 (1999): 189–197.

Phanstiel, D., Unwin, R., McAlister, G. C. & Coon, J. J. "Peptide Quantification Using 8-Plex Isobaric Tags and Electron Transfer Dissociation Tandem Mass Spectrometry." Analytical Chemistry, 81 (2009): 1693–1698.

Phanstiel, D., Zhang, Y., Marto, J. A. & Coon, J. J. "Peptide and Protein Quantification Using iTRAQ with Electron Transfer Dissociation." Journal of the American Society for Mass Spectrometry, 19 (2008): 1255–1262.

Pichler, P., Kocher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G. & Mechtler, K. "Peptide Labeling with Isobaric Tags Yields Higher Identification Rates Using iTRAQ 4-Plex Compared to TMT 6-Plex and iTRAQ 8-Plex on LTQ Orbitrap." Analytical Chemistry, 82 (2010): 6549–6558.

Powell, B. S., Enama, J. T., Ribot, W. J., Webster, W., Little, S., Hoover, T., Adamovicz, J. J. & Andrews, G. P. "Multiple asparagine deamidation of Bacillus anthracis protective antigen causes charge Isoforms whose complexity correlates with reduced biological activity." Proteins-Structure Function and Bioinformatics, 68 (2007): 458–479.

Priego-Capote, F. & Sanchez, J. C. "Strategies for proteomic analysis of non–enzymatically glycated proteins." Mass Spectrometry Reviews, 28 (2009): 135–146.

Priego-Capote, F., Scherl, A., Muller, M., Waridel, P., Lisacek, F. & Sanchez, J. C. "Glycation Isotopic Labeling with C-13-Reducing Sugars for Quantitative Analysis of Glycated Proteins in Human Plasma." Molecular & Cellular Proteomics, 9 (2010): 579–592.

Quan,C., Alcala, E., Petkovska, I., Matthews, D., Canova-Davis, E., Taticek, R. & Ma, S. "A study in glycation of a therapeutic recombinant humanized monoclonal antibody: Where it is, how it got there, and how it affects charge-based behavior." Analytical Biochemistry, 373 (2008): 179–191.

Radestock, S. & Gohlke, H.. "Protein rigidity and thermophilic adaptation." Proteins-Structure Function amd Bioinformatics, 79 (2011): 1089–1108.

Ragunath, C., Manuel, S. G. A., Venkataraman, V., Sait, H. B. R., Kasinathan, C. & Ramasubbu, N. "Probing the Role of Aromatic Residues at the Secondary Saccharide-Binding Sites of Human Salivary alpha-Amylase in Substrate Hydrolysis and Bacterial Binding." Journal of Molecular Biology, 384 (2008): 1232–1248.

Ren, D., Pipes, G. D., Liu, D. J., Shih, L. Y., Nichols, A. C., Treuheit, M. J., Brems, D. N. & Bondarenko, P. V. "An improved trypsin digestion method minimizes digestion-induced modifications on proteins." Analytical Biochemistry, 392 (2009): 12–21.

Robert, X., Haser, R., Gottschalk, T. E., Ratajczak, F., Driguez, H., Svensson, B. & Aghajari, N. "The structure of barley alpha-amylase isozyme 1 reveals a novel role of domain C in substrate recognition and binding: a pair of sugar tongs." Structure, 11 (2003): 973–984.

Robinson, N. E. "Protein Deamidation." Proceedings of the National Academy of Sciences of the United States of America, 99 (2002): 5283–5288.

Robinson, N. E. & Robinson, A. B. "Prediction of Protein Deamidation Rates from Primary and Three-Dimensional Structure." Proceedings of the National Academy of Sciences of the United States of America, 98 (2001): 4367–4372.

Roepstorff, P. & Fohlman, J. "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomedical Mass Spectrometry, 11 (1984): 601.

Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents." Molecular & Cellular Proteomics, 3 (2004): 1154–1169.

Sadygov, R. G., Cociorva, D. & Yates, J. R. "Large-scate database searching using tandem mass spectra: Looking up the answer in the back of the book." Nature Methods, 1 (2004): 195–202.

Scherl, A., Shaffer, S. A., Taylor, G. K., Hernandez, P., Appel, R. D., Binz, P. A. & Goodlett, D. R. "On the Benefits of Acquiring Peptide Fragment Ions at High Measured Mass Accuracy." Journal of the American Society for Mass Spectrometry, 19 (2008): 891–901.

Schnolzer, M., Jedrzejewski, P. & Lehmann, W. D.. "Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry." Electrophoresis, 17 (1996): 945–953.

Shaw, A., Bott, R. & Day, A. G. "Protein engineering of alpha-amylase for low pH performance." Current Opinion in Biotechnology, 10 (1999): 349–352.

Steen, H. & Mann M. "The ABC's (and XYZ's) of peptide sequencing." Nature Reviews Molecular Cell Biology, 5 (2004): 699–711.

Stratton, L. P., Kelly, R. M., Rowe, J., Shively, J. E., Smith, D. D., Carpenter, J. F. & Manning, M. C. "Controlling deamidation rates in a model peptide: Effects of temperature, peptide concentration, and additives." Journal of Pharmaceutical Sciences, 90 (2001): 2141–48.

Svendsen, I., Hejgaard, J. & Mundy J. "Complete amino acid sequence of the alpha-amylase/subtilisin inhibitor from barley." Carlsberg Research Communications, 51 (1986): 43–50.

Svensson, B. "Regional distant sequence homology between amylases, alpha-glucosidases and transglucanosylases." FEBS Letters, 230 (1988): 72–76.

Svensson, B. "Protein engineering in the alpha-amylase family: catalytic mechanism, substrate specificity, and stability." Plant Molecular Biology, 25 (1994): 141–157.

Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., Syka, J. E. P. & Coon, J. J. "Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors." Analytical Chemistry, 79 (2007): 477–485.

Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry." Proceedings of the National Academy of Sciences of the United States of America, 101 (2004): 9528–9533.

Terashima, I., Koga, A. & Nagai A. "Identification of deamidation and isomerization sites on pharmaceutical recombinant antibody using $H_2{}^{18}O$." Analytical Biochemistry, 368 (2007): 49–60.

Tester, R. F. & Karkalas, K. "Starch." In Polysaccharides and polyamides in the food industry: properties, production and patents. Editors Steinbüchel, A. & Rhee, S. K. Wiley-VCH, Weinheim (2005): Chapter 12.

Thermo Scientific. "Instructions TMT Mass Tagging Kits and Reagents." Pierce Biotechnology, Rockford (2012).

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T. & Hamon, C. "Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS." Analytical Chemistry, 75 (2003): 1895–1904.

Thorpe, S. R. & Baynes, J. W. "Role of the maillard reaction in diabetes mellitus and diseases of aging." Drugs & Aging, 9 (1996): 69–77.

Tomazic, S. J. & Klibanov, A. M. "Mechanisms of irreversible thermal inactivation of *Bacillus* alpha-amylases." The Journal of Biological Chemistry, 263 (1988a): 3086–3091.

Tomazic, S. J. & Klibanov, A. M. "Why is one *Bacillus* alpha-amylase more resistant against irreversible thermoinactivation than another?" The Journal of Biological Chemistry, 263 (1988b): 3092–3096.

Valle, B.L., Stein, E. A., Sumerwell, W. N. & Fischer, E. H. "Metal content of alpha-amylases of various origins." The Journal of Biological Chemistry, 234 (1959): 2901–2905.

Vallée, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K. W., Svensson, B. & Haser, R. "Barley alpha-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 A resolution." Structure, 6 (1998): 649–659.

van der Maarel, M. J. E. C., van der Veen, B., Uitdehaag, J. C. M., Leemhuis, H., & Dijkhuizen, L. "Properties and applications of starch-converting enzymes of the alpha-amylase family." Journal of Biotechnology, 94 (2002): 137–155.

van Wandelen, C. & Cohen, S. A. "Using quaternary high-performance liquid chromatography eluent systems for separating 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate-derivatized amino acid mixtures." Journal of Chromatography A, 763 (1997): 11–22.

Vieille, C. & Zeikus G. J. "Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability." Microbiology and Molecular Biology Reviews, 65 (2001): 1–43.

Vihinen, M. & Mantsala, P. "Microbial amylolytic enzymes." Critical Reviews in Biochemistry and Molecular Biology, 24 (1989): 329–418.

Viner, R. I., Zhang, T., Second, T. & Zabrouskov, V. "Quantification of post-translationally modified peptides of bovine alpha-crystallin using tandem mass tags and electron transfer dissociation." Journal of Proteomics, 72 (2009): 874–885.

Wilm, M. & Mann, M. "Analytical Properties of the Nanoelectrospray Ion Source." Analytical Chemistry, 68 (1996): 1–8.

Wysocki, V. H., Tsaprailis, G., Smith, L. L. & Breci, L. A. "Mobile and localized protons: a framework for understanding peptide dissociation." Journal of Mass Spectrometry, 35 (2000): 1399–1406.

Xie, M. & Schowen, R. L.. "Secondary structure and protein deamidation." Journal of Pharmaceutical Sciences, 88 (1999): 8–13.

Yang, H. Q. & Zubarev, R. A. "Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides." Electrophoresis, 31 (2010): 1764–1772.

Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N. & Udaka, S. "Complete nucleotide sequence of a gene coding for heat- and pH-stable alpha-amylase of *Bacillus licheniformis*: comparison of the amino acid sequences of three bacterial liquefying alpha-amylases deduced from the DNA sequences." Journal of Biochemistry, 98 (1985): 1147–1156.

Zhang, Q., Frolov, A., Tang, N., Hoffmann, R., van de Goor, T., Metz, T. O. & Smith, R. D. "Application of electron transfer dissociation mass spectrometry in analyses of non-enzymatically glycated peptides." Rapid Communications in Mass Spectrometry, 21 (2007a): 661–66.

Zhang, Q., Tang, N., Brock, J. W. C., Mottaz, H. M., Ames, J. M., Baynes, J. W., Smith, R.D. & Metz, T. O. "Enrichment and analysis of nonenzymatically glycated peptides: boronate affinity chromatography coupled with electron-transfer dissociation mass spectrometry." Journal of Proteome Research, 6 (2007b): 2323–2330.

Zhang, Q., Schepmoes, A. A., Brock, J. W. C., Wu, S., Moore, R. J., Purvine, S. O., Baynes, J. W., Smith, R. D. & Metz, T. O. "Improved Methods for the Enrichment and Analysis of Glycated Peptides." Analytical Chemistry, 80 (2008a): 9822–9829.

Zhang, Q., Petyuk, V. A., Schepmoes, A. A., Orton, D. J., Monroe, M. E., Yang, F., Smith, R. D. & Metz, T. O."Analysis of non-enzymatically glycated peptides: neutral-loss-triggered MS(3) versus multi-stage activation tandem mass spectrometry." Rapid Communications in Mass Spectrometry, 22 (2008b): 3027–3034.

Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., Rush, J., Lauffenburger, D. A. & White, F. M. "Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules." Molecular & Cellular Proteomics, 4 (2005): 1240–1250.

Zubarev, R. A. "Electron-capture dissociation tandem mass spectrometry." Current Opinion in Biotechnology, 15 (2004): 12–16.

Zubarev, R. A., Kelleher, N. L. & McLafferty, F. W. "Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process." Journal of the American Chemical Society, 120 (1998): 3265–3266.

# 7 Appendices

## 7.1 Appendix I - Supplementary figures

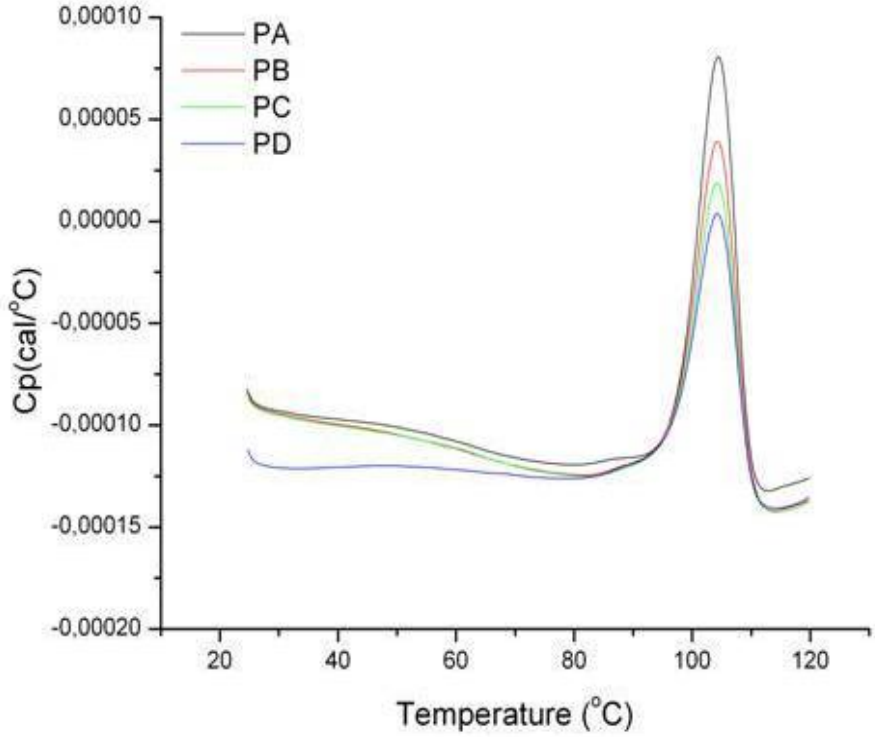



**Figure 1.** Unfolding of the four partly separated pools of BLA (PA – PD) determined by DSC.

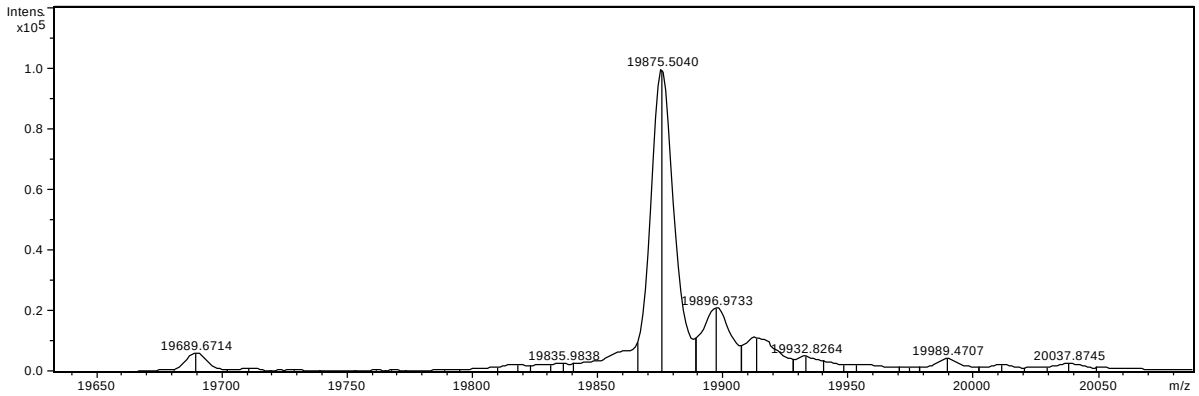
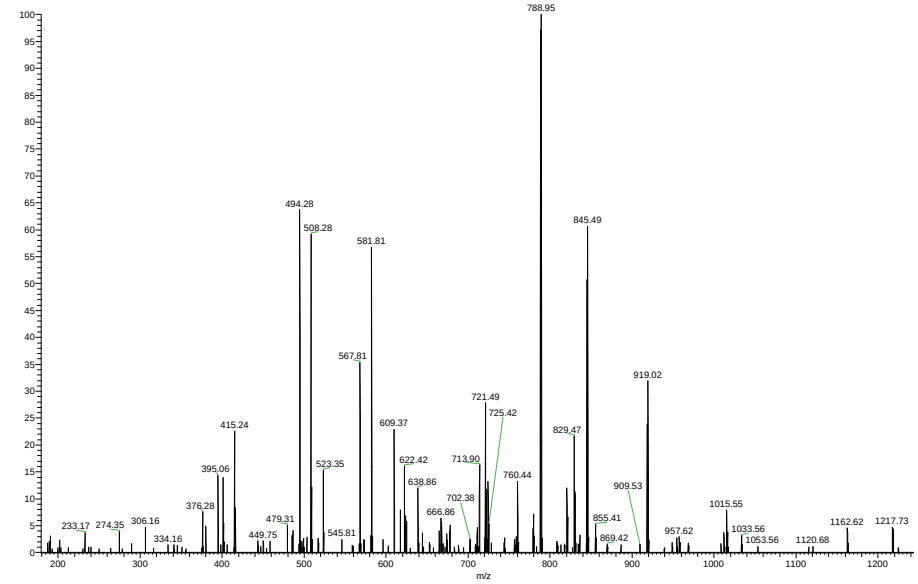
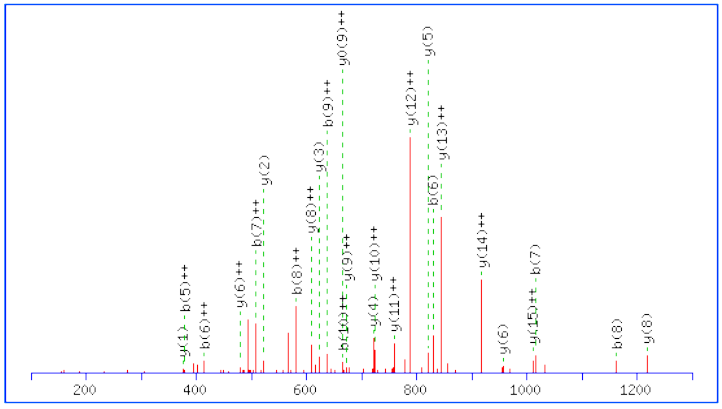



**Figure 2.** Full-length MS of BASI.

**A)**




**B)**

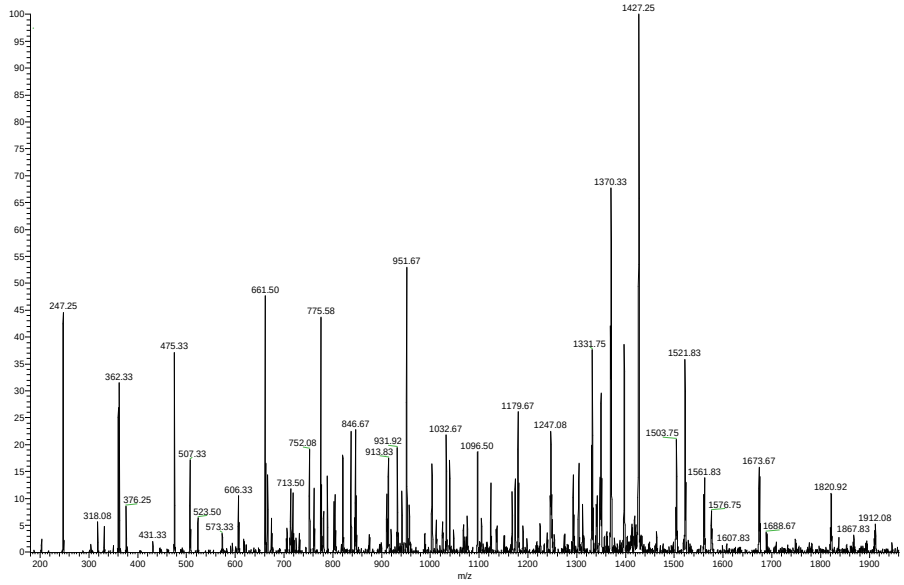



**C)**

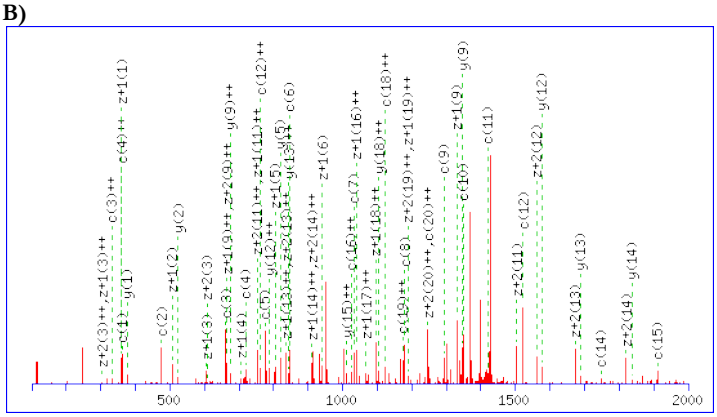
| # | b | $b^{++}$ | b* | $b^{*++}$ | $b^{0}$ | $b^{0++}$ | Seq. | y | $y^{++}$ | y* | $y^{*++}$ | $y^{0}$ | $y^{0++}$ | # |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 345.1972 | 173.1022 | | | 327.1866 | 164.0969 | D | | | | | | | 21 |
| 2 | 458.2812 | 229.6442 | | | 440.2706 | 220.6390 | L | 2506.3696 | 1253.6884 | 2489.3430 | 1245.1752 | 2488.3590 | 1244.6831 | 20 |
| 3 | 644.3817 | 322.6945 | 627.3551 | 314.1812 | 626.3711 | 313.6892 | K | 2393.2855 | 1197.1464 | 2376.2590 | 1188.6331 | 2375.2750 | 1188.1411 | 19 |
| 4 | 701.4031 | 351.2052 | 684.3766 | 342.6919 | 683.3926 | 342.1999 | G | 2207.1851 | 1104.0962 | 2190.1585 | 1095.5829 | 2189.1745 | 1095.0909 | 18 |
| 5 | 758.4246 | **379.7159** | 741.3980 | 371.2027 | 740.4140 | 370.7106 | G | 2150.1636 | 1075.5854 | 2133.1371 | 1067.0722 | 2132.1531 | 1066.5802 | 17 |
| 6 | **829.4617** | **415.2345** | 812.4352 | 406.7212 | 811.4511 | 406.2292 | A | 2093.1422 | 1047.0747 | 2076.1156 | 1038.5614 | 2075.1316 | 1038.0694 | 16 |
| 7 | **1015.5410** | **508.2741** | 998.5145 | 499.7609 | 997.5304 | 499.2689 | W | 2022.1050 | 1011.5562 | 2005.0785 | 1003.0429 | 2004.0945 | 1002.5509 | 15 |
| 8 | **1162.6094** | **581.8084** | 1145.5829 | 573.2951 | 1144.5989 | 572.8031 | F | 1836.0257 | 918.5165 | 1818.9992 | 910.0032 | 1818.0152 | 909.5112 | 14 |
| 9 | 1275.6935 | **638.3504** | 1258.6669 | 629.8371 | 1257.6829 | 629.3451 | L | 1688.9573 | 844.9823 | 1671.9308 | 836.4690 | 1670.9467 | 835.9770 | 13 |
| 10 | 1332.7150 | **666.8611** | 1315.6884 | 658.3478 | 1314.7044 | 657.8558 | G | 1575.8732 | 788.4403 | 1558.8467 | 779.9270 | 1557.8627 | 779.4350 | 12 |
| 11 | 1403.7521 | 702.3797 | 1386.7255 | 693.8664 | 1385.7415 | 693.3744 | A | 1518.8518 | 759.9295 | 1501.8252 | 751.4163 | 1500.8412 | 750.9242 | 11 |
| 12 | 1504.7997 | 752.9035 | 1487.7732 | 744.3902 | 1486.7892 | 743.8982 | T | 1447.8147 | 724.4110 | 1430.7881 | 715.8977 | 1429.8041 | 715.4057 | 10 |
| 13 | 1633.8423 | 817.4248 | 1616.8158 | 808.9115 | 1615.8318 | 808.4195 | E | 1346.7670 | 673.8871 | 1329.7404 | 665.3739 | 1328.7564 | 664.8819 | 9 |
| 14 | 1730.8951 | 865.9512 | 1713.8686 | 857.4379 | 1712.8845 | 856.9459 | P | ***1217.7244*** | 609.3658 | 1200.6978 | 600.8526 | | | 8 |
| 15 | 1893.9584 | 947.4829 | 1876.9319 | 938.9696 | 1875.9479 | 938.4776 | Y | 1120.6716 | 560.8395 | 1103.6451 | 552.3262 | | | 7 |
| 16 | 2031.0173 | 1016.0123 | 2013.9908 | 1007.4990 | 2013.0068 | 1007.0070 | H | ***957.6083*** | 479.3078 | 940.5818 | 470.7945 | | | 6 |
| 17 | 2130.0858 | 1065.5465 | 2113.0592 | 1057.0332 | 2112.0752 | 1056.5412 | V | ***820.5494*** | 410.7783 | 803.5228 | 402.2651 | | | 5 |
| 18 | 2229.1542 | 1115.0807 | 2212.1276 | 1106.5674 | 2211.1436 | 1106.0754 | V | ***721.4810*** | 361.2441 | 704.4544 | 352.7309 | | | 4 |
| 19 | 2328.2226 | 1164.6149 | 2311.1960 | 1156.1017 | 2310.2120 | 1155.6096 | V | ***622.4126*** | 311.7099 | 605.3860 | 303.1966 | | | 3 |
| 20 | 2475.2910 | 1238.1491 | 2458.2645 | 1229.6359 | 2457.2804 | 1229.1439 | F | ***523.3442*** | 262.1757 | 506.3176 | 253.6624 | | | 2 |
| 21 | | | | | | | K | ***376.2757*** | 188.6415 | 359.2492 | 180.1282 | | | 1 |

**Figure 3.** MS/MS results obtained by HCD for the peptide D156 – K176 glycated with the AGE carboxymethyl on K158. A) The MS/MS spectra for the peptide. B) The MS/MS spectra from Mascot showing the peptide ions matched by Mascot. C) The theoretical peptide ions calculated for the peptide by Mascot. The Peptide ions matched by Mascot in the MS/MS data are shown in red.

**A)**




**B)**

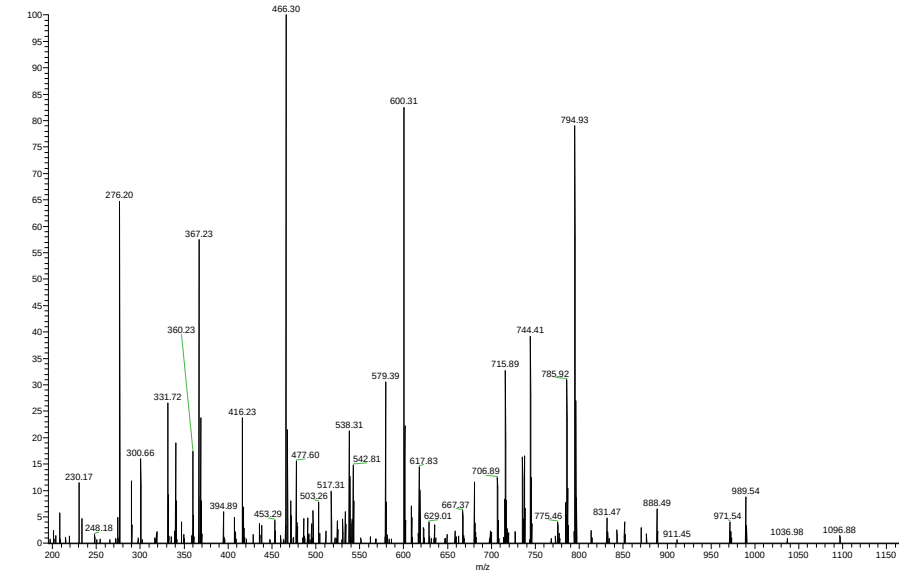



**C)**

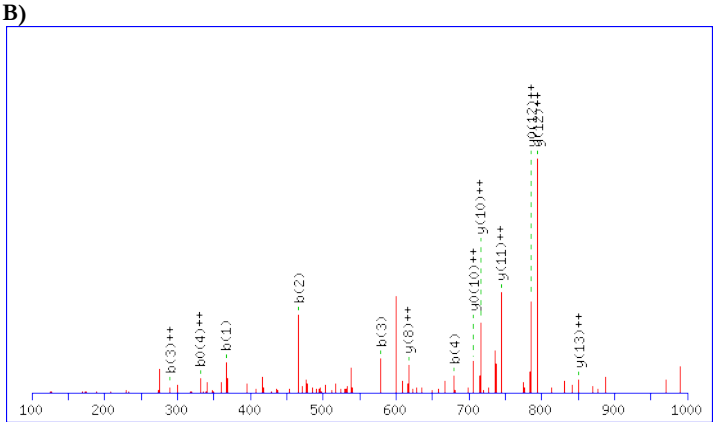
| # | c | c$^{++}$ | Seq. | y | y$^{++}$ | z+1 | z+1$^{++}$ | z+2 | z+2$^{++}$ | # |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ***362.2237*** | 181.6155 | D | | | | | | | 21 |
| 2 | ***475.3078*** | 238.1575 | L | 2506.3696 | 1253.6884 | 2490.3509 | 1245.6791 | 2491.3587 | **1246.1830** | 20 |
| 3 | ***661.4082*** | **331.2077** | K | 2393.2855 | 1197.1464 | 2377.2668 | **1189.1370** | 2378.2746 | **1189.6409** | 19 |
| 4 | ***718.4297*** | **359.7185** | G | 2207.1851 | 1104.0962 | 2191.1664 | **1096.0868** | 2192.1742 | 1096.5907 | 18 |
| 5 | ***775.4511*** | 388.2292 | G | 2150.1636 | 1075.5854 | 2134.1449 | **1067.5761** | 2135.1527 | 1068.0800 | 17 |
| 6 | ***846.4882*** | 423.7478 | A | 2093.1422 | 1047.0747 | 2077.1234 | **1039.0654** | 2078.1313 | 1039.5693 | 16 |
| 7 | ***1032.5676*** | 516.7874 | W | 2022.1050 | 1011.5562 | 2006.0863 | 1003.5468 | 2007.0941 | 1004.0507 | 15 |
| 8 | ***1179.6360*** | 590.3216 | F | **1836.0257** | 918.5165 | 1820.0070 | **910.5071** | **1821.0148** | **911.0111** | 14 |
| 9 | ***1292.7200*** | 646.8637 | L | **1688.9573** | 844.9823 | 1672.9386 | **836.9729** | **1673.9464** | **837.4768** | 13 |
| 10 | ***1349.7415*** | 675.3744 | G | **1575.8732** | 788.4403 | 1559.8545 | 780.4309 | **1560.8623** | 780.9348 | 12 |
| 11 | ***1420.7786*** | 710.8929 | A | 1518.8518 | 759.9295 | 1502.8331 | **751.9202** | **1503.8409** | **752.4241** | 11 |
| 12 | ***1521.8263*** | **761.4168** | T | 1447.8147 | 724.4110 | 1431.7959 | 716.4016 | 1432.8038 | 716.9055 | 10 |
| 13 | 1650.8689 | 825.9381 | E | **1346.7670** | 673.8871 | ***1330.7483*** | **665.8778** | 1331.7561 | **666.3817** | 9 |
| 14 | ***1747.9217*** | 874.4645 | P | 1217.7244 | 609.3658 | 1201.7057 | 601.3565 | 1202.7135 | 601.8604 | 8 |
| 15 | ***1910.9850*** | 955.9961 | Y | 1120.6716 | 560.8395 | 1104.6529 | 552.8301 | 1105.6607 | 553.3340 | 7 |
| 16 | 2048.0439 | **1024.5256** | H | 957.6083 | 479.3078 | ***941.5896*** | 471.2984 | 942.5974 | 471.8023 | 6 |
| 17 | 2147.1123 | 1074.0598 | V | **820.5494** | 410.7783 | ***804.5307*** | 402.7690 | 805.5385 | 403.2729 | 5 |
| 18 | 2246.1807 | **1123.5940** | V | 721.4810 | 361.2441 | ***705.4623*** | 353.2348 | 706.4701 | 353.7387 | 4 |
| 19 | 2345.2491 | **1173.1282** | V | 622.4126 | 311.7099 | ***606.3938*** | **303.7006** | **607.4017** | **304.2045** | 3 |
| 20 | 2492.3175 | **1246.6624** | F | **523.3442** | 262.1757 | ***507.3254*** | 254.1664 | 508.3333 | 254.6703 | 2 |
| 21 | | | K | **376.2757** | 188.6415 | ***360.2570*** | 180.6321 | 361.2648 | 181.1361 | 1 |

**Figure 4.** MS/MS results obtained by ETD for the peptide D156 – K176 glycated with the AGE carboxymethyl on K158. A) The MS/MS spectra for the peptide. B) The MS/MS spectra from Mascot showing the peptide ions matched by Mascot. C) The theoretical peptide ions calculated for the peptide by Mascot. The Peptide ions matched by Mascot in the MS/MS data are shown in red.

**A)**




**B)**

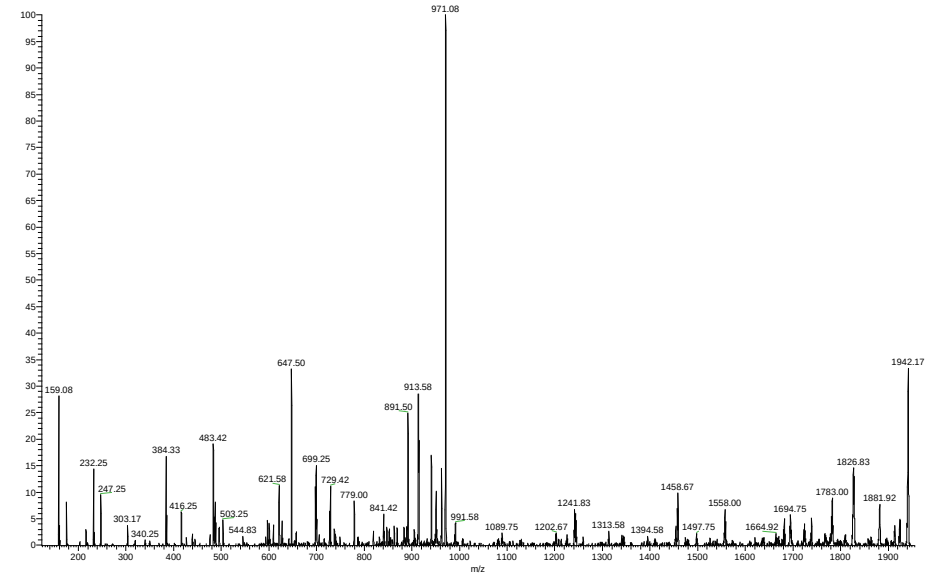



**C)**

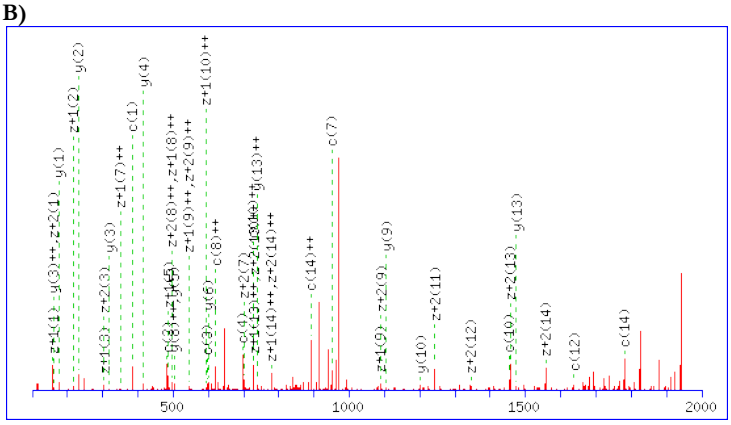
| # | b | $b^{++}$ | b* | $b^{*++}$ | $b^{0}$ | $b^{0++}$ | Seq. | y | $y^{++}$ | y* | $y^{*++}$ | $y^{0}$ | $y^{0++}$ | # |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ***367.2291*** | 184.1182 | | | | | H | | | | | | | 15 |
| 2 | ***466.2975*** | 233.6524 | | | | | V | 1800.9905 | 900.9989 | 1783.9639 | 892.4856 | 1782.9799 | 891.9936 | 14 |
| 3 | ***579.3816*** | **290.1944** | | | | | I | 1701.9220 | 851.4647 | 1684.8955 | 842.9514 | 1683.9115 | 842.4594 | 13 |
| 4 | ***680.4293*** | 340.7183 | | | 662.4187 | 331.7130 | T | 1588.8380 | 794.9226 | 1571.8114 | 786.4094 | 1570.8274 | 785.9173 | 12 |
| 5 | 737.4507 | 369.2290 | | | 719.4402 | 360.2237 | G | 1487.7903 | 744.3988 | 1470.7638 | 735.8855 | 1469.7797 | 735.3935 | 11 |
| 6 | 834.5035 | 417.7554 | | | 816.4929 | 408.7501 | P | 1430.7688 | 715.8881 | 1413.7423 | 707.3748 | 1412.7583 | 706.8828 | 10 |
| 7 | 933.5719 | 467.2896 | | | 915.5614 | 458.2843 | V | 1333.7161 | 667.3617 | 1316.6895 | 658.8484 | 1315.7055 | 658.3564 | 9 |
| 8 | 1290.8298 | 645.9185 | 1273.8033 | 637.4053 | 1272.8192 | 636.9133 | K | 1234.6477 | 617.8275 | 1217.6211 | 609.3142 | 1216.6371 | 608.8222 | 8 |
| 9 | 1405.8568 | 703.4320 | 1388.8302 | 694.9187 | 1387.8462 | 694.4267 | D | 877.3898 | 439.1985 | 860.3632 | 430.6852 | 859.3792 | 430.1932 | 7 |
| 10 | 1502.9095 | 751.9584 | 1485.8830 | 743.4451 | 1484.8990 | 742.9531 | P | 762.3628 | 381.6850 | 745.3363 | 373.1718 | 744.3523 | 372.6798 | 6 |
| 11 | 1589.9415 | 795.4744 | 1572.9150 | 786.9611 | 1571.9310 | 786.4691 | S | 665.3101 | 333.1587 | 648.2835 | 324.6454 | 647.2995 | 324.1534 | 5 |
| 12 | 1686.9943 | 844.0008 | 1669.9678 | 835.4875 | 1668.9837 | 834.9955 | P | 578.2780 | 289.6427 | 561.2515 | 281.1294 | 560.2675 | 280.6374 | 4 |
| 13 | 1774.0263 | 887.5168 | 1756.9998 | 879.0035 | 1756.0158 | 878.5115 | S | 481.2253 | 241.1163 | 464.1987 | 232.6030 | 463.2147 | 232.1110 | 3 |
| 14 | 1831.0478 | 916.0275 | 1814.0213 | 907.5143 | 1813.0372 | 907.0223 | G | 394.1932 | 197.6003 | 377.1667 | 189.0870 | | | 2 |
| 15 | | | | | | | R | 337.1718 | 169.0895 | 320.1452 | 160.5763 | | | 1 |

**Figure 5.** MS/MS results obtained by HCD for the peptide H108 – R122 glycated with glycation on K115. A) The MS/MS spectra for the peptide. B) The MS/MS spectra from Mascot showing the peptide ions matched by Mascot. C) The theoretical peptide ions calculated for the peptide by Mascot. The Peptide ions matched by Mascot in the MS/MS data are shown in red.

**A)**



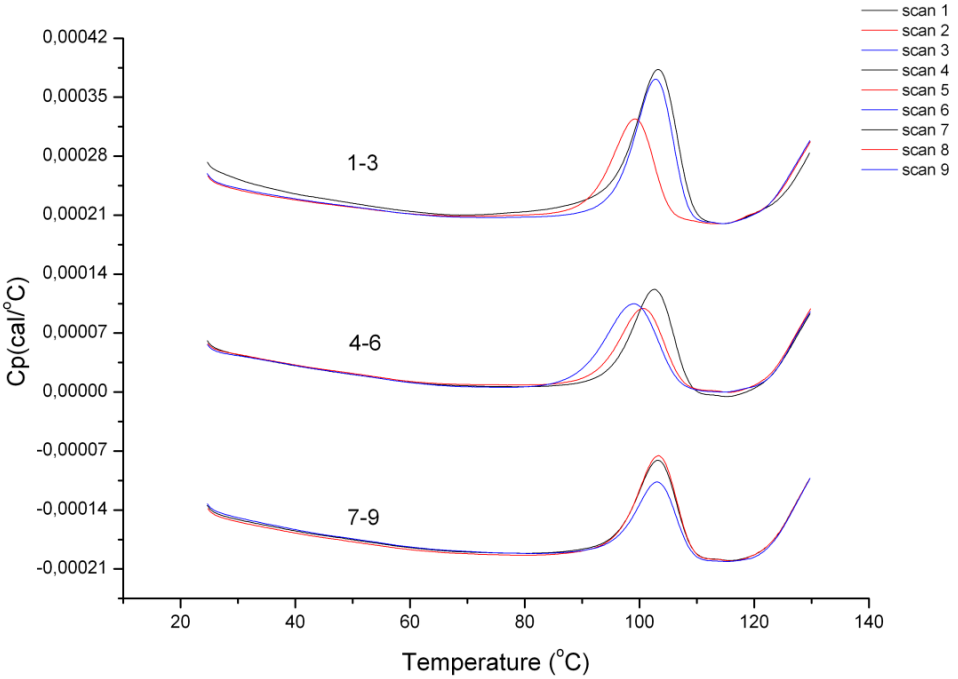

**B)**




**C)**

| # | c | $c^{++}$ | Seq. | y | $y^{++}$ | z+1 | $z+1^{++}$ | z+2 | $z+2^{++}$ | # |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ***384.2557*** | 192.6315 | H | | | | | | | 15 |
| 2 | ***483.3241*** | 242.1657 | V | 1571.8275 | 786.4174 | 1555.8088 | **778.4080** | **1556.8166** | **778.9120** | 14 |
| 3 | ***596.4081*** | 298.7077 | I | ***1472.7591*** | 736.8832 | 1456.7404 | **728.8738** | **1457.7482** | **729.3777** | 13 |
| 4 | ***697.4558*** | 349.2316 | T | 1359.6751 | 680.3412 | 1343.6563 | 672.3318 | **1344.6642** | 672.8357 | 12 |
| 5 | 754.4773 | 377.7423 | G | 1258.6274 | 629.8173 | 1242.6086 | 621.8080 | **1243.6165** | 622.3119 | 11 |
| 6 | 851.5301 | 426.2687 | P | ***1201.6059*** | 601.3066 | 1185.5872 | **593.2972** | 1186.5950 | 593.8011 | 10 |
| 7 | ***950.5985*** | 475.8029 | V | ***1104.5531*** | 552.7802 | **1088.5344** | **544.7708** | **1089.5422** | **545.2748** | 9 |
| 8 | 1240.7463 | **620.8768** | K | 1005.4847 | 503.2460 | 989.4660 | **495.2366** | 990.4738 | **495.7406** | 8 |
| 9 | 1355.7732 | 678.3902 | D | 715.3369 | 358.1721 | 699.3182 | **350.1627** | **700.3260** | 350.6667 | 7 |
| 10 | ***1452.8260*** | **726.9166** | P | ***600.3100*** | 300.6586 | 584.2913 | 292.6493 | 585.2991 | 293.1532 | 6 |
| 11 | 1539.8580 | 770.4326 | S | ***503.2572*** | 252.1323 | **487.2385** | 244.1229 | 488.2463 | 244.6268 | 5 |
| 12 | ***1636.9108*** | 818.9590 | P | ***416.2252*** | 208.6162 | 400.2065 | 200.6069 | 401.2143 | 201.1108 | 4 |
| 13 | 1723.9428 | 862.4750 | S | ***319.1724*** | 160.0899 | **303.1537** | 152.0805 | **304.1615** | 152.5844 | 3 |
| 14 | ***1780.9642*** | **890.9858** | G | ***232.1404*** | 116.5738 | **216.1217** | 108.5645 | 217.1295 | 109.0684 | 2 |
| 15 | | | R | ***175.1190*** | 88.0631 | **159.1002** | 80.0538 | **160.1081** | 80.5577 | 1 |

**Figure 6.** MS/MS results obtained by HCD for the peptide H108 – R122 glycated with glycation on K115. A) The MS/MS spectra for the peptide. B) The MS/MS spectra from Mascot showing the peptide ions matched by Mascot. C) The theoretical peptide ions calculated for the peptide by Mascot. The Peptide ions matched by Mascot in the MS/MS data are shown in red.




| | |
|---|---|
| Scan 1 | BLA |
| Scan 2 | Q264E |
| Scan 3 | N265D |
| Scan 4 | Q360E |
| Scan 5 | Q9E |
| Scan 6 | N172D |
| Scan 7 | K306M |
| Scan 8 | K383M |
| Scan 9 | R437L |

**Figure 7.** Unfolding of BLA and the BLA variants determined by DSC.

## 7.2 Appendix II – Manuscript 1

Manuscript in preparation for Applied Biochemistry and Biotechnology

**Characterisation of chemical modifications on *Bacillus licheniformis* α-amylase and the implications on detergent wash performance**

C. Nielsen, S. G. Kaasgaard, C. Andersen, C. P. Sönksen, S. Jacobsen, B. Svensson

C. Nielsen, S. G. Kaasgaard, C. Andersen, C. P. Sönksen
Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

C. Nielsen, S. Jacobsen, B. Svensson B
Enzyme and Protein Chemistry, Department of Systems Biology, Soeltofts Plads Building 24, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

**Abstract**

*Bacillus licheniformis* α-amylase (BLA) is an important industrial enzyme used in the detergent industry. The enzyme becomes modified during the fermentation resulting in multiple forms varying in pI value. Deamidation of asparagine and glutamine residues and glycations on lysine and argenines are partly responsible for these multiple forms. Tandem mass spectrometry (MS/MS) was used to identify deamidation and glycation sites on the BLA forms. The modifications were mainly found on the surface of the enzyme and some of the modifications are located in sites involved in stability, activity and substrate binding. The influence of some of the specific modifications were determined by producing BLA mutants with site specific mutations in selected modification sites in the area around the active site and in sites suspected to be involved in binding of longer substrates and secondary binding sites. $K_M$ and $k_{cat}$ values were determined on maltoheptaose substrate and and on amylose for the mutants along with the wash performance in a liquid and a powder detergent to determine the influence of the single modifications. The mutations mimicking deamidations (N→D and Q→E) influenced the $K_M$ and $k_{cat}$ values in the sites around the active site differently, though they generally improved the wash performance especially the BLA mutant N265D showed great improvement. Removing the glucose reactive sites (K→M and R→L) in sites related to binding of longer substrates and secondary binding sites did not influence the kinetic values $K_M$ and $k_{cat}$ but it improved the wash performance especially in liquid detergent.

**Introduction**

α-Amylases are a family of enzymes which catalyse the hydrolysis of α-D-(1,4) glucosidic linkages in starch and related compounds hereby releasing maltooligosaccharides and glucose in the α-anomeric configuration [1, 2]. α-Amylases are among the most important industrial enzymes with widespread use in various industries. The enzyme is used for starch liquefaction and saccharification in the starch and fuel alcohol industries, for starch stain removal in the detergent industry as well as in industries such as baking and brewing [3–5].

*Bacillus licheniformis* α-amylase (BLA) is widely used in industrial processes as it is thermostable and stable at alkaline conditions, the latter being an important feature in the detergent industry where α-amylases are the second most used enzyme after proteases [4–6].

The structure of BLA is divided into the 3 main domains; A, B and C where domain A make up the core of the enzyme with domain B and C located on each side. Domain A contains the active site with the three active site residues Asp231, Glu261 and Asp328 [7]. BLA contains several substrate binding sites on the surface for interaction with the polymeric substrate. The binding site in the active site cleft consists of at least eight subsites each binding a glucosyl residue of the substrate. This site is formed by domain A and B [1, 8]. The interface between domain A and B also contains a quite unique Ca-Na-Ca metal triad. The three metal ions mainly situated in domain B are linearly arranged and involved in interactions with surrounding amino acid residues [1, 7]. The metal triad stabilises the structure around the active site and is important for the thermostability of BLA [9, 10].

Structural studies of amylases have revealed that non-catalytic binding of carbohydrates also can occur in secondary binding sites (SBS). These SBS are generally located on the surface more distant from the substrate binding cleft and the active site. SBS generally contain amino acids with aromatic side chains such as Trp and Tyr which are able to hydrophobically interact with carbohydrates. The exact role of these sites is not fully elucidated though they have been proposed to cooperate with the active site in degradation of different substrate, to be initial substrate recognition sites and to retaining reaction product for further processing by other proteins [11, 12].

Industrial α-amylases are produced by fermentation of genetically modified organisms such as bacteria or fungi. However, during the fermentation of recombinant proteins these can become chemically modified and the modifications can influence the enzyme stability and activity. Some of the most prevalent chemical modifications are deamidation, oxidation and glycations [13, 14].

Glycation is a non-enzymatic reaction between primary amines on proteins and the carbonyl group of a reducing sugar so that lysine and arginine side chains are prone to the reaction. The glycations form a ketoamine which further can undergo a series of rearrangement reactions to form a group of more reactive carbonyl compounds termed advanced glycation end products (AGEs) [15, 16]. These AGEs

include 3-deoxyglucosone, carboxymethyl, carboxyethyl, and methylglyoxal-derived hydroimidazolone (MG-HI) as some of the most described [17–20].

It has been observed that BLA becomes chemically modified during the fermentation resulting in multiple forms of BLA varying in pI value. The purpose of the present study is to unravel the modifications on BLA responsible for the multiple forms varying in pI with focus on glycations and deamidations. The influence of specific modifications at specific sites is evaluated for some of the most interesting modification sites. BLA variants are prepared and characterised to prove the functional roles of the specific amino acids and thus validate the role of specific modified sites in the function and stability of the enzyme at the conditions use in laundry detergent wash.

## Materials and Methods

### *Bacillus lichenisformis* α-Amylase

BLA expressed in recombinant *B. licheniformis* were provided by Novozymes.

### Isoelectric focusing (IEF)

FocusGel 4–6 (Gelcompany) were used with Amersham Isoelectric Focusing Calibration Kit (GE Healthcare) low range (pH 2.5–6.5) as standards and gels stained in Coomasie blue.

### In-gel digestion with $H_2{}^{18}O$

Bands of BLA were cut out of the IEF gel (twice) and destained twice in 50% ethanol. The gel pieces were dehydrated in 100% acetonitrile for 15 min. The following solutions were prepared with water-$^{18}O$, 97atom% $^{18}O$ (Sigma-Aldrich) to differentiate between deamidations present prior to sample preparation and those occurring during the digestion [21, 22]. The digest was performed with trypsin and Asp-N, adding 0.4 μg trypsin Sequencing grade (Roche Diagnostics) or 0.04 μg Endoproteinase

Asp-N Sequencing grade (Roche Diagnostics) in 50 mM $NH_4HCO_3$ and stored for 45 min at 5°C. Remaining protease solution was removed and 50 mM $NH_4HCO_3$ was added covering the gel piece and incubated at 37°C over night. Peptides were extracted from the gel pieces with 70% acetonitrile 0.1% triflouroacetic acid three times. The peptide samples were dried and dissolved in 5% formic acid.

Tandem mass spectrometry

The LC-MS/MS experiments were performed fragmenting with Higher energy C-trap dissociation (HCD) and electron transfer dissociation (ETD) as the latter method leaves labile modifications intact during the fragmentation hereby being better at identifying glycations. The experiments were run on a LTQ Orbitrap Velos (Thermo Scientific) for HCD fragmentation and on a LTQ Orbitrap XL (Thermo Scientific) for ETD fragmentation. Both instruments have a nano-ESI source (Nanospray Flexion Source, Thermo Scientific) coupled to an EASY-nLC II (Thermo Scientific). The peptides were separated on Thermo Scientific EASY-Column (C18) 2 cm x 100 µm i.d. column followed by a Thermo Scientific EASY-Column (C18) 10 cm x 75 µm i.d. column, eluting from 5–50% acetonitrile in 0.1% formic acid over 60 min at a flow rate of 250 nl/min.

Experiments on the Orbitrap Velos were conducted in the orbitrap with automatic gain control (AGC) target of 1,000,000 with 30,000 resolution and a maximum ion injection time of 10 ms in full MS and for MS/MS AGC target was 50,000 with injection time of 100 ms and resolution of 7500. The HCD normalized collision energy was 40% and the 10 most intense precursor ions in the range 350–1750 m/z were selected for MS/MS.

For analysis on the Orbitrap XL $MS^1$ was performed in the orbitrap mass analyser with a resolution of 60,000 (at 400 m/z). AGC target was 500,000 with a maximum ion injection time of 300 ms for the orbitrap in full MS. MS/MS was performed in the ion trap with AGC target of 40,000. ETD reactions were performed for 85 ms detecting fragment ions in the ion trap with a maximum injection time of 200 ms using 3 microscans selecting precursor ions in the range 350–1200 m/z, selecting the four most abundant precursor ions, rejecting charge state 1 and 2.

The data analysis was performed with Mascot (Matrix Science). The searches were conducted against a custom made database containing 14,100 sequences with decoy option enabled. The searches were conducted with trypsin or Asp-N as enzyme allowing 5 missed cleavages with 10 ppm precursor mass tolerance, 0.5 Da fragments mass tolerance for ETD and 0.02 Da mass tolerance for HCD. The following variable modifications were chosen: Deamidated (NQ), Deamidated: 18O(1) (NQ), Label: 18O(1)(C-term), Label: 18O(2)(C-term), Hex (KR), 3-deoxyglucosone (KR), Carboxymethyl (KR), Carboxyethyl (KR), MG-HI (KR), Oxidation (M), Carbamyl (KR), Acetyl (K).

Expression of BLA mutants

BLA mutants and BLA for comparison were produced by Novozymes. The constructs were made by site directed mutagenesis according to the principles described by others [23]. BLA and its variants was expressed in recombinant *B. subtilis*, fermented at 37°C for 96 h in a medium consisting of 100 mg/ml sucrose, 40 mg/ml soya flour, 10 mg/ml $Na_2HPO_4 \cdot 12H_2O$, 0.01 % v/v Dowfax 63N10, 5 mg/ml $CaCO_3$.

Purification of BLA mutants

The fermented samples were centrifuged at 5400 g for 50 min and filtered through 0.2 μm PES filters. MOPS, $CaCl_2$ and $(NH_4)_2SO_4$ were added to concentrations of 20 mM, 1 mM and 1 M respectively, and pH adjusted to 7.5. Samples were loaded to a phenyl sepharose column (Phenyl sepharose 6 fast flow, GE Healthcare) pre-equilibrated in 20 mM MOPS, 1 mM $CaCl_2$, 1 M $(NH_4)_2SO_4$, pH 7.5. Elution was done with $(NH_4)_2SO_4$ gradient (1M–0M). Amylase containing fractions were determined on NuPAGE 4–12% Bis-Tris gradient gel (Invitrogen). These fractions were pooled and dialysed over night in 10 mM Tris, 1 mM $CaCl_2$, pH 8.9.

The dialysed sample were loaded on to a Source 30Q column (GE healthcare) pre-equllibrated with 20 mM Tris, 1 mM $CaCl_2$, pH 8.9. Elution was done with NaCl gradient (0–600 mM) in 20 mM Tris, 1 mM $CaCl_2$, pH 8.9. Eluted fractions only containing BLA were pooled, concentrated and buffer

changed to 20 mM HEPES, 1 mM $CaCl_2$, pH 7.0 on Vivaspin 20 10,000 MWCO PES (Sartorius Stedim) spin columns. Protein concentration was determined by amino acid analysis [24, 25]. Homogeneity was checked on NuPAGE 4–12% Bis-Tris gel.

Differential scanning caliometry

Differential scanning calorimetry (DSC) measurements were carried out using a Microcal VP-capillary DSC (GE Healthcare). Scan rate of 1.5°C $min^{-1}$ was used to a final temperature of 130°C. Samples for analysis were approx. 1 mg/ml in 25 mM HEPES, 1 mM $CaCl_2$, pH 7.0. $T_{max}$ was determined as the temperature at the peak maximum of the transition from the folded to unfolded state. For data analysis the Origin software package (MicroCal, GE Healthcare) was used.

Maltoheptaose *p*-nitrophenol ($G_7$pNP) assay

The amylase kit AMYL α-amylase liquid (Roche) contains an α-glucosidase solution and a substrate solution (4,6-ethylidene-(G7)-pNP) was used in the assay. The initial rate of hydrolysis was determined at eleven substrate concentrations (0.025–3 mM) in 52.5 mM HEPES pH 7.0.
The assay were run on a KONELAB ARENA 30 (Thermo Electron Corporation) at 30 ºC mixing 147 µl α-glucosidase solution with 83 µl substrate, adding 20 µl enzyme solution (in 52.5 mM HEPES pH 7.0). After 3 min the colour formation was measured 7 times over 2 min at 405 nm. pNP standard (0–0.6 mM in 52.5 mM HEPES) was made from *p*-Nitrophenol (pNP) standard solution 10 mM (Sigma). $K_M$ and $k_{cat}$ were obtained by fitting to the Michaelis-Menten equation with the program CurveExpert 1.3 (http://www.curveexpert.net/).

Amylose assay

The initial rate of formation of reducing power was determined at 30 ºC at ten different amylose DP440 (Amylose type III: from potato, Sigma) solutions 0.1–2.5 mg/ml in 20 mM HEPES pH 7.0, 5

mM $CaCl_2$, 4.44% v/v dimethyl sulfoxide, 0.005% w/v bovine serum albumin (BSA) by amylase (in 20 mM HEPES pH 7.0, 5 mM $CaCl_2$, 0.005% w/v BSA) were determined using copper-bicinchoninate with maltose as standard [26, 27]. The colour formation was measured in spectrophotometer at $A_{540}$. $K_M$ and $k_{cat}$ were obtained as above.

Wash performance assay

Wash performance was determined on a mini wash robot (Novozymes) on coloured rice starch cotton CS-28 (Center For Testmaterials BV). CS-28 strips were mounted on racks which during the wash are moving up and down into the wash beakers (50 times per min). Washes were performed at 40°C for 30 min at six enzyme concentrations (0–1.0 mg /L) followed by 5 min rinsing in cold water. Washes were performed in 0.8 g/l liquid detergent pH 8.5 (20% linear alkylbenzene sulfonate (LAS, Nacconol 90G), 10% alkyl ether sulfate (STEOL CS-370E), 10% alcohol ethoxylate (AEO, Bio-soft N25-7, Stepan Company), 4% oleic acid, 0.5% Ethanol, 1.5% sodium xylene sulfonate, 7% 1,2-propanediol, 8% tri-Na-citrat, 0.5% triethanolamine, 1.5% boric acid, pH 8.5) and a powder detergent pH 10.0 of 1.715 g/L with 0.035 g/L AEO. The powder detergent had the following composition 17% LAS, 20% Sodium carbonate, 12% Sodium disilicate, 15% Zeolite A, 1% PCA, 34% Sodium sulphate. Both detergents were prepared with water hardness of 15°dH (Ca/Mg 4:1). All percentages stated are in w/w. After wash, rinse and drying the colour remission of the CS-28 strips was measured at 460 nm on TIDAS (Zeiss spectrophotometer).

**Results**

The multiple forms of BLA are shown on an IEF gel in Figure 1, where several bands occur over 1–1.5 pH units. The multiple forms were characterised by LC-MS/MS on twelve of the bands digested with trypsin and Asp-N to determine the modifications responsible for the forms varying in pI values with focus on the modifications deamidations and glycations.

Figure 1.

LC-MS/MS of the digested samples identified multiple deamidations, glycations, carbamylations and oxidations on Met. The observed deamidation and glycation sites are listed in Table 1. The identified glycations also included the different AGEs 3-deoxyglucosone, carboxymethyl, carboxyethyl and MG-HI. The modifications were mainly located on the surface of BLA though a few of the deamidations were located in the area around the active site. The focus of the project was on the deamidation and glycations and several of these were found in sites relevant for stability, activity and substrate binding. To determine the effect of some of the glycations and deamidations in specific sites suspected to be involved in BLAs activity and substrate binding nine BLA mutants was expressed with site specific mutations in single sites. The possible deamidation sites selected were changed to the corresponding acid to mimic the deamidation (Q→E, N→D), hereby mimicking the influence of the specific deamidation. For the selected glycation sites Lys was substituted with Met and Arg with Leu to evaluate the effect of removing the glucose reactive site.

The mutants were expressed in *Bacillus subtilis* along with a wild type BLA. After purification the enzymes were analysed by SDS-PAGE revealing single bands of approximately 55 kDa. The mutants and BLA were also analysed by IEF gel to examine the influence of the mutations on the pI values. This revealed that most of the mutations resulted in the same pI decrease except the two mutations N172D and K306M which did not result in a change in pI value compared to BLA, Figure 2.

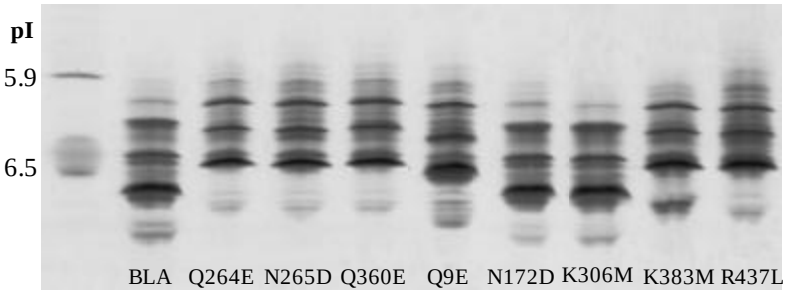


Figure 2

The mutations influence on the stability of BLA was evaluated by DSC, Table 2. The $T_{max}$ value for BLA was determined to 103.2 °C and the mutations in general did not influence the stability as $T_{max}$ only decreased slightly for the three mutants Q9E, N172D and Q264E decreasing by 2.6, 4.3 and 4 °C, respectively.

The kinetic parameters $k_{cat}$, $K_M$ and the catalytic efficiency ($k_{cat}/K_M$) for BLA and the mutants are listed in Table 3. The parameters were determined both on a short maltoheptaose (G7) substrate and on amylose at 30°C pH 7.0.

Alterations in the kinetic parameters on G7 substrates occur for the Q264E, Q360E and Q9E mutants. Q264E markedly decreased in $K_M$ which resulted in an increase in $k_{cat}/K_M$. Both $k_{cat}$ and $K_M$ increased for Q360 which meant that $k_{cat}/K_M$ overall was unaffected by the mutation. The Q9E mutation had a negative effect on $K_M$ as this value almost doubled which decreased $k_{cat}/K_M$. The $k_{cat}$ value was slightly higher for both N265D and R437L though the variation was quite high for the determination on R437L so this was not significantly different from BLA.

The kinetic parameters obtained on amylose revealed the same tendencies as the determinations on the G7 substrate. The changes were primarily observed for Q264E, Q360E and Q9E. For Q264E and Q360E the changes were the same as for the G7 substrate though the decrease in $K_M$ for Q264E was less pronounced. The changes in kinetic parameters were, however, different for Q9E. On amylose $k_{cat}$ decreased, whereas on the G7 substrate $K_M$ increased. The different changes observed on the different substrates for Q9E did, however result in a decrease in the catalytic efficiency for both substrates. Again R437L had a higher $k_{cat}$ value, with a high variation.

The wash performance of the mutants was determined in a liquid and a powder detergent as the components of the two detergents differ considerably and so does the pH during the wash with a pH of

approximately 8 in liquid detergent and 10 in powder. The washes were conducted at 40 °C as this is a standard wash temperature in Europe. The wash performance was determined on starch cotton swatches where the starch is coloured pink. The wash performance on the swatches was then determined spectrophotometrically by subjecting the cotton swatch to light and measuring the light reflecting from the swatch (the remission). The wash performance is determined in delta remission units which are the difference between the absorbance of the remission of a blank (a starch cotton swatch washed without enzyme) and the sample. The higher the delta remission the better the wash performance of the enzyme.

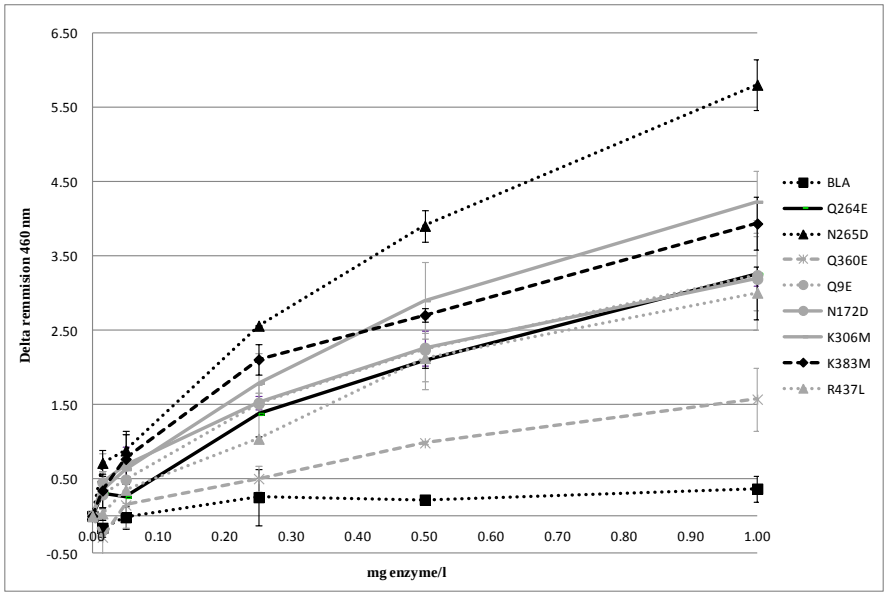


Figure 3

The wash performance for the BLA mutants in liquid detergent were all improved compared to BLA, Figure 3. Though, the wash performance for Q360E was approximately 1 delta remission higher than BLA, which is not enough to conclude an increased wash performance, the difference has to be at least 1.5 delta remission units. The N265D mutant stood out from the rest with the best wash performance. The remaining variants had a better wash performance than BLA and Q360E but not as good a performance as N265D. The fact that more or less all the variants except Q360E increased in wash performance at these conditions was unexpected.

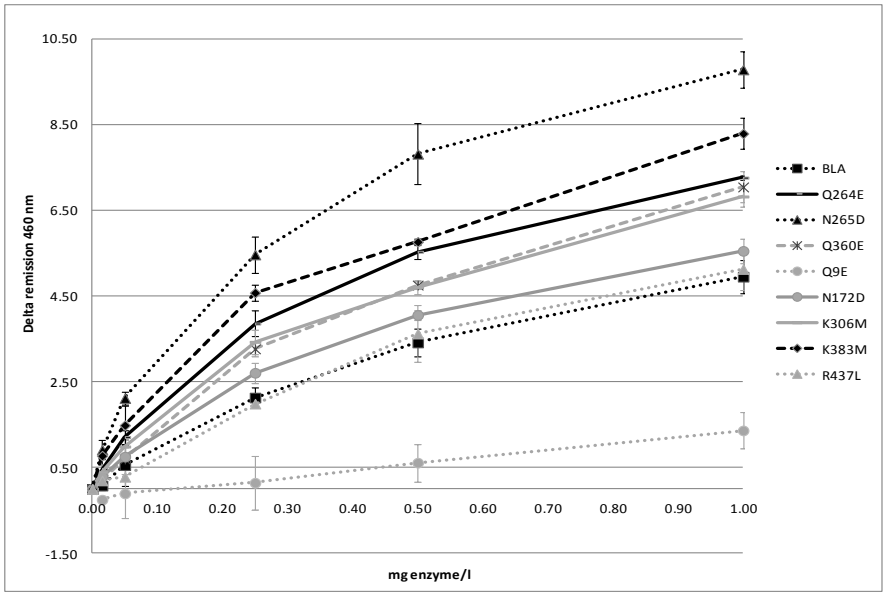


Figure 4

The wash performances for the mutants and BLA in powder detergent are illustrated in Figure 4. Again N265D had the highest wash performance. The wash performance of K306M, Q360E, Q264E and K383M was better than of BLA, but not as good as N265D. The variants N172D and R437L had the same wash performance as BLA at these conditions, whereas Q9E had a lower wash performance. In general there was no strong correlation between the kinetic parameters obtained on G7 and amylose with the wash performance data, thus the structural factors influencing the kinetic parameters in the assay are not necessarily the same as the ones affecting the wash performance.

## Discussion

### Identification of deamidation and glycation sites

Multiple deamidations of Asn and Gln residues as well as glycations of Lys and Arg residues were identified which all can be responsible for the creation of the multiple forms of varying pI values formed during the fermentation. The majority of modifications were observed on the surface of BLA, Figure 5, though a few deamidations were present inside the structure in the area around the active site. As mentioned previously several of the modification sites are suspected to be involved in the stability, activity and substrate binding.

In the fermentation of BLA sucrose was used as substrate but this carbohydrate is not a reducing sugar and therefore not able to form glycations directly. However, sucrose can be hydrolysed during sterilisation into fructose and glucose [28] and the glycations then stem from the sucrose hydrolysis products.

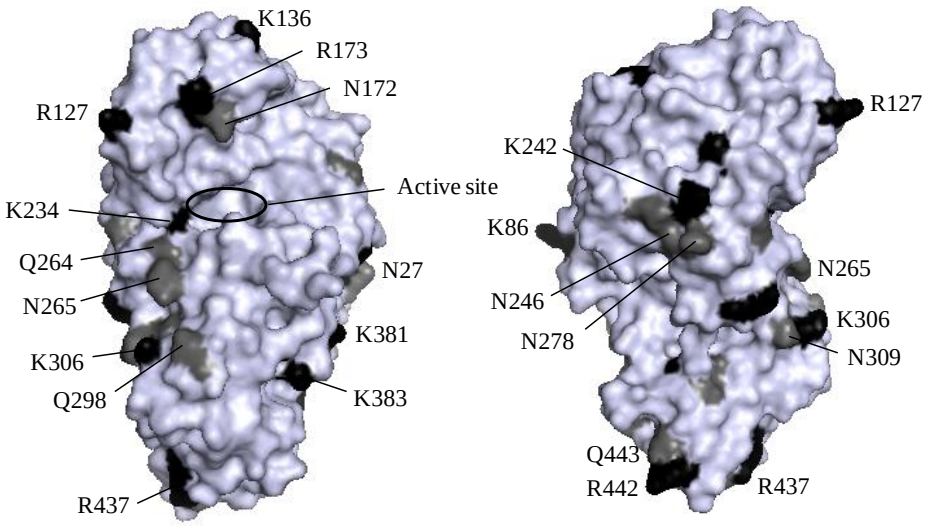



Figure 5

Both deamidations and modifications of Lys residues have in other studies shown to have a negative effect on the stability of BLA and modifications on Lys residues were also found to influence the activity and substrate specificity [29–33].

Several residues related to the stability of BLA were also found to be modified especially the modification sites R127, N172, R173, K237 and N272 are likely to be important for the enzymes thermostability as they are located in the interface between domain A and B containing the Ca-Na-Ca triad proposed to play a vital role in the thermostability of BLA [1, 7, 9, 10, 34]. The sites N172 and R173 are also located in a part of Domain B which also is speculated to contain part of an extended substrate binding site from the active site, however, the sites involved are not fully elucidated [1].

The deamidations on the residues Q9, N326 and Q360 are all located in the active site and are proposed to be involved in the activity and the substrate binding. The residues Q264 and K234 are proposed to contribute to two of the subsites in the substrate binding site [1, 5, 7]. As N265 is located right next to Q264 it is also likely that this site can influence the substrate binding. The results obtained for the deamidation on N265 is not very good as the score is quite low and the modified site is not mapped by the MS/MS results. Though the results for the deamidation in Q264 were quite good

and it is likely that N265 also become deamidated. The modifications in these sites located in the area of the active site and the substrate binding site will influence the activity and substrate binding affinity for BLA.

The glycation site K306 is located in what could be the elongation of the substrate binding site involved in binding of longer substrates. The glycations in this site most likely affect substrate binding for longer substrates such as starch.

Secondary carbohydrate binding sites have been found on *Bacillus halmapalus* α-amylase (BHA) [35]. The equivalence to these sites can be found on BLA. From the findings in BHA it can be deduced by sequence alignment and structural analysis that the residues W342, P432 and D430 in BLA may compose a maltose binding site and R437 and W467 a glucose binding sites.

The residue R437 composing one part of the glucose binding site becomes glycated and deamidation occur in N463 which is located right next to this binding site. The glycation site K383 is located so close to W342 that a glycation in the site could interfere with the maltose binding site. These modifications especially the glycations could decrease the efficiency of the secondary binding sites or completely inactivate the sites. As the exact role of these sites are not fully understood the effect of the modifications in these sites are not known. Though, if the sites function as initial substrate recognition sites or substrate binding for longer substrates, glycations at the secondary binding sites could decrease the efficiency of the amylase.

## Characterisation of BLA mutants

On the basis of the modifications and modification sites identified in the previous section BLA mutants were produced to determine the possible impact of some of the modifications on BLA. The modification sites chosen for mutagenesis were selected based on their possible impact on the active site, the substrate binding site or on secondary carbohydrate binding sites.

All the mutations result in substitutions to an amino acid with a lower pI value, therefore a decrease in pI value should occur for all the variants. This is, however, not the case. It appears that the same amino acid substitutions at different parts of the molecule can result in different changes in pI value or

no change at all. Several microenvironmental factors can influence the pKa values of the amino acid residues in proteins such as electrostatic and hydrophobic effects, hydrogen bond formation and helix dipole interactions [36, 37]. This can explain why the same amino acid residue substitution or chemical modification at different positions of the amylase has different effects on the pI value.

The Q264E mutation resulted in a great improvement on the catalytic efficiency compared to BLA and the other variants but in the wash performance the Q264E variant was quite average. The increased catalytic efficiency was caused by improved substrate binding. The crystal structure of a BLA variant (not from this project) with acarbose has been solved by Novozymes (structure is not published). In the crystal structure of the active site of BLA in complex with acarbose Q264 is oriented away from the substrate, Figure 6. In theory the carbonyl or amide groups on Gln should be able to make hydrogen bonds with hydroxyl groups on the substrate. The Gln side chain is very flexible in this structure so it is possible that hydrogen bonding does occur. The Q264E mutation increases the substrate binding. This strongly suggests that Glu is hydrogen bonding with hydroxyl groups on the substrate, whereas Gln does not. If Gln does hydrogen bond to the substrate, the bonds formed by Glu appear to be considerably stronger resulting in increased substrate binding. The improved substrate binding is accompanied by a positive effect on the wash performance at pH 8 and 10.

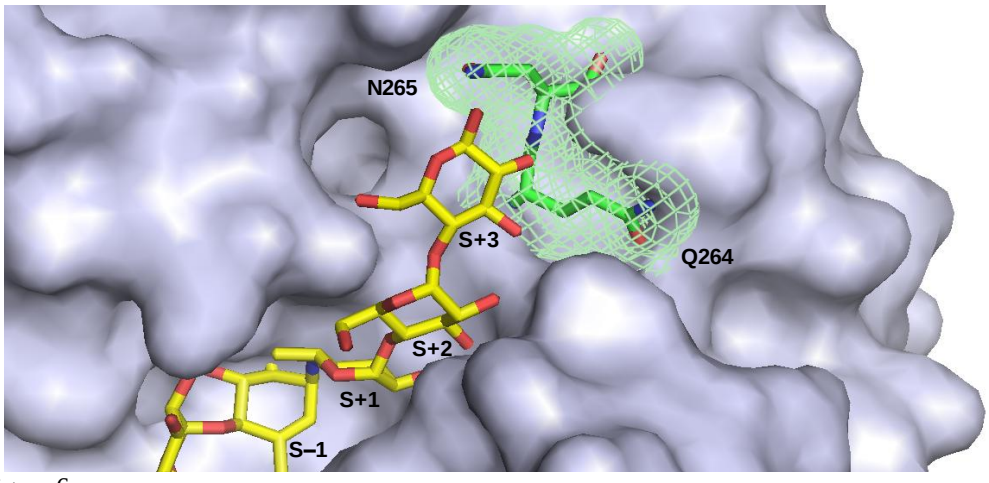


Figure 6

The kinetic parameters for the N265D variant did not show an effect of this mutation on the enzyme but it outperformed all the other variants in wash performance in both liquid and powder detergent. N265 is located right next to residue 264 in which Q264E has shown to interact with the substrate. The crystal structure of BLA in complex with acarbose, Figure 6, reveals that N265 is located close to the

substrate and it should be able to form hydrogen bonds with this. Since Q264E increased substrate binding it would be expected that the N265D mutation had a similar effect especially on a longer substrate as it looks like a longer substrate could bind with that side chain. This is, however, not confirmed by the $K_M$ values for the G7 substrate and amylose. The substantial increase in wash performance for N265D does show that the mutation has a positive effect on the performance. This indicates that it has improved substrate binding to the rice starch on the textile. Another possibility is that it is able to interact with the fabric hereby possibly improving the substrate amylase interaction. The kinetic parameters for Q9E revealed a lower catalytic efficiency on both substrates. Q9 is located right behind H327 which can interact with the substrate in subsite -1, Figure 7. This region is therefore very important for both substrate binding and catalytic activity. Q9 is most likely able to hydrogen bond with H327, though when the substrate binds in the active site H327 also hydrogen bond to one of the hydroxyl groups in the substrate. The Q9E mutation most likely results in salt bridge formation between the carboxy group on Q9E with the imidazole group in H327. This bond is stronger than the hydrogen bond so the interaction between H327 and the substrate will be weakened or even non-existing. This can explain the decreased substrate binding for the G7 substrate resulting in lower catalytic efficiency. The mutation does not have a negative effect on the substrate affinity for the longer substrate probably because this binding is stabilised by more binding sites on BLA than the short substrate. The mutation does, however, result in a decreased substrate turnover. The possible loss of hydrogen bonding between His327 and amylose could result in a different conformation of the substrate in the active site, hereby lowering the catalytic efficiency.

Despite the decrease in catalytic efficiency the wash performance of Q9E increased in liquid detergent. The wash performance in powder detergent was considerably lower than of BLA and all the other variants. This indicates that the site influence the pH profile of the amylase as the performance is good at pH 8 but decreases significantly at pH 10. This also correlates with other studies showing that pKa values of amino acid residues around the catalytic site can influence the pH activity profile of an enzyme [37]. The differences in wash performance could, however, also be due to some in-wash stability problems with the powder detergent possibly caused by one or several detergent components as for instance chelators can influence the stability of amylases [38]. Overall the different assays

reveal that the Q9 site has an influence on the substrate catalysis influencing substrate affinity and substrate turnover as well as possibly the pH activity profile.

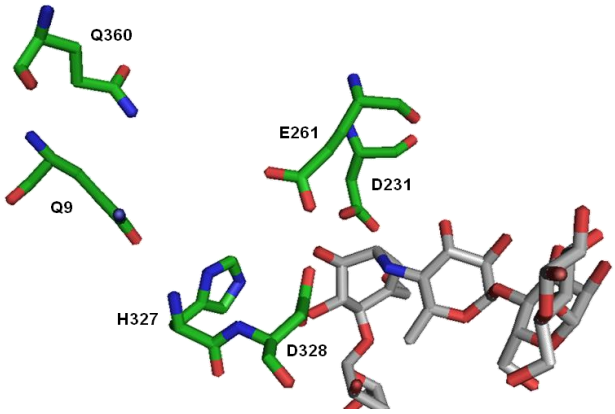



Figure 7

Q360 is located very close to Q9 in the area around the active site, Figure 7. Q360 is surrounded by several hydrophilic amino acid residues. It is therefore possible that this site is involved in the overall packing of the surrounding area as it could be loosening up the packing of this area. The Q360E mutation which is more polar at neutral pH could further loosen up the packing, thus affecting the area around the active site. The position of Q360 is approximately 5.9 Å from the active site residue E261. The Q360E mutation is therefore likely to influence the pKa of E261. The kinetic parameters revealed that the Q360E mutation improved the substrate turnover but lowered $k_{cat}$ thus not influencing the overall catalytic efficiency. This indicates that the Q360E mutation gives rise to some slightly larger conformational changes in the area that changes the general substrate binding and catalytic turnover. The Q360E mutation does not influence the wash performance in liquid detergent but the performance is slightly improved in powder detergent. This could indicate that the site influences the pH activity profile of the amylase as no change in performance occurs at pH 8 but performance increases at pH 10.

The N172D mutation did not have any effect on the kinetic parameters. The crystal structure of the active site in complex with acarbose, Figure 8, reveals that N172 is relatively close to the substrate binding, but approximately 5 Å from the substrate which is too far to form hydrogen bonds. Therefore, it has no effect on the substrate binding. The wash performance in powder detergent was the same as for BLA though the performance improved in the liquid detergent. The $T_{max}$ determinations showed that the mutation resulted in a small decrease in the stability of BLA. As previously mentioned it has

been hypothesised that for thermostabile enzymes decreasing the stability might increase activity at lower temperatures. It is also possible that the mutation influences the interaction with the fabric at pH 8 hereby improving wash performance. Though, it could also be other parameters responsible for the change such as the different components in the detergents.

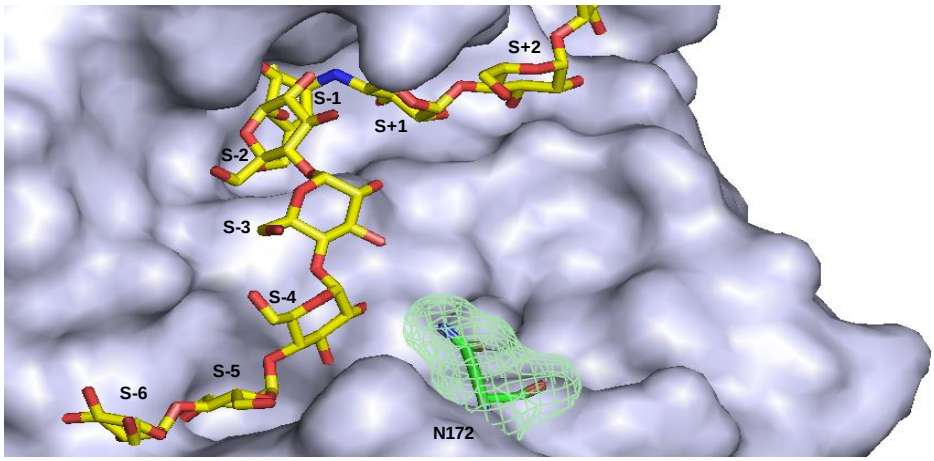



Figure 8

The position of K306 in relation to the substrate binding site is illustrated in Figure 9A. It appears that this site could be located in an area in which longer substrates bind. Though, the K306M mutation did not have any effect on the kinetic parameters thus indicating that K306 is not important for interaction with amylose. The mutation does, however, increase wash performance. The K → M mutation on the enzyme surface would be expected to have a lower binding affinity toward the substrate on the textile as the positively charged Lys can be involved in these interactions. The amylose used in the kinetic assay was gelatinised; the starch on the textile in the wash experiments might have a different structure which could explain the different results. The improved wash performance could be caused by improved interaction with the substrate on the fabric or by lover unproductive binding of the enzyme to other surfaces such as the textile.

K383 is located very close to the secondary binding site at W342, Figure 9B, so K383 could be involved in carbohydrate binding in this site. The mutation K383M did not result in any changes in the kinetic parameters for BLA. It did, however, result in improved wash performance. As described above the increased wash performance could be because the K383M mutation improves interaction of the enzyme with the substrate on the fabric. It is also possible that the mutation hinder the glycation in this site which could block the secondary binding site.

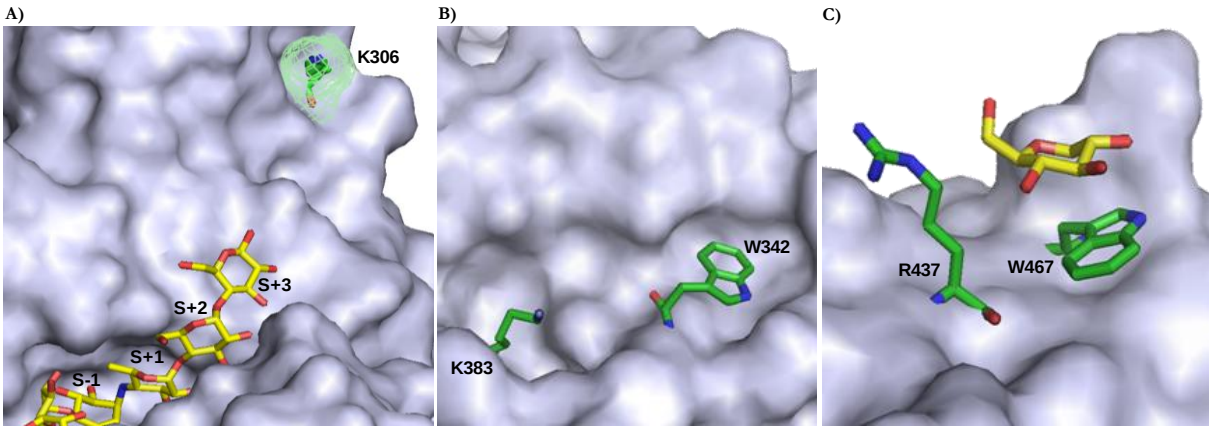


Figure 9

R437 is a part of a secondary binding site in domain C along with W467, Figure 9C. The interaction with glucose shown in the figure is an estimated modelling of the position of R437 after the position of W439 in BHA which structure was solved with glucose in this binding site [35]. From this structure it could appear that R437 is able to form hydrogen bonds to the carbohydrate. The R437L mutation may have a small positive effect on the $k_{cat}$ and hereby a slight improvement of the catalytic efficiency, but the variation of this result is very high. These results indicate that the site may be involved in substrate binding possibly as an initial substrate recognition site. The R437L mutation also has a positive effect on the wash performance in liquid detergent. The fact that the improvement only occur in one of the detergents could indicate that the improvement is due to some interaction or lack of interaction with a detergent ingredient which is present in one of the detergents.

The analysis of the variants shows that deamidations in the area close to the active site influence the kinetic parameters on different substrates and hereby the activity and substrate binding. Deamidations a bit further from the active site does not seem to have much effect on the kinetic parameters but they can influence factors such as stability. The Q→E and N→D variants also indicated that deamidations could affect wash performance generally positively though in the area close to the active site the modifications are also likely to influence the pH activity profile which can have a negative effect on wash performance.

The removal of possible reducing sugar reactive sites on the surface of BLA (K→M and R→L) did not seem to have any effect on the kinetic parameters though all the substitutions seemed to have a positive effect on the wash performance in liquid detergent as well as in powder detergent for the

K→M mutations. This may be the result of changes in the interaction of the amylase surface with the starch on the texile.

## Conclusion

Deamidations and glycations were found to contribute to the formation of the multiple forms of BLA varying in pI values created during the fermentation. These modifications were found in sites which are important to the enzymes stability, activity and substrate binding. BLA mutants were produced with site specific mutations in identified modification sites related to the active site, substrate binding and secondary binding sites. The mutants mimicking deamidation in the area around the active site influenced the activity and the substrate binding in different ways though they generally resulted in an improved wash performance for the single mutations especially for the BLA mutant N265D. Removing the glycation sites (K→M and R→L) in sites related to binding of longer substrates and secondary binding sites did not influence catalytic efficiency for the mutants K306M, K383M and R437L though it resulted in improved wash performance especially in liquid detergent.

**Acknowledgments**

The authors would like to thank the Danish Agency for Science Technology and Innovation for funding the research.

**Reference List**


1. Machius, M., Declerck, N., Huber, R. & Wiegand, G. (1998). *Structure*, 6, 281–292.

2. Vihinen, M. & Mantsala, P. (1989). *Critical Reviews in Biochemistry And Molecular Biology*, 24, 329–418.

3. de Souza, P. M. & Magalhaes, P. D. E. (2010). *Brazilian Journal of Microbiology*, 41, 850–861.

4. Kirk, O., Borchert, T. V. & Fuglsang, C. C. (2002). *Current Opinion in Biotechnology*, 13, 345–351.

5. Nielsen, J. E. & Borchert, T. V. (2000). *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1543, 253–274.

6. Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N. & Udaka, S. (1985). *Journal of Biochemistry*, 98,1147–1156.

7. Machius, M., Wiegand, G. & Huber, R. (1995). *Journal of Molecular Biology*, 246, 545–559.

8. Kandra, L., Gyémánt, G., Remenyik, J., Hovánszki, G. & Lipták, A. (2002). *FEBS Letters*, 518, 79–82.

9. Declerck, N., Machius, M., Wiegand, G., Huber, R. & Gaillardin, C. (2000). *Journal of Molecular Biology*, 301, 1041–1057.

10. Declerck, N., Machius, M., Joyet, P., Wiegand, G., Huber, R. & Gaillardin, C. (2003). *Protein Engineering (Oxford)*, 16, 287–294.

11. Cuyvers, S., Dornez, E., Delcour, J. A. & Courtin, C. M. (2012). *Critical Reviews in Biotechnology*, 32, 93–107.

12. Nielsen, M. M., Seo, E. S., Dilokpimol, A., Andersen, J., Hachem, M. A., Naested, H., Willemoes, M., Bozonnet, S., Kandra, L., Gyemant, G., Haser, R., Aghajari, N. & Svensson, B. (2008). *Biocataalysis and Biotransformation*, 26, 59–67.

13. Jenkins, N., Murphy, L. & Tyther, R. (2008). *Molecular Biotechnology*, 39, 113–118.

14. Manning, M., Chou, D., Murphy, B., Payne, R. & Katayama, D. (2010). *Pharmaceutical Research*, 27, 544–575.

15. Baynes, J. W., Watkins, N. G., Fisher, C. I., Hull, C. J., Patrick, J. S., Ahmed, M. U., Dunn, J. A. & Thorpe, S. R. (1989). *Progress in Clinical and Biological Research*, 304, 43–67.

16. Brownlee, M., Cerami, A. & Vlassara, H. (1988). *The New England Journal of Medicine*, 318, 1315–1321.

17. Ames, J. M. (2008). *Biochemical Society Transactions*, 36, 1051–154.

18. Barnaby, O. S., Cerny, R. L., Clarke, W. & Hage, D. S. (2011). *Clinica Chimica Acta*, 412, 277–285.

19. Kislinger, T., Humeny, A. & Pischetsrieder, M. (2004). *Current Medicinal Chemistry*, 11, 2185–2193.

20. Kislinger, T., Humeny, A., Peich, C. C., Becker, C. M. & Pischetsrieder, M. (2005). *Annals of the New York Academy of Sciences*, 1043, 249–259.

21. Gaza-Bulseco, G., Li, B. Q., Bulseco, A. & Liu, H. C. (2008). *Analytical Chemistry*, 80, 9491–9498.

22. Li, X. J., Cournoyer, J. J., Lin, C. & O'Cormora, P. B. (2008). *Journal of the American Society for Mass Spectrometry*, 19, 855–864.

23. Deng, W. P. & Nickoloff, J. A. (1992). *Analytical Biochemistry*, 200, 81–88.

24. Barkholt, V. & Jensen, A. L. (1989). *Analytical Biochemistry*, 177, 318–322.

25. van Wandelen, C. & Cohen, S. A. (1997). *Journal of Chromatography A*, 763, 11–22.

26. Bozonnet, S., Jensen, M. T., Nielsen, M. M., Aghajari, N., Jensen, M. H., Kramhøft, B., Willemoës, M., Tranier, S., Haser, R. & Svensson, B. (2007). *FEBS Journal*, 274, 5055–5067.

27. Mori, H., Bak-Jensen, K. S., Gottschalk, T. E., Motawia, M. S., Damager, I., Møller, B. L. & Svensson, B. (2001). *European Journal of Biochemistry*, 268, 6545–6558.

28. Petersen, K. K., Hansen, J. & Krogstrup, P. (1999). *Plant Cell Tissue and Organ Culture*, 58, 189–197.

29. Habibi, A. E., Khajeh, K. & Nemat-Gorgani, M. (2004). *Journal of Biochemistry and Molecular Biology*, 37, 642-647.

30. Khajeh, K., Naderi-Manesh, H., Ranjbar, B., Moosavi-Movahedi, A. A. & Nemat-Gorgani, M. (2001). *Enzyme and Microbial Technology*, 28, 543–549.

31. Khajeh, K., Ranjbar, B., Naderi-Manesh, H., Habibi, A. E. & Nemat-Gorgani, M. (2001). *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1548, 229–237.

32. Tomazic, S. J. & Klibanov, A. M. (1988). *The Journal of Biological Chemistry*, 263, 3092–3096.

33. Tomazic, S. J. & Klibanov, A. M. (1988). *The Journal of Biological Chemistry*, 263, 3086–3091.

34. Machius, M., Declerck, N., Huber, R. & Wiegand, G. (2003). *Journal of Biological Chemistry*, 278, 11546–11553.

35. Lyhne-Iversen, L., Hobley, T. J., Kaasgaard, S. G., & Harris, P. Structure of *Bacillus halmapalus* alpha-amylase crystallized with and without the substrate analogue acarbose and maltose (2006). *Acta Crystallographica: Section F*, 62, 849–854.

36. Li, H., Robertson, A. D. & Jensen, J. H. (2005). *Proteins*, 61, 704–721.

37. Nielsen, J. E., Borchert, T. V. & Vriend, G. (2001). *Protein Engineering*, 14, 505–512.

38. Lund, H., Kaasgaard, S., Skagerlind, P., Jorgensen, L., Jørgensen, C. & Weert, M. (2012). *Journal of Surfactants & Detergents*, 15, 265–276.

## Figure Captions

Figure 1:
IEF of BLA where 12 bands over the entire pI span were selected for in-gel digest with trypsin and Asp-N followed by LC-MS/MS to identify the modifications responsible for the multiple forms.

Figure 2:
IEF gel of the 8 BLA variants and wt BLA expressed in *Bacillus subtilis*.

Figure 3:
The wash performance of the BLA variants on rice starch cotton patches in liquid detergent. Standard deviation for the determinations is shown.

Figure 4:
The wash performance of the BLA variants on rice starch cotton patches in powder detergent. Standard deviation for the determinations is shown.

Figure 5:
The surface of BLA showing the Asn and Gln found to become deamidated during the fermentation in grey and the glycated Lys and Arg residues are shown in black. PyMOL was used for molecular rendering (The PyMOL molecular Graphics System Vers. 1.3, Schrödinger, LCC)(PDB 1BLI).

Figure 6:
The active site of BLA in complex with acarbose showing the possible substrate binding residues of Q264 and N265 (Unpublished structure of BLA). The subsites +1, +2, +3 and -1 in the substrate binding site are shown as S+1, S+2, S+3 and S-1. (The PyMOL molecular Graphics System Vers. 1.3, Schrödinger, LCC).

Figure 7:
Close-up of acarbose in the active site showing the three active site residues D231, E261 and D328 along with the residues Q9, Q360 and His327 (Unpublished structure of BLA) (The PyMOL molecular Graphics System Vers. 1.3, Schrödinger, LCC).

Figure 8:
BLA in complex with acarbose in the active site showing the residue N172 (Unpublished structure of BLA). The subsites in the binding site are indicated by S and the site number. (The PyMOL molecular Graphics System Vers. 1.3, Schrödinger, LCC).

Figure 9:
Different views of the crystal structure of BLA (Unpublished structure of BLA). A) The structure of acarbose in the active site showing K306 in what could be an elongation of the substrate binding. The subsites in the binding site are indicated by S and the site number. B) K383 and the secondary binding site W342. C) R437 as part of a secondary binding site with W467 with glucose bound. The position of R437 is estimated by modelling after the position of W439 in BHA solved with glucose in the binding site.

Table 1:
Modifications observed in the 12 digested BLA bands analysed by LC-MS/MS fragmenting with HCD and ETD. Hex: Glycation with hexose, 3DG: 3-Deoxyglucosone, Cbm: Carboxymethyl, Cbe: Carboxyethyl, MG-HI: methylglyoxal-derived hydroimidazolone, Dea: Deamidation, Covered is whether or not the modification site is directly mapped by the MS/MS results.

| Site | Modification | Score | Expect | Covered |
|---|---|---|---|---|
| Q9 | Dea | 61 | 7.8E-7 | + |
| R23 | Cbm/Cbe | 32 | 0.0031 | + |
| R24 | MGHI | 15 | 0.041 | + |
| N27 | Dea | 36 | 0.0016 | + |
| Q51 | Dea | 46 | 2.3E-5 | - |
| K88 | Cbe | 28 | 0.0091 | + |
| R127 | MGHI | 38 | 0.00015 | + |
| K136 | Cbm | 23 | 0.025 | + |
| K154 | Cbe | 43 | 0.0012 | + |
| N172 | Dea | 39 | 0.021 | + |
| R173 | 3DG | 34 | 0.017 | + |
| K234 | Cbm | 53 | 0.00023 | + |
| K237 | Cbm/Cbe | 39 | 0.0011 | + |
| R242 | Cbe/MGHI | 71 | 2.5E-6 | + |
| N246 | Dea | 37 | 0.07 | + |
| Q264 | Dea | 52 | 5.8E-6 | + |
| N265 | Dea | 39 | 0.00038 | - |
| N272 | Dea | 26 | 0.0026 | - |
| N278 | Dea | 78 | 7.8E-5 | + |
| N280 | Dea | 43 | 0.028 | + |
| Q291 | Dea | 72 | 0.00034 | + |
| Q298 | Dea | 78 | 6.8E-7 | + |
| K306 | Cbm | 57 | 0.00033 | + |
| N309 | Dea | 42 | 6.5E-5 | + |
| K315 | Hex/Cbm | 50 | 0.0009 | + |
| N326 | Dea | 92 | 7.8E-8 | + |
| Q360 | Dea | 41 | 8.4E-5 | + |
| K381 | Cbm | 29 | 0.0011 | + |
| K383 | Cbe | 20 | 0.011 | + |
| K389 | Cbm/Cbe | 49 | 0.00057 | + |
| K392 | Cbe | 79 | 8.7E-5 | + |
| Q393 | Dea | 79 | 8.7E-5 | + |
| Q399 | Dea | 77 | 7.3E-5 | + |
| R413 | Cbe | 79 | 0.00012 | + |
| N421 | Dea | 67 | 0.0006 | + |
| R437 | MGH1 | 30 | 0.001 | + |
| R442 | MGH1/Cbe/Hex | 54 | 0.00014 | + |
| Q443 | Dea | 39 | 0.00014 | - |
| N444 | Dea | 39 | 0.00014 | + |
| N463 | Dea | 49 | 1.2E-5 | - |

Table 2:
$T_{max}$ values for the BLA variants determined by DSC.

| | BLA | Q9E | N172D | Q264E | N265D | Q360E | K306M | K383M | R437L |
|---|---|---|---|---|---|---|---|---|---|
| $T_{max}$ (°C) | 103.2 | 100.6 | 98.9 | 99.2 | 102.9 | 102.7 | 103.4 | 103.2 | 103.1 |

Table 3:
The kinetic parameters $k_{cat}$, $K_M$ and $k_{cat}/K_M$ for the BLA mutants and BLA determined at 30°C pH 7.0 on maltoheptaose (G7)-pNP substrate and amylose.

| | G7-pNP | | | Amylose | | |
|---|---|---|---|---|---|---|
| Sample | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ ($s^{-1}$ $mM^{-1}$) | $k_{cat}$ ($s^{-1}$) | $K_M$ (mg/ml) | $k_{cat}/K_M$ ($s^{-1}$ $mg^{-1}$ ml) |
| BLA | 65 ± 12 | 0.011 ± 0.003 | 5922 | 200 ± 8 | 0.140 ± 0.03 | 1455 |
| Q9E | 63 ± 14 | 0.027 ± 0.008 | 2394 | 117 ± 12 | 0.112 ± 0.04 | 1099 |
| N172D | 70 ± 9 | 0.013 ± 0.005 | 5462 | 201 ± 6 | 0.129 ± 0.01 | 1555 |
| Q264E | 60 ± 5 | 0.005 ± 0.001 | 13013 | 228 ± 15 | 0.089 ± 0.02 | 2632 |
| N265D | 83 ± 5 | 0.011 ± 0.002 | 7277 | 232 ± 28 | 0.128 ± 0.03 | 1852 |
| Q360E | 155 ± 24 | 0.026 ± 0.005 | 5996 | 416 ± 17 | 0.261 ± 0.06 | 1630 |
| K306M | 70 ± 4 | 0.010 ± 0.001 | 7335 | 213 ± 38 | 0.125 ± 0.04 | 1740 |
| K383M | 69 ± 10 | 0.010 ± 0.002 | 7133 | 241 ± 33 | 0.139 ± 0.02 | 1743 |
| R437L | 81 ± 18 | 0.010 ± 0.003 | 8153 | 307 ± 67 | 0.137 ± 0.02 | 2249 |

## 7.3 Appendix III – Manuscript 2

Manuscript in preparation for Journal of The American Society for Mass Spectrometry:

# Identification and Quantification of Non-Enzymatically Glycated Peptides by Electron Transfer Dissociation with Isobaric Tags

Connie Nielsen,[1,2] Svend G. Kaasgaard,[1] Susanne Jacobsen,[2] Birte Svensson,[2] Carsten P. Sönksen[1]

[1]Novozymes A/S, Bagsvaerd, Denmark
[2]Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark

**Abstract**

Non-enzymatic glycation of proteins can occur during production and storage of recombinant proteins. The modification can result in stability and activity loss. The glycation can undergo a series of rearrangement reactions to form a group of more reactive carbonyl compounds termed advanced glycation end products (AGEs). Mass spectrometry (MS) based methods for identification and quantification of the glycations could be a useful tool in the characterisation of the relative level of glycation during the various production stages and the storage of proteins. Specific glycation sites on peptides can be determined by liquid chromatography (LC) - tandem mass spectrometry (MS/MS), however, fragmentation techniques such as collision induced dissociation (CID) results in extensive patterns of neutral loss and loss of labile modifications such as glycations. Electron transfer dissociation (ETD) leaves labile modifications intact allowing for direct mapping of peptide modifications such as glycations. This report asses the utility of isobaric stable isotope tagging and ETD for relative quantification of glycations and the following AGEs. ETD is compared to higher energy collision dissociation (HCD) for the identification and quantification of glycated peptides. The ETD fragmentation allowed for identification of a considerably higher number of glycated peptides compared to HCD. Quantification of glycated peptides was performed with ETD though the method is not as efficient as HCD in the formation of reporter ions suitable for quantification. It was also observed that in some cases the glycated peptides could influence the quantification results for both of the fragmentation techniques.

**Introduction**

The production of enzymes for technical application in a wide range of different industries has been growing for the last couple of decades [1]. However, during production and storage proteins can be modified by non-enzymatic glycation which can result in loss of stability and activity [2–4]. A mass spectrometry (MS) based method for identification and relative quantification of the glycations could be a useful tool in the characterisation of the glycation levels during the various production stages and storage of enzymes.

The labile modification glycation is a non-enzymatic reaction between primary amines on proteins and the carbonyl group of a reducing sugar. Since the reaction occurs at amino groups, the N-terminal ends of proteins along with lysine and argenine residues are prone to the reaction. The glycation forms a ketoamine also known as the Amadori compound which can undergo a series of rearrangement reactions to forms a group of more reactive carbonyl compounds termed advanced glycation end products (AGEs) [5, 6]. These AGEs include 3-deoxyglucosone, carboxymethyl, carboxyethyl and Methylglyoxal-derived hydroimidazolone (MG-HI) as some of the most described [7–9].

Specific glycation sites on peptides can be determined by liquid chromatography (LC) - tandem mass spectrometry (MS/MS), however, some of the most common fragmentation techniques results in extensive patterns of neutral loss and loss of labile modifications such as glycations [10-13]. Electron transfer dissociation (ETD) is a soft fragmentation method which is very suitable for characterisation of peptides containing post-translational modifications (PTM) as it leaves labile modifications intact allowing for direct mapping of peptide modifications [10, 12, 14–16]. ETD does, however, also have some limitations as it is most effective on triply or above charged ions and on ions below 850 m/z [17, 18].

Quantification has become an important tool in MS based proteomics which has resulted in numerous different methods for either absolute or relative quantification [19–21]. However, only a few methods exists for relative quantification of glycations by LC-MS and these are generally based on addition of internal standards of the corresponding isotopically labelled glycated peptides [22, 23].

Isobaric stable isotope tagging allows for relative quantifications of proteins or peptides between two or more samples (up to 8) by use of MS/MS. The isobaric tags are made from light and heavy isotopes which mean that all the tags have the same total mass as well as the same physiochemical properties. The tags consist of a reporter group, a balance group and an amino reactive end which react with the amino groups of lysine and the N-terminus of peptides. Each sample is digested and labelled where after all samples are mixed and the differentially labelled peptides are selected for MS/MS as a single precursor ion. Upon fragmentation the specific reporter ions for each sample are released allowing for relative quantification from the reporter ion ratios. The MS/MS spectrum of the peptide is not altered by the tags and can therefore also be used to identify the peptide sequence. One of the advantages of

this method is the low level of noise generally obtained in MS/MS [24–27]. It has previously been shown that ETD can be used for relative quantification with isobaric tags. However, the reporter ions produced by ETD differ from those produced by CID due to the different fragmentation mechanism, producing c- and z-ions. This has resulted in some of the isobaric tags producing the same reporter ions so for instance the TMT 6-plex tags could previously only quantify up to four samples. However, very recently the position of some of the heavy isotopes have been changed for the TMT 6-plex tags so each tag now generates a unique reporter ions in ETD fragmentation (114, 115, 116, 117, 118 and119 from the 126-131 TMT tags, respectively) [28–31].

In this report we examine the possibility of identifying and performing relative quantification of the labile modifications glycation and AGE's with TMT tags fragmenting with ETD. The results obtained by the identification and quantification of glycations and AGEs by ETD are compared to results obtained by higher energy collision dissociation (HCD). For these experiments Barley α-amylase/subtilisin inhibitor (BASI) is used as model protein as it is a smaller protein (20 kDa) consisting of 7 Lys and 12 Arg residues [32, 33].

## Materials and methods

*Sample preparation*

Barley α-amylase/subtilisin inhibitor (BASI) was provided by Novozymes (Bagsvaerd, Denmark). The TMTsixplex isobaric mass tagging kit was purchased from Thermo Scientific (Rockford, IL). BASI (2 mg/ml) incubated 4 days (BASIG4) and 8 days (BASIG8) in 5% glucose 25 mM $NaHCO_3$ 0.1 mM $CaCl_2$ 0.01% Triton-x 100 pH 9.5 at 40°C. Glycated BASI was digested, reduced, alkylated and labelled with TMT tags according to manufacturer protocol. Two TMT labelled samples were prepared mixing 1:1 of BASIG4 and BASIG8 so the tags used were 126:128 and 127:130 (BASIG4:BASIG8). Labelled samples were purified on Solid phase extraction cartridge (C18 octadecyl standard density) purchased from 3M (St. Paul, MN) according to manufactures protocol. The samples were dried and diluted in 5% formic acid.

*Liquid chromatography and mass spectrometry*

All experiments were performed on a LTQ Orbitrap XL mass spectrometer (Thermo Scientific) with nano-ESI source (Nanospray Flexion Source, Thermo Scientific) couple to an EASY-nLC II (Thermo Scientific). For LC-MS/MS the peptides were separated on Thermo Scientific EASY-Column (C18) 2 cm x 100 µm i.d. column. The elution was performed from 5–50% acetonitril in 0.1% formic acid over 60 min at a flow rate of 250 nl/min. $MS^1$ was performed in the orbitrap mass analyser with 60,000 resolution, automatic gain control (AGC) target was 500,000 with a maximum ion injection time of 300 ms. AGC target for MS/MS was 40,000 for the ion trap and 100,000 for the orbitrap. Normalized collision energy of 55% were used for HCD and fragment ions were detected in the orbitrap with 7500 resolution with a maximum injection time of 300 ms using 2 microscans selecting precursor ions in the range 300–2000 m/z. ETD reactions were performed for 85 ms detecting fragment ions in the ion trap with a maximum injection time of 100 ms using 1 microscan selecting precursor ions in the range 350–1200 m/z. For both HCD and ETD the four most abundant precursors rejecting charge state 1 for HCD and charge state 1 and 2 for ETD.

*Data analysis*

The data analysis was performed using Mascot Distiller (Matrix Science, UK), Mascot Daemon (Matrix Science, UK) and the Mascot search engine (Matrix Science, UK) for spectra processing, search automation, protein/peptide identification and quantification. The searches were conducted against a custom made database containing 14100 sequences (mainly enzyme sequences from different organisms) searching against tryptic peptides allowing 5 missed cleavages with 10 ppm precursor mass tolerance, 0.5 Da fragments mass tolerance for ETD and 0.02 Da mass tolerance for HCD with decoy database enabled. Static modification TMT6plex of N-terminus and carbamidomethyl (C) was set and the following variable modifications were chosen: Deamidation (NQ), Hex (KR), 3-deoxyglucosone (KR), carboxymethyl (KR), carboxyethyl (KR), MG-HI (KR), Oxidation (M). For the TMT quantification normalization was not used, auto outlier removal was chosen and peptide threshold was set to above homology level. Each sample was determined by 3 runs. The quantification results were assessed by a 3-factor full factorial ANOVA model where outlier detection was used.

## Results and discussion

*Identification of glycation sites*

BASI is a smaller protein (20 kDa) consisting of 181 amino acids with 7 Lys and 12 Arg residues. This makes it a good model protein for the study of glycations as it relatively small but there are still multiple possible glycation sites. Therefore digest of the glycated protein provides multiple glycated peptides.

BASI was glycated by incubation with glucose for four (BASIG4) and eight days (BASIG8). The two samples were digested and labelled with TMT; 126 and 127 for BASIG4 and TMT 128 and 130 for BASIG8. The labelled samples were mixed 1:1 (126:128 and 127:130) and analysed by ETD and HCD.

**Table 1.** Glycated peptides from BASI identified (ID) and quantified by ETD and HCD. Covered is the mapping of the glycated residue in the obtained peptide sequence coverage. The glycated residue is shown in **bold**. Hex: glucose, CBM: carboxymethyl, CBE: carboxyethyl, MG-H1:methylglyoxal-derived hydroimidazolone.

| Amino acid | Peptide | Glycation | ID | Covered | Quantified |
|---|---|---|---|---|---|
| 1-27 | ADPPPVHDTDGHEL**R**ADANYYVLSANR | CBE | ETD | ETD | |
| 16-41 | ADANYYVLSAN**R**AHGGGLTMAPGHGR | MG-H1 | ETD/HCD | ETD | ETD/HCD |
| 16-41 | ADANYYVLSAN**R**AHGGGLTMAPGHGR | CBM | ETD/HCD | ETD/HCD | HCD |
| 16-41 | ADANYYVLSAN**R**AHGGGLTMAPGHGR | CBE | ETD/HCD | ETD | HCD |
| 28-61 | AHGGGLTMAPGHG**R**HCPLFVSQDPNGQHDGFPVR | MG-H1 | ETD | ETD | |
| 42-72 | HCPLFVSQDPNGQHDGFPV**R**ITPYGVAPSDK | MG-H1 | ETD | ETD | |
| 62-75 | ITPYGVAPSD**K**IIR | Hex | ETD/HCD | ETD/HCD | HCD |
| 62-75 | ITPYGVAPSD**K**IIR | CBM | ETD/HCD | ETD/HCD | HCD |
| 76-85 | LSTDV**R**ISFR | MG-H1 | HCD | HCD | HCD |
| 76-85 | LSTDV**R**ISFR | CBM | ETD/HCD | ETD | HCD |
| 76-85 | LSTDV**R**ISFR | CBE | ETD/HCD | ETD | HCD |
| 86-107 | AYTTCLQSTEWHIDSELAAG**R**R | CBE | ETD/HCD | ETD | HCD |
| 107-122 | **R**HVITGPVKDPSPSGR | CBM | ETD | ETD | |
| 107-122 | **R**HVITGPVKDPSPSGR | MG-H1 | ETD | ETD | |
| 108-122 | HVITGPV**K**DPSPSGR | Hex | ETD/HCD | ETD | ETD/HCD |
| 108-122 | HVITGPV**K**DPSPSGR | CBM | ETD/HCD | ETD/HCD | ETD/HCD |
| 123-130 | ENAF**R**IEK | MG-H1 | ETD/HCD | ETD/HCD | HCD |
| 128-140 | IE**K**YSGAEVHEYK | Hex | ETD/HCD | ETD | ETD/HCD |
| 128-140 | IE**K**YSGAEVHEYK | CBM | ETD/HCD | ETD/HCD | HCD |
| 131-155 | YSGAEVHEY**K**LMSCGDWCQDLGVFR | Hex | ETD/HCD | ETD | ETD/HCD |
| 141-158 | LMSCGDWCQDLGVF**R**DLK | MG-H1 | ETD/HCD | ETD/HCD | HCD |
| 141-158 | LMSCGDWCQDLGVF**R**DLK | CBM | ETD/HCD | ETD | HCD |
| 141-158 | LMSCGDWCQDLGVF**R**DLK | CBE | ETD/HCD | ETD | HCD |
| 156-176 | DL**K**GGAWFLGATEPYHVVVFK | Hex | ETD/HCD | ETD | ETD/HCD |
| 156-176 | DL**K**GGAWFLGATEPYHVVVFK | CBM | ETD/HCD | ETD/HCD | ETD/HCD |
| 159-177 | GGAWFLGATEPYHVVVF**K**K | Hex | ETD/HCD | ETD | ETD/HCD |
| 159-177 | GGAWFLGATEPYHVVVF**K**K | CBM | ETD/HCD | ETD/HCD | ETD/HCD |
| 159-181 | GGAWFLGATEPYHVVVFKKAPPA | MG-H1 | ETD/HCD | | HCD |

The analysis of the samples resulted in identification of multiple glycated Lys and Arg residues, Table 1, where several of the glycations were the advanced glycation end-products (AGEs): carboxymethyl,

carboxyethyl and methylglyoxal-derived hydroimidazolone (MG-HI). As expected ETD identified more glycation sites than HCD. ETD fragmentation also provided better sequence coverage of the glycated peptides. This is because ETD does not cleave of the modification during the fragmentation so the peptide fragments carrying the glycation can be identified by the Mascot search engine hereby giving higher peptide sequence coverage. Many of the glycated peptides were also identified with HCD though for most of these the glycated residues were not mapped. Though for some of the glycated peptides peaks corresponding to glycated peptide fragments were found in the HCD spectra. When the sequence coverage and the Lys and Arg residues are not covered the Mascot search engine assigns the glycation to the Arg or Lys residue in the C-terminal instead of the missed cleavage site. Trypsin does not cleave at a glycation site so the glycation should be at the missed cleavage site. This inability to assign the correct glycation site for glycated peptides with low sequence coverage were also observed in other experiments with SEQUEST as search engine (data not shown). The peptide G159 – A181was found to contain the MG-HI glycation but the sequence coverage was not very good with either of the fragmentation methods so it is not possible to determine which of the two Lys residues are modified (K176 or K177), though Mascot assigns the modification to K177.

These data supports that ETD is an efficient fragmentation method for sequencing peptides with labile modifications such as glycations compared to methods such as HCD. Though HCD can be used for the identification of glycations however, the identification rate is lower as well as the sequence coverage of the glycated peptides and hereby identification of the specific glycation site.

*Relative quantification of glycation sites*

The two samples BASIG4 and BASIG8 was digested, labelled with TMT and mixed as described in the previous section. The intensities of the reporter ions are important for the quantification as reporter ions with low intensity exhibit poor quantifications providing high error rates. The reporter ions in ETD has to exceed 1,000 and for HCD 100,000 to achieve average errors below 10% [30]. Therefore, quantifications were only used for spectra with reporter ions above these intensities.

The glycated peptides are quantified by ETD and HCD, Table 1. Quantification results are obtained by HCD for all the glycated peptides identified by this method whereas only around a third of the

glycated identified by ETD are quantified by this method. The lower number of peptide quantifications by ETD have also been observed and the reason for this is explained by others [30, 31]. The ETD fragmentation produced lower reporter ion intensities and some of the ETD spectra for the identified glycated peptides did not contain any reporter ions. Viner and co-workers [31] speculate that the lower intensities for the reporter ions are due to ETD cleaving randomly along the peptide backbone unlike HCD and therefore produces lower yield of reporter ions.

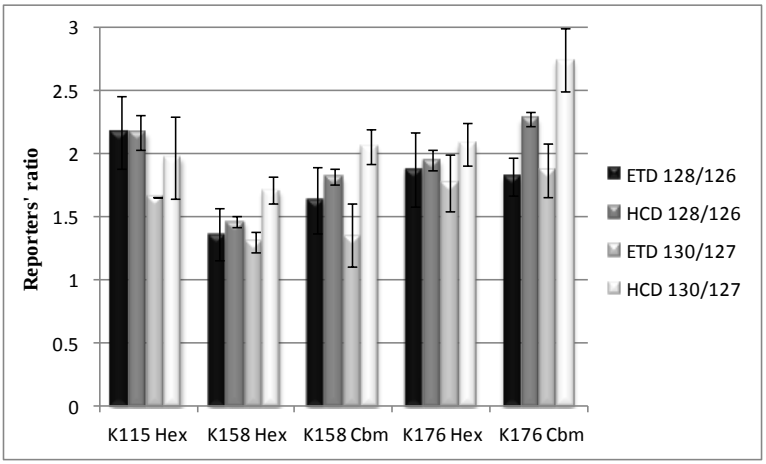



**Figure 1.** Relative quantification of glycated peptides between BASI incubated four days (BASIG4) and eight days (BASIG8) in glucose. The ratio is are for BASIG8/BASIG4 mixed 1:1 after labelling with the reporter ions 128 and 130 for BASIG8, 126 and 127 for BASIG4. The relative quantification was determined with LC-MS/MS fragmenting with ETD and HCD.

The significance of the quantification results were evaluated by statistical analysis using a 3-factor full factorial ANOVA model on the peptides with the following glycations: K115, K158, K176 glycated with glucose as well as K158 and K176 glycated with carboxymethyl. These glycated peptides were chosen for the statistical analysis as they had more than one quantification determination for each of the label combinations (130/127 and 128/126) with both fragmentation methods. Outlier removal was used to remove six determinations from the dataset.

The quantification results for the five glycated peptides obtained for the different label combinations with the two fragmentation methods are shown in Figure 1 and the overall quantification results obtained by ETD and HCD for the five peptides are listed in Figure 2. The figures reveal that the quantification results obtained by HCD generally are higher than the ones obtained by ETD.

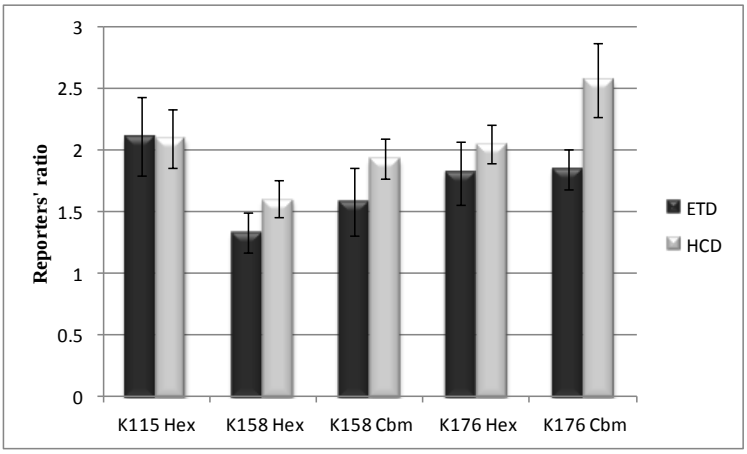



**Figure 2.** Comparison of relative quantification results for glycated peptides from BASI obtained by LC-MS/MS fragmenting with HCD and ETD. The ratios are for BASIG8/BASIG4.

The statistical analysis revealed that the performance of the fragmentation methods depends on the glycated peptide analysed. HCD and ETD perform similarly on the peptides containing K115 and K176 glycated with glucose but the methods are giving significantly different results on the peptides containing K158 glycated with glucose, K158 and K176 glycated with carboxymethyl. The analysis also agreed with the observations from Figure 1 and 2 that the quantification ratios obtained by HCD are higher than the ones obtained by ETD. The performance of the fragmentation methods depends on the glycated peptides. This could at least partly be caused by the extensive pattern of neutral loss which HCD fragmentation of glycated peptides can result in. This neutral loss pattern could interfere with the reporter ions providing the variations in the ratio calculations between the two methods.
The statistical analysis also revealed that the performance of the labels can depend on the glycated peptides. This was observed for the peptide with glycated K115 and the peptide containing K176 glycated with carboxymethyl where the 128/126 and the 130/127 ratios differed significantly. The glycated peptides influence on the performance of the labels could be caused by the neutral losses occurring in HCD described previously. It is also possible that the fragmentation of some peptides form peptide ions close to or with the same mass as the reporter ions hereby interfering with the intensity of the reporter ions resulting in too high or too low quantification ratios.
It is therefore necessary to used different label combinations for the quantification of glycated peptides to evaluate if the results are influenced by the glycated peptides. It has been proposed by others [31] to

used ETD for the identification of labile modifications and HCD for the quantification. Though as the glycated peptides can interfere with the quantification results it might be advisable to use both methods to determine if the results are reproducible by both.

A larger investigation of the influence of the glycated peptides using more TMT labels and the different fragmentation methods might provide more insight into how the glycated peptides affect the performance of the fragmentation methods and the tags. This could also provide better insight into how to plan quantification experiment with TMT for the quantification of glycated peptides.

## Conclusion

ETD outperforms HCD for the identification of glycation sites by LC-MS/MS as it identifies more glycated peptides and with higher sequence coverage. ETD is able to perform quantification of the glycated peptides with TMT labelling though the reporter ion intensities obtained by ETD are low so for many of the glycated peptides identified quantification cannot be performed. The reporter ion intensities produced by HCD are all high enough for quantification. Statistical analysis revealed that the glycated peptides can influence the quantification results. It is therefore important to perform the quantifications with different label combinations and also different fragmentation methods to determine if the results are correct.

## Acknowledgement

The authors would like to thank the Danish Agency for Science Technology and Innovation for funding the research. We would also like to thank Lene Bjorg Cesar for performing the statistical analysis.

**References**


1. Kirk, O., Borchert, T.V., Fuglsang, C.C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351 (2002)

2. Bruins, M.E., Van Hellemond, E.W., Janssen, A.E.M., Boom, R.M. Maillard reactions and increased enzyme inactivation during oligosaccharide synthesis by a hyperthermophilic glycosidase. *Biotechnol. Bioeng.* **81**, 546–552 (2003)

3. Li, S., Patapoff, T.W., Overcashier, D., Hsu, C., Nguyen, T.H., Borchardt, R.T. Effects of reducing sugars on the chemical stability of human relaxin in the lyophilized state. *J. Pharm. Sci.* **85**, 873–877 (1996)

4. Quan, C., Alcala, E., Petkovska, I., Matthews, D., Canova-Davis, E., Taticek, R., Ma, S. A study in glycation of a therapeutic recombinant humanized monoclonal antibody: Where it is, how it got there, and how it affects charge-based behavior. *Anal. Biochem.* **373**, 179–191 (2008)

5. Baynes, J.W., Watkins, N.G., Fisher, C.I., Hull, C.J., Patrick, J.S., Ahmed, M.U., Dunn, J.A., Thorpe, S.R. The Amadori product on protein: structure and reactions. *Prog. Clin. Biol. Res.* **304**, 43–67 (1989)

6. Brownlee, M., Cerami, A., Vlassara, H. Advanced Glycosylation End Products in Tissue and the Biochemical Basis of Diabetic Complications. *N. Engl. J. Med.* **318**, 1315–1321 (1988)

7. Ames, J.M. Mass spectrometry to detect the site specificity of advanced glycation/lipoxidation end-product formation on protein: some challenges and solutions. *Biochem. Soc. Trans.* **36**, 1051–1054 (2008)

8. Barnaby, O.S., Cerny, R.L., Clarke, W., Hage, D.S. Comparison of modification sites formed on human serum albumin at various stages of glycation. *Clin. Chim. Acta* **412**, 277–285 (2011)

9. Kislinger, T., Humeny, A., Pischetsrieder, M. Analysis of protein glycation products by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Curr. Med. Chem.* **11**, 2185–2193 (2004)

10. Kjeldsen, F., Giessing, A.M.B., Ingrell, C.R., Jensen, O.N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **79**, 9243–9252 (2007)

11. Lapolla, A., Fedele, D., Reitano, R., Aricò, N.C., Seraglia, R., Traldi, P., Marotta, E., Tonani, R. Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides. *J. Am. Soc. Mass Spectrom.* **15**, 496–509 (2004)

12. Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E.P., Shabanowitz, J., Hunt, D.F. The utility of ETD mass spectrometry in proteomic analysis. *BBA - Proteins Proteomics* **1764**, 1811–1822 (2006)

13. Priego-Capote, F., Sanchez, J.C. Strategies for proteomic analysis of non-enzymatically glycated proteins. *Mass Spectrom. Rev.* **28**, 135–146 (2009)

14. Chi, A., Huttenhower, C., Geer, L.Y., Coon, J.J., Syka, J.E.P., Bai, D.L., Shabanowitz, J., Burke, D.J., Troyanskaya, O.G., Hunt, D.F. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci.* **104**, 2193–2198 (2007)

15. Hogan, J.M., Pitteri, S.J., Chrisman, P.A., McLuckey, S.A. Complementary Structural Information from a Tryptic N-Linked Glycopeptide via Electron Transfer Ion/Ion Reactions and Collision-Induced Dissociation. *J. Proteome Res.* **4**, 628–632 (2005)

16. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.USA* **101**, 9528–9533 (2004)

17. Good, D.M., Wirtala, M., McAlister, G.C., Coon, J.J. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cel. Proteomics* **6**, 1942–1951 (2007)

18. Swaney, D.L., McAlister, G.C., Wirtala, M., Schwartz, J.C., Syka, J.E.P., Coon, J.J. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal. Chem.* **79**, 477–485 (2007)

19. Aebersold, R., Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003)

20. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007)

21. Ong, S.E., Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252–262 (2005)

22. Delatour, T., Hegele, J., Parisod, V., Richoz, J., Maurer, S., Steven, M., Buetler, T. Analysis of advanced glycation endproducts in dairy products by isotope dilution liquid chromatography-electrospray tandem mass spectrometry. *J. Chromatogr. A* **1216**, 2371–2381 (2009)

23. Priego-Capote, F., Scherl, A., Muller, M., Waridel, P., Lisacek, F., Sanchez, J.C. Glycation Isotopic Labeling with C-13-Reducing Sugars for Quantitative Analysis of Glycated Proteins in Human Plasma. *Mol. Cell. Proteomics* **9**, 579–592 (2010)

24. Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D.F., Burkhard, P.R., Sanchez, J.C. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Anal. Chem.* **80**, 2921–2931 (2008)

25. Ross, P.L., Huang, Y.L.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004)

26. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Hamon, C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003)

27. Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A., White, F.M. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240–1250 (2005)

28. Han, H., Pappin, D.J., Ross, P.L., McLuckey, S.A. Electron Transfer Dissociation of iTRAQ Labeled Peptide Ions. *J. Proteome Res.* **7**, 3643-3648 (2008)

29. Phanstiel, D., Unwin, R., McAlister, G.C., Coon, J.J. Peptide Quantification Using 8-Plex Isobaric Tags and Electron Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **81**, 1693–1698 (2009)

30. Phanstiel, D., Zhang, Y., Marto, J.A., Coon, J.J. Peptide and Protein Quantification Using iTRAQ with Electron Transfer Dissociation. *J. Am. Soc.Mass Spectrom.* **19**, 1255–1262 (2008)

31. Viner, R.I., Zhang, T., Second, T., Zabrouskov, V. Quantification of post-translationally modified peptides of bovine alpha-crystallin using tandem mass tags and electron transfer dissociation. *J. Proteomics* **72**, 874–885 (2009)

32. Svendsen, I., Hejgaard, J., Mundy, J. Complete amino acid sequence of the alpha-amylase/subtilisin inhibitor from barley. *Carlsberg Res. Comm.* **51**, 43–50 (1986)

33. Vallée, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K.W., Svensson, B., Haser, R. Barley alpha-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 A resolution. *Structure* **6**, 649–659 (1998)

## 7.4 Appendix IV – List of publications

***Manuscripts in preparation:***

C. Nielsen, S. G. Kaasgaard, C. Andersen, C. P. Sönksen, S. Jacobsen, B. Svensson. Characterisation of chemical modifications on *Bacillus licheniformis* α-amylase and the implications on detergent wash performance.

Connie Nielsen, Svend G. Kaasgaard, Susanne Jacobsen, Birte Svensson, Carsten P. Sönksen. Identification and Quantification of Non-Enzymatically Glycated Peptides by Electron Transfer Dissociation with Isobaric Tags.

***Poster presentations:***

9th Carbohydrate Bioengineering Meeting (CBM9), May 15-18 2011, Lisbon, Portugal.

4th Workshop in Protein.DTU on Technologies in Protein Science, November 12th 2010, Kgs. Lyngby, Denmark.

25th International Carbohydrate Symposium (ICS2010), August 1-6 2010, Tokyo, Japan.

11th Summer Course in Glycoscience, May 17-20 2010, Wageningen, Holland.

***Patent application:***

Co-inventor on patent application filed on June 30th 2011.

## 7.5 Appendix V – Poster

# Characterization of multiple forms of α-amylase created by post-translational modifications

**Connie Nielsen*,**, Svend G. Kaasgaard**, Carsten Sönksen**, Susanne Jacobsen* and Birte Svensson***

*Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark
**Novozymes, Bagsvaerd, Denmark

## Introduction

α-Amylases are among the most important industrial enzymes. They are used in a number of different industries such as the detergent, starch, biofuel and baking industry. *Bacillus licheniformis* α-amylase (BLA) is a thermostable enzyme which is stable between pH 6-11 and these qualities make it a very important industrial enzyme.
During the fermentation of BLA by the bacterium, multiple forms of the enzyme can be identified which differ in pI value. These forms are expressed from the same gene, so this bacterial amylase must undergo post-translational modifications during the fermentation.

## Summary

Kinetic studies on partially separated forms of BLA revealed only small differences in the catalytic efficiency for the pools A-C toward maltoheptaose whereas it decreases for pool D. On amylose the catalytic efficiency decreased from pool A to pool D indicating that the more modified the pool the lower the catalytic efficiency.

LC-MS on the pools indicated that the most abundant form is unmodified BLA. The other forms are a result of deamidations in combinations with carbamylations or acetylations and most likely also other modifications which were not identified in the complex samples.

## *Bacillus licheniformis* α-amylase

BLA consists of 483 amino acids (55.2 kDa) and it catalyses the hydrolysis of α-D-(1,4) glycosidic linkages in starch and related compounds. Starch is composed of the two polysaccharides, amylose and amylopectin. Amylose is basically linear consisting of mainly α-(1,4) bonds and a few α-(1,6) bonds. Amylopectin is branched with many more α-(1,6) bonds than amylose.

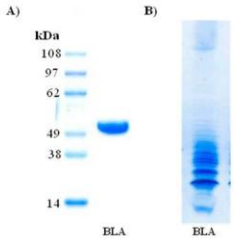



**Figure 1. A)** SDS-PAGE of multiple forms of BLA **B)** Isoelectric focusing gel of multiple forms of BLA.

During fermentation the α-amylase undergoes post-translational modifications. These modifications result in multiple forms with varying pI values. These forms can be visualised on an isoelectric focusing gel whereas on a SDS gel the α-amylases appear homogeneous (Figure 1).

## Separation of BLA forms

The separation of the different forms is difficult due to the high similarity between these. A partial separation was, however, achieved using anion exchange chromatography providing four pools A-D containing multiple forms but having different pI ranges (Figure 2).

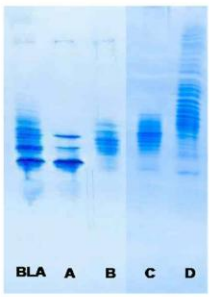



**Figure 2.** Isoelectric focusing gel with the multiple forms of BLA and pools A – D from anion exchange chromatography.

## Kinetics for multiple forms of BLA

The kinetics for the four pools were determined at 30°C and pH 7 on a maltoheptaose derivate (G7 substrate) and on amylose. On the G7 substrate the catalytic efficiency ($k_{cat}/K_M$) slowly declined from pool A to C with only small differences. The catalytic efficiency for Pool D is much lower than for the other pools (Table 1).

**Table 1.** Enzymatic activity for the four pools A-D toward maltoheptaose derivate determined at 10 different substrate concentrations (0.05-3 mM) at 30°C and pH 7.

| Pool | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ |
|---|---|---|---|
| A | 74 ± 7 | 0.011 ± 0.002 | 7117 |
| B | 92 ± 6 | 0.013 ± 0.002 | 6968 |
| C | 85 ± 4 | 0.014 ± 0.002 | 6336 |
| D | 68 ± 3 | 0.013 ± 0.002 | 5072 |

The tendencies observed for the catalytic efficiency on the short substrate were also observed on the amylose substrate though the decline in the catalytic efficiency were more pronounced for the long substrate (Table 2).

**Table 2.** Enzymatic activity for the four pools A-D toward amylose DP440. Determined at 10 different concentrations (0.1-2.5 mg/ml) of amylose at 30°C and pH 7.

| Pool | $k_{cat}$ ($s^{-1}$) | $K_M$ (mg/ml) | $k_{cat}/K_M$ |
|---|---|---|---|
| A | 301 ± 21 | 0.192 ± 0.050 | 1625 |
| B | 313 ± 10 | 0.208 ± 0.020 | 1515 |
| C | 279 ± 12 | 0.229 ± 0.060 | 1293 |
| D | 234 ± 20 | 0.213 ± 0.004 | 1101 |

Overall it appears that the different modifications occurring in the pools B-D have a negative effect on the catalytic efficiency especially on the long substrate. The modifications does, however, not seem to have much effect on $K_M$.

## Mass spectrometry

The pools were analysed by MS on a LC-ESI-TOF. In pool A unmodified BLA and BLA with two deamidations were observed. Pool B resulted in a broad peak corresponded to a mass increase of 46 Da compared to unmodified BLA. Hereby indicating the presence of various modification like carbamylation/acetylation combined with deamidations. Results from MS on pool C and D were of poor quality indicating very complex sample heterogeneity.

## Further work

The different forms of BLA are to be characterised by peptide mapping and tandem mass spectrometry to identify the modified amino acids and the modifications responsible for these changes. This should provide insight into the correlation between modification sites and the decrease in catalytic efficiency.

DTU Systems Biology
Department of Systems Biology
EPC Enzyme and Protein Chemistry

novozymes
Rethink Tomorrow